Project\_Tennis

Effy Fang

August 19, 2019

library(fpp)

## Loading required package: forecast

## Loading required package: fma

## Loading required package: expsmooth

## Loading required package: lmtest

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: tseries

library(TSA)

##   
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':  
##   
## acf, arima

## The following object is masked from 'package:utils':  
##   
## tar

library(ggplot2)  
library(vars)

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following objects are masked from 'package:fma':  
##   
## cement, housing, petrol

## Loading required package: strucchange

## Loading required package: sandwich

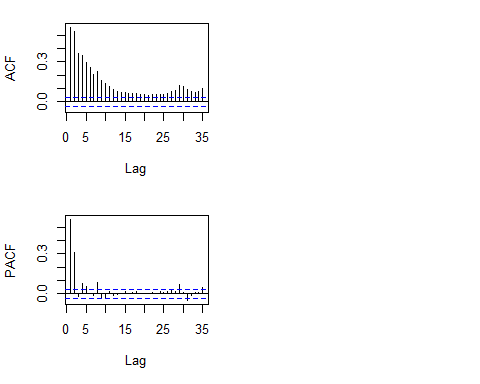
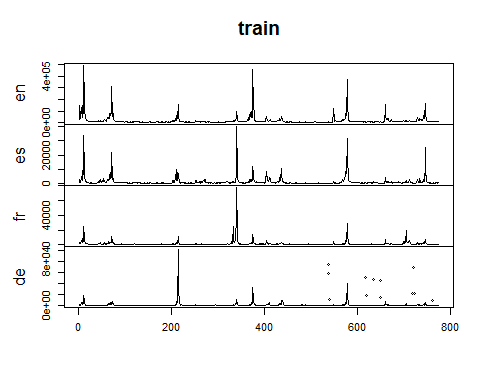
## Loading required package: urca

## data preparation

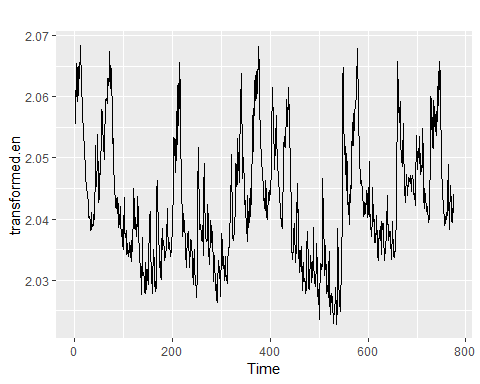
datapath <- "C:/Users/effyf/Documents/MSCA/time series/project/TimeSeries-Project/TSProject"  
project.dat<-readRDS(file=paste(datapath,"new\_train.rda",sep="/"))  
head(project.dat)

## rick\_y\_morty rick\_and\_morty travis\_fimmel vikings   
## 2015-07-01 " 181" "15887" " 2880" " 8440"   
## 2015-07-02 " 198" "12968" " 4693" " 8710"   
## 2015-07-03 " 148" " 8424" " 3368" " 8356"  
## 2015-07-04 " 108" " 7574" " 3068" " 8513"   
## 2015-07-05 " 137" " 7819" " 3603" "10155"   
## 2015-07-06 " 144" " 8597" " 3515" " 10275"  
## angeilna\_jolie serena\_williams\_en serena\_williams\_fr  
## 2015-07-01 "16024" "50829" " 1802"   
## 2015-07-02 "16436" "22283" " 1199"   
## 2015-07-03 " 15832" "147384" " 3407"   
## 2015-07-04 "18941" "49251" " 1850"   
## 2015-07-05 "19026" "37265" " 1508"   
## 2015-07-06 " 20362" "113394" " 4476"   
## serena\_williams\_es serena\_williams\_de appendicitis\_zh  
## 2015-07-01 " 878" " 1349" " 210"   
## 2015-07-02 " 666" " 1204" " 219"   
## 2015-07-03 " 2894" " 3204" " 190"   
## 2015-07-04 " 1225" " 1701" " 181"   
## 2015-07-05 " 780" " 803" " 238"   
## 2015-07-06 " 3298" " 6236" " 221"   
## jiakang\_zh shenjieshi\_zh  
## 2015-07-01 " 34" " 301"   
## 2015-07-02 " 31" " 294"   
## 2015-07-03 " 33" " 271"   
## 2015-07-04 " 62" " 307"   
## 2015-07-05 " 61" " 260"   
## 2015-07-06 " 37" " 269"

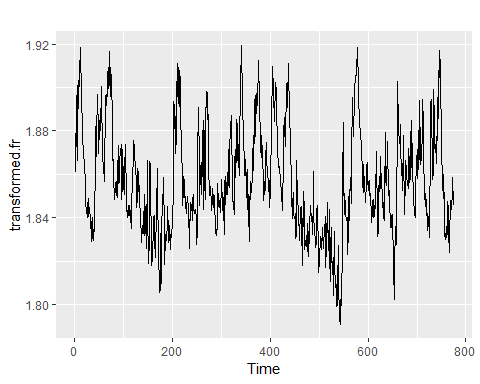
tennis.en=as.integer(project.dat[,6])  
tennis.fr=as.integer(project.dat[,7])  
tennis.es=as.integer(project.dat[,8])  
tennis.de=as.integer(project.dat[,9])  
tennis=cbind(en=tennis.en,es=tennis.es,fr=tennis.fr,de=tennis.de)  
  
