SEASONAL TIME SERIES

STAT485

GROUP 16:

Jane

Harkomal Kaur

Bhavana Kaushik

Yongliang Ye

Executive Summary

Problem:

In our project, we will focus on the seasonal time series and we use a dataset of the price of apple in 5 cities in Russia, but we will be focusing on the price of apples specifically in Moscow city. We are going to do a more theoretical project so we are going to use the data set of apple prices, fit the data, simulate the fitted model and then forecast the future values.

Plan:

1. Check for seasonality in the model
2. Visualize the dataset to see if the model is deterministic or stochastic
3. Seasonality
    a. If the model is deterministic, we will check the data and see if the price of apple products follow the same trend yearly so our model will be:
        $Y_t = \mu_t + X_t$
        $\mu_t = \mu_{t-12}$ + for all t
        X_t is stationary series with mean 0
    b. If the model is stochastic, we will use the multiplicative seasonal ARMA model in our project:
        Define ARIMA
        AR-autoregressive model
        I-differencing
        MA-Moving Average Model
4. Use the model to predict apple prices in Moscow for the next 2 years

Key Findings:

- Looking at time series plot and ACF plot, we can see that the model is not stationary and not much seasonality is visible in the plot just that maximum sales are in May throughout 2017-2020.
- After we apply the differencing to the data, we find out that the p-value obtained by the ADF test is lower than 0.05, which means we have enough evidence to reject the null hypothesis which stated that this model is stationary
- After comparing multiple ARIMA models, we choose the one with the least AIC which in this case is ARIMA(1,1,3)
- Prediction of apple prices in Moscow from 2020 April to 2022 April

```
$pred
Time Series:
Start = 88
End = 111
Frequency = 1
 [1] 110.6759 109.5344 109.1112 108.8668 108.7256 108.6440 108.5969 108.5697
 [9] 108.5540 108.5449 108.5396 108.5366 108.5349 108.5338 108.5333 108.5329
[17] 108.5327 108.5326 108.5325 108.5325 108.5325 108.5325 108.5325 108.5325
```

Data Description

The data we used in this project is the apple price in five cities in Russia. In our project, we are going to focus on the apple price in the city, Mosco.

The data contains 1 row, which is the monthly apple price in Mosco; and it contains 88 columns, which are months from January 2013 to March 2020.

Goal

Our goal is finding a better fitted model for the data set and use it to predict the future value of the apple prices in Moscow.

Stationary

By viewing the time series plot(figure 1), we cannot see a clearly seasonal trend, except the seasonalities that are kindly shown from 2017 to 2020 as the highest apple price is around May each year.

We form an ACF plot and by viewing the ACF plot(figure 2), we can see the seasonal trend is kindly shown, and most of the points are outside of the blue line; therefore, we can determine that the data is not stationary. Also, from the ADF test(figure 3), we can see that the p-value is 0.06447, which is larger than 0.05, so we do not have enough evidence to reject the null hypothesis, which also means that our data is not stationary. Therefore, we need to do a differencing on our data.

Differencing

As the data is not stationary, we are going to differentiate the data.

After we apply the first order differencing to the data, we find out that the p-value obtained by the ADF test(figure 4) is lower than 0.05, which means we have enough evidence to reject the null hypothesis which states that this model is stationary, and the data after differencing can be considered as stationary.

By viewing the ACF(figure 5) and PACF(figure 6) of the data after differencing,we can predict the fitted ARIMA model may be ARIMA(1,1,3).

Also by viewing the EACF(figure 7)of the data after differencing, ARIMA(1,1,3) or ARIMA(4,1,2) may be a good fitted model.

Model Comparison

By fitting several models and comparing their AIC as seen below, we found that the model ARIMA(1,1,3) had the smallest AIC, which means ARIMA(1,1,3) fits better than other fitted models.

