DNSC 6319
TIME SERIES FORECASTING FOR ANALYTICS

GROUP 4:
Abdul Haleem Abdul Salam
Meghana Nekkanti
Jinning Zhang
Zhixing Liu
Yidi Wu

Section 1: Introduction and Overview

The dataset we are using is the Shanghai License Plate Price Dataset. The time series data spans from January 2002 to January 2019, consisting of monthly averages for plate prices. The target variable is the average price of the license plate over the years. Additional variables comprise the total count of license plates, the minimum price, and the total count of applicants. The data does not contain a Month column, so we created it. Figure 1.1 shows the dataset we are working with. The data set has 204 observations; we chose to have 24 of the observations as a hold-out sample.

```
> head(shangai)
  Total.number.of.license.issued lowest.price avg.price Total.number.of.applicants Month
1                           1400        13600     14735                       3718     1
2                           1800        13100     14057                       4590     2
3                           2000        14300     14662                       5190     3
4                           2300        16000     16334                       4806     4
5                           2350        17800     18357                       4665     5
6                           2800        19600     20178                       4502     6
```

Figure 1.1: First 5 rows of the dataset

The time series plot (Figure 1.2) reveals an overall upward trend, but there are some significant spikes and drops throughout the series. The Time component here is Month.

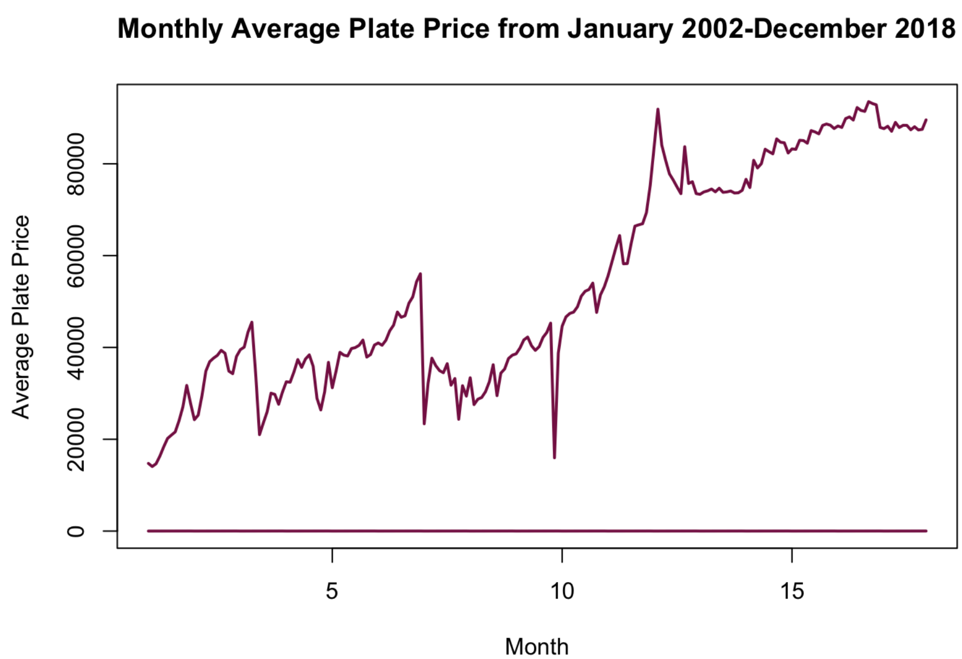

Figure 1.2: Time series plot of the average plate price

The box plots (Figure 1.3) indicate little changes over the months and no significant seasonality.

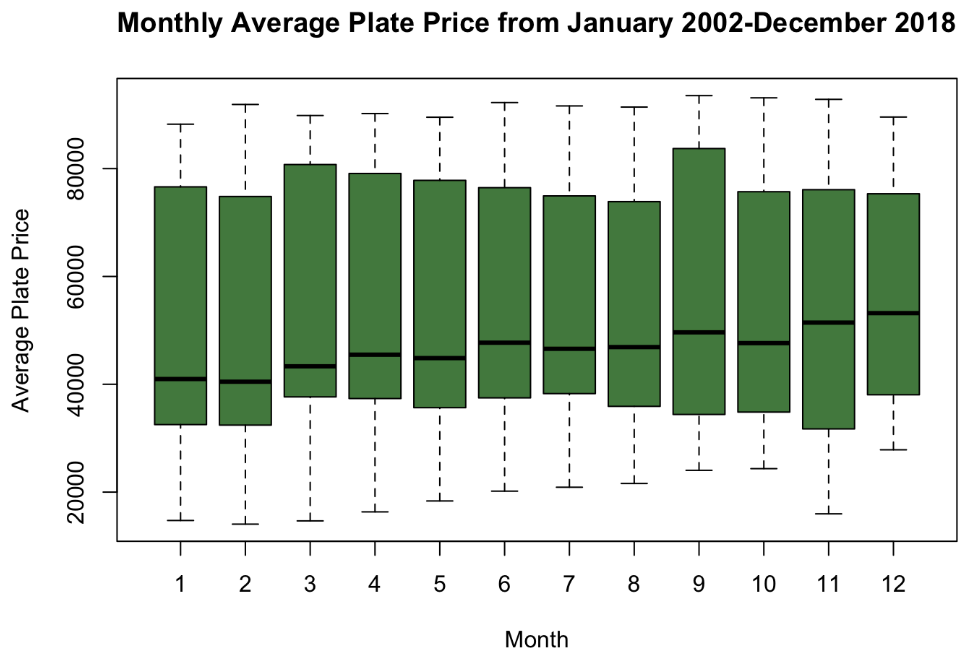

Figure 1.3: Box plot of the Average Plate Price

The periodogram (Figure 1.4) shows the significant periods in the data. There are approximately 6 significant periods. They are periods 108, 216, 54, 36, 24 and 72 with the associated harmonics 2, 1, 4, 6, 9, and 3 respectively.

```
> head(all[order(-all[, "amplitude"]), ],7)
     harmonic period   frequency  amplitude
[1,]        2  108.0 0.009259259 5793439405
[2,]        1  216.0 0.004629630 5243963756
[3,]        4   54.0 0.018518519 2079951958
[4,]        6   36.0 0.027777778  789626058
[5,]        9   24.0 0.041666667  546777583
[6,]        3   72.0 0.013888889  487658962
```

Figure 1.4: Top 6 periods with harmonics and amplitude

Figure 1.5 shows the periodogram vs the periods and figure 1.6 shows the periodogram with the frequencies.

