R Notebook

```r
library(TSA)
```

```
## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1    2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-22. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##
## Attaching package: 'TSA'

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

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

```r
library(tseries)
library(astsa)
library(imputeTS)
```

```
##
## Attaching package: 'imputeTS'

## The following object is masked from 'package:tseries':
##
##     na.remove
```

```r
library(tsoutliers)
library(xts)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following object is masked from 'package:imputeTS':
##
##     na.locf

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

Original dataset

```
terror10 <-read.csv("input/editted_input1.csv")$num.attacks.with.kill.thresh
pdf("image/og_og_og.pdf")
plot(as.xts(ts(terror10, frequency = 12, start=1970)), main = "Number of Terrorist Attacks (unfiltered)
dev.off()
```

Interpolating Data

Here we interpolate our missing data with a linear model.

```
terror2 <- read.csv("input/og_num_casualities_greater_than_10.csv")
terror3 <- na.interpolation(terror2$num.attacks.with.kill.thresh, option="linear")
plot(as.xts(ts(terror3, frequency = 12, start=1970)), main = "Number of Terrorist Attacks (w/ Linear Im
pdf("image/og_ts.pdf")
lines(as.xts(ts(terror2$num.attacks.with.kill.thresh, frequency = 12, start=1970)), col="black", lwd=1.
dev.off()
```

Removing outliers

```
outlier_terror3 <- tso(ts(terror3), types = c("TC", "AO", "IO"))
```

```
## Warning in locate.outliers.oloop(y = y, fit = fit, types = types, cval =
## cval, : stopped when 'maxit.oloop = 4' was reached
```

```
plot(outlier_terror3)

#plot outlier effects
pdf("image/outlier_effects.pdf")
plot(as.xts(ts(outlier_terror3$effects, frequency = 12, start=1970)), main = "Outlier Effects", major.f
dev.off()

#Plot outlier time series
xts.terror3 <- as.xts(ts(terror3, frequency = 12, start=1970))
plot(as.xts(ts(terror3, frequency = 12, start=1970)), main = "Number of Terrorist Attacks (Outliers Rem
lines(as.xts(ts(outlier_terror3$yadj, frequency = 12, start=1970)), col="blue")
points(xts.terror3[427], col="red",pch=19, cex=1)
points(xts.terror3[516], col="red",pch=19, cex=1)
points(xts.terror3[521], col="red",pch=19, cex=1)

points(xts.terror3[523], col="red",pch=19, cex=1)
points(xts.terror3[547], col="red",pch=19, cex=1)
pdf("image/outlier_comparison.pdf")
points(xts.terror3[556], col="red",pch=19, cex=1)
dev.off()
```

Making, Training, and Validation set

```
terror4 <- outlier_terror3$yadj

#terror3 <- na.kalman(terror2$num.attacks, model="auto.arima")
cuttoff.index <- length(terror4) - 48 #floor(0.1 * length(terror3))
cuttoff.index2 <- length(terror4) - 12
terror4.valid <- terror4[(cuttoff.index+1) :cuttoff.index2]
terror4.testing <- terror4[(cuttoff.index2 + 1): length(terror4)]
terror4 <- terror4[1: cuttoff.index]

#plot(as.xts(ts(terror4, frequency = 12, start=1970)), main = "Number of Terrorist Attacks (Training Se

#plot(as.xts(ts(terror4.valid, frequency = 12, start=1970)), main = "Number of Terrorist Attacks (Valid
```

Chasing Stationarity

```
#log_terror4 <- log(outlier_terror3$yadj)
adf.test(terror4)
adf.test(diff(terror4))
```

```
## Warning in adf.test(diff(terror4)): p-value smaller than printed p-value
```

```
adf.test(diff(diff(terror4)))
```

```
## Warning in adf.test(diff(diff(terror4))): p-value smaller than printed p-
## value
```

```
pdf("image/first_diff.pdf")
plot(as.xts(ts(diff(terror4), frequency = 12, start=1970)), main = "Number of Terrorist Attacks (First
dev.off()

pdf("image/second_diff.pdf")
plot(as.xts(ts(diff(diff(terror4)), frequency = 12, start=1970)), main = "Number of Terrorist Attacks (
dev.off()

#ts.plot(diff(terror4))
#ts.plot(diff(diff(terror4)))
```

```
pdf("image/acf_og.pdf")
acf(terror4, main="ACF of Training Data")
dev.off()
pdf("image/pacf_og.pdf")
pacf(terror4, main="PACF of Training Data")
dev.off()

pdf("image/acf_first_diff.pdf")
acf(diff(terror4), main="ACF of First Diff Training Data")
dev.off()
pdf("image/pacf_first_diff.pdf")
pacf(diff(terror4), main="PACF of First Diff Training Data")
dev.off()
```

```
pdf("image/acf_second_diff.pdf")
acf(diff(terror4), main="ACF of Second Diff Training Data")
dev.off()
pdf("image/pacf_second_diff.pdf")
pacf(diff(terror4), main="PACF of Second Diff Training Data")
dev.off()
```

