Project

Mamun Mohammed (s3571301)

15 October 2017

## R Markdown

**library**(TSA)

**library**(forecast)

**library**(tseries)

library(dynlm)

library(ggplot2)

library(AER)

library(Hmisc)

library(dLagM)

library(readr)

library(car)

library(expsmooth)

library(dplyr)

library(xtable)

# Since we spotted intervention for the Mexico and Rest of the world series (others) we have run the dynamic linear models for those to run the intervention analysis as well

# for exponential smoothing methods and state space models--

We have used state space models with the automatic selection of model (ZZZ) and enhanced the model selection using the five different optimization criteria such as mse,amse,sigma,mae and lik for these three different series and here is the output for MASE,AIC and BIC

dl= length(data)  
  
accuracy.holder = c()  
for(i in seq(length(data))){   
 fit1 =ets(data[[i]],opt.crit = "mse")  
 fit2 =ets(data[[i]],opt.crit = "amse")  
 fit3 =ets(data[[i]],opt.crit = "sigma")  
 fit4 =ets(data[[i]],opt.crit = "mae")  
 fit5 =ets(data[[i]],opt.crit = "lik")  
   
   
 mase.fit1 = accuracy(fit1)[1,6]  
 mase.fit2 = accuracy(fit2)[1,6]  
 mase.fit3 = accuracy(fit3)[1,6]  
 mase.fit4 = accuracy(fit4)[1,6]  
 mase.fit5 = accuracy(fit5)[1,6]  
 AIC.fit1=AIC(fit1)  
 AIC.fit2=AIC(fit2)  
 AIC.fit3=AIC(fit3)  
 AIC.fit4=AIC(fit4)  
 AIC.fit5=AIC(fit5)  
 BIC.fit1=BIC(fit1)  
 BIC.fit2=BIC(fit2)  
 BIC.fit3=BIC(fit3)  
 BIC.fit4=BIC(fit4)  
 BIC.fit5=BIC(fit5)  
   
 accuracy.holder = c(accuracy.holder,mase.fit1 , mase.fit2 , mase.fit3 , mase.fit4 , mase.fit5 , AIC.fit1 , AIC.fit2,AIC.fit3,AIC.fit4, AIC.fit5,  
 BIC.fit1 , BIC.fit2,BIC.fit3,BIC.fit4, BIC.fit5)  
}  
print(accuracy.holder)

## [1] 0.6136585 0.6205520 0.6136585 0.6180026 0.6136597  
## [6] 5470.5670859 5479.9307190 5470.5670859 5473.0243248 5470.5671075  
## [11] 5519.2724469 5538.3771523 5519.2724469 5521.7296858 5519.2724686  
## [16] 0.5149088 0.5168390 0.5157364 0.4986766 0.5148549  
## [21] 5274.3315642 5276.8093195 5270.8505797 5278.0344099 5270.5860399  
## [26] 5323.0369253 5332.0087288 5326.0499889 5333.2338191 5325.7854491  
## [31] 0.4249198 0.4243551 0.4136072 0.4116296 0.4236739  
## [36] 5480.9681261 5481.5396306 5474.0156379 5481.1711236 5477.2805449  
## [41] 5539.4145594 5539.9860639 5522.7209990 5539.6175569 5535.7269782

models = c("ets.mse", "ets.amse","ets.sigma","ets.mae","ets.lik")  
exp.checks = c(" mase.fit1 ", " mase.fit2 ", " mase.fit3 ", " mase.fit4 ", " mase.fit5 ","AIC.fit1 ", " AIC.fit2 ", " AIC.fit3 ", " AIC.fit4 ", " AIC.fit5 ",   
 " BIC.fit1 ", " BIC.fit2 ", " BIC.fit3 ", " BIC.fit4 ", " BIC.fit5 ")  
  
Col1 = rep(1:dl,each = length(exp.checks)) # this gives the first column  
Col2 = rep(models,dl) # this gives column 2  
Col3 = c(rep("MASE",length(models)),rep("AIC",length(models)),rep("BIC",length(models))) # this gives column 3  
Col4 = accuracy.holder # this gives column 4  
  
expFINAL =do.call(rbind, Map(data.frame, Data.No=Col1, Model.type=Col2,Model.check = Col3,Value=Col4)) ## this creates the data frame  
expFINAL

## Data.No Model.type Model.check Value  
## 1 1 ets.mse MASE 0.6136585  
## 2 1 ets.amse MASE 0.6205520  
## 3 1 ets.sigma MASE 0.6136585  
## 4 1 ets.mae MASE 0.6180026  
## 5 1 ets.lik MASE 0.6136597  
## 6 1 ets.mse AIC 5470.5670859  
## 7 1 ets.amse AIC 5479.9307190  
## 8 1 ets.sigma AIC 5470.5670859  
## 9 1 ets.mae AIC 5473.0243248  
## 10 1 ets.lik AIC 5470.5671075  
## 11 1 ets.mse BIC 5519.2724469  
## 12 1 ets.amse BIC 5538.3771523  
## 13 1 ets.sigma BIC 5519.2724469  
## 14 1 ets.mae BIC 5521.7296858  
## 15 1 ets.lik BIC 5519.2724686  
## 16 2 ets.mse MASE 0.5149088  
## 17 2 ets.amse MASE 0.5168390  
## 18 2 ets.sigma MASE 0.5157364  
## 19 2 ets.mae MASE 0.4986766  
## 20 2 ets.lik MASE 0.5148549  
## 21 2 ets.mse AIC 5274.3315642  
## 22 2 ets.amse AIC 5276.8093195  
## 23 2 ets.sigma AIC 5270.8505797  
## 24 2 ets.mae AIC 5278.0344099  
## 25 2 ets.lik AIC 5270.5860399  
## 26 2 ets.mse BIC 5323.0369253  
## 27 2 ets.amse BIC 5332.0087288  
## 28 2 ets.sigma BIC 5326.0499889  
## 29 2 ets.mae BIC 5333.2338191  
## 30 2 ets.lik BIC 5325.7854491  
## 31 3 ets.mse MASE 0.4249198  
## 32 3 ets.amse MASE 0.4243551  
## 33 3 ets.sigma MASE 0.4136072  
## 34 3 ets.mae MASE 0.4116296  
## 35 3 ets.lik MASE 0.4236739  
## 36 3 ets.mse AIC 5480.9681261  
## 37 3 ets.amse AIC 5481.5396306  
## 38 3 ets.sigma AIC 5474.0156379  
## 39 3 ets.mae AIC 5481.1711236  
## 40 3 ets.lik AIC 5477.2805449  
## 41 3 ets.mse BIC 5539.4145594  
## 42 3 ets.amse BIC 5539.9860639  
## 43 3 ets.sigma BIC 5522.7209990  
## 44 3 ets.mae BIC 5539.6175569  
## 45 3 ets.lik BIC 5535.7269782

