A Predictor Based Method for Signal Detection in Time Series Data, Case Study: Financial Markets of Iran in Corona Virus Pandemi.

Amir Hossein Ghatari* Seyed Reza Seyed Ali Mortezaei†

Abstract

In this paper, we study the detection of signals in time series data, particularly within the context of the financial markets. By analyzing historical data from auxiliary markets, such as foreign exchange rates, stock indices, and cryptocurrency prices, we aim to identify significant signals affecting the cash price of refined gold in the Iranian market. Our approach leverages various time series models, including ARIMA, SARIMA, and neural networks, as well as regression-based time series methods to predict fluctuations in gold prices. We focus on a period of 45 working days before and after the onset of the COVID-19 outbreak in Iran on February 23, 2020.

To achieve this, we propose a predictor-based algorithm for signal detection that utilizes both traditional time series and regression models. This algorithm identifies auxiliary markets that correlate with the target market, fits appropriate models to predict future values, and then determines cloud confidence intervals around these predictions. Observations that deviate significantly from these intervals are flagged as potential signals, suggesting unexpected changes or trends in the target market.

Our method not only enhances the ability to detect significant signals in financial markets but also provides a valuable tool for investors and analysts to anticipate and respond to market fluctuations, particularly during periods of economic instability, ultimately contributing to more informed decision-making and risk mitigation strategies.

Keywords: Cloud Interval, Signal Detection, Time Series Data.

1 Introduction

In the concept financial modeling, analyzing and predicting fluctuations in financial markets have always been of great interest to investors, professionals, and market participants. Fluctuations in financial markets, including currencies, gold, stocks, and even cryptocurrencies, are not only indicators of economic health but can also signal potential opportunities or threats in the future. Accurately forecasting these fluctuations can reduce risks for market agents and brokers, helping them make better decisions. This importance has led to the identification and analysis of patterns and signals in time series data becoming one of the essential tools in financial decision-making [9].

Statistical and machine learning models, particularly time series models, play a significant role in analyzing complex patterns and predicting financial market fluctuations. One of the main challenges in this field is understanding and identifying important signals within time series data. A signal is defined as a sudden or unexpected change in data that could indicate a new trend or serve as a warning for the future [15]. In this regard, time series models such as ARIMA and SARIMA, as well as neural networks, have become common tools for market analysis due to their forecasting capabilities and ability to learn complex patterns [5].

Various financial markets in each country, such as the gold, currency, and stock markets, can be influenced by different variables, including domestic and global economic indicators. For instance, changes in the dollar exchange rate, the Tehran Stock Exchange index, alongside global prices of Bitcoin and gold, directly or indirectly impact the price of gold in Iran. Numerous studies have explored the relationship between these market fluctuations and their interdependencies [12].

For implementing the proposed algorithm in this paper, various time series models, including ARIMA, SARIMA, and neural networks, are utilized. These models, by leveraging historical data, have the ability to identify hidden patterns and latent signals, which can lead to better decision-making in the gold market. Alongside these models, regression time series is used to analyze and explain the complex dependencies between variables [1], [13].

Previous studies have shown that analyzing fluctuations and relationships between financial variables can lead to a deeper understanding of market behavior and enable more optimal decisions. [3] presented machine learning techniques for time series regressions, emphasizing their application in now-casting, which is particularly useful for real-time economic analysis. They demonstrated

*Department of Statistics, Amirkabir University of Technology, Tehran, Iran, a.h.ghatari@aut.ac.ir

†Department of Statistics, Shahid Beheshti University, Tehran, Iran, rezamortezaei90@yahoo.com

how machine learning models could improve the accuracy of short-term predictions by efficiently capturing non-linear dependencies in the data. The detection of outliers in time series data remains a critical task, especially in the presence of anomalies that may signify unusual events or errors in data collection. [4] provided a comprehensive review of anomaly detection methods, highlighting various techniques ranging from statistical methods to machine learning algorithms, emphasizing their respective strengths and weaknesses. Similarly, [6] offered an earlier survey focusing on outlier detection for temporal data, detailing the evolution of techniques and the challenges associated with handling temporal structures, such as seasonality and trends.

In the concept of financial time series, the relationship between external factors and market volatility has been extensively studied. [7] investigated the influence of partisan conflict on U.S. stock market volatility using a quantile-on-quantile regression model. Their findings suggested that political uncertainty could be a predictor of changes in market volatility, providing insights into the integration of external socio-political factors in financial models. Another emerging field within time series analysis is the use of graph neural networks (GNNs) for tasks such as forecasting, classification, and anomaly detection. [10] surveyed the application of GNNs to time series data, emphasizing the advantages of leveraging graph structures to capture complex relationships between time series features. This approach has shown promise in improving predictive performance, especially when the data exhibits strong interdependencies.

