Assignment 4

Tai Lo Yeung, Pan Yiming, Qiwei Liu

Barcelona Graduate School of Economics

Universitat Pompeu Fabra

Nov 20, 2020

Panel Data Example

a

Estimate the equation by pooled OLS using l_roe and l_e_p as the third explanatory variable (X_{it}) (so estimate the equation twice). Is there evidence of return predictability? What variables seem to have information about the bank's subsequent return?

Using l_roe:

```
Pooling Model

Call:
plm(formula = ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("pooling"))

Unbalanced Panel: n = 866, T = 1-94, N = 32743

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.3210765 -0.0842020 -0.0076864  0.0790847  2.0946404

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept) -0.0172221  0.0019231 -8.9554  < 2e-16 ***
l_ret       -0.0093574  0.0054031 -1.7319  0.08331 .
l_btm        0.0226434  0.0014272 15.8659  < 2e-16 ***
l_roe        0.7527838  0.0218903 34.3890  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    897.21
Residual Sum of Squares: 865.29
R-Squared:      0.035572
Adj. R-Squared: 0.035483
F-statistic: 402.512 on 3 and 32739 DF, p-value: < 2.22e-16
```

Using l_e_p:

```
Pooling Model

Call:
plm(formula = ret ~ l_ret + l_btm + l_e_p, data = paneldf, model = c("pooling"))

Unbalanced Panel: n = 866, T = 1-94, N = 32743

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.8906927 -0.0832014 -0.0075331  0.0786108  2.0779823

Coefficients:
              Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  0.0055757  0.0015717  3.5475 0.0003894 ***
l_ret       -0.0053487  0.0053971 -0.9910 0.3216813
l_btm        0.0194750  0.0014026 13.8852 < 2.2e-16 ***
l_e_p        0.5200984  0.0154009 33.7707 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    897.21
Residual Sum of Squares: 866.37
R-Squared:      0.034372
Adj. R-Squared: 0.034284
F-statistic: 388.454 on 3 and 32739 DF, p-value: < 2.22e-16
```

There in no evidence of return predictability because the lagged return variable is not significance in both tests, and the Adjusted R-squared are both too low. The l_btm, l_e_p, and l_roe both have information about the bank's subsequent return.

b

Estimate now the model (from now on, choose l_roe as X_{it}) again as pooled OLS but where you:
1. Include bank specific intercepts (bank dummies) and test for the joint significance of the effects. Remember that these estimates are inconsistent, but we do not want to worry much about that. Also, you may want to be careful here, since there are lots of banks, so if you use Stata the program might ask you to increase the memory allocated to the analysis.
2. Include time effects but not bank effects (i.e. include a dummy for each period of time). Test for the joint significance of the time effects.

c

Estimate now the model as a random-effects panel and a fixed-effects panel, with no time effects. Check that the point estimates of β_1, β_2 and β_3 in fixed effects are the same as those in b-1. Do the conclusions on predictability change now?

d

Perform the Breusch-Pagan test for random effects versus no random effects. What is the conclusion of the test?

e

Do the Hausman test of fixed effects (always consistent) versus random effects (efficient under random individual heterogeneity, inconsistent otherwise). Interpret the outcome of the test.

f

Given the analysis above, offer a “story” for your findings. That is, if you did find some predictability, are there any interesting theoretical arguments for such predictability?

code

```
library(sandwich)
library(ggplot2)
library(lmtest)
library(zoo)
library(corrplot)
library(car)

# 1

# a
gasoline[,1] <- gasoline[,1] - 1960
gasoline[,2] <- gasoline[,2]/gasoline[,11]

reg_g <- lm(g ~ year + pg + y + pnc + puc + ppt + pd + pn + ps, data = gasoline)
summary(reg_g)

# b
H_0 <- c('pnc = puc')
linearHypothesis(reg_g, H_0)

# c

ela_pg <- coef(reg_g)[3] * mean(gasoline$pg)/mean(gasoline$g)
ela_y <- coef(reg_g)[4] * mean(gasoline$y)/mean(gasoline$g)
ela_ppt <- coef(reg_g)[7] * mean(gasoline$ppt)/mean(gasoline$g)


# d
log_g <- lm(g ~ log(pg) + year + log(y) + log(pnc) + log(puc) + log(ppt) + log(pd) + log(pn) + log(ps
summary(log_g)

# 2

# a
reg_va <- nls(valueadd ~ (alpha * labor^beta * capital^gama), data = Production, start=list(alpha=1.5
summary(reg_va)

# b
log_va <- lm(log(valueadd) ~ log(labor) + log(capital), data = Production)
summary(log_va)

# c
```

```
H_1 = c('log(labor) = 0.6')
H_2 = c('log(labor) + log(capital) = 1')
H_3 = c('log(labor) = 0.6', 'log(labor) + log(capital) = 1')

linearHypothesis(log_va, H_1)
linearHypothesis(log_va, H_2)
linearHypothesis(log_va, H_3)

# d
lny <- log(Production$valueadd)
lnl <- log(Production$labor)
lnk <- log(Production$capital)
lnl2 <- 0.5 * lnl^2
lnk2 <- 0.5 * lnk^2

trans_log <- lm(lny ~ lnl + lnk + lnl2 + lnk2 + lnl * lnk)
summary(trans_log)

library(ggfortify)
autoplot(trans_log)

resettest(trans_log, power = 3, type = 'fitted')
resettest(trans_log, power = 3, type = 'regressor')
```