In-class Exercise 2: Fitting Accident Data with seasonal ARIMA

Robert Smith

October 31, 2013

1. Read in acci.txt file from the course web site. Plot the time series. Does the plot looks stationary? Test for the stationarity using adf, pp, and kpss tests. What is your conculusion?

```
library(TSA)

acci_df <- read.csv("acci.txt")
acci <- ts(acci_df, frequency = 12)

plot(acci)
```

```
temp <- adf.test(acci)
pp.test(acci)
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  acci
## Dickey-Fuller Z(alpha) = -26.23, Truncation lag parameter = 3,
## p-value = 0.0108
## alternative hypothesis: stationary
```

```
kpss.test(acci)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  acci
## KPSS Level = 0.2922, Truncation lag parameter = 1, p-value = 0.1
```
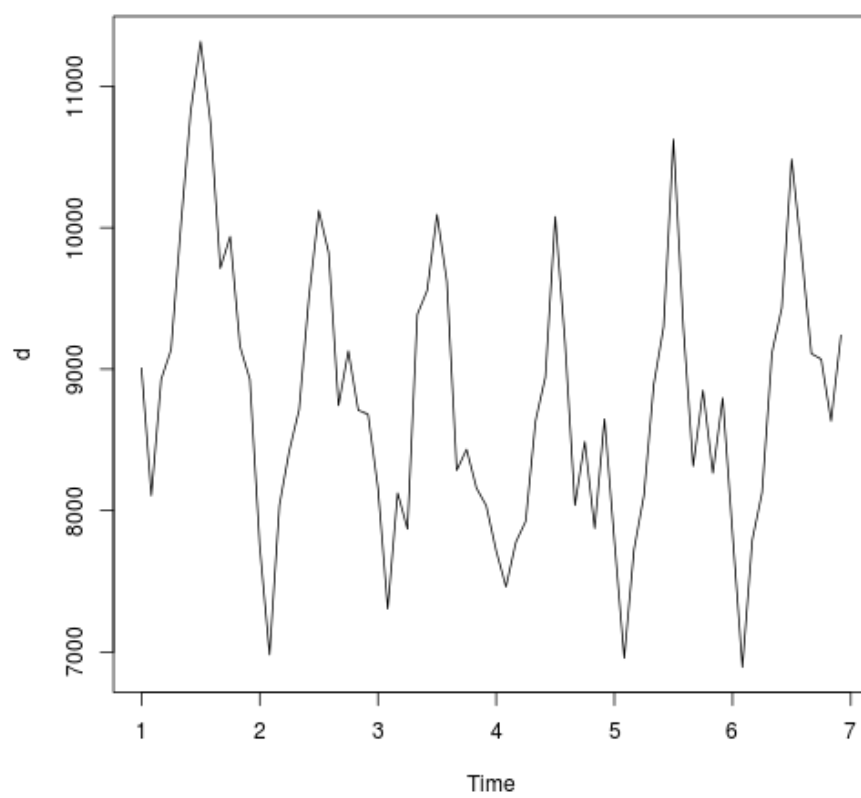
Figure 1: plot of chunk Q1

Based on the plot I cannot see any clear trends by eye, and all of the tests (ADF, PP & KPSS) also confirm that the data set is stationary.

2. Model the data using ARIMA(p,1,q)x(P,1,Q)12 model. Is the model fit adequate? How did you deteremine values of p,q,P,and Q?

```
library(forecast)

(est <- auto.arima(acci, d = 1, D = 1))

## Series: acci
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##        -0.426  -0.558
## s.e.    0.123   0.179
##
## sigma^2 estimated as 99480:  log likelihood=-425.5
## AIC=857.1   AICc=857.5   BIC=863.3

plot(forecast(est))
```

When I ran `auto.arima` to find the values of p, q, P & Q I found that the function was not able to find a reasonalble forecast. Based on the σ coefficient being 0.123 & 0.179 and an AIC of 857.06 I do not believe the model fit to be adequate.