  
# convert to ts  
ts.tennis=ts(tennis,frequency = 1) #,start=c(2015,7))   
train <- window(ts.tennis, start=1, end=775)  
test <-window(ts.tennis,start=776) # test on the last 4 weeks  
  
ts.weekly <- ts(tennis, frequency = 7)  
train.weekly <- window(ts.weekly, end=c(111,5))  
test.weekly <- window(ts.weekly, start=c(111,6))  
tsdisplay(train)

 the variance is not stable, need boxcox transformation

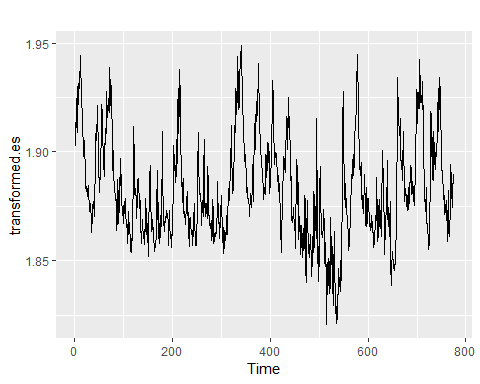
lambda1 <- BoxCox.lambda(train[,1])  
transformed.en <- BoxCox(train[,1],lambda=lambda1)  
lambda2 <- BoxCox.lambda(train[,2])  
transformed.fr <- BoxCox(train[,2],lambda=lambda2)  
lambda3 <- BoxCox.lambda(train[,3])  
transformed.es <- BoxCox(train[,3],lambda=lambda3)  
lambda4 <- BoxCox.lambda(train[,4])  
transformed.de <- BoxCox(train[,4],lambda=lambda4)  
autoplot(transformed.en)



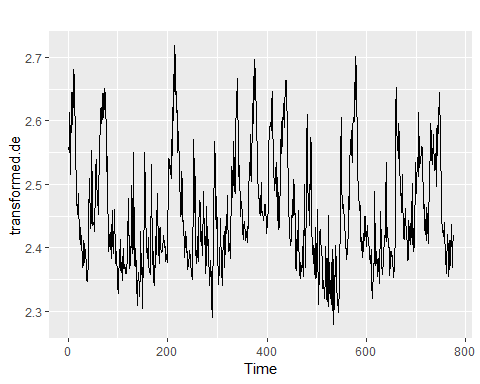
autoplot(transformed.fr)



autoplot(transformed.es)



autoplot(transformed.de)

 ## stationary test

kpss.test(transformed.en) #is level stationary

##   
## KPSS Test for Level Stationarity  
##   
## data: transformed.en  
## KPSS Level = 0.378, Truncation lag parameter = 6, p-value =  
## 0.08664

adf.test(transformed.en)

##   
## Augmented Dickey-Fuller Test  
##   
## data: transformed.en  
## Dickey-Fuller = -4.996, Lag order = 9, p-value = 0.01  
## alternative hypothesis: stationary

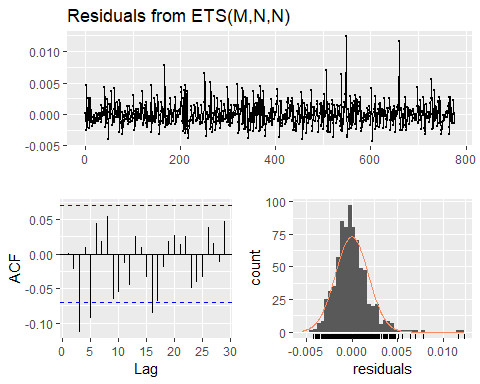
# model fitting

## exponential smoothing

#m.ses <- ses(transformed.en)  
#summary(m.ses)  
#m.holt <- holt(transformed.en)  
#summary(m.holt)  
m.ets <- ets(transformed.en) # best amoung the 3  
summary(m.ets)

## ETS(M,N,N)   
##   
## Call:  
## ets(y = transformed.en)   
##   
## Smoothing parameters:  
## alpha = 0.9542   
##   
## Initial states:  
## l = 2.0608   
##   
## sigma: 0.0018  
##   
## AIC AICc BIC   
## -3523.282 -3523.251 -3509.324   
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE  
## Training set -2.672513e-05 0.003687297 0.002711013 -0.001472833 0.1325645  
## MASE ACF1  
## Training set 0.999373 0.0004279761

checkresiduals(m.ets)



##   
## Ljung-Box test  
##   
## data: Residuals from ETS(M,N,N)  
## Q\* = 27.231, df = 8, p-value = 0.0006449  
##   
## Model df: 2. Total lags used: 10

residuals do not look good. we’ll not consider this model.

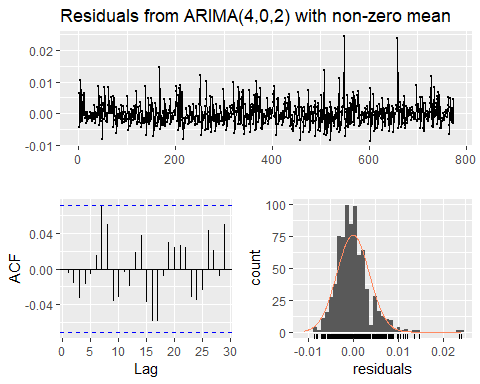
f.ets=forecast(m.ets,h=28)  
error.ets=InvBoxCox(f.ets$mean,lambda=lambda1)-test[,1]  
#unique(InvBoxCox(f.ets$mean,lambda=lambda1))  
#tail(tennis.en)

