Cross Correlation Analysis: Relationship between Diesel Prices and Commercial Auto Premiums

Bob McPherson

February 24, 2017

# Examine Lags

In this section, we apply the functions above, to examine the possible lags. The time series is offset by one period, and the correlation is tested. This process is repeated, until the maximum number of lags is reached. A p-value is calculated for each correlation, to test the statistical significance. The candidate lag with the most significant p-value is selected as the best lag to use in offsetting the time series data. The smallest p-value corresponds to the most statistically significant value.

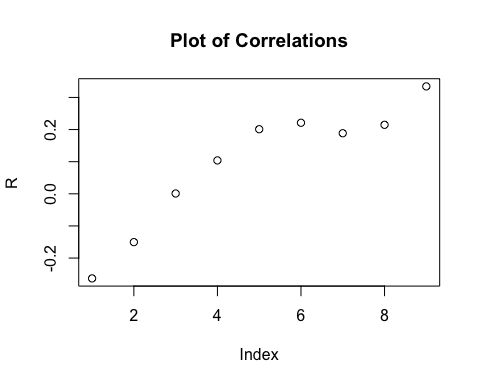
(R <- crosscor(x,y))

## [,1]  
## [1,] -0.2633018789  
## [2,] -0.1503578036  
## [3,] 0.0009792564  
## [4,] 0.1039573783  
## [5,] 0.2010701076  
## [6,] 0.2211319016  
## [7,] 0.1886030674  
## [8,] 0.2147961749  
## [9,] 0.3343344057

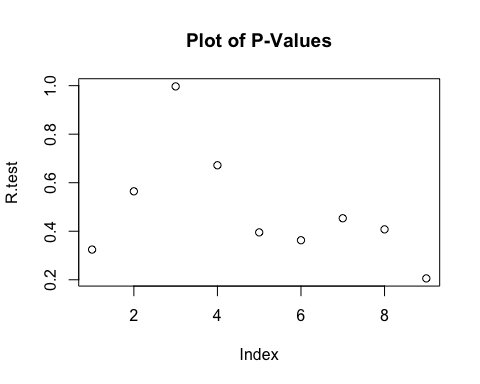
(R.test <- crosscor.test(x,y))

## [,1]  
## [1,] 0.3244735  
## [2,] 0.5646076  
## [3,] 0.9969231  
## [4,] 0.6719106  
## [5,] 0.3953004  
## [6,] 0.3629335  
## [7,] 0.4535587  
## [8,] 0.4077281  
## [9,] 0.2056336

plot(R,main="Plot of Correlations")



plot(R.test,main="Plot of P-Values")



(best.pval.index <- which.min(R.test))

## [1] 9

(best.pval <- R.test[best.pval.index])

## [1] 0.2056336

if(best.pval <= .05){  
 pval.comment <- "This p-value suggests that the best lag is statistically significant at the 95% confidence level."  
} else {  
 pval.comment <- "This p-value suggests that the best lag is not statistically significant at the 95% confidence level."  
 }

# Absolute Correlation Value Comparisons

Instead of comparing the p-values, we can also select the best lag based on comparing the correlation coefficients, or r-values. In this test, we select the largest absolute value of the correlation coefficients. In most cases, the best lag should agree with the value that was chosen using the p-value comparison. This exhibit is merely shown to reveal the correlation coefficients behind the p-values.

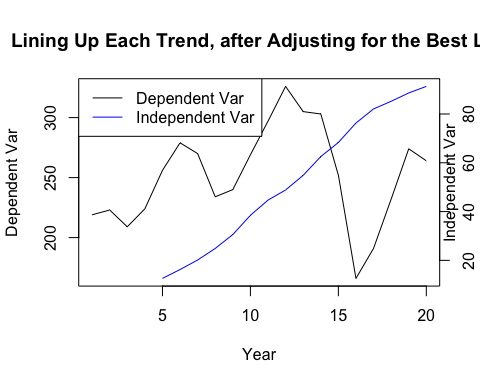
if(abs(min(R))>=max(R)){  
 (max.cor <- which.min(R))  
 direction <- "negative, meaning when one trend goes up, the trend for the other variable tends to move in the opposite direction"  
} else {  
 (max.cor <- which.max(R))  
 direction <- "positive, meaning when one trend goes up or down, the trend for the other variable tends to move in the same direction"  
}  
  
(best.lag <- ((-1\*min.vector):(min.vector))[max.cor])

## [1] 4

best.correlation <- R[max.cor]

The best lag for the data is 4 years. The correlation coefficient for the best lag is 0.3343344. The sign of the correlation coefficient is positive, meaning when one trend goes up or down, the trend for the other variable tends to move in the same direction. The p-value for the best lag is 0.2056336. This p-value suggests that the best lag is not statistically significant at the 95% confidence level.

