# Forecasting Gold Monthly Returns using ARIMA

Finnegan Nguyen

20 May 2019


## Contents

## 1. Introduction

### 1.1. Gold and Question of Interest

In the past 10 years, the risks and instabilities that have occurred all over the financial world have made gold a more essential investment than ever before. However, gold itself carries in it many complicated and unpredictable movements. Since then, the need to grasp developments and income from gold has become extremely important (Azzutti, 2016).

The aim of this report is to predict the monthly return of gold in 12 months of 2018 base on the LBMA's monthly gold prices from January 2002 to December 2018 by choosing an adequate ARIMA model.

Various researchers have forecasted gold using the ARIMA method. Some of the studies are: Abdullah (2012), Guha and Bandyopadhyay (2016), Yang (2019).

### 1.2. The Data

The data used in this report relates to the prices of gold. The price of gold is decided through trading in gold and also its derivatives markets. The London Gold Fix which establishes in September 1919 listed a daily benchmark price for gold. The London Bullion Market Association (LBMA) Gold Price which sprang from March 2015 to replace the London Gold Fix (London, 2019). The price is set twice daily, morning and evening, in US dollars. The monthly evening close price from January 2002 to December 2018 (204 observed prices) is used for the analysis. The data is obtained from www.quandl.com.


Figure 1. Monthly price per one troy ounce of Gold in USD from January 2002 to December 2018.

Figure 1 shows that the price of gold is $282.30 dollars/Oz in 2002, then the price starts to increase faster with a peak on 2011. During the last 7 years, gold price dropped of more than $310 dollars/Oz. The time series of monthly price of gold is not stationary.

The ACF plot in Figure 2 shows there are significant autocorrelations with many lags in the time series. In Figure 3, the PACF plot only shows a spike at lag 1. By viewing the ACF and PACF, there is no evidence for seasonality.

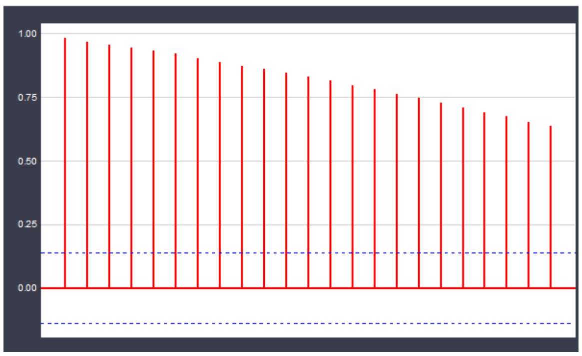



Figure 2. ACF of monthly price of gold.

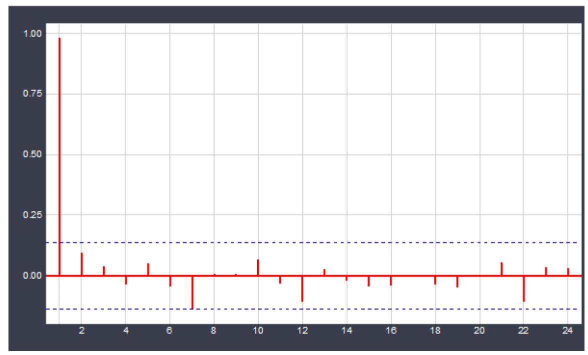



Figure 3. PACF of monthly price of gold.

### 1.3. Outliers

Outliers are atypical observations that become apparent because of measurement errors or short-term changes in the underlying process (Cryer, 2011). Figure 4 displays outliers in 2002, 2008 and 2009. In 2008, the financial crisis had large impact on the world's economy, however, it set stage for the gold to rise for the next few years. The outliers are not considered extreme; they can be replaced by potential replacements values which are suggested by a special function that detect outliers in R.

