Gold Price Prediction Using ARIMA

Nurul Ariessa Binti Norramli

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak

Abstract

Autoregressive integrated moving average (ARIMA) is a widely known statistical method for time series forecasting. As such, I used ARIMA model to perform gold price prediction on a time series dataset. ARIMA is a type of stochastic time series model that takes error values and lag terms in calculation. It functions by relying on regressing a variable on past values.

Introduction

When it comes to predicting gold price, having a time series dataset is a must. Time series datasets are those that requires a prediction in numerical or categorical value. Even so, the rows of data are indexed by time. In this project, the rows of data were organized by date and the type of dataset used is a multivariate time series dataset. This is because there are two features (columns that affect prediction target):

- 1 Day Moving Average
- 2 Days Moving Average

Model Preparation

The first step of model preparation is preparing dataset. I gathered data of gold prices in USD for the year 1978 - 2019 from World Gold Council. The data was taken up until 5th of July 2019. The second step involves loading the data into a Pandas DataFrame and setting the column date as index. Then, I proceed by cleaning the dataset. I dropped rows with missing values such as NaN, which stands for Not A Number. Figure 1 shows the result of using Pandas head() method, which returns the top 5 rows of a DataFrame.

Date	Price	USD / MYR Price
12/29/1978	226.0	NaN
1/1/1979	226.0	NaN
1/2/1979	226.8	NaN
1/3/1979	218.6	NaN
1/4/1979	223.2	NaN

Date	Price	USD / MYR Price
12/26/1979	473.1	2.1790
12/27/1979	473.1	2.1665
12/28/1979	512.0	2.1620
12/31/1979	512.0	2.1880
1/2/1980	559.5	2.1770

Figure 1. Before and after dropping rows with missing values

Features

As mentioned before, features are columns that affect prediction target (column to predict). The features were created based on the moving average of the gold price. After creating features, I cleaned the dataset again by dropping rows with missing values. Then, I assembled the features into variable X.

```
X = ['1_Day_Moving_Average', '2_Days_Moving_Average']
```

Date	1_Day_Moving_Average	2_Days_Moving_Average
12/28/1979	473.1	473.10
12/31/1979	512.0	492.55
1/2/1980	512.0	512.00
1/3/1980	559.5	535.75
1/4/1980	634.0	596.75

Figure 2. Features

Check Stationarity

In order to use ARIMA model, the stationarity of the time series need to be checked. A time series is stationary if it has constant mean and constant variance over time. There are two major reasons on why a time series might not be stationary: trend, and seasonality. Those two elements need to be removed in a time series so that it will be easier to model. Aside from that, some models assume or require time series to be stationary. In this project, I checked for stationarity by using a statistical test called Augmented Dickey-Fuller Test.

```
ADF Statistic: -0.048233341976880784
p-value: 0.9543567594525035
Critical Values:
	1%: -3.431
	5%: -2.862
	10%: -2.567
```

Figure 3. Result of Augmented Dickey-Fuller Test on time series

Based on figure 3, it can be observed that the ADF statistic is bigger than the 5% critical values. Therefore, it can be said with 95% confidence that the data is not stationary.

Differentiate

In order to achieve stationarity, I differentiated the time series to remove trends and seasonality. Then, I checked for stationarity again using Augmented Dickey-Fuller test.

```
ADF Statistic: -11.905312067244587
p-value: 5.45372371491237e-22
Critical Values:
	1%: -3.431
	5%: -2.862
	10%: -2.567
```

Figure 4. Result of Augmented Dickey-Fuller test on differenced time series

Based on figure 4, it can be observed that the ADF statistic is smaller than 5% critical values. Thus, it can be said with 95% confidence that the time series is stationary.

Filter Out Validation Sample

I used Train-Test-Split to split the data into train and test set. The training data is used to train the model and the testing data used for evaluating the model. The data was split into 60:40 ratio, with training data being the larger one in ratio.

