EXCERSISES IN APPLIED PANEL DATA ANALYSIS #3

CHRISTOPHER F. PARMETER

1. Introduction

This R example will introduce you to estimation of the unobserved effects model under the fixed effects framework. Recall that the fixed effects framework allows for the covariates in the model to be uncorrelated with the unobserved effects. We will be using the plm library exclusively to discuss two priminent applied examples.

2. Estimating the Unoberved Effects Model Under the Fixed Effects Framework

2.1. **Public Capital Productivity Puzzle.** Munnell (1990) and Baltagi & Pinnoi (1995) use a balanced panel to estimate the impact of state level public and private capital stocks on state level output. Munnell's (1990) work ignored the panel structure of the data and found a statistically significant positive effect of public capital on state output levels. Baltagi & Pinnoi (1995) was one of the first papers to provide an in-depth look at the problem using a range of panel data methods. Baltagi & Pinnoi's (1995) seminal finding of a negative coefficient on public capital has spawned a cottage industry of papers seeking to explain why we might obtain a negative (or statistically equivalent to 0) estimate. For our purposes we will focus exclusively in this exercise on the within estimator and the models in Baltagi & Pinnoi (1995).

```
> library(plm)
> data("Produc")
> pubcap.data <- Produc
> ## Replicate models in Baltagi and Pinnoi (1995, Table 1.)
> ## We will estimate both the pooled models and
> ## the within transformation
>
> ## Take logs of pcap, hwy, water, util, pc, gsp and emp
> pubcap.data$lpubc <- log(pubcap.data$pcap)
> pubcap.data$lhwy  <- log(pubcap.data$hwy)
> pubcap.data$lwatr <- log(pubcap.data$water)
> pubcap.data$lutil <- log(pubcap.data$util)
```

University of Miami
Date: September 10, 2013.
Christopher F. Parmeter, Department of Economics, University of Miami; e-mail: cparmeter@bus.miami.edu.
*These exercises have been prepared for the "Applied Panel Data Econometrics" course in Dakar, Senegal sponsored by IFPRI and AGRODEP..

```
> pubcap.data$lprvc <- log(pubcap.data$pc)
> pubcap.data$lgsp  <- log(pubcap.data$gsp)
> pubcap.data$lemp  <- log(pubcap.data$emp)
> ## Pooled OLS, no public capital,
> ## Cobb-Douglas specification
> model.pool1 <- plm(lgsp~lprvc+lemp+unemp,
+                    data=pubcap.data,model="pooling")
> ## Pooled OLS, aggregate public capital,
> ##Cobb-Douglas specification
> model.pool2 <- plm(lgsp~lprvc+lemp+lpubc+unemp,
+                    data=pubcap.data,model="pooling")
> ## Pooled OLS, decomposed public capital,
> ## Cobb-Douglas specification
> model.pool3 <- plm(lgsp~lprvc+lemp+lhwy+lwatr+lutil+unemp,
+                    data=pubcap.data,model="pooling")
> ## Fixed Effects Framework, no public capital,
> ## Cobb-Douglas specification
> model.wn1 <- plm(lgsp~lprvc+lemp+unemp,
+                  data=pubcap.data,
+                  model="within",effect="individual")
> ## Fixed Effects Framework, aggregate public capital,
> ## Cobb-Douglas specification
> model.wn2 <- plm(lgsp~lprvc+lemp+lpubc+unemp,
+                  data=pubcap.data,
+                  model="within",effect="individual")
> ## Fixed Effects Framework, decomposed public capital,
> ## Cobb-Douglas specification
> model.wn3 <- plm(lgsp~lprvc+lemp+lhwy+lwatr+lutil+unemp,
+                  data=pubcap.data,
+                  model="within",effect="individual")
```

We can see that our estimates align with those in Baltagi & Pinnoi (1995, Table 1) and the corresponding estimates in Munnell (1990, Tables 4-6). When we ignored the panel structure of the data we have a positive and significant impact of public capital on state output. When we account for unobserved state leel effects under the fixed effects transformation, the model that includes aggregate public capital, `model.wn2` has a negative and insignificant estimate whereas the disaggregated model, `model.wn3`, two of the three components have positive and significant impacts on state output, while a third, has the largest (in magnitude) effect, but is negative and statistically significant.

```
> sum.pool1 <- summary(model.pool1)
> sum.pool2 <- summary(model.pool2)
> sum.pool3 <- summary(model.pool3)
> sum.wn1 <- summary(model.wn1)
> sum.wn2 <- summary(model.wn2)
> sum.wn3 <- summary(model.wn3)
> ## Print out Coefficient estimates and t-statistics
> sum.pool1$coefficients[,c(1,3)]
                Estimate   t-value
(Intercept)  1.947921847 39.791513
lprvc        0.355413109 38.054099
lemp         0.694710029 82.424135
unemp       -0.006070062 -4.092927
> sum.pool2$coefficients[,c(1,3)]
                Estimate   t-value
(Intercept)  1.643302263 28.535869
lprvc        0.309190167 30.100327
lemp         0.593934898 43.203240
lpubc        0.155007005  9.036324
unemp       -0.006732976 -4.753664
> sum.pool3$coefficients[,c(1,3)]
                Estimate    t-value
(Intercept)  1.926004375 36.6835744
lprvc        0.312023086 28.1418784
lemp         0.549695456 35.3800440
lhwy         0.058881719  3.8206480
lwatr        0.118580557  9.5965597
lutil        0.008555123  0.6924966
unemp       -0.007270503 -5.2546498
> sum.wn1$coefficients[,c(1,3)]
          Estimate   t-value
lprvc  0.287906487 11.655390
lemp   0.756453250 27.868542
unemp -0.005652743 -6.233435
> sum.wn2$coefficients[,c(1,3)]
          Estimate    t-value
lprvc  0.292006925 11.6246309
lemp   0.768159473 25.5272539
```

```
lpubc -0.026149654 -0.9016632
unemp -0.005297741 -5.3581508
> sum.wn3$coefficients[,c(1,3)]
        Estimate   t-value
lprvc  0.23503554  8.966114
lemp   0.80112516 26.922978
lhwy   0.07675379  2.456711
lwatr  0.07868485  5.244764
lutil -0.11477816 -6.325128
unemp -0.00517948 -5.287121
```