## arima

# fit auto.arima on en  
m.auto=auto.arima(train[,1],lambda=lambda1)  
summary(m.auto)

## Series: train[, 1]   
## ARIMA(4,0,2) with non-zero mean   
## Box Cox transformation: lambda= -0.4826064   
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ma1 ma2 mean  
## -0.7326 0.7675 0.7293 0.0042 1.6715 0.8172 2.0428  
## s.e. 0.1684 0.0685 0.1372 0.0473 0.1647 0.1291 0.0019  
##   
## sigma^2 estimated as 1.308e-05: log likelihood=3259.97  
## AIC=-6503.93 AICc=-6503.75 BIC=-6466.71  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 3185.942 30567.34 7144.078 -3.723988 22.4529 0.7655936  
## ACF1  
## Training set -0.1083828

checkresiduals(m.auto)



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(4,0,2) with non-zero mean  
## Q\* = 9.2117, df = 3, p-value = 0.0266  
##   
## Model df: 7. Total lags used: 10

residuals look okay, pass ljung-box test if we lower the significance level.

f.auto <- forecast(m.auto,h=28)  
error.auto <- f.auto$mean-test[,1]  
#plot(forecast(m.auto,h=7),include=100)

## arima

eacf(transformed.en)

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

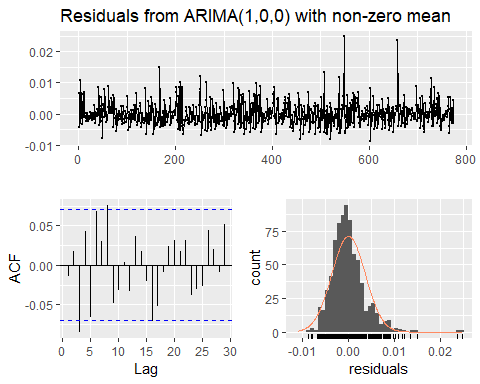
m.1.0 <- Arima(transformed.en, order=c(1,0,0))#, lambda=lambda1)  
m.1.1 <- Arima(transformed.en, order=c(1,0,1))#, lambda=lambda1)  
summary(m.1.0) # better

## Series: transformed.en   
## ARIMA(1,0,0) with non-zero mean   
##   
## Coefficients:  
## ar1 mean  
## 0.9327 2.0428  
## s.e. 0.0129 0.0019  
##   
## sigma^2 estimated as 1.322e-05: log likelihood=3253.29  
## AIC=-6500.58 AICc=-6500.55 BIC=-6486.62  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE  
## Training set -3.636497e-05 0.003631728 0.002658642 -0.002091913 0.1300157  
## MASE ACF1  
## Training set 0.9800672 -0.01469448

summary(m.1.1)

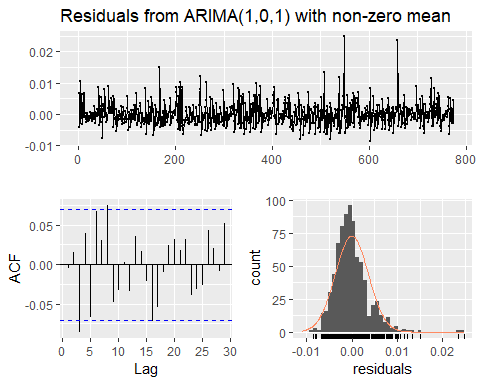
## Series: transformed.en   
## ARIMA(1,0,1) with non-zero mean   
##   
## Coefficients:  
## ar1 ma1 mean  
## 0.9344 -0.0117 2.0427  
## s.e. 0.0137 0.0381 0.0019  
##   
## sigma^2 estimated as 1.324e-05: log likelihood=3253.33  
## AIC=-6498.67 AICc=-6498.62 BIC=-6480.06  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE  
## Training set -3.023698e-05 0.003631508 0.00265909 -0.001791309 0.1300372  
## MASE ACF1  
## Training set 0.9802326 -0.004621347

checkresiduals(m.1.0)



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(1,0,0) with non-zero mean  
## Q\* = 22.304, df = 8, p-value = 0.004383  
##   
## Model df: 2. Total lags used: 10

checkresiduals(m.1.1)



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(1,0,1) with non-zero mean  
## Q\* = 21.892, df = 7, p-value = 0.002652  
##   
## Model df: 3. Total lags used: 10

AIC from arima(1,0,0) is slightly lower but not different by much. However the redisual p-value is far from passing the ljung-box test. so we’ll not consider this model.

f.arima.100 <- forecast(m.1.0,h=28)  
error.arima.100 <- InvBoxCox(f.arima.100$mean,lambda=lambda1)-test[,1]

## sarima

# reconstruct ts with 365, assume annual seasonality   
### not good!   
#annual.en=ts(transformed.en,frequency=365)  
# fit auto.arima on en  
#m2=auto.arima(annual.en,seasonal=T, D=1)  
#summary(m2)  
  
# reconstruct ts with 30, assume weekly seasonality   
### weekly is better.  
#monthly.en=ts(transformed.en,frequency=30)  
# fit auto.arima on en  
#m.monthly=auto.arima(monthly.en,seasonal=T, D=1)  
#summary(m.monthly)  
#checkresiduals(m.monthly)

