FORECASTING

MATH 1307

Analysis Report of a Call Centre Abandoned Calls

BY

RESHMA TARUNI GADALA

Table of Contents

Introduction ... **3**

Data Description ... **3**

Methodology ... **3**

Descriptive Analysis ... **4**

Transformation and Differencing Target Series ... **6**

Decomposition ... **7**

Correlation Check ... **8**

Parameter Estimation and Diagnostic Checking ... **8**

- **Distributed Lag Models** ... **8**
  - Finite DLM's ... 8
  - Polynomial DLM's ... 10
  - The Koyck Transformation ... 11
  - ARDLM ... 12
- **Dynamic Linear Models** ... **14**
- **Exponential Smoothing Methods -Holt winters seasonal Methods** ... **15**
- **Innovation state space models** ... **17**

Forecasting ... **19**

- **Finite Dlm** ... **20**
- **ARDLM** ... **20**

Reference ... **23**

Appendix ... **24**

Introduction

It is seen that the Calls Received and Response Rate to the Call Centres is a crucial part of the performance of the call centres and also analysing them to increase the overall performance.
The main aim of this project is to forecast the abandoned calls for next 10 units in order to meet the call demand without over-staffing or under-staffing and in order to plan on resources that are needed to handle the centre's contact volume.
By forecasting the correct number of calls a company can predict staffing needs, meet service level requirements, improve customer satisfaction, and benefit from many other optimizations.

Data Description

The data used is about the "Call Centre Metrics for the Health Service System" of city of San Francisco, which records the monthly data from the phone system and includes metrics pertaining to

Inbound Calls - Total number of Calls Answered

Average Speed - Average speed of Calls Answered

In-person visit- In person Assistance

Abandoned Calls - Calls that are Abandoned

Abandonment rate - Abandonment rate of Total Calls offered.

Using the given data one more column was created in the flat file which is

Total Calls -Total number of calls been offered to the company.

The original dataset includes metrics from Jan 2011 to Oct 2018 , In this report different models will be analysed using abandoned calls series between Jan 2011 and Dec 2017, further Total number of calls received within the same time span are used as a predictor variable to forecast the future of abandoned calls which will help the company in maintaining optimal staff levels and plan accordingly to minimise the abandonment rate for the better performance of the Company. As the metrics belong to health service system call centre it is expected to get more calls in the flu seasons, and flu seasons of San Francisco are October to December.

Methodology

The methods used for the analysis of abandoned calls are:

- Time series plots are analysed to check for the basic characteristics of the series and applied statistical tests to confirm the observations.
- Augmented Dickey–Fuller test (ADF Test) is used to check the stationarity of the series.
- Decomposition is applied to have a clear understanding of the trend and seasonality component in the target series.
- Necessary transformation and differencing is applied to stabilise variation in the target series.
- Parameter Estimation and diagnostic checking for -
    - Distributed Lag Models (Finite DLM's, Polynomial DLM's, Koyck Transformation, ARDL Models),
    - Dynamic Models (Intervention analysis using Pulse Function) and
    - Innovations State Space Models (Linear homoscedastic state-space models, Linear heteroscedastic state-space models, Some nonlinear seasonal state-space models)
- Model Selection and forecasting 10 months ahead.

Descriptive Analysis

From time series plot of the ‘Monthly Total Calls Received by Call Centre’ (Figure 1), we can infer that

1. There is no trend.
2. There is clear seasonality as we can see peaks for the October month from 2013.
3. There is an apparent changing variance in the series.
4. Because of the fluctuations in the series we can consider Moving average behaviour and because of the succeeding points it can be considered Auto Regressive behaviour for this series.
5. There might be an intervention point in June 2011 and January 2013 as they have more number of incoming calls other than the flu seasons .

.

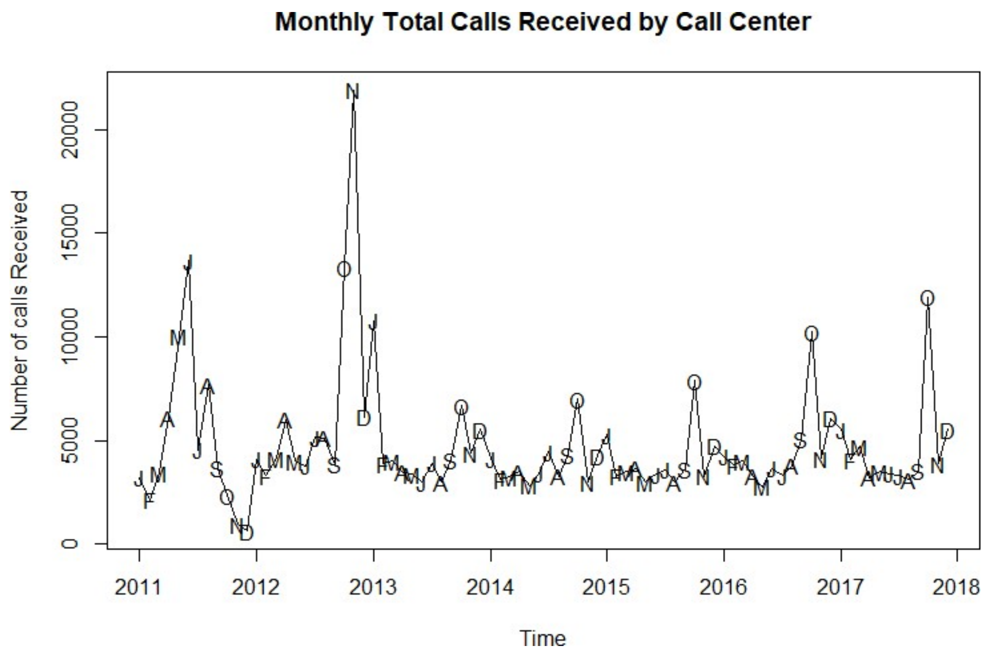

Figure 1. Time Series Plot of Monthly Total Calls Received

From the **ACF plot** (*Figure 2*) of the ‘Monthly Total Calls Received by Call Centre’ it was observed that there is no seasonality in the series and there is probably trend.

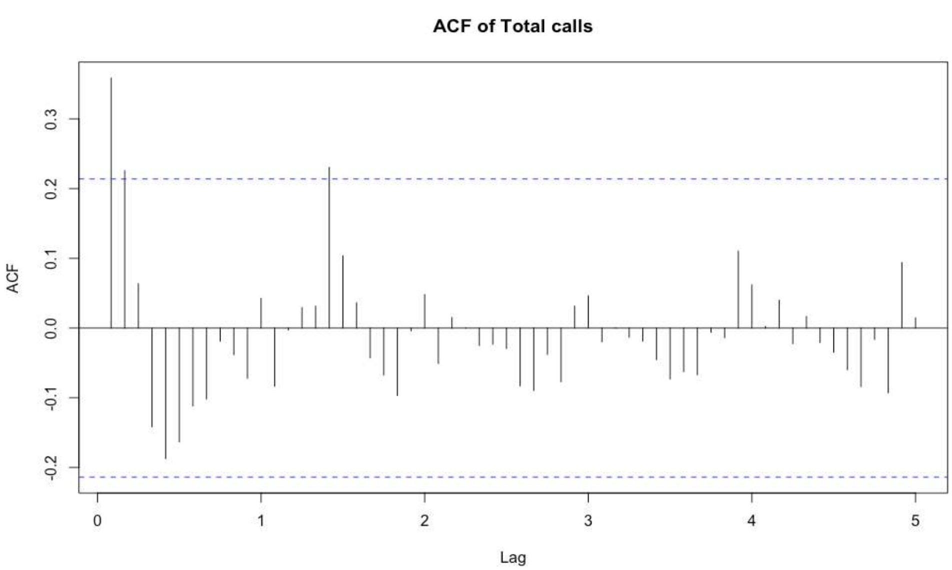

Figure 2. ACF Plot of Monthly Total Calls Received

According to the **ADF test** for ‘Total Calls Series’ as shown below,the p value is less than 0.05 we conclude the series is **stationary** at 5% level of significance.

```
        Augmented Dickey-Fuller Test

data:  Total_Calls.ts
Dickey-Fuller = -4.5296, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

From the time series plot (Figure 3)of ‘Monthly Abandoned Calls of the Call Centre’, it can be observed that

1. There is no trend.
2. There is clear seasonality as we can see peaks for the October months.
3. There is an apparent changing variance in the series.
4. Because of the fluctuations in the series we can consider Moving average behaviour and because of the succeeding points it can be considered Auto Regressive behaviour for this series.
5. There is clear intervention point in the series in January 2013 as high abandonment rate is recorded in that month , and may be the company did not have enough staff or resources at that particular period.

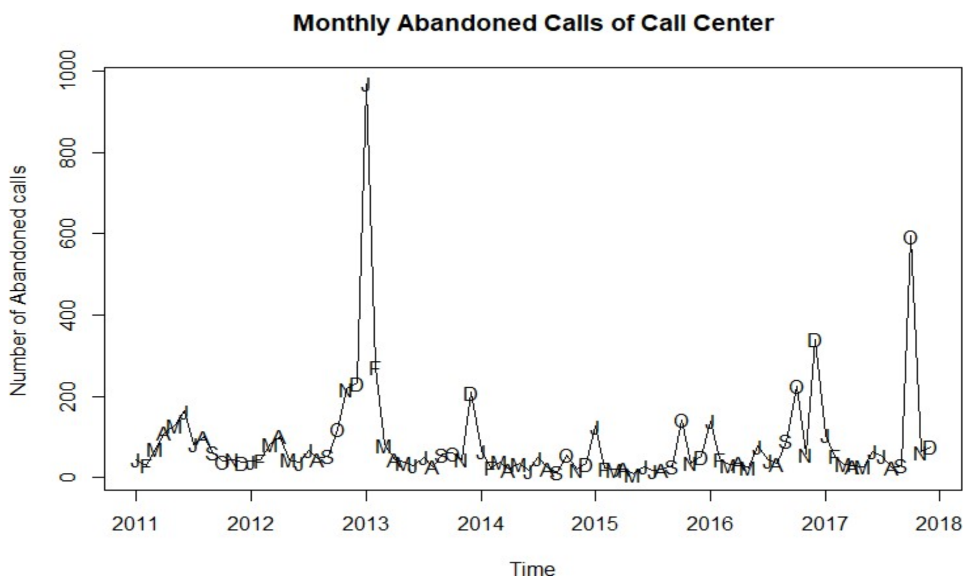

Figure 3. Time Series Plot of Monthly Abandoned Calls

From the **ACF plot** (Figure 4.)of ‘Monthly Abandoned Calls of the Call Centre’ it is observed that there is seasonality but there is no trend in the series.

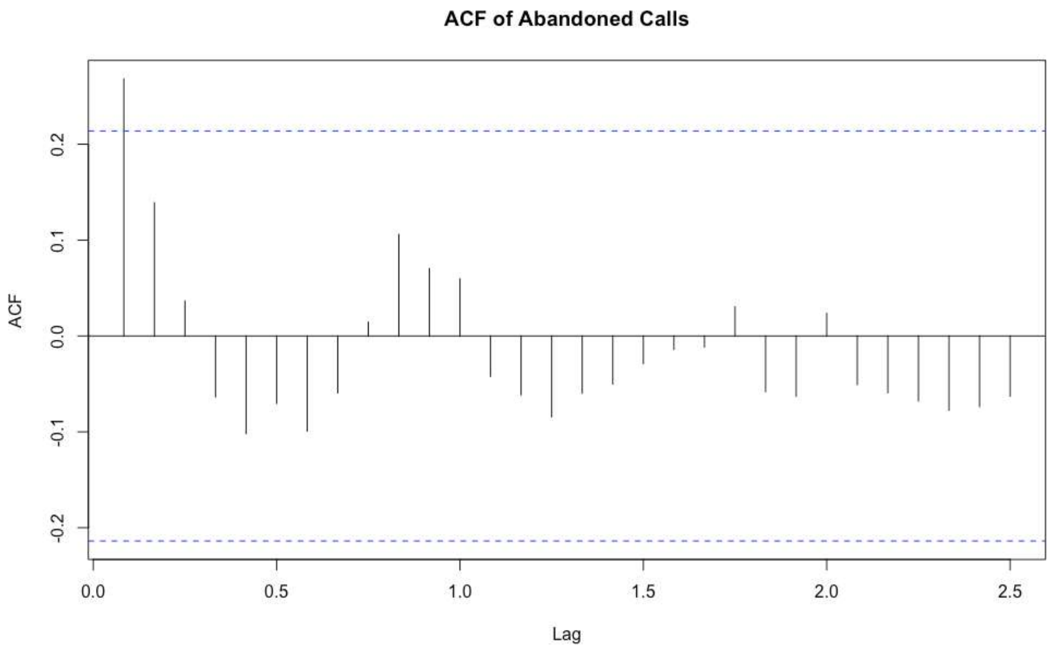

Figure 4. ACF Plot of Monthly Abandoned Calls

According to **ADF test** of the 'Abandoned Calls' as shown below , the p value is less than 0.05 , we conclude the series is **stationary** at 5% level of significance.

