University of Waterloo
Faculty of Mathematics
Department of Statistics and Actuarial Science

STAT 443 PROJECT REPORT

Predicting Monthly Gold Prices (CHF)

Group 2: Xinjie Qiao, Catherine Zhou, Zoe Zou
December 4, 2022

Contents

List of Tables

List of Figures

1. Introduction

Throughout the history and development of economic systems, various goods and materials have played the role of money. Among them, the most representative one is gold. Even until now, Gold has always been considered a custodian of value and its basic function is to preserve purchasing power in times of great uncertainty. Gold price fluctuation trend prediction is an important issue in the financial world. Even small improvements in predictive performance can make lots of profits.

There are many methods in the literature that make predictions of the price of gold based on historical data, like random forest [1]. For another example, one team also use combines information from NASDAQ index, Hang sen Index to adjust the prediction [2].

In this project, the topic is to have some naive attempt of prediction of gold price within the scope of STAT 443. The methods that we will use are regression, smoothing methods, and Box-Jenkins models. Moreover, we will evaluate the prediction power by mean square of errors(MSE) and pick the best models.

In this project, our teammates are responsible for the following:

Xinjie Qiao Box-Jenkins models, SARIMA model selection, analysis on the validity of the model

Catherine Zhou Extraction of data, power transformation, regression models, project report

Zoe Zou Smoothing methods, introduction, presentation slides, presentation recording

---

2. Analysis

2.1. Data

We find the data from Kaggle, which contains the monthly gold price per gram from January 1979 to July 2021 from 18 different countries. We choose the Switzerland CHF currency as the data we want to analyze, because the price of the gold includes the fluctuations of the currency exchange rate. In order to make sure that the data only related to the nature volatility of gold price, we decided to choose the currency with more stable exchange rate, in which Switzerland is chosen, as that the country's zero-inflation policy, combined with its political independence, makes CHF an extremely powerful and stable currency [3].

Date	Switzerland.CHF.
31-01-1979	379.3
28-02-1979	413.6
30-03-1979	406.2
30-04-1979	420.0
31-05-1979	478.0
29-06-1979	457.7

Table 2.1: First 5 rows of the monthly gold price

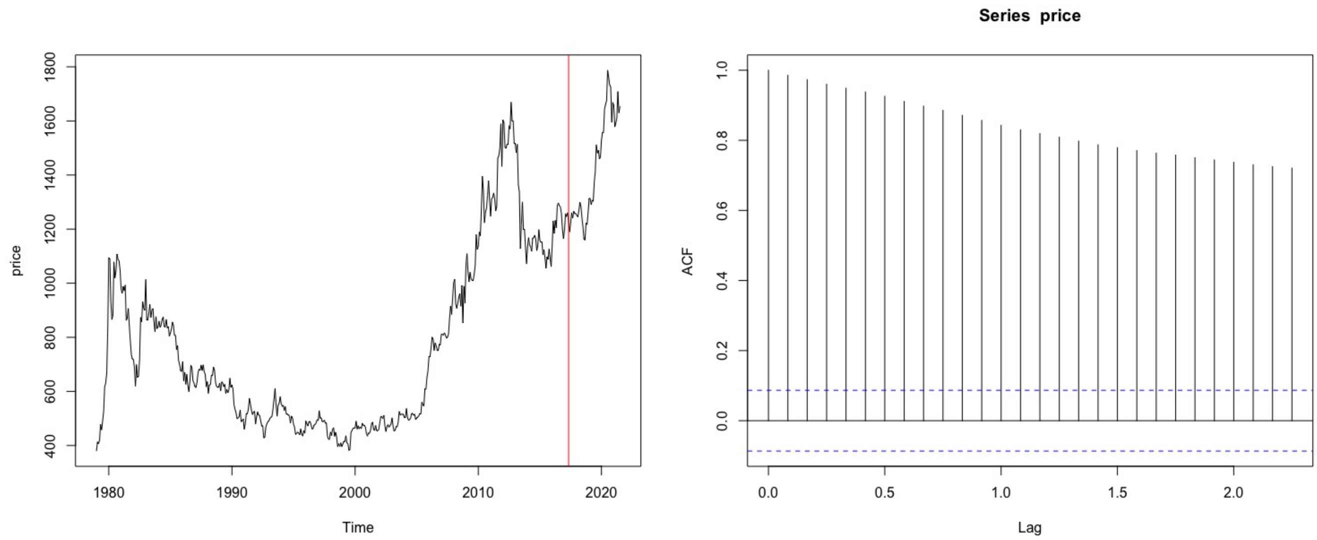


Figure 2.1: Time Series plot and ACF plot

Table 2.1 is the first 5 rows of our time-series data. We remove the date component from the first column and create a time series object by month. We want to fit a forecasting model on the data and predict the next 2 years' monthly gold price. From figure 2.1, we can see that our data has trend but no seasonality, since the ACF plot has a slow decay but not signs of periodic patterns or slow decay on seasonal lag. We also used classic decomposition on the data and observe significant trend and insignificant seasonality (see figure A.1 and figure A.2 for details).

Additionally, we didn't observe any non-constant variance in the plot. For academic rigour, we performed the Fligner-Killeen test on the data set and extremely small p-value has been observe, implicating that non-constant variance exists in our data-set. However, when we try to use power transformation and Box-Cox method to remove the non-constant variance, neither method is helpful (see Appendix B.1 for the result of F-test). Hence, we decide to move forward to try to fit models, but note that the results might not be valid since the data is non-stationary.

2.2. Modelling

We split the data-set into a training set and a test set, and use the test set MSE to select the best model within each category. The 90% training set is from 1979-01 to 2017-03, and the 10% test set is from 2017-04 to 2021-06 (see the red line in figure 2.1). Our goal is to predict the future 30 months of the gold price (up until 2023-12).

Three different types of models were applied to our data-set: regression (polynomial and Shrinkage methods), smoothing methods (exponential smoothing and Holt-Winters), and Box-Jenkins models (ARMA, ARIMA/SARIMA, etc).

2.2.1. Regression

We fit the polynomial regression with degree 2 to 10 on the training set and apply the model on the test set to predict. We then calculate the MSE and compare them across different degrees. The result and corresponding plot is shown in figure B.3 and table B.1. Degree 2 has the smallest test set MSE. However, we know that our data is pretty complex looking at figure 2.1, and whether a polynomial with only degree 2 can be a good fit remains questionable. The reason why our higher degree polynomial works terribly on the test set is due to bias-variance trade off (see appendix B.2 for details).

Then, we fitted shrinkage methods on the training data to see if they do better than the polynomial regression. We used 10-fold cross validation to find the best λ value.

We find that

- the best ridge model is at degree 9 with $\lambda = 1.0000$
- the best lasso model is at degree 10 with $\lambda = 0.2231$
- the best elastic model is at degree 9 with $\lambda = 0.0302$

The detailed results refers to appendix B.3.