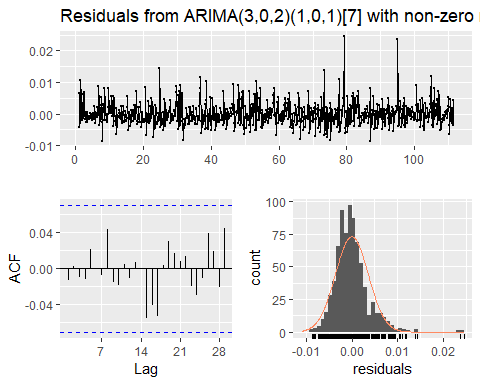
## weekly

# fit auto.arima on weekly ts   
m.weekly=auto.arima(train.weekly[,1],seasonal=T,lambda=lambda1)  
summary(m.weekly)

## Series: train.weekly[, 1]   
## ARIMA(3,0,2)(1,0,1)[7] with non-zero mean   
## Box Cox transformation: lambda= -0.4826064   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1 ma2 sar1 sma1 mean  
## -0.7796 0.7492 0.7451 1.7197 0.8654 0.5562 -0.4573 2.0428  
## s.e. 0.0734 0.0352 0.0653 0.0598 0.0543 0.2532 0.2709 0.0019  
##   
## sigma^2 estimated as 1.296e-05: log likelihood=3263.86  
## AIC=-6509.72 AICc=-6509.48 BIC=-6467.84  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 3126.547 30712.47 7157.469 -3.735335 22.26676 0.4569681  
## ACF1  
## Training set -0.1147682

AIC is lower and loglikelyhood is better. -6720 to -6726

checkresiduals(m.weekly)



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(3,0,2)(1,0,1)[7] with non-zero mean  
## Q\* = 2.8275, df = 6, p-value = 0.8302  
##   
## Model df: 8. Total lags used: 14

the residuals also look like white noise and pass ljung-box test.

f.sarima <- forecast(m.weekly,h=28)  
error.sarima <- f.sarima$mean-test.weekly[,1]

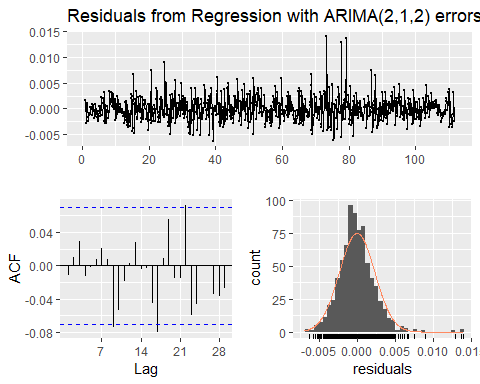
## arima error regression

adding the influence from es, fr, and de pages.

m.reg <- auto.arima(train.weekly[,1],xreg =cbind(transformed.es, transformed.fr, transformed.de),seasonal=T, lambda=lambda1)  
summary(m.reg)

## Series: train.weekly[, 1]   
## Regression with ARIMA(2,1,2) errors   
## Box Cox transformation: lambda= -0.4826064   
##   
## Coefficients:  
## ar1 ar2 ma1 ma2 transformed.es transformed.fr  
## -0.3616 0.4353 0.0044 -0.5897 0.1276 0.1076  
## s.e. 0.1661 0.0908 0.1616 0.0829 0.0101 0.0092  
## transformed.de  
## 0.0198  
## s.e. 0.0025  
##   
## sigma^2 estimated as 5.481e-06: log likelihood=3593.54  
## AIC=-7171.08 AICc=-7170.89 BIC=-7133.86  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set 311.7833 12379.86 3344.256 -1.655122 13.51809 0.2135138  
## ACF1  
## Training set 0.05532041

checkresiduals(m.reg)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(2,1,2) errors  
## Q\* = 8.9192, df = 7, p-value = 0.2585  
##   
## Model df: 7. Total lags used: 14

further lowers AIC, residuals look good.

transformed.test=cbind(transformed.es=BoxCox(test[,2],lambda=lambda2),  
 transformed.fr=BoxCox(test[,3],lambda=lambda3),  
 transformed.de=BoxCox(test[,4],lambda=lambda4))  
f.xreg <- forecast(m.reg,xreg=transformed.test,h=28)  
error.xreg <- f.xreg$mean-test.weekly[,1]

## VAR

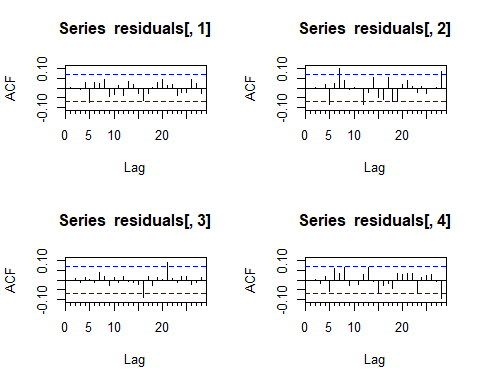
transformed <- cbind(transformed.en, transformed.es, transformed.fr, transformed.de)  
VARselect(transformed)

## $selection  
## AIC(n) HQ(n) SC(n) FPE(n)   
## 4 2 2 4   
##   
## $criteria  
## 1 2 3 4  
## AIC(n) -3.726759e+01 -3.736971e+01 -3.738896e+01 -3.739569e+01  
## HQ(n) -3.722089e+01 -3.728565e+01 -3.726754e+01 -3.723692e+01  
## SC(n) -3.714628e+01 -3.715136e+01 -3.707357e+01 -3.698326e+01  
## FPE(n) 6.529701e-17 5.895814e-17 5.783466e-17 5.744728e-17  
## 5 6 7 8  
## AIC(n) -3.736920e+01 -3.736076e+01 -3.736741e+01 -3.734996e+01  
## HQ(n) -3.717307e+01 -3.712727e+01 -3.709656e+01 -3.704175e+01  
## SC(n) -3.685973e+01 -3.675424e+01 -3.666385e+01 -3.654935e+01  
## FPE(n) 5.899093e-17 5.949320e-17 5.910225e-17 6.014687e-17  
## 9 10  
## AIC(n) -3.734614e+01 -3.732560e+01  
## HQ(n) -3.700057e+01 -3.694268e+01  
## SC(n) -3.644849e+01 -3.633091e+01  
## FPE(n) 6.038239e-17 6.164173e-17