3. Model the data using ARIMA(p,0,q)x(P,1,Q)12 model. Is the model fit adequate? How did you deteremine values of p,q,P,and Q?

```
(est <- auto.arima(acci, d = 0, D = 1))

## Series: acci
## ARIMA(2,0,0)(1,1,0)[12] with drift
##
## Coefficients:
##          ar1     ar2     sar1    drift
##        0.592   0.258   -0.350   -13.71
## s.e.   0.138   0.143    0.142    18.56
##
## sigma^2 estimated as 138798:  log likelihood=-348
## AIC=705.9   AICc=707   BIC=716.4
```
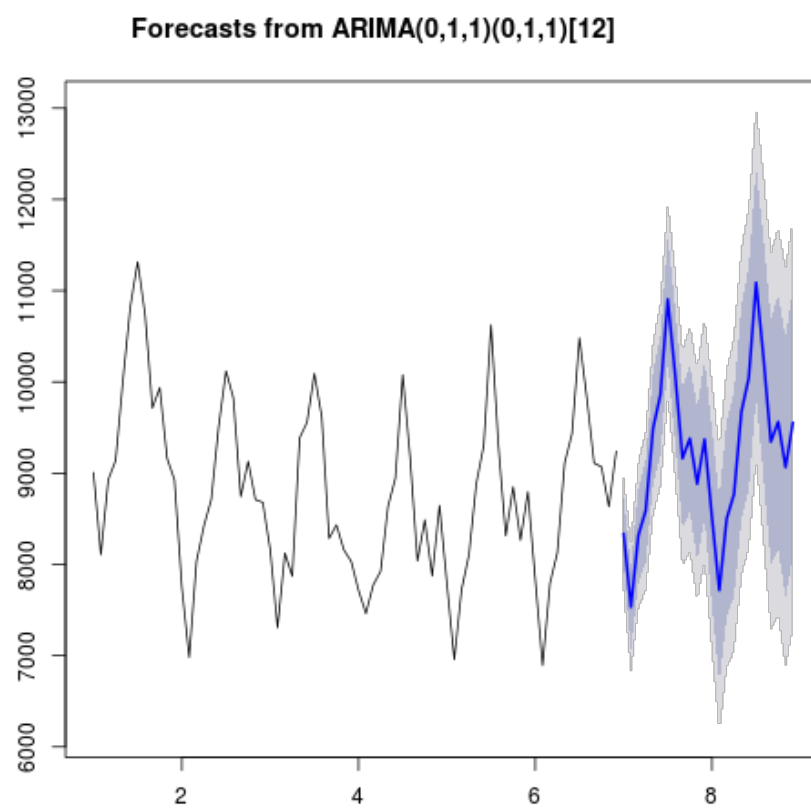
3

Figure 2: plot of chunk Q2

```
plot(forecast(est))
```
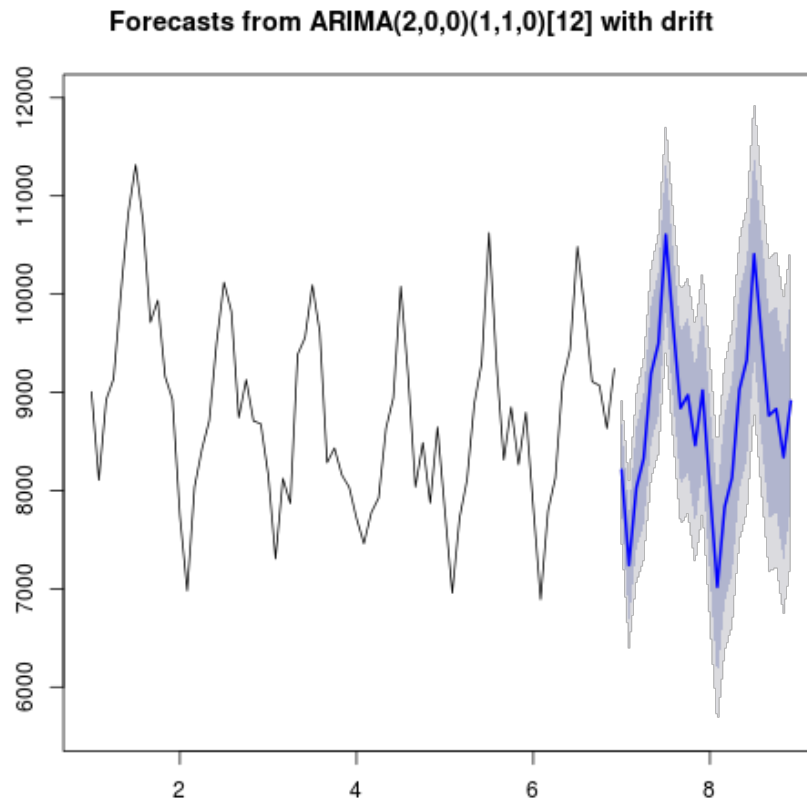
Forecasts from ARIMA(2,0,0)(1,1,0)[12] with drift