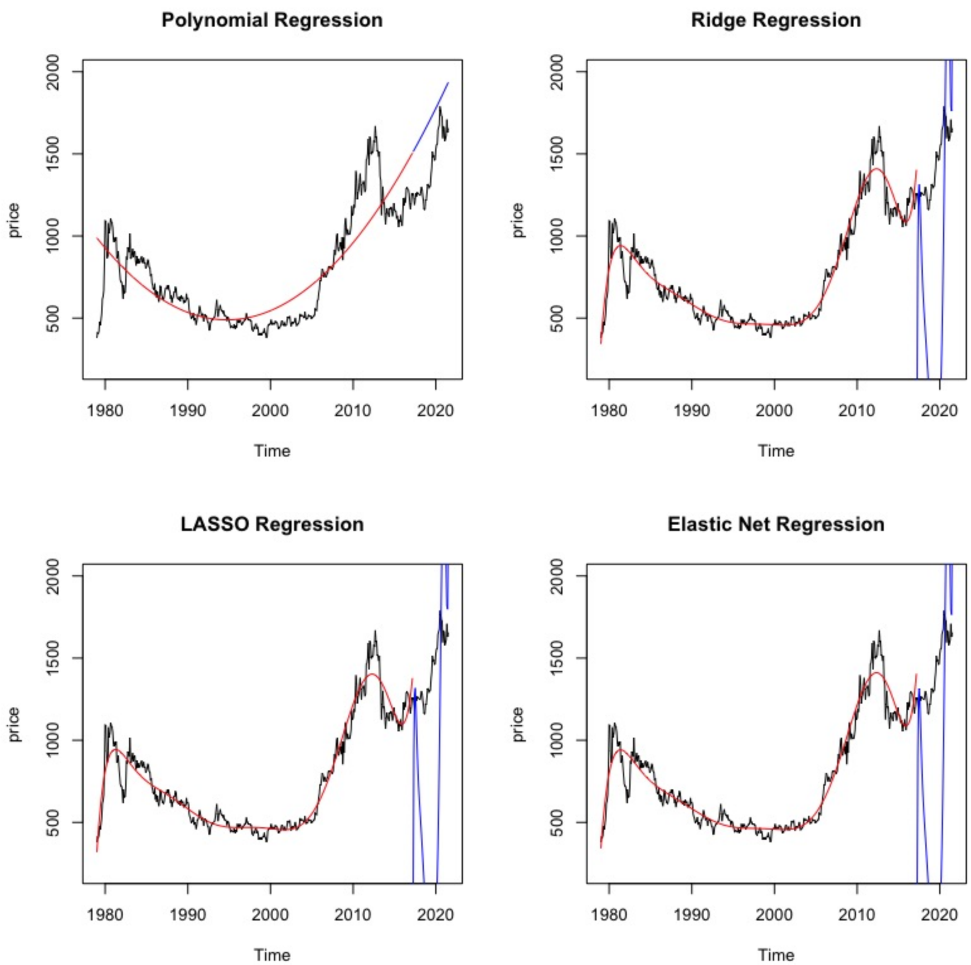


Figure 2.2: Regressions Training vs. Test set

With the chosen degree and λ, we compared the prediction power of the four proposed models (polynomial, ridge, LASSO, and elastic net) with MSE_{pred} on the test set.

polynomial	ridge	lasso	elastic
95478.66	1108226	1109892	1112580

Table 2.2: Regression Prediction MSE

It's clear that the polynomial regression with degree of 2 has the smallest MSE_{pred}. We observed that this was an example of bias-variance trade off, since the polynomial regression had the worst fit, and all of the shrinkage methods had pretty descent fit. The high bias brought low variance in prediction, thus the polynomial regression surprisingly had a much better prediction on the test set than any of the shrinkage methods.

Hence, we chose it as our best regression model and moved on to see how it fitted on the entire data-set, and whether the model assumptions would be met.

Now, we fitted the optimal polynomial regression on the entire data-set (see Appendix B.2 for summary of the model) and calculated its residual of the whole data set. We also performed a model diagnostic to check whether the residuals follows the model assumptions.

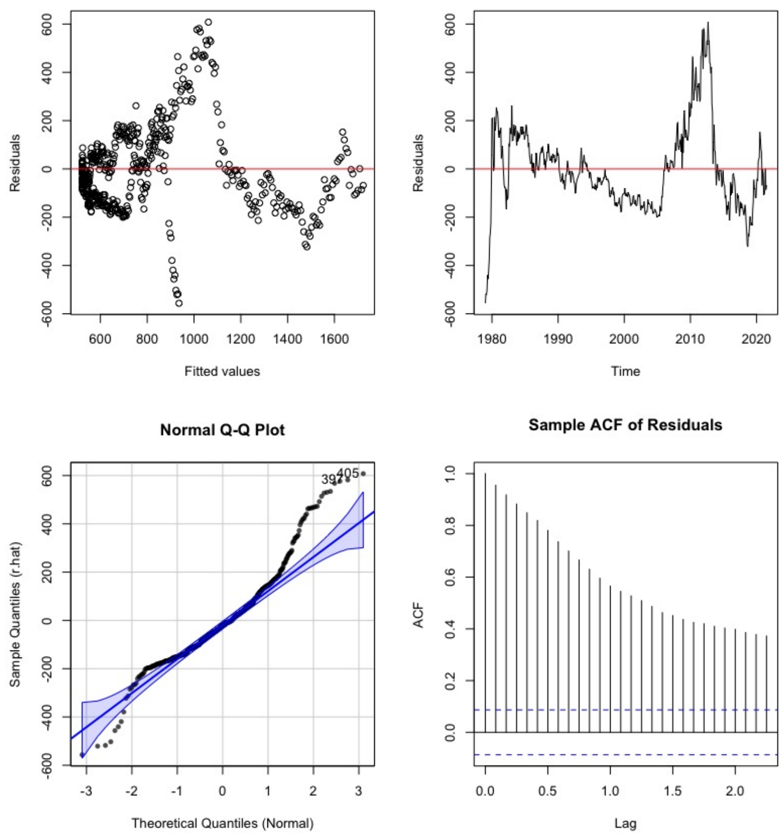


Figure 2.3: Polynomial Regression Residual Diagnostics

From figure 2.3, we could see that the residuals

- did'nt have constant mean: trend was observed in the residuals plot,
- had correlation with the fitted values: trend was observed in the residuals vs. fitted values plot,
- was not normal: the Q-Q plot had a significant amount of data points falling out of the confidence interval,
- was correlated with each other: the ACF plot had very significant lags and has exponential decay.

2.2.2. Smoothing Methods

We performed single exponential smoothing, double exponential smoothing, additive and multiplicative Holt-Winters algorithm on the training set, and applied the model on the test set to see how strong the fit and prediction power is. The model selection criteria was MSE_{pred} as well.

