Panal_data

1 Qa solution:

Estimate the equation by pooled OLS using l_roe and l_e_p as the third explanatory variable. Is there evidence of return predictability? What variables seem to have information about the bank's subsequent return?

```
panel_fin = read.dta13('./panel_fin.dta')
paneldf <- pdata.frame(panel_fin, index = c('entity', 'quarter'))
# Qa 1:
pd_roe <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("pooling"))
summary(pd_roe)

# Qa 2:
pd_lep <- plm(ret ~ l_ret + l_btm + l_e_p, data = paneldf, model = c("pooling"))
summary(pd_lep)
```

1.1 Qa 1 output: pooled OLS 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
>
```

1.2 Qa 2 output: pooled OLS 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 have information about the bank’s subsequent return.

2 Qb solution:

Estimate now the model (from now on, choose l roe as Xit) again as pooled OLS but where you:

2.1 using bank dummy variables

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.

```
# Qb1:
## add dummy variable for every bank:

paneldf_bank_dummy <- dummy_cols(paneldf,select_columns='entity',remove_first_dummy = TRUE)
#paneldf_bank_dummy <- as.data.frame(paneldf_bank_dummy)
## get the formula:
formula_b1 = 'ret ~ l_ret + l_btm + l_roe'
i=1
while (i < 868) {
  i = i + 1
  formula_b1 <- paste(formula_b1, '+entity_', i, sep = "")

}
roe_bank_dummy <- plm(formula_b1, data = paneldf_bank_dummy, model = c("pooling"), index = c("entity","quarter"))
## test for the bank dummy
```

```
Hb10 = c()
i=1
while (i < 868) {
  i = i + 1
  test_i <- paste('entity_', i, '=0',sep = "")
  Hb10 <- append(Hb10, test_i)

}
test_resb1 = linearHypothesis(roe_bank_dummy, Hb10)
test_resb1
```

output:

```
  Res.Df  Df  Chisq Pr(>Chisq)
33128
32261 867 977.78   0.005077 **
```

Joint significance of the effects of individual fixed effect is low, which is 0.5% and the F-value is 977.78. So we can reject the null-hypothesis. So there exits the individual fixed effect.

2.2 using time dummy variables

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.

```
# Qb2:
## add dummy variable for every time peroid
paneldf_time_dummy <- dummy_cols(paneldf,select_columns='quarter',remove_first_dummy = TRUE)
## get the formula:
formula_b2 = 'ret ~ l_ret + l_btm + l_roe'
i=1
while (i < 94) {
  i = i + 1
  formula_b2 <- paste(formula_b2, '+quarter_', i, sep = "")

}
roe_time_dummy <- plm(formula_b2, data = paneldf_time_dummy, model = c("pooling"), index = c("entity","quarter"))
## test for the bank dummy
Hb20 = c()
i=1
while (i < 94) {
  i = i + 1
  test_i <- paste('quarter_', i, '=0',sep = "")
  Hb20 <- append(Hb20, test_i)
}

test_resb2 = linearHypothesis(roe_time_dummy, Hb20)
test_resb2
```

output

```
  Res.Df Df Chisq Pr(>Chisq)
32739
32646 93 12696  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The effect of the time trend is significant, from the F-test , we can see the F-value is 12696 and P-value is very low. So we can say that there exits very abvious individual variation in different time peroid.

3 Qc solotion:

Estimate now the model as a random-effects panel and a fixed-effects panel, with no time effects. Check that the point estimates of β1, β2 and β3 in fixed effects are the same as those in b-1. Do the conclusions on predictability change now?

```
# Qc:
pd_roe_rand <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("random"))
pd_roe_fixed <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("within"))
```

output:

estimates of β1, β2 and β3 in fixed effects are the same as those in b-1. The coefficients of l_ret is siginificantly different from 0, so the l_ret has information about the bank's subsequent return. But the fixed-effect model' R2 is also very low. So the predictability does not change a lot.

```

> pd_roe_fixed

Model Formula: ret ~ l_ret + l_btm + l_roe

Coefficients:
    l_ret      l_btm      l_roe
-0.022614  0.029920  0.729007
> summary(roe_bank_dummy)
Call:
plm(formula = ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("within"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.2497745 -0.0844652 -0.0062537  0.0798545  2.0614517

Coefficients:
        Estimate Std. Error t-value  Pr(>|t|)
l_ret -0.0226142  0.0054786 -4.1277 3.673e-05 ***
l_btm  0.0299200  0.0016015 18.6827 < 2.2e-16 ***
l_roe  0.7290068  0.0235140 31.0031 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    867.35
Residual Sum of Squares: 838.98
R-Squared:      0.032714
Adj. R-Squared: 0.0063723
F-statistic: 359.327 on 3 and 31874 DF, p-value: < 2.22e-16
```

4 Qd solotion:

From the fixed-effect model, we can see that l_ret, l_btm and l_roe all have effect on the bank's return, and l_roe has the most siginificant effect, the coefficient is 0.729, which means the closed relationship between the return on equity and the return of the bank.

```r
#All the code
rm(list = ls())
setwd("Desktop/courses_term1/econometrics/problem_sets/assignment4/")
#install.packages('plm')
library(zoo)
library(sandwich)
library(lmtest)
library(plm)
library(car)
library(readstata13)
library(fastDummies)

panel_fin = read.dta13('./panel_fin.dta')
#panel_fin = read.csv('./panel_fin.csv')
paneldf <- pdata.frame(panel_fin, index = c('entity', 'quarter'))

# Qa 1:
pd_roe <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("pooling"))
summary(pd_roe)

# Qa 2:
pd_lep <- plm(ret ~ l_ret + l_btm + l_e_p, data = paneldf, model = c("pooling"))
summary(pd_lep)

# Qb1:
## add dummy variable for every bank:

paneldf_bank_dummy <- dummy_cols(paneldf,select_columns='entity',remove_first_dummy = TRUE)
#paneldf_bank_dummy <- as.data.frame(paneldf_bank_dummy)
## get the formula:
formula_b1 = 'ret ~ l_ret + l_btm + l_roe'
i=1
while (i < 868) {
  i = i + 1
  formula_b1 <- paste(formula_b1, '+entity_', i, sep = "")

}
roe_bank_dummy <- plm(formula_b1, data = paneldf_bank_dummy, model = c("pooling"), index = c("entity","quarter"))
## test for the bank dummy
Hb10 = c()
i=1
while (i < 868) {
  i = i + 1
  test_i <- paste('entity_', i, '=0',sep = "")
  Hb10 <- append(Hb10, test_i)

}
test_resb1 = linearHypothesis(roe_bank_dummy, Hb10)
test_resb1



# Qb2:
## add dummy variable for every time peroid
```

```r
paneldf_time_dummy <- dummy_cols(paneldf,select_columns='quarter',remove_first_dummy = TRUE)
## get the formula:
formula_b2 = 'ret ~ l_ret + l_btm + l_roe'
i=1
while (i < 94) {
  i = i + 1
  formula_b2 <- paste(formula_b2, '+quarter_', i, sep = "")

}
roe_time_dummy <- plm(formula_b2, data = paneldf_time_dummy, model = c("pooling"), index = c("entity","quarter"))
## test for the bank dummy
Hb20 = c()
i=1
while (i < 94) {
  i = i + 1
  test_i <- paste('quarter_', i, '=0',sep = "")
  Hb20 <- append(Hb20, test_i)
}

test_resb2 = linearHypothesis(roe_time_dummy, Hb20)
test_resb2


# Qc:
pd_roe_rand <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("random"))
pd_roe_fixed <- plm(ret ~ l_ret + l_btm + l_roe, data = paneldf, model = c("within"))

plmtest(pd_roe_fixed, effect = c("individual"), type = c( "bp"))

```