kdodson Draft Model

Karlos Dodson

8/18/2020

```r
#Load the provided data and split into train and test sets.
library(xts)
```

```
##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last
```

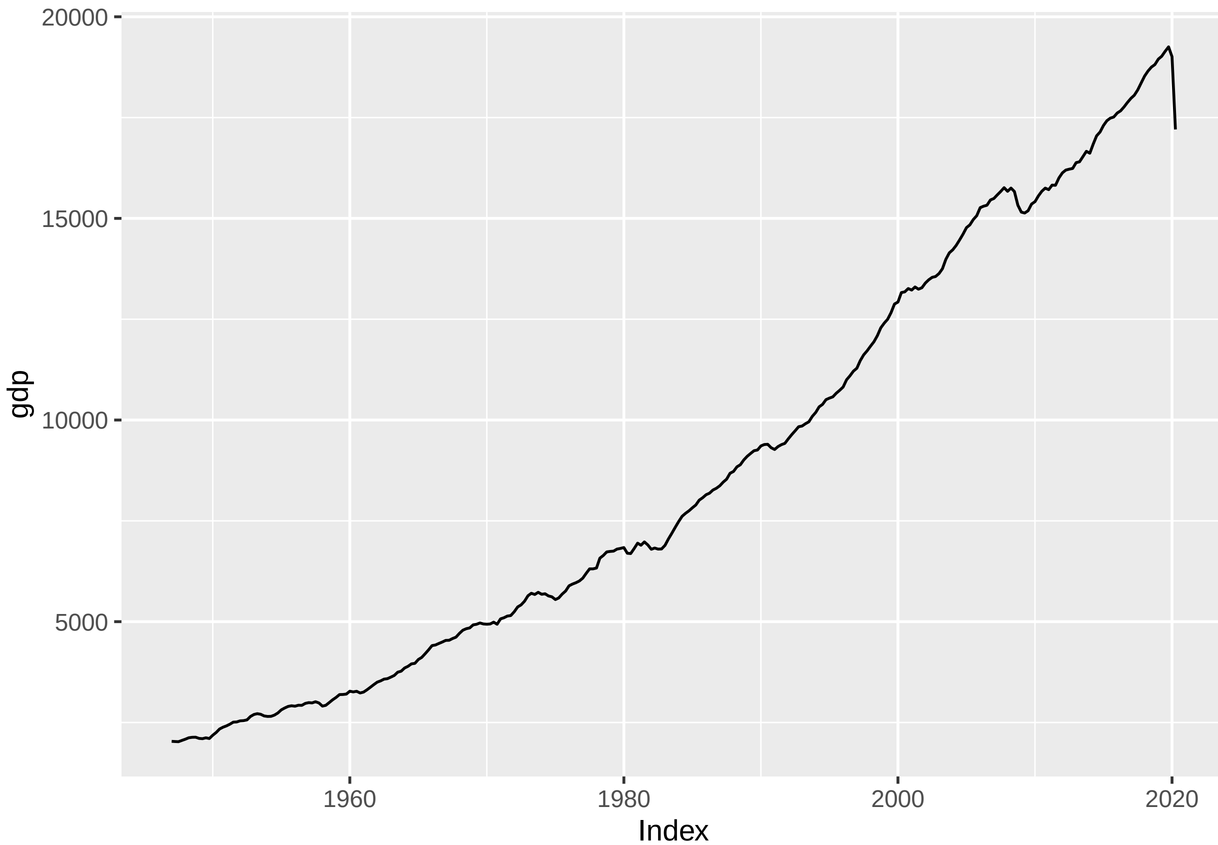
```r
gdp <- read.csv('~/Desktop/GDPC1.csv')
head(gdp)
```

```
##         DATE    GDPC1
## 1 1947-01-01 2033.061
## 2 1947-04-01 2027.639
## 3 1947-07-01 2023.452
## 4 1947-10-01 2055.103
## 5 1948-01-01 2086.017
## 6 1948-04-01 2120.450
```