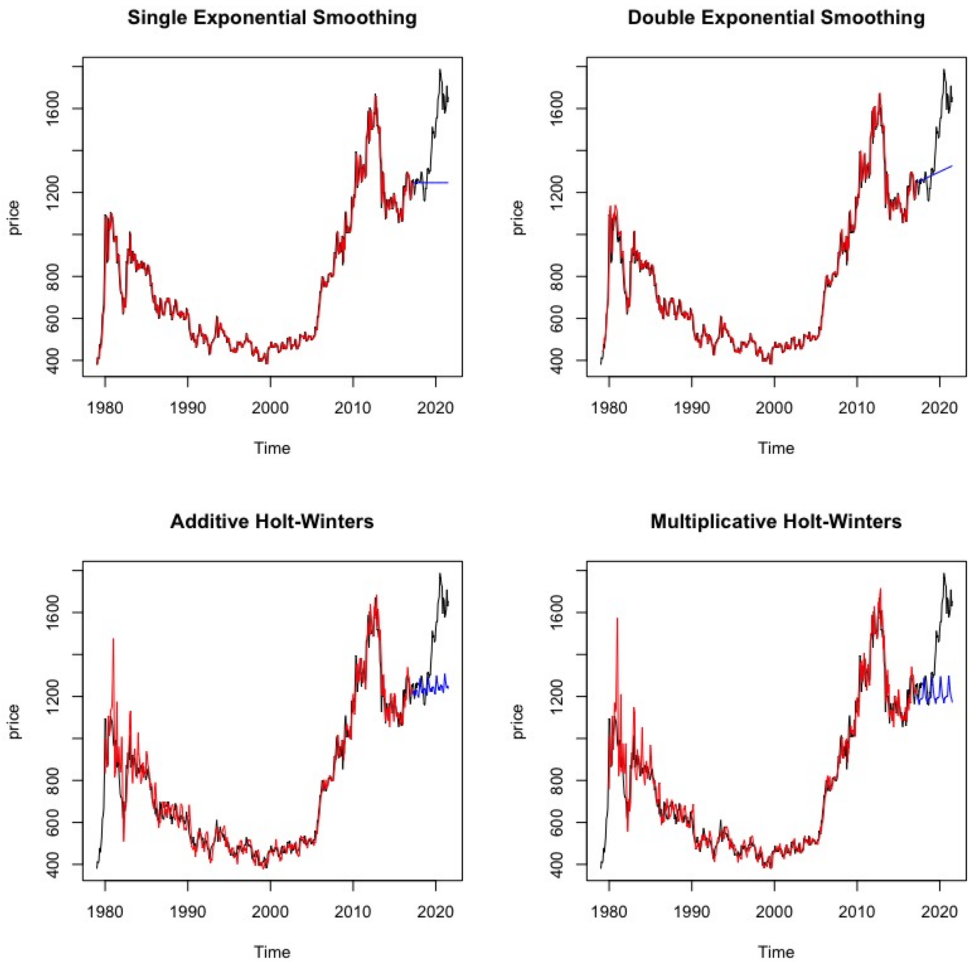


Figure 2.4: Smoothing Methods Training vs. Test

single	double	additive HW	multiplicative HW
68041.7	47731.91	67930.53	83083.73

Table 2.3: Smoothing Methods Prediction MSE

From the R output, the smallest MSE resulted in the double exponential smoothing. By figure 2.4, we could see that even the model is almost perfect fit, the prediction for the test set was quite terrible. We concluded it as an extreme example of bad bias-variance trade-off. Obviously, smoothing methods were not good choices for predicting the trend of the gold price.

To check the model assumption, we used the same four plots as in the regression section 2.2.1.

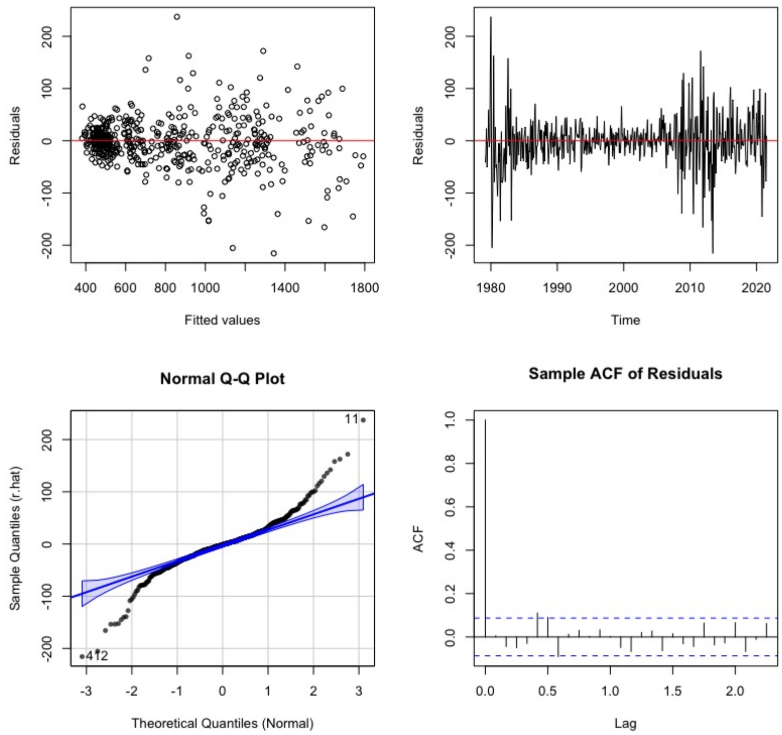


Figure 2.5: Double Exponential Smoothing Residual Diagnostics

Figure 2.5 showed that the residuals

- were not correlated with the fitted values
- had constant mean of 0, as seen in the residual plot. However, this plot also showed the non-constant variance in the residuals. This was due to the overall non-constant variance issues we'd discussed earlier (B.1),
- were not normal, which might be influenced by the non-constant variance,
- had no trend or seasonality, as illustrated in the sample ACF plot.

Therefore, the only source of non-stationarity was the non-constant variance.

2.2.3. Box-Jenkins Models

From figure 2.1, we saw that the data contains significant trend and no seasonality. Hence, we performed regular differencing to first remove non-stationarity from the data.

After performing one regular differencing on the data, the trend seemed to be removed from the data since there was no linear decay in the ACF plot. To be sure that no more differencing was needed, we performed another differencing, and figure B.7 confirmed this conclusion. To avoid over-differencing, we stopped here and moved on to propose models.

Based on the ACF and PACF plot (2.6), we proposed several ARIMA models with no seasonality:

ARIMA(0,1,1) We observed that the ACF plot cuts off after lag 1, and the PACF plot had a damped sine wave. The spikes in the ACF plot after lag 1 was due to the 95% confidence interval.

ARIMA(1,1,0) We observed that the ACF plot has a damped sine wave, and the PACF plot cut off after lag 1 (we can see a little spike on lag 1). The spikes in the PACF plot after lag 1 was due to the 95% confidence interval.