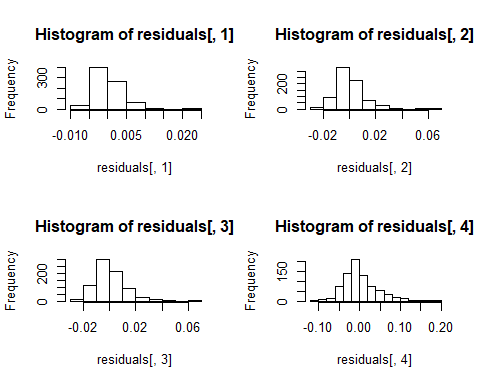
# 4 selected by aic, 2 by bic  
m.var <- VAR(transformed,p=4,type = "both")  
summary(m.var)

##   
## VAR Estimation Results:  
## =========================   
## Endogenous variables: transformed.en, transformed.es, transformed.fr, transformed.de   
## Deterministic variables: both   
## Sample size: 771   
## Log Likelihood: 10112.041   
## Roots of the characteristic polynomial:  
## 0.9306 0.8803 0.8803 0.8536 0.58 0.5522 0.5522 0.5103 0.5103 0.4884 0.4884 0.3527 0.3148 0.3148 0.2097 0.1475  
## Call:  
## VAR(y = transformed, p = 4, type = "both")  
##   
##   
## Estimation results for equation transformed.en:   
## ===============================================   
## transformed.en = transformed.en.l1 + transformed.es.l1 + transformed.fr.l1 + transformed.de.l1 + transformed.en.l2 + transformed.es.l2 + transformed.fr.l2 + transformed.de.l2 + transformed.en.l3 + transformed.es.l3 + transformed.fr.l3 + transformed.de.l3 + transformed.en.l4 + transformed.es.l4 + transformed.fr.l4 + transformed.de.l4 + const + trend   
##   
## Estimate Std. Error t value Pr(>|t|)   
## transformed.en.l1 8.103e-01 5.615e-02 14.431 < 2e-16 \*\*\*  
## transformed.es.l1 2.876e-02 1.742e-02 1.651 0.099144 .   
## transformed.fr.l1 -6.321e-03 1.584e-02 -0.399 0.690026   
## transformed.de.l1 8.780e-03 4.150e-03 2.116 0.034688 \*   
## transformed.en.l2 9.223e-03 6.634e-02 0.139 0.889463   
## transformed.es.l2 1.463e-02 1.953e-02 0.749 0.454030   
## transformed.fr.l2 1.276e-03 1.807e-02 0.071 0.943733   
## transformed.de.l2 -5.518e-03 4.740e-03 -1.164 0.244749   
## transformed.en.l3 -2.373e-02 6.699e-02 -0.354 0.723289   
## transformed.es.l3 -2.905e-02 1.947e-02 -1.492 0.136243   
## transformed.fr.l3 3.003e-03 1.802e-02 0.167 0.867655   
## transformed.de.l3 6.811e-04 4.760e-03 0.143 0.886257   
## transformed.en.l4 2.095e-01 5.626e-02 3.723 0.000211 \*\*\*  
## transformed.es.l4 -1.545e-02 1.750e-02 -0.883 0.377365   
## transformed.fr.l4 -1.552e-02 1.580e-02 -0.982 0.326337   
## transformed.de.l4 -8.714e-03 4.161e-03 -2.094 0.036576 \*   
## const 3.568e-02 5.166e-02 0.691 0.489998   
## trend 7.077e-08 6.014e-07 0.118 0.906356   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
##   
## Residual standard error: 0.00357 on 753 degrees of freedom  
## Multiple R-Squared: 0.8745, Adjusted R-squared: 0.8716   
## F-statistic: 308.6 on 17 and 753 DF, p-value: < 2.2e-16   
##   
##   
## Estimation results for equation transformed.es:   
## ===============================================   
## transformed.es = transformed.en.l1 + transformed.es.l1 + transformed.fr.l1 + transformed.de.l1 + transformed.en.l2 + transformed.es.l2 + transformed.fr.l2 + transformed.de.l2 + transformed.en.l3 + transformed.es.l3 + transformed.fr.l3 + transformed.de.l3 + transformed.en.l4 + transformed.es.l4 + transformed.fr.l4 + transformed.de.l4 + const + trend   
##   
## Estimate Std. Error t value Pr(>|t|)   
## transformed.en.l1 7.348e-01 1.679e-01 4.377 1.37e-05 \*\*\*  
## transformed.es.l1 5.819e-01 5.209e-02 11.171 < 2e-16 \*\*\*  
## transformed.fr.l1 2.295e-02 4.738e-02 0.484 0.62820   
## transformed.de.l1 8.058e-03 1.241e-02 0.649 0.51629   
## transformed.en.l2 -7.279e-01 1.984e-01 -3.670 0.00026 \*\*\*  
## transformed.es.l2 1.753e-01 5.839e-02 3.003 0.00277 \*\*   
## transformed.fr.l2 4.071e-02 5.404e-02 0.753 0.45147   
## transformed.de.l2 1.449e-03 1.417e-02 0.102 0.91859   
## transformed.en.l3 1.924e-02 2.003e-01 0.096 0.92353   
## transformed.es.l3 4.932e-03 5.823e-02 0.085 0.93252   
## transformed.fr.l3 -3.637e-02 5.387e-02 -0.675 0.49976   
## transformed.de.l3 8.120e-03 1.423e-02 0.570 0.56852   
## transformed.en.l4 3.506e-01 1.682e-01 2.084 0.03748 \*   
## transformed.es.l4 3.038e-02 5.232e-02 0.581 0.56165   
## transformed.fr.l4 -1.635e-02 4.726e-02 -0.346 0.72952   
## transformed.de.l4 -3.285e-02 1.244e-02 -2.640 0.00845 \*\*   
## const -3.615e-01 1.545e-01 -2.340 0.01954 \*   
## trend -1.513e-06 1.798e-06 -0.841 0.40059   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
##   
## Residual standard error: 0.01068 on 753 degrees of freedom  
## Multiple R-Squared: 0.7981, Adjusted R-squared: 0.7936   
## F-statistic: 175.1 on 17 and 753 DF, p-value: < 2.2e-16   
##   
##   
## Estimation results for