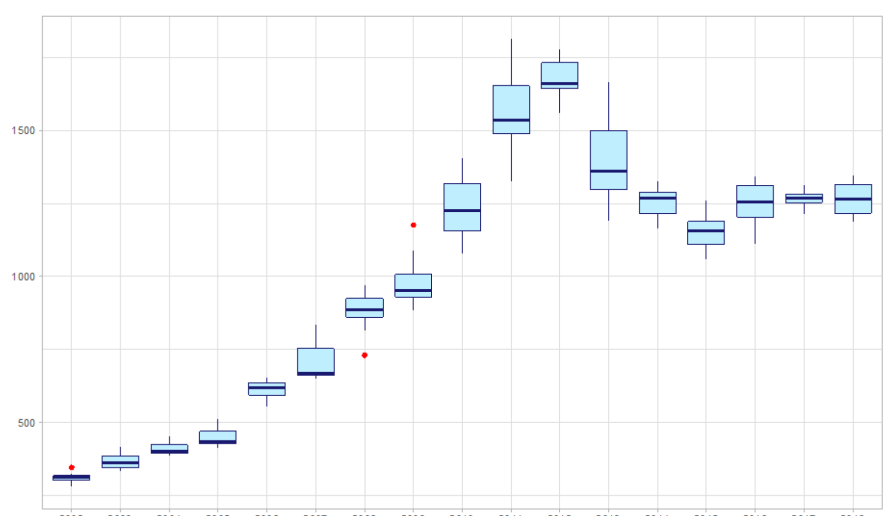


Figure 4. Box plot of monthly gold price with outliers (red dot) from January to December 2018.

## 1.4 Monthly Return of Gold

The interest of the report is the (continuously compounded) return on the *t*th month of gold is defined as

$$r_t = log(p_t) - log(p_{t-1}) \tag{1}$$

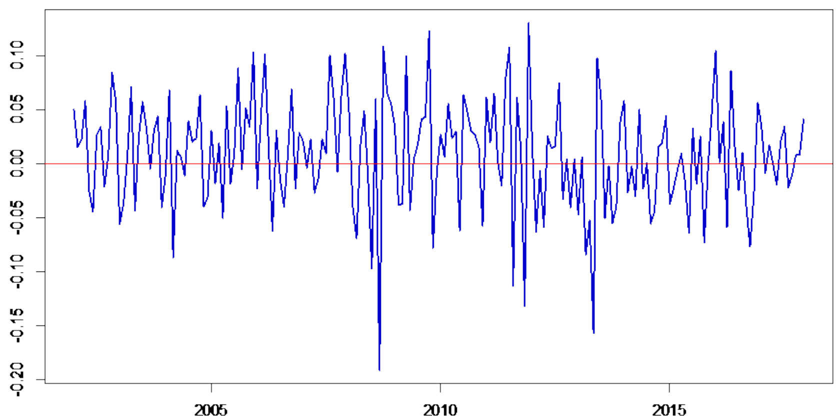
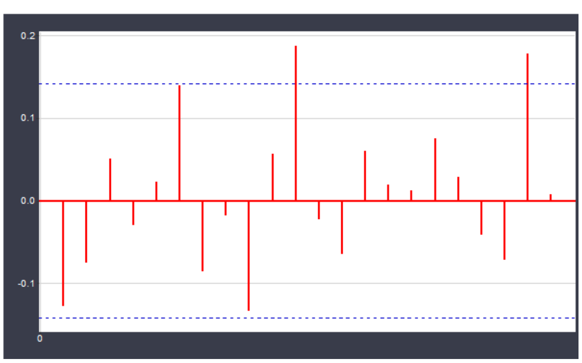



Figure 5. Monthly return of gold from January 2002 to December 2017 (Training Sample).

**Training Sample:** It consists of 192 time observations corresponding to 192 monthly data points from January 2002 to December 2017.
**Testing Sample:** It consists of 12 time observations corresponding to 12 monthly data points from January 2018 to December 2018.

By taking the first different of the logarithm of the data, the mean and variance are constant over time. Thus, the time series in Figure 5 is considered stationary and the Augmented Dickey-Fuller Test also confirm the stationarity with the p-value less than 0.05.



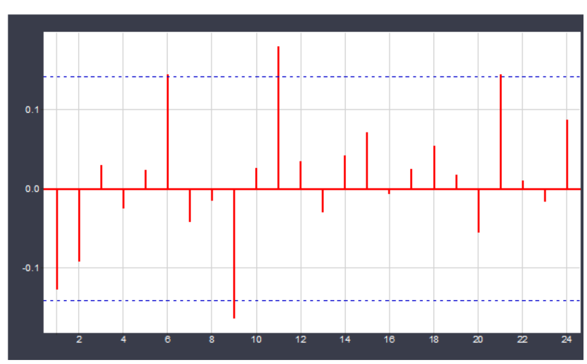



Figure 6. ACF plot of gold monthly return.