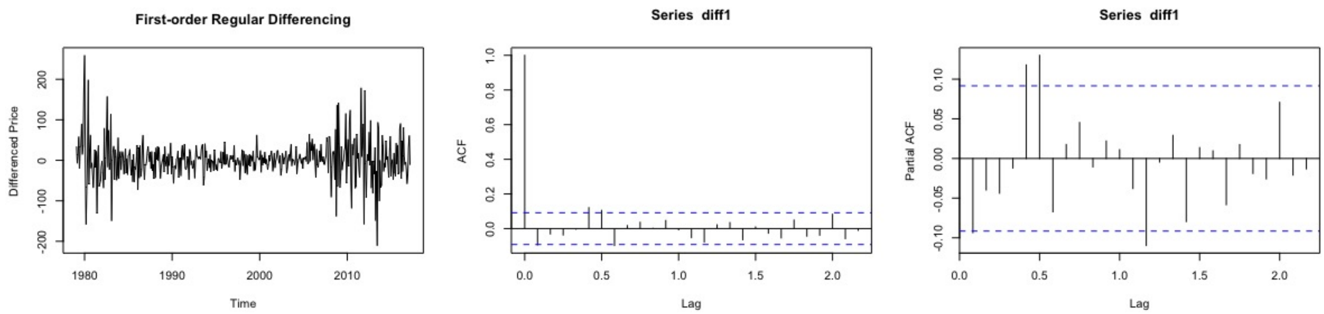


Figure 2.6: First order Regular Differencing

ARIMA(0,1,5) We observed that the ACF plot cut off after lag 5, and the PACF plot had a damped sine wave. The spikes in the ACF plot after lag 5 was due to the 95% confidence interval.

ARIMA(0,1,6) We observed that the ACF plot cut off after lag 6, and the PACF plot had a damped sine wave. The spikes in the ACF plot after lag 6 was due to the 95% confidence interval.

ARIMA(6,1,0) We observed that the ACF plot had a damped sine wave, and the PACF plot cut off after lag 6. The spike in the PACF plot after lag 6 was due to the 95% confidence interval.

ARIMA(1,1,1) As above, we observed damped sine wave in both ACF and PACF plots.

ARIMA(p,1,q) We can also tried other ARMA methods with larger p and q values.

Model	MSE_{pred}	AIC	AICc	BIC
ARIMA(0,1,1)	43795.86	10.48568	10.48573	**10.51271**
ARIMA(1,1,0)	**43692.13**	10.48642	10.48648	10.51345
ARIMA(0,1,5)	46688.48	10.47991	10.48032	10.54299
ARIMA(0,1,6)	49255.66	10.47709	10.47764	10.54918
ARIMA(6,1,0)	47231.58	**10.47324**	**10.47378**	10.54532
ARIMA(1,1,1)	44388.76	10.48873	10.48884	10.52477
ARIMA(1,1,2)	44479.14	10.49295	10.49314	10.53800
ARIMA(2,1,1)	44517.89	10.49288	10.49308	10.53794
ARIMA(2,1,2)	44396.06	10.49741	10.49770	10.55148

Table 2.4: Box-Jenkins Methods Comparison

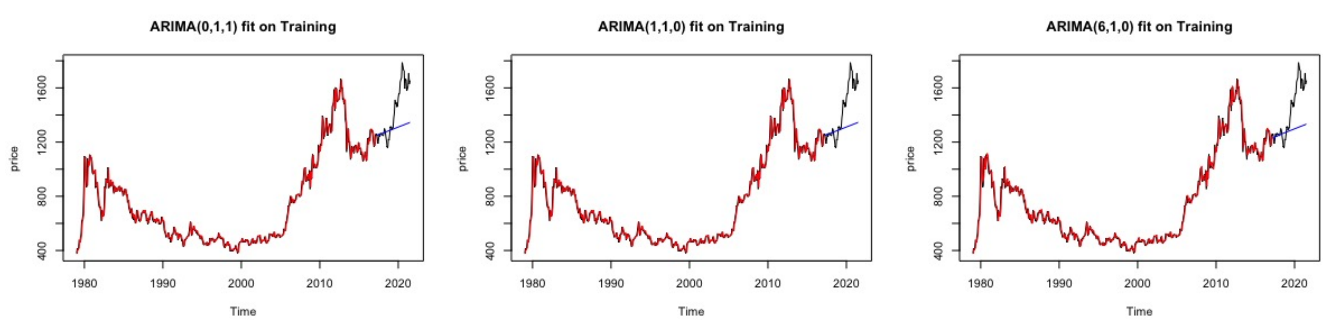


Figure 2.7: ARIMA models fit on Training set

After calculating the MSE, AIC, AICc and BIC, which are shown in table 2.4, we found that ARIMA(6,1,0), ARIMA(0,1,1), and ARIMA(1,1,0) all had better prediction power than the rest proposed models. We included the later two here because ARIMA(6,1,0) had 6 parameters comparing to the other two, and we tried to avoid choosing a model with too many parameters. Looking at figure 2.7, we saw that all three models have similar fit and prediction, so we would perform residual analysis on all three models to find the best model.

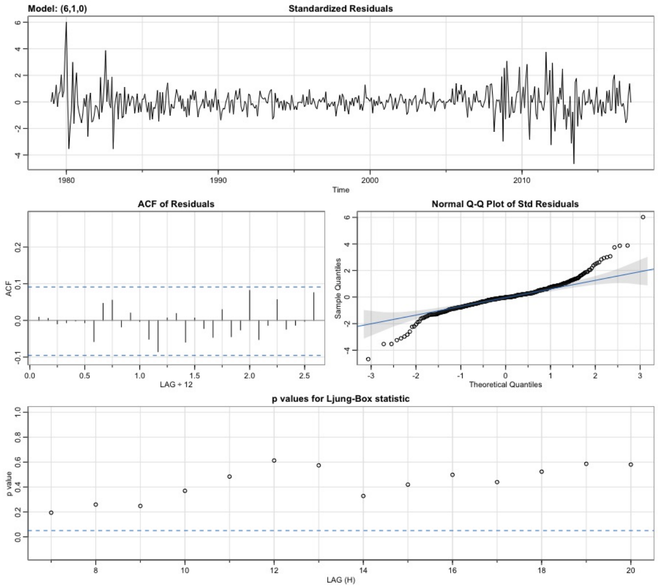


Figure 2.8: ARIMA(6,1,0) on Training set

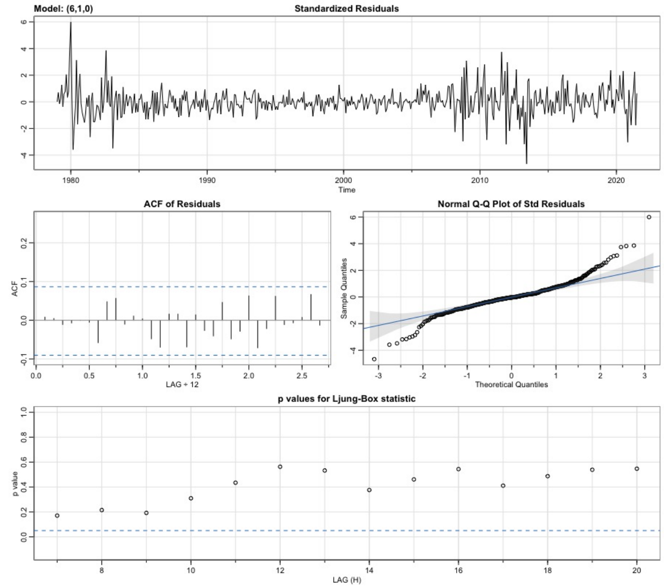


Figure 2.9: Final ARIMA Model Residual Analysis

In Figure 2.8, we could see that the residuals have constant mean, no seasonality, not normal (might due to the variance issue B.1), and the p values seemed great. Other than non-constant variance, the residuals were pretty much stationary. We also included the residual analysis for ARIMA(1,1,0) (figure B.9) and ARIMA(0,1,1) (figure B.8) in Appendix, and they didn't look as good as this model.