## minimum MASE,AIC,BIC value--

The below are lowest MASE,AIC,and BIC value for all three series using five different optimization criteria such as mse,amse,sigma,mae and lik

expFINAL %>%   
 group\_by(Data.No,Model.check) %>%   
 slice(which.min(Value))

## # Groups: Data.No, Model.check [9]  
## Data.No Model.type Model.check Value  
## <int> <fctr> <fctr> <dbl>  
## 1 1 ets.mse MASE 0.6136585  
## 2 1 ets.mse AIC 5470.5670859  
## 3 1 ets.mse BIC 5519.2724469  
## 4 2 ets.mae MASE 0.4986766  
## 5 2 ets.lik AIC 5270.5860399  
## 6 2 ets.mse BIC 5323.0369253  
## 7 3 ets.mae MASE 0.4116296  
## 8 3 ets.sigma AIC 5474.0156379  
## 9 3 ets.sigma BIC 5522.7209990

From the output,we have found lowest MASE,AIC and BIC for our SERIES-1(MEXICO) based on optimization criteria mse,for SERIES-2(CANADA) we have found lowest MASE,AIC and BIC based on three different optimization criteria such as mse,mae and lik. For our last series which is the visitors from rest of the world we have found two different optimizatin critetia for lowest AIC,BIC and MASE values.

The table for lowest MASE,AIC and BIC Values

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MASE | AIC | BIC |
| Model\_Mexico1 | 0.4471108 | 4959.447 | 5014.736 |
| Model\_Mexico2 | 0.4420263 | 4931.233 | 4989.679 |
| Model\_Mexico3 | 0.4384837 | 4905.562 | 4967.156 |
| Model\_Mexico4 | 0.4323638 | 4880.137 | 4944.866 |
| Model\_others1 | 0.4263643 | 5073.584 | 5128.873 |
| Model\_others2 | 0.4769229 | 5109.334 | 5164.533 |
| Model\_others3 | 0.4384837 | 4905.562 | 4967.156 |
| Model\_others4 | 0.4179781 | 4994.787 | 5059.516 |

## dynlm for MEXICO--

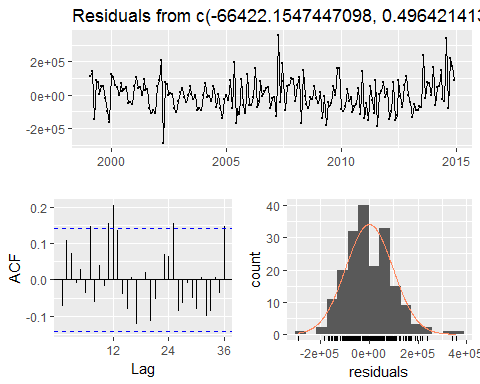
Since there are intervention in MEXICO SERIES, and other series,we have fit dynamic Linear modelfor both these series.

Y.t=Mexico[1:192]  
Y.t = ts(Y.t,start=c(1999,01),frequency = 12)  
T = 73 # The time point when the intervention occurred   
s.t = 1\*(seq(Y.t) >= T)  
s.t.1 = Lag(s.t,+1)   
  
model\_Mexico1 = dynlm(Y.t ~ L(Y.t , k = 1 ) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
summary(model\_Mexico1)

##   
## Time series regression with "ts" data:  
## Start = 1999(2), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t ~ L(Y.t, k = 1) + s.t.1 + s.t + trend(Y.t) +   
## season(Y.t))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -289909 -70807 -5773 55002 358540   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.642e+04 4.583e+04 -1.449 0.149037   
## L(Y.t, k = 1) 4.964e-01 6.704e-02 7.405 5.31e-12 \*\*\*  
## s.t.1 -4.474e+05 1.181e+05 -3.787 0.000209 \*\*\*  
## s.t 7.924e+05 1.052e+05 7.534 2.52e-12 \*\*\*  
## trend(Y.t) 1.058e+04 3.208e+03 3.297 0.001184 \*\*   
## season(Y.t)Feb 5.300e+04 4.266e+04 1.242 0.215777   
## season(Y.t)Mar 1.542e+05 4.678e+04 3.296 0.001189 \*\*   
## season(Y.t)Apr 3.381e+05 4.490e+04 7.531 2.57e-12 \*\*\*  
## season(Y.t)May 9.734e+04 3.866e+04 2.518 0.012697 \*   
## season(Y.t)Jun 1.345e+05 4.239e+04 3.172 0.001786 \*\*   
## season(Y.t)Jul 4.555e+05 4.347e+04 10.479 < 2e-16 \*\*\*  
## season(Y.t)Aug 1.079e+05 3.690e+04 2.924 0.003911 \*\*   
## season(Y.t)Sep 9.490e+04 4.004e+04 2.370 0.018874 \*   
## season(Y.t)Oct 2.171e+05 4.349e+04 4.993 1.43e-06 \*\*\*  
## season(Y.t)Nov 2.648e+05 4.126e+04 6.419 1.25e-09 \*\*\*  
## season(Y.t)Dec 4.303e+05 3.911e+04 11.004 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 100600 on 175 degrees of freedom  
## Multiple R-squared: 0.9537, Adjusted R-squared: 0.9498   
## F-statistic: 240.4 on 15 and 175 DF, p-value: < 2.2e-16

checkresiduals(model\_Mexico1)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
##   
2014.66666666667, 2014.75, 2014.83333333333, 2014.91666666667)Residuals from 12Residuals from FALSE  
## LM test = 46.639, df = 24, p-value = 0.003706

accuracy(model\_Mexico1) #MASE 0.4471108

## ME RMSE MAE MPE MAPE MASE  
## Training set 1.147545e-12 96269.03 75294.12 -0.8598329 12.89771 0.4471108  
## ACF1  
## Training set -0.07293843

AIC(model\_Mexico1) # 4959.447

## [1] 4959.447

BIC(model\_Mexico1)#5014.736

## [1] 5014.736

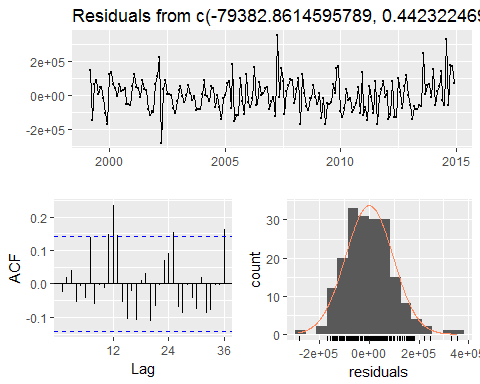
Our model and most of the coefficients are significant at 5% level of significance. Although most of the residual diagnostics are suitable, there is still some serial correlation left in the residuals. To overcome this issue, we add Yt-2 to the model as another predictor variable.