2.2. **The Impact of Unions on Wages.** Vella & Verbeek (1998) estimate a dynamic model of wage determination to determine the impact that unionization has on wages. They propose a complicated two-stage model which involves a high level interaction type between the unobserved worker fixed effects and union status. For our purposes we will replicate their baseline regression results using both pooled OLS and the within estimator. The data of Vella & Verbeek (1998) were obtained from the *Journal of Applied Econometrics*' data archive.

```
> data <- read.table("VV-DATA.DAT",h=F)
> names(data) <- c("NR", "YEAR", "AG", "BLACK", "BUS", "CON",
+                  "ENT", "EXPER", "FIN", "HISP", "HLTH",
+                  "HOURS", "MAN", "MAR", "MIN", "NC", "NE",
+                  "OCC1", "OCC2", "OCC3", "OCC4", "OCC5",
+                  "OCC6", "OCC7", "OCC8", "OCC9", "PER",
+                  "PRO", "PUB", "RUR", "SOUTH", "SCHOOL",
+                  "TRA", "TRAD", "UNION", "WAGE")
> ## Replicate columns 1-4 in Vella and Verbeek (1998, Table III)
>
> model1 <- lm(WAGE~UNION+SCHOOL+EXPER+I(EXPER^2)+HISP+
+                BLACK+RUR+MAR+HLTH+AG+MIN+CON+TRAD+
+                TRA+FIN+BUS+PER+ENT+MAN+PRO+SOUTH+NC+
+                NE+factor(YEAR),
+              data=data)
> model2 <- lm(WAGE~UNION+SCHOOL+EXPER+I(EXPER^2)+HISP+
+                BLACK+RUR+MAR+HLTH+AG+MIN+CON+TRAD+
+                TRA+FIN+BUS+PER+ENT+MAN+PRO+SOUTH+NC+
+                NE+factor(YEAR)+OCC1+OCC2+OCC3+OCC4+
+                OCC5+OCC6+OCC7+OCC8,,
+              data=data)
> ## Construct pdata.frame
```

```
> pdata <- pdata.frame(data,index=c("NR","YEAR"))
> model3 <- plm(WAGE~UNION+EXPER+I(EXPER^2)+
+                 RUR+MAR+HLTH+AG+MIN+CON+TRAD+
+                 TRA+FIN+BUS+PER+ENT+MAN+PRO+
+                 SOUTH+NC+NE,
+               data=pdata,
+               effect="individual",model="within")
> model4 <- plm(WAGE~UNION+EXPER+I(EXPER^2)+
+                 RUR+MAR+HLTH+AG+MIN+CON+TRAD+
+                 TRA+FIN+BUS+PER+ENT+MAN+PRO+
+                 SOUTH+NC+NE+OCC1+OCC2+OCC3+
+                 OCC4+OCC5+OCC6+OCC7+OCC8,
+               data=pdata,
+               effect="individual",model="within")
```

```
> sum.pool1 <- summary(model1)
> sum.pool2 <- summary(model2)
> sum.wn1 <- summary(model3)
> sum.wn2 <- summary(model4)
> ## Print out Coefficient estimates and t-statistics
> sum.pool1$coefficients[1:10,c(1,3)]
```

```
               Estimate   t value
(Intercept)  0.31976677  3.633711
UNION        0.14754101  8.723342
SCHOOL       0.08426294 16.443377
EXPER        0.05885123  4.477087
I(EXPER^2)  -0.00185478 -2.362501
HISP        -0.05850864 -2.668268
BLACK       -0.15009480 -6.521455
RUR         -0.12897143 -6.968970
MAR          0.11003118  7.200727
HLTH        -0.05484089 -1.016804
```

```
> sum.pool2$coefficients[1:10,c(1,3)]
```

```
                Estimate    t value
(Intercept)  0.388128267  4.3345798
UNION        0.177360361 10.3696552
SCHOOL       0.073067253 13.8181392
EXPER        0.056718032  4.3281191
I(EXPER^2)  -0.001782778 -2.2784044
```

```
HISP        -0.046866260 -2.1397990
BLACK       -0.126224733 -5.4774513
RUR         -0.114442701 -6.1740513
MAR          0.101677601  6.6708853
HLTH        -0.032117863 -0.5999495
> sum.wn1$coefficients[1:6,c(1,3)]
               Estimate    t-value
UNION       0.079345882  4.0823227
EXPER       0.112461265 13.2690197
I(EXPER^2) -0.004145225 -6.8295368
RUR         0.050129571  1.7296917
MAR         0.039772637  2.1780267
HLTH       -0.016687395 -0.3545404
> sum.wn2$coefficients[1:6,c(1,3)]
               Estimate    t-value
UNION       0.080371240  4.1203989
EXPER       0.110943740 12.9354448
I(EXPER^2) -0.004081806 -6.6870488
RUR         0.047845256  1.6491513
MAR         0.038268379  2.0921020
HLTH       -0.010024233 -0.2125491
```