After checking the assumptions, we fitted the model on the entire data-set and estimate the parameters.

Model	MSE	AIC	AICc	BIC
ARIMA(6,1,0)	2012.242	10.47854	10.47897	10.54496

Table 2.5: Final ARIMA Fit and Prediction Power

We could see that the residuals (figure 2.9) matched the model assumptions, except non-constant variance and deviance in normality.

3. Conclusions

3.1. Statistical Conclusions

We used the following criteria to compare the selected regression, smoothing, and Box-Jenkins model:

1. MSE_{pred} on the testing set
2. MSE on the entire data-set
3. Residual analysis

The first analyzed the prediction power of the model; the second analyzed the fit of the model, and the third showed the validity of the model.

Model	MSE_{fit}	MSE_{pred}
Polynomial	30239.665	95478.66
Smoothing	2124.056	47731.91
Box-Jenkins	**2012.242**	**47231.58**

Table 3.1: Final Comparison between 3 models

The above table shows that the Box-Jenkins model (2.2.3) had strongest fit and prediction power among all three. And if we look back at the residual analysis of the 3 models (figure 2.3, figure 2.5, figure 2.9), the ARIMA model's residuals are the closest to the model assumption. Therefore, we conclude that ARIMA(6,1,0) is the best model to predict the monthly gold price data we have.

Now, we would use ARIMA(6,1,0) to predict the next 30 months' gold price.

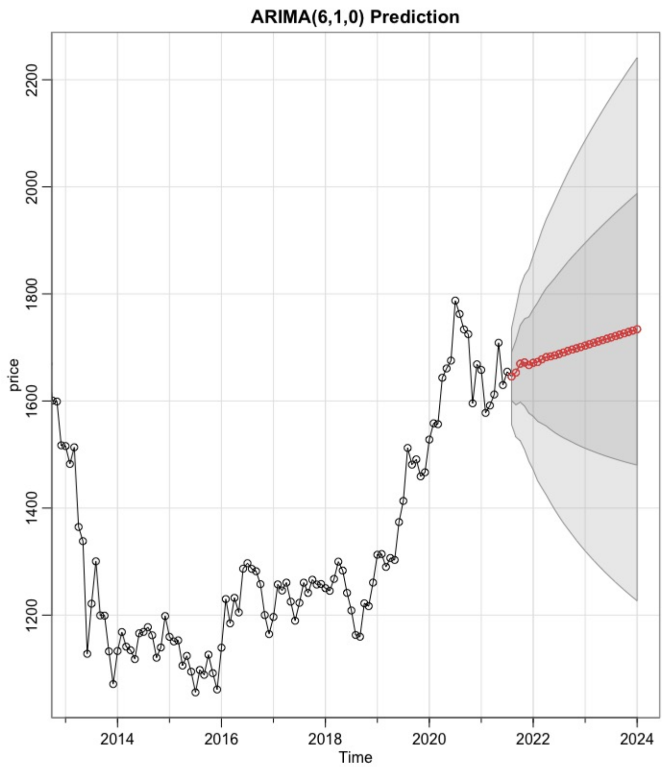


Figure 3.1: ARIMA(6,1,0) Prediction

year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2021								1645.730	1653.005	1669.634	1672.018	1667.086
2022	1671.276	1672.802	1677.870	1681.929	1683.312	1685.172	1687.807	1690.540	1693.578	1696.108	1698.402	1700.877
2023	1703.453	1706.077	1708.667	1711.164	1713.662	1716.197	1718.749	1721.302	1723.837	1726.362	1728.893	1731.431

Table 3.2: ARIMA(6,1,0) Prediction values

However, due to the non-constant variance within the residuals, the validity of the prediction interval remained questionable.

3.2. Connect to the Context

In this project, we found that the gold price was pretty much hard to predict using a time-series model due to the complexity and cluster variation existing within this topic. The price is not only tied to time but also many other factors. As mentioned in the introduction, some research has already put effort in this prediction and include parameters such as NASDAQ index, which relates to data from the stock market.

As a result, there were several change points in the data-set, e.g. around the 1980s, 2010s, and 2020s. We could guess that the first is due to the stock market recession in 1980-1981, the second is due to the financial crisis in 2007-2008, and the last is due to the COVID-19 pandemic. All these global incidents cause the high variation and non-predictable pattern in the data-set, which is another reason why it's hard to remove the non-constant variance (B.1).

Our predictions above are the monthly gold prices from 2021 August to 2023 December. The monthly gold price would have a steady and slowly increasing trend, but the variation in our results exists and was pretty huge as shown.