Figure 3: plot of chunk Q3

When I ran `auto.arima` to find the values of p, q, P & Q I found that the function was not able to find a reasonalble forecast. Based on the σ coefficient being 0.138, 0.143, 0.142 & 18.56 and an AIC of 705.92 I do not believe the model fit to be adequate either.

4. Is there a non-seasonal model ARIMA(p,d,q) that fits the data?

```
(est <- auto.arima(acci, seasonal = FALSE))

## Series: acci
```

```
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##           ar1      ar2      ma1     ma2   intercept
##         1.522   -0.748   -0.808   0.195      8783.3
## s.e.    0.137    0.115    0.211   0.150       129.4
##
## sigma^2 estimated as 391976:  log likelihood=-566.3
## AIC=1145    AICc=1146    BIC=1158
```
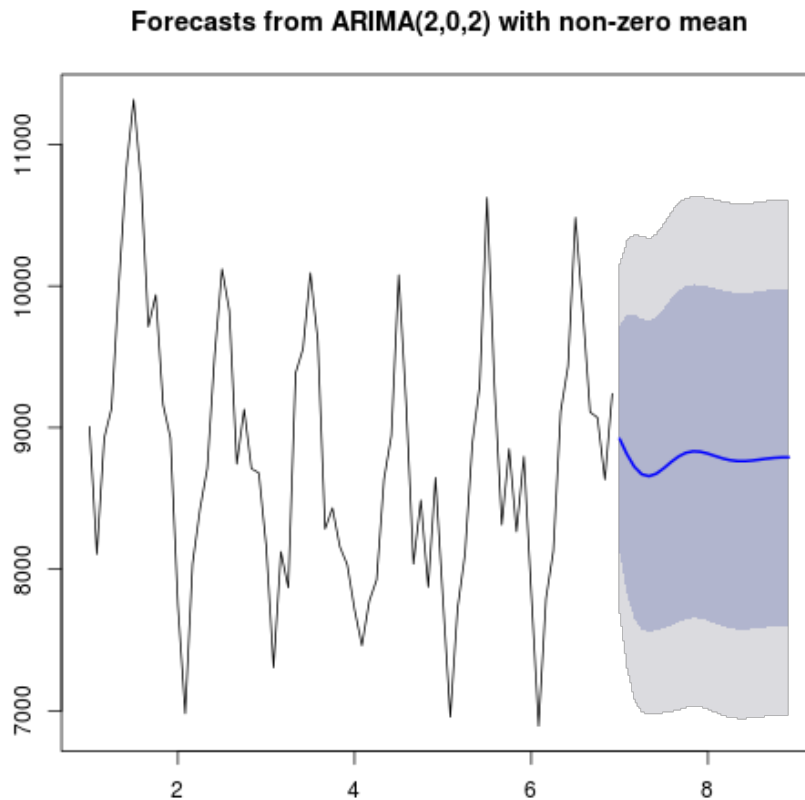
```
plot(forecast(est))
```


Figure 4: plot of chunk Q4

When I ran `auto.arima` to find the values of p, q, P & Q I found that the function was not able to find a reasonalble forecast with the 'd' term equal to

0. Based on the 's.e' coefficient showing a poor fit and an AIC of 1144.67 I do not believe the model fit to be adequate, less so than the previous models which include the assumption of seasonality.

5. What is the conceptual difference between model (2) and model (3)? Which model do you like better and why?

The conceptual difference between models #2 and #3 are that #2 differences both month-over-month and year-over-year, and #3 differences only year-over-year, neglecting the month-over-month information.

In this instance, I prefer the model created by #2 because I believe that this time series has a relationship with both year-over-year data (for seasonality considerations) and it also requires detrending month-over-month to add stationarity. While comparing the AIC and related statistics show the model from #3 is somewhat better, σ^2 and log likelihood both favor model #2 which make me believe that this model has greater predictive power.