ARIMA(1,1,1)

```
Call:
arima(x = yy, order = c(1, 1, 1), method = "ML")

Coefficients:
         ar1     ma1
      0.2733  0.4147
s.e.  0.1382  0.1120

sigma^2 estimated as 24.37:  log likelihood = -259.58,  aic = 523.16
```

ARIMA(2,1,1)

```
Call:
arima(x = yy, order = c(2, 1, 1), method = "ML")

Coefficients:
         ar1      ar2      ma1
      1.2853  -0.6450  -0.7789
s.e.  0.1029   0.0792   0.1156

sigma^2 estimated as 21.03:  log likelihood = -253.54,  aic = 513.08
```

ARIMA(4,1,2)

```
Call:
arima(x = yy, order = c(4, 1, 2), method = "ML")

Coefficients:
         ar1     ar2      ar3     ar4     ma1      ma2
      0.5760  0.4434  -0.6864  0.2345  0.0296  -0.7816
s.e.  0.1832  0.1804   0.1177  0.1214  0.1536   0.1276

sigma^2 estimated as 20.07:  log likelihood = -251.59,  aic = 515.19
```

ARIMA(1,1,3)

```
Call:
arima(x = yy, order = c(1, 1, 3), method = "ML")

Coefficients:
         ar1     ma1      ma2      ma3
      0.5776  0.0366  -0.3770  -0.4243
s.e.  0.1347  0.1351   0.0919   0.0896

sigma^2 estimated as 20.29:  log likelihood = -252.01,  aic = 512.01
```

ARIMA(1,0,0)

```
Call:
arima(x = yy, order = c(1, 0, 0), method = "ML")

Coefficients:
         ar1  intercept
      0.9225     97.987
s.e.  0.0392      7.335

sigma^2 estimated as 35.69:  log likelihood = -279.91,  aic = 563.81
```

Check Residual

From the histogram of the residuals of the fitted model(figure 13), we can see that the standard residuals of the fitted model is nearly normally distributed.

From this Q-Q plot(figure 14), even though one or two outliers occur, we can still determine that it is normal distribution.

From the ACF plot(figure 15), we can see that there is a significant spike at lag 15; however, we think it can be ignored as it is at lag 15 and not very large.