A. Additional Figures

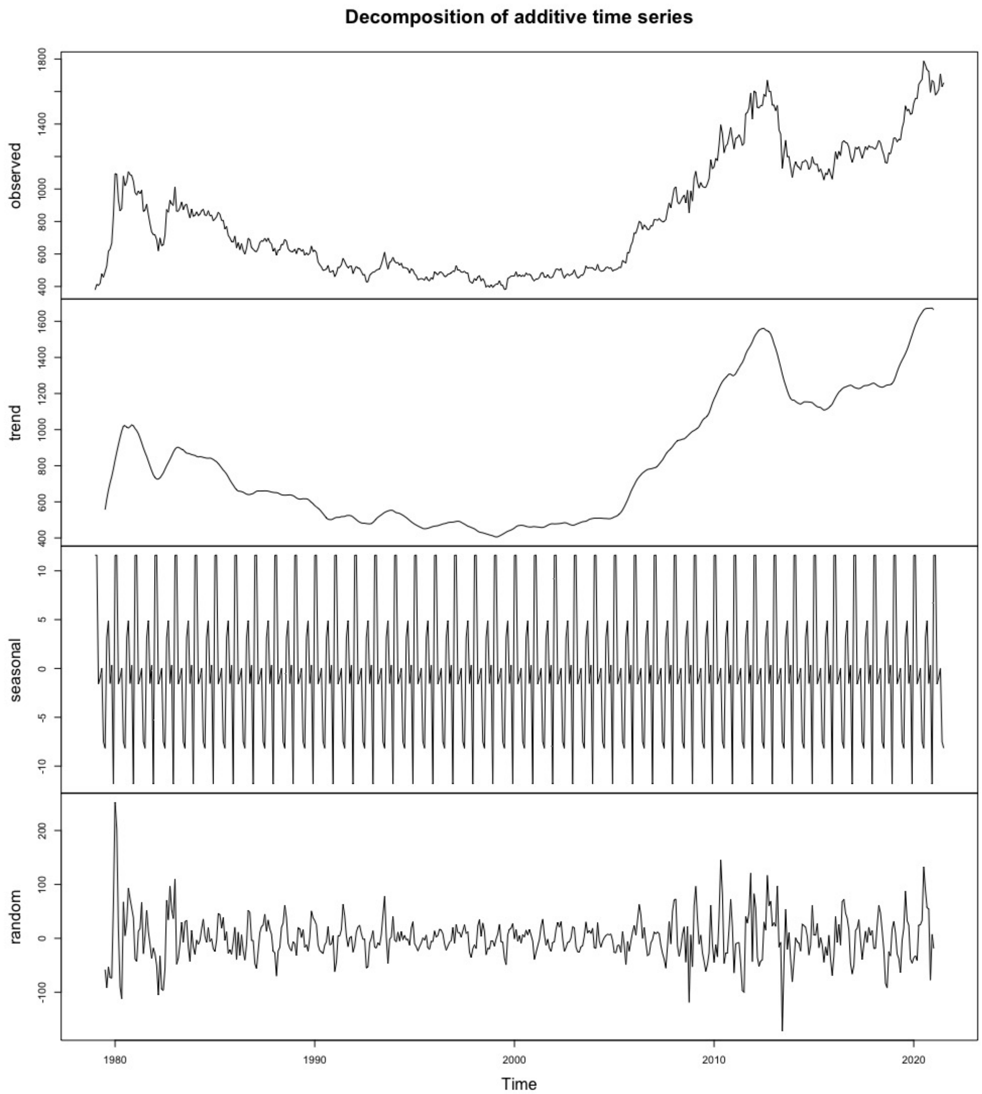


Figure A.1: Additive decomposition plot

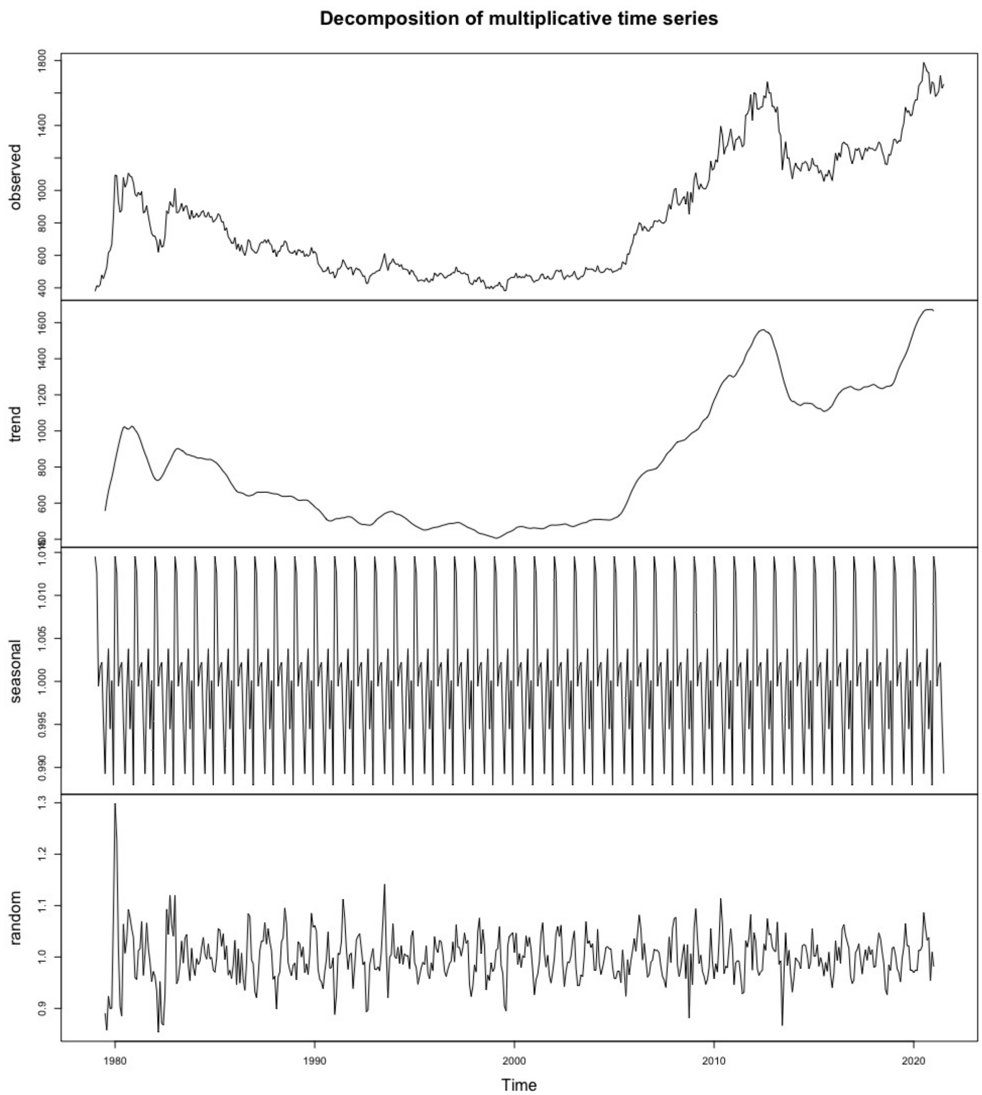


Figure A.2: Multiplicative decomposition plot

B. Supplementary Material

B.1. Non-constant Variance

We got the following when performing the Fligner test on the entire data:

```
Fligner-Killeen test of homogeneity of variances
data:  price and seg
Fligner-Killeen:med chi-squared = 197.01, df = 4, p-value < 2.2e-16
```

We want to choose α from $\{-2, -1.5, -1, -0.5, 0, 0.5, 1.5, 2\}$. When we use the Boxcox method, we get the following graph:

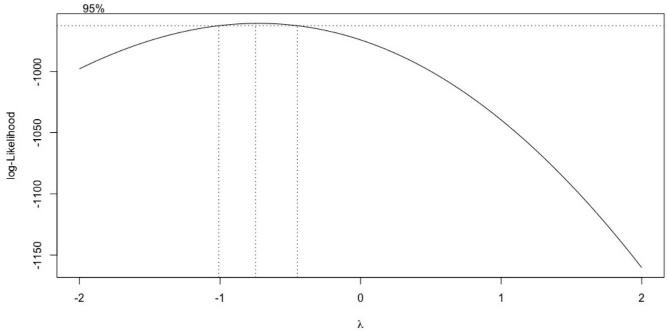


Figure B.1: Boxcox result

When we again test on the power transformed data with $\alpha = -0.7474747$, we get

```
Fligner-Killeen test of homogeneity of variances
data:  price^bx$x[which.max(bx$y)] and seg
Fligner-Killeen:med chi-squared = 107.15, df = 4, p-value < 2.2e-16
```

The extremely small p-value doesn't improve our results above.