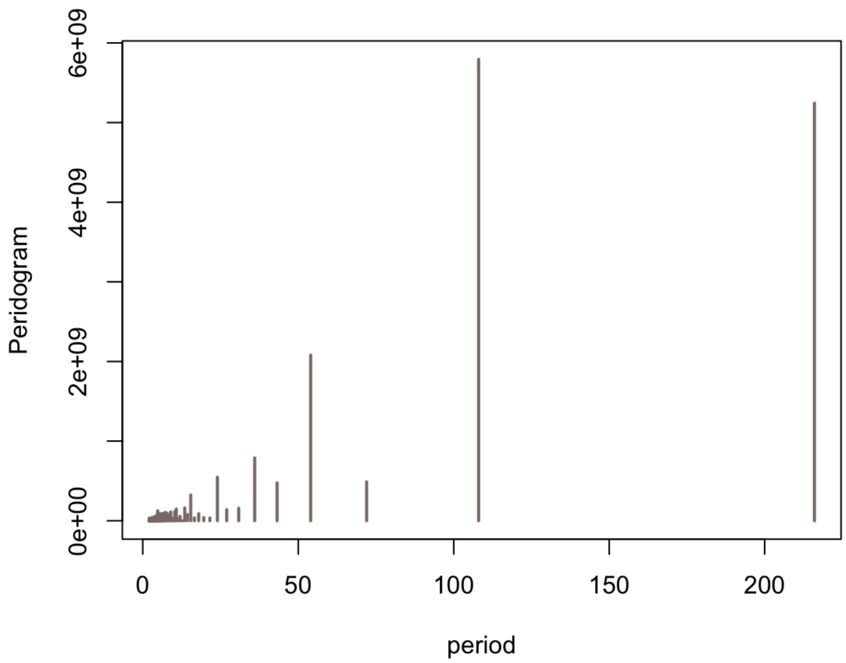

Figure 1.5: Periodogram vs periods

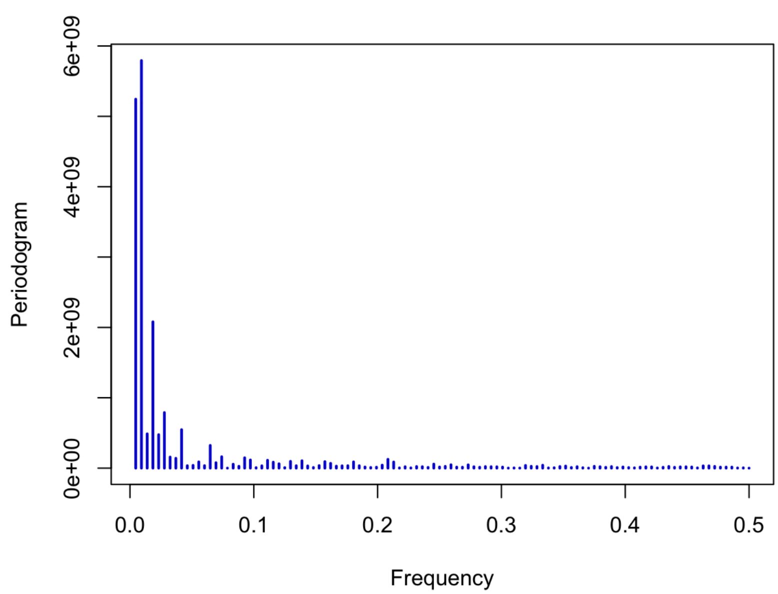

Figure 1.6: Periodogram vs frequencies

Section 2: Univariate Time-series Models

Section 2.1: Comparison of "candidate" models in terms of fit and hold-out sample

Our objective for this part was to investigate which model is appropriate for describing the series. We used three models: model without a trend, model with a trend, and a cyclic model. We have not used a seasonal model since we have no evidence of seasonality based on the box plots (Figure 1.3).

<u>*Model without trend:*</u>

```
Call:
lm(formula = n_avgprice ~ poly(time, 20) + nd1 + nd2 + nd3 +
    nd4 + nd5)

Residuals:
    Min      1Q  Median      3Q     Max
-8859.1 -1385.2    46.3  1773.5  8639.9

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)         38594       1127  34.257  < 2e-16 ***
poly(time, 20)1    421830      16766  25.160  < 2e-16 ***
poly(time, 20)2     -6768       9116  -0.743  0.45891
poly(time, 20)3      3224       7366   0.438  0.66223
poly(time, 20)4    -38303       4664  -8.213 8.15e-14 ***
poly(time, 20)5      6285       3995   1.573  0.11766
poly(time, 20)6     15720       5214   3.015  0.00301 **
poly(time, 20)7     16118       3638   4.431 1.77e-05 ***
poly(time, 20)8     -6334       3739  -1.694  0.09229 .
poly(time, 20)9    -23674       3331  -7.107 4.15e-11 ***
poly(time, 20)10    -4377       3788  -1.156  0.24961
poly(time, 20)11    10711       3400   3.150  0.00196 **
poly(time, 20)12     7976       3124   2.553  0.01166 *
poly(time, 20)13     3278       3373   0.972  0.33267
poly(time, 20)14     5201       3285   1.583  0.11538
poly(time, 20)15   -11106       3369  -3.296  0.00122 **
poly(time, 20)16   -15634       2996  -5.217 5.77e-07 ***
poly(time, 20)17     2965       3272   0.906  0.36623
poly(time, 20)18    10826       3231   3.351  0.00101 **
poly(time, 20)19    -1260       3114  -0.405  0.68634
poly(time, 20)20    -1890       3095  -0.611  0.54229
nd1                 38267       3793  10.089  < 2e-16 ***
nd2                 21193       2597   8.161 1.10e-13 ***
nd3                -29157       3082  -9.462  < 2e-16 ***
nd4                 15907       3097   5.136 8.39e-07 ***
nd5                  6391       3113   2.053  0.04176 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2962 on 154 degrees of freedom
Multiple R-squared:  0.9834,	Adjusted R-squared:  0.9807
F-statistic: 364.4 on 25 and 154 DF,  p-value: < 2.2e-16
```

Figure 2.1.1: Output for Model without Trend

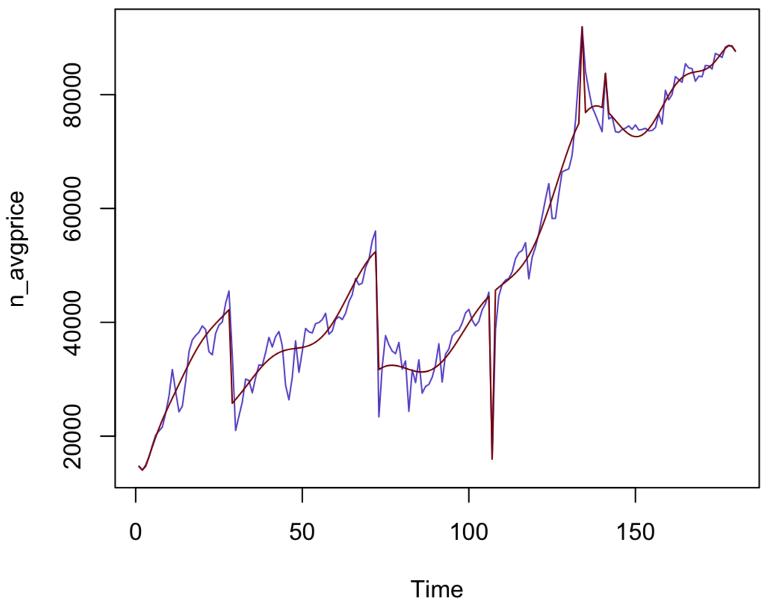

Figure 2.1.2: Actual vs predicted fit of Model w/o Trend (Purple = actual, Maroon = predicted)

The model has a high R-squared value of 0.9834 (as seen above in Figure 2.1.1), which means that the model predicts 98.34% of the data. The p-value associated is less than 0.05; this indicates that the model is statistically significant. The model also shows a good fit of the actual vs predicted series (Figure 2.1.2). However, the model has a MAPE of 113.5396, which indicates that the mean average error of the model is over 11353.96%, which could indicate a possible overfit.

<u>*Model with trend:*</u>

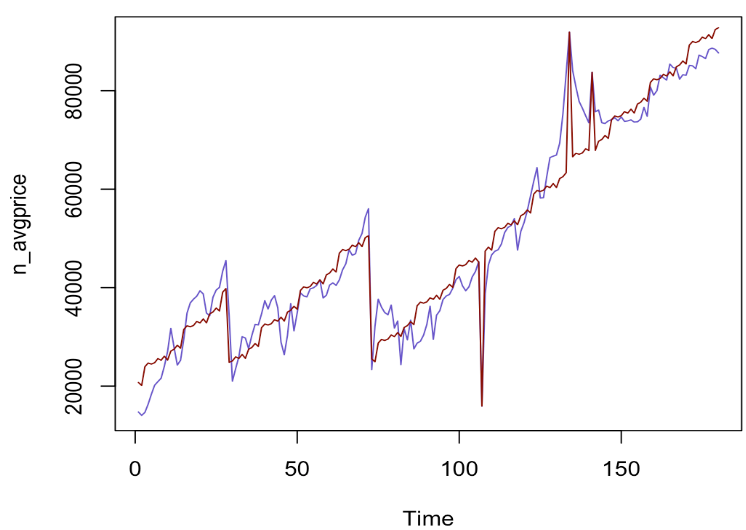

Figure 2.1.3: Actual vs predicted fit of Model with Trend (Purple = actual, Maroon = predicted)

```
Residuals:
    Min      1Q  Median      3Q     Max 
-9287.3 -3067.3  -840.6  3062.0 20217.4 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)          -20481.30    2263.43  -9.049 4.34e-16 ***
time                    630.34      14.95  42.151  < 2e-16 ***
as.factor(n_Month)2   -1232.29    1858.19  -0.663  0.50817    
as.factor(n_Month)3    1953.66    1825.98   1.070  0.28625    
as.factor(n_Month)4    2058.92    1826.29   1.127  0.26125    
as.factor(n_Month)5    1227.88    1827.70   0.672  0.50266    
as.factor(n_Month)6     865.94    1827.96   0.474  0.63634    
as.factor(n_Month)7    1065.07    1828.35   0.583  0.56102    
as.factor(n_Month)8     114.60    1828.85   0.063  0.95011    
as.factor(n_Month)9     306.59    1861.38   0.165  0.86938    
as.factor(n_Month)10  -1119.87    1830.23  -0.612  0.54148    
as.factor(n_Month)11     74.76    1865.22   0.040  0.96808    
as.factor(n_Month)12   -176.35    1832.10  -0.096  0.92344    
nd1                   40585.90    1980.83  20.489  < 2e-16 ***
nd2                   25815.42    1452.16  17.777  < 2e-16 ***
nd3                  -31069.59    5199.27  -5.976 1.41e-08 ***
nd4                   29146.34    5188.88   5.617 8.28e-08 ***
nd5                   15020.10    5189.99   2.894  0.00433 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5000 on 162 degrees of freedom
Multiple R-squared:  0.9502,	Adjusted R-squared:  0.9449 
F-statistic: 181.7 on 17 and 162 DF,  p-value: < 2.2e-16
```