```
        Augmented Dickey-Fuller Test

data:  Abandoned_Calls.ts
Dickey-Fuller = -3.9682, Lag order = 4, p-value = 0.01506
alternative hypothesis: stationary
```

Transformation and Differencing Target Series

It is observed that there is changing variance in the series. To stabilise the series Box cox transformation (*Figure 5*) and first seasonal differencing (*Figure 6*)are applied. Boxcox transformation does not have much effect on the data as changing variance still exists. Same with the Seasonal differencing, there is still changing variance in the differenced series, but the series is stable.

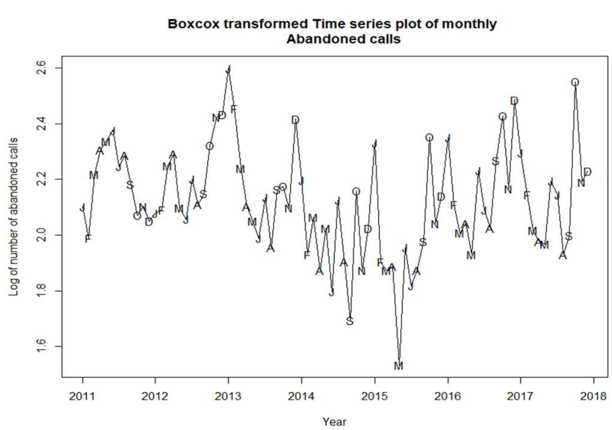

Figure 5: Transformed time series plot

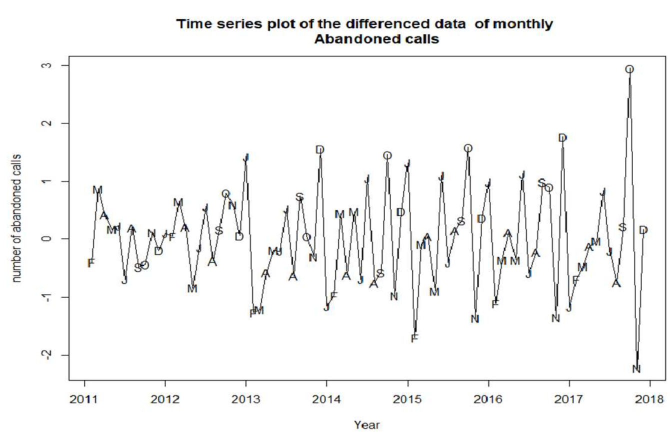

Figure 6: Differenced time series plot.

Decomposition

Decomposition is applied to have a view on the trend and seasonal components, which will give a better insight of the series. Here for the target variable Classic decomposition, X12 decompositions are applied which plots the characters individually.

In **Classic Decomposition** (Figure 7) the observed component shows the original series, in the trend component it can be seen that there is a slight upward trend in the series and in the random component there is a highest peak in 2013 which could be an intervention point.

From **X12 Decomposition (**Figure 8 **)** it can inferred that there is no trend in the series, next, the seasonally adjusted line shows there is seasonality in the plot and at some point the seasonally adjusted line does not fit or follow the original series, it means there is something else happening and not seasonality.

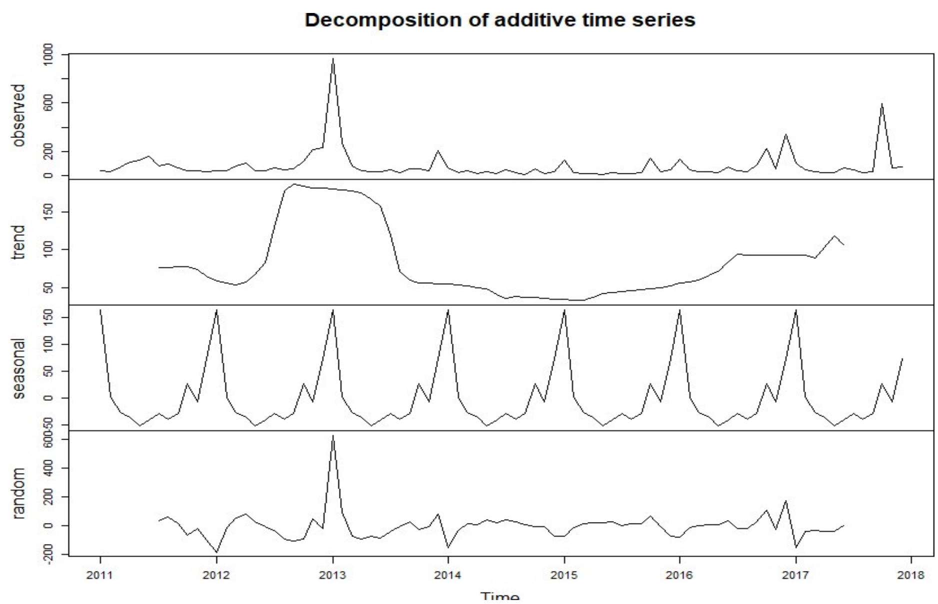

Figure 7: Classic Decomposition

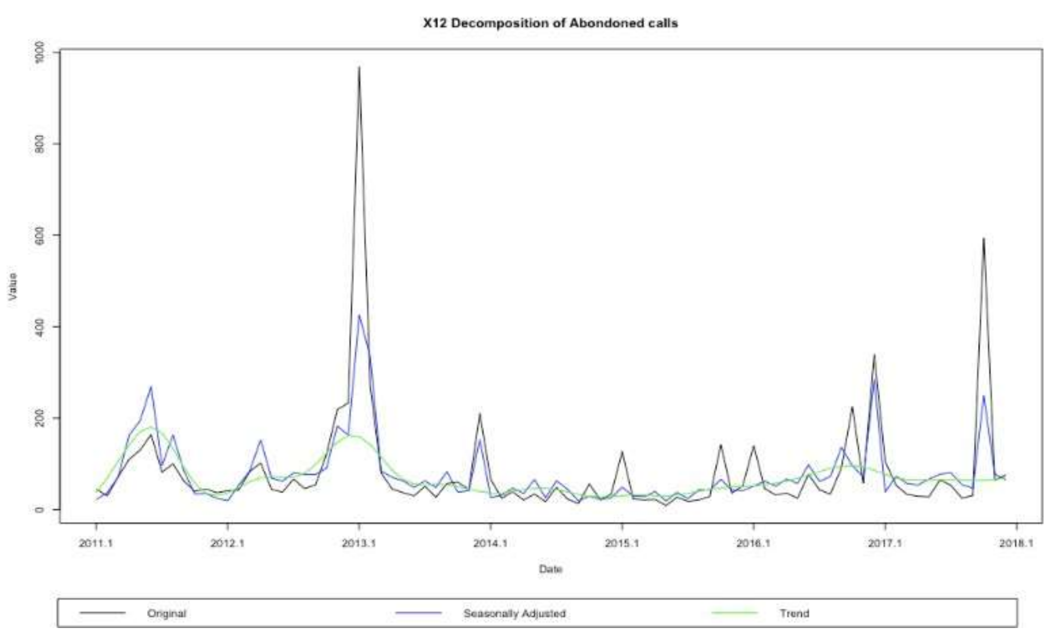

Figure 8: X12 Decomposition of Abandoned Calls

Correlation Check

From both the series (*Figure 9*) it can be observed that X and Y series follow each other but not very closely. 54% of correlations explained between the 'Total Calls' and 'Abandoned calls' which is a medium correlation.

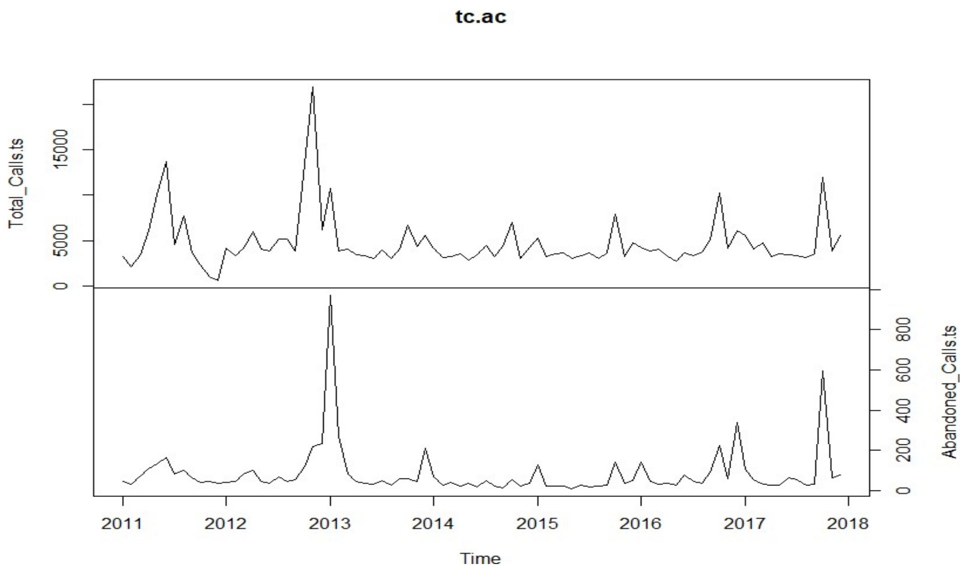

Figure 9: Plot showing both the series.

Output of Correlation:

	Abandoned Calls	Total Calls
Abandoned Calls	1.0000000	0.5425359
Total Calls	0.5425359	1.0000000

Parameter Estimation and Diagnostic Checking

Distributed Lag Models

Finite DLM's

In finite Dlag Models q value is changed to find the finite lag length by comparing AIC and BIC values.

The best model is always the model with lowest AIC and BIC values, because the information lost for this model is low.

The following table (TABLE 1)shows the values obtained from the analysis. The model with parameter q=10 gives the lowest AIC 871.8091 and BIC 901.7619.

For the model with parameter q=10 (Shown in Output of Finite DLM) all the coefficients are not significant, it has adjusted R-squared value is 0.6547 which is low but better than the other models, and the p-value was significant at 5% level, the MASE value is 0.5667.

Using the checkresiduals function we get the Residual plots (*Figure 10.1*), the time series plot of the standardised residuals has random pattern and the ACF plot has one highly significant lag which implies that there is some serial auto correlation left in the residuals and the histogram has a long tail at the end with outliers.

To check the overall significance of the model Breusch-Godfrey test is used, where p-value is greater than alpha, means greater than 0.05 we do not reject the null hypothesis, concluding that there is no serial correlation left in residuals. In Variance Inflation Factors it can be observed that the estimates of the finite DLM are not suffering from the multicollinearity. Hence, this model can be used for forecasting.

q	AIC	BIC	R^2	MASE
1	1020.745	1030.42	0.2784	0.6773328
2	977.3596	989.3932	0.517	0.7048363
3	961.8096	976.1763	0.5495	0.6900045
4	948.6113	965.2855	0.5677	0.6528492
5	939.2641	958.2197	0.5643	0.6483907
6	927.5661	948.7765	0.5724	0.6467594
7	916.0717	939.5097	0.5812	0.6054886
8	890.041	915.679	0.6612	0.5675623
9	880.5469	908.3568	0.66	0.5653771
10	871.8091	901.7619	0.6547	0.5667715

TABLE 1: AIC, BIC, MASE VALUE AND RSQUARE of Finite DLM

Output of Finite DLM:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.632e+02  4.163e+01  -3.919 0.000224 ***
x.t          1.960e-02  3.564e-03   5.498 7.69e-07 ***
x.1         -5.934e-03  3.683e-03  -1.611 0.112248
x.2          2.626e-02  3.738e-03   7.027 1.93e-09 ***
x.3          1.598e-02  3.998e-03   3.997 0.000173 ***
x.4         -7.188e-03  3.989e-03  -1.802 0.076364 .
x.5         -6.420e-03  3.811e-03  -1.684 0.097132 .
x.6          2.438e-03  3.773e-03   0.646 0.520598
x.7          3.440e-03  3.715e-03   0.926 0.357972
x.8          2.238e-03  3.695e-03   0.606 0.546841
x.9          1.663e-03  3.658e-03   0.455 0.650917
x.10         1.467e-03  3.474e-03   0.422 0.674244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 80.19 on 62 degrees of freedom
Multiple R-squared:  0.7067,    Adjusted R-squared:  0.6547
F-statistic: 13.58 on 11 and 62 DF,  p-value: 9.748e-13

AIC and BIC values for the model:
       AIC      BIC
1 871.8091 901.7619
> checkresiduals(model1.10$model)

        Breusch-Godfrey test for serial correlation of order up to 15

data:  Residuals
LM test = 20.853, df = 15, p-value = 0.1416

> MASE(model1.10)
              MASE
model1.10 0.5667715
> VIF.model1.10 = vif(model1.10$model)
> VIF.model1.10 > 10
  x.t   x.1   x.2   x.3   x.4   x.5   x.6   x.7   x.8   x.9  x.10
FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

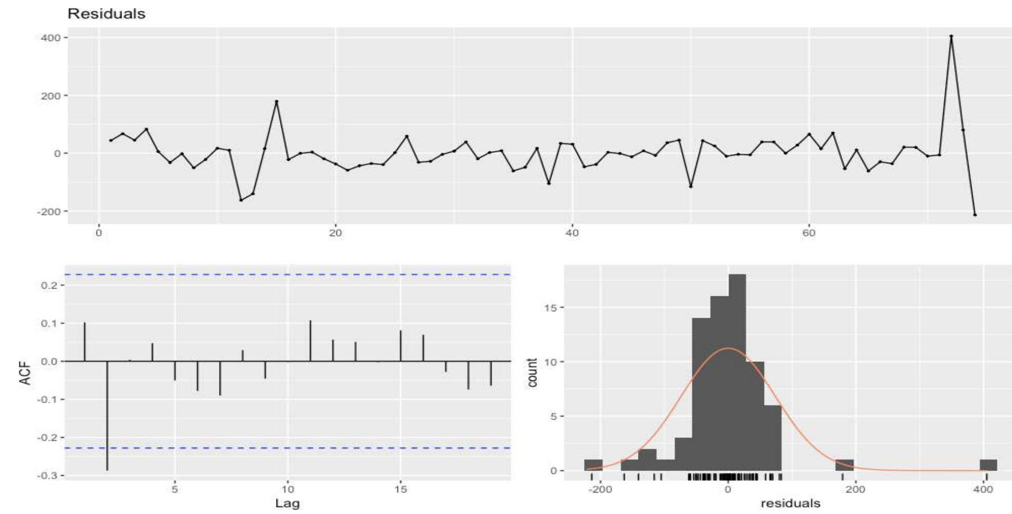

Figure 10.1: Residuals of finite DLM

Polynomial DLM's

Using 2nd order Polynomial DLM , the estimates of the original parameters are found.

The model with parameters q=10 and k=2 gives the lowest AIC 902.5478 and BIC 914.0681.

The following table (*Table 2*)are the values obtained from the analysis which shows the AIC and BIC

Many different models were analysed, and the best model had the following results.

From the summary of the model using q = 10, it was observed that all the coefficients are significant, the p-value is significant at 5% level of significance, the Adjusted R-squared value is 0.4249 which is low, MASE value is 0.6901(as shown below in Output of Polynomial DLM and *Table 2*))

q	AIC	BIC	R^2	MASE
2	977.3596	989.3932	0.517	0.7048363
3	980.089	992.0612	0.4288	0.6897832
4	974.1602	986.0703	0.391	0.7202781
5	963.9686	975.8158	0.3831	0.7292902
6	953.3517	965.1352	0.3762	0.7225411
7	940.0599	951.7789	0.3934	0.7067812
8	926.8389	938.4925	0.4098	0.6853801
9	913.515	925.1024	0.4267	0.6741808
10	902.5478	914.0681	0.4249	0.6901556

Table 2: AIC, BIC, MASE VALUE AND RSQUARE of Polynomial DLM

In the diagnostic check (*Figure 10.2*) the time series plot of the standardised residuals there is obvious non-random pattern , ACF plot shows one highly significant lag which implies that there is some serial auto correlation left in the residuals, the histogram long tail at the ends with outliers.

In Breusch-Godfrey test p-value is greater than 0.05 we do not reject the null hypothesis, concluding that the re is no serial correlation left in residuals. The Variance Inflation Factors estimate that the Poly DLM are slig htly affected from the multicollinearity. Hence this model is not used for forecasting.

Output of Polynomial DLM:

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.747e+02  5.330e+01  -3.277 0.001634 ** 
z.t0         1.746e-02  2.457e-03   7.108 7.94e-10 ***
z.t1        -4.980e-03  1.068e-03  -4.661 1.46e-05 ***
z.t2         3.581e-04  1.027e-04   3.487 0.000848 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 103.5 on 70 degrees of freedom
Multiple R-squared:  0.4485,    Adjusted R-squared:  0.4249 
F-statistic: 18.98 on 3 and 70 DF,  p-value: 4.133e-09

> checkresiduals(model2.102$model)

	Breusch-Godfrey test for serial correlation of order up to 10

data:  Residuals
LM test = 9.506, df = 10, p-value = 0.4849

> MASE(model2.102)
                MASE
model2.102 0.6901556
> VIF.model2.102 = vif(model2.102$model)
> VIF.model2.102 > 10
 z.t0  z.t1  z.t2 
FALSE  TRUE  TRUE 
```