Figure 7. PACF plot of gold monthly return

## 2. Methods

### 2.1. Model Specification

The ACF in Figure 6 shows significant autocorrelation at lags 11 and 21. The PACF in Figure 7 shows significant autocorrelation at lags 9 and 11.

The plot of lag 1 in Figure 8 reflects a weak negative autocorrelation. The lack of autocorrelation at lag 2 is apparent from the scatterplot in Figure 9.

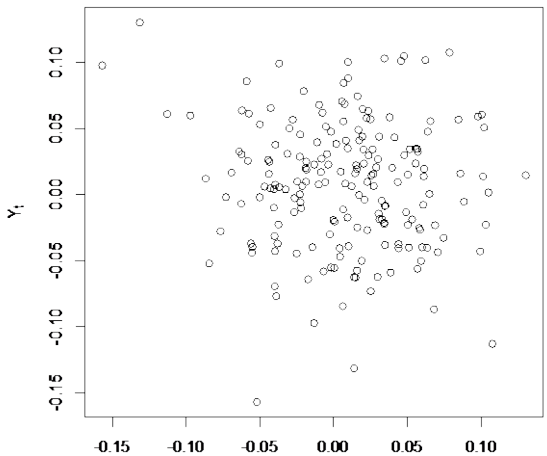



Figure 8. Plot of $Y_t$ - $Y_{t-1}$ for gold monthly return.

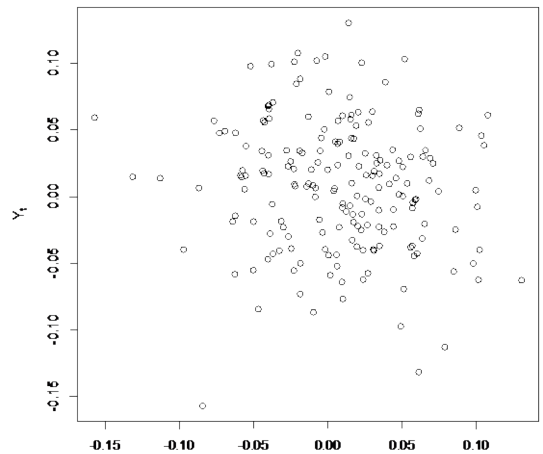



Figure 9. Plot of $Y_t$ - $Y_{t-2}$ for gold monthly return

According to suggestions of the EACF in Figure 10, the algorithm is built to try sets of different coefficient values. Based on comparing the Alkaike Information Criteria (AIC), the best model to forecast monthly gold return is ARIMA (0,0,2) with the lowest AIC. The auto-arima function suggests ARIMA (1,1,3) but the AIC of the model is higher than other models.

| ARIMA (p,d,q) | AIC |
|---|---|
| 0,0,1 | -617.48 |
| 0,0,2 | -617.59 |
| 1,0,1 | -616.14 |
| 1,0,3 | -616.51 |
| 1,1,3 | -609.53 |
| 1,0,2 | -616.56 |

Table 1. Comparison of 6 models have the best AIC.

