Modeling Amd using Time Series Methods

Dante Miller, University of Minnesota-Morris

December 16, 2021

Abstract

The purpose of the project is to create a polynomial regression, arima, and garch time series models from AMD stock data provided by yahoo finance, and determine if the models adequately capture the information from the data. By analyzing the models, we can gain a better understanding on what methods generate the best results in capturing the information from AMD stock data.

Project Background

Advanced Micro Devices is a semiconductor company that specializes in computer processors and other related technologies. The stock or security that represents ownership of a fraction of Advanced Micro Devices has increased in prices over the last couple of years with the recent price being over 100 dollars. This study uses data on the past stock closing prices across time of Advanced Micro Devices to determine if we can create models that adequately capture the datas information. This study is important in understand whether polynomial regression, arima, and garch models are capable of capturing the information of stock data. Further research can be done on exploring other machine learning models along with improving the models used in this study by manipulating the methods.

Data Explanation

The data used in this study was attained from the Yahoo Finance. There is only one variable, the closing price. The indexes of the data goes by time, and is within the range of 2020-08-01 to 2021-12-16.

Data Manipulation

During the manipulation phase, I manipulated the indexes of the data to be month based instead of daily as the daily indexes skipped a lot of days. The closing price variable from the data was transformed through the equation below:

$$diff(\log(closingprice))$$

Data Exploration

During the data exploration phase, I explored the data through summary statics, time series lot, box plot, and histogram plot. Table 1 - Table 2 give us a glimpse at the distribution of the index and closing price variable in the data through summary statistics. It can be noticed that the mean value is 94.51 and there is a variance of 383.1923. Figure 1 shows a box plot of the closing price variable, and it can be noticed that there is a lot of outliers in the data, and the majority of the data is in between 70 to 100. Figure 2 shows a histogram of the closing price variable, and it can be noticed that the data is right-skewed as majority of the data is in between 70 and 120. Figure 3 shows a time series plot of the closing price variable before transformation, and it can be noticed that it amds closing price has increased over the years while flunctuating. Figure shows a time series plot of the closing price variable after transformation.

Table 1: Summary statistics for amd.

Index	Amd
Min. :2020-08-03	Min. : 73.09
1st Qu.:2020-12-03	1st Qu.: 81.56
Median :2021-04-10	Median : 87.15
Mean :2021-04-08	Mean : 94.51
3rd Qu.:2021-08-11	3rd Qu.:103.59
Max. :2021-12-14	Max. :161.91

Table 2: Mean, variance, skewness, and kurtosis for amd.