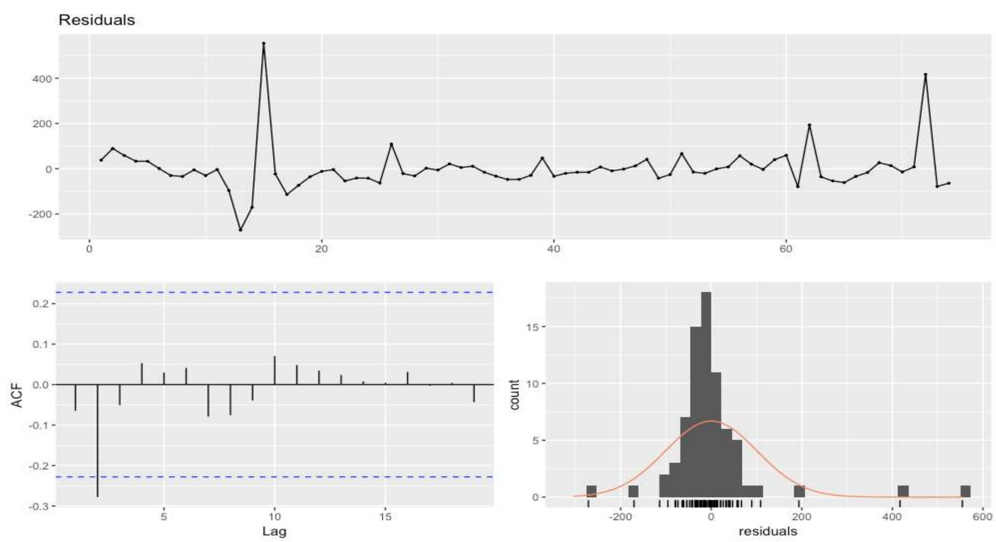

Figure 10.2: Residuals of polynomial DLM.

The Koyck Transformation

Another way to deal with Infinite DLM is to use Koyck Transformation, from the summary results shown below in the output of Koyck transformation model it is observed that the coefficient X.t is not significant, the of Adjusted R-Squared value is 0.277 which is very low, the p-value is significant at 5% level of significance. MASE value is 0.6697 which is moderate.

The Variance Inflation Factors estimate that the Koyck Transformation is not affected from the multicollinearity.

In the residual analysis (Figure 11) the time series plot is not random, the ACF plot has white noise but the histogram has a very long tails at the ends and not normally distributed.

Adjusted R square value is very low and not all coefficients are significant and in the diagnostic check the histogram is not normally distributed. Hence, this model is not used for forecasting.

Output of Koyck transformation model:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.52765   50.98307   0.069   0.9450
Y.1          0.24617    0.09561   2.575   0.0119 *
X.t          0.01271    0.01054   1.205   0.2317
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 110.1 on 80 degrees of freedom
Multiple R-Squared: 0.2946,     Adjusted R-squared: 0.277
Wald test: 4.809 on 2 and 80 DF,  p-value: 0.01066

                           alpha       beta       phi
Geometric coefficients:  4.679615 0.01270695 0.2461664
> checkresiduals(model3$model)
> MASE(model3)
            MASE
model3 0.6697285
> VIF.model3 = vif(model3$model)
> VIF.model3 > 10
  Y.1   X.t
FALSE FALSE
```

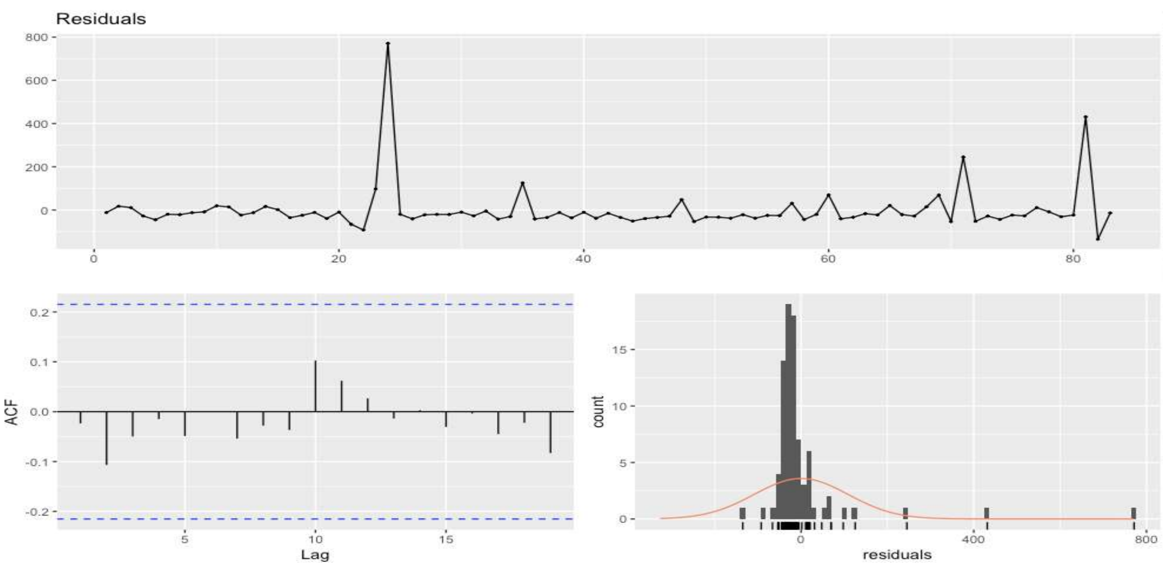

Figure 11: Residuals of Koyck transformation model

ARDLM

Autoregressive Distributed Lag Model are used when we are not able to find a suitable model with the Poly DLMs and Kyock DLMs, by using the large values of p and q we can possibly find a good model for the series.

The following AIC and BIC along with p and q values were obtained after fitting the model.

According to both AIC and BIC, ARDL(5,2) model is the most accurate one as it has the lowest AIC 933.2973 and BIC 956.9918(shown in table 3)
From the summary of the model ARDL(5,2) shown below in output of ARDLM, it is observed that all the coefficients are significant, the p-value is significant at 5% level of significance, the Adjusted R-squared value is 0.6151 which is moderarte, MASE value is 0.6329.

The Variance Inflation Factors estimate that the ARDL(5,2) is not affected from the multicollinearity.

In the diagnostic plots of residuals (*Figure 12*) for ARDL(5,2) ,the time series plot is fluctuating around zero and is random , the ACF plot shows there is no correlation left in the residuals ,the histogram is nearly normally distributed but has a very long tails at the ends with outliers .Hence this model is used for forecasting.

p	q	AIC	BIC	R^2	MASE
5	2	933.2973	956.9918	0.605	0.6216514
2	5	933.8896	957.584	0.602	0.6240863
5	3	935.2948	961.3587	0.5993	0.6222979
3	5	935.7412	961.8051	0.597	0.6211908
5	4	937.2646	965.6979	0.5936	0.6212069

Table 3: AIC, BIC, MASE VALUE AND RSQUARE of Finite DLM

Output of ARDLM:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -87.749297  21.166794  -4.146 8.73e-05 ***
X.t           0.020916   0.003226   6.484 8.09e-09 ***
X.1          -0.014691   0.004009  -3.664 0.000457 ***
X.2           0.029574   0.003839   7.703 4.06e-11 ***
Y.1           0.335770   0.086680   3.874 0.000225 ***
Y.2          -0.312348   0.086703  -3.602 0.000560 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 80.73 on 76 degrees of freedom
Multiple R-squared:  0.6389,	Adjusted R-squared:  0.6151 
F-statistic: 26.89 on 5 and 76 DF,  p-value: 1.501e-15

> checkresiduals(model4.52$model)

	Breusch-Godfrey test for serial correlation of order up to 10

data:  Residuals
LM test = 6.675, df = 10, p-value = 0.7557

> MASE(model4.52)
              MASE
model4.52 0.6329252
> VIF.model4.52<-vif(model4.52$model)
> VIF.model4.52 > 10
      X.t L(X.t, 1) L(X.t, 2) L(y.t, 1) L(y.t, 2) 
    FALSE     FALSE     FALSE     FALSE     FALSE 
```