We also compare the fligner test result on each α in the above list to find the optimal α. The p-values from each test is as follows:

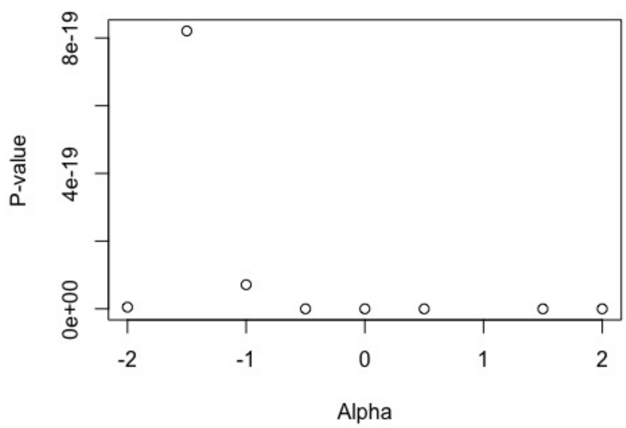


Figure B.2: Fligner p-values

All the p values are extremely small, which means that any of α doesn't remove non-constant variance in our data.

B.2. Polynomial Regression Model

Model selection results:

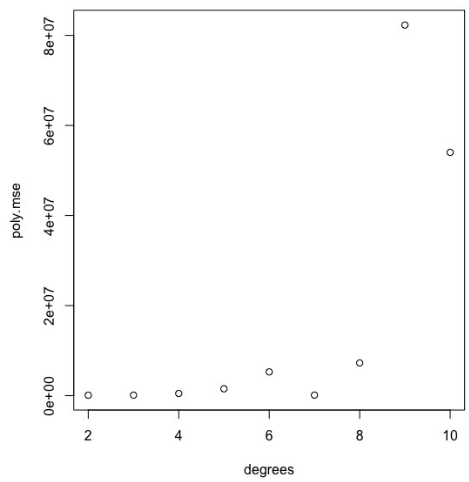
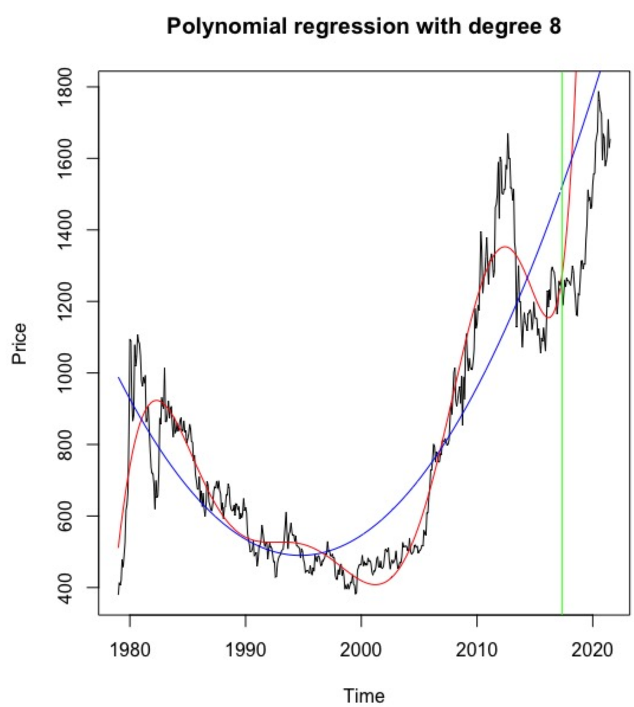


Figure B.3: Polynomial regression test set MSE

degree	MSE_{pred}
2	95478.66
3	116093.35
4	489939.41
5	1521066.09
6	5275901.68
7	120539.58
8	7270130.42
9	82291940.62
10	54021934.33

Table B.1: Polynomial regression test set MSE

We observed that higher polynomial degree result in a higher prediction MSE. This is because models with higher polynomial degree fits better on the training set (smaller bias) but consequently has a higher variance.



The above plot shows the model with degree 8 (red) and degree 2 (blue). Degree 8 model fits the data much better than the degree 2 model, but it has a much worse prediction comparing to degree 2 model.