model\_Mexico2 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
summary(model\_Mexico2)

##   
## Time series regression with "ts" data:  
## Start = 1999(3), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t ~ L(Y.t, k = 1) + L(Y.t, k = 2) + s.t.1 +   
## s.t + trend(Y.t) + season(Y.t))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -283205 -67945 -3390 56344 353823   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.938e+04 4.583e+04 -1.732 0.08501 .   
## L(Y.t, k = 1) 4.423e-01 7.245e-02 6.105 6.56e-09 \*\*\*  
## L(Y.t, k = 2) 1.259e-01 6.613e-02 1.903 0.05866 .   
## s.t.1 -5.124e+05 1.218e+05 -4.206 4.16e-05 \*\*\*  
## s.t 8.110e+05 1.047e+05 7.745 7.71e-13 \*\*\*  
## trend(Y.t) 9.170e+03 3.304e+03 2.775 0.00612 \*\*   
## season(Y.t)Feb 8.920e+03 4.692e+04 0.190 0.84946   
## season(Y.t)Mar 1.455e+05 4.663e+04 3.121 0.00211 \*\*   
## season(Y.t)Apr 3.458e+05 4.468e+04 7.739 8.01e-13 \*\*\*  
## season(Y.t)May 1.103e+05 3.891e+04 2.834 0.00514 \*\*   
## season(Y.t)Jun 1.139e+05 4.345e+04 2.621 0.00955 \*\*   
## season(Y.t)Jul 4.505e+05 4.319e+04 10.432 < 2e-16 \*\*\*  
## season(Y.t)Aug 1.234e+05 3.748e+04 3.293 0.00120 \*\*   
## season(Y.t)Sep 6.135e+04 4.348e+04 1.411 0.16005   
## season(Y.t)Oct 2.023e+05 4.384e+04 4.614 7.68e-06 \*\*\*  
## season(Y.t)Nov 2.675e+05 4.093e+04 6.535 6.85e-10 \*\*\*  
## season(Y.t)Dec 4.289e+05 3.878e+04 11.058 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 99720 on 173 degrees of freedom  
## Multiple R-squared: 0.9545, Adjusted R-squared: 0.9503   
## F-statistic: 226.8 on 16 and 173 DF, p-value: < 2.2e-16

checkresiduals(model\_Mexico2)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
##   
## LM test = 44.234, df = 24, p-value = 0.007166

accuracy(model\_Mexico2) #MASE 0.4420263

## ME RMSE MAE MPE MAPE MASE  
## Training set 7.752395e-14 95154.13 74520.64 -0.9072927 12.85146 0.4420263  
## ACF1  
## Training set -0.02564606

AIC(model\_Mexico2) #AIC= 4931.233

## [1] 4931.233

BIC(model\_Mexico2) # BIC=4989.679

## [1] 4989.679

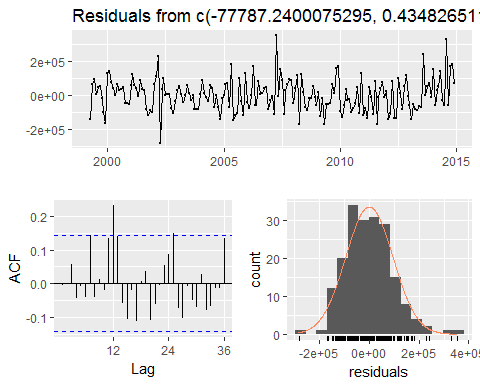
In the new model there are still some serial correlation left in the resduals,but in this model there is an improvement in AIC value,BIC Value and MASE Value as well.

model\_Mexico3 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 ) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
summary(model\_Mexico3)

##   
## Time series regression with "ts" data:  
## Start = 1999(4), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t ~ L(Y.t, k = 1) + L(Y.t, k = 2) + L(Y.t,   
## k = 3) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -282938 -69959 -2365 55799 352597   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.779e+04 4.577e+04 -1.699 0.09106 .   
## L(Y.t, k = 1) 4.348e-01 7.265e-02 5.986 1.23e-08 \*\*\*  
## L(Y.t, k = 2) 1.455e-01 7.432e-02 1.958 0.05185 .   
## L(Y.t, k = 3) -2.882e-02 6.365e-02 -0.453 0.65123   
## s.t.1 -5.029e+05 1.225e+05 -4.104 6.26e-05 \*\*\*  
## s.t 8.114e+05 1.045e+05 7.763 7.24e-13 \*\*\*  
## trend(Y.t) 9.936e+03 3.389e+03 2.932 0.00383 \*\*   
## season(Y.t)Feb 4.383e+03 4.694e+04 0.093 0.92572   
## season(Y.t)Mar 1.415e+05 5.167e+04 2.738 0.00684 \*\*   
## season(Y.t)Apr 3.466e+05 4.520e+04 7.668 1.26e-12 \*\*\*  
## season(Y.t)May 1.085e+05 3.890e+04 2.790 0.00586 \*\*   
## season(Y.t)Jun 1.083e+05 4.378e+04 2.475 0.01432 \*   
## season(Y.t)Jul 4.534e+05 4.485e+04 10.110 < 2e-16 \*\*\*  
## season(Y.t)Aug 1.252e+05 3.756e+04 3.333 0.00105 \*\*   
## season(Y.t)Sep 5.472e+04 4.422e+04 1.237 0.21763   
## season(Y.t)Oct 2.074e+05 4.744e+04 4.372 2.13e-05 \*\*\*  
## season(Y.t)Nov 2.696e+05 4.190e+04 6.435 1.20e-09 \*\*\*  
## season(Y.t)Dec 4.271e+05 3.873e+04 11.029 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 99530 on 171 degrees of freedom  
## Multiple R-squared: 0.9549, Adjusted R-squared: 0.9504   
## F-statistic: 212.7 on 17 and 171 DF, p-value: < 2.2e-16

checkresiduals(model\_Mexico3)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
##   
  
## LM test = 42.672, df = 24, p-value = 0.01084

accuracy(model\_Mexico3) #MASE 0.4384837

## ME RMSE MAE MPE MAPE MASE  
## Training set 5.007033e-12 94675.51 74238.51 -0.8597698 12.74043 0.4384837  
## ACF1  
## Training set -0.005144415

AIC(model\_Mexico3) #4905.562

## [1] 4905.562

BIC(model\_Mexico3) #4967.156

## [1] 4967.156

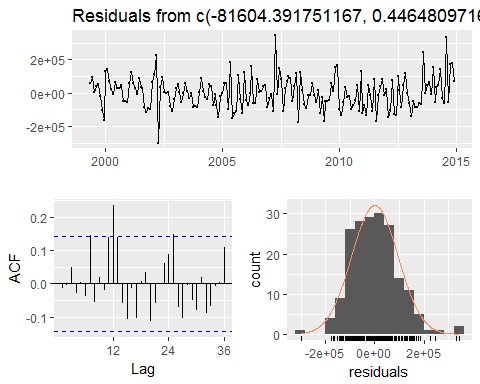
In the new model there are still some serial correlation left in the resduals,but in this model there is an improvement in AIC value,BIC Value and MASE value.

