Homework2 - Data analysis on sales data

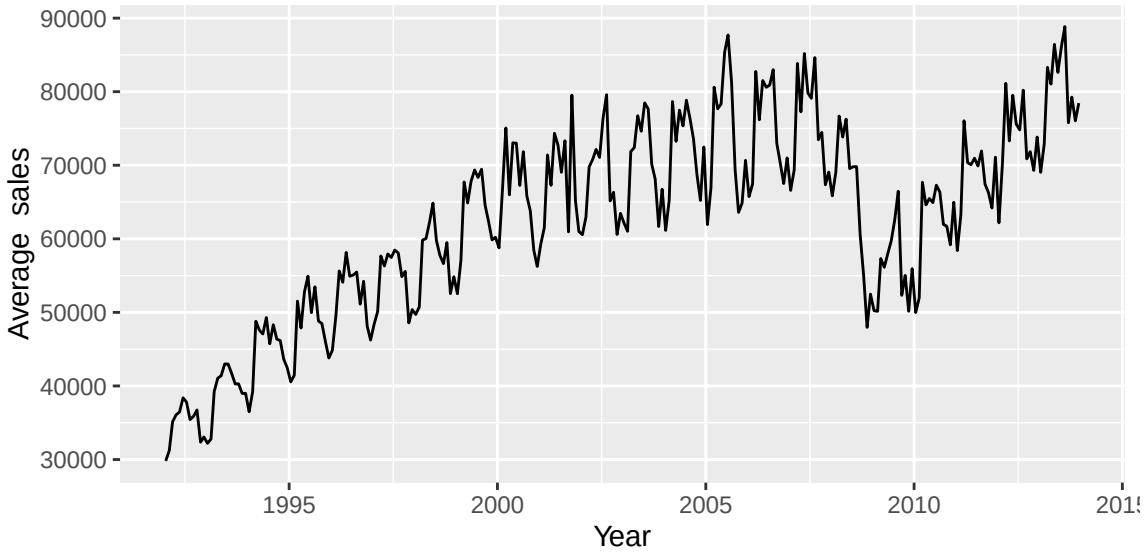
Data analysis for motor vehicle and parts dealers

Tadesse Zemicheal

January 20, 2016

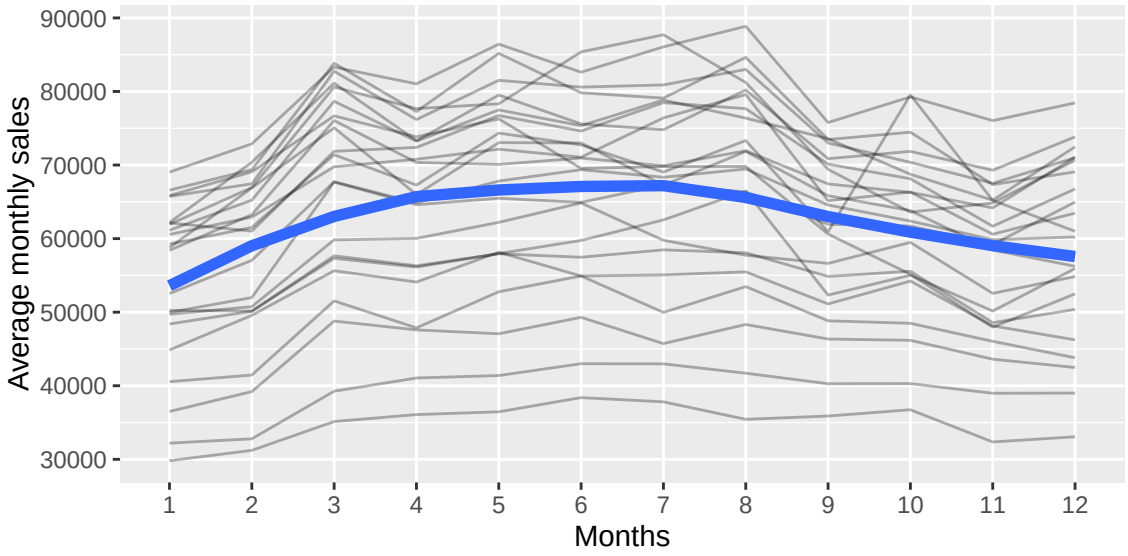
Introduction

The estimates of the total monthly sales for motor vehicle and parts dealers in the USA from 1992 to 2014 of monthly average sales is shown below. The main goal of the analysis is to describe seasonality, trend and residual variation in the data. First we will start with seasonal pattern.