equation transformed.fr:   
## ===============================================   
## transformed.fr = transformed.en.l1 + transformed.es.l1 + transformed.fr.l1 + transformed.de.l1 + transformed.en.l2 + transformed.es.l2 + transformed.fr.l2 + transformed.de.l2 + transformed.en.l3 + transformed.es.l3 + transformed.fr.l3 + transformed.de.l3 + transformed.en.l4 + transformed.es.l4 + transformed.fr.l4 + transformed.de.l4 + const + trend   
##   
## Estimate Std. Error t value Pr(>|t|)   
## transformed.en.l1 7.287e-01 1.749e-01 4.167 3.44e-05 \*\*\*  
## transformed.es.l1 7.616e-02 5.425e-02 1.404 0.160802   
## transformed.fr.l1 5.476e-01 4.934e-02 11.098 < 2e-16 \*\*\*  
## transformed.de.l1 1.745e-02 1.292e-02 1.350 0.177266   
## transformed.en.l2 -7.588e-01 2.066e-01 -3.673 0.000257 \*\*\*  
## transformed.es.l2 1.098e-01 6.082e-02 1.805 0.071430 .   
## transformed.fr.l2 9.881e-02 5.628e-02 1.756 0.079533 .   
## transformed.de.l2 1.287e-03 1.476e-02 0.087 0.930548   
## transformed.en.l3 2.035e-02 2.086e-01 0.098 0.922312   
## transformed.es.l3 -8.536e-02 6.065e-02 -1.407 0.159706   
## transformed.fr.l3 6.185e-02 5.611e-02 1.102 0.270666   
## transformed.de.l3 -4.691e-03 1.482e-02 -0.316 0.751740   
## transformed.en.l4 4.127e-01 1.752e-01 2.356 0.018745 \*   
## transformed.es.l4 -6.813e-02 5.449e-02 -1.250 0.211546   
## transformed.fr.l4 5.654e-02 4.922e-02 1.149 0.251032   
## transformed.de.l4 -2.713e-02 1.296e-02 -2.094 0.036615 \*   
## const -4.145e-01 1.609e-01 -2.576 0.010186 \*   
## trend -2.321e-06 1.873e-06 -1.239 0.215743   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
##   
## Residual standard error: 0.01112 on 753 degrees of freedom  
## Multiple R-Squared: 0.7944, Adjusted R-squared: 0.7898   
## F-statistic: 171.2 on 17 and 753 DF, p-value: < 2.2e-16   
##   
##   
## Estimation results for equation transformed.de:   
## ===============================================   
## transformed.de = transformed.en.l1 + transformed.es.l1 + transformed.fr.l1 + transformed.de.l1 + transformed.en.l2 + transformed.es.l2 + transformed.fr.l2 + transformed.de.l2 + transformed.en.l3 + transformed.es.l3 + transformed.fr.l3 + transformed.de.l3 + transformed.en.l4 + transformed.es.l4 + transformed.fr.l4 + transformed.de.l4 + const + trend   
##   
## Estimate Std. Error t value Pr(>|t|)   
## transformed.en.l1 2.035e+00 6.335e-01 3.213 0.001370 \*\*   
## transformed.es.l1 3.729e-01 1.965e-01 1.897 0.058164 .   
## transformed.fr.l1 7.226e-02 1.788e-01 0.404 0.686131   
## transformed.de.l1 6.091e-01 4.682e-02 13.010 < 2e-16 \*\*\*  
## transformed.en.l2 -1.819e+00 7.484e-01 -2.431 0.015293 \*   
## transformed.es.l2 1.857e-01 2.203e-01 0.843 0.399465   
## transformed.fr.l2 1.187e-01 2.039e-01 0.582 0.560487   
## transformed.de.l2 6.557e-02 5.348e-02 1.226 0.220530   
## transformed.en.l3 1.084e+00 7.558e-01 1.434 0.151911   
## transformed.es.l3 -6.155e-01 2.197e-01 -2.801 0.005220 \*\*   
## transformed.fr.l3 -1.063e-01 2.033e-01 -0.523 0.601117   
## transformed.de.l3 9.907e-02 5.370e-02 1.845 0.065443 .   
## transformed.en.l4 -7.347e-02 6.347e-01 -0.116 0.907876   
## transformed.es.l4 9.380e-02 1.974e-01 0.475 0.634759   
## transformed.fr.l4 -4.619e-02 1.783e-01 -0.259 0.795672   
## transformed.de.l4 -1.644e-02 4.695e-02 -0.350 0.726358   
## const -2.050e+00 5.829e-01 -3.518 0.000461 \*\*\*  
## trend -3.870e-07 6.785e-06 -0.057 0.954536   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
##   
## Residual standard error: 0.04028 on 753 degrees of freedom  
## Multiple R-Squared: 0.792, Adjusted R-squared: 0.7873   
## F-statistic: 168.6 on 17 and 753 DF, p-value: < 2.2e-16   
##   
##   
##   
## Covariance matrix of residuals:  
## transformed.en transformed.es transformed.fr transformed.de  
## transformed.en 1.275e-05 2.590e-05 2.541e-05 8.362e-05  
## transformed.es 2.590e-05 1.140e-04 6.726e-05 2.351e-04  
## transformed.fr 2.541e-05 6.726e-05 1.236e-04 2.319e-04  
## transformed.de 8.362e-05 2.351e-04 2.319e-04 1.623e-03  
##   
## Correlation matrix of residuals:  
## transformed.en transformed.es transformed.fr transformed.de  
## transformed.en 1.0000 0.6794 0.6400 0.5815  
## transformed.es 0.6794 1.0000 0.5666 0.5468  
## transformed.fr 0.6400 0.5666 1.0000 0.5177  
## transformed.de 0.5815 0.5468 0.5177 1.0000