In this paper, we aim to develop and implement an algorithm for signal detection within time series data, specifically focusing on identifying significant signals affecting the cash price of refined gold in the Iranian market. By integrating various time series models, including SARIMA, ARIMA, and neural networks, alongside regression-based time series models, we seek to analyze and predict fluctuations in gold prices. Our analysis will consider key auxiliary variables such as the USD to IR Rial exchange rate, the Tehran Stock Exchange index, and Bitcoin prices, to capture complex interactions and dependencies. The study focuses on a 45 working-day period before and after the onset of the COVID-19 outbreak in Iran on February 23, 2020. By applying our proposed algorithm to this case study, we aim to provide deeper insights into the behavior of gold prices during periods of market instability, helping to better understand the impact of external shocks and identify patterns that can aid in future market predictions.

2 Basic Concepts

As mentioned in Section 1, we want to detect signals in a target variable and based on the prediction method, we may use some features in the model structure. In this section, we recall the requirements of prediction for the study including time series, ARIMA models, Neural Network time series and etc.

2.1 Time Series

In general, a time series is a sequence of observations recorded over time. The key characteristics of time series can include mean functions, variance, autocovariance, and autocorrelation functions. A time series can be thought of as the values of a random variable X_t, observed or to be observed at different times. Formally, a time series can be defined as a sequence of random variables, $\{X_t, t \in T\}$, where T is a set of indices. In this context, T can be any subset of real numbers, but we assume it is a set of natural numbers. We also assume that the time series $\{X_t\}$ has finite first and second moments:

$$\mu_X(t) = E(X_t),\ t \in T$$

$$\sigma^2_X(t) = Var(X_t) = E[(X_t - \mu_X)^2],\ t \in T$$

$$C_X(t,s) = E[(X_t - \mu_X)(X_s - \mu_X)],\ t,s \in T$$

$$R_X(t,s) = \frac{C_X(t,s)}{\sigma_X \sigma_X},\ t,s \in T$$

Where the mean function, variance function, autocovariance function (ACVF), and autocorrelation function (ACF) describe the time series $\{X_t\}$. For a more comprehensive explanation, refer to [16].

2.2 AutoRegressive Integrated Moving Average Moldes

The **ARIMA** (AutoRegressive Integrated Moving Average) model is a popular statistical method used for analyzing and forecasting time series data. It combines three key components: autoregression (AR), differencing (I), and moving average (MA). Below, we provide the mathematical formulation of the ARIMA model. An ARIMA model is defined by three parameters: (p, d, q):

- p: Order of the **AutoRegressive (AR)** part. This refers to the number of lagged values of the series included in the model.
- d: Degree of **differencing**. This is the number of times the data is differenced to achieve stationarity.
- q: Order of the **Moving Average (MA)** part. This indicates the number of lagged forecast errors included in the model.

The general form of the ARIMA model can be expressed as:

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t$$

where:

- y_t: The observed time series value at time t.
- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the **AR** polynomial of order p.
- $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$ is the **MA** polynomial of order q.
- B is the **backshift operator**, defined by $By_t = y_{t-1}$.
- $(1-B)^d$ represents the **differencing** operator, applied d times to make the series stationary.
- ϵ_t is **white noise** (random error term), with a mean of zero and constant variance.

The ARIMA model combines all these components to create a flexible approach for modeling time series data that can accommodate both trends and noise. The parameters p, d, and q control how the model fits the data. Note that for $d = 0$, we have AutoRegressive-Moving Average (ARMA) model which is a stationary time series model. For more information, see [16] and [14].

2.3 SARIMA Models

The **SARIMA** (Seasonal AutoRegressive Integrated Moving Average) model is an extension of the **ARIMA** model that supports the modeling of seasonality in time series data. A SARIMA model extends ARIMA by including additional seasonal terms. The SARIMA model is defined by seven parameters: $(p, d, q) \times (P, D, Q)_s$, where:

- (p, d, q): Parameters for the **non-seasonal** part (as defined in the ARIMA model).
- (P, D, Q): Parameters for the **seasonal** part.
  - P: Order of the **seasonal autoregressive** part.
  - D: Degree of **seasonal differencing**.
  - Q: Order of the **seasonal moving average** part.
- s: Length of the **seasonal cycle** (e.g., $s = 12$ for monthly data with yearly seasonality).

The general equation for the SARIMA model is:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = \theta(B)\Theta(B^s)\epsilon_t$$

where:

- $\phi(B)$ and $\theta(B)$ are the **non-seasonal** AR and MA polynomials.
- $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$ is the **seasonal AR** polynomial.
- $\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}$ is the **seasonal MA** polynomial.
- $(1-B^s)^D$ applies **seasonal differencing** D times.