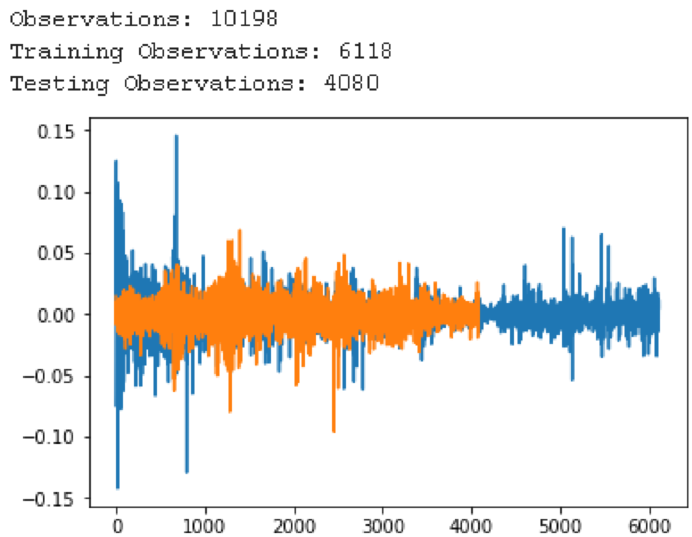

Figure 5. Result of Train-Test-Split

Select AR and MA Terms

An ARIMA model is made up of AR (Auto-Regressive), I (Integrated), and MA (Moving Average) components. I plotted Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) to find the value of AR and MA.

