Time Series – Student Project
Validation by Experience or Education (VEE)
February 2016

# Review of Historical Price of Gold

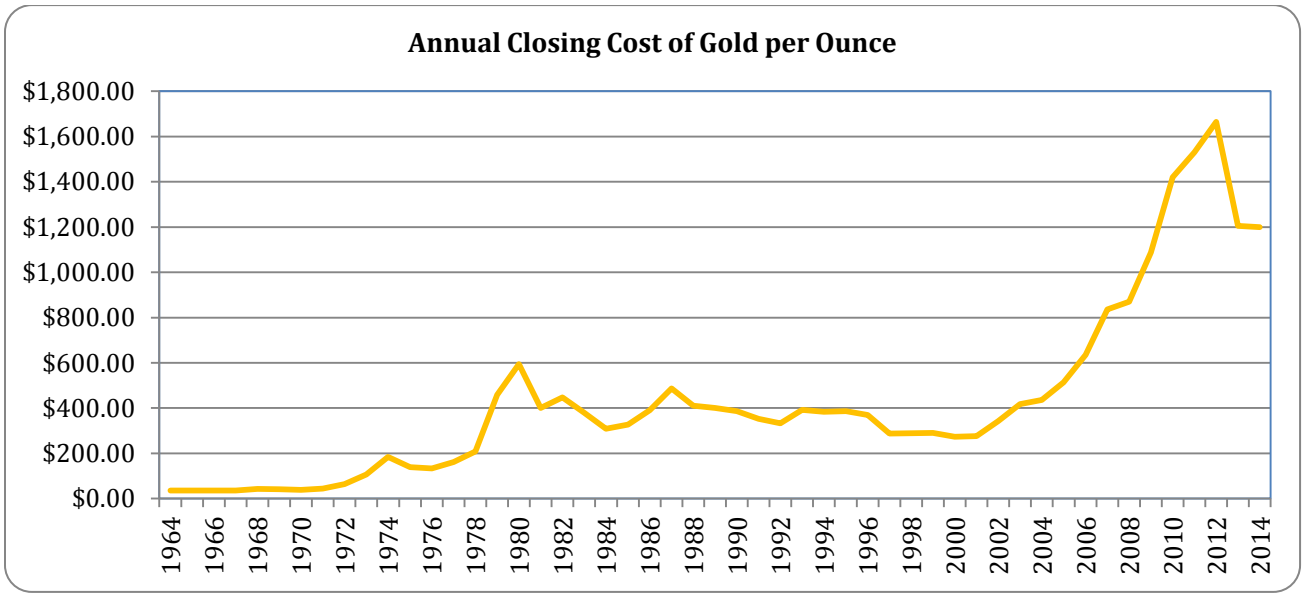
## Introduction

Of all the precious metals, gold is the most popular as an investment. Investors generally buy gold as a way of diversifying risk. The gold market is subject to speculation and volatility as are other markets. Gold is shown to have the most effective safe haven and hedging properties across a number of countries out of all precious metals available for investment. Gold is the 79th element and represented by Au. Additionally, to quote Goldmember, "I love gooooooooold".