residuals=residuals(m.var)  
par(mfrow=c(2,2))  
Acf(residuals[,1])  
Acf(residuals[,2])  
Acf(residuals[,3])  
Acf(residuals[,4])



par(mfrow=c(2,2))  
hist(residuals[,1])  
hist(residuals[,2])  
hist(residuals[,3])  
hist(residuals[,4])



Box.test(residuals[,1],type="Ljung-Box")

##   
## Box-Ljung test  
##   
## data: residuals[, 1]  
## X-squared = 3.9248e-05, df = 1, p-value = 0.995

Box.test(residuals[,2],type="Ljung-Box")

##   
## Box-Ljung test  
##   
## data: residuals[, 2]  
## X-squared = 6.9759e-05, df = 1, p-value = 0.9933

Box.test(residuals[,3],type="Ljung-Box")

##   
## Box-Ljung test  
##   
## data: residuals[, 3]  
## X-squared = 0.00018852, df = 1, p-value = 0.989

Box.test(residuals[,4],type="Ljung-Box")

##   
## Box-Ljung test  
##   
## data: residuals[, 4]  
## X-squared = 0.0033276, df = 1, p-value = 0.954

all the residuals look good

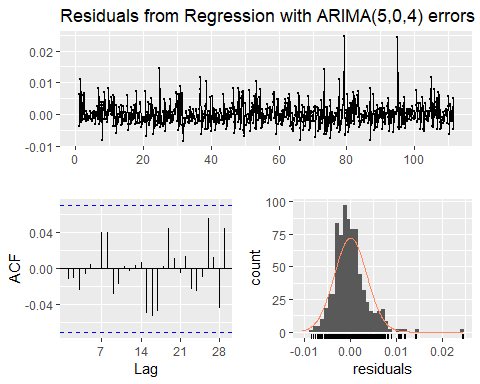
f.var <- forecast(m.var, h=28)  
error.var = InvBoxCox(f.var$forecast$transformed.en$mean, lambda=lambda1)-test[,1]

## fourier

transformed.weekly.en=ts(transformed.en,frequency = 7)  
harmonic <- fourier(transformed.weekly.en,K=3)  
m.fourier <- auto.arima(transformed.weekly.en,xreg=harmonic, seasonal=F)  
summary(m.fourier)

## Series: transformed.weekly.en   
## Regression with ARIMA(5,0,4) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 ar4 ar5 ma1 ma2 ma3  
## -1.3313 0.1642 1.0876 0.5685 0.1222 2.2728 1.9899 0.7598  
## s.e. 0.4157 NaN 0.6251 NaN NaN 0.3921 NaN NaN  
## ma4 intercept S1-7 C1-7 S2-7 C2-7 S3-7 C3-7  
## 0.1373 2.0429 -4e-04 8e-04 -2e-04 -3e-04 1e-04 1e-04  
## s.e. NaN 0.0020 2e-04 2e-04 1e-04 1e-04 1e-04 1e-04  
##   
## sigma^2 estimated as 1.275e-05: log likelihood=3274.47  
## AIC=-6514.94 AICc=-6514.13 BIC=-6435.84  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE  
## Training set -3.823015e-05 0.003533254 0.002577319 -0.0021633 0.1260188  
## MASE ACF1  
## Training set 0.413209 -0.01226784

checkresiduals(m.fourier)



##   
## Ljung-Box test  
##   
## data: Residuals from Regression with ARIMA(5,0,4) errors  
## Q\* = 11.817, df = 3, p-value = 0.008039  
##   
## Model df: 16. Total lags used: 19

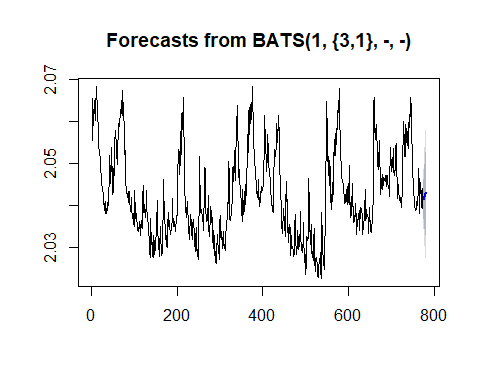
new.harmonic <- fourier(transformed.weekly.en,K=3,h=28)  
f.fourier <- forecast(m.fourier,xreg=new.harmonic, h=28)  
error.fourier <- InvBoxCox(f.fourier$mean, lambda=lambda1)-as.numeric(test[,1])