model\_Mexico4 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 )+ L(Y.t , k = 4 ) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
summary(model\_Mexico4)

##   
## Time series regression with "ts" data:  
## Start = 1999(5), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t ~ L(Y.t, k = 1) + L(Y.t, k = 2) + L(Y.t,   
## k = 3) + L(Y.t, k = 4) + s.t.1 + s.t + trend(Y.t) + season(Y.t))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -299969 -66545 -2876 56290 345556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -8.160e+04 4.583e+04 -1.781 0.07677 .   
## L(Y.t, k = 1) 4.465e-01 7.321e-02 6.099 7.06e-09 \*\*\*  
## L(Y.t, k = 2) 1.522e-01 7.477e-02 2.035 0.04339 \*   
## L(Y.t, k = 3) -2.284e-02 7.343e-02 -0.311 0.75618   
## L(Y.t, k = 4) -1.935e-02 6.338e-02 -0.305 0.76054   
## s.t.1 -5.086e+05 1.230e+05 -4.134 5.61e-05 \*\*\*  
## s.t 8.130e+05 1.044e+05 7.787 6.58e-13 \*\*\*  
## trend(Y.t) 9.590e+03 3.481e+03 2.755 0.00651 \*\*   
## season(Y.t)Feb 7.117e+03 4.705e+04 0.151 0.87996   
## season(Y.t)Mar 1.476e+05 5.200e+04 2.839 0.00508 \*\*   
## season(Y.t)Apr 3.683e+05 5.048e+04 7.297 1.10e-11 \*\*\*  
## season(Y.t)May 1.131e+05 3.964e+04 2.853 0.00487 \*\*   
## season(Y.t)Jun 1.108e+05 4.392e+04 2.522 0.01259 \*   
## season(Y.t)Jul 4.568e+05 4.547e+04 10.046 < 2e-16 \*\*\*  
## season(Y.t)Aug 1.301e+05 3.968e+04 3.278 0.00127 \*\*   
## season(Y.t)Sep 5.740e+04 4.422e+04 1.298 0.19611   
## season(Y.t)Oct 2.103e+05 4.864e+04 4.323 2.62e-05 \*\*\*  
## season(Y.t)Nov 2.795e+05 4.547e+04 6.148 5.47e-09 \*\*\*  
## season(Y.t)Dec 4.325e+05 3.973e+04 10.886 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 99410 on 169 degrees of freedom  
## Multiple R-squared: 0.9551, Adjusted R-squared: 0.9503   
## F-statistic: 199.7 on 18 and 169 DF, p-value: < 2.2e-16

checkresiduals(model\_Mexico4)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
  
## LM test = 42.275, df = 24, p-value = 0.01202

accuracy(model\_Mexico4) #0.4323638

## ME RMSE MAE MPE MAPE MASE  
## Training set -4.569281e-12 94255.78 73534.34 -0.8319326 12.53444 0.4323638  
## ACF1  
## Training set -0.01254671

AIC(model\_Mexico4) #AIC 4880.137

## [1] 4880.137

BIC(model\_Mexico4) #BIC 4944.866

## [1] 4944.866

In the new model there is a very good improvement in seril coorelation in the residuals,only at one point there is a serial correlation left in the resduals,AIC value ,BIC value and MASE Value.

The residuals are randomly distributed and in the time plot the residuals are bouncing over mean value.

# Considering all these criteria model\_Mexico4 is the best model for dynamic linear method.

## dynlm for others

Y.t.others=others[1:192]  
which(Y.t.others < 1200000)

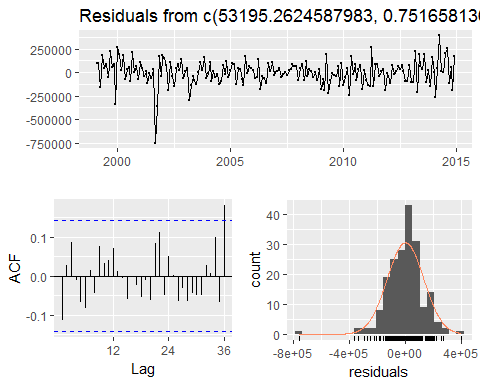
## [1] 35 37

Y.t.others = ts(Y.t.others,start=c(1999,01),frequency = 12)  
T = 35 # The time point when the intervention occurred   
s.t.others = 1\*(seq(Y.t.others) >= T)  
s.t.others.1 = Lag(s.t,+1)   
  
model\_others1 = dynlm(Y.t.others ~ L(Y.t.others , k = 1 ) + s.t.others.1 + s.t.others + trend(Y.t.others) + season(Y.t.others))  
summary(model\_others1)

##   
## Time series regression with "ts" data:  
## Start = 1999(2), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t.others ~ L(Y.t.others, k = 1) + s.t.others.1 +   
## s.t.others + trend(Y.t.others) + season(Y.t.others))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -752296 -73476 12714 77974 396114   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.320e+04 1.253e+05 0.425 0.671600   
## L(Y.t.others, k = 1) 7.517e-01 6.156e-02 12.210 < 2e-16 \*\*\*  
## s.t.others.1 -3.041e+04 3.801e+04 -0.800 0.424726   
## s.t.others -1.641e+05 6.221e+04 -2.638 0.009088 \*\*   
## trend(Y.t.others) 2.792e+04 8.141e+03 3.430 0.000754 \*\*\*  
## season(Y.t.others)Feb 2.458e+05 5.434e+04 4.523 1.12e-05 \*\*\*  
## season(Y.t.others)Mar 5.991e+05 5.602e+04 10.695 < 2e-16 \*\*\*  
## season(Y.t.others)Apr 4.578e+05 4.953e+04 9.243 < 2e-16 \*\*\*  
## season(Y.t.others)May 4.411e+05 4.887e+04 9.025 3.17e-16 \*\*\*  
## season(Y.t.others)Jun 4.581e+05 4.878e+04 9.391 < 2e-16 \*\*\*  
## season(Y.t.others)Jul 8.189e+05 4.892e+04 16.738 < 2e-16 \*\*\*  
## season(Y.t.others)Aug 4.730e+05 5.650e+04 8.372 1.76e-14 \*\*\*  
## season(Y.t.others)Sep 1.865e+05 5.520e+04 3.379 0.000898 \*\*\*  
## season(Y.t.others)Oct 3.655e+05 4.918e+04 7.432 4.56e-12 \*\*\*  
## season(Y.t.others)Nov 6.678e+04 4.907e+04 1.361 0.175322   
## season(Y.t.others)Dec 6.063e+05 5.135e+04 11.807 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 135600 on 175 degrees of freedom  
## Multiple R-squared: 0.9305, Adjusted R-squared: 0.9246   
## F-statistic: 156.2 on 15 and 175 DF, p-value: < 2.2e-16

checkresiduals(model\_others1)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
## LM test = 25.546, df = 24, p-value = 0.3766

accuracy(model\_others1)# MASE 0.4263643

## ME RMSE MAE MPE MAPE MASE  
## Training set -6.701571e-12 129792.2 95308.51 -0.3466987 4.998633 0.4263643  
## ACF1  
## Training set -0.1146601