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o o o x  o  o  o
1 x o o o o o o o o o x  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x o o x o o o o o o o  o  o  o
5 x x x x x o o o o o o  o  o  o
6 x x x x x o o o o o o  o  o  o
7 x o x x o x o o o o o  o  o  o
```

Figure 10. EACF of gold monthly return.

## 2.2 Model Fitting

The estimated parameters of the ARIMA (0,0,2) reflect the equation for MA(2) process

$$Y_t = e_t + 0.0784e_{t-1} + 0.1041e_{t-2} \tag{2}$$

ARIMA (0,0,2) with non-zero mean

Coefficients:

| | ma1 | ma2 | mean |
|---|---|---|---|
| | -0.0784 | -0.1041 | 0.081 |
| s.e. | 0.0738 | 0.0707 | 0.0028 |

sigma^2 estimated as 0.002287: log likelihood = 312.8, AIC = -617.59, BIC = -604.56

Maximum likelihood estimates: $\theta_1 = -0.0784, \theta_2 = -0.1041 \; and \; \mu = 0.081$

## 2.4 Model Diagnostics

The Model Diagnostics is concerned with testing the goodness of fit of a model which describes how well it fits a set of observations. Measures of goodness of fit is a summary of the discrepancy between observed values and the expected value under the model (Cryer, 2011). In this report, the analysis of residuals need to examine the Residual plots, Density histogram of residuals, Normal plot of residuals, Quantile-Quantile plot of residuals, ACF of residuals, Ljung-Box Test and Shapiro-Wilk normality test.

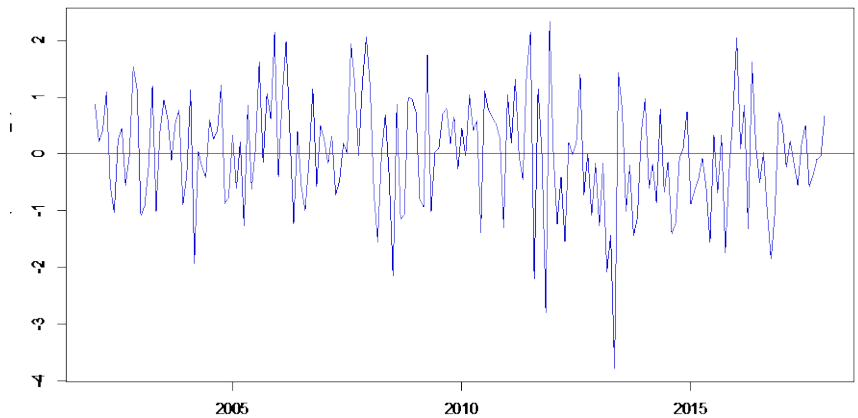



Figure 11. Plot of the standardized residuals (blue) from a fitted ARIMA (0,0,2) and a zero horizontal level (red).

The standardized residuals in Figure 11 looks random with no trend whatsoever. The parameters were estimated using maximum likelihood. There are one residuals with magnitude larger than 3 - unusual in a standard normal distribution. The Bonferroni critical values with n = 192 and $\alpha$ = 0.05 are $\pm$3.65, so the large negative standardized residual is not outlier. Also, 95% of the residuals lie between -1.96 and +1.96 and nearly all of them lie between -3.29 and +3.29.

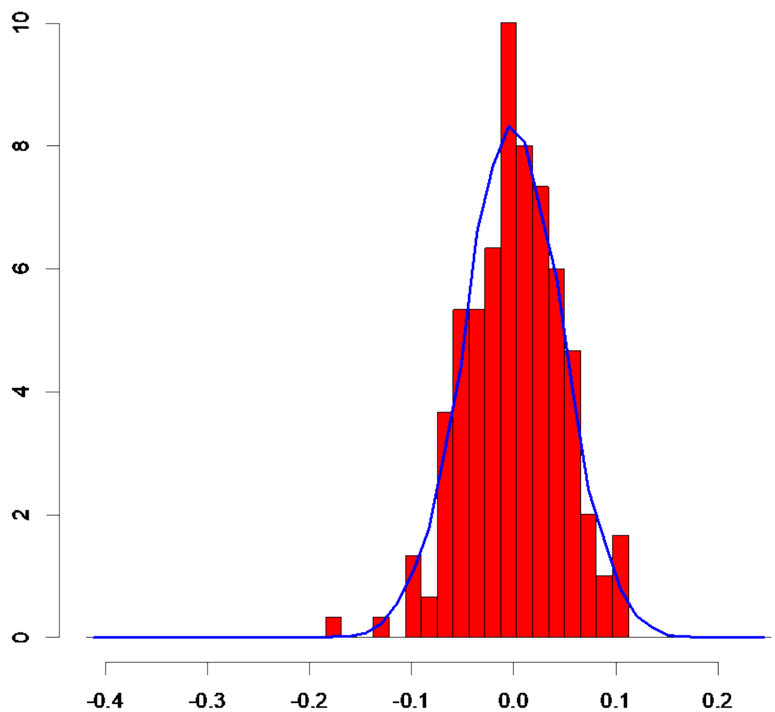