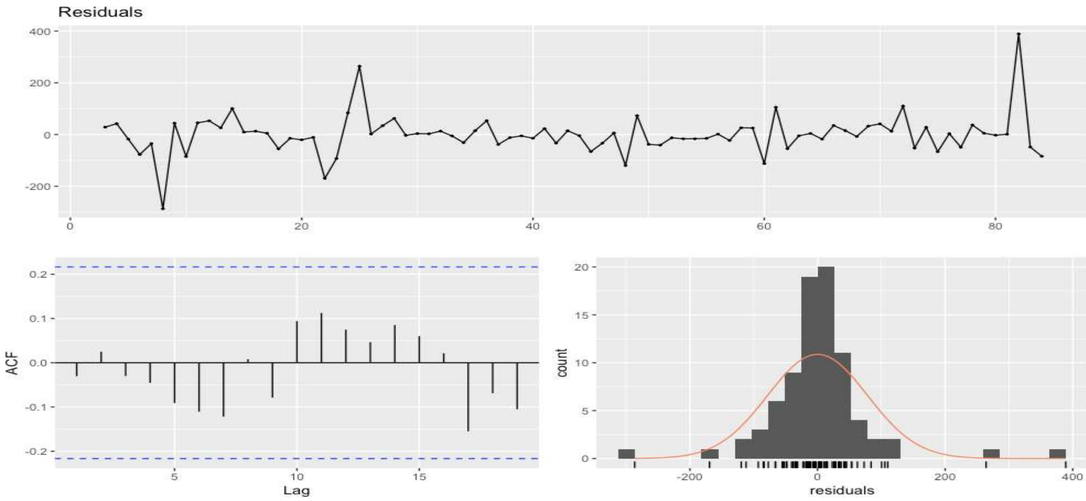

Figure 12: Residuals of ARDLM

Dynamic Linear Models:

We will use the intervention analysis to find the right model for our data as there is an intervention in our series in the year 2013 of January for Abandoned calls series.

When the mean level of the series does not go back to its previous mean level after intervention step function is used on the other hand from the Abandoned Calls series it is observed that after the intervention the mean level falls back to the previous mean level for which the pulse function is used.

The pulse function analysis (shown in output of Dynamic Linear Model) the series and the following AIC values were obtained from the fitted models which resulted in the model with first lag of the series and P.t and seasonal component being the best model.

From summary it is observed that first lag of the series P.t and October month are significant, the other months are not significant, the p-value is significant at 5% level of significance, the Adjusted R-squared value is 0.6572 which is moderate.

In Breusch-Godfrey test p-value is 0.4622 which is greater than 0.05 so we do not reject the null hypothesis, concluding that there is no serial correlation left in residuals.

In the residual analysis (*Figure 13*) it is observed that series is randomly distributed , but in the ACF there is one highly significant lag but is not having any seasonality ,where as the histogram is not normally distributed.

The AIC and BIC values of the fitted intervention models are shown below: .

Model	AIC	BIC
5.7	960.5363	999.0438
5.8	961.3103	1002.2245
5.1	968.6933	1004.9759
5.2	968.6933	1007.5986
5.6	969.6236	1008.3250
5.3	971.5473	1012.6676
5.5	974.4731	1073.7295

Table 4: AIC, BIC VALUEs of Intervention Models

Output of Dynamic Linear Model.

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      67.40926   35.26952   1.911  0.06013 .
L(Y.t, k = 1)     0.21093    0.07225   2.920  0.00473 **
P.t             852.44481   83.34753  10.228 1.83e-15 ***
season(Y.t)Feb  -42.25407   44.75131  -0.944  0.34836
season(Y.t)Mar  -30.77569   44.63238  -0.690  0.49280
season(Y.t)Apr  -26.11406   44.79363  -0.583  0.56180
season(Y.t)May  -34.55044   44.78673  -0.771  0.44308
season(Y.t)Jun  -16.94562   44.87028  -0.378  0.70684
season(Y.t)Jul  -28.14746   44.71726  -0.629  0.53113
season(Y.t)Aug  -38.91875   44.78948  -0.869  0.38790
season(Y.t)Sep  -27.58294   44.91814  -0.614  0.54119
season(Y.t)Oct   99.15041   44.82609   2.212  0.03029 *
season(Y.t)Nov  -35.11160   44.49293  -0.789  0.43273
season(Y.t)Dec   58.20198   44.63670   1.304  0.19660
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 75.81 on 69 degrees of freedom
Multiple R-squared:  0.7115,    Adjusted R-squared:  0.6572
F-statistic: 13.09 on 13 and 69 DF,  p-value: 6.321e-14

> checkresiduals(model5.1)

        Breusch-Godfrey test for serial correlation of order up to 17

data:  Residuals
LM test = 16.885, df = 17, p-value = 0.4622
```

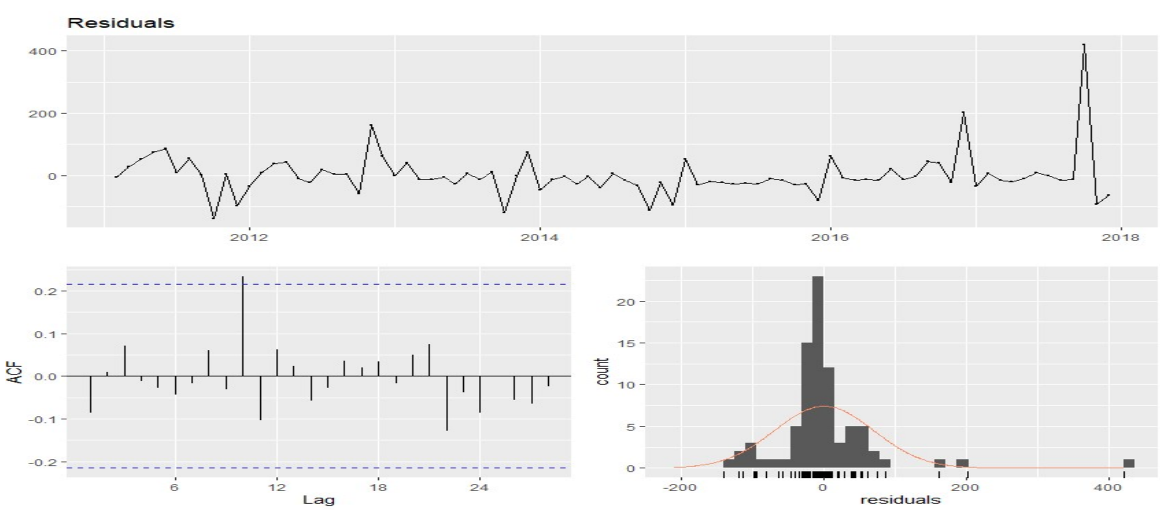

Figure 13: Residuals of Dynamic Linear Model

Exponential Smoothing Methods -Holt winters seasonal Methods

Exponential smoothing methods are used to model time series data by fitting different trend and seasonality pattern which are

Simple Exponential Smoothing.

Holt's Linear Method.

Holt-Winters trend and seasonality method.

The above methods consist different components which are used in different conditions according to the data or the series on which the analysis is performed.

Since there is no existence of trend in the series and there is no high seasonality in the series the Simple exponential smoothing, Holt's-linear, damped, Exponential Trend Models are not being considered and use of Holt-Winters’ additive errors method is used to analyse the series.

From the analysis of the Holt-Winters’ additive method the models were not good models from the residual analysis.

The model Holt-Winters' multiplicative method has the lowest MASE value 0.7715 and summary of this model shows AIC = 1107.791 , AICc =1117.064 , BIC =1149.115 (shown in below output of holt winters method), but from the residuals check (*Figure 14*) it is observed that the series is not random, ACF plot has a significant seasonal lag and histogram is not normally distributed .

`Output of Holt winter’s method:`

```
     AIC     AICc      BIC 
1107.791 1117.064 1149.115 

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set -5.333804 107.6944 58.64766 -51.22871 83.57982 0.7715387 -0.09336848

Forecasts:
         Point Forecast       Lo 80     Hi 80      Lo 95     Hi 95
Jan 2018      235.74983  -115.26638  586.7660  -301.0831  772.5828
Feb 2018       89.95389   -76.47901  256.3868  -164.5833  344.4910
Mar 2018       79.19006   -98.05692  256.4370  -191.8858  350.2659
Apr 2018      124.66416  -207.42646  456.7548  -383.2246  632.5529
May 2018       34.35024   -73.40436  142.1048  -130.4462  199.1467
Jun 2018       53.24639  -142.02618  248.5190  -245.3972  351.8900
Jul 2018       55.25550  -180.44554  290.9566  -305.2181  415.7291
Aug 2018       71.54870  -282.15293  425.2503  -469.3913  612.4887
Sep 2018       71.56751  -337.35949  480.4945  -553.8324  696.9674
Oct 2018      194.95374 -1090.05433 1479.9618 -1770.2965 2160.2040
> checkresiduals(fit7.hw)

        Ljung-Box test

data:  Residuals from Holt-Winters' multiplicative method
Q* = 24.339, df = 3, p-value = 2.122e-05

Model df: 16.   Total lags used: 19
```

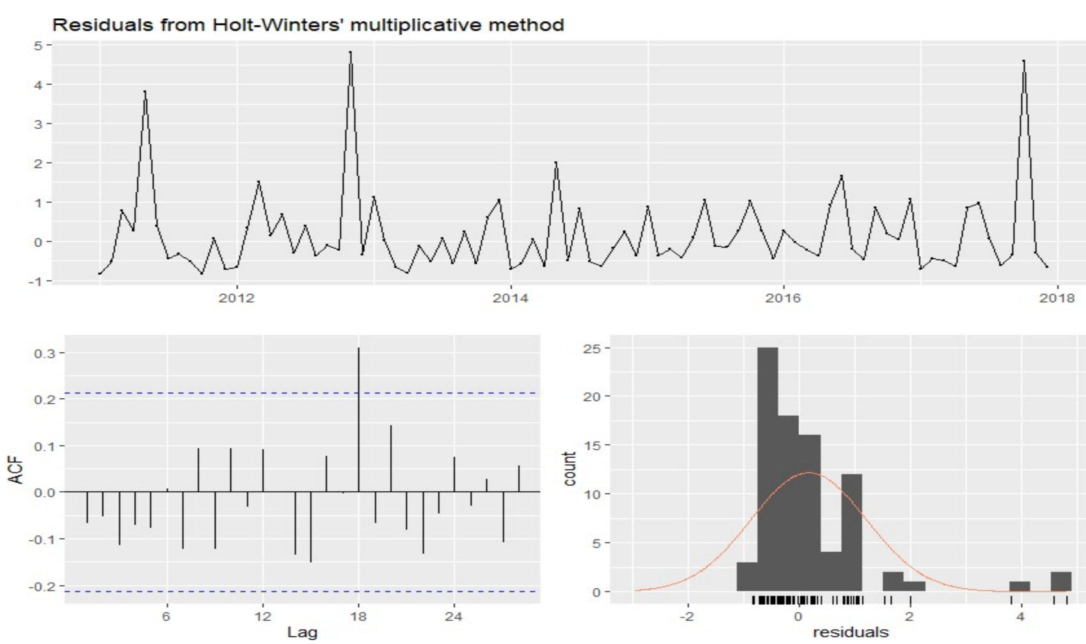

Figure 14: Residuals of Holts winters multiplicative method

Innovation state space models:

Further Innovation State Space Models were used to analyse the series.
Models using Additive Errors, Multiplicative errors, Multiplicative trend, Additive trend, Multiplicative Seasonality, Additive Seasonality and damped component and all the combinations were used to get the best model.
From analysing each models it was concluded that the model with the Additive errors, no trend and additive seasonality (ANA) has the lowest MASE value 0.729(shown in Output of ETS ANA model) from all the other models but the residuals check failed as the ACF plot has one significant lag) (*Figure 15*). The model with Additive Errors, Additive Damped Trend and Additive seasonality (AAdA) where the bounds are admissible has the next best MASE value 0.8091 (shown in Output of ETS(A,Ad,A))and the residual analysis (*Figure 16*) shows the series is not random, ACF plot has no significant lag , histogram has long tails as previous models ,comparatively good from other models. Hence this model can be considered for forecasting.The following table 5 shows some of the fitted models AIC,BIC,AICc and MASE values.

Output of ETS ANA model