AIC(model\_others1) # AIC 5073.584

## [1] 5073.584

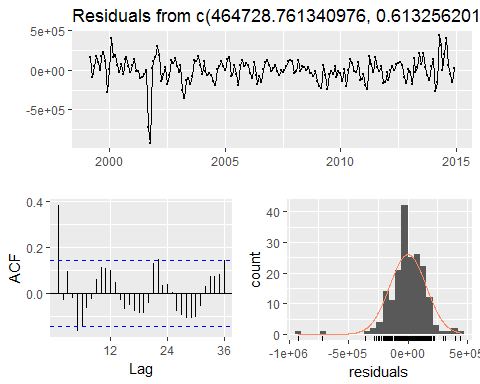
BIC(model\_others1) # BIC 5128.873

## [1] 5128.873

Our model and most of the coefficients are significant at 5% level of significance. Although most of the residual diagnostics are suitable, there is still some serial correlation left in the residuals. To overcome this issue, we add Yt???2 to the model as another predictor variable.

model\_others2 = dynlm(Y.t.others ~ L(Y.t.others , k = 2 ) + s.t.others.1 + s.t.others + trend(Y.t.others) + season(Y.t.others))  
checkresiduals(model\_others2)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



##   
## Breusch-Godfrey test for serial correlation of order up to 24  
## LM test = 66.755, df = 24, p-value = 6.706e-06

accuracy(model\_others2)# MASE 0.4769229

## ME RMSE MAE MPE MAPE MASE  
## Training set -7.969784e-12 152848.5 106277.6 -0.5242155 5.6763 0.4769229  
## ACF1  
## Training set 0.3816218

AIC(model\_others2) #AIC 5109.334

## [1] 5109.334

BIC(model\_others2) #BIC 5164.533

## [1] 5164.533

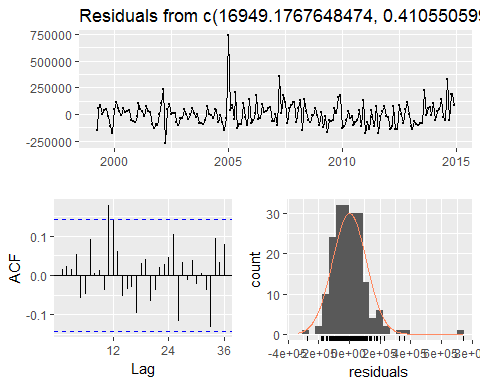
In the new model there are still some serial correlation left in the resduals,but in this model AIC value,BIC Value has increased which is not good,however MASE value has improved

model\_others3 = dynlm(Y.t ~ L(Y.t , k = 1 ) + L(Y.t , k = 2 ) + L(Y.t , k = 3 ) + s.t.others.1 + s.t.others + trend(Y.t.others) + season(Y.t.others))  
summary(model\_others3)

##   
## Time series regression with "ts" data:  
## Start = 1999(4), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t ~ L(Y.t, k = 1) + L(Y.t, k = 2) + L(Y.t,   
## k = 3) + s.t.others.1 + s.t.others + trend(Y.t.others) +   
## season(Y.t.others))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -271054 -73891 -8066 53643 740019   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.695e+04 5.763e+04 0.294 0.76904   
## L(Y.t, k = 1) 4.106e-01 8.533e-02 4.811 3.28e-06 \*\*\*  
## L(Y.t, k = 2) 9.178e-02 8.638e-02 1.063 0.28947   
## L(Y.t, k = 3) -3.350e-02 7.426e-02 -0.451 0.65243   
## s.t.others.1 3.409e+05 7.399e+04 4.607 7.96e-06 \*\*\*  
## s.t.others -2.414e+04 3.123e+04 -0.773 0.44069   
## trend(Y.t.others) 1.457e+04 4.359e+03 3.344 0.00102 \*\*   
## season(Y.t.others)Feb -4.694e+04 5.407e+04 -0.868 0.38651   
## season(Y.t.others)Mar 7.366e+04 5.944e+04 1.239 0.21692   
## season(Y.t.others)Apr 2.719e+05 5.213e+04 5.215 5.26e-07 \*\*\*  
## season(Y.t.others)May 4.061e+04 4.458e+04 0.911 0.36361   
## season(Y.t.others)Jun 4.818e+04 5.050e+04 0.954 0.34142   
## season(Y.t.others)Jul 3.858e+05 5.157e+04 7.481 3.69e-12 \*\*\*  
## season(Y.t.others)Aug 6.242e+04 4.281e+04 1.458 0.14667   
## season(Y.t.others)Sep 3.225e+03 5.097e+04 0.063 0.94962   
## season(Y.t.others)Oct 1.439e+05 5.456e+04 2.638 0.00911 \*\*   
## season(Y.t.others)Nov 2.022e+05 4.790e+04 4.221 3.94e-05 \*\*\*  
## season(Y.t.others)Dec 3.645e+05 4.422e+04 8.243 4.25e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 115500 on 171 degrees of freedom  
## Multiple R-squared: 0.9392, Adjusted R-squared: 0.9331   
## F-statistic: 155.3 on 17 and 171 DF, p-value: < 2.2e-16

checkresiduals(model\_others3)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



## LM test = 22.484, df = 24, p-value = 0.5504

accuracy(model\_Mexico3) #MASE 0.4384837

## ME RMSE MAE MPE MAPE MASE  
## Training set 5.007033e-12 94675.51 74238.51 -0.8597698 12.74043 0.4384837  
## ACF1  
## Training set -0.005144415

AIC(model\_Mexico3) # AIC 4905.562

## [1] 4905.562

BIC(model\_Mexico3) # BIC 4967.156

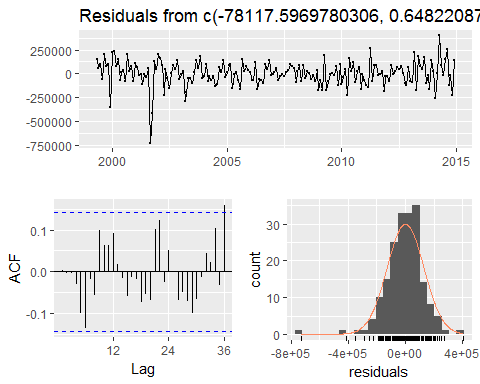
## [1] 4967.156

model\_others4 = dynlm(Y.t.others ~ L(Y.t.others , k = 1 ) + L(Y.t.others , k = 2 ) + L(Y.t.others , k = 3 )+ L(Y.t.others , k = 4 ) + s.t.others.1 + s.t.others + trend(Y.t.others) + season(Y.t.others))  
summary(model\_others4)