Mean	Variance	Skewness	Kurtosis
94.51159	383.1923	1.645352	2.146588

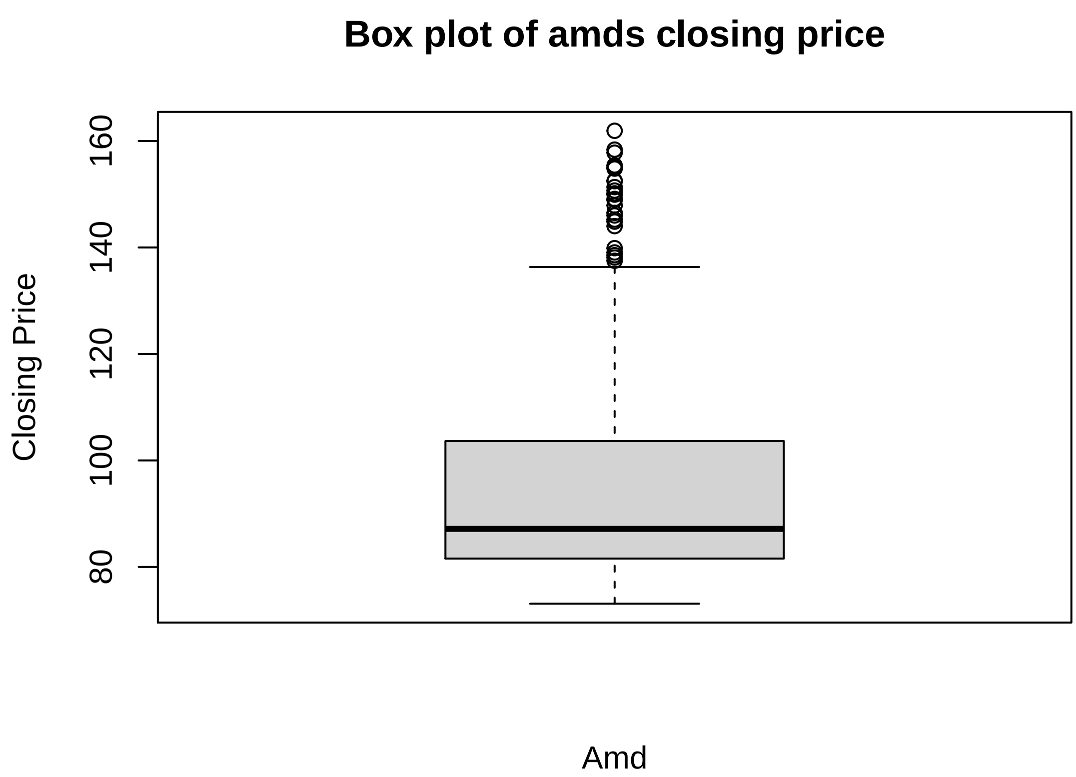

Figure 1: Box plot showing the distribution of amds closing price.

Figure 2: Histogram showing the distribution of amds closing price.

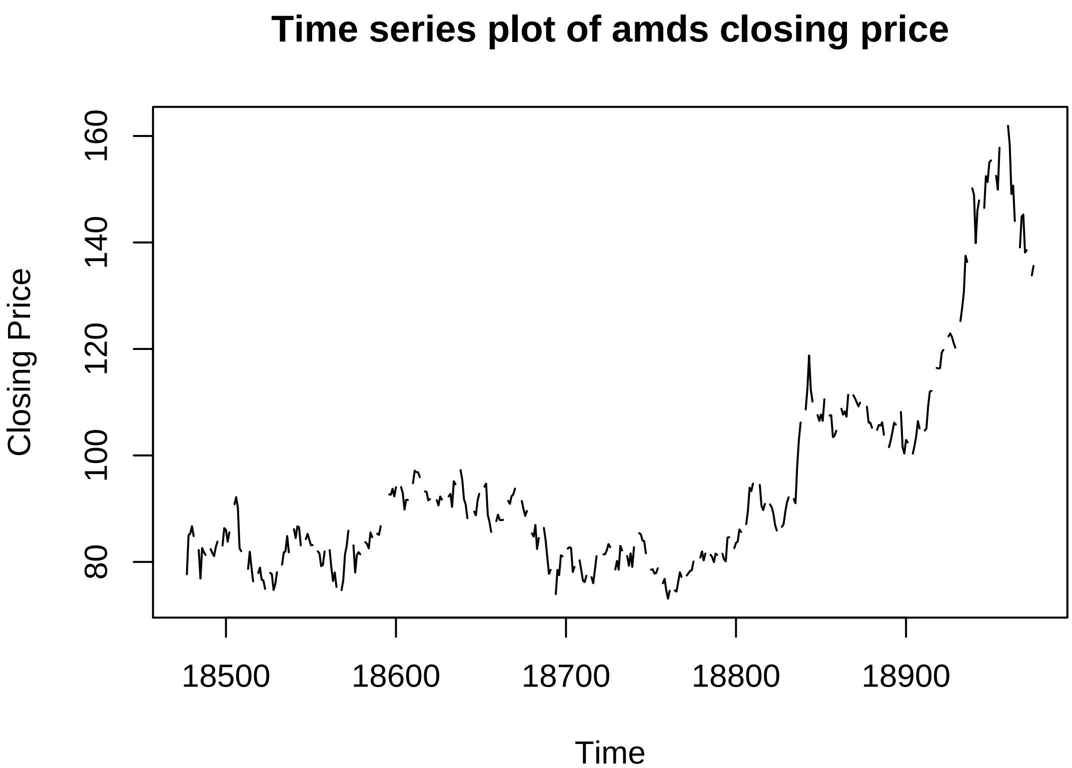

Figure 3: Time series plot showing amds closing price across time.