Different parts of SARIMA models:

- **Non-seasonal part**: Handles the general trend and patterns in the data.
- **Seasonal part**: Models the repeating seasonal patterns over time.
- **Seasonal differencing** $(1-B^s)$: Removes the seasonality by differencing the data at lag s.

SARIMA can capture both long-term trends (using non-seasonal components) and cyclical patterns that repeat at regular intervals (using seasonal components). For more information, see [16] and [14].

2.4 Time Series Regression

The simplest form of a linear regression model is as follows:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, \ t = 1, ..., n,$$

where Y is the response variable (the one we aim to predict), X is the explanatory variable (the one used for modeling), and ϵ_t is the noise in the model. If the noise follows a time series pattern, the response variable also becomes a time series, and the usual regression methods cannot be applied directly. The general form of a regression time series model is:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \ldots + \beta_K X_{Kt} + \epsilon_t, \ t = 1, ..., n$$

Or equivalently:

$$Y_t = \beta' X_t + \epsilon_t, \ t = 1, ..., n,$$

where $X_t = (X_{1t}, ..., X_{Kt})'$ and $\beta = (\beta_0, \beta_1, ..., \beta_K)'$. To convert a regular regression model into a time series regression model, we introduce lagged variables (e.g., past values of the market) based on statistical tests. The general procedure for fitting regression time series models involves:

- Identify the variables that show a linear correlation with the response variable using Pearson correlation tests.
- Fit a standard regression model as follows:

  $$\hat{Y}_t = \hat{\beta}' X_t, \ t = 1, ..., n,$$

  where $\hat{Y}_t$ represents the predicted values of Y_t. Calculate the residuals of the model using:

  $$e_t = Y_t - \hat{Y}_t, \ t = 1, ..., n.$$

  The residuals e_t are considered as estimates of the noise ε_t.

- Examine the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots of the residuals to determine whether time series characteristics have influenced the model.

- If time series effects are present, identify potential structures for the time series model related to the residuals based on the ACF and PACF plots. Modify the regression model according to these structures and reevaluate it.

- If the model is not validated, consider other time series structures (e.g., stability assessment, variance stabilization, etc.).

For more information about time series regression models refer to [13] and [11] and [16].

2.5 Neural Networks for Time Series Forecasting

Neural networks have become a powerful tool for modeling and forecasting time series data, especially when dealing with non-linear and complex patterns. Unlike traditional statistical models (like ARIMA), neural networks can learn from data without making strong assumptions about the underlying patterns. Here, we provide a brief overview of how neural networks are applied to time series forecasting.
A neural network is a computational model inspired by the way biological neural networks in the brain process information. It consists of interconnected layers of nodes (neurons) that transform inputs to outputs through weighted connections. In the context of time series, the neural network learns patterns from historical data to predict future values.

2.5.1 Structure of a Neural Network

A typical neural network used for time series forecasting consists of:

- **Input Layer**: Receives input values (e.g., past observations of the time series).

- **Hidden Layers**: Consists of one or more layers where neurons process the input data using weights and activation functions to capture complex patterns.

- **Output Layer**: Produces the final forecast (e.g., the predicted value of the next time step).

The network is trained by adjusting the weights through backpropagation to minimize the error between predicted and actual values.

2.5.2 Common Types of Neural Networks for Time Series

Feedforward Neural Network (FNN)
Feedforward neural networks are the simplest form of neural networks where data flows in one direction, from input to output. They can capture non-linear patterns but are limited in their ability to model temporal dependencies effectively.
Recurrent Neural Network (RNN)
Recurrent neural networks are designed specifically for sequential data, where each neuron's output can loop back as input to itself. This allows the network to retain information from previous time steps, making it suitable for time series forecasting:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where:

- h_t is the hidden state at time t.

- x_t is the input at time t.

- W_h and W_x are weight matrices, and b is a bias term.

- f is the activation function (e.g., tanh or ReLU).

Long Short-Term Memory (LSTM)
LSTM is a special type of RNN that overcomes the problem of long-term dependencies by introducing memory cells that can retain information for longer periods. LSTMs are effective for time series data where the dependence on past values is complex.

2.5.3 Training Process in Neural Networks

The training of neural networks for time series involves:

- **Loss Function**: Measures the difference between predicted and actual values (e.g., Mean Squared Error).

- **Backpropagation**: Adjusts the weights to minimize the loss.

- **Optimization Algorithms**: Techniques like gradient descent or Adam to update weights efficiently.

Neural networks, especially advanced architectures like RNNs and LSTMs, are capable of capturing complex and non-linear patterns in time series data. They provide flexibility in modeling, but require careful tuning and sufficient data for effective forecasting. To obtain general information about the application of neural networks in time series data analysis, see [2], [8] and [17].