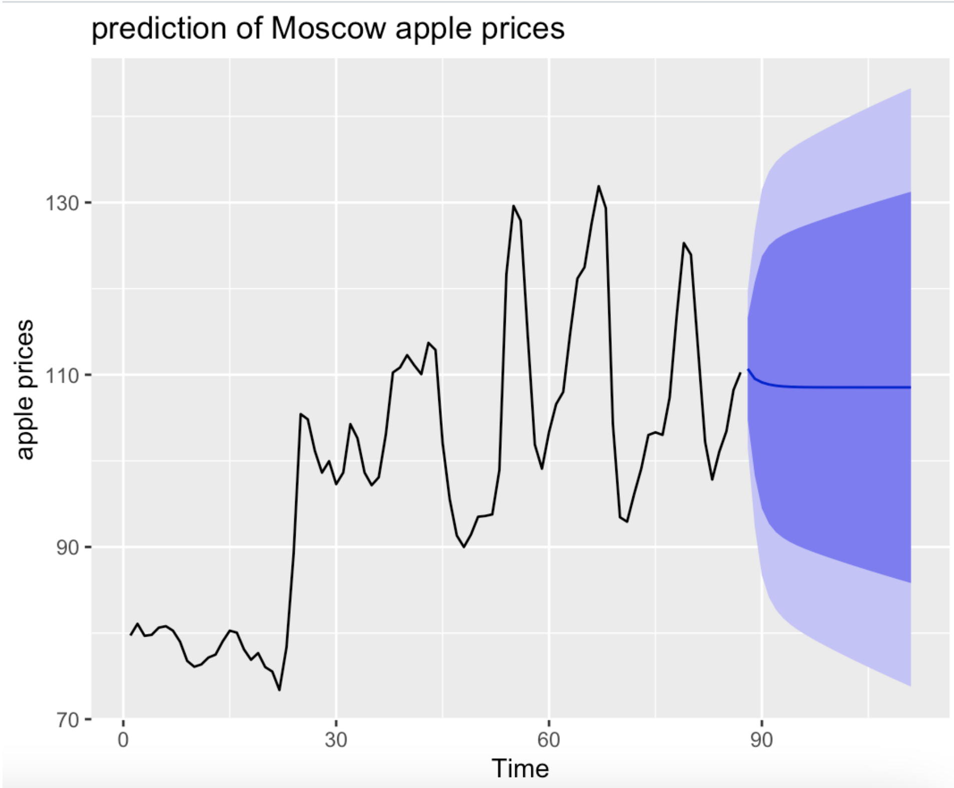
Estimation

This table is the prediction we make for the next 24 months of the apple prices in Mosco.

```
$pred
Time Series:
Start = 88
End = 111
Frequency = 1
 [1] 110.6759 109.5344 109.1112 108.8668 108.7256 108.6440 108.5969 108.5697
 [9] 108.5540 108.5449 108.5396 108.5366 108.5349 108.5338 108.5333 108.5329
[17] 108.5327 108.5326 108.5325 108.5325 108.5325 108.5325 108.5325 108.5325

$se
Time Series:
Start = 88
End = 111
Frequency = 1
 [1]  4.612654  8.758783 11.429692 12.609972 13.279042 13.729782 14.075509
 [8] 14.366327 14.626473 14.868433 15.098927 15.321688 15.538846 15.751642
[15] 15.960824 16.166857 16.370044 16.570595 16.768667 16.964378 17.157831
[22] 17.349112 17.538298 17.725459
```

And this is the graphical prediction.



We think the prediction is good as the 95% confidence interval included all the prices in the previous three or four years. If no serious event happens in Moscow or Russia, for example, we can see there is apple price soar in 2015 by weak ruble of import from Europe (The Moscow Time,2015), our forecast range is reliable.

CONCLUSION

After fitting a model of our dataset which was a dataset of apple prices in Moscow from 2013 January to 2020 March, we found that our dataset was not stationary so we did differencing to find a better fit for our model. We compared multiple ARIMA models as shown in Figures 8-12 in the appendix and as seen, the best fitted model for the dataset is ARIMA(1,1,3) as it has the lowest AIC. Then, we use ARIMA(1,1,3) to forecast the next 24 months' apple prices as shown by figures 16,17, which seems acceptable.