```
ETS(A,N,A)

Call:
 ets(y = Abandoned_Calls.ts, model = "ANA")

  Smoothing parameters:
    alpha = 1e-04
    gamma = 1e-04

  Initial states:
    l = 77.8342
    s = 64.8116 -8.025 56.9691 -37.3848 -40.6577 -44.4462
           -42.8563 -44.9667 -14.4543 -30.0146 1.0083 140.0166

  sigma:  127.3966

     AIC     AICc      BIC
1201.221 1208.280 1237.683

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 6.843276 116.2966 60.21141 -53.26734 87.23013 0.7921106 0.2703156
> checkresiduals(fit1.etsA)

        Ljung-Box test

data:  Residuals from ETS(A,N,A)
Q* = 14.809, df = 3, p-value = 0.001988

Model df: 14.   Total lags used: 17
```

ETS model	AIC	BIC	AICc	MASE
A,N,A	1201.221	1237.683	1208.280	0.7921
AAdA(bounds=Ad)	1205.337	1249.092	1215.860	0.8091
AAA	1208.592	1249.916	1217.864	0.8955
MNA	1099.769	1136.231	1106.828	0.8211
MAdM	1039.941	1083.695	1050.464	0.8391
MAM	1036.051	1077.375	1045.324	0.8669
MAN	1171.477	1186.062	1172.568	0.8686
MAdN	1171.477	1186.062	1172.568	0.8686
MAA	1111.348	1152.672	1120.620	1.0568
MAdA	1158.55	1202.710	1169.478	0.9474

Table 5:AIC, BIC,AICs and MASE values of space Models

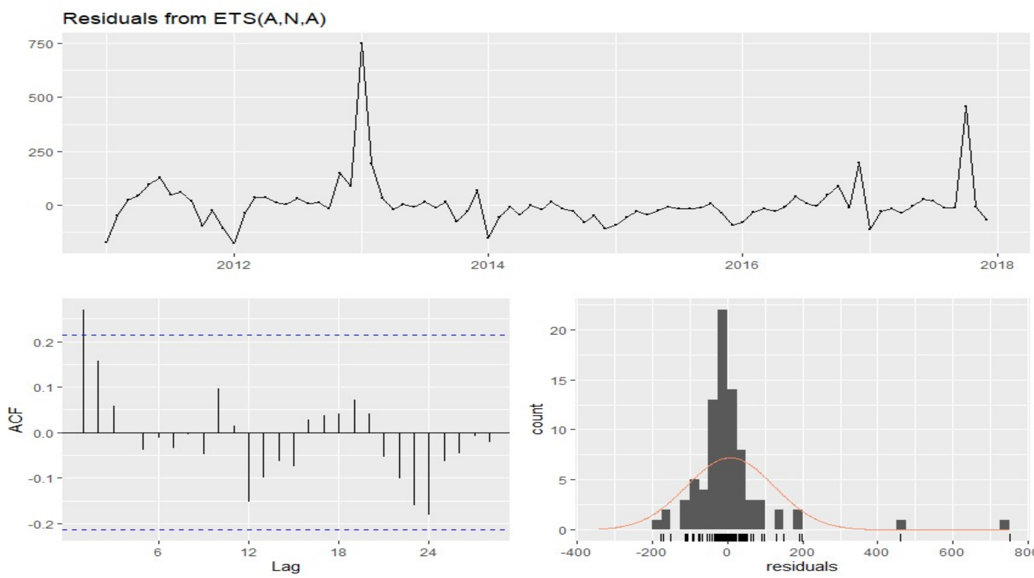

Figure 15: Residuals of ETS(A,N,A)

Output of ETS(A,Ad,A)

```
ETS(A,Ad,A)

Call:
 ets(y = Abandoned_Calls.ts, model = "AAA", damped = TRUE, bounds = "admiss")

  Smoothing parameters:
    alpha = 0.1977
    beta  = -0.0304
    gamma = 0.001
    phi   = 0.8463

  Initial states:
    l = 57.8346
    b = 3.2135
    s = 68.284 -12.2153 32.5161 -30.5406 -40.9698 -29.9016
           -37.9457 -53.9976 -31.1495 -32.9093 4.9849 163.8443

  sigma:  128.7657

     AIC     AICc      BIC
1205.337 1215.860 1249.092

Training set error measures:
                   ME RMSE      MAE       MPE     MAPE      MASE      ACF1
Training set 3.685842  115 61.50642 -42.27548 85.17789 0.8091471 0.1019839
> checkresiduals(fit.AAA1)#GOOD MODEL

