Lab 11 - Time Series (summary) and Panel data

Julieth Santamaria

April 19, 2019

Time Series

The dataset we're going to work on has the following variables.

`dw.nsa` = number of discouraged workers, not seasonally adjusted (thousands)

`un.nsa` = number of unemployed workers, not seasonally adjusted (thousands)

`avgdur` = average weeks unemployed

`cs` = U. of Michigan consumer sentiment index (higher means more optimistic)

`dw.un` = `dw.nsa`/`un.nsa`

1. Load the dataset and run the following lines.

```
setwd("G:/My Drive/U of M/TA/TA APEC3003/APEC 3003 - 2019/APEC 3003 R work/labs/")

load("../data/discouraged.rdata")

source("../functions/convenience.r")
source("../functions/summaryHC.r")

library("car")
```

```
## Warning: package 'car' was built under R version 3.5.3
```

```
## Loading required package: carData
```

2. Make a plot of the variable unemp.nsa

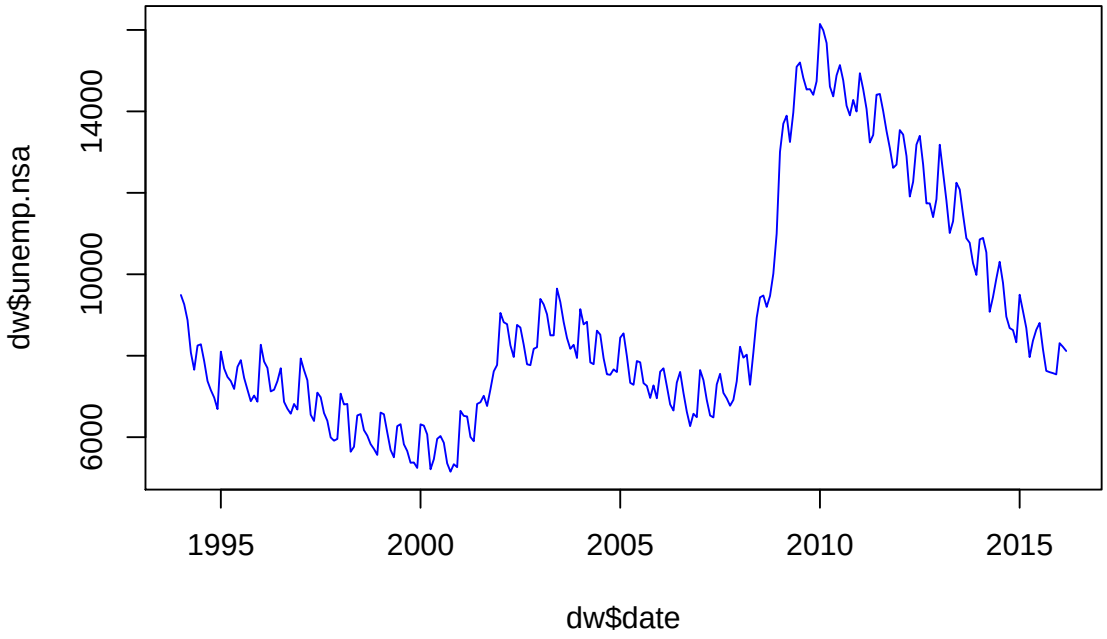
   Recall that the Dicky-Fuller test is:

   $H_0 : \delta = 0 \equiv \rho = 1$ (unit root) $H_1 : \delta < 0 \equiv \rho < 1$ (stationary)

   The critical values are:

   1%: -3.99 5%: -3.43 10%: -3.13

```
plot(dw$date, dw$unemp.nsa, type="l", col="blue")
```



3. Regress unemp.nsa on consumer sentiment (Use two lags) and correct for autocorrelated errors.

```
reg1 <- lm(unemp.nsa~cs+L(cs)+L(cs,2), data=dw)
print(summaryHC(reg1))
```

```
## Warning: package 'sandwich' was built under R version 3.5.3
## Warning: package 'lmtest' was built under R version 3.5.3
## Loading required package: zoo
## Warning: package 'zoo' was built under R version 3.5.3
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Call:
## lm(formula = unemp.nsa ~ cs + L(cs) + L(cs, 2), data = dw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4755.7  -893.5    62.8   949.7  4625.0
##
## Coefficients:
##              Estimate Robust SE t value Pr(>|t|)
## (Intercept) 23874.788   969.163  24.634  < 2e-16 ***
```

```
## cs            -41.725    33.838  -1.233    0.219
## L(cs)           1.137    43.228   0.026    0.979
## L(cs, 2)     -131.344    31.103  -4.223 3.34e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1670 on 261 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6383, Adjusted R-squared:  0.6341
## F-statistic: 102.5 on 3 and 261 DF,  p-value: < 2.2e-16
```