APPENDIX

STATIONARY

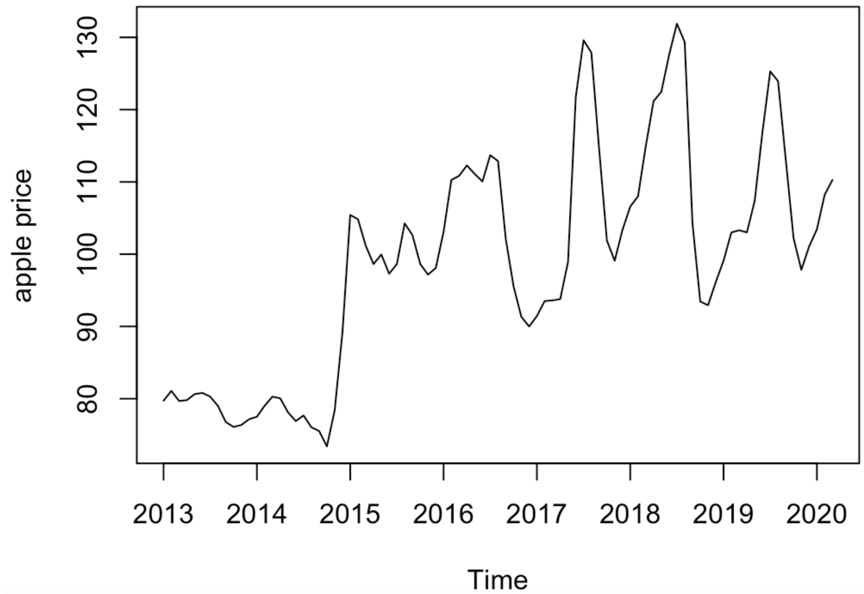


FIGURE 1. TIME SERIES PLOT OF APPLE PRICES DATASET

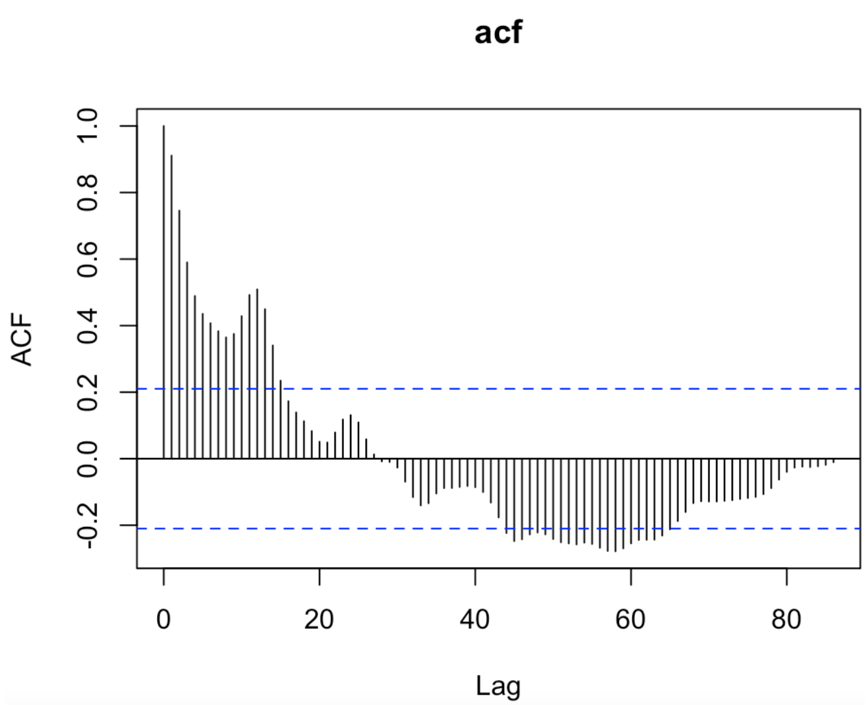


FIGURE 2. AUTOCORRELATION FUNCTION PLOT OF THE APPLE DATASET

```
	Augmented Dickey-Fuller Test

data:  df.ts
Dickey-Fuller = -3.3756, Lag order = 4, p-value = 0.06447
alternative hypothesis: stationary
```

FIGURE 3. AUGMENTED DICKERY FULL TEST ON APPLE DATASET

DIFFERENCING

```
	Augmented Dickey-Fuller Test

data:  diffyy
Dickey-Fuller = -5.2802, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diffyy) : p-value smaller than printed p-value
```