        Ljung-Box test

data:  Residuals from ETS(A,Ad,A)
Q* = 11.472, df = 3, p-value = 0.009431

Model df: 17.   Total lags used: 20
```

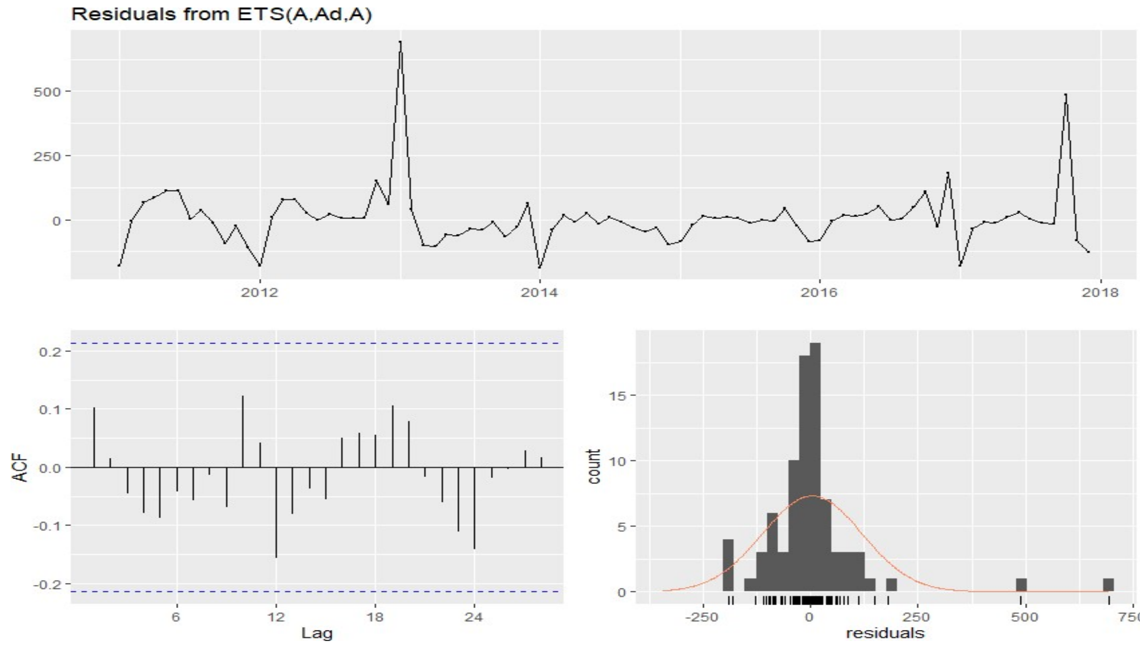

Figure 16: Residuals of ETS(A,Ad,A)

Forecasting

Our main data set has values up to 2018 October so we have used values from January 2018 to October 2018 of the total calls to plot the forecast for DLM along with original abandoned call series. After obtaining the best models from the modelling methods, these models are used to forecast their series.

The forecasts for Finite DLM (*Figure 17*) and ARDLM (*Figure 18*) are similar for next 10 months, but the MASE value for Finite DLM is 0.5653 which is less than MASE value of ARDLM which is 0.6329.The forecast for the Intervention Model is not good in the residuals check and also the forecast (*Figure 19*) is not good.

The 1000 Simulated future sample paths are shown in *Figure 20.*

The forecasts for model ETS(AAdA) with bounds admissible (*Figure 21*) shows the mean prediction with 5% upper limit and 95% lower limit.

As it can be observed from the forecasts for Finite DLM (*Figure 17*) it captures the seasonality and follows the trend of the original series and looks better than the other models , hence this model can be considered as the best fit model for forecasting.

The following are the forecasts:

Finite Dlm

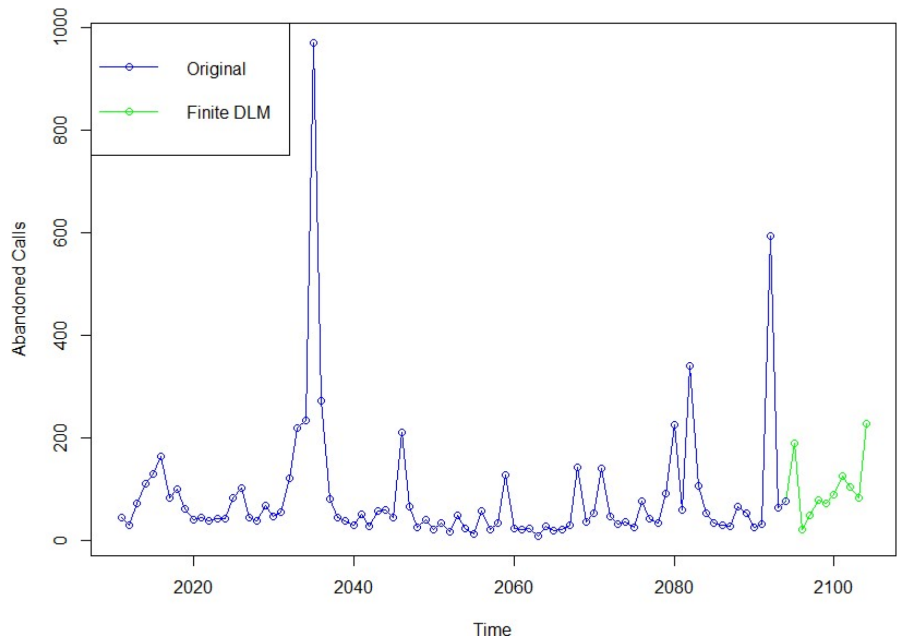

Figure 17: Forecast of Finite DLm

ARDLM

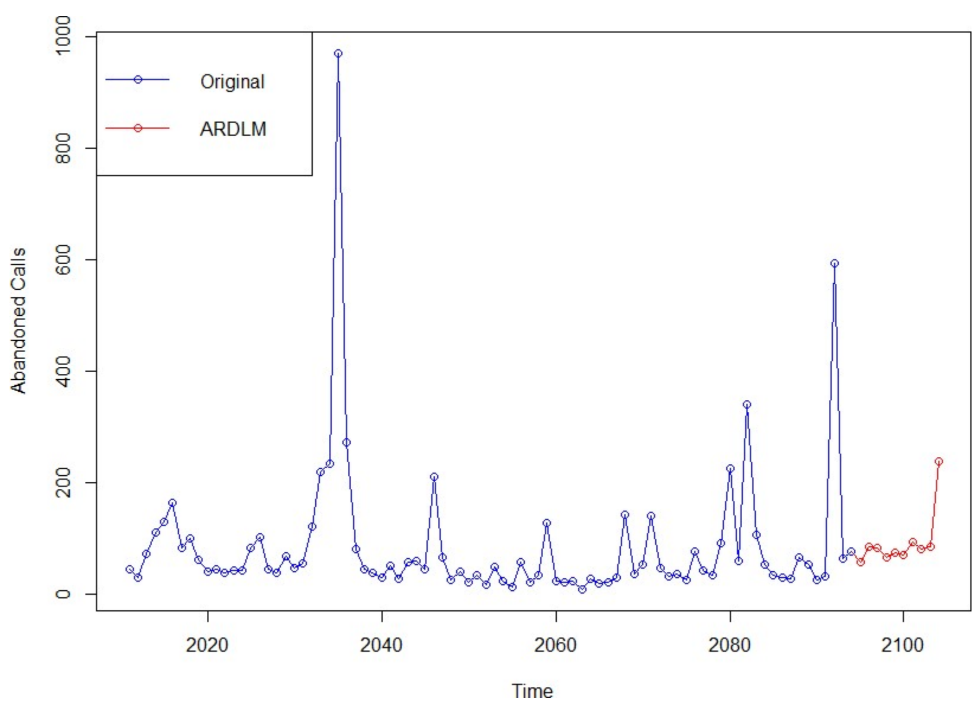

Figure 18: Forecast of ARDLM

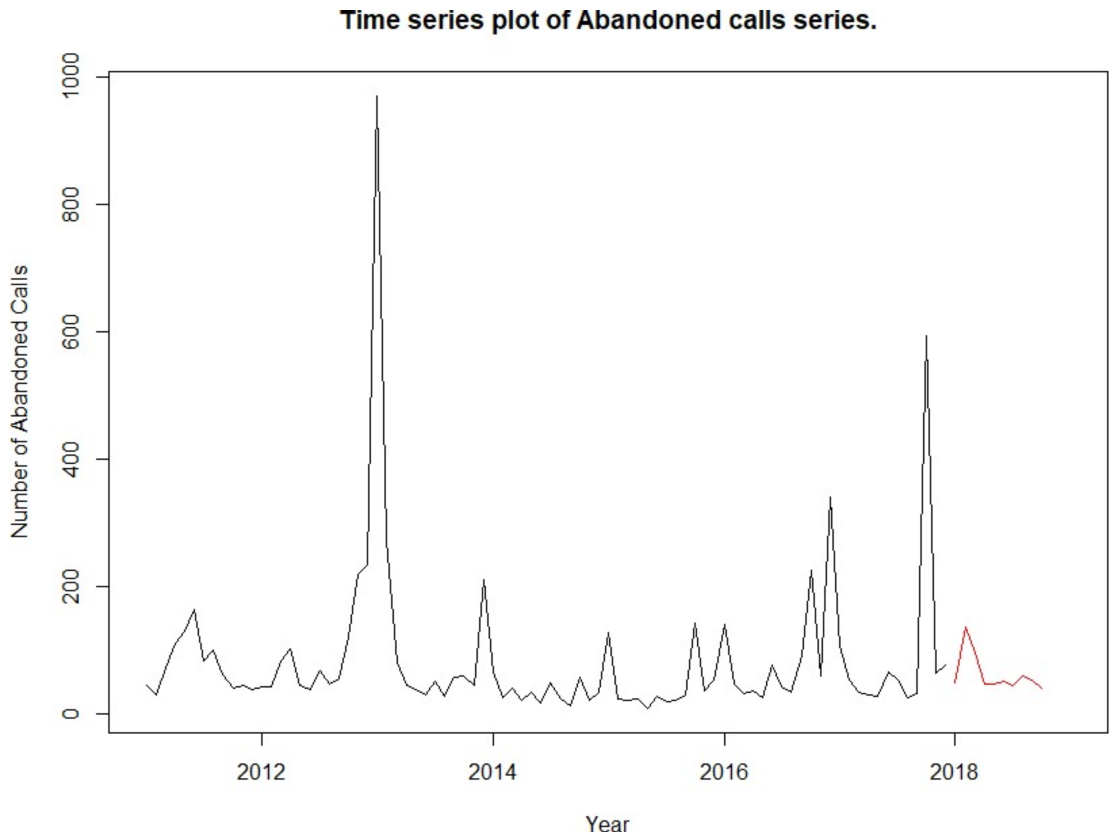

Figure 19: Forecast of Dynamic Linear Model

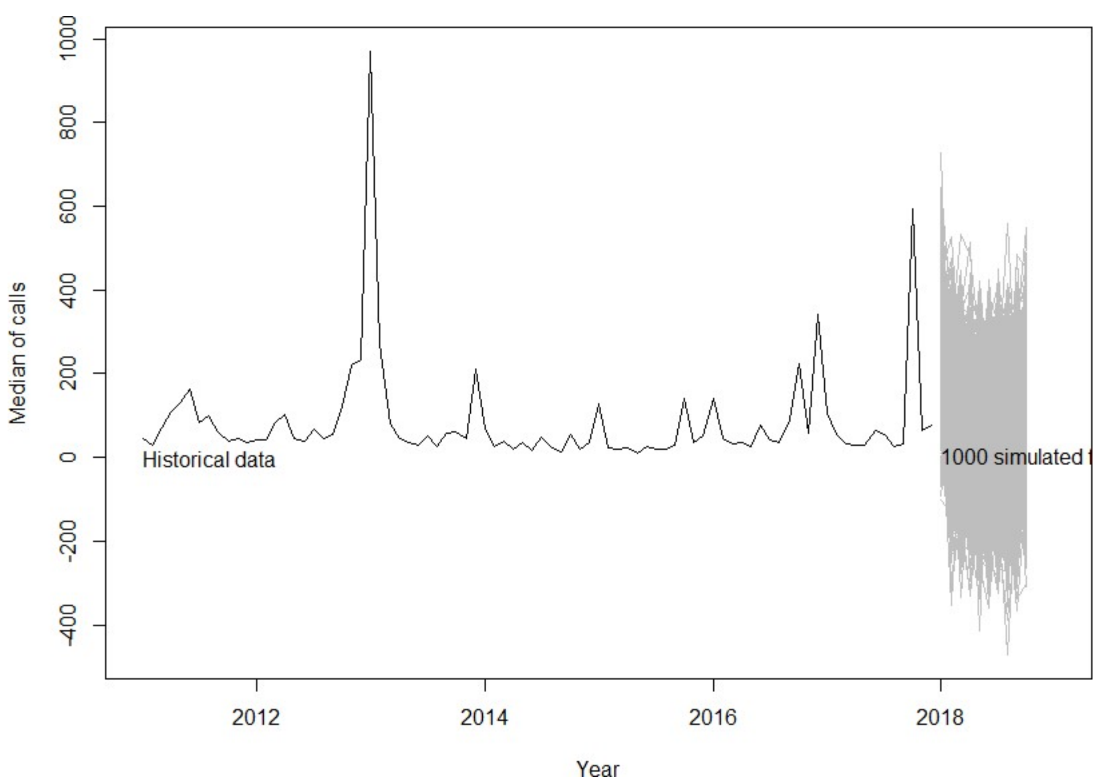

Figure 20: Forecast of plots with 1000 simulations

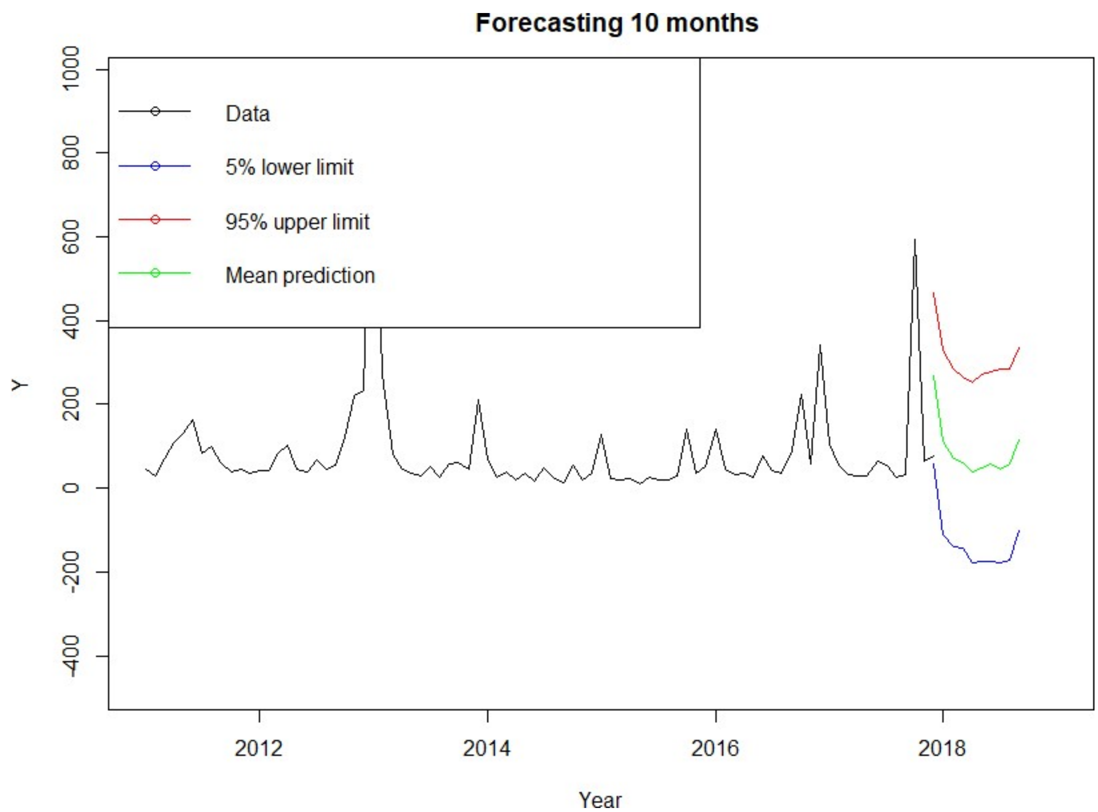

Figure 21: Forecast With ETS(A,Ad,A)

Reference

U.S. Department of Health & Human Services,(2018),Call Centre Metrics for the Health Service System: Data and Resources, Available at:https://healthdata.gov/dataset/call-center-metrics-health-service-system

Article: SFGATE,(2018),'California's deadly flu season could be worst in a decade', Available

https://www.sfgate.com/health/article/It-s-early-but-California-s-deadly-flu-12485751.php

https://www.ucsf.edu/news/2013/02/13486/2013-flu-season-how-protect-yourself
Flu seasons :San Francisco

https://www.sfchronicle.com/health/article/Troubling-early-signs-as-flu-season-approaches-14499535.php

https://cxcentral.com.au/kpis/popular-call-centre-metrics-and-kpis/

Appendix

```
library(TSA)
library(tseries)
library(car)
library(dLagM)
library(forecast)
library(dplyr)
library(tidyr)
library(Hmisc)
library(reshape)
library(lubridate)
library(dynlm)
library(expsmooth)
library(x12)
library(urca)
library(readr)

source('C:/Users/gadal/Desktop/forecasting_MATH1307/week3/MATH1307_utilityFunctions.R')



#Reading the data
Call_Metrics <- read_csv("Call_Metrics.csv")
#View(Call_Metrics)
dim(Call_Metrics)
head(Call_Metrics, 5)
class(Call_Metrics)
Call_Metrics <- Call_Metrics[order(as.Date(Call_Metrics$Month, format="%m/%d/%Y")),]
Call_Metrics <- Call_Metrics[-c(85:94),]

Call_Metrics.ts<-ts(Call_Metrics[,c(4,7)], start = c(2011,1), end=c(2017,12), frequency = 12)
head(Call_Metrics.ts)
dim(Call_Metrics.ts)
class(Call_Metrics.ts)
```

```
Total_Calls.ts <- ts(Call_Metrics[,7], start = c(2011,1), end=c(2017,12), frequency = 12)

class(Total_Calls.ts)

plot(Total_Calls.ts,type='o',ylab='Number of calls Received', main = 'Monthly Total Calls Received by Call Center')

plot(Total_Calls.ts,ylab='Number of calls Received', main = 'Monthly Total Calls Received by Call Center')

points(y=Total_Calls.ts,x=time(Total_Calls.ts),pch=as.vector(season(Total_Calls.ts)))

#no trend, seasonality not clear ,clear changing variance , intervention point not clear, moving average behaviour.

par(mfrow=c(1,2))

acf(Total_Calls.ts, lag.max =60,main="ACF of Total calls") #trend, seasonality may exist as there is a pattern

pacf(Total_Calls.ts, lag.max =60 ,main="PACF of Total calls")

par(mfrow=c(1,1))

adf.test(Total_Calls.ts)#stationary

#and how there has been variations in the calls during and flu seasons.


