Chapter 11 - Further Issues in Using OLS with Time Series Data

Thalles Quinaglia Liduares

06/04/2022

Exercise 11.9

Upload packages

```
library(lmreg)
library(wooldridge)
library(tseries)
#library(forecast)
```

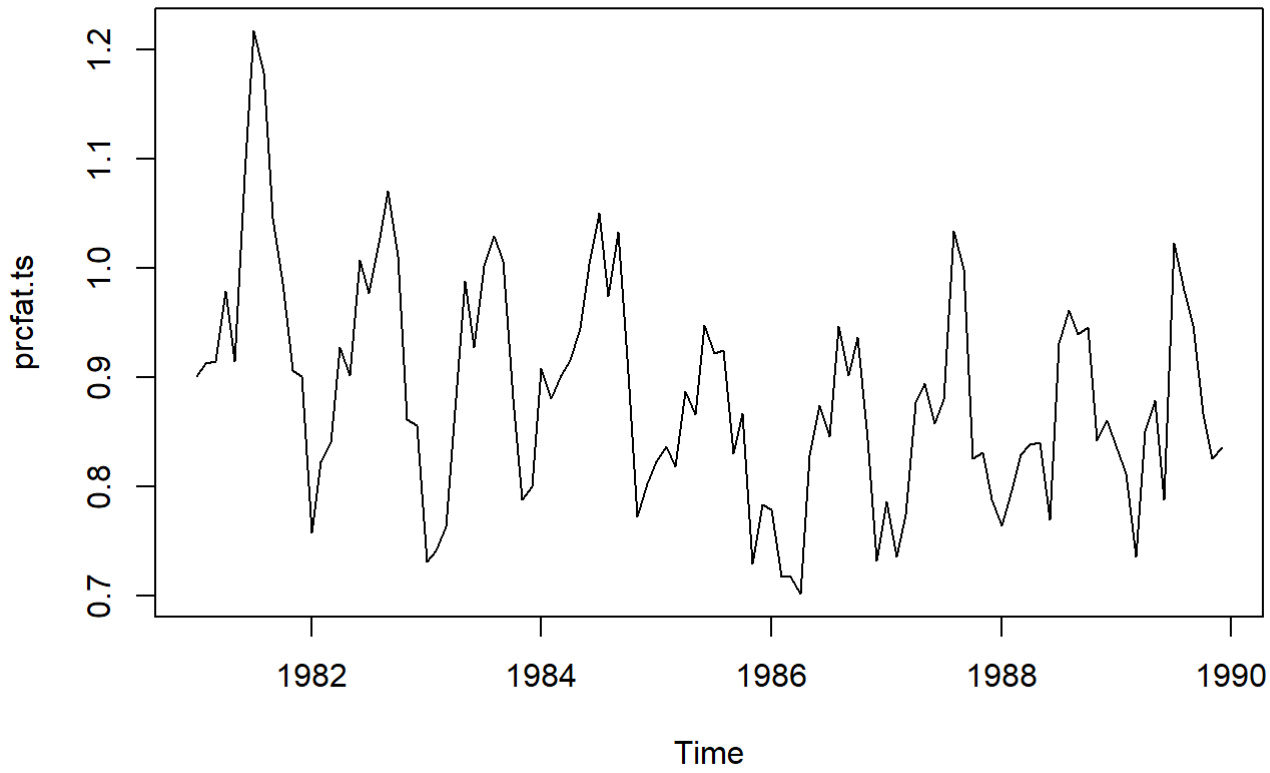
Upload database

```
data<-wooldridge::traffic2

attach(data)
```

Use the data in TRAFFIC2.RAW for this exercise. Computer Exercise C11 in Chapter 10 previously asked for an analysis of these data.

(i) Compute the first order autocorrelation coefficient for the variable `prcfat`. Are you concerned that `prcfat` contains a unit root? Do the same for the unemployment rate.

```
prcfat.ts<-ts(prcfat, start=c(1981,1), end=c(1989,12), frequency = 12)

plot.ts(prcfat.ts)
```