Figure 2.1.4: Output for Model with Trend

The model has a high R-squared value of 0.9502 (as seen above in Figure 2.1.4), which means that the model predicts 95.02% of the data. The p-value associated is less than 0.05; this indicates that the model is statistically significant. The model also shows a good fit of the actual vs predicted series (Figure 2.1.3). The model has a MAPE of 0.1342845, which indicates that the mean average error of the model is over 13.42845%, which is relatively low and suggests a good fit.

Cyclic model:

The model has a high R-squared value of 0.9502 (as seen below in Figure 2.1.6), which means that the model predicts 95.02% of the data. The p-value associated is less than 0.05; this indicates that the model is statistically significant. The model also shows a good fit of the actual vs predicted series (Figure 2.1.5). The model has a MAPE of 0.1342845, which indicates that the mean average error of the model is over 13.42845%, which is relatively low and suggests a good fit.

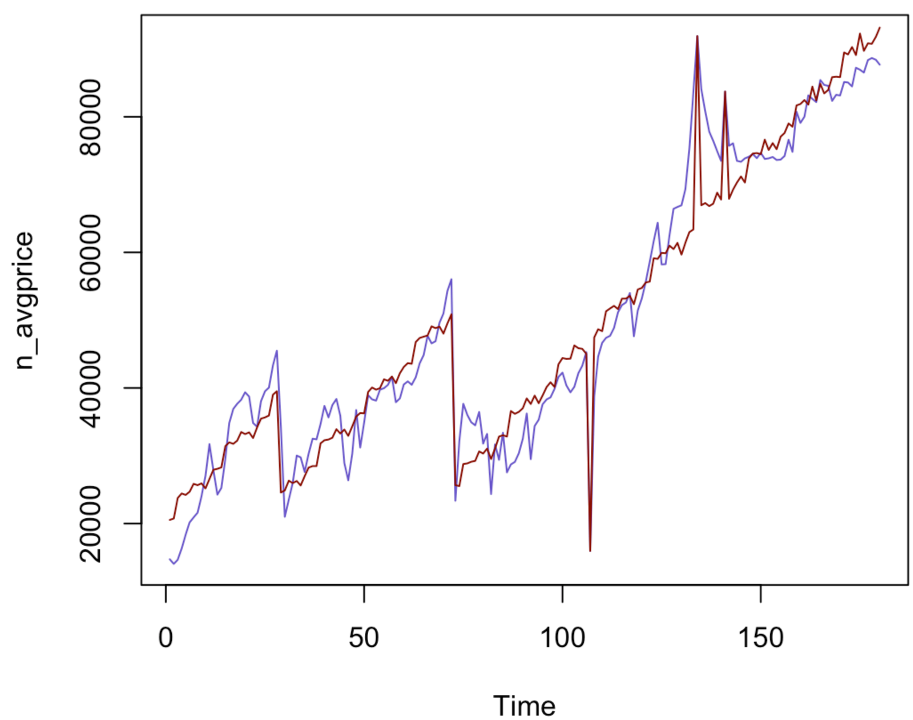

Figure 2.1.5: Actual vs predicted fit of Cyclic Model (Purple = actual, Maroon = predicted)

```
Residuals:
    Min      1Q  Median      3Q     Max 
-9110.0 -3286.3  -928.7  2948.3 20191.6 

Coefficients: (2 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.005e+04  1.941e+03 -10.330  < 2e-16 ***
time         6.304e+02  1.485e+01  42.447  < 2e-16 ***
cos1        -8.704e+02  5.275e+02  -1.650  0.10088    
sin1         5.402e+02  5.335e+02   1.013  0.31271    
cos2        -4.740e+01  5.290e+02  -0.090  0.92872    
cos4         3.274e+02  5.291e+02   0.619  0.53696    
sin4         2.062e+02  5.340e+02   0.386  0.69985    
cos6        -1.506e+02  4.484e+02  -0.336  0.73737    
sin6         1.627e+16  2.151e+16   0.756  0.45052    
cos9         6.055e+02  5.286e+02   1.145  0.25375    
sin9         2.785e+02  5.314e+02   0.524  0.60090    
cos3                NA         NA      NA       NA    
sin3                NA         NA      NA       NA    
nd1          4.058e+04  1.968e+03  20.614  < 2e-16 ***
nd2          2.581e+04  1.442e+03  17.896  < 2e-16 ***
nd3         -3.074e+04  5.119e+03  -6.006 1.19e-08 ***
nd4          2.890e+04  5.110e+03   5.655 6.77e-08 ***
nd5          1.450e+04  5.171e+03   2.803  0.00567 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4969 on 164 degrees of freedom
Multiple R-squared:  0.9502,    Adjusted R-squared:  0.9456 
F-statistic: 208.5 on 15 and 164 DF,  p-value: < 2.2e-16
```

Figure 2.1.6:Output for Cyclic Model

Section 2.2: Comparison of "candidate" models in terms of fit and hold-out sample

	Model without Trend	**Model with Trend**	**Cyclic Model**
p-value	Statistically significant	Statistically significant	Statistically significant
R-squared	0.9834	0.9502	0.9502
MAPE	113.5396	0.1342845	0.1342845
Variance estimate	2747.279	4756.639	4756.711

Table 2.2.1: Comparison of the models

Based on Table 2.2.1, we can see that both the 'Model with trend' and the 'Cyclic model' perform well and have the same MAPE and R-squared. Upon comparing square root of variance estimates we chose the 'Model with trend' as the deterministic time series model due to it being a bit lower.

Section 2.3: Looking at residuals of the model(s)

Model with trend:

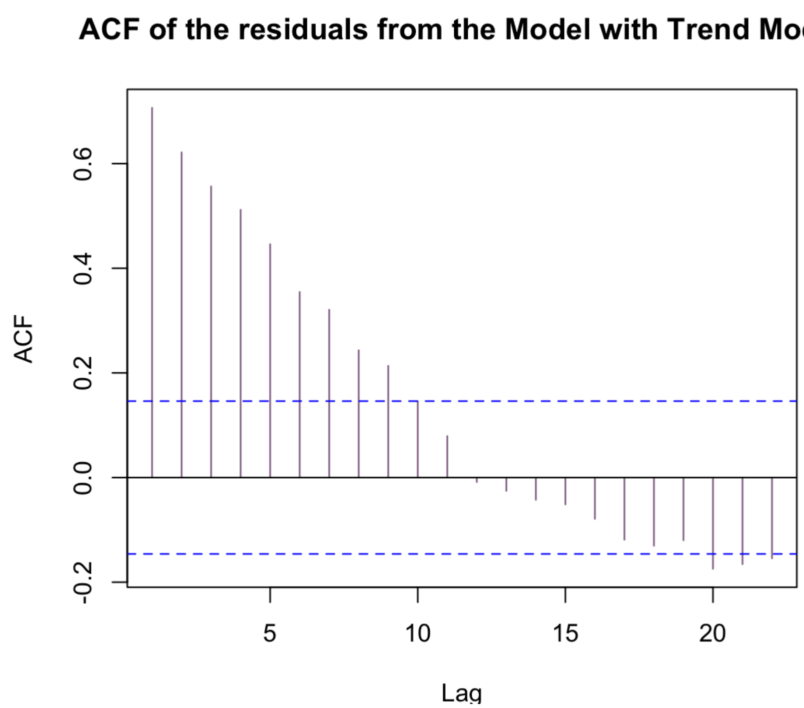

Figure 2.3.1 ACF of residuals of Model with Trend

```
        Box-Pierce test

data:  fit_wT$resid
X-squared = 378.89, df = 20, p-value < 2.2e-16
```

Figure 2.3.2 Box-Pierce test of Model with Trend

The ACF decays slowly meaning that the series is non-stationary. Based on the ACF (Figure 2.3.1) we can see that most of the lags are outside the 2 standard error bounds, this rejects the null hypothesis that the series is a white noise series. The Box-Pierce test (Figure 2.3.2), has a p-value less that 0.05, which means it also rejects the null hypothesis that the series is a white noise series.