BC1= BoxCox.ar(Total_Calls.ts)

BC1$ci # lambda value is nearly zero , so going with log transformation

Total_Calls.tr = log(Total_Calls.ts)

plot(Total_Calls.tr,ylab='Log of number of Total calls',xlab='Year',main = "Time series plot of the logarithm of
monthly

   Total calls")

points(y=Total_Calls.tr,x=time(Total_Calls.tr), pch=as.vector(season(Total_Calls.tr)))

#transformation did not work well for this series

adf.test(Total_Calls.tr)

acf(Total_Calls.tr, lag.max =60,main="ACF of Total calls")


Abandoned_Calls.ts <- ts(Call_Metrics[,4], start = c(2011,1), end=c(2017,12), frequency = 12)

class(Abandoned_Calls.ts)

head(Abandoned_Calls.ts)

par(mfrow=c(1,1))

plot(Abandoned_Calls.ts,type='o',ylab='Number of calls Abandoned', main = 'Monthly Abandoned Calls of Call
Center')

plot(Abandoned_Calls.ts,ylab='Number of Abandoned calls ', main = 'Monthly Abandoned Calls of Call Center')

points(y=Abandoned_Calls.ts,x=time(Abandoned_Calls.ts),pch=as.vector(season(Abandoned_Calls.ts)))

#clear presence of intervention point at January 2013, no trend , seasonality not clear , there is changing variance ,
moving average behaviour
```

```
par(mfrow=c(1,2))

acf(Abandoned_Calls.ts, lag.max = 30,main="ACF of Abandoned Calls") #no trend & seasonality may exist as there is a pattern

pacf(Abandoned_Calls.ts, lag.max = 48,main="PACF of Abandoned Calls")

par(mfrow=c(1,1))

adf.test(Abandoned_Calls.ts)#stationary, p value leass than 0.05


BC= BoxCox.ar(Abandoned_Calls.ts)

BC$ci

Abandoned_Calls.BC = BoxCox(Abandoned_Calls.ts, lambda = -0.35 )

Abandoned_Calls.tr = log(Abandoned_Calls.ts)#doing log transformation to smoothen it a little bit .

plot(Abandoned_Calls.BC,ylab='Log of number of abandoned calls',xlab='Year',main = "Boxcox transformed Time series plot of monthly

    Abandoned calls")

points(y=Abandoned_Calls.BC,x=time(Abandoned_Calls.BC), pch=as.vector(season(Abandoned_Calls.BC)))

#transformation did not work well for this series

adf.test(Abandoned_Calls.BC)#stationary, p value leass than 0.05

acf(Abandoned_Calls.BC, lag.max = 30,main="ACF of Abandoned Calls")#looks like there is little seasonality.


#Doing ordinary differencing

Abandoned_Calls.diff = diff(Abandoned_Calls.BC)

plot(Abandoned_Calls.diff,ylab='number of abandoned calls',xlab='Year',main = "Time series plot of the differenced data of monthly

    Abandoned calls")

points(y=Abandoned_Calls.diff,x=time(Abandoned_Calls.diff), pch=as.vector(season(Abandoned_Calls.diff)))

#changing variance still exists, cannot see any seasonality.

acf(Abandoned_Calls.diff, lag.max = 30,main="ACF of Abandoned Calls")

adf.test(Abandoned_Calls.diff)#stationary, p value leass than 0.05


#Doing seasonal differencing

Abandoned_Calls.diff1 = diff(Abandoned_Calls.tr, differences=1, lags=12)

plot(Abandoned_Calls.diff1,ylab='number of abandoned calls',xlab='Year',main = "Time series plot of the differenced data of monthly

    Abandoned calls")

points(y=Abandoned_Calls.diff1,x=time(Abandoned_Calls.diff1), pch=as.vector(season(Abandoned_Calls.diff1)))

#same as the ordinary differencing
```

```
acf(Abandoned_Calls.diff1, lag.max = 30,main="ACF of Abandoned Calls")

adf.test(Abandoned_Calls.diff1)#stationary, p value leass than 0.05



#Decomposition :

##classic Decomposition

calls.decom.classic1=decompose(Abandoned_Calls.ts,type ="additive")

par(mfrow=c(1,1))

plot(calls.decom.classic1)


#calls.decom.classic2=decompose(Abandoned_Calls.ts,type ="multiplicative")

#par(mfrow=c(1,1))

#plot(calls.decom.classic2)


calls.decom.x12 = x12(Abandoned_Calls.ts)

par(mfrow=c(1,1))

plot(calls.decom.x12 , sa=TRUE , trend=TRUE)

#the fluctuations are higher than the seasonally adjusted line , it means there is something else happening and not
seasonality.



# We can scale and center both series to see the same plots clearly

data1.scaled = scale(Call_Metrics.ts)

plot(data1.scaled, plot.type="s",col = c("blue","black"), main = "Scaled Series")

legend("topright",lty=1, pch = 1, text.width = 11, col=c("blue","black"), c("Abandoned Calls","Total Calls"))


tc.ac=ts.intersect(Total_Calls.ts,Abandoned_Calls.ts)

plot(tc.ac,yax.flip=T)


#Checking Correlation

cor(Call_Metrics.ts)#not very highly correlated but correlated

#positive correlation 0.5425359 implies when Total Calls increase the Abandoned Calls will increase.


#DLag Models
```

```
#for dlm models we do not need stationarity in the series


##1.Finite DLM’s:

# No need to supply ts object. The function looks for vectors.

# Finding the finite lag length based on AIC and BIC for this data

# We can change q and compare AIC/BIC values

for ( i in 1:10){

 model1.i = dlm( x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts), q = i )

 cat("q = ", i, "AIC = ", AIC(model1.i$model), "BIC = ", BIC(model1.i$model),"\n")

}

# We go for the smallest AIC/BIC , here q = 10 has lowest AIC=871.8091 and BIC=901.7619

#the best model is always the model with lowest aic and lowest bic ,because the information lost for this model is low.

#checking the coefficients and doing residuals check for the lowest AIC and BIC model

model1.10 = dlm( x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts), q = 9)

summary(model1.10)#Adjusted R-squared= 0.6547 which is low, pvalue is less than 0.05 which is good ,

checkresiduals(model1.10$model)#one significant lag in ACF , long tail in histogram

MASE(model1.10)#0.5667715

summary(model1.10)

checkresiduals(model1.10$model)

MASE(model1.10)

VIF.model1.10 = vif(model1.10$model)

VIF.model1.10 > 10

VIF.model1.10 = vif(model1.10$model)# variance inflation factors

# Notice again we are getting the model object out of the model fitting

# by using "$model" in addition to "model1.10"

VIF.model1.10

VIF.model1.10 > 10

#From the VIF values, it is obvious that the estimates of the finite DLM are not suffering from the multicollinearity.

#We can use this model for forecasting


##2.Polynomial DLM’s :

# Fitting a polynomial DLM and Displaying the estimates of original parameters and their significance tests.

#we used a 2nd order polynomial
```

```
for (i in 2:10) {
  model2 =polyDlm(x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts) , q = i , k =2)
  cat("q = ", i, "AIC = ", AIC(model2$model),
      "BIC = ", BIC(model2$model),"\n")
}

#q=10 has lowest AIC = 902.5478 BIC = 914.0681
model2.102 = polyDlm( x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts), q = 10,k=2)
summary(model2.102)##all the coefficients are significant,pvalue less than 0.05, rsquared value is 0.429 which is low.
#Now, the model and all parameters are significant at 5% level of significance.
checkresiduals(model2.102$model)#one significant lag in ACF and a long tails in histogram.
MASE(model2.102)#0.6901556
VIF.model2.102 = vif(model2.102$model)
VIF.model2.102 > 10
#This model is slightly affected by the multicollinearity.
#We couldn’t find a better model in terms of overall significance tests,and adjusted r-squared measures with polynomial models.



summary(model2.102)
checkresiduals(model2.102$model)
MASE(model2.102)
VIF.model2.102 = vif(model2.102$model)
VIF.model2.102 > 10
##3.The Koyck transformation:

model3 = koyckDlm(x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts) )
summary(model3)#r squared value is very low 0.277,coefficient X.t not significant , pvalue less than 0.05 which is good
checkresiduals(model3$model)
MASE(model3)#0.6697285
#We got a significant model (p-value = 0.01066) with one significant coefficient at 5% and a lower
#adjusted r-squared value with the Koyck’s transformation.BUt with the residuals analysis this model is ot good
```

```r
summary(model3)

checkresiduals(model3$model)

MASE(model3)

VIF.model3 = vif(model3$model)

VIF.model3 > 10


##4.ARDL models


for (i in 1:5){

  for(j in 1:5){

    model4 = ardlDlm(x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts), p = i , q = j )

    cat("p = ", i, "q = ", j,

    "AIC = ",  AIC(model4$model),

    "BIC = ", BIC(model4$model),"\n")

  }

}


#p = 5 q = 2 has lowest AIC = 933.2973 ,BIC = 956.9918 values

model4.52 = ardlDlm(x = as.vector(Total_Calls.ts) , y = as.vector(Abandoned_Calls.ts) ,p = 2 , q = 2)

summary(model4.52)#Adjusted R-squared: 0.605 which is low ,pvalue is less than 0.05

checkresiduals(model4.52$model)#residual check histogram has long tails and there are outliers, ACF Plot shows
there is no serial auto correlation , BG test confirms the same .

MASE(model4.52)#0.6216514

#According to ARDL(5,2) model, we can conclude that abandoned calls might be related with its levels in the past
five years and total calls in the past two years.

#However, this is not a strong concLusion due to the moderate level of the adjusted R-square,which tells us that
there are some other variables realted with abandoned calls need to be added to the model.

summary(model4.52)

checkresiduals(model4.52$model)

MASE(model4.52)

VIF.model4.52<-vif(model4.52$model)

VIF.model4.52 > 10

#From the VIF values, it is obvious that the estimates of the ardlDlm are not suffering from the multicollinearity.

#We can use this model for forecasting
```

#Dynamic Linear Models :

#INTERVENTION ANALYSIS:

#intervention affected the process by changing the mean level

#if the mean level does not goes back to mean level as it was before ,after the intervention ,we use step function or else use pulse function.

#Models for pluse response intervention

```
Y.t = Abandoned_Calls.ts

T = 25 # the time intervention occured

P.t = 1*(seq(Y.t)== T)

P.t.1 = Lag(P.t,+1)


#First lag of Y.t

model5.1 = dynlm(Y.t ~ L(Y.t , k = 1 ) + P.t + season(Y.t))

summary(model5.1)

checkresiduals(model5.1)
```

#months are Not significant .

#P.t & season(Y.t) are significant,pvalue less than 0.05, but rsquared value is moderate.

bgtest p-value = 0.4622 implies no serial correlation left .

#but ACF lot has one significant lag , plot is not random and histogram not normally distributed and has avery long tail.