FIGURE 4. AUGMENTED DICKERY FULL TEST OF APPLE DATASET AFTER DIFFERENCING

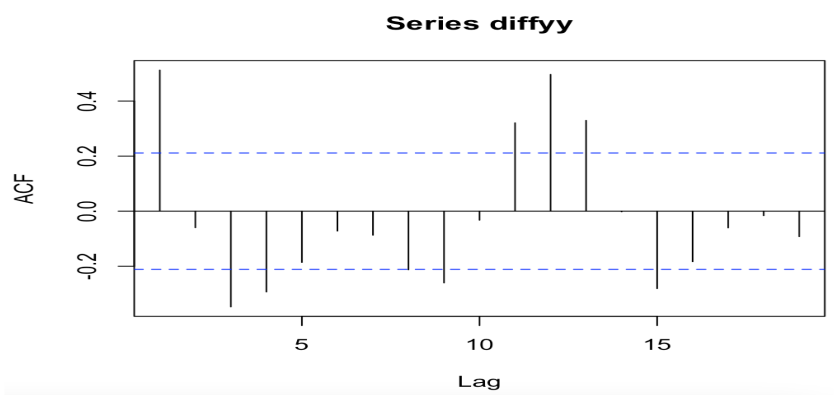


FIGURE 5. AUTOCORRELATION FUNCTION of APPLE DATA SET AFTER DIFFERENCING

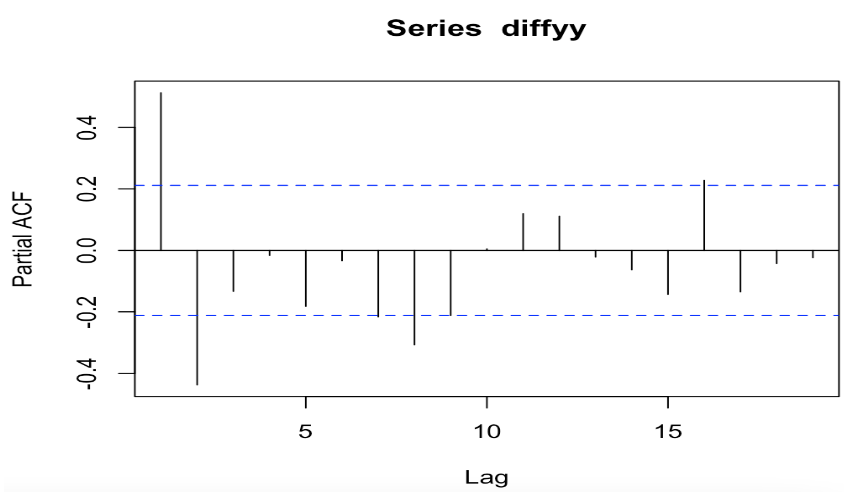


FIGURE 6. PARTIAL AUTOCORRELATION FUNCTION OF APPLE DATASET AFTER DIFFERENCING

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

FIGURE 7. EXTENDED AUTOCORRELATION FUNCTION OF APPLE DATASET AFTER DIFFERENCING

ARIMA MODELS

```
Call:
arima(x = yy, order = c(1, 1, 1), method = "ML")

Coefficients:
         ar1     ma1
      0.2733  0.4147
s.e.  0.1382  0.1120

sigma^2 estimated as 24.37:  log likelihood = -259.58,  aic = 523.16
```

FIGURE 8. ARIMA(1,1,1) MODEL

```
Call:
arima(x = yy, order = c(2, 1, 1), method = "ML")

Coefficients:
         ar1      ar2      ma1
      1.2853  -0.6450  -0.7789
s.e.  0.1029   0.0792   0.1156

sigma^2 estimated as 21.03:  log likelihood = -253.54,  aic = 513.08
```

FIGURE 9. ARIMA(2,1,1) MODEL

```
Call:
arima(x = yy, order = c(4, 1, 2), method = "ML")

Coefficients:
         ar1     ar2      ar3     ar4     ma1      ma2
      0.5760  0.4434  -0.6864  0.2345  0.0296  -0.7816
s.e.  0.1832  0.1804   0.1177  0.1214  0.1536   0.1276

sigma^2 estimated as 20.07:  log likelihood = -251.59,  aic = 515.19
```

FIGURE 10. ARIMA(4,1,2) MODEL

```
Call:
arima(x = yy, order = c(1, 1, 3), method = "ML")

Coefficients:
         ar1     ma1      ma2      ma3
      0.5776  0.0366  -0.3770  -0.4243
s.e.  0.1347  0.1351   0.0919   0.0896

sigma^2 estimated as 20.29:  log likelihood = -252.01,  aic = 512.01
```