#plot line graphs with no lag  
#plot(x,type="l")  
#par(new=T)  
#plot(y,type="l")  
#axis(side=4)  
#par(new=F)  
  
##plot the variables after adjusting for the lag  
plot(y,type="l",col=1,xlab="Year",ylab="Dependent Var",main="Lining Up Each Trend, after Adjusting for the Best Lag")  
par(new=T)  
plot(Lag(x,shift=best.lag),type="l",col="blue",xlab='',ylab='',axes=F)  
axis(side=4)  
mtext(side=4,line=0,"Independent Var")  
legend("topleft",legend=c("Dependent Var","Independent Var"),lty=c(1,1),col=c("black","blue"))



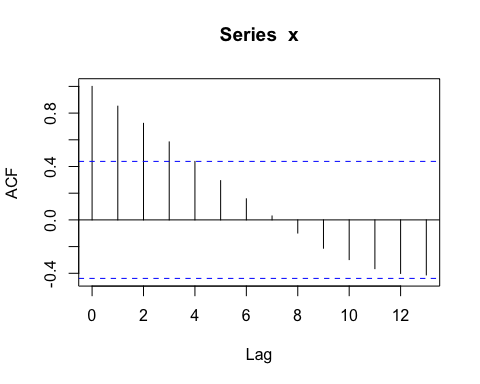
par(new=F)

# Auto Regressive Tendencies and First Order Differencing

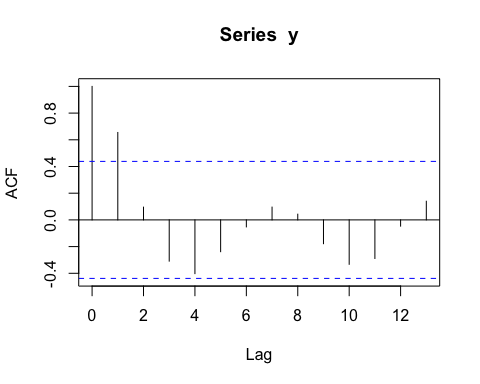
Data from a time series can often have a tendency to be correlated to itself. A common situation that results in this phenomenon is seasonality. This is called, auto regression. The ACF function in R tests for auto regression. We will use this function in this section.

We will also detrend the data using first order differencing. This merely refers to subtracting each precending value from the value that follows.

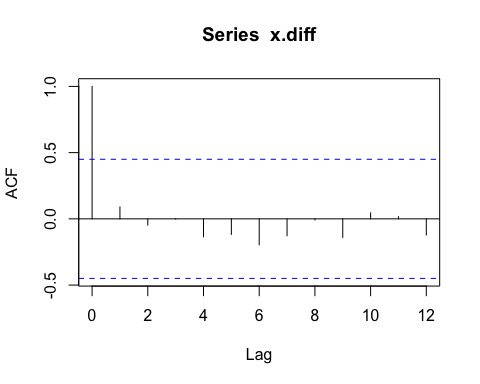
##Rerun with first order differencing  
  
##run ACFs  
acf(x)



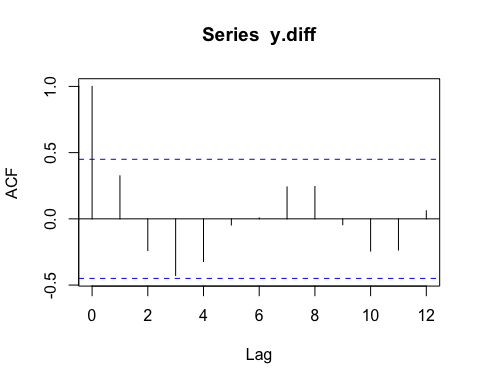
acf(y)



##first order differencing  
x.diff <- diff(x,1)  
y.diff <- diff(y,1)  
  
##ACF on differenced data  
acf(x.diff)



acf(y.diff)

 Next, we run p-value significance tests on the cross correlations of the differenced data. The results are also plotted.

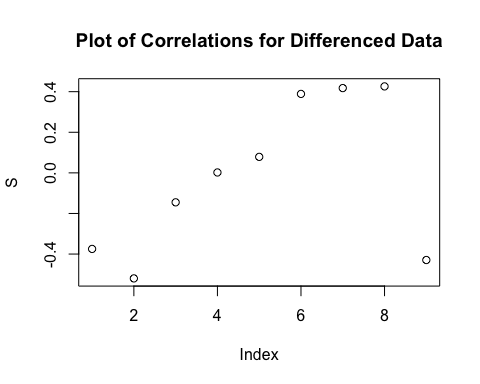
(S <- crosscor(x.diff,y.diff))

## [,1]  
## [1,] -0.37478727  
## [2,] -0.51992571  
## [3,] -0.14516029  
## [4,] 0.00198793  
## [5,] 0.07877634  
## [6,] 0.38894590  
## [7,] 0.41770938  
## [8,] 0.42597898  
## [9,] -0.42931473

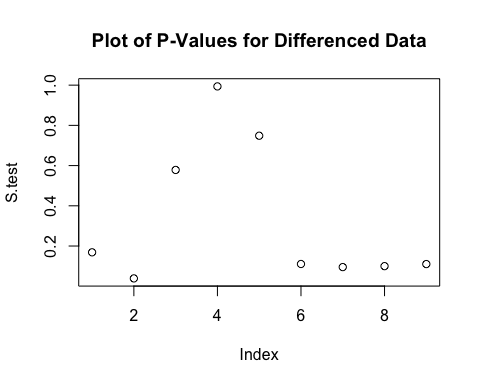
(S.test <- crosscor.test(x.diff,y.diff))