4. Test the joint significance of your coefficients

```
print(linearHypothesis(reg1, c("cs","L(cs)","L(cs, 2)"), vcov=vcovHAC))
```

```
## Linear hypothesis test
##
## Hypothesis:
## cs = 0
## L(cs) = 0
## L(cs, 2) = 0
##
## Model 1: restricted model
## Model 2: unemp.nsa ~ cs + L(cs) + L(cs, 2)
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1    264
## 2    261  3 42.711 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Panel Data

Panel Data characteristics:

- They have cross-sectional and time-series dimensions
- Contains, N individuals observed in T periods
- A *balanced panel* has observations of individuals in all T periods. An *unbalanced panel* contains information of individuals in some of the T periods.

You will see three types of regressors:

- Varying regressors x_{it}: Annual income of a person, monthly consumption of a person
- Time-invariant regressors x_i: Sex, race, education
- Individual-invariant regressors x_t: economy trends, unemployment rate.

Models can estimate:

- Overall variation: Variation across individuals over time
- Between variation: Variation between individuals
- **Within variation**: Variation within individuals (over time)

Note: 1. Time-invariant regressors have zero within variation. 2. Individual-invariat regressors have zero between variation.

Pooled Model: Assumes constant coefficients. $y_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_{it}$

Fixed effects model: Captures "unobserved heterogeneity" $y_{it} = \beta_0 + \alpha_i + \beta_1 X_{it} + \varepsilon_{it}$

- De-meaned approach: $y_{it} - \bar{y_i} = \beta_1(x_{it} - \bar{x}_i) + v_{it}$

First difference estimator: There are any regressors that don't vary over time.

Difference in differences approach

Main assumption of these models is the **Parallel trends assumption**

Application

PSID example taken from the Econometrics Academy webpage.

1. Load the data and set the dataset as panel

```
library(plm)
```

```
## Warning: package 'plm' was built under R version 3.5.3
```

```
## Loading required package: Formula
```

```
setwd("G:/My Drive/U of M/TA/TA APEC3003/APEC 3003 - 2019/APEC 3003 R work/labs/")
load("../data/panelwage.Rdata")
panel_wage <- pdata.frame(panel_wage, index=c("id","t"))
```

2. Estimate the effect of experience using a pooled model

```
# Pooled OLS estimator
pooled <- plm(lwage ~ exp+exp2+wks+ed, data=panel_wage, model= "pooling")
summary(pooled)
```

```
## Pooling Model
##
## Call:
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = panel_wage,
##     model = "pooling")
##
## Balanced Panel: n = 595, T = 7, N = 4165
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -2.16057670 -0.25034526  0.00027256  0.26792139  2.12969386
##
## Coefficients:
##                Estimate  Std. Error  t-value  Pr(>|t|)
## (Intercept)  4.9080e+00  6.7330e-02  72.8945 < 2.2e-16 ***
## exp          4.4675e-02  2.3929e-03  18.6701 < 2.2e-16 ***
## exp2        -7.1563e-04  5.2794e-05 -13.5552 < 2.2e-16 ***
## wks          5.8270e-03  1.1826e-03   4.9271 8.673e-07 ***
## ed           7.6041e-02  2.2266e-03  34.1511 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    886.9
## Residual Sum of Squares: 635.41
## R-Squared:      0.28356
## Adj. R-Squared: 0.28287
## F-statistic: 411.624 on 4 and 4160 DF, p-value: < 2.22e-16
```

3. How do your results change using fixed effects?

```
fixed <- plm(lwage ~ exp+exp2+wks+ed, data=panel_wage, model= "within")
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = panel_wage,
##     model = "within")
##
## Balanced Panel: n = 595, T = 7, N = 4165
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -1.8120879 -0.0511128  0.0037112  0.0614250  1.9434065
##
## Coefficients:
##         Estimate  Std. Error t-value  Pr(>|t|)
## exp   1.1379e-01  2.4689e-03 46.0888 < 2.2e-16 ***
## exp2 -4.2437e-04  5.4632e-05 -7.7678 1.036e-14 ***
## wks   8.3588e-04  5.9967e-04  1.3939    0.1634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    240.65
## Residual Sum of Squares: 82.632
## R-Squared:      0.65663
## Adj. R-Squared: 0.59916
## F-statistic: 2273.74 on 3 and 3567 DF, p-value: < 2.22e-16
```

4. Estimate now the first differences estimator

```
firstdiff <- plm(lwage ~ exp+exp2+wks+ed, data=panel_wage, model= "fd")
summary(firstdiff)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = panel_wage,
##     model = "fd")
##
## Balanced Panel: n = 595, T = 7, N = 4165
## Observations used in estimation: 3570
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -2.1131555 -0.0654718 -0.0095751  0.0483881  2.3295637
##
## Coefficients:
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept)  0.11706540  0.00631057 18.5507 < 2.2e-16 ***
## exp2        -0.00053212  0.00013927 -3.8207 0.0001354 ***
## wks         -0.00026826  0.00056483 -0.4749 0.6348525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Total Sum of Squares:    118.06
## Residual Sum of Squares: 117.58
## R-Squared:      0.004108
## Adj. R-Squared: 0.0035496
## F-statistic: 7.35691 on 2 and 3567 DF, p-value: 0.0006479
```