B.3. Shrinkage methods

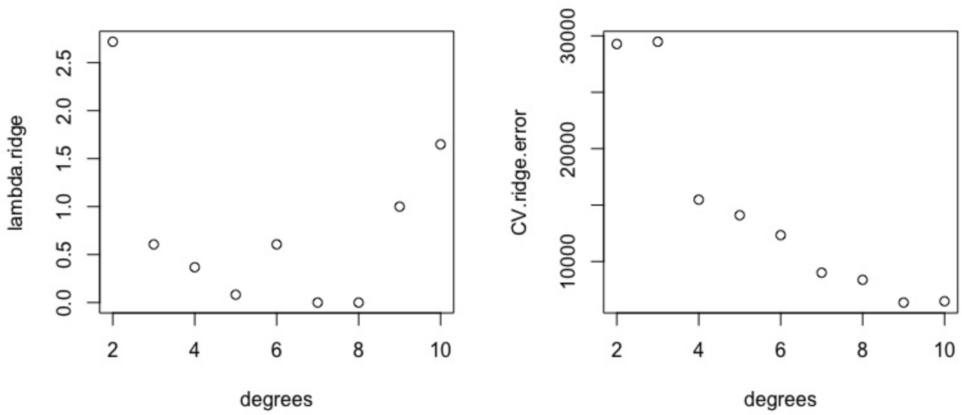


Figure B.4: Ridge Regression Lambda Values

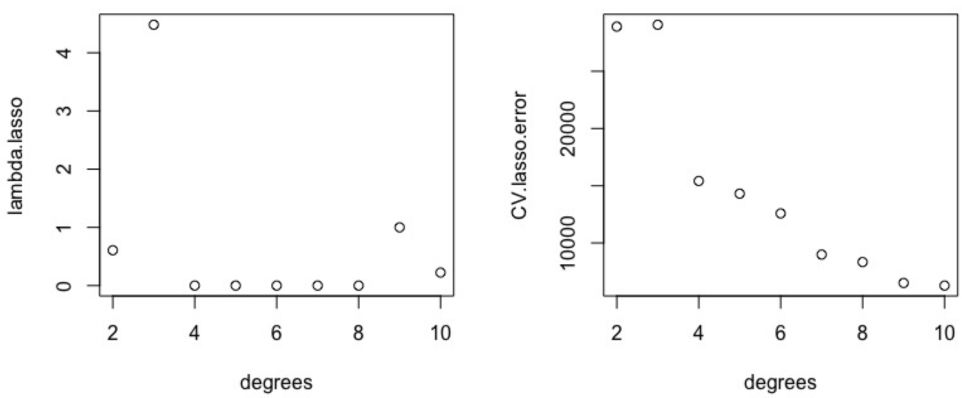


Figure B.5: LASSO Regression Lambda Values

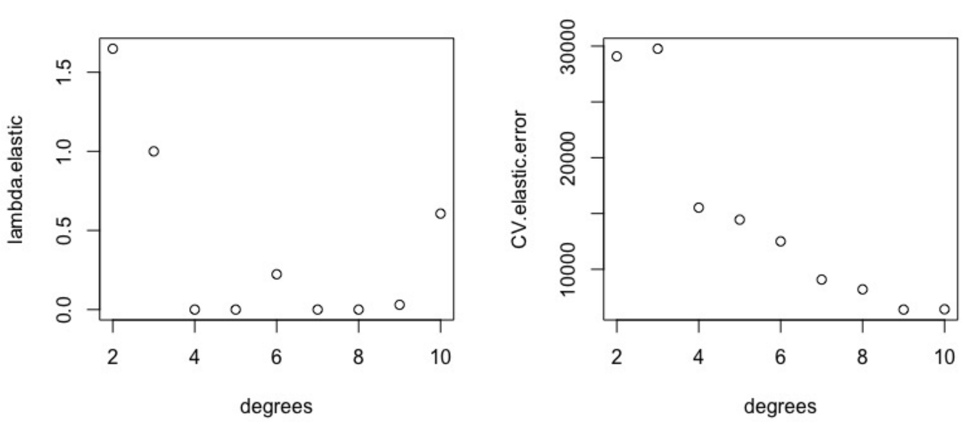


Figure B.6: Elastic Net Regression Lambda Values

B.4. Box-Jenkins

Second-order regular differencing:

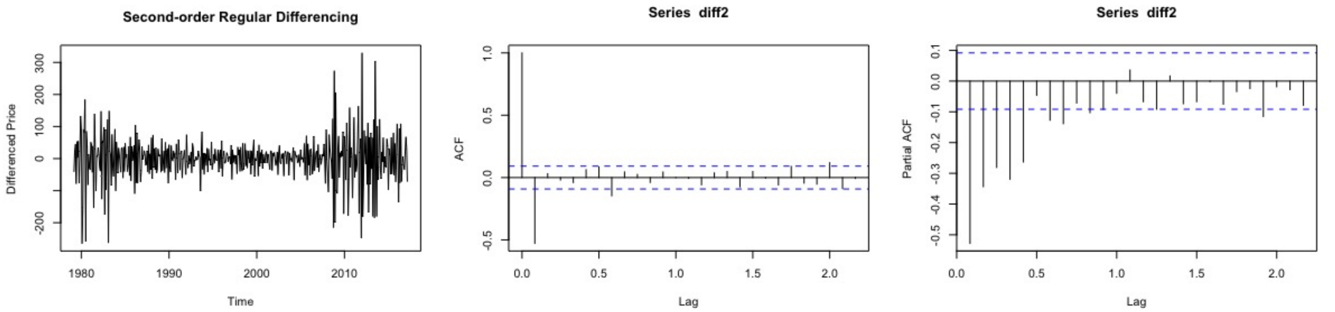


Figure B.7: Second Regular Differencing

Analysis for ARIMA(1,1,0) and ARIMA(0,1,1)

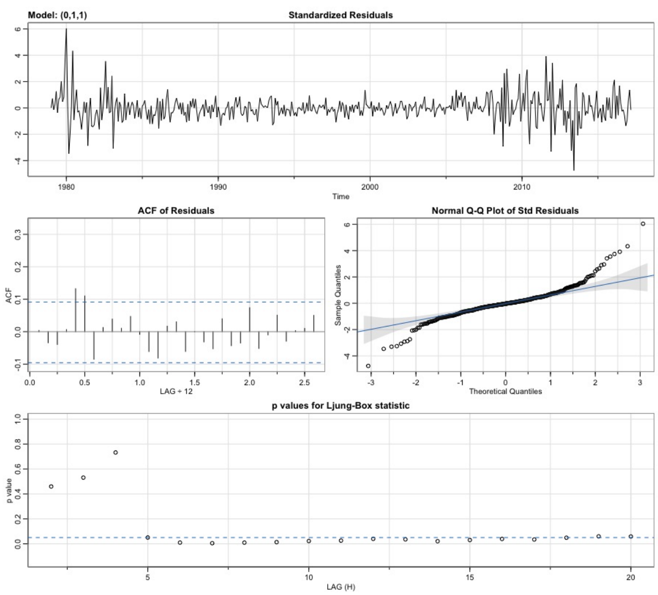


Figure B.8: ARIMA(0,1,1) on Training set

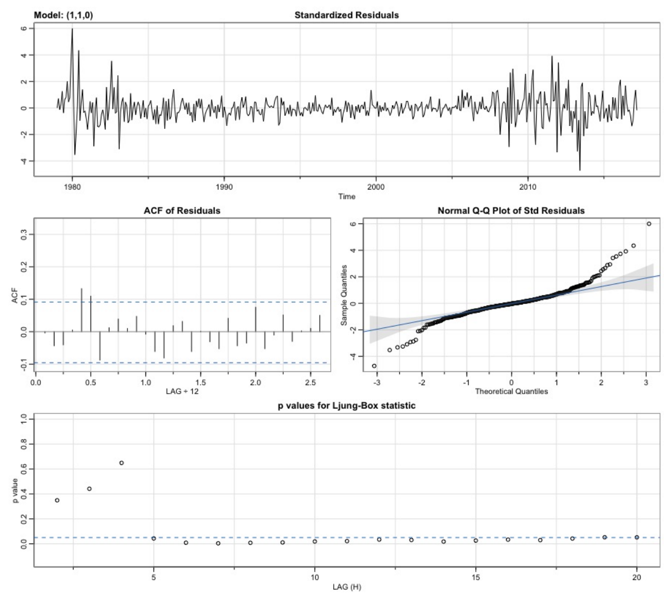


Figure B.9: ARIMA(1,1,0) on Training set

References

[1] Zhi Liu Dan & Li. "Gold Price Forecasting and Related Influence Factors Analysis Based on Random Forest". In: *The European Journal of Applied Economics* 502 (2017), pp. 711–723. DOI: 10.1007/978-981-10-1837-4_59.

[2] Ana Stokanovic Sevic Jovana & Jovancai-Stakić. "Prediction of gold price movement considering the number of infected with the Covid 19". In: *The European Journal of Applied Economics* 19 (2022), pp. 71–83. DOI: 10.5937/EJAE19-39258.

[3] CFI Team. *Swiss Franc (CHF)*. URL: https://corporatefinanceinstitute.com/resources/foreign-exchange/swiss-franc-chf/.