Periodogram; Figuring out Seasonality

```
m = floor(sqrt(length(diff(terror4))))
pdf("image/raw_periodogram.pdf")
mvspec(diff(terror4), log="no", main="Raw Periodogram (First Difference)", cex.main=1.5)
dev.off()

pdf("image/smooth_tapered_periodogram.pdf")
mvspec(diff(terror4), kernel('modified.daniell', m), log="no", taper=0.1, main="Smoothed and Tapered Pe
dev.off()

pdf("image/raw_periodogram_2.pdf")
m = floor(sqrt(length(diff(diff(terror4)))))
mvspec(diff(diff(terror4)), log="no", main="Raw Periodogram (First Difference)", cex.main=1.5)
dev.off()

pdf("image/smooth_tapered_periodogram_2.pdf")
mvspec(diff(diff(terror4)), kernel('modified.daniell', m), log="no", taper=0.1, main="Smoothed and Tape
dev.off()
```

Finding which model to use

```
eacf(diff(terror4))
eacf(diff(diff(terror4)))
```

```
sarima(terror4, 0, 1, 1)
sarima(terror4, 0, 1, 1, 1, 0, 1, 4)
sarima(terror4, 0, 1, 1, 1, 1, 1, 4)
sarima(terror4, 0, 1, 1, 1, 1, 2, 4)

sarima(terror4, 1, 1, 2)
sarima(terror4, 1, 1, 2, 1, 0, 1, 4)
sarima(terror4, 1, 1, 2, 1, 1, 1, 4)
sarima(terror4, 1, 1, 2, 1, 1, 2, 4)

sarima(terror4, 3, 2, 1)
sarima(terror4, 3, 2, 1, 1, 0, 1, 4)

sarima(terror4, 0, 1, 1, 1, 0, 1, 3)
sarima(terror4, 3, 2, 1, 1, 0, 1, 3)
```

```
pdf("image/best_model.pdf")
sarima(terror4, 0, 1, 1, 0, 0, 0, 0)
```

```
dev.off()
```

MSE calculations

change the model to see mse of different models.

```
#eacf(diff(diff(log_terror4)))
terror5 <- terror4
total_error <- 0
start_pts <- c(1, 13, 25)
for (i in start_pts)
{
  actual <- terror4.valid[i : (i + 11)]
  predicted <- sarima.for(terror5, 12, 3, 2, 1, 1, 0, 1, 3)$pred
  total_error <- total_error + sum((actual - predicted)^2)
  terror5 <- c(terror5, terror4.valid[i : (i + 11)])
}

mse <- total_error / length(start_pts)
mse
```

Predicting the future using our best model

```
val <- sarima.for(c(terror4, terror4.valid), 12, 0, 1, 1, 0, 0, 0, 0)
pred <-val$pred
err  <-val$se
total <- c(terror4, terror4.valid, terror4.testing)
par(cex.main = 2)
plot(as.xts(ts(total, frequency = 12, start=1970))[492:length(total)], main = "Number of Terrorist Attac
points(as.xts(ts(total, frequency = 12, start=1970)),col="lightgray",pch="o")
lines(as.xts(ts(c(terror4, terror4.valid), frequency = 12, start=1970)),col="black")
points(as.xts(ts(c(terror4, terror4.valid), frequency = 12, start=1970)),col="black",pch="o")
lines(as.xts(ts(pred, frequency = 12, start=2016)),col="blue")
lines(as.xts(ts(pred + err, frequency = 12, start=2016)),col="blue", lty="dashed")
lines(as.xts(ts(pred - err, frequency = 12, start=2016)),col="blue", lty="dashed")
lines(as.xts(ts(pred + 2*err, frequency = 12, start=2016)),col="blue", lty="dotted")
pdf("image/prediction_on_testing.pdf")
lines(as.xts(ts(pred - 2*err, frequency = 12, start=2016)),col="blue", lty="dotted")
dev.off()

mse <- sum((pred - terror4.testing)^2)
mse
```