3 Predictor Based Algorithm for Signal Detection in Time Series Data

3.1 Approach Expression

Based on the theoretical definition of a signal in statistics, it is essential to seek a rare occurrence that is not expected. In financial markets, it is necessary to determine whether a data point is truly a rare signal or merely a predictable market shock. The main objective of this research plan is to provide a procedure for identifying signals in domestic markets based on performance and signals issued by related foreign markets. Detecting a signal in one market using other influential markets requires the use of regression time series models.
Suppose that there is a target market and K auxiliary markets associated with this target market in the study. The goal is to detect signals issued by the target market. There are two approaches available:

1. Modeling the target market using time series models.

2. Modeling the target market using a regression time series model based on existing auxiliary variables.

In both approaches, predictions of the target variable are computed, and signal detection is performed. The procedure is as follows:

1. Collect sufficient data from all markets involved in the research. The data should follow a regular steps in the time

2. Based on exploratory data analysis, select $d < K$ foreign markets that are statistically confirmed to have a linear correlation with the domestic market and are suitable for signal analysis and modeling.

3. Fit the model to the available data.

4. In the case of a regression time series model, examine the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the residuals. If correlations are present, fit a suitable time series model (based on these plots) to the residuals and derive an adjusted regression model (regression time series).

5. Estimate the values for the target market.

6. By obtaining the distribution of estimated values, calculate the confidence interval cloud for each observation using the following equation:

$$\hat{Y}_t \in \{\hat{\mu}_{\hat{Y}_t} \pm t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\},$$

   where $\hat{\mu}_{\hat{Y}_t}$ represents the mean of the estimated values $\hat{Y}_t$, $\hat{\sigma}$ is the variance, and $t_{\alpha/2}$ is the quantile of the t-distribution used to compute the confidence interval.

7. Any data point from the collected sample for the target market that falls outside the calculated confidence interval cloud is considered a potential signal.

8. If the identified data point is also confirmed based on the nature of the market and the theoretical definition of a signal, it is considered a chosen signal, indicating that the target market may undergo future changes. The d selected auxiliary markets should remain under further research scrutiny for predicting future market behavior.

Note that the collected data in the first step must be temporally regular, with predetermined intervals (e.g., daily, weekly, or monthly) to ensure consistency and reliability in time series analysis. For this study, daily data was selected to capture short-term market fluctuations effectively. Also, the above algorithm is not dependent on a specific time period. Additionally, if no specific signal is identified in the present period, the algorithm can be rerun with updated data in the near future, potentially identifying new related foreign markets that were not initially recognized. Moreover, the data utilized in this study consists of daily observations about 45 days before and after February 23, 2020, the official date marking the onset of the COVID-19 pandemic in Iran.

3.2 Algorithmic Structure

In this subsection, we present the algorithm designed for signal detection within time series data, based on the descriptive explanations provided in subsection 3.1. Algorithm 1 outlines the step-by-step process and methodology to detecting signals within the data.

When a data point is identified as a potential signal by the algorithm, it undergoes expert evaluation to determine its significance. Confirmed signals are reported to planners and market analysts for actionable insights and strategic decision-making.

4 Numerical Analysis

The present research is a type of statistical modeling. The statistical population includes all discoverable data within the specified time period, and there is no need for sampling from the available data. Additionally, the data used in this research will be obtained from sources that publish information related to capital markets and economic indicators, which are legally authorized to operate online within the country. The two main sources of data for this research are the following websites:

- `https://www.tgju.org/`
- `https://www.bourseview.com/`

Algorithm 1 Procedure for Signal Detection in Time Series Data

Input: Data from target market and K auxiliary markets
Output: Detected signals for the target market
Collect sufficient data from all markets involved in the research.
if using a regression time series model **then**
    Select $d < K$ confirmed foreign markets having a linear correlation with the domestic market for signal modeling.
end if
Fit the model to the available data.
if using a regression time series model **then**
    Examine the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the residuals.
    if correlations exists **then**
        Fit a suitable time series model to the residuals and derive an adjusted regression model.
    end if
end if
Estimate the values for the target market $\hat{Y}_t$.
Obtain the distribution of $\hat{Y}_t$ and calculate the confidence interval cloud for each observation using:

$$\hat{Y}_t \pm t_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}},$$

Detect any data point from the target market that falls outside the confidence interval cloud as a potential signal.
if the identified data point is confirmed based on the nature of the market and the theoretical definition of a signal **then**
    Mark it as a chosen signal.
end if
Note: The algorithm is not dependent on a specific time period. If no specific signal is identified, rerun the algorithm with updated data.