FIGURE 11. ARIMA(1,1,3) MODEL

```
Call:
arima(x = yy, order = c(1, 0, 0), method = "ML")

Coefficients:
         ar1  intercept
      0.9225     97.987
s.e.  0.0392      7.335

sigma^2 estimated as 35.69:  log likelihood = -279.91,  aic = 563.81
```

FIGURE 12. ARIMA(1,1,0) MODEL

RESIDUALS

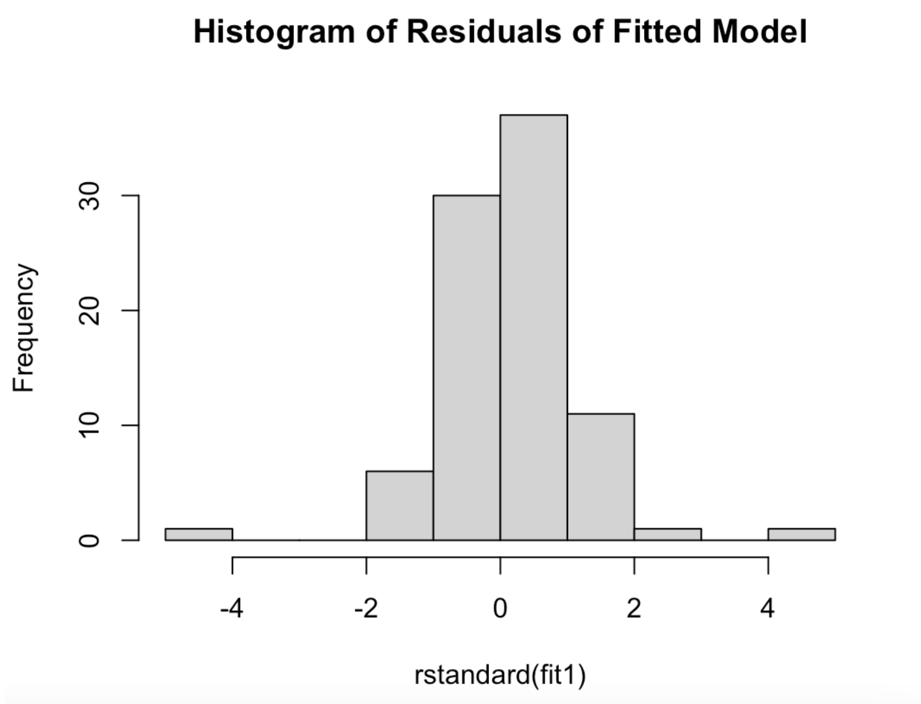


FIGURE 13. HISTOGRAM OF RESIDUALS OF ARIMA(1,1,3) FITTED MODEL

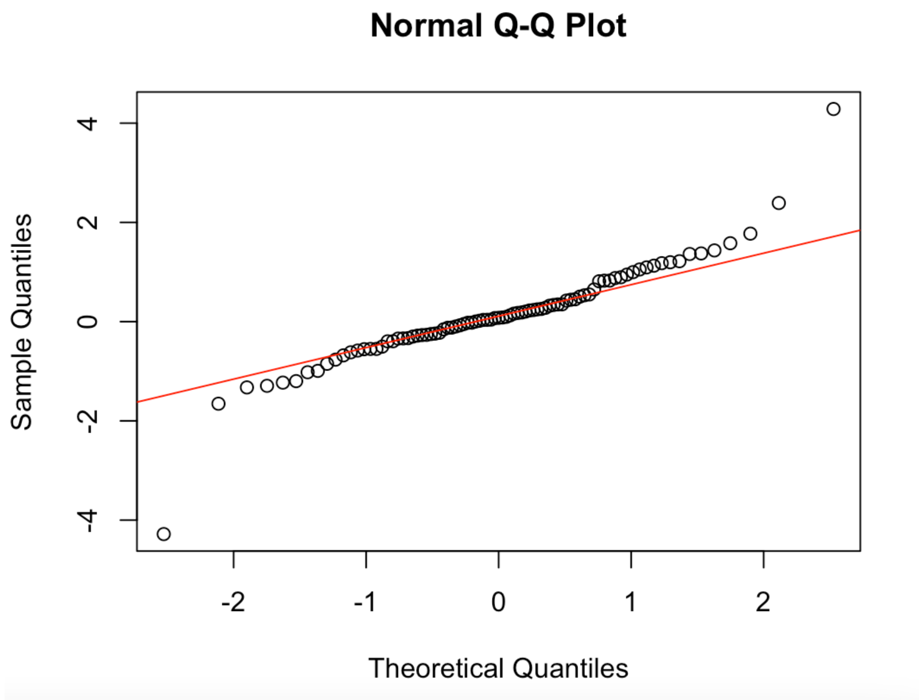


FIGURE 14.NORMAL Q-Q PLOT OF ARIMA(1,1,3) FITTED MODEL

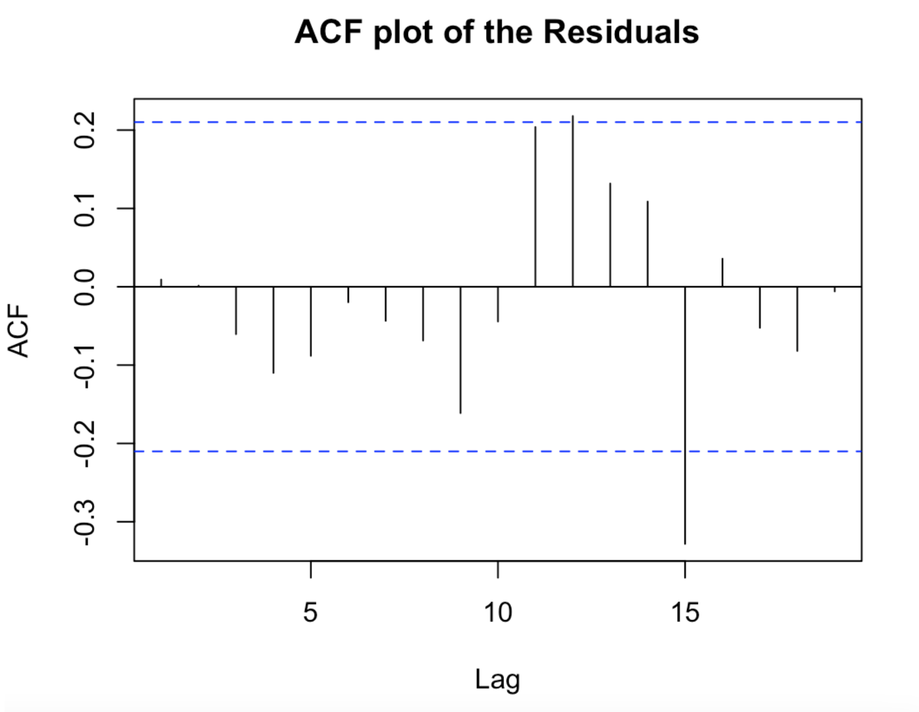


FIGURE 15. AUTOCORRELATION FUNCTION PLOT OF THE RESIDUALS OF ARIMA(1,1,3) FITTED MODEL

<u>ESTIMATION</u>

```
$pred
Time Series:
Start = 88
End = 111
Frequency = 1
 [1] 110.6759 109.5344 109.1112 108.8668 108.7256 108.6440 108.5969 108.5697
 [9] 108.5540 108.5449 108.5396 108.5366 108.5349 108.5338 108.5333 108.5329
[17] 108.5327 108.5326 108.5325 108.5325 108.5325 108.5325 108.5325 108.5325

$se
Time Series:
Start = 88
End = 111
Frequency = 1
 [1]  4.612654  8.758783 11.429692 12.609972 13.279042 13.729782 14.075509
 [8] 14.366327 14.626473 14.868433 15.098927 15.321688 15.538846 15.751642
[15] 15.960824 16.166857 16.370044 16.570595 16.768667 16.964378 17.157831
[22] 17.349112 17.538298 17.725459
```