```r
gdp$DATE <- as.Date(gdp$DATE)

gdp <- xts(gdp$GDPC1, gdp$DATE)

autoplot(gdp)
```




```r
ntrain <- round(nrow(gdp)*.8)
train <- gdp[1:ntrain, ]
test <- gdp[-c(1:ntrain), ]
ntest <- nrow(test)

# Model 1
aa_model <- auto.arima(train, trace=TRUE)
```

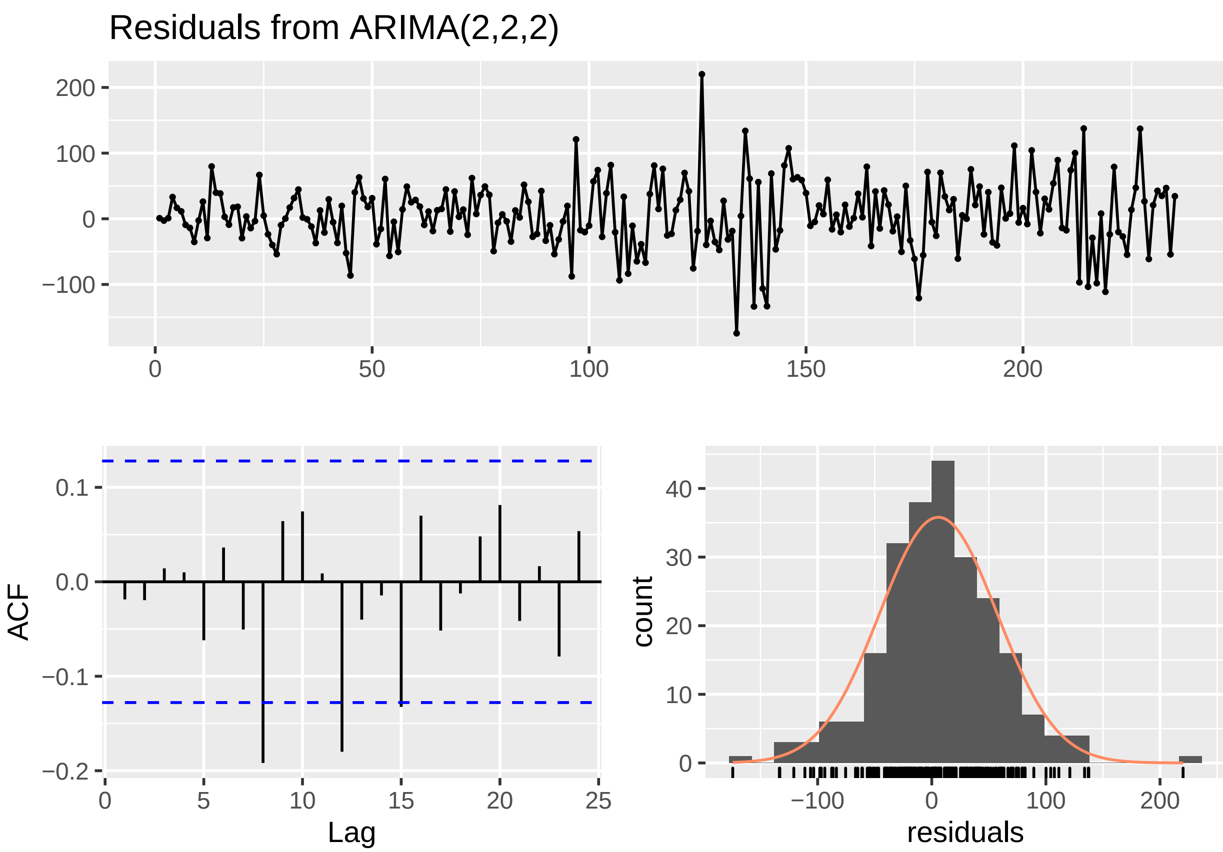
```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,2,2)                    : 2498.512