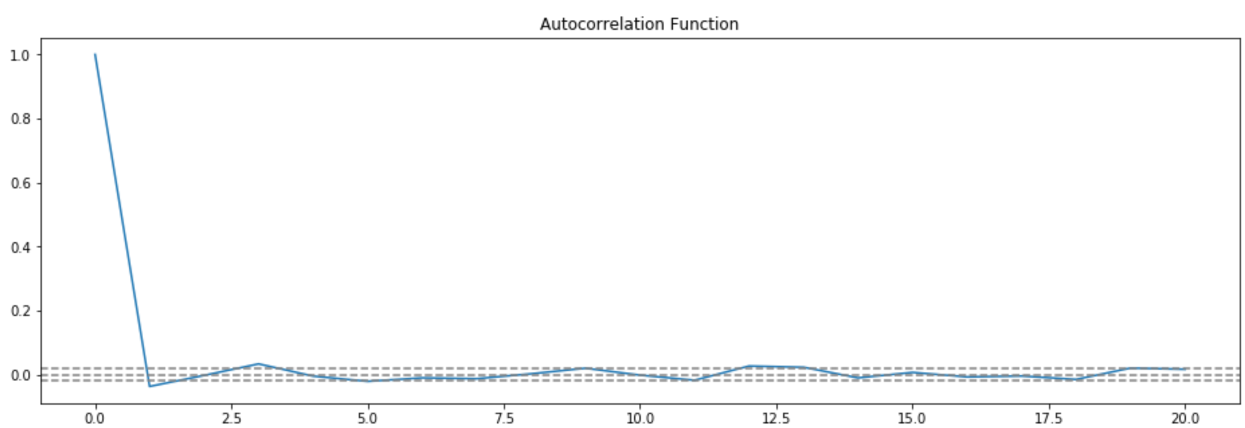

Figure 6. ACF Plot

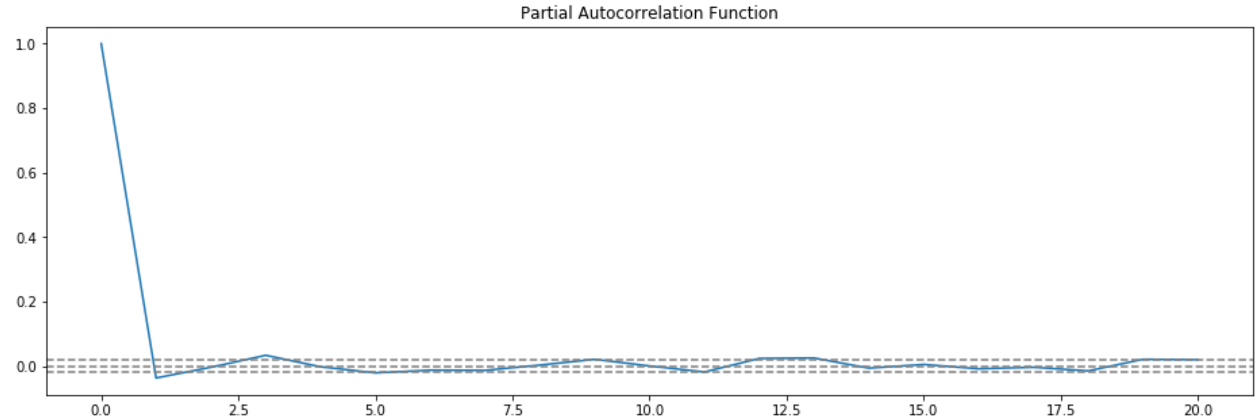

Figure 7. PACF Plot

Based on the ACF and PACF plots, it can observe that the value where the both charts crossed the upper confidence interval for the first time is 1. Therefore, the value for AR and MA is 1.

Build and Forecast

The best ARIMA model was built using the following order (2, 0, 2). Then, the model is fitted and validated with the testing data. After validating, the fitted model was used to make gold price prediction for the next 20 days. Note that some of the actual prices for certain dates were not available. The accuracy of the model was measured using mean absolute error (MAE) and was recorded to be at 22.230. MAE refers to the average of prediction errors.

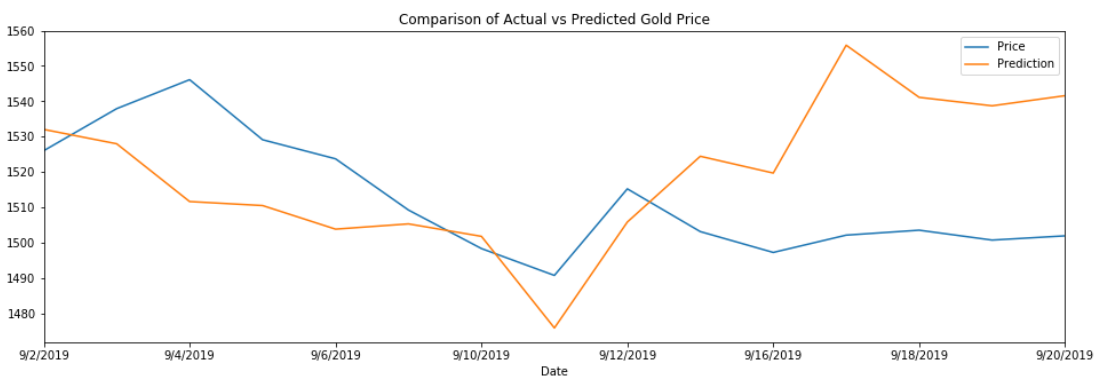

Figure 8. Comparison of Actual vs Predicted Gold Price

Date	Price	Prediction
9/2/2019	1526.0	1532.022897
9/3/2019	1537.9	1527.958214
9/4/2019	1546.1	1511.600948
9/5/2019	1529.1	1510.472832
9/6/2019	1523.7	1503.793872
9/9/2019	1509.2	1505.282388
9/10/2019	1498.3	1501.755070
9/11/2019	1490.7	1475.827025
9/12/2019	1515.2	1505.811838
9/13/2019	1503.1	1524.421624
9/16/2019	1497.2	1519.667093
9/17/2019	1502.1	1555.857606
9/18/2019	1503.5	1541.101239
9/19/2019	1500.7	1538.704842
9/20/2019	1501.9	1541.574108

References

[1] Askari, M., & Askari, H. (2011). Time series grey system prediction-based models: Gold price forecasting. *Trends in Applied Sciences Research*, *6*(11), 1287-1292.

[2] Guha, B., & Bandyopadhyay, G. (2016). Gold price forecasting using ARIMA model. *Journal of Advanced Management Science Vol*, *4*(2).

[3] Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. *Resources policy*, *35*(3), 178-189.