In this section, we will first introduce the data and perform a preliminary exploratory data analysis (EDA) on it. Then, using the time series forecasting methods introduced in Section 2, we will proceed to obtain $\hat{y}_t$ for implementing the algorithm presented in Section 3 for signal detection.

4.1 Data Description and EDA

The outbreak of the COVID-19 pandemic was a process that began with its media coverage in the fall of 2019 in China and, by the end of winter of the same year, had spread worldwide, affecting the entire globe. It is evident that such an outbreak impacts the global economy, making the months before and after the spread of COVID-19 significant for examining signals in economic markets. For this reason, we use the collected data related to this period as a model to implement the signal detection algorithm introduced.

Since the outbreak of COVID-19 did not have an exact date, and different countries recorded different dates, and because the primary geography of this research is Iran and its markets, we have chosen February 23, 2020, as the starting date. The reason for this selection is the official announcement by the Ministry of Health and the temporary closure of institutions and educational facilities in the country on this date.

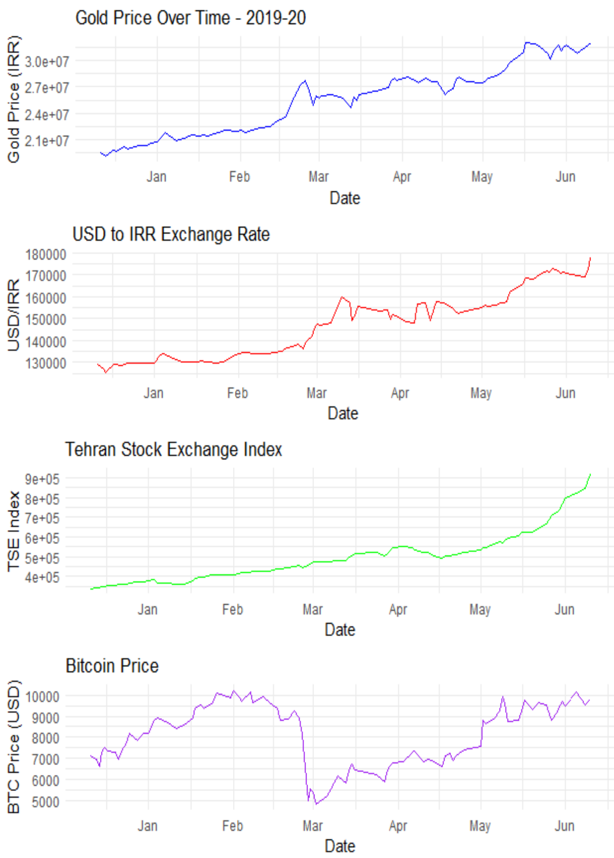

Figure 1: Time series plots of Gold price, USD to IRR exchange rate, Tehran Stock Exchange index, and Bitcoin price during the period starting from the outbreak of COVID-19 in Iran (February 23, 2020 onwards).

Figure 1 shows the time series plots of the price of gold in Iran, along with three other related markets: USD to IRR exchange rate, the Tehran Stock Exchange index, and the Bitcoin price. As observed, all three domestic markets exhibit an upward trend and a somewhat seasonal pattern. Additionally, the Bitcoin price displays two distinct seasons and an overall upward trend, with

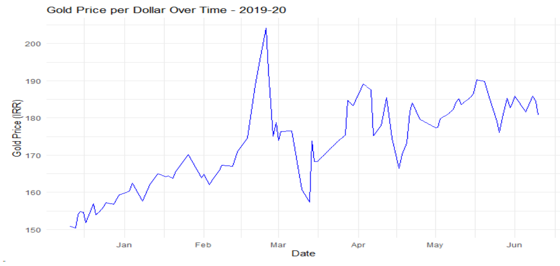

Figure 2: Time series plot of Gold price in the Iranian market (in USD) during the period starting from the outbreak of COVID-19 in Iran.

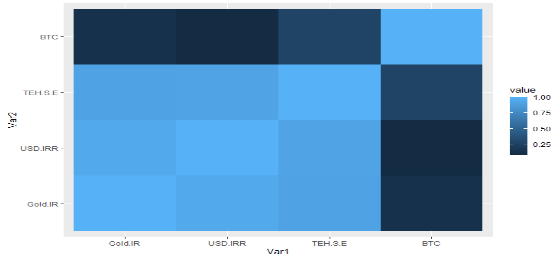

Figure 3: Heatmap of Correaltion Between the Four Markets in the Study. The Shaded Cells in the Table Represent the Level of Correlation: Darker Cells Indicate Weak correlation, while Lighter Cells Denote Strong Correlation.

a temporary drop around the global announcement of the COVID-19 pandemic.
To better identify the data trends, we use the Phillips-Perron unit root test, which is specifically designed to examine stationarity in the presence of drift and trend.
Table 1 presents the results of the Phillips-Perron test

Table 1: Results of Phillips-Perron Unit Root Test for Stationary of Time Series.