##  ARIMA(0,2,0)                    : 2590.527
##  ARIMA(1,2,0)                    : 2535.979
##  ARIMA(0,2,1)                    : 2514.011
##  ARIMA(1,2,2)                    : 2501.687
##  ARIMA(2,2,1)                    : 2499.753
##  ARIMA(3,2,2)                    : 2501.971
##  ARIMA(2,2,3)                    : 2500.595
##  ARIMA(1,2,1)                    : 2504.854
##  ARIMA(1,2,3)                    : 2501.295
##  ARIMA(3,2,1)                    : 2500.475
##  ARIMA(3,2,3)                    : 2501.385
##
##  Now re-fitting the best model(s) without approximations...
##
```

```
##  ARIMA(2,2,2)                    : 2515.577
##
##  Best model: ARIMA(2,2,2)
```

```r
aa_model
```

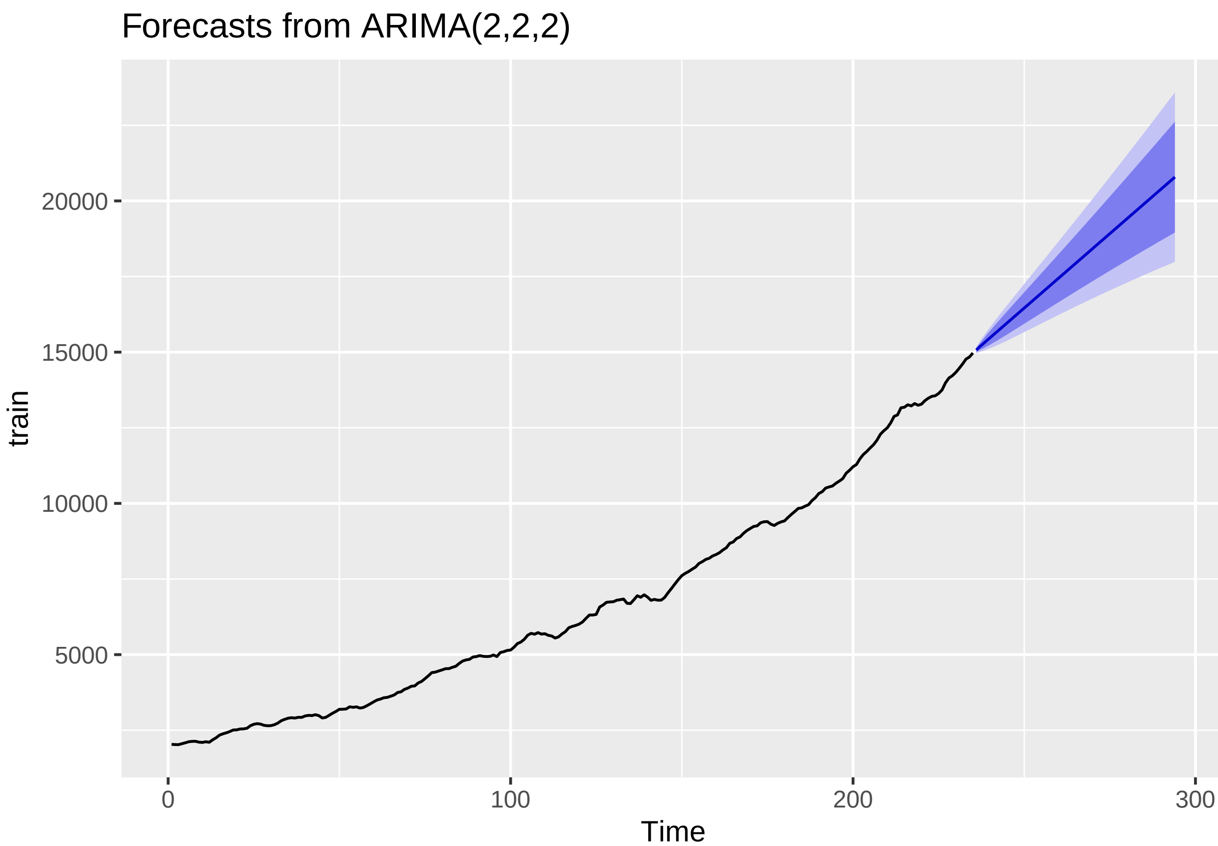
```
## Series: train
## ARIMA(2,2,2)
##
## Coefficients:
##           ar1     ar2      ma1      ma2
##       -0.2202  0.3194  -0.4747  -0.4697
## s.e.   0.2292  0.0764   0.2371   0.2252
##
## sigma^2 estimated as 2762:  log likelihood=-1252.66
## AIC=2515.31   AICc=2515.58   BIC=2532.57
```