##   
## Time series regression with "ts" data:  
## Start = 1999(5), End = 2014(12)  
##   
## Call:  
## dynlm(formula = Y.t.others ~ L(Y.t.others, k = 1) + L(Y.t.others,   
## k = 2) + L(Y.t.others, k = 3) + L(Y.t.others, k = 4) + s.t.others.1 +   
## s.t.others + trend(Y.t.others) + season(Y.t.others))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -724494 -72392 7703 78915 401595   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.812e+04 1.647e+05 -0.474 0.63595   
## L(Y.t.others, k = 1) 6.482e-01 7.801e-02 8.309 3.02e-14 \*\*\*  
## L(Y.t.others, k = 2) 1.455e-01 9.511e-02 1.530 0.12790   
## L(Y.t.others, k = 3) 9.758e-02 9.298e-02 1.050 0.29543   
## L(Y.t.others, k = 4) -5.368e-02 8.039e-02 -0.668 0.50519   
## s.t.others.1 -2.315e+04 3.820e+04 -0.606 0.54534   
## s.t.others -8.379e+04 7.827e+04 -1.071 0.28591   
## trend(Y.t.others) 1.816e+04 1.010e+04 1.798 0.07400 .   
## season(Y.t.others)Feb 1.896e+05 7.726e+04 2.454 0.01513 \*   
## season(Y.t.others)Mar 5.507e+05 6.358e+04 8.661 3.61e-15 \*\*\*  
## season(Y.t.others)Apr 5.198e+05 5.960e+04 8.721 2.51e-15 \*\*\*  
## season(Y.t.others)May 4.417e+05 7.131e+04 6.194 4.32e-09 \*\*\*  
## season(Y.t.others)Jun 4.175e+05 6.952e+04 6.006 1.13e-08 \*\*\*  
## season(Y.t.others)Jul 7.843e+05 5.962e+04 13.155 < 2e-16 \*\*\*  
## season(Y.t.others)Aug 4.717e+05 6.314e+04 7.471 4.07e-12 \*\*\*  
## season(Y.t.others)Sep 1.195e+05 7.939e+04 1.506 0.13400   
## season(Y.t.others)Oct 2.355e+05 7.729e+04 3.048 0.00268 \*\*   
## season(Y.t.others)Nov -2.246e+03 6.014e+04 -0.037 0.97026   
## season(Y.t.others)Dec 5.382e+05 7.022e+04 7.664 1.34e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 134900 on 169 degrees of freedom  
## Multiple R-squared: 0.9332, Adjusted R-squared: 0.9261   
## F-statistic: 131.2 on 18 and 169 DF, p-value: < 2.2e-16

checkresiduals(model\_others4)

## Warning in if (method == "Missing") main <- "Residuals" else main <-  
## paste("Residuals from", : the condition has length > 1 and only the first  
## element will be used



## LM test = 24.727, df = 24, p-value = 0.4207

accuracy(model\_others4) #MASE 0.4179781

## ME RMSE MAE MPE MAPE MASE  
## Training set -2.352834e-11 127859.3 93352.58 -0.3514305 4.914019 0.4179781  
## ACF1  
## Training set 0.002840409

AIC(model\_others4) # AIC 4994.787

## [1] 4994.787

BIC(model\_others4) #BIC 5059.516

## [1] 5059.516

In the new model there is a very good improvement in serial coorelation in the residuals,only at one point there is a serial correlation left in the resduals,AIC value ,BIC value didnot imorove much compare to first model but MASE Value is the lowest

The residuals are randomly distributed and in the time plot the residuals are bouncing over mean value.

# Considering all these criteria model\_others4 is the best model for dynamic linear method.

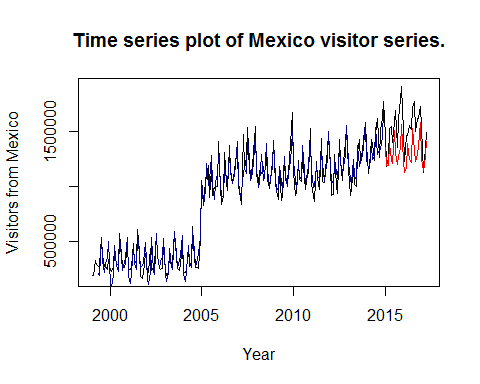
### Prediction validation

For the predication Validation procedure we have considered 190 observation out of 220 observation as a training data set and the rest as a test data set to check which models perform better.

# From the Exponential and state space models and the intervention analysis the best model that was found using AIC,BIC and MASE were ets.mse and model\_Mexico4  
mexico\_mse = forecast(ets(Mexico\_reduced,opt.crit = "mse"),h=30)  
mexico\_pv\_mse= accuracy(mexico\_mse,Mexico[191:220])  
mexico\_pv\_mase\_mse =mexico\_pv\_mse[2,6]# MASE is 0.6719396  
mexico\_pv\_mape\_mse =mexico\_pv\_mse[2,5]# MAPE is 7.527646 we will compare this value with the dynlm forecast test set accuracy MAPE number

### mexico dynlm forecast

q = 28  
n = nrow(model\_Mexico4$model)  
mexico.frc = array(NA , (n + q))  
mexico.frc[1:n] = Y.t[5:length(Y.t)]  
  
trend = array(NA,q)  
trend.start = model\_Mexico4$model[n,"trend(Y.t)"]  
trend = seq(trend.start , trend.start + q/12, 1/12)  
  
for (i in 1:q){  
 months = array(0,11)  
 months[(i-1)%%12] = 1  
 data.new = c(1,mexico.frc[n-1+i],mexico.frc[n-2+i],mexico.frc[n-3+i],mexico.frc[n-4+i], s.t.1[n] , s.t[n] ,trend[i],months)   
 mexico.frc[n+i] = as.vector(model\_Mexico4$coefficients) %\*% data.new  
}  
  
par(mfrow=c(1,1))  
  
plot(Mexico,ylab='Visitors from Mexico',xlab='Year',main = "Time series plot of Mexico visitor series.")  
lines(model\_Mexico4$fitted.values,col = "darkblue")  
lines(ts(mexico.frc[(n+1):(n+q)],start=c(2015,1),frequency = 12),col="red")



fcast = ts(mexico.frc[(n+1):(n+q)],start=c(2015,1),frequency = 12)  
actual = ts(Mexico[193:210],start=c(2015,1),frequency = 12)  
mexico\_dynlm\_MAPE = accuracy(fcast,actual)[1,5]  
mexico\_dynlm\_MAPE#13.87927