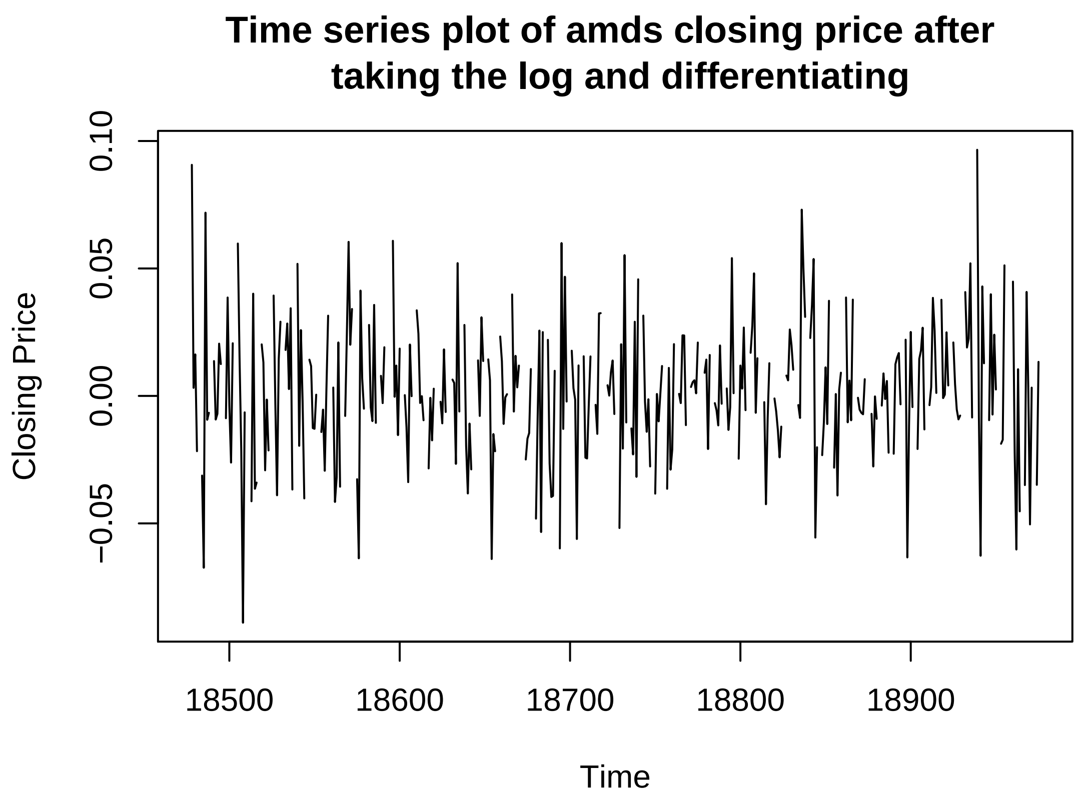

Figure 4: Time series plot of amds log and difference amds closing price.

Method

In this study, we analyzed and compared a polynomial regression, arima, and garch time series model on capture the information of the Advanced Micro Devices stock data by looking at the model diagnostics for each one. Polynomial regression is a type of regression analysis in which the independent variable and dependent variables is modeled as an nth degree polynomial in the independent variable. Arima is a type of model fitted to time series data for the purpose of understanding or forecasting future values. Garch is another type of model fitted to time series data for the purpose of understanding its variance or forecasting future variance values. To perform this study, we used the programming language R.

Result

In the model development phase, we created the three types of models using the code in the appendix. The polynomial regression is a model in the format below:

$$closingprice \sim time + time^2 + time^3 + time^4 + time^5$$

Arima is a model with the number of auto regressive terms, nonseasonal differences needed for stationary, and lagged forecast errors in the prediction equation equaled 1,0,0, respectively. Garch is a model with the how many autoregressive lags and moving lags equaled 1,1, respectively.