Test Type	Market	test statistics	p.value
Without drift and trend	GOLD.IR	0.47	0.79
	USD.IRR	0.33	0.76
	TEH.S.E	1.34	0.95
	BITCOIN	0.16	0.72
With drift and trend	GOLD.IR	-22.14	0.04
	USD.IRR	-18.54	0.08
	TEH.S.E	7.83	0.99
	BITCOIN	-6.21	0.71

for the four financial markets discussed in this paper. According to the results, no stationarity was observed in any of the cases. However, when the data are assumed to have drift and trend, stationarity is confirmed only for the gold price. This observation provides a logical basis for using the SARIMA time series model to model the target variable, which is the gold price.
An alternative approach, based on Figure 1 and the findings related to Table 1, is also presented. Instead of expressing the gold price in IRR, we can calculate it in USD within the Iranian market by dividing the daily gold price by the USD to IRR exchange rate for that day. Figure 2 shows the time series plot of this new variable, i.e., the gold price in USD in the Iranian market.
Based on the time series plot of this variable, and after performing a stationarity test, it can be concluded that an ARIMA model should be used for this variable to obtain forecast values. Here, after fitting the ARIMA model and calculating the forecasted values of the target variable, these values are multiplied by the USD to IRR exchange rate, thus providing a form of forecasted value for the cash gold price, which is the main target variable in this study.
The next idea we intend to explore in the model fitting section is the use of regression relationships among the data and fitting a regression time series model. Such problems require auxiliary variables alongside the target variable. In this study, the USD to IRR exchange rate, the Tehran Stock Exchange index, and the Bitcoin price are considered auxiliary variables for the regression time series analysis.

Table 2: Correlation Coefficients between the Four Markets in the Study.

	Gold.IR	USD.IRR	TEH.S.E	BTC
Gold.IR	1.00	0.95	0.91	0.13
USD.IRR	0.95	1.00	0.92	0.08
TEH.S.E	0.91	0.92	1.00	0.27
BTC	0.13	0.08	0.27	1.00

Table 2 and Figure 3 illustrate the degree of correlation between the gold price and other financial markets presented in this paper. Additionally, Table 3 shows the results of the correlation test between the three auxiliary variables and the target market. As observed, Bitcoin does not have a statistically significant direct correlation with the target market. However, the decision to exclude it from the model depends on its significance in the fitted regression model. On the other hand, the two domestic markets—stocks and USD—exhibit a significant and strong correlation with the target market, which was not unexpected. Based on the findings from

Table 3: Results of Correlation Test Between Gold price and other Markets.

Market	USD.IRR	TEH.S.E	BTC
P-value	0.00	0.00	0.21

the correlation analysis, the regression time series model to be examined in the following subsection can be represented as:

$$\text{Gold.IR} \sim \beta_0 + \beta_1 \cdot \text{USD.IRR} + \beta_2 \cdot \text{TEH.S.E} + \beta_3 \cdot \text{BTC} + \varepsilon_t$$

In addition, we also utilize neural networks to forecast the target market, i.e., the gold price. These four approaches will be evaluated from the perspective of signal detection.

4.2 Model Fitting and Signal Detection

In subsection 4.1, we examined the data through exploratory data analysis (EDA) and discussed the rationale behind using different methods to forecast the target market, i.e., the gold price in Iran. In this subsection, we proceed to calculate the values of $\hat{Y}_t$ for the various methods and implement the signal detection algorithm for each.
Now, we proceed with fitting the models and calculating the values of $\hat{Y}_t$ using the methods of FNN, the SARIMA model, and the time series regression model for gold prices, as well as the ARIMA model for gold prices in USD. For the ARIMA model, after fitting, the predicted values are multiplied by the current exchange rate of USD to obtain the actual price.
For the FNN, the data related to gold prices along with the lags for the previous three days in the market are defined as inputs, and the number of hidden layers is set to 5. The selected SARIMA model, based on appropriate functions in R software, is considered as $SARIMA(1,1,1) \times (1,1,1)_4$. The $ARIMA(0,1,2)$ model is fitted to the gold prices in USD.
For the regression model, the residuals follow an autoregressive model of order one (based on the 'auto.arima' command in R software). Therefore, the modified model is fitted by applying a one-day lag to the data. The results for $\hat{Y}_t$ obtained from the four methods mentioned above are presented in Figure 4. Figure 5 shows the time series curve and the charts of cloud confidence intervals obtained from FNN, SARIMA Model, ARIMA Model, and Time Series Regression. The red dots indicate data points that fall outside the cloud intervals and are potential signals, which should be analyzed further to understand their nature. The vertical dashed line marks February 23, 2020, the day when Iran's Ministry of Health, Treatment, and Medical Education officially announced the beginning of the COVID-19 pandemic in the country.
Based on the observations from Figure 5, it can be stated that all four methods detected certain data points as signals in mid-February. The SARIMA model detected the highest number of potential signals. This does not necessarily indicate the superiority of the method; rather, it can be seen as a less conservative approach. This is because a confidence interval is equivalent to a hypothesis test, and if a data point falls out-