```
#second lag of Y.t

model5.2= dynlm(Y.t ~ L(Y.t , k = 2 ) + P.t + trend(Y.t) + season(Y.t))

summary(model5.2)


#first lag of P.t

model5.3 = dynlm(Y.t ~ L(Y.t , k = 1 ) + P.t + P.t.1 + trend(Y.t) + season(Y.t))

summary(model5.3)
```

#dropping P.t.1 as it is not significant in the previous modes.

#checking the trend and seasonality in seperate models

```
model5.4 = dynlm(Y.t ~ L(Y.t , k = 1 ) + P.t + season(Y.t))
```

```
summary(model5.4)#everything is significant except for few months and P.t is not significant,adjusted rsquared value is low.

model5.5 = dynlm(Y.t ~ L(Y.t , k = 1 ) + P.t + trend(Y.t) )

summary(model5.5)

#taking first lag of P.t and season component

model5.6 = dynlm(Y.t ~ L(Y.t , k = 1 ) +P.t + P.t.1 + season(Y.t))

summary(model5.6)

#taking second lag of y.t and first lag of P.t

model5.7 = dynlm(Y.t ~ L(Y.t , k = 2 ) + P.t + P.t.1+ season(Y.t) )

summary(model5.7)

#adding another lag

model5.8 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + P.t+ P.t.1 + season(Y.t))

summary(model5.8)

model5.9 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + P.t.1 + season(Y.t))

summary(model5.9)

aic = AIC(model5.1, model5.2, model5.3, model5.4, model5.5, model5.6, model5.7, model5.8,model5.9)

bic = BIC(model5.1, model5.2, model5.3, model5.4, model5.5, model5.6, model5.7, model5.8,model5.9)

sort.score(aic,score = "aic")

sort.score(bic,score = "bic")

checkresiduals((model5.1))

###TEsting extra models:

#adding third lag

model5.91 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 ) + S.t.1 + season(Y.t))

summary(model5.91)

#adding everything , lags,trend ,seasonality.

model5.92 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 ) + P.t + P.t.1 + trend(Y.t) + season(Y.t))

summary(model5.92)

#adding 4th lag

model5.93 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 ) + L(Y.t , k = 4 ) + P.t + P.t.1 + trend(Y.t) + season(Y.t))
```

```
summary(model5.93)

#going back to model from the beginning and get rid of the seasonality from there,

#so do seasonal differencing of y.t

Y.t.1 = diff(Y.t, lag = 12, differences = 1)

P.t2 = 1*(seq(Y.t.1) == T)

P.t.12 = Lag(P.t2,+1)


model5.94 = dynlm(Y.t.1 ~ L(Y.t.1 , k = 1 ) + P.t2 + season(Y.t.1))

summary(model5.94)


aic = AIC(model5.1, model5.4, model5.9)

bic = BIC(model5.1, model5.4, model5.9)

sort.score(aic,score = "aic")

sort.score(bic,score = "bic")


#Checking for spurious correlation :

total <- Call_Metrics.ts[,2]

abandoned<-Call_Metrics.ts[,1]


calls.joint=ts.intersect(total,abandoned)#intersect them

plot(calls.joint,yax.flip=T)

#no trend, changing variance, moving average and auto regressive behaviour, seasonality, intervention point


ccf(as.vector(calls.joint[,1]), as.vector(calls.joint[,2]),ylab='CCF', main = "Sample CCF between Total Calls and 
Abandoned Calla")


me.dif=ts.intersect(diff(diff(total,12)),diff(diff(abandoned,12)))

prewhiten(as.vector(me.dif[,1]),as.vector(me.dif[,2]),ylab='CCF', main="Sample CCF after prewhitening ")

not spurious correlation ,means correlated to each other.


#Using exponential smoothing methods to compare trend and seasonal component in time series data:

#considering there is no existence of trend we are ignoring simple exponential smoothing,and

#holt's - linear , damped ,exponential trend models

#Fitting holt-winter's seasonality methods:
```

```
#additive seasonality method:
fit5.hw <- hw(Abandoned_Calls.ts,seasonal="additive", h=10)
summary(fit5.hw)
checkresiduals(fit5.hw)

fit6.hw <- hw(Abandoned_Calls.ts,seasonal="additive",damped = TRUE, h=10)
#additive seasonality model which is damped
summary(fit6.hw)
checkresiduals(fit6.hw)

#Mutiplicative seasonality method:
fit7.hw <- hw(Abandoned_Calls.ts,seasonal="multiplicative", h=10)
summary(fit7.hw)
checkresiduals(fit7.hw)

#Innovation State Space Models :
#Linear Homoscedastic state space models
#Models with additive errors

fit.ANN=ets(Abandoned_Calls.ts, model="ANN")#Additive errors, no trend and no seasonality
summary(fit.ANN)#0.871512
checkresiduals(fit.ANN)#

fit.AAN1=ets(Abandoned_Calls.ts, model="AAN")#Additive errors with additve trend and no seasonality
summary(fit.AAN1)#MASE is 0.8524016,AIC is 1197.736
checkresiduals(fit.AAN1)#

fit.AAN2=ets(Abandoned_Calls.ts, model="AAN",damped = TRUE)#Additive errors with additve trend which is
damped and no seasonality
summary(fit.AAN2)#MASE is 0.8854373,AIC is 1199.345
checkresiduals(fit.AAN2)#same as previous model

fit1.etsA = ets(Abandoned_Calls.ts, model="ANA")
summary(fit1.etsA)
```

```
checkresiduals(fit1.etsA)

fit2.etsA = ets(Abandoned_Calls.ts, model="AAA")
summary(fit2.etsA)#MASE=0.8955713,AIC=1208.592
checkresiduals(fit2.etsA)#residuals ok ,longtail at end

fit3.etsA = ets(Abandoned_Calls.ts, model="AAA",damped = TRUE)
summary(fit3.etsA)#MASE=0.8879911,AIC=1210.027
checkresiduals(fit3.etsA)#residuals ok

fit.AAA1 = ets(Abandoned_Calls.ts, model = "AAA", damped = TRUE, bounds="admiss" )
summary(fit.AAA1)#MASE=0.8091471,AC=1205.337
checkresiduals(fit.AAA1)#GOOD MODEL

fit.AAA3 = ets(Abandoned_Calls.ts, damped = FALSE, model = "AAA", upper=rep(1,4))
summary(fit.AAA3)#MASE=0.8956307,AIC =1208.592
checkresiduals(fit.AAA3)

fit.AAA4 = ets(Abandoned_Calls.ts, damped = TRUE, model = "AAA", opt.crit = "mse")
summary(fit.AAA4)#MASE =0.8879911,AIC =1208.592
checkresiduals(fit.AAA4)#GOOD

fit.AAA5 = ets(Abandoned_Calls.ts, damped = FALSE, model = "AAA", opt.crit = "mae")
summary(fit.AAA5)#MASE=0.8922778,AIC=1210.025
checkresiduals(fit.AAA5)

#Linear heteroscedastic state-space models:
#Models with multiplicative errors:

fit.MNN = ets(Abandoned_Calls.ts, model="MNN")
summary(fit.MNN)#MASE=0.8828339,AIC=1169.103
checkresiduals(fit.MNN)

fit6.etsA = ets(Abandoned_Calls.ts, model="MNA")
```

```
summary(fit6.etsA)#MASE=0.8211372,AIC=1099.769

checkresiduals(fit6.etsA)#residuals not ok


fit8.etsA = ets(Abandoned_Calls.ts, model="MAN")

summary(fit8.etsA)#MASE=0.8686499,AIC=1171.477

checkresiduals(fit8.etsA)#residuals not ok


fit9.etsA = ets(Abandoned_Calls.ts, model="MAN",damped = TRUE)

summary(fit9.etsA)#MASE=0.8686499,AIC=1171.477

checkresiduals(fit9.etsA)#same as previous model


fit.MAA = ets(Abandoned_Calls.ts, model="MAA")

summary(fit.MAA)#MASE=0.8686499,AIC=1171.477

checkresiduals(fit.MAA)#residuals ok


fit.MAA1 = ets(Abandoned_Calls.ts, model="MAA",damped = TRUE)

summary(fit.MAA1)#MASE=0.9474353,AIC=1158.955

checkresiduals(fit.MAA1)#same as previous model


#Nonlinear seasonal state-space models

fit7.etsA = ets(Abandoned_Calls.ts, model="MNM")

summary(fit7.etsA)#MASE=0.8425208,AIC=1028.623

checkresiduals(fit7.etsA)#residuals ok, significant lags in ACF plot


fit10.etsA = ets(Abandoned_Calls.ts, model="MAM")

summary(fit10.etsA)#MASE=0.8669716,AIC=1036.051

checkresiduals(fit10.etsA)#residuals ok


fit11.etsA = ets(Abandoned_Calls.ts, model="MAM",damped = TRUE)

summary(fit11.etsA)#MASE=0.8391306,AIC=1039.941

checkresiduals(fit11.etsA)#same as previous model


#Models with multiplicative errors, multiplicative trend, and either no seasonality or multiplicative seasonality
```

fit12.etsA = ets(Abandoned_Calls.ts, model="MMN")

summary(fit12.etsA)#MASE=0.8685622,AIC=1162.103

checkresiduals(fit12.etsA)#residuals not ok , significant lag in ACF but it is too far so we can ignore it ,long tail in histogram,

fit13.etsA = ets(Abandoned_Calls.ts, model="MMN",damped = TRUE)

summary(fit13.etsA)#MASE=0.8311464,AIC=1164.510

checkresiduals(fit13.etsA)#same as previous model

fit14.etsA = ets(Abandoned_Calls.ts, model="MMM")

summary(fit14.etsA)#MASE=0.9308495,AIC=1039.962

checkresiduals(fit14.etsA)#residuals ok

fit15.etsA = ets(Abandoned_Calls.ts, model="MMM",damped = TRUE)

summary(fit15.etsA)#MASE=0.9308495,AIC=1039.962

checkresiduals(fit15.etsA)#same as previous model

MASE(model1.10,model4.52)

aic = AIC(model1.10,model4.52)

bic = BIC(model1.10,model4.52)

sort.score(aic,score = "aic")

sort.score(bic,score = "bic")

#Generate forecasts of Abandoned calls series over finite DL and ARDL models for 10 months.

model1.10.forecast=dLagM::forecast(model1.10,x=c(5104,3768,3929,4688,5155,4545,4388,5000,5250,12123),h=10)

model1.10.forecast

plot(ts(c(as.vector(Abandoned_Calls.ts), model1.10.forecast$forecasts), start = c(2011,1)),xlimtype="o", col="green", ylab = "Abandoned Calls")

lines(ts(as.vector(Abandoned_Calls.ts), start = c(2011,1)),col="Blue",type="o")

legend("topleft",lty=1, pch = 1, text.width = 11, col=c("blue","green"), c("Original","Finite DLM"))

model4.52.forecast=dLagM::forecast(model4.52,x=c(5104,3768,3929,4688,5155,4545,4388,5000,5250,12123),h=10)

```
model4.52.forecast


plot(ts(c(as.vector(Abandoned_Calls.ts), model4.52.forecast$forecasts), start = c(2011,1)),type="o", col="Red", ylab = "Abandoned Calls")

axis(1, at = seq(2011, 2018, by = 1), las=10)

lines(ts(as.vector(Abandoned_Calls.ts), start = c(2011,1)),col="Blue",type="o")

legend("topleft",lty=1, pch = 1, text.width = 11, col=c("blue","red"), c("Original","ARDLM"))


#Forecasting for Intervention model

q =10

n =nrow(model5.1$model)

abandoned.frc = array(NA , (n +q))

abandoned.frc[1:n] =Y.t[2:length(Y.t)]



for (i in 1:q) {

  months =array(0,11)

  months[(i+9)%%12] =1

  data.new =c(1,abandoned.frc[n-1+i], P.t[n] ,months)

  abandoned.frc[n+i] =as.vector(model5.1$coefficients) %*%data.new

}

par(mfrow=c(1,1))


plot(Y.t,xlim=c(2011,2019),ylab='Number of Abandoned Calls',xlab='Year',main ="Time series plot of Abandoned calls series.")

lines(ts(abandoned.frc[(n+1):(n+q)],start=c(2018,1),frequency =12),col="red")


# FORECASTING FOR ETS MODELS

A = ts(matrix(NA,10,1000),start=c(2018,1),frequency = 12)

M = 1000

for (i in 1:M){

  A[,i] = simulate(fit.AAA1 , initstate = fit.AAA1 $states[25,] , nsim=10)

  # Generate random epsilons and apply model formulation

}
```

```
par(mfrow=c(1,1))

plot(Abandoned_Calls.ts , ylim=range(Abandoned_Calls.ts,A) , xlim=c(2011,2019) , ylab="Median of calls" , xlab="Year")

for(i in 1:1000){

  lines(A[,i],col="gray")

}

text(2011,0,"Historical data",adj=0)

text(2018,0,"1000 simulated future sample paths",adj=0)


N = 10

data = Abandoned_Calls.ts

xlim=c(2011,2019)

Pi = array(NA, dim=c(N,2))

avrg = array(NA, N)

# Calcualte the interval estimates and mid point

for (i in 1:N){

  Pi[i,] = quantile(A[i,],type=8,prob=c(.05,.95))

  avrg[i] = mean(A[i,]) # This would be median as well

}

# Create ts objects for plotting

Pi.lb = ts(Pi[,1],start=end(data),f=12)

Pi.ub = ts(Pi[,2],start=end(data),f=12)

avrg.pred = ts(avrg,start=end(data),f=12)


plot(data,xlim=xlim , ylim=range(data,A),ylab="Y",xlab="Year", main="

    Forecasting 10 months")

lines(Pi.lb,col="blue", type="l")

lines(Pi.ub,col="red", type="l")

lines(avrg.pred,col="green", type="l")

legend("topleft", lty=1, pch=1, col=c("black","blue","red","green"), text.width = 4,

      c("Data","5% lower limit","95% upper limit","Mean prediction"))
```