When looking at the Box- Ljung test for each of the models, the p-value is greater than 0.05 for them all which means that the residuals are independent. The residuals in the the time plots shown in Figure 5 - Figure 7 show that there is a lot of variation over time. In Figure 5 - Figure 6, the ACF plots do not have any significant spikes whereas Figure 7 does. The histograms in Figure 5 - 6 show that the residuals have a uni modal type shape. In Figure 8, the QQ plot of the residuals for each of the three models show that the tail of the distribution are slightly further from the mean than what would be expected if the data was actually a normal distribution. In Figure 9 - Figure 10, it can be noticed that in the residual vs fitted plot that the values are scattered around zero with quite a bit of outliers.

```
##
## Call:
## lm(formula = Amd ~ poly(time, 5, raw = TRUE), data = amd)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.090323 -0.014378 -0.001324  0.016216  0.092263
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                5.801e-03  8.939e-03   0.649    0.517
## poly(time, 5, raw = TRUE)1 -3.032e-04  5.194e-04  -0.584    0.560
## poly(time, 5, raw = TRUE)2  5.805e-06  9.269e-06   0.626    0.532
## poly(time, 5, raw = TRUE)3 -4.869e-08  6.781e-08  -0.718    0.473
## poly(time, 5, raw = TRUE)4  1.788e-10  2.159e-10   0.828    0.408
## poly(time, 5, raw = TRUE)5 -2.320e-13  2.484e-13  -0.934    0.351
##
## Residual standard error: 0.02697 on 339 degrees of freedom
## Multiple R-squared:  0.0111, Adjusted R-squared:  -0.003482
## F-statistic: 0.7613 on 5 and 339 DF,  p-value: 0.5783
```

```
## Series: amd$Amd
## ARIMA(1,0,0) with zero mean
##
## Coefficients:
##           ar1
##       -0.0609
## s.e.   0.0592
##
## sigma^2 estimated as 0.0007249:  log likelihood=757.92
## AIC=-1511.84   AICc=-1511.82   BIC=-1503.42
##
## Training set error measures:
##                      ME       RMSE        MAE     MPE     MAPE      MASE
## Training set 0.00168521 0.02688437 0.02032372 94.9945 110.4532 0.6598154
##                      ACF1
## Training set 0.0007938746
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(data = amd$Amd, dist = "std")
##
## Mean and Variance Equation:
##  data ~ garch(1, 1)
## <environment: 0x55e2297d19b0>
##  [data = amd$Amd]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu       omega      alpha1       beta1
## 1.5692e-03  9.8858e-05  7.1442e-02  7.8754e-01
```

```
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu     1.569e-03   1.395e-03    1.125    0.261
## omega  9.886e-05   7.678e-05    1.288    0.198
## alpha1 7.144e-02   4.134e-02    1.728    0.084 .
## beta1  7.875e-01   1.335e-01    5.897  3.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  761.6222    normalized:  2.207601
##
## Description:
##  Thu Dec 16 21:35:02 2021 by user:
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  9.501242  0.008646325
##  Shapiro-Wilk Test  R    W      0.9915452 0.04607617
##  Ljung-Box Test     R    Q(10)  8.009744  0.6278851
##  Ljung-Box Test     R    Q(15)  13.2074   0.5862819
##  Ljung-Box Test     R    Q(20)  17.33813  0.6309207
##  Ljung-Box Test     R^2  Q(10)  6.564478  0.7658212
##  Ljung-Box Test     R^2  Q(15)  8.894557  0.8829658
##  Ljung-Box Test     R^2  Q(20)  10.05994  0.9670719
##  LM Arch Test       R    TR^2   9.815138  0.6321746
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -4.392013 -4.347450 -4.392278 -4.374266
```