Section 3: Time Series Regression Models

Section 3.1: Discussion of independent variables: correlation analysis and scatter plots

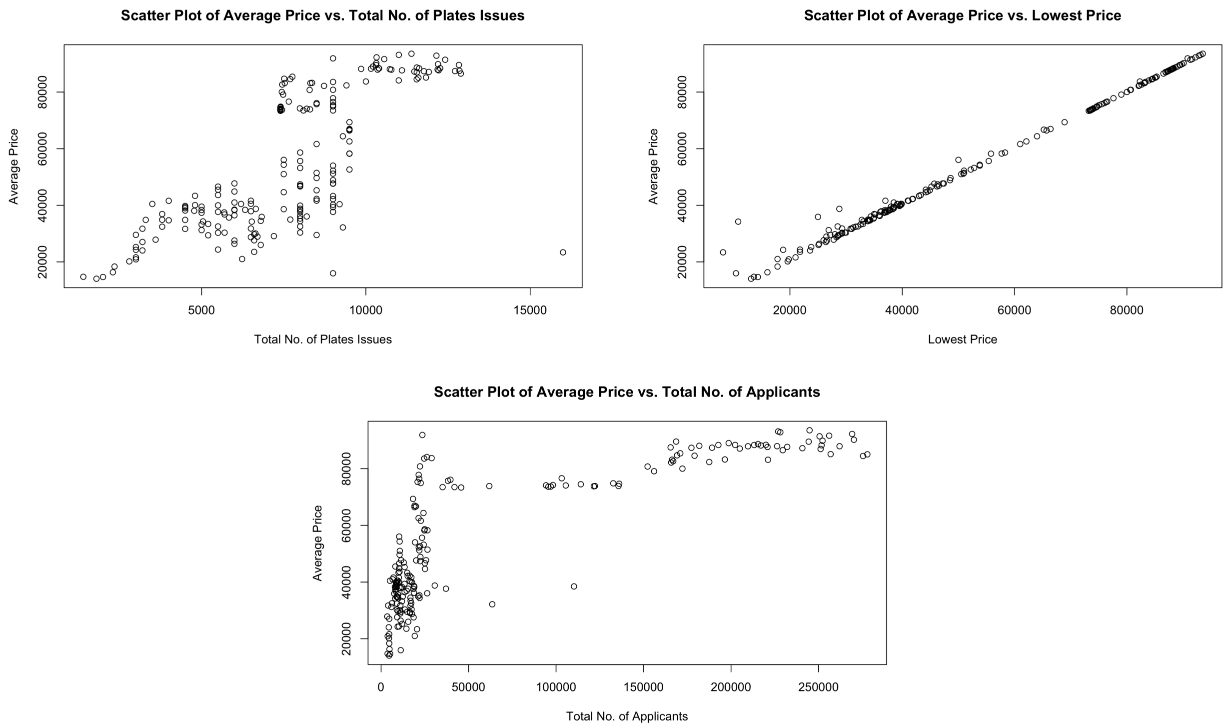

Figure 3.1.1: Scatter plots of independent variables vs target variable (Average Price)

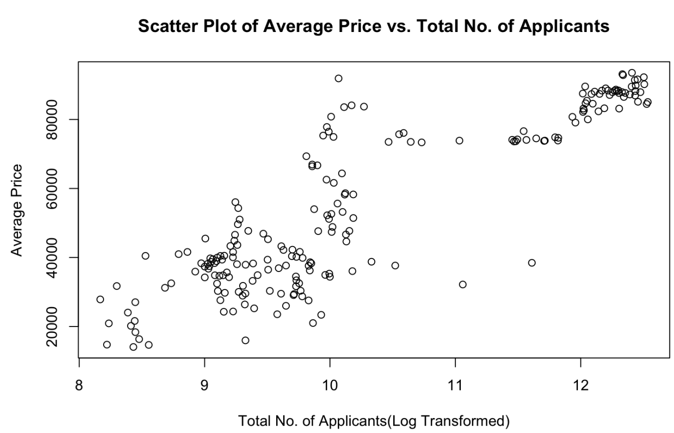

Figure 3.1.2: Scatter plot of Log(Total number of applicants) vs target variable (Average price)

From Figure 3.1.1, in the scatter plot of Average Price vs. Total number of plates issued, there is a positive correlation between the two variables, though there is also a lot of scatter in the data. This means that as the total number of plates issued increases, the average price also tends to increase.

In the scatter plot of Average Price vs. Lowest Price, there is no clear correlation between the two variables. The data points are spread out in a cloud with no apparent pattern.

The scatter plot of Average Price vs. Total number of Applicants. In this plot there is a weak positive correlation between the two variables. There is a lot of scatter in the data, and the correlation is not very strong. Due to this we are doing a logarithm of the model. Post logarithm (Figure 3.1.2) of the model, we are again plotting the scatter plot there is a positive correlation between the two variables, though there is also a lot of scatter in the data. This means that as the total number of applicants increases, the average price also tends to increase.

	Total.number.of.license.issued	lowest.price	avg.price	Total.number.of.applicants	Month	log_Total_number_of_applicants
Total.number.of.license.issued	1.0000000	0.72045090	0.73874089	0.649151871	0.127166677	0.74535195
lowest.price	0.7204509	1.00000000	0.99595950	0.814284593	0.047202063	0.87712956
avg.price	0.7387409	0.99595950	1.00000000	0.817585000	0.045124233	0.87882389
Total.number.of.applicants	0.6491519	0.81428459	0.81758500	1.000000000	0.003639173	0.93485738
Month	0.1271667	0.04720206	0.04512423	0.003639173	1.000000000	0.01855809
log_Total_number_of_applicants	0.7453520	0.87712956	0.87882389	0.934857384	0.018558095	1.00000000

Figure 3.1.3: Correlation between the variables

(Based on Figure 3.1.3)
Total number of license issues vs Average Price: The correlation coefficient between these two variables is likely positive. This means that as the total number of licenses issued increases, the average price of the licenses also tends to increase. This could be because there is a higher demand for licenses when more are issued, driving up the price.

Lowest price and Average Price: The correlation coefficient between these two variables is also likely positive. This means that when the lowest price of a license plate increases, the average price also tends to increase. This relationship makes sense as the average price is likely influenced by the lowest priced licenses as well.

Total number of applicants and Average Price: The correlation coefficient between these two variables is likely positive. This means that as the total number of applicants for licenses increases, the average price of the licenses also tends to increase. This could be because a higher number of applicants indicates a higher demand for licenses, which can drive up the price.

Average Price vs. Total number of Applicants (after log): The correlation coefficient is 0.649, which indicates a weak positive correlation. There is a slight upward trend in the data, but there is also a lot of scatter, meaning that there are many exceptions to the trend.