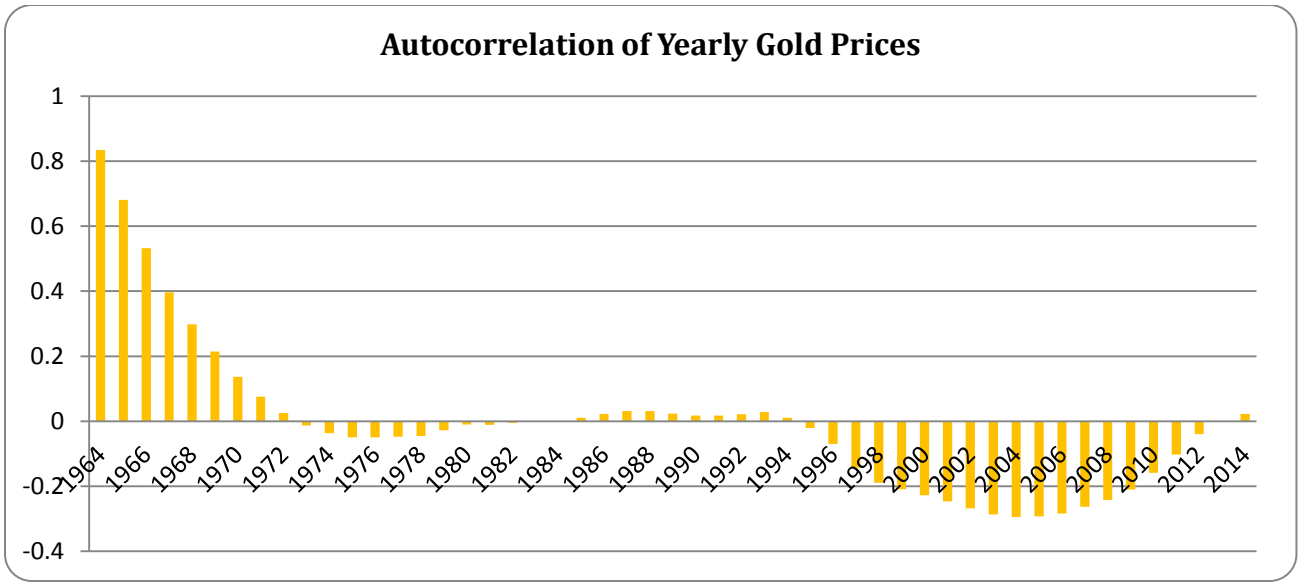
Data was obtained for the closing price at year-end of gold from 1792 to 2014 from http://onlygold.com/Info/Historical-Gold-Prices.asp. While all data and information on the internet should be trusted (you can't possibly post inaccurate information on something as sacred as the internet), the information has not been reviewed further or validated.

## Analysis

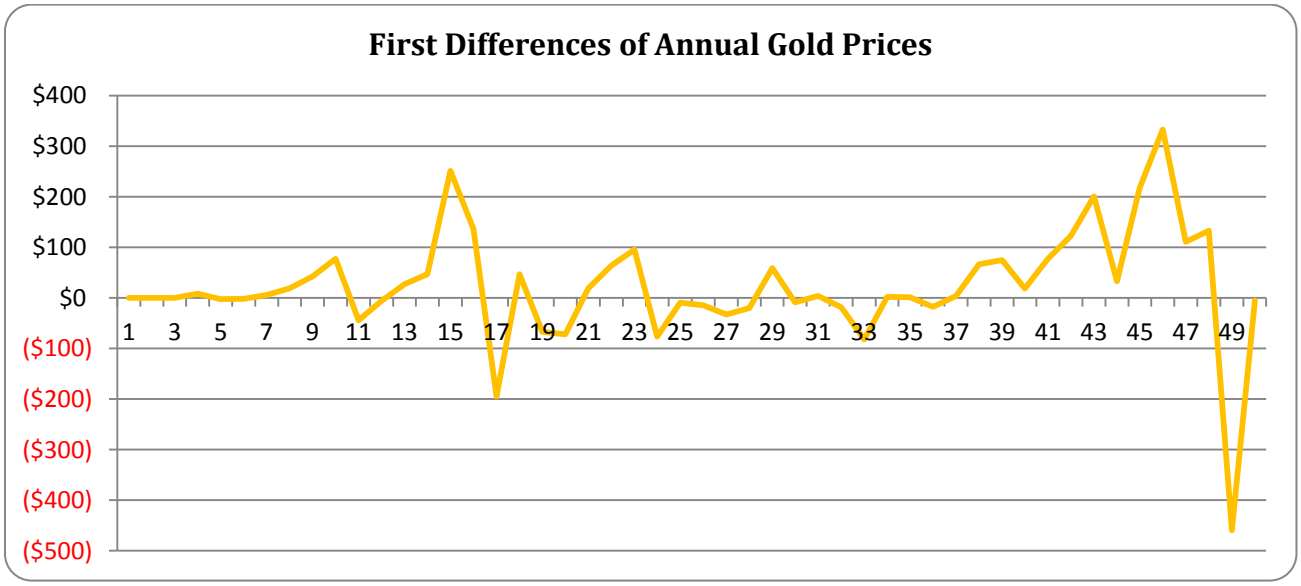
While data was available for over 200 years, we will review the price of gold for the past 50 years of data (1964-2004). The prices below have not been adjusted for CPI or any other external factors. Based on the increasing trend, we can conclude that this is not a stationary process.