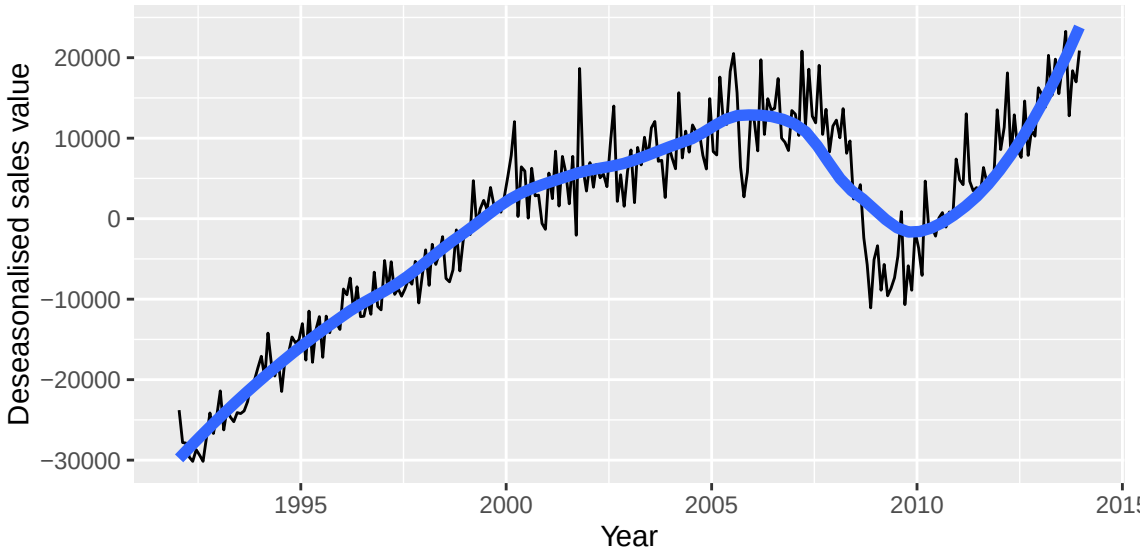
Seasonality

To check seasonality the monthly average sales is plotted against month of the year. The blue curve is a smooth through the data using local weighted regression (loess). The plot shows sales increase around spring to end of summer season. However, the sales decreases starting from fall to winter season.



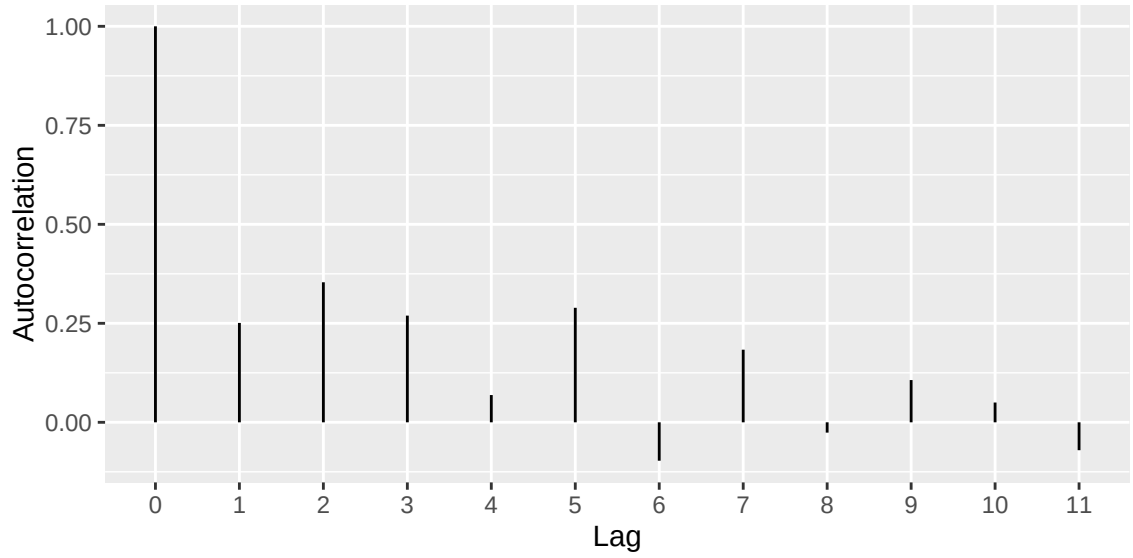
Trend

To get seasonally adjusted sales value, the previous smoothed value is subtracted from the monthly sales resulting in seasonally adjusted sales value. These are plotted along smoothed blue line below. As we can see there is still small monthly variation in sales, with the largest in the order of 10000 in 2003 and 2009. Besides, there is a gradual annual sales increase from 1992 to 2007. However, in year 2007 there is sudden decrease in sales for 3 years till it recovers its annual increase trend in sales from 2010 to 2015.