Figure 12. Density histogram of the residuals (red) and the density curve of the normal distribution with the same mean and variance (blue line).

Figure 12 reflects the residuals are normally distributed. The Quantile-Quantile plot of the residuals in Figure 13 is approximately linear supporting the condition that the errors terms are normally distributed. The Shapiro-Wilk normality test returns W = 0.9909 and p-value = 0.2672 > 0.05 implying that the distribution of the residuals are not significantly different from normal distribution. Finally, we can assume the normality of the residuals.

The ACF in Figure 14 displays the residuals look like white noise with other higher lags fall within the 95% confidence interval. The p-value in Figure 15 all above 0.05 indicates that there is no evidence to against the null of no autocorrelation in the errors of the Ljung-Box Test. Therefore, ARIMA(0,0,2) is an adequate model.

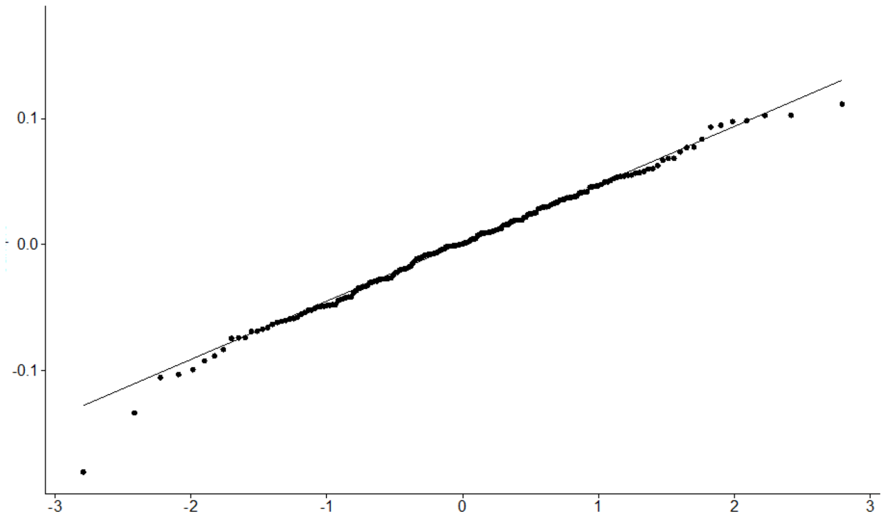



Figure 13. Quantile-Quantile plot: Residuals from ARIMA (0,0,2) Model with a 45-degree reference line.

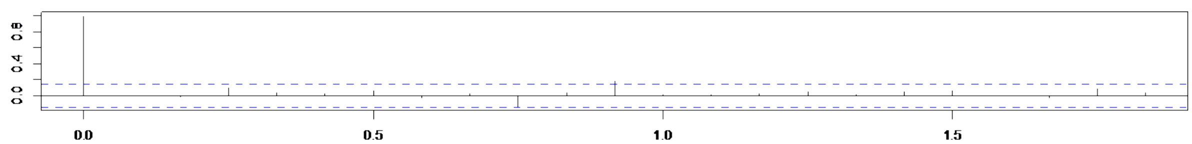



Figure 14. ACF plot of residuals from a fitted ARIMA (0,0,2) model.

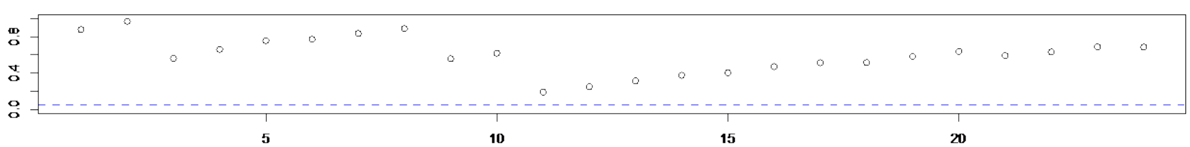



Figure 15. P-value for Ljung-Box statistic.

## 3. Results

According to the methods above, ARIMA (0,0,2) which is an MA(2) model:

$$Y_t = e_t + 0.0784e_{t-1} + 0.1041e_{t-2}$$

is the best fitting model for forecasting gold monthly returns.

Looking at Figure 16 and the values on Table 2, the actual values all fall within 95% forecast limits. The forecast range of 12 months from January 2018 to December

2018 indicate that the monthly return of gold is set to increasing in January and then decreasing in Febuary, follow by a constant value in the next 10 months. The actual values also reflect a rise in January then drop until May, the returns increase again in the next 7 months.

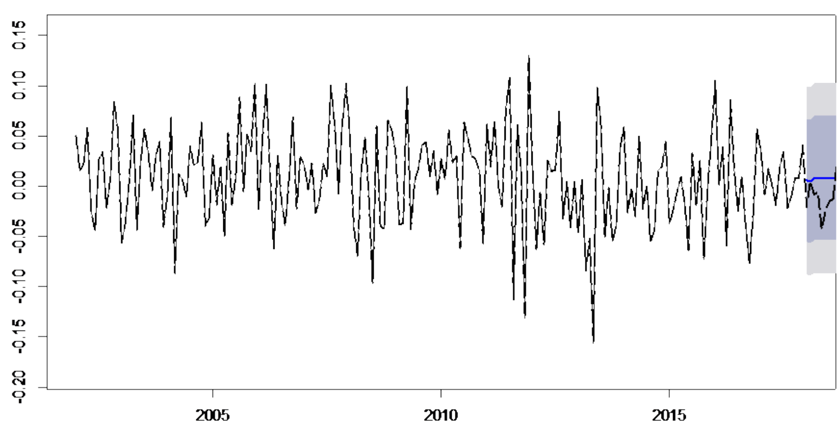



Figure 16. Forecasts from ARIMA(0,0,2) with non-zero mean (blue line), the testing sample of monthly gold return (dash line), the 80% forecast limits (blue), and the 90% forecast limits (gray).