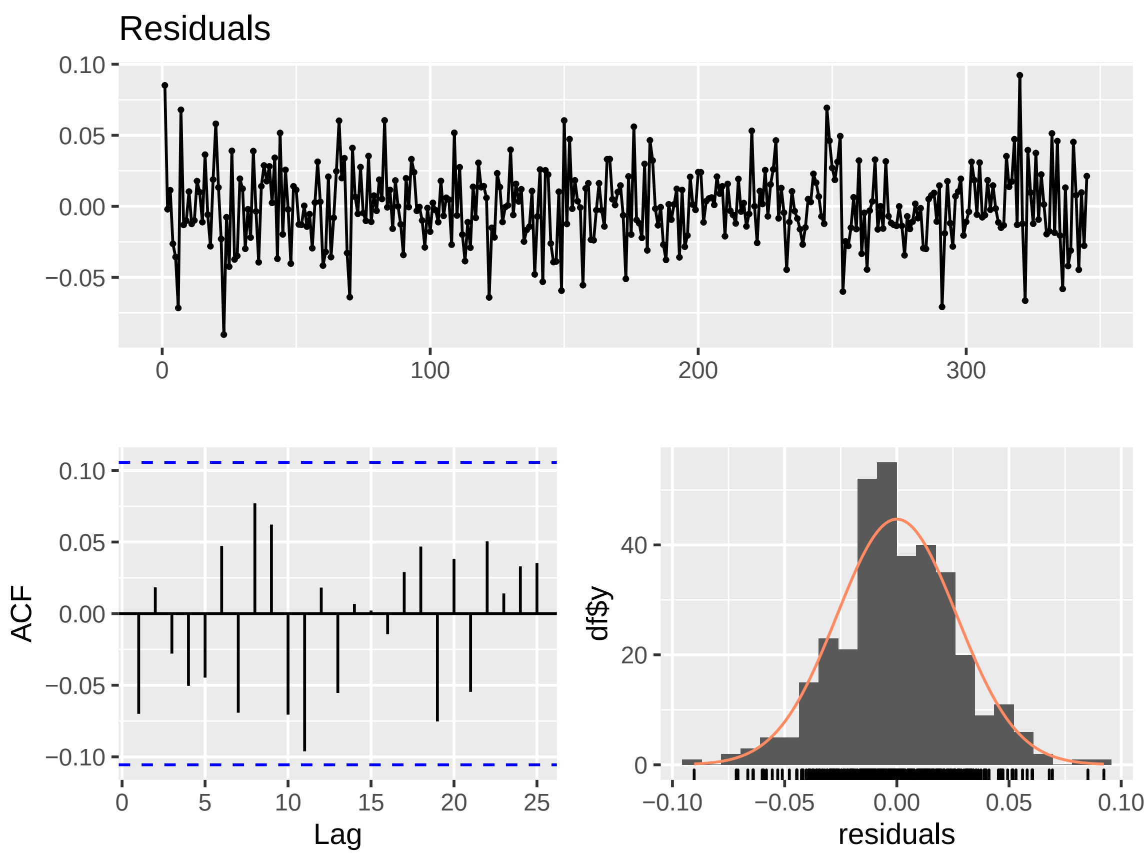

Figure 5: Model diagostics for the polynomial regression model.

```
##
##  Breusch-Godfrey test for serial correlation of order up to 10
##
## data:  Residuals
## LM test = 10.824, df = 10, p-value = 0.3714
```

```
##
##  Box-Ljung test
##
## data:  residuals(poly_amd)
## X-squared = 1.7026, df = 1, p-value = 0.1919
```