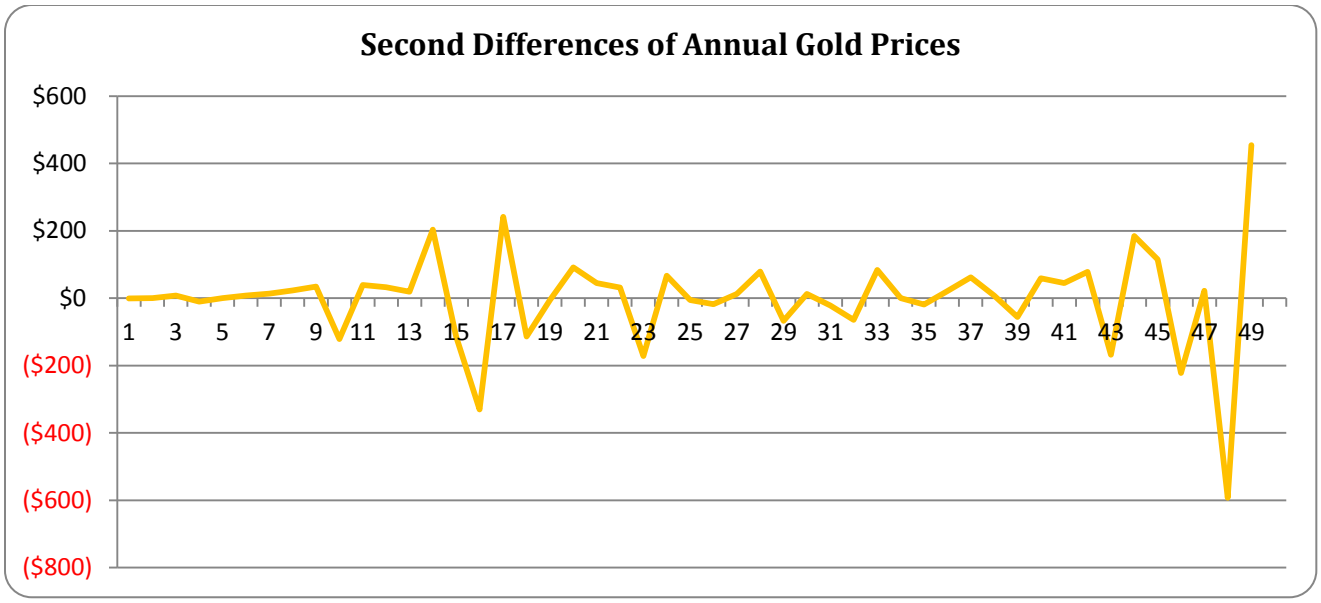

We are also able to review whether the series is stationary with the autocorrelation, shown below:

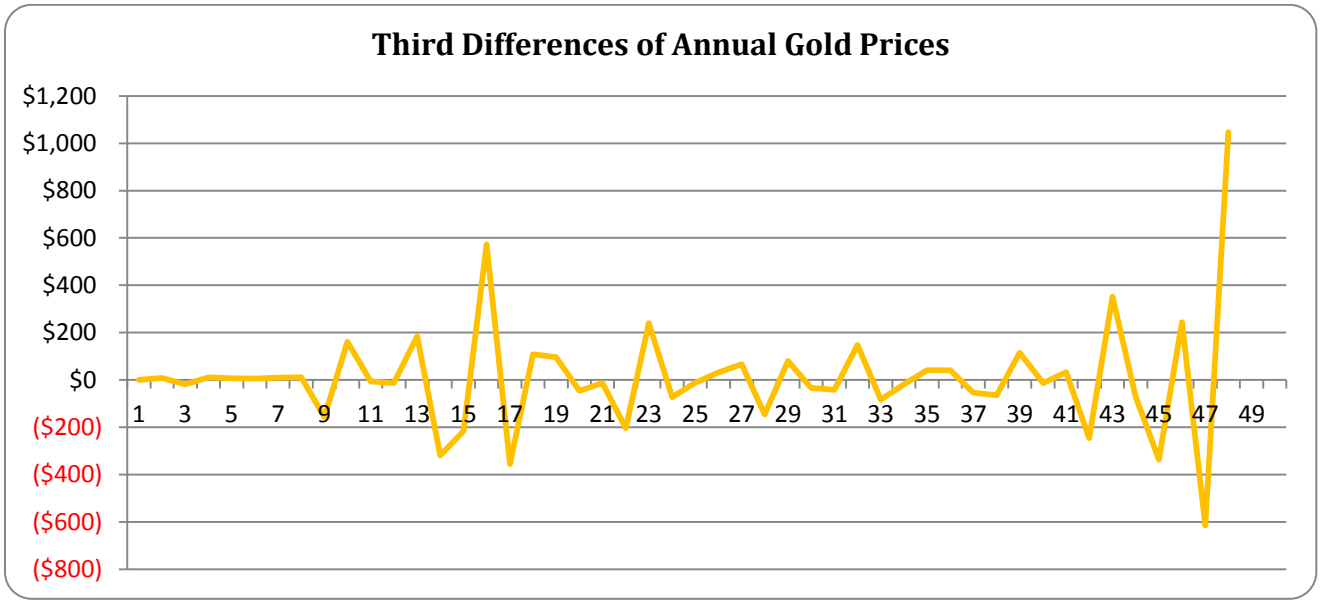



The results of the analysis also support the times series being non-stationary. Next we will review the first, second and third differences of the series.