| Time | Actual | Forecast | 95% lower | 95% upper |
|---|---|---|---|---|
| Jan-18 | -0.020429566 | 0.005707722 | -0.0873 | 0.0987 |
| Feb-18 | 0.004542537 | 0.004682455 | -0.0886 | 0.0981 |
| Mar-18 | -0.008077251 | 0.008052056 | -0.0857 | 0.1018 |
| Apr-18 | -0.005995703 | 0.008052056 | -0.0857 | 0.1018 |
| May-18 | -0.042967718 | 0.008052056 | -0.0857 | 0.1018 |
| Jun-18 | -0.023874242 | 0.008052056 | -0.0857 | 0.1018 |
| Jul-18 | -0.015268102 | 0.008052056 | -0.0857 | 0.1018 |
| Aug-18 | -0.012721434 | 0.008052056 | -0.0857 | 0.1018 |
| Sep-18 | 0.023063215 | 0.008052056 | -0.0857 | 0.1018 |
| Oct-18 | 0.002137719 | 0.008052056 | -0.0857 | 0.1018 |
| Nov-18 | 0.04923788 | 0.008052056 | -0.0857 | 0.1018 |
| Dec-18 | 0.050256626 | 0.008052056 | -0.0857 | 0.1018 |

Table 2. Monthly return of gold forecasts with actual values, forecast values from ARIMA (0,0,2) and 95% forecast limits.

At the result, this report has succesfully find the adequate ARIMA model to predict the monthly return of gold in 12 months of 2018 base on the LBMA's monthly gold prices from January 2002 to December 2018.

## References

[1] Abdullah, Lazim. (2012). ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. International Journal of Advances in Applied Sciences. 1. 10.11591/ijaas.v1i4.1495.

[2] Azzutti, Alessio. (2016). Forecasting Gold Price: A Comparative Study. 10.13140/RG.2.1.4206.5686.

[3] Cryer, J. D., & Chan, K. (2011). Time series analysis with applications in R. New York: Springer.

[4] Guha, Banhi & Bandyopadhyay, Gautam. (2016). Gold Price Forecasting Using ARIMA Model. Journal of advance Management Journal. 10.12720/joams.4.2.117-121.

[5] London bullion market. (2019). Retrieved from https://en.wikipedia.org/wiki/London_bullion_market

[6] Quandl. (2019). Retrieved from https://www.quandl.com/data/LBMA-London-Bullion-Market-Association

[7] Yang, Xiaohui. (2019). The Prediction of Gold Price Using ARIMA Model. 10.2991/ssphe-18.2019.66.