Section 3.2: Comparison of "candidate" models in terms of fit and hold-out sample

Multiple regression model - all three independent variables:

```
Call:
lm(formula = n_avg ~ n_month + n_issued + n_lowest + n_applicants)

Residuals:
    Min      1Q  Median      3Q     Max 
-2472.3  -907.0  -252.7   463.6 19919.2 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1945.75440 2109.14550   0.923    0.358    
n_month       -40.28970   46.95884  -0.858    0.392    
n_issued        0.42325    0.09199   4.601 8.04e-06 ***
n_lowest        0.93171    0.01313  70.974  < 2e-16 ***
n_applicants  -29.15987  267.26791  -0.109    0.913    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2152 on 175 degrees of freedom
Multiple R-squared:   0.99,	Adjusted R-squared:  0.9898 
F-statistic:  4344 on 4 and 175 DF,  p-value: < 2.2e-16
```

Figure 3.2.1: Output of Multiple regression model - all three independent variables

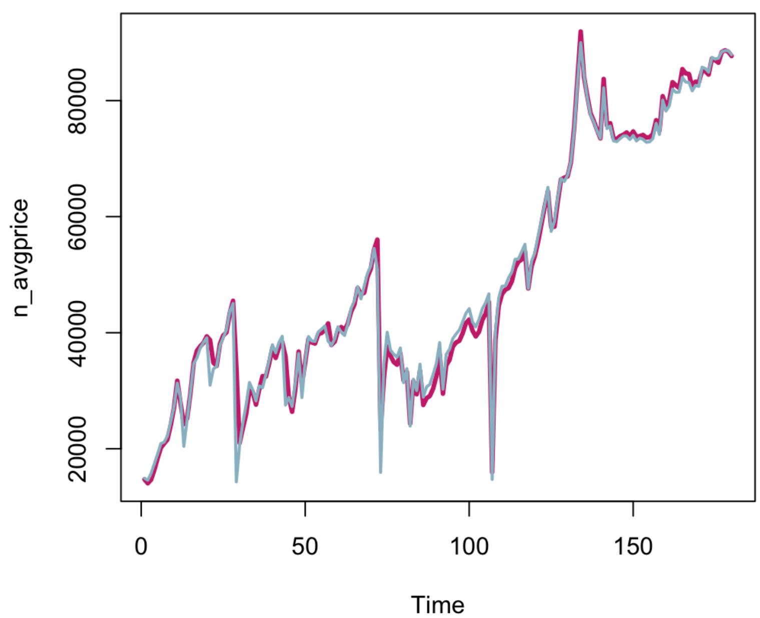

Figure 3.2.2: Actual vs predicted fit for Multiple regression model model - all three independent variables (Pink = actual, Blue = predicted)

The regression model (Figure 3.2.1) is built using the 3 independent variables: Total number or license issued, Total number of applicants (after logarithm) and lowest price of the plates for the dependent variable - Average price of the plates. The model has a very high R-squared of 0.99, which means that the model predicts 99% of the target variable. The fit of the actual vs predicted (Figure 3.2.2) is also very good. The MAPE of this model is 0.003989538, which indicates that the mean average error of the model is over 0.3989538%, which is very low and suggests a good fit.

Multiple regression model - Total number of applicants + Total number of plates issued:

```
Residuals:
   Min     1Q Median     3Q    Max 
-40060  -6312    819   5963  38379 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -9.565e+04  8.703e+03 -10.991  < 2e-16 ***
n_month       4.470e+01  2.555e+02   0.175 0.861317    
n_issued      1.685e+00  4.911e-01   3.431 0.000749 ***
n_applicants  1.330e+04  1.035e+03  12.853  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11710 on 176 degrees of freedom
Multiple R-squared:  0.703,	Adjusted R-squared:  0.698 
F-statistic: 138.9 on 3 and 176 DF,  p-value: < 2.2e-16
```

Figure 3.2.3: Output of Multiple regression model - without lowest price

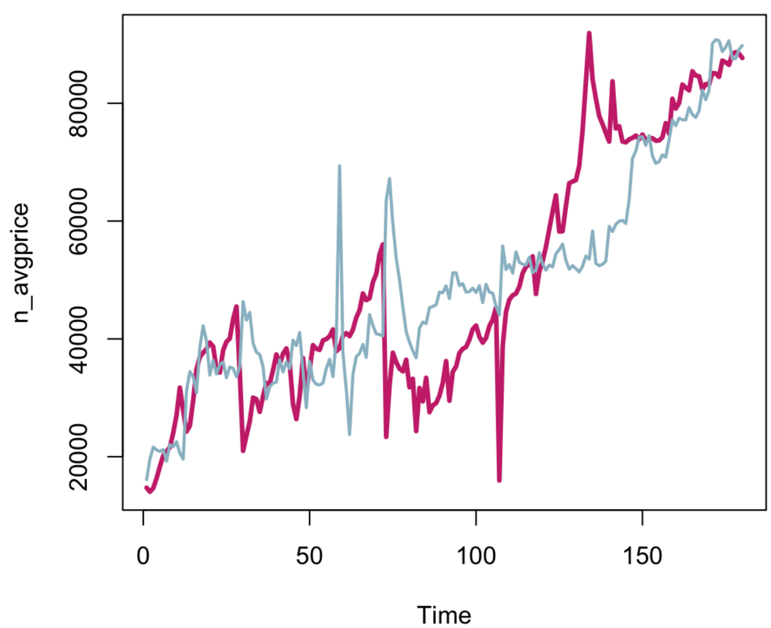

Figure 3.2.4: Actual vs predicted fit of multiple regression model - without lowest price (Pink = actual, Blue = predicted)

The regression model (Figure 3.2.3) is built using the 2 independent variables: Total number or license issued, Total number of applicants (after logarithm) for the dependent variable - Average price of the plates. The model has a very high R-squared of 0.703, which means that the model predicts 70.3% of the target variable. The model fits the predicted over the actual (Figure 3.2.4) pretty well at some points but not all. The MAPE of this model is 0.02796903, which indicates that the mean average error of the model is over 2.796903%, which is very low and suggests a good fit.

	MLR: all 3 independent variables	**MLR: excluding lowest price**
p-value	Statistically significant	Statistically significant
R-squared	0.99	0.703
MAPE	0.003989538	0.02796903
Variance estimate	2132.237	11613.34

Table 3.2.1: Comparison of the models

Based on Table 3.2.1, we can see that the 'MLR: all 3 independent variables' model performs better as it has a lower MAPE, higher R-squared and lower square root of variance estimate. So we choose that model for further analysis of the data.

Section 3.3: Looking at residuals of the model(s)

Multiple regression model - all three independent variables:

Figure 3.3.1: ACF of the residuals from multiple regression model

```
        Box-Pierce test

data:  fit_cyc$resid
X-squared = 380.58, df = 20, p-value < 2.2e-16
```

Figure 3.3.2: Box-Pierce test of multiple regression model

The ACF decays slowly meaning that the series is non-stationary. Based on the ACF (Figure 3.3.1) we can see that most of the lags are outside the 2 standard error bounds, this rejects the null hypothesis that the series is a white noise series. The Box-Pierce test (Figure 3.3.2), has a p-value less that 0.05, which means it also rejects the null hypothesis that the series is a white noise series.