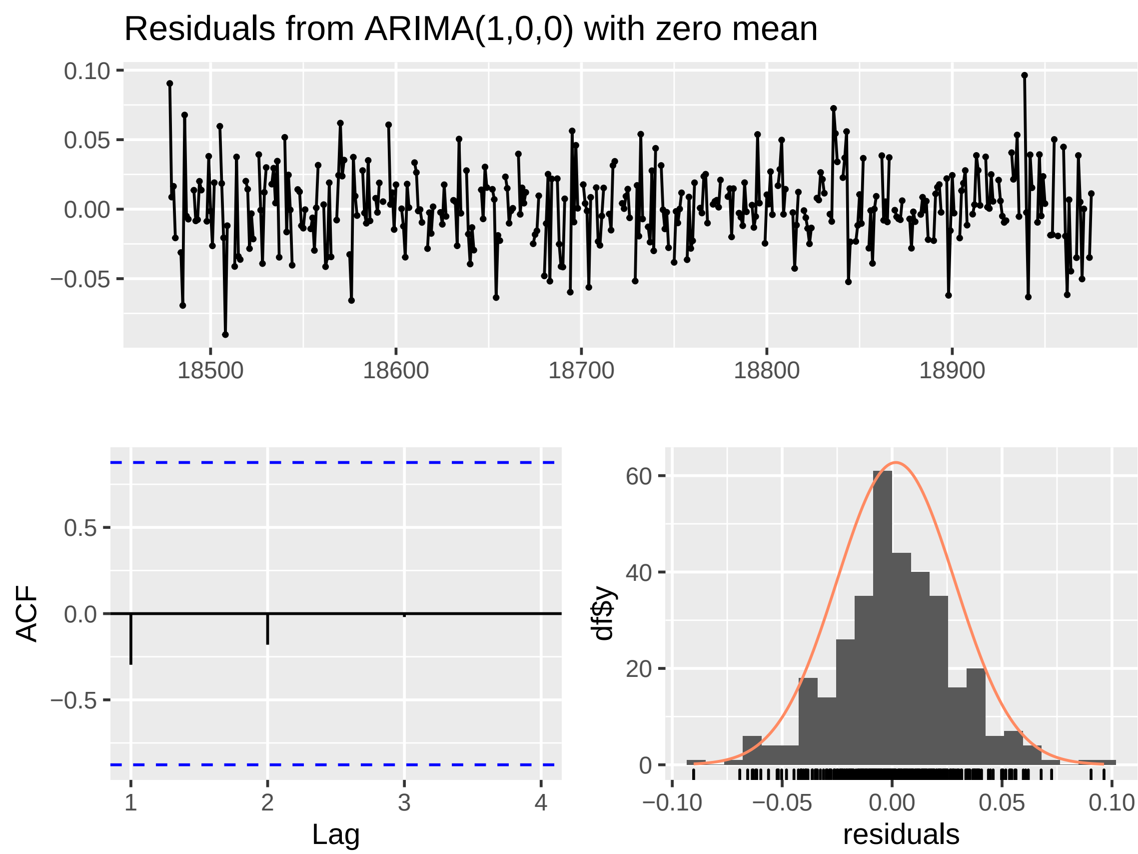

Figure 6: Model diagostics for the auto arima model.

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with zero mean
## Q* = 46.092, df = 9, p-value = 5.789e-07
##
## Model df: 1.   Total lags used: 10
```

```
##
##  Box-Ljung test
##
## data:  residuals(arima_amd)
## X-squared = 0.00021933, df = 1, p-value = 0.9882
```

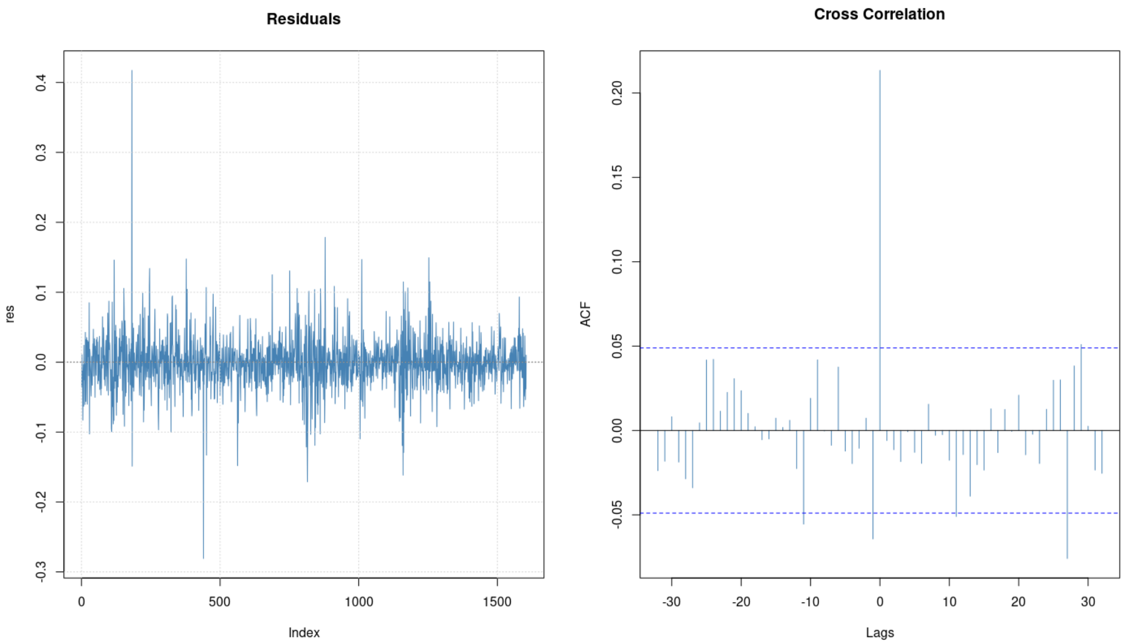

Figure 7: Model diagostics for the garch model.