## [,1]  
## [1,] 0.16869374  
## [2,] 0.03898333  
## [3,] 0.57829209  
## [4,] 0.99375380  
## [5,] 0.74854002  
## [6,] 0.11065652  
## [7,] 0.09524272  
## [8,] 0.09993208  
## [9,] 0.11027465

plot(S,main="Plot of Correlations for Differenced Data")



plot(S.test,main="Plot of P-Values for Differenced Data")



if(abs(min(S))>=max(S)){  
 (max.cor <- which.min(S))  
 direction <- "negative, meaning when one trend goes up, the trend for the other variable tends to move in the opposite direction"  
} else {  
 (max.cor <- which.max(S))  
 direction <- "positive, meaning when one trend goes up or down, the trend for the other variable tends to move in the same direction"  
}  
  
best.lag <- ((-1\*min.vector):(min.vector))[max.cor]   
  
best.correlation <- S[max.cor]  
  
(best.pval.index <- which.min(S.test))

## [1] 2

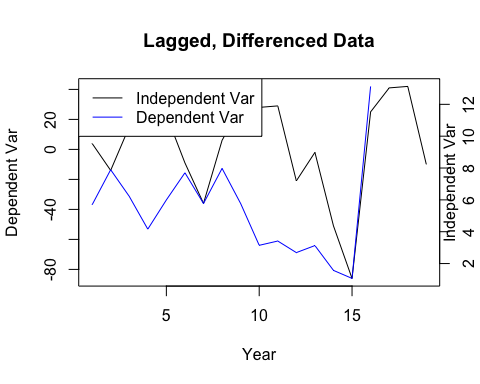
(best.pval <- S.test[best.pval.index])

## [1] 0.03898333

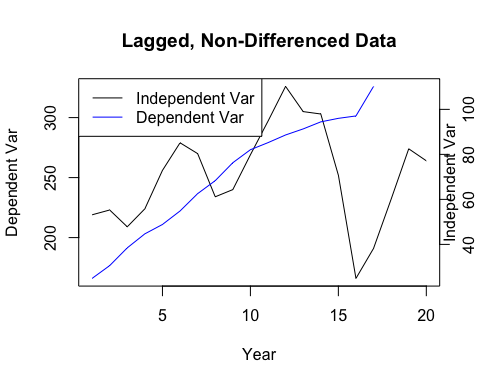
if(best.pval <= .05){  
 pval.comment <- "This p-value suggests that the best lag is statistically significant at the 95% confidence level."  
} else {  
 pval.comment <- "This p-value suggests that the best lag is not statistically significant at the 95% confidence level."  
 }

The best lag for the data is -3 years. The correlation coefficient for the best lag is -0.5199257. The sign of the correlation coefficient is negative, meaning when one trend goes up, the trend for the other variable tends to move in the opposite direction. The p-value for the best lag is 0.0389833. This p-value suggests that the best lag is statistically significant at the 95% confidence level.

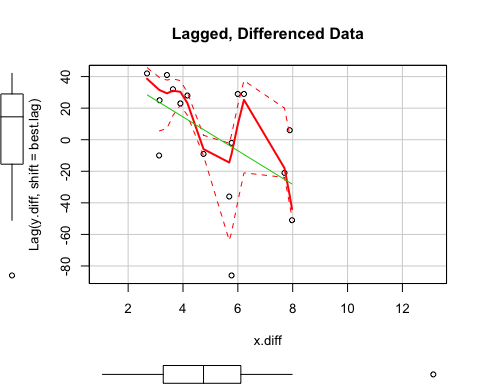
##plot line graph of lagged, differenced data  
plot(y.diff,type="l",col=1,xlab="Year",ylab="Dependent Var",main="Lagged, Differenced Data")  
par(new=T)  
plot(Lag(x.diff,shift=best.lag),type="l",col="blue",xlab='',ylab='',axes=F)  
axis(side=4)  
mtext(side=4,line=0,"Independent Var")  
legend("topleft",legend=c("Independent Var","Dependent Var"),lty=c(1,1),col=c("black","blue"))



par(new=F)  
  
##plot line graph of lagged, NON-differenced data;   
plot(y,type="l",col=1,xlab="Year",ylab="Dependent Var",main="Lagged, Non-Differenced Data")  
par(new=T)  
plot(Lag(x,shift=best.lag),type="l",col="blue",xlab='',ylab='',axes=F)  
axis(side=4)  
mtext(side=4,line=0,"Independent Var")  
legend("topleft",legend=c("Independent Var","Dependent Var"),lty=c(1,1),col=c("black","blue"))



par(new=F)  
  
scatterplot(x.diff,Lag(y.diff,shift=best.lag),main="Lagged, Differenced Data")



scatterplot(x,Lag(y,shift=best.lag),main="Lagged, Non-Differenced Data")