Section 4: Stochastic Time Series Models

Section 4.1: Analysis and modeling of deterministic time series model residuals

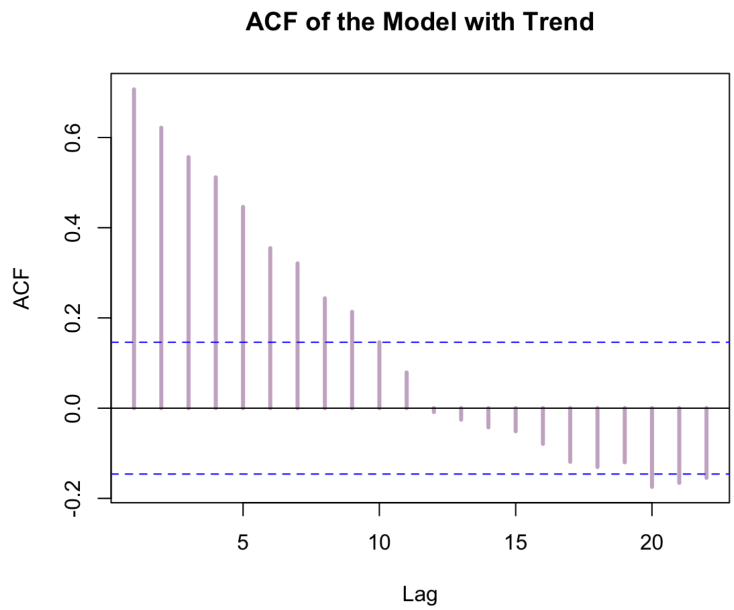

Figure 4.1.1: ACF of the residuals from the Model with Trend

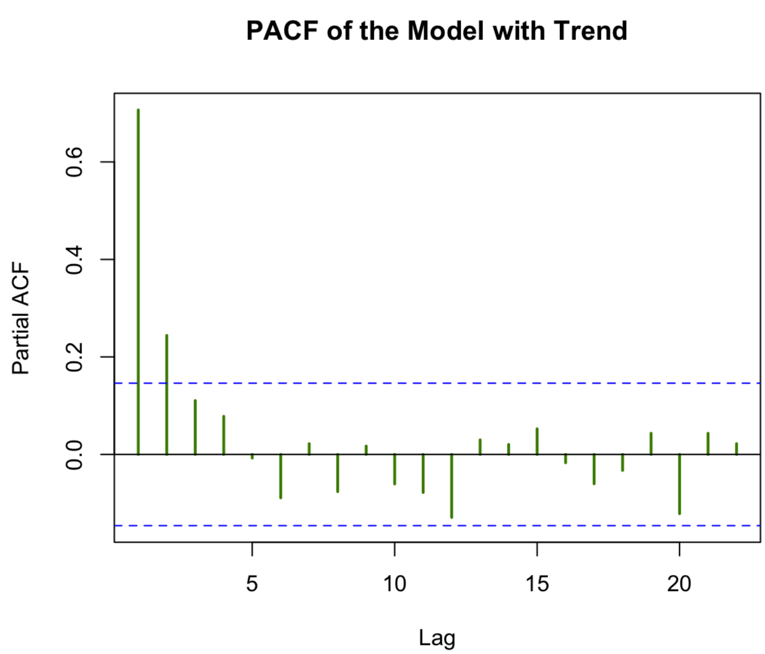

Figure 4.1.2: PACF of the residuals from the Model with Trend

The ACF (Figure 4.1.1) of the deterministic time series (Model with trend) decays slowly, meaning that it is a non stationary and the PACF (Figure 4.1.2) cuts off after lag 2 indicating that it is an Auto Regressive process of order 2 (AR(2)).

```
Series: n_avgprice
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1     ar2      mean
      0.8017  0.1879  53527.31
s.e.  0.0686  0.0690  21298.86

sigma^2 = 21468897:  log likelihood = -2011.8
AIC=4031.6   AICc=4031.8   BIC=4044.87

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE        ACF1
Training set 410.5026 4599.258 2600.077 -0.6237172 7.244234 0.9975311 -0.03745203
```

Figure 4.1.3: Output of AR(2) process for deterministic time series

Figure 4.1.3 shows us the coefficients of the AR(2) process along with the RMSE, MAPE, etc.

```
        Box-Pierce test

data:  four.1_ar$resid
X-squared = 14.712, df = 20, p-value = 0.7926
```

Figure 4.1.4: Box-Pierce test for AR(2) process for deterministic time series

The Box-Pierce test (Figure 4.1.4) for the AR(2) process has a p-value greater than 0.05, which means that we fail to reject the null hypothesis that the series is white noise. Based on the ACF of the residuals (Figure 4.1.5) for the AR(2) process, we also fail to reject the null hypothesis because all the lags - but one - are in between the 2 standard error bounds.

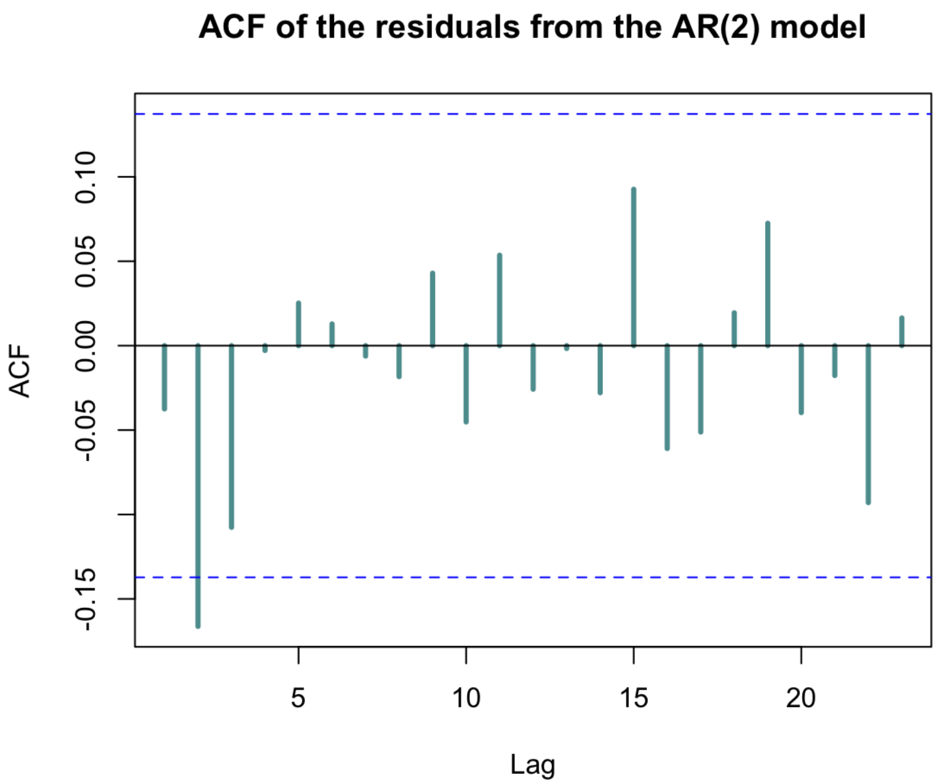

Figure 4.1.5: ACF of the residuals for AR(2) process for deterministic time series

Section 4.2: Analysis and modeling of regression model residuals

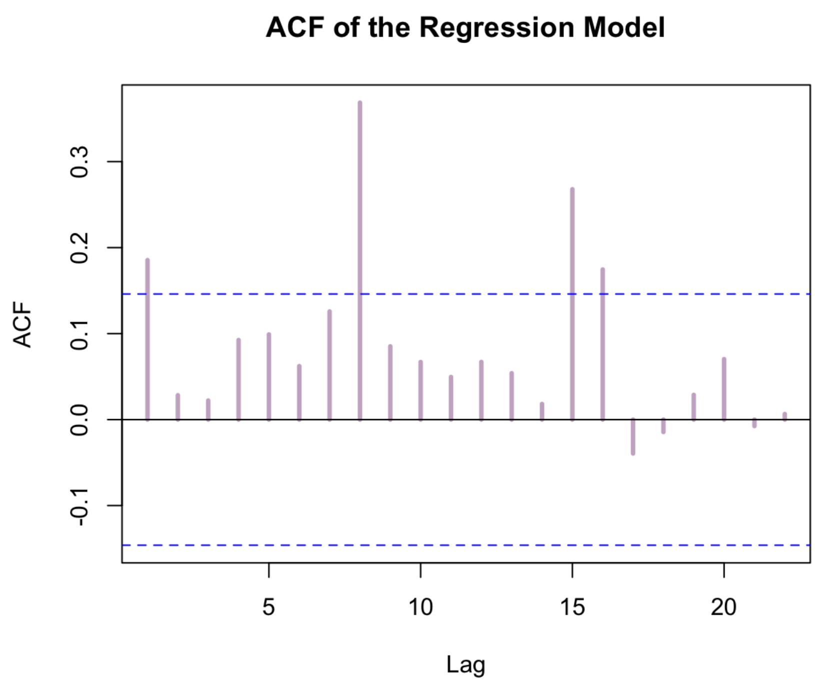

Figure 4.2.1: ACF of the residuals from the Regression Model

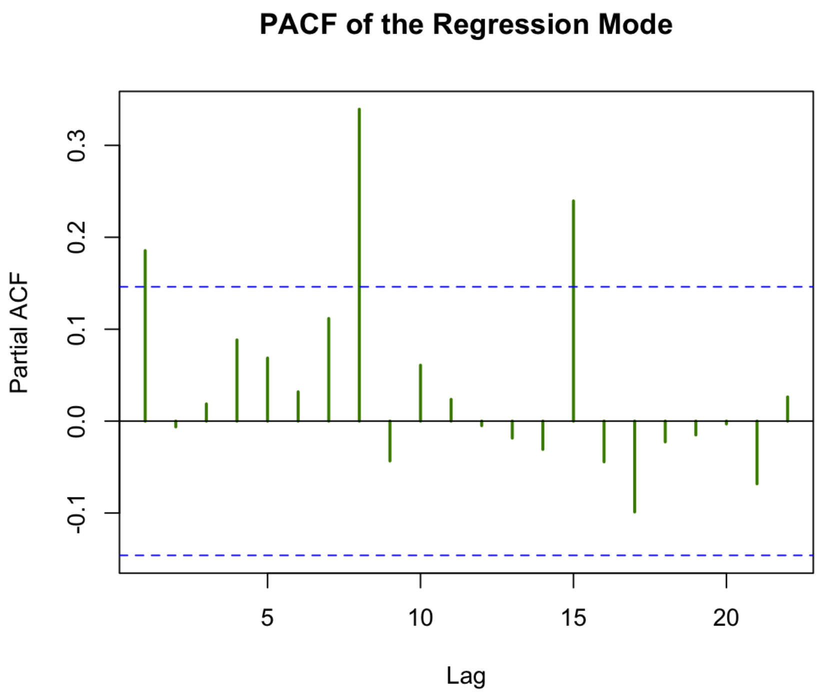

Figure 4.2.2: PACF of the residuals from the Regression Model

The ACF (Figure 4.2.1) and PACF (Figure 4.2.2) of the residuals for the regression model do not exhibit any stationary behavior so we difference them and plot the ACF and PACF of the difference residuals.