## TBATS

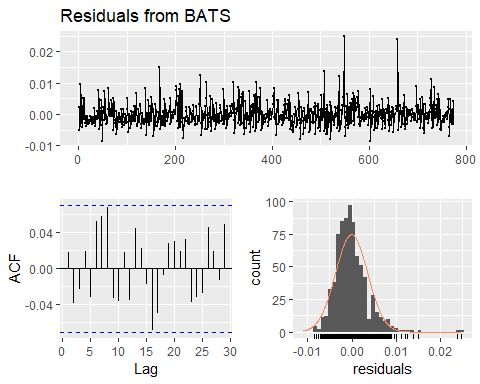
m.tbats <- tbats(transformed.en)  
print(m.tbats)

## BATS(1, {3,1}, -, -)  
##   
## Call: tbats(y = transformed.en)  
##   
## Parameters  
## Alpha: 0.01801944  
## AR coefficients: 0.198666 0.700155 -0.047834  
## MA coefficients: 0.704137  
##   
## Seed States:  
## [,1]  
## [1,] 2.058643  
## [2,] 0.000000  
## [3,] 0.000000  
## [4,] 0.000000  
## [5,] 0.000000  
##   
## Sigma: 0.003629454  
## AIC: -3532.974

plot(forecast(m.tbats, h=7))



checkresiduals(m.tbats)



##   
## Ljung-Box test  
##   
## data: Residuals from BATS  
## Q\* = 16.11, df = 3, p-value = 0.001077  
##   
## Model df: 10. Total lags used: 13

## model comparison

logLik(m.var)

## 'log Lik.' 10112.04 (df=72)

aicc <- cbind(ets=m.ets$aicc, auto.arima=m.auto$aicc, arima.100=m.1.0$aicc,   
 sarima=m.weekly$aicc, xreg=m.reg$aicc, varma.AIC =AIC(m.var), fourier = m.fourier$aicc)   
aicc

## ets auto.arima arima.100 sarima xreg varma.AIC fourier  
## [1,] -3523.251 -6503.746 -6500.547 -6509.48 -7170.887 -20080.08 -6514.131

sort(aicc)

## [1] -20080.082 -7170.887 -6514.131 -6509.480 -6503.746 -6500.547  
## [7] -3523.251

bic <- cbind(ets=m.ets$bic, auto.arima=m.auto$bic, arima.100=m.1.0$bic,   
 sarima=m.weekly$bic, xreg=m.reg$bic, varma =BIC(m.var), fourier = m.fourier$bic)   
bic

## ets auto.arima arima.100 sarima xreg varma fourier  
## [1,] -3509.324 -6466.711 -6486.62 -6467.84 -7133.863 -19745.45 -6435.841

ll <- cbind(ets=m.ets$loglik, auto.arima=m.auto$loglik, arima.100=m.1.0$loglik,   
 sarima=m.weekly$loglik, xreg=m.reg$loglik, varma =logLik(m.var), fourier = m.fourier$loglik)   
ll

## ets auto.arima arima.100 sarima xreg varma fourier  
## [1,] 1764.641 3259.967 3253.289 3263.858 3593.538 10112.04 3274.47

rmse.ets <- sqrt(mean(error.ets^2,na.rm=TRUE))  
rmse.auto <- sqrt(mean(error.auto^2,na.rm=TRUE))  
rmse.arima.100 <- sqrt(mean(error.arima.100^2,na.rm=TRUE))  
rmse.weekly <- sqrt(mean(error.sarima^2,na.rm=TRUE))  
rmse.xreg <- sqrt(mean(error.xreg^2,na.rm=TRUE))  
rmse.varma <- sqrt(mean(error.var^2,na.rm=TRUE))  
rmse.fourier <- sqrt(mean(error.fourier^2,na.rm=TRUE))  
  
rmse <- c(ets=rmse.ets, auto.arima = rmse.auto, arima.100=rmse.arima.100,  
 sarima=rmse.weekly, xreg=rmse.xreg, varma=rmse.varma, fourier=rmse.fourier)  
rmse

## ets auto.arima arima.100 sarima xreg varma   
## 17392.61 17044.36 17040.78 17080.81 11406.77 16716.11   
## fourier   
## 17132.87

data <- rbind(aicc,bic,ll,rmse)  
comparison <- as.data.frame(t(data),row.name=c('ets','auto.arima','arima(1,0,0)','sarima(weekly)', 'regression with arima error', 'varma', 'fourier'))  
colnames(comparison)<-c('aic', 'bic', 'loglik', 'rmse')  
comparison[order(comparison$aic),]

## aic bic loglik rmse  
## varma -20080.082 -19745.448 10112.041 16716.11  
## regression with arima error -7170.887 -7133.863 3593.538 11406.77  
## fourier -6514.131 -6435.841 3274.470 17132.87  
## sarima(weekly) -6509.480 -6467.840 3263.858 17080.81  
## auto.arima -6503.746 -6466.711 3259.967 17044.36  
## arima(1,0,0) -6500.547 -6486.620 3253.289 17040.78  
## ets -3523.251 -3509.324 1764.641 17392.61

comparison[order(comparison$rmse),]

## aic bic loglik rmse  
## regression with arima error -7170.887 -7133.863 3593.538 11406.77  
## varma -20080.082 -19745.448 10112.041 16716.11  
## arima(1,0,0) -6500.547 -6486.620 3253.289 17040.78  
## auto.arima -6503.746 -6466.711 3259.967 17044.36  
## sarima(weekly) -6509.480 -6467.840 3263.858 17080.81  
## fourier -6514.131 -6435.841 3274.470 17132.87  
## ets -3523.251 -3509.324 1764.641 17392.61