## [1] 13.87927

# prediciton valuation for canada i.e. Data series 2

# the best models that were found using AIC,BIC and MASE were ets.mse, ets.mae, ets.sigma  
canada\_mse = forecast(ets(Canada\_reduced,opt.crit = "mse"),h=30)  
canada\_mae = forecast(ets(Canada\_reduced,opt.crit = "mae"),h=30)  
canada\_lik = forecast(ets(Canada\_reduced,opt.crit = "lik"),h=30)  
  
canada\_pv\_mase\_mse= accuracy(canada\_mse,Canada[191:220])[2,6]  
canada\_pv\_mase\_mae= accuracy(canada\_mae,Canada[191:220])[2,6]  
canada\_pv\_mase\_lik= accuracy(canada\_lik,Canada[191:220])[2,6]  
  
canada\_pv\_mase\_mse#.6970713 This is the best for Canada

## [1] 0.6970713

canada\_pv\_mase\_mae# 0.8997252

## [1] 0.8997252

canada\_pv\_mase\_lik#0.9206229

## [1] 0.9206229

# the best models that were found using AIC,BIC and MASE were ets.mae, ets.sigma and model\_others4  
  
others\_mae = forecast(ets(others\_reduced,opt.crit = "mae"),h=30)  
others\_sigma = forecast(ets(others\_reduced,opt.crit = "sigma"),h=30)  
  
others\_pv\_mase\_mae= accuracy(others\_mae,others[191:220])[2,6]  
others\_pv\_mape\_mae= accuracy(others\_mae,others[191:220])[2,5]  
others\_pv\_mase\_sigma= accuracy(others\_sigma,others[191:220])[2,6]  
others\_pv\_mape\_sigma= accuracy(others\_sigma,others[191:220])[2,5]  
  
others\_pv\_mase\_mae# MASE 0.8995914

## [1] 0.8995914

others\_pv\_mase\_sigma# MASE 0.8914436

## [1] 0.8914436

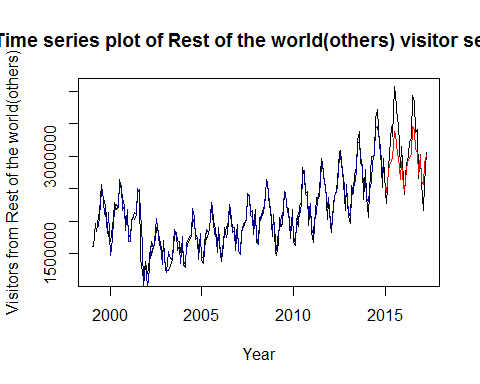
others\_pv\_mape\_mae# MAPE 6.137864

## [1] 6.137864

others\_pv\_mape\_sigma# MAPE 6.054912

## [1] 6.054912

# others->dynlm   
  
q = 28  
n = nrow(model\_others4$model)  
others.frc = array(NA , (n + q))  
others.frc[1:n] = Y.t.others[5:length(Y.t.others)]  
  
trend = array(NA,q)  
trend.start = model\_others4$model[n,"trend(Y.t.others)"]  
trend = seq(trend.start , trend.start + q/12, 1/12)  
  
for (i in 1:q){  
 months = array(0,11)  
 months[(i-1)%%12] = 1  
 data.new = c(1,others.frc[n-1+i],others.frc[n-2+i],others.frc[n-3+i],others.frc[n-4+i], s.t.others.1[n] , s.t.others[n] ,trend[i],months)   
 others.frc[n+i] = as.vector(model\_others4$coefficients) %\*% data.new  
}  
  
par(mfrow=c(1,1))  
  
plot(others,ylab='Visitors from Rest of the world(others)',xlab='Year',main = "Time series plot of Rest of the world(others) visitor series.")  
lines(model\_others4$fitted.values,col = "darkblue")  
lines(ts(others.frc[(n+1):(n+q)],start=c(2015,1),frequency = 12),col="red")



fcast.others = ts(others.frc[(n+1):(n+q)],start=c(2015,1),frequency = 12)  
actual = ts(others[193:210],start=c(2015,1),frequency = 12)  
  
others\_dynlm\_MAPE = accuracy(fcast.others,actual)[1,5]  
others\_dynlm\_MAPE#8.104702

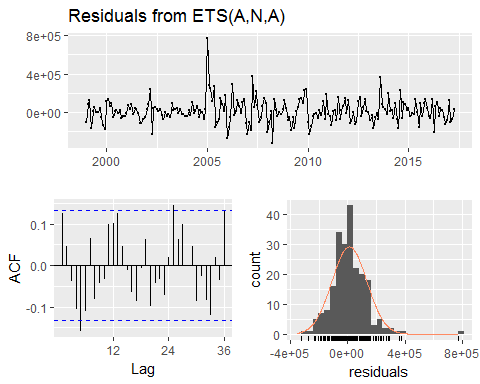
## [1] 8.104702

So, now we get 1 top model for each of the series after the prediction validation and they are 1)for Mexico based on MAPE ets.mse - ANA 2)for Canada based on MASE was ets.mse - MNM and 3)for Others based on MASE and MAPE was ets.sigma -MAdM

The Table below shows the MASE and MAPE values for models of three time series

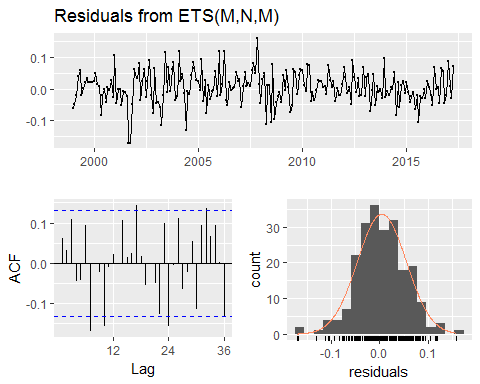
|  |  |
| --- | --- |
| Canada | |
| Model | MASE |
| ets.mse | 0.6970713 |
| ets.mae | 0.8997252 |
| ets.lik | 0.9206229 |
|  |  |
| Mexico | |
| Model | MAPE |
| ets.mse | 7.527646 |
| dynlm | 13.87927 |
|  |  |
| Others | |
| Model | MAPE |
| ets.mae | 6.137864 |
| ets.sigma | 6.054912 |
| dynlm | 8.104702 |