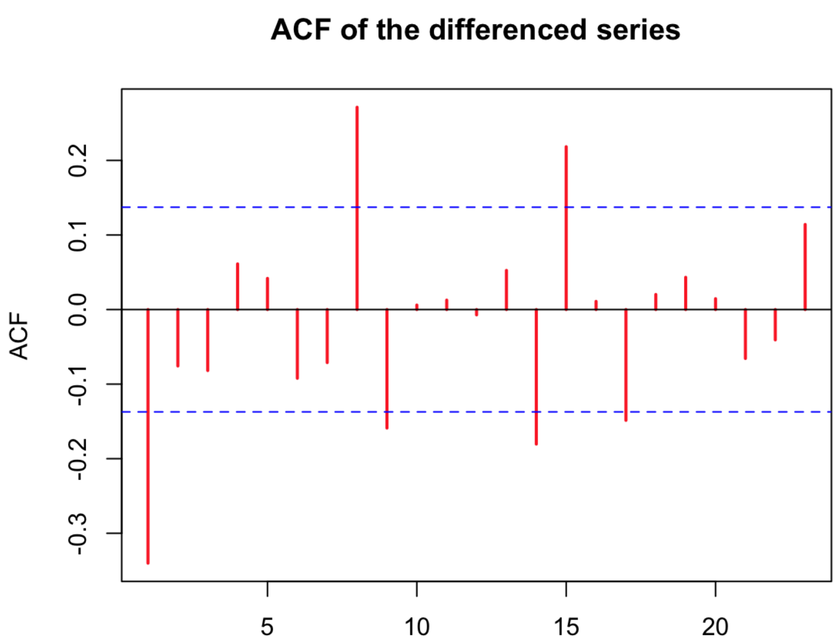

Figure 4.2.3: ACF of the differenced series from the Regression Model

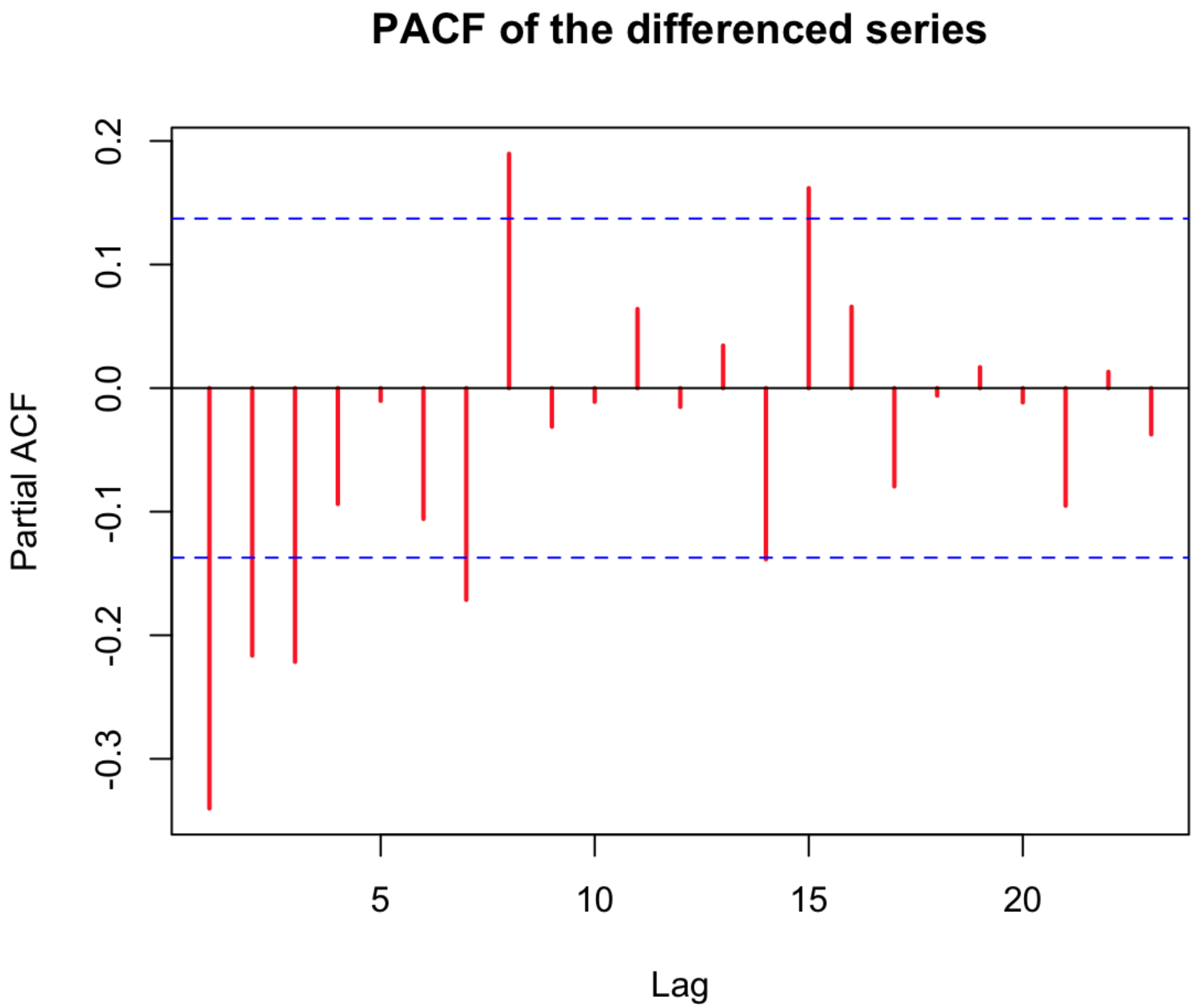

Figure 4.2.4: PACF of the differenced series from the Regression Model

The ACF in the differenced series cuts off after lag 1 (Figure 4.2.3) and the PACF (Figure 4.2.4) decays slowly, indicating that it is an Moving Average process of order 1 (MA(1)).

Figure 4.2.5 shows us the coefficients of the MA(1) process of the corrected regression model along with the RMSE, MAPE, etc.

The Box-Pierce test (Figure 4.2.6) for the MA(1) process has a p-value less than 0.05, which means that reject the null hypothesis that the series is white noise. Based on the ACF of the residuals (Figure 4.2.7) for the MA(1) process, we also reject the null hypothesis because few lags are outside the 2 standard error bounds.

```
Call:
arima(x = n_avgprice, order = c(0, 1, 1), xreg = x)

Coefficients:
          ma1  Total.number.of.license.issued  lowest.price  Total.number.of.applicants
       -0.751                          0.5863        0.8198                     -0.0086
s.e.    0.070                          0.1236        0.0317                      0.0080

sigma^2 estimated as 3543552:  log likelihood = -1819.14,  aic = 3646.29
```

Figure 4.2.5: Output of the MA(1) process of the corrected regression model

```
	Box-Pierce test

data:  regarima$resid
X-squared = 37.136, df = 20, p-value = 0.01127
```

Figure 4.2.6: Box-Pierce test of the MA(1) process of the corrected regression model