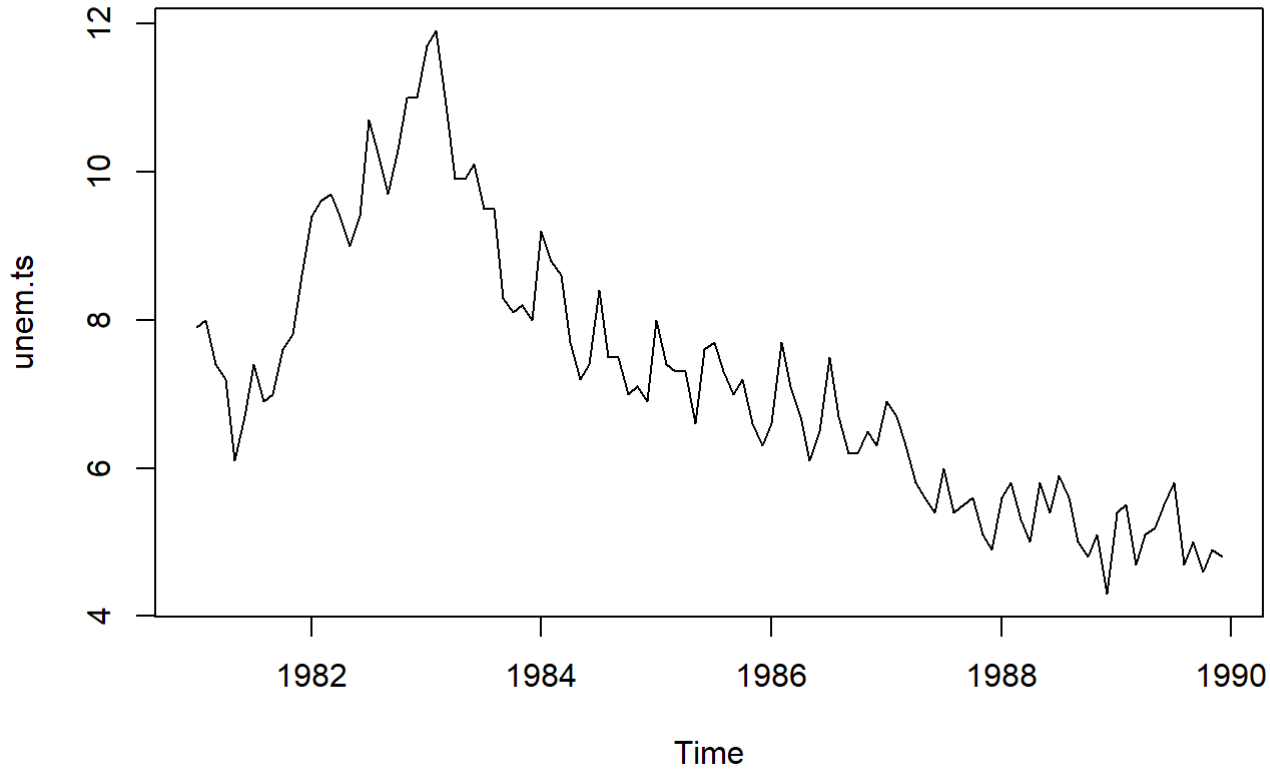



```
acf(prcfat.ts, pl=F, lag=5)
```

```
##
## Autocorrelations of series 'prcfat.ts', by lag
##
## 0.0000 0.0833 0.1667 0.2500 0.3333 0.4167
##  1.000  0.708  0.444  0.180 -0.056 -0.230
```

Based on time serie plot of variable `prcfat` and in the respective ACF function, there's evidence that this variable is not stationary.

```
unem.ts<-ts(unem, start=c(1981,1), end=c(1989,12), frequency = 12)

plot.ts(unem.ts)
```




```
acf(unem, pl=F, lag=5)
```

```
##
## Autocorrelations of series 'unem', by lag
##
##     0     1     2     3     4     5
## 1.000 0.941 0.893 0.860 0.837 0.817
```

Similarly for the variable `unem` .

(ii) Estimate a multiple regression model relating the first difference of prcfat, $\Delta prcfat$, to the same variables in part (vi) of Computer Exercise C11 in Chapter 10, except you should first difference the unemployment rate, too. Then, include a linear time trend, monthly dummy variables, the weekend variable, and the two policy variables; do not difference these. Do you find any interesting results?

```
unem_1<-diff(unem)

unem_1<-append(unem_1, values=NA, after=length(unem_1)) # To equal the length of
                                                        #  unem_1 to the length
                                                        # of other variables

lm1<-lm(prcfat~prcfat_1+unem+unem_1+t+feb+mar+
                  apr+may+jun+jul+aug+sep+oct+nov+dec+wkends+
                  spdlaw+beltlaw)


summary(lm1)
```

```
##
## Call:
## lm(formula = prcfat ~ prcfat_1 + unem + unem_1 + t + feb + mar +
##     apr + may + jun + jul + aug + sep + oct + nov + dec + wkends +
##     spdlaw + beltlaw)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.115521 -0.031755  0.003955  0.030300  0.113623
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7097354  0.1585718   4.476 2.30e-05 ***
## prcfat_1     0.3022481  0.1045733   2.890  0.00486 **
## unem        -0.0116317  0.0057633  -2.018  0.04665 *
## unem_1      -0.0028833  0.0139553  -0.207  0.83679
## t           -0.0016141  0.0004943  -3.266  0.00156 **
## feb          0.0036190  0.0295537   0.122  0.90282
## mar          0.0043734  0.0282046   0.155  0.87713
## apr          0.0621882  0.0284706   2.184  0.03163 *
## may          0.0612395  0.0279746   2.189  0.03127 *
## jun          0.0835468  0.0292800   2.853  0.00540 **
## jul          0.1454946  0.0303758   4.790 6.80e-06 ***
## aug          0.1441164  0.0323517   4.455 2.49e-05 ***
## sep          0.1066706  0.0331442   3.218  0.00181 **
## oct          0.0559355  0.0314546   1.778  0.07885 .
## nov         -0.0137649  0.0294798  -0.467  0.64172
## dec          0.0050779  0.0305185   0.166  0.86824
## wkends       0.0014954  0.0062443   0.239  0.81129
## spdlaw       0.0444481  0.0213836   2.079  0.04060 *
## beltlaw     -0.0187579  0.0228291  -0.822  0.41351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05568 on 87 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.7462, Adjusted R-squared:  0.6937
## F-statistic: 14.21 on 18 and 87 DF,  p-value: < 2.2e-16
```

Based on coefficients associated to months of july and august, seems to have a greater impact on traffic accidents.

(iii) Comment on the following statement: “We should always first difference any time series we suspect of having a unit root before doing multiple regression because it is the safe strategy and should give results similar to using the levels.” [In answering this, you may want to do the regression from part (vi) of Computer Exercise C11 in Chapter 10, if you have not already.]

Non-stationary series may not satisfy the CLM assumptions. Generally, a differentiation process is required to make the series stationary and perform the multiple regression estimations.