```r
checkresiduals(aa_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,2,2)
## Q* = 13.533, df = 6, p-value = 0.03531
##
## Model df: 4.   Total lags used: 10
```

```r
aa_pred <- aa_model %>% forecast(h = ntest)
aa_pred %>% autoplot()
```




```r
accuracy(aa_pred, test)
```

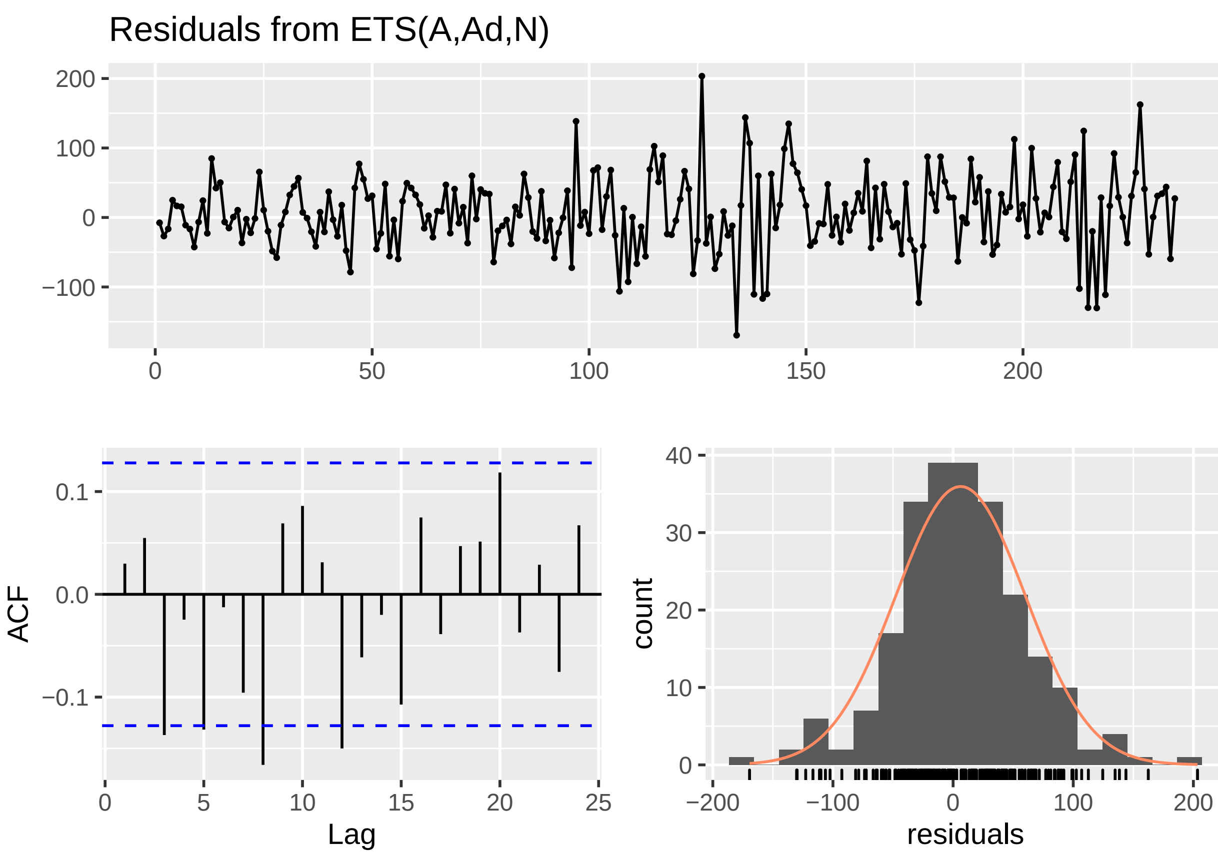
```
##                       ME       RMSE        MAE         MPE      MAPE
## Training set    5.762517   51.88394   39.06632  0.09629045 0.6857568
## Test set    -1207.194021 1361.89061 1211.13456 -7.11874242 7.1445387
##                   MASE        ACF1
## Training set  0.5911897 -0.01866561
## Test set     18.3280693          NA
```

```r
# Model 2
ets_model <- ets(train)
ets_model
```

```
## ETS(A,Ad,N)
##
## Call:
##  ets(y = train)
##
##   Smoothing parameters:
##     alpha = 0.9999
##     beta  = 0.3234
##     phi   = 0.9674
```