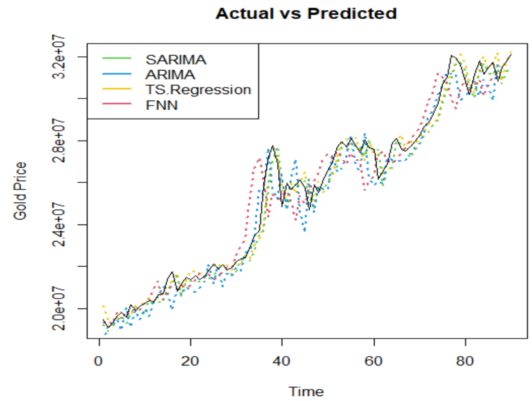

Figure 4: Comparison of Gold Prices and the Obtained Predicted Gold Prices Using FNN, SARIMA, ARIMA, and Time Series Regression Models

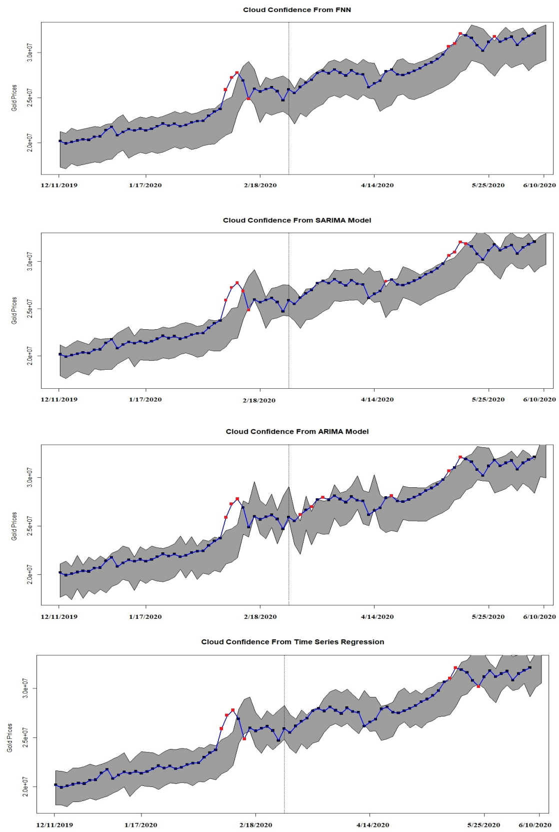

Figure 5: The Time series curve of Gold Price with cloud confidence intervals obtained from FNN, SARIMA Model, ARIMA Model, and Time Series Regression, Respectively. The red dots indicate data points that fall outside the cloud intervals and are potentially signals. The vertical dashed line marks February 23, 2020.

side the interval, it effectively confirms the hypothesis of being a signal. This means SARIMA is less strict in identifying a data point as a signal.
However, an important point pertains to the other model, ARIMA. While all methods detected a local minimum before the official announcement of the COVID-19 pandemic as a signal, ARIMA did not. It appears that this method performed better in predicting this local minimum. All four methods under study detected at least seven data points as potential signals. One form of evidence that a detected point is indeed a signal is when it is identified as such by all methods. In such cases, it indicates that all time series modeling approaches agree that the data point deviates from the expected trend. In this study, three data points before the onset of the pandemic, in mid-February 2020, were commonly identified by all four methods. Additionally, the data point related to May 20, 2020, was also commonly detected as a signal by all four methods.
Although other data points were detected around this date, only this one was consistently marked by all four methods. Thus, it can be concluded that the three data points detected in mid-February 2020, along with the single data point on May 20, 2020, can be considered as signals in this dataset (i.e., the cash price of gold in Iran). The focus has been on detecting signals. However, the question arises: can each statistically confirmed signal also be considered an indicator of changes in that financial market? The answer to this question falls within the realm of future studies, but we will briefly discuss the significance of the data identified as signals.
If we look at the trend of data following the three signals identified in February, it is understandable that the world was, at that time, engulfed in rumors and concerns related to the COVID-19 pandemic. In such a situation, and from a non-expert perspective, these fluctuations are not surprising. This brings us to the concept of weak signals, as proposed by Ansoff in 1975—rare data points that deviate from the trend but may indicate significant changes in the future. As can be observed, the three identified signals are essentially relative maxima in the data for that period. However, after these three data points, the market experienced a brief fluctuation followed by a period of stagnation, during which even the official announcement of the COVID-19 pandemic did not lead to any significant change in the market. But following this temporary stagnation, the gold market (along with other financial indices) began a general upward trend, which continued for nearly six months. The same is true for the data point identified on May 20, 2020; however, since the scope of potential changes following this signal lies outside the domain of our research, we will refrain from delving into it.
The most important application of such signals, which are essentially detected as outliers and can indicate future changes, is in the domain of market safety and for economic experts. These signals can help them stay alert to major shifts or even potential downturns, enabling them to manage financial markets and related factors with scientific evidence (information obtained from the signals) before substantial changes occur.