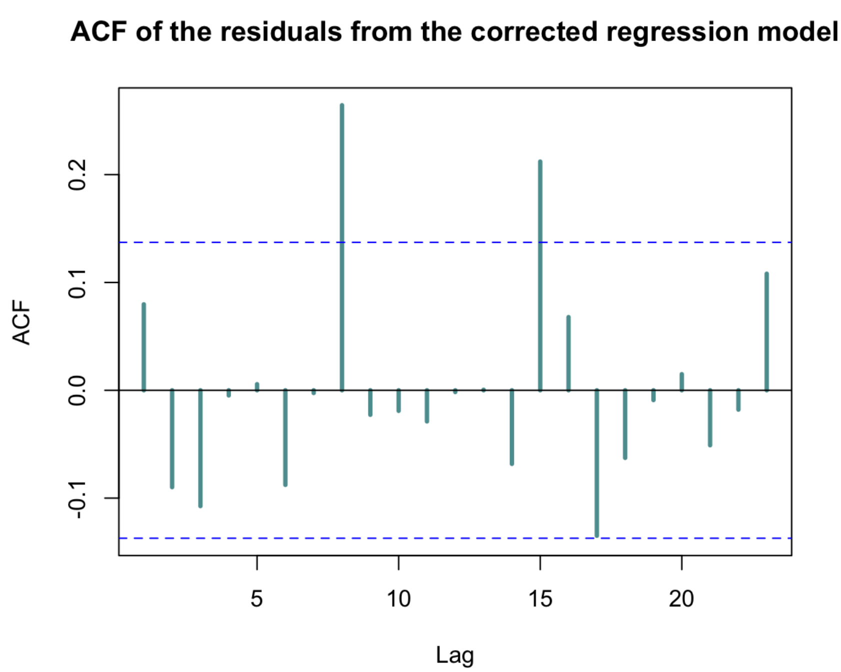

Figure 4.6.7: ACF of the residuals of the MA(1) process of the corrected regression model

Section 4.3: Analysis and modeling of regression model residuals

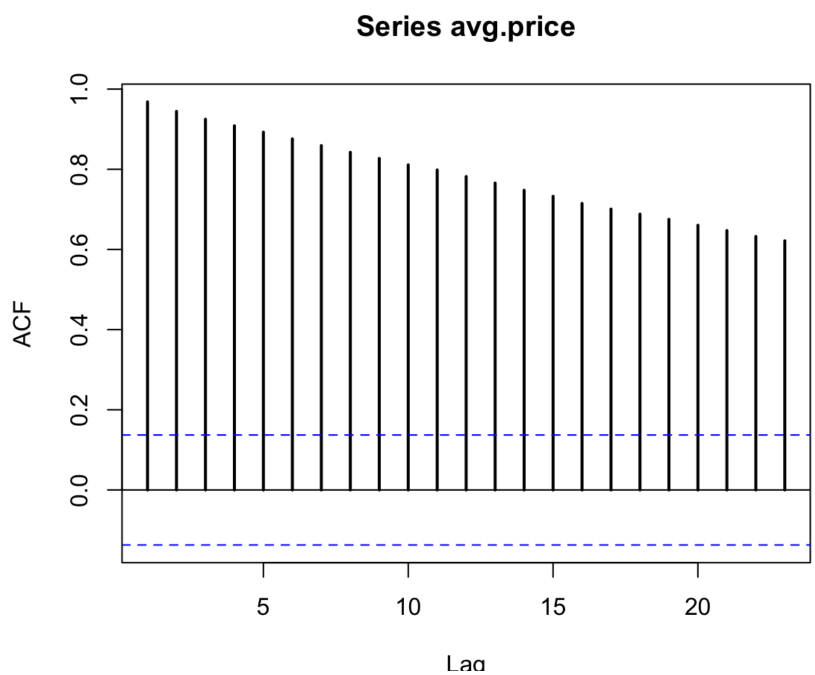

Figure 4.3.1: ACF of the series (Average price)

The ACF for the series (Figure 4.3.1) is non stationary, so we have to difference the series to make it stationary.

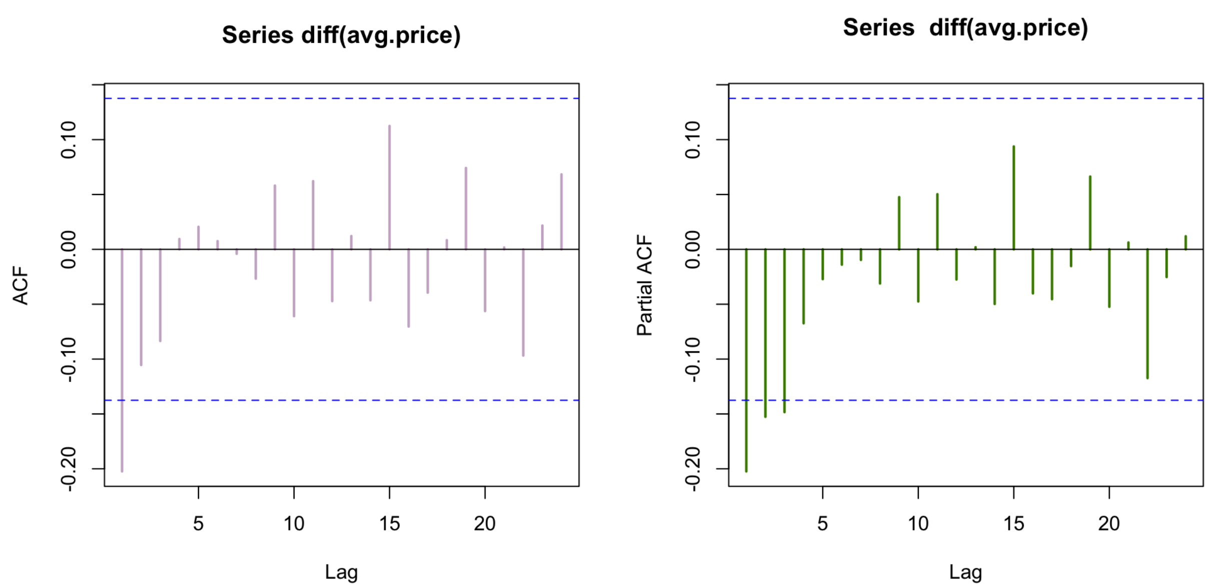

Figure 4.3.2: ACF and PACF of difference of the series

After differencing the series the ACF and the PACF (Figure 4.3.2) both decay slowly indicating an ARMA process.

```
Series: avg.price
ARIMA(1,1,1)

Coefficients:
         ar1      ma1
      0.3712  -0.6285
s.e.  0.1593   0.1295

sigma^2 = 20669633:  log likelihood = -1996.78
AIC=3999.57   AICc=3999.69   BIC=4009.51

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE         ACF1
Training set 614.5839 4512.834 2656.951 0.1191895 7.241835 1.019351 -0.008383722
```

Figure 4.3.3: Output of the ARMA process

Figure 4.3.3 shows us the coefficients of the ARMA process along with the RMSE, MAPE, etc.

```
        Box-Pierce test

data:  four.3_arima$residuals
X-squared = 7.5493, df = 20, p-value = 0.9945
```

Figure 4.3.4: Box-Pierce test of the ARMA process

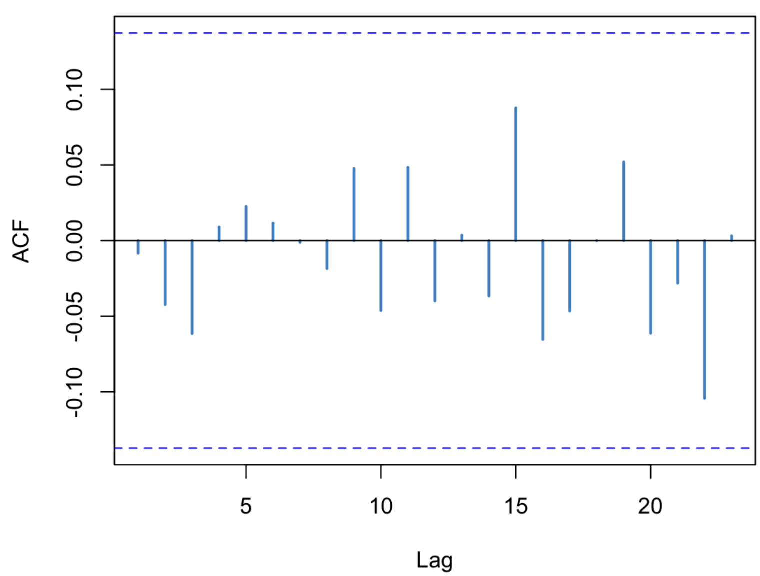

Figure 4.3.5: ACF of the residuals from the ARMA model

The Box-Pierce test (Figure 4.3.4) for the ARMA process has a p-value greater than 0.05, which means that we fail to reject the null hypothesis that the series is white noise. Based on the ACF of the residuals (Figure 4.3.4) for the ARMA process, we also fail to reject the null hypothesis because all the lags are in between the 2 standard error bounds.