FIGURE 16. PREDICTED VALUE OF APPLE PRICES FOR THE NEXT 24 MONTHS(APRIL 2020-APRIL 2022)

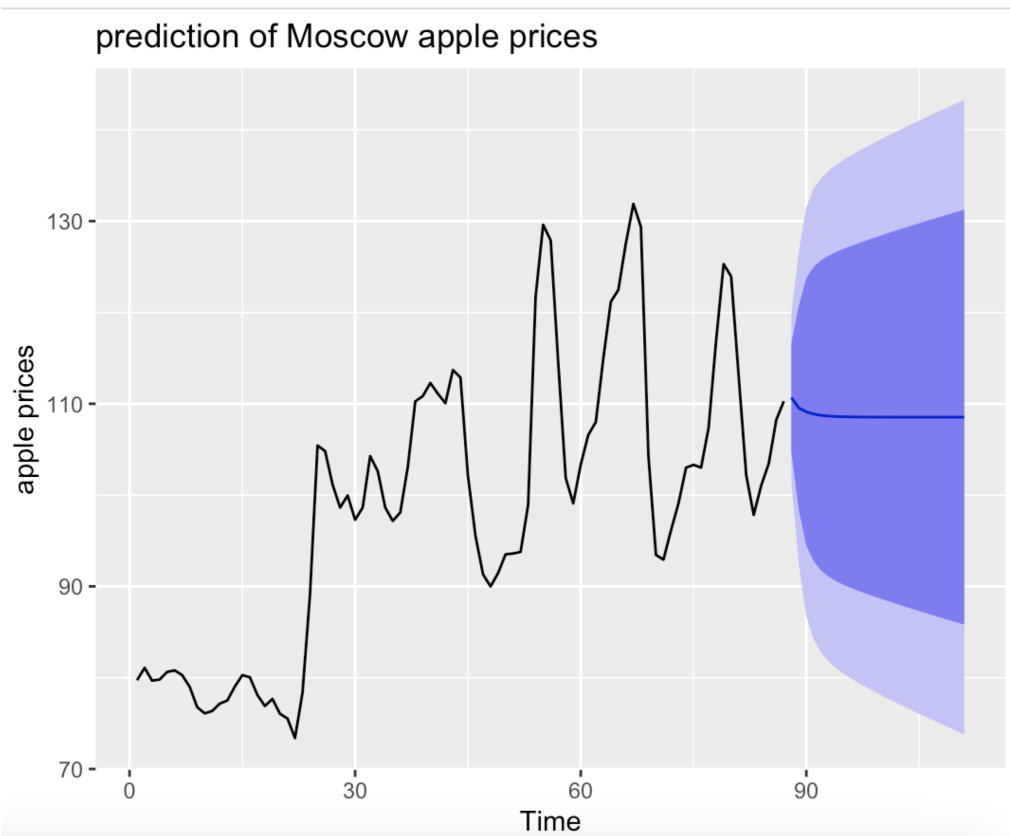


FIGURE 17. GRAPHICAL REPRESENTATION OF PREDICTION FOR THE NEXT 24 MONTHS(APRIL 2020-APRIL 2022)

Citation

-GeeksforGeeks. (2022, March 22). *How to Perform an Augmented Dickey-Fuller Test in R*. https://www.geeksforgeeks.org/how-to-perform-an-augmented-dickey-fuller-test-in-r/

-Times, T. M. (2015, October 14). *Weak Ruble Sees Fruit Prices Soar in Russia*. The Moscow Times. https://www.themoscowtimes.com/2015/10/14/weak-ruble-sees-fruit-prices-soar-in-russia-a50252

-https://www.kaggle.com/datasets/kapatsa/apple-prices-in-russian-regions?select=apples_ts.csv