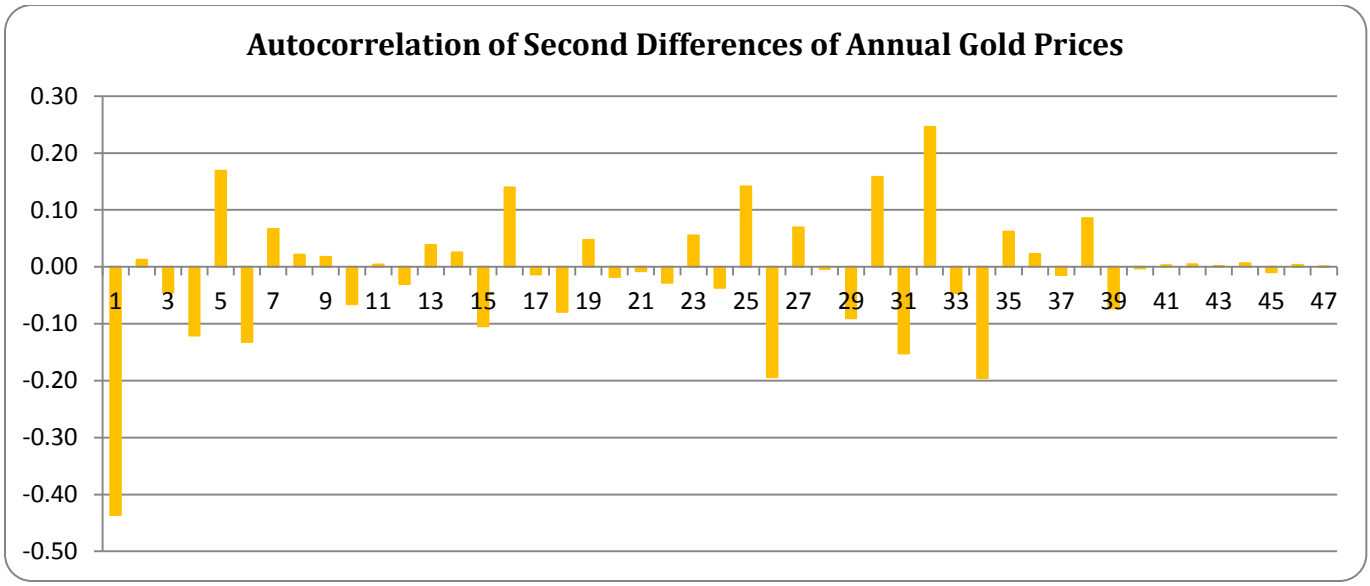


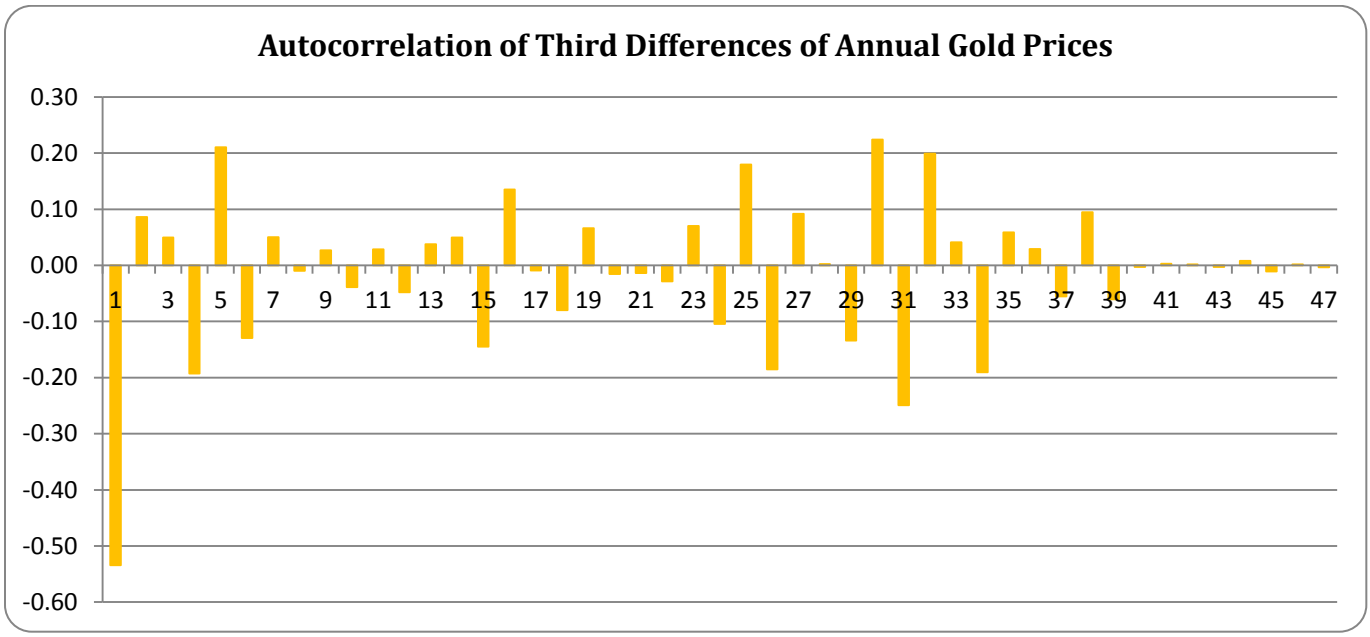

The graph of the first differences of the price of gold suggests that this may be a stationary process, counter to our original hypothesis. We will also review the second and third differences.




The graphs of the differences also suggest that the time series may be stationary. We can also review the autocorrelation of the second and third differences.




Finally, we will determine which regression model is the best fit for the series.

The following outlines the results of a regression analysis for the various models.