#So, we will run a residual check on these models  
  
# Check residuals  
Mexico\_final\_model =ets(Mexico,opt.crit = "mse")#ANA  
checkresiduals(Mexico\_final\_model)



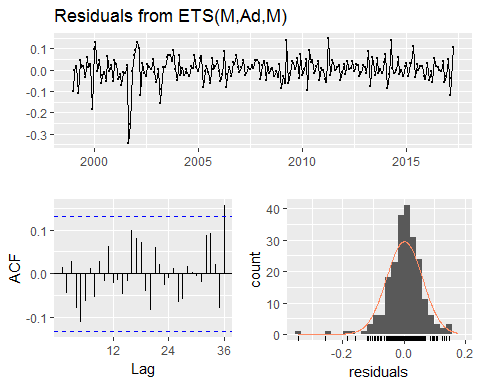
##   
## Ljung-Box test  
##   
## data: Residuals from ETS(A,N,A)  
## Q\* = 35.595, df = 10, p-value = 9.88e-05  
##   
## Model df: 14. Total lags used: 24

Canada\_final\_model = ets(Canada,opt.crit = "mse")#MNM  
checkresiduals(Canada\_final\_model)



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(M,N,M)  
## Q\* = 41.058, df = 10, p-value = 1.102e-05  
##   
## Model df: 14. Total lags used: 24

others\_final\_model=ets(others,opt.crit = "sigma")  
checkresiduals(others\_final\_model)#MAdM



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(M,Ad,M)  
## Q\* = 16.881, df = 7, p-value = 0.01818  
##   
## Model df: 17. Total lags used: 24

## Residuals check for MEXICO--

After we run the diagonistic check we can see that there is not much serial auto correlation left in the residuals for ANA model,in time series plot the residuals are bouncing over mean and histograms for residuals fairly bell shaped

## Residuals check for CANADA--

After we run the diagonistic check The results are very similar to the those obtained with ANA model.

## Residuals check for others--

After we run the diagonistic check we can see that there is not much serial auto correlation left in the residuals for MAdM model,in time series plot the residuals are bouncing over mean and histograms for residuals are bell shaped

# Forecasts and plots frc.Mexico.ANA = forecast(Mexico\_final\_model, h =10) frc.Canada.MNM = forecast(Canada\_final\_model, h =10) frc.Others.MAdM = forecast(others\_final\_model, h =10) # Mexico Plot plot(frc.Mexico.ANA, ylab="Mexico visitors",plot.conf=FALSE, type="l", fcol="red", xlab="Year", main="Fits and forecasts for Mexico series")

## Warning in plot.window(xlim, ylim, log, ...): "plot.conf" is not a  
## graphical parameter

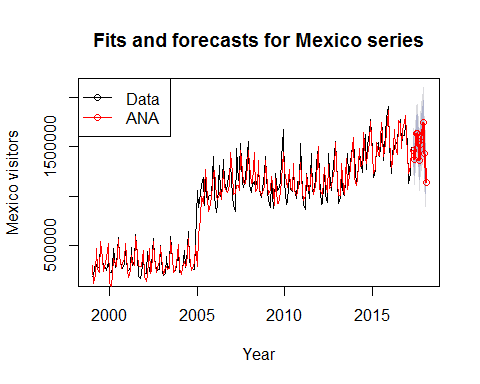
## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "plot.conf"  
## is not a graphical parameter

## Warning in axis(1, ...): "plot.conf" is not a graphical parameter

## Warning in axis(2, ...): "plot.conf" is not a graphical parameter

## Warning in box(...): "plot.conf" is not a graphical parameter

lines(fitted(frc.Mexico.ANA), col="red", lty=1)  
lines(frc.Mexico.ANA$mean,col="red", type="o")  
legend("topleft",lty=1, pch=1, col=1:3, c("Data","ANA "))



# Canada Plot  
plot(frc.Canada.MNM, ylab="Canada visitors",plot.conf=FALSE, type="l", fcol="red", xlab="Year",main="Fits and forecasts for Canada series")

## Warning in plot.window(xlim, ylim, log, ...): "plot.conf" is not a  
## graphical parameter

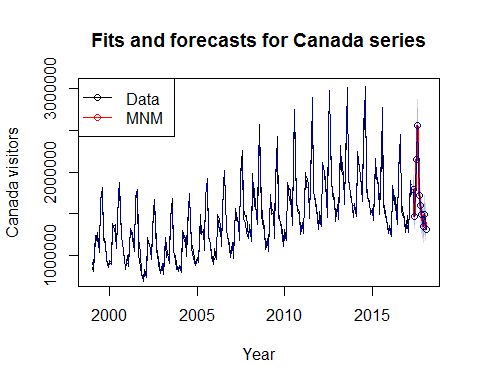
## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "plot.conf"  
## is not a graphical parameter

## Warning in axis(1, ...): "plot.conf" is not a graphical parameter

## Warning in axis(2, ...): "plot.conf" is not a graphical parameter

## Warning in box(...): "plot.conf" is not a graphical parameter

lines(fitted(frc.Canada.MNM), col="darkblue", lty=1)  
lines(frc.Canada.MNM$mean,col="darkblue", type="o")  
legend("topleft",lty=1, pch=1, col=1:3, c("Data","MNM "))



plot(frc.Others.MAdM, ylab="Rest of the wotld (others) visitors",plot.conf=FALSE, type="l", fcol="purple", xlab="Year",main="Fits and forecasts for Rest of the world (others) series")

## Warning in plot.window(xlim, ylim, log, ...): "plot.conf" is not a  
## graphical parameter

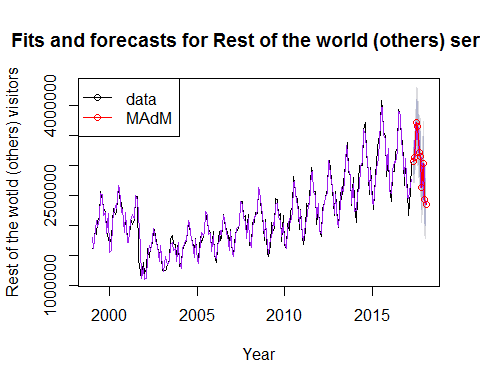
## Warning in title(main = main, xlab = xlab, ylab = ylab, ...): "plot.conf"  
## is not a graphical parameter

## Warning in axis(1, ...): "plot.conf" is not a graphical parameter

## Warning in axis(2, ...): "plot.conf" is not a graphical parameter

## Warning in box(...): "plot.conf" is not a graphical parameter

lines(fitted(frc.Others.MAdM), col="purple", lty=1)  
lines(frc.Others.MAdM$mean,col="red", type="o")  
legend("topleft",lty=1, pch=1, col=1:3, c("data","MAdM"))



after we run 10 months ahead forecast We get a very good fit with original data series and forecasts from ANA model for our mexico time series,MNM model from Canada Series and for other series additive damped trend Holt-Winters' model with the additive errors state space model.

## Conclusion--

After we go through all steps to find best models for our three time series for visitors visiting to USA from neighboring countries like Mexico,Canada and fromrest of the worlds we found our best three models for three time series to predict for 10 months ahead forecast ,and in all cases our forecasts lies within 95% confidence interval level.