5 Discussion

This study presents a systematic approach to detecting signals in financial markets, focusing on identifying rare occurrences that deviate from typical trends. Utilizing a combination of neural networks, SARIMA, ARIMA, and regression-based time series models, the algorithm was developed to process and analyze gold price data in Iran, particularly during the onset of the COVID-19 pandemic. Each model was employed to predict values and identify data points falling outside established confidence intervals, marking them as potential signals. The results indicate that all four methods were effective in identifying specific signals before significant market events. The use of SARIMA, ARIMA, and regression models provided varying degrees of sensitivity, with SARIMA showing a tendency to identify more signals due to its less conservative confidence intervals. On the other hand, the ARIMA model, which adjusted gold prices based on the USD exchange rate, demonstrated better performance in predicting specific market movements, particularly during the onset of the pandemic.
A key finding was the detection of three significant signals in mid-February 2020, coinciding with the early rumors and concerns about COVID-19. These were identified by all models, highlighting their potential as robust indicators of future market behavior. Similarly, a signal detected on May 20, 2020, was consistently marked across models, suggesting its reliability as a forecast of subsequent market trends. The analysis suggests that the identified signals could be early indicators of future shifts, supporting the notion of "weak signals".
The methodology and algorithm developed in this study provide a structured and adaptable framework for detecting signals in financial markets. With the ability to incorporate regular updates with new data, the algorithm can refine its predictions and identify emerging trends, offering a dynamic tool for analyzing market behavior. This capability holds significant value for market analysts and economists, enabling them to manage financial markets proactively through data-driven insights. Future research could expand on these findings by exploring the relationship between identified signals and subsequent market behavior, using time series regression with neural networks, and also more validating the practical utility of the algorithm in simulation and real-world datasets.

References

[1] S. Aliahmadi and M. Ahmadlou. Forecast the gold coin future contracts prices by arima models in iran mercantile exchange (ime). *Financial Knowledge of Securities Analysis*, 4(9):61–74, 2011.

[2] E. M. Azoff. *Neural network time series forecasting of financial markets*. John Wiley & Sons, Inc., 1994.

[3] A. Babii, E. Ghysels, and J. Striaukas. Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106, 2022.

[4] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano. A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR)*, 54(3):1–33, 2021.

[5] M. Dehghani, M. Mohammad, and H. Ansarisamani. Machine learning algorithms for time series in financial markets. *Journal of Soft Computing and Information Technology*, 8(3):60–67, 2019.

[6] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering*, 26(9):2250–2267, 2013.

[7] R. Gupta, C. Pierdzioch, R. Selmi, and M. E. Wohar. Does partisan conflict predict a reduction in us stock market (realized) volatility? evidence from a quantile-on-quantile regression models. *The North American Journal of Economics and Finance*, 43:87–96, 2018.

[8] T. Hill, M. O'Connor, and W. Remus. Neural network models for time series forecasts. *Management science*, 42(7):1082–1092, 1996.

[9] M. Holopainen and M. Toivonen. Weak signals: Ansoff today. *Futures*, 44(3):198–205, 2012.

[10] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20, 2024.

[11] B. Kedem and K. Fokianos. *Regression models for time series analysis*. John Wiley & Sons, 2005.

[12] A. Mohammadi and H. Pashootanizadeh. Scenario planning the effect of foreign gold price and exchange rate on financial market with using system dynamics approach. *Financial Management Perspective*, 7(19):27–50, 2017.

[13] C. W. Ostrom. *Time series analysis: Regression techniques*. Number 9. Sage, 1990.

[14] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer, 2011.

[15] L. Su, L. Deng, W. Zhu, and S. Zhao. Statistical detection of weak pulse signal under chaotic noise based on elman neural network. *Wireless Communications and Mobile Computing*, 2020(1):9653586, 2020.

[16] R. S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, 2010.

[17] G. P. Zhang and M. Qi. Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160(2):501–514, 2005.