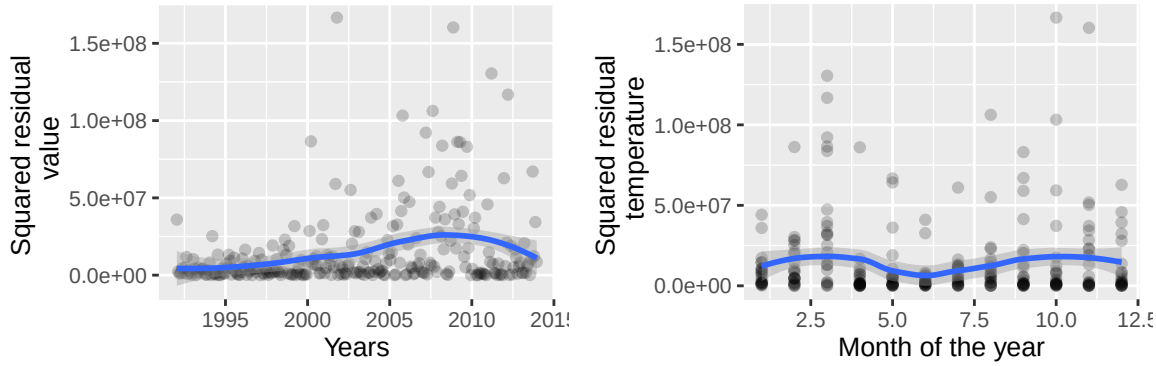


Residual

The autocorrelation function for the estimated residual is show below, where the unit lag is equal to one month. The correlation between two months is low `{r} round(sales_acf$acf[2], 2)`. In general, the monthly sales between longer months is low or with unpredictable correlations.



The stationary of the residual is examined by plotting the squared residual against the month. It is shown, there is a long term trend in sales from 2000 to 2015. Additionally, there is high variance in sales between month of the same and different years.



In summary,

- The greatest source of variation in monthly sales amount.
- Monthly average sales tend to increase from Spring to Winter months and eventually decrease in the cold seasons.
- The annual sales of motor vehicles and part sales tend to increase rapidly with exception of from 2007 to 2010.

Appendix

```
knitr::opts_chunk$set(echo=FALSE, message = FALSE,
  warning = FALSE, results = "hide", fig.height = 3, fig.width = 6)
library(ggplot2)
library(dplyr)
library(lubridate)
options(stringsAsFactors = FALSE)
```

```r
#load(url("http://stat565.cwick.co.nz/data/sales.rda")) #load data
load('sales.rda')
#format date #
sales <- sales %>% mutate(yday=yday(date),year=year(date),month=month(date))
qplot(date,value,data=sales,geom="line") + ylab("Average  sales")+xlab("Year")
qplot(factor(month), value, data = sales, geom = "line", group = year, alpha = I(.3)) +
    geom_smooth(aes(group = 1), method = "loess", se = FALSE, size = 2) +
    ylab("Average monthly sales") +
    xlab("Months")
# fit the seasonal model
lo_fit <- loess(value ~ month, data = sales, na.action = na.exclude)
sales$seasonal_smooth <- fitted(lo_fit)
# subtract off seasonal pattern
sales$deseasonalised <- sales$value - sales$seasonal_smooth
qplot(date, deseasonalised, data = sales, geom = "line") +
    geom_smooth(se = FALSE, method = "loess", span = 0.4, size = 2)+
    ylab("Deseasonalised sales value") +
    xlab("Year")
#Detrend
lo_fit_trend <- loess(deseasonalised ~ as.numeric(date), data = sales, na.action = na.exclude,
                                            span = 0.4)
sales$trend_smooth <- fitted(lo_fit_trend) #fit smoothed data
sales$residual <- sales$deseasonalised - sales$trend_smooth #remove trends
#plot the residual
qplot(date, residual, data = sales, geom  = "line")
# force to all be in order still
sales <- sales[order(sales$date), ]
# correlation of residuals
sales_acf <- acf(sales$residual, na.action = na.pass,lag.max = 11)
# qplotize it, for uniform look.
acf_df <- data.frame(ACF = sales_acf$acf, Lag = sales_acf$lag)
ci <- qnorm((1 + 0.95)/2)/sqrt(sales_acf$n.used)

qplot(factor(Lag), ymin = 0, ymax = ACF, data = acf_df, geom = "linerange")+
    #geom_hline(aes(yintercept = c(-ci, ci)), color = "blue", linetype ="dashed")+
     ylab("Autocorrelation") + xlab("Lag")

qplot(date, residual^2, data = sales, alpha = I(0.2)) +
    geom_smooth(method = "loess") +
    ylab("Squared residual \n value") +
    xlab("Years") +
    theme_grey(10)

qplot(month, residual^2, data = sales, alpha = I(0.2)) +
    geom_smooth(method = "loess") +
    ylab("Squared residual \n temperature") +
    xlab("Month of the year") +
    theme_grey(10)
```