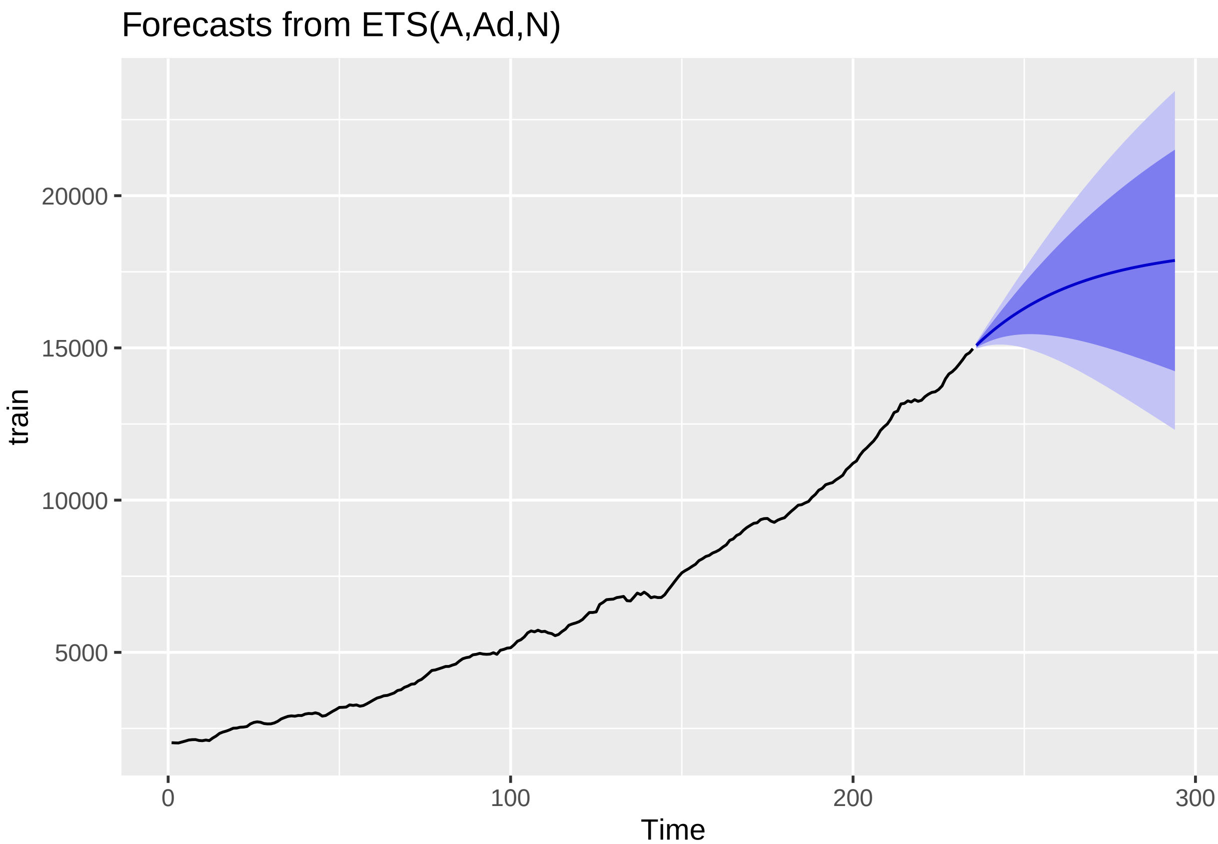
```
##
##   Initial states:
##     l = 2016.1438
##     b = 25.3683
##
##   sigma:  54.8557
##
##      AIC     AICc      BIC
## 3172.160 3172.529 3192.918
```

```r
checkresiduals(ets_model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(A,Ad,N)
## Q* = 21.814, df = 5, p-value = 0.000568
##
## Model df: 5.   Total lags used: 10
```

```r
ets_pred <- ets_model %>% forecast(h = ntest)
ets_pred %>% autoplot()
```




```r
accuracy(ets_pred, test)
```

```
##                       ME      RMSE       MAE         MPE      MAPE      MASE
## Training set    6.250539  54.26894  41.19406  0.09533967 0.7379589 0.6233887
## Test set     -164.332276 734.14816 601.97122 -1.24595856 3.5851146 9.1096156
##                    ACF1
## Training set 0.02984813
## Test set             NA
```

```r
# Model 3
#didn't work well
#var_model <- VAR(train)
#var_model
#checkresiduals(var_model)
```