***AR(1)***

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.958369 |
| R Square | 0.918471 |
| Adjusted R Square | 0.916772 |
| Standard Error | 113.686 |
| Observations | 50 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 6988860 | 6988860 | 540.7446 | 8.88E-28 |
| Residual | 48 | 620376.6 | 12924.51 | | |
| Total | 49 | 7609237 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 28.63991 | 23.93679 | 1.196481 | 0.237386 |
| X Variable 1 | 0.987164 | 0.042452 | 23.25392 | 8.88E-28 |

***AR(2)***

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.959092 |
| R Square | 0.919858 |
| Adjusted R Square | 0.916374 |
| Standard Error | 113.8621 |
| Observations | 49 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 2 | 6845068 | 3422534 | 263.9911 | 6.15E-26 |
| Residual | 46 | 596370.8 | 12964.58 | | |
| Total | 48 | 7441438 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 31.79134 | 24.51413 | 1.296858 | 0.201148 |
| X Variable 1 | 1.170259 | 0.144605 | 8.092771 | 2.16E-10 |
| X Variable 2 | -0.20039 | 0.149844 | -1.33733 | 0.187689 |

***AR(3)***

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.958727 |
| R Square | 0.919157 |
| Adjusted R Square | 0.913645 |
| Standard Error | 115.547 |
| Observations | 48 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 6679115 | 2226372 | 166.7556 | 4.8E-24 |
| Residual | 44 | 587448.5 | 13351.1 | | |
| Total | 47 | 7266564 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 38.974 | 26.49366 | 1.471069 | 0.148389 |
| X Variable 1 | 1.137222 | 0.152459 | 7.459213 | 2.44E-09 |
| X Variable 2 | -0.01022 | 0.290297 | -0.0352 | 0.972079 |
| X Variable 3 | -0.18662 | 0.242167 | -0.77062 | 0.445053 |

***AR(4)***

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.957857 |
| R Square | 0.917489 |
| Adjusted R Square | 0.909631 |
| Standard Error | 117.9724 |
| Observations | 47 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 6499800 | 1624950 | 116.756 | 3.58E-22 |
| Residual | 42 | 584534.5 | 13917.49 | | |
| Total | 46 | 7084334 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 38.06402 | 29.01411 | 1.311914 | 0.196675 |
| X Variable 1 | 1.140903 | 0.157017 | 7.266102 | 6.07E-09 |
| X Variable 2 | -0.00164 | 0.296985 | -0.00551 | 0.995629 |
| X Variable 3 | -0.27067 | 0.347018 | -0.77999 | 0.439769 |
| X Variable 4 | 0.081655 | 0.249116 | 0.32778 | 0.744707 |

***AR(5)***

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.956976 |
| R Square | 0.915803 |
| Adjusted R Square | 0.905279 |
| Standard Error | 120.5256 |
| Observations | 46 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 5 | 6320112 | 1264022 | 87.01545 | 2.08E-20 |
| Residual | 40 | 581056.6 | 14526.41 | | |
| Total | 45 | 6901169 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 34.95293 | 32.47528 | 1.076294 | 0.288245 |
| X Variable 1 | 1.13874 | 0.160712 | 7.085609 | 1.43E-08 |
| X Variable 2 | -0.00753 | 0.304378 | -0.02475 | 0.980376 |
| X Variable 3 | -0.24301 | 0.360316 | -0.67444 | 0.503912 |
| X Variable 4 | -0.03126 | 0.361405 | -0.08649 | 0.931508 |
| X Variable 5 | 0.112359 | 0.26391 | 0.425748 | 0.672574 |

## Conclusion

The following summarizes the results of the models:

| | P-values | | | | |
|---|---|---|---|---|---|
| | **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **AR(5)** |
| **Intercept** | 0.2374 | 0.2011 | 0.1484 | 0.1967 | 0.2882 |
| **ϕ1** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| **ϕ2** | | 0.1877 | 0.9721 | 0.9956 | 0.9804 |
| **ϕ3** | | | 0.4451 | 0.4398 | 0.5039 |
| **ϕ4** | | | | 0.7447 | 0.9315 |
| **ϕ5** | | | | | 0.6726 |
| | | | | | |
| **R Square** | 0.9185 | 0.9199 | 0.9192 | 0.9175 | 0.9158 |
| **Adjusted R Square** | 0.9168 | 0.9164 | 0.9136 | 0.9096 | 0.9053 |

Based on the above, the AR(2) model appears to be the most appropriate model.

With that, the resulting equation is as follows:

$$\mathbf{Y_t = 0.2011 + 0.000*Y_{t-1} + 0.1877*Y_{t-2}}$$