```
##
##  Box-Ljung test
##
## data:  residuals(garch_amd)
## X-squared = 1.2051, df = 1, p-value = 0.2723
```

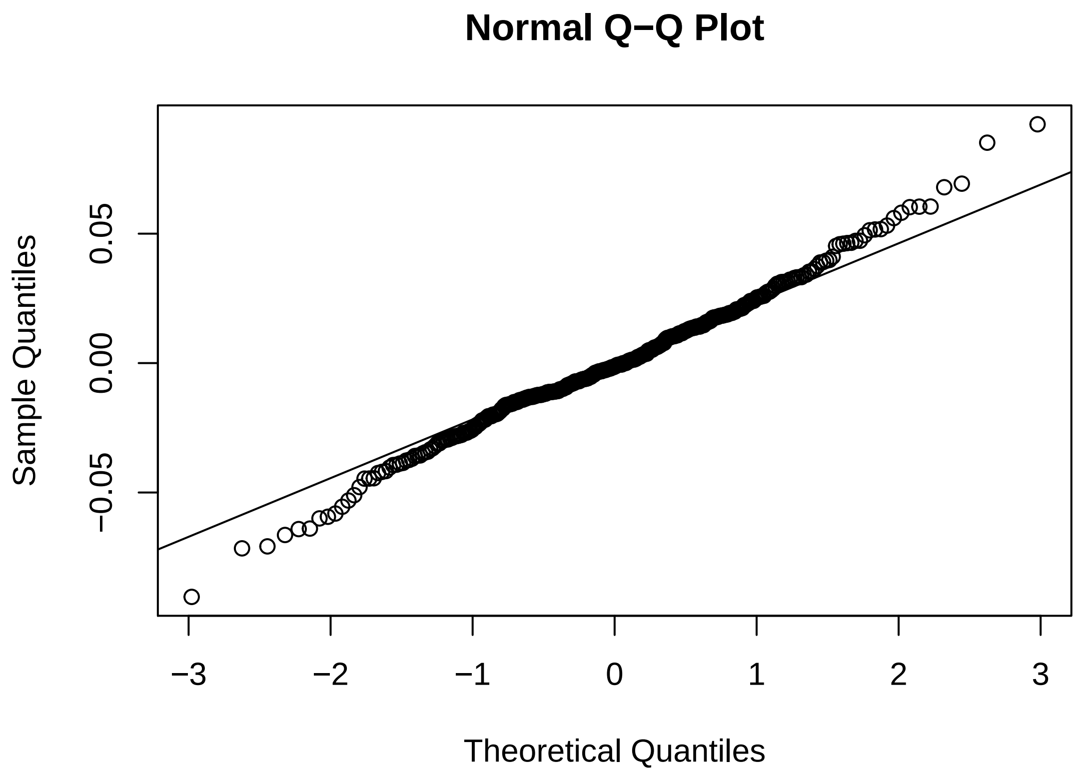

Figure 8: QQ plot of the residuals for each of the three models.

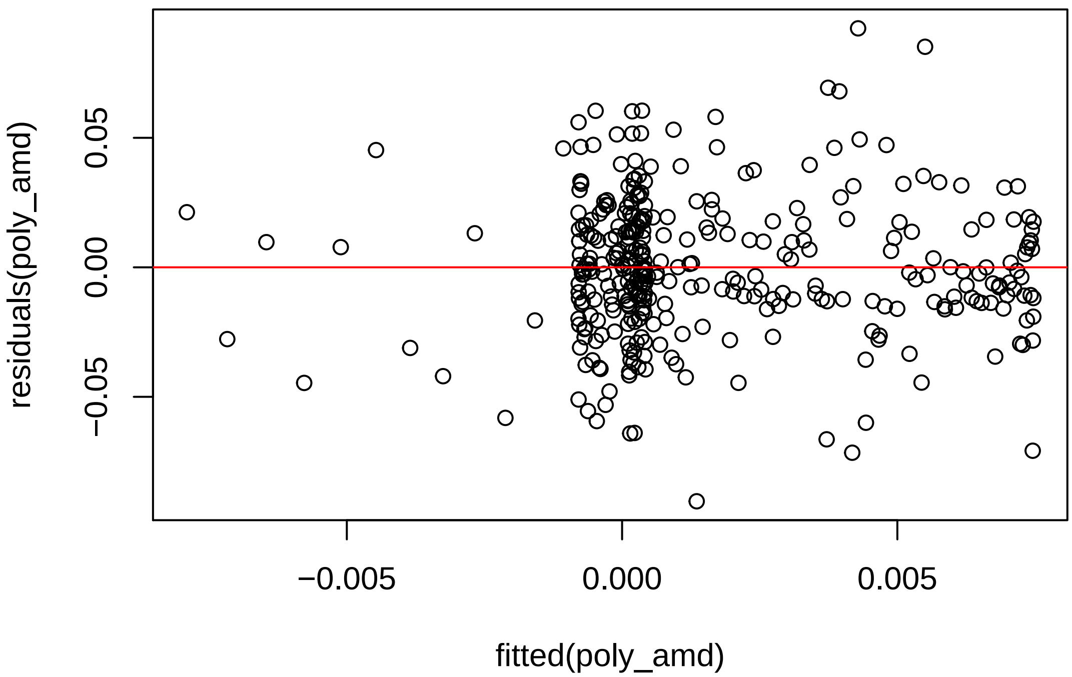

Figure 9: Residual vs fitted plot for polynomial regression model.

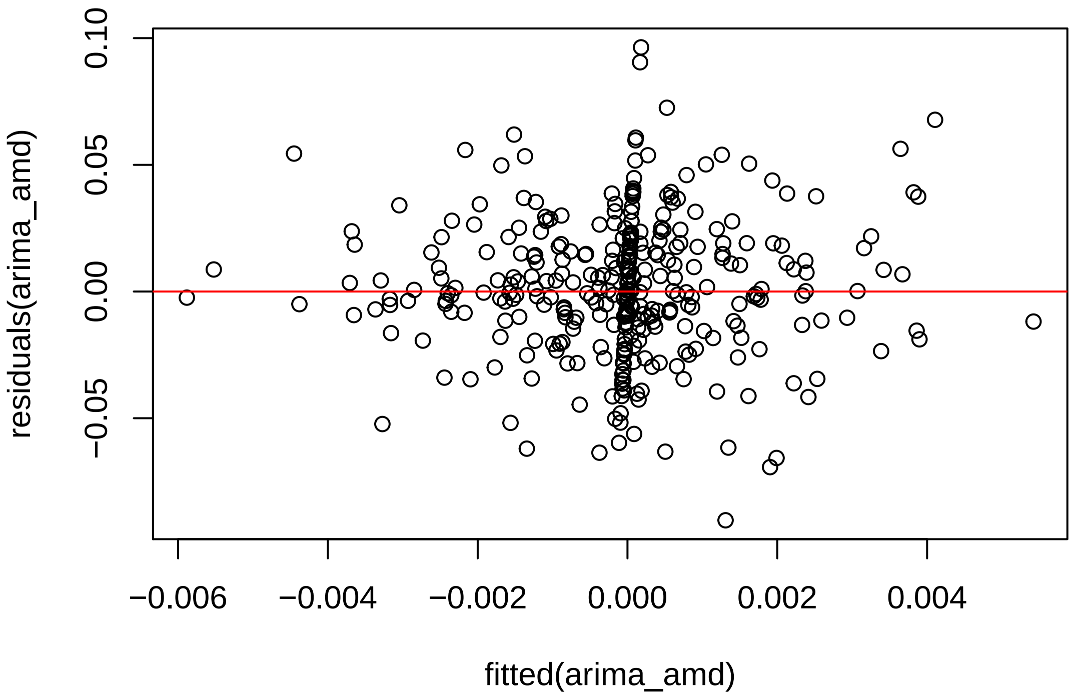

Figure 10: Residual vs fitted plot for arima model.

Conclusion

Previously, we mentioned how the residual for the models that the residuals are independent. The residuals for each of the models show large amounts of variation over time. The ACF plot for garch is shown to have significant spikes whereas arima and polynomial do not. When looking at the QQ plot, the tail of distribution for the residuals are shown to be slightly further from the mean than what is expected from a normal distribution. The residual vs fitted plots for arima and polynomial regression are shown to have outliers and be linear as the values are scattered around 0. The results from this study begged the question on whether the data needs to be reduced due to the large increase in values over time which may have an impact on the models generated in this study. Even though the results seem to be similar for each of the models, the garch model would be the best overall as it accommodates for data that has a lot of volatility which fits our scenario as the data was transformed using log and diff, making it more volatile.

Appendix

```
amd$time <-  1:length(amd$Amd)
poly_amd <- lm(Amd ~ poly(time ,5, raw=TRUE), amd)
arima_amd <- auto.arima(amd$Amd, ic = "aic")
garch_amd <- garchFit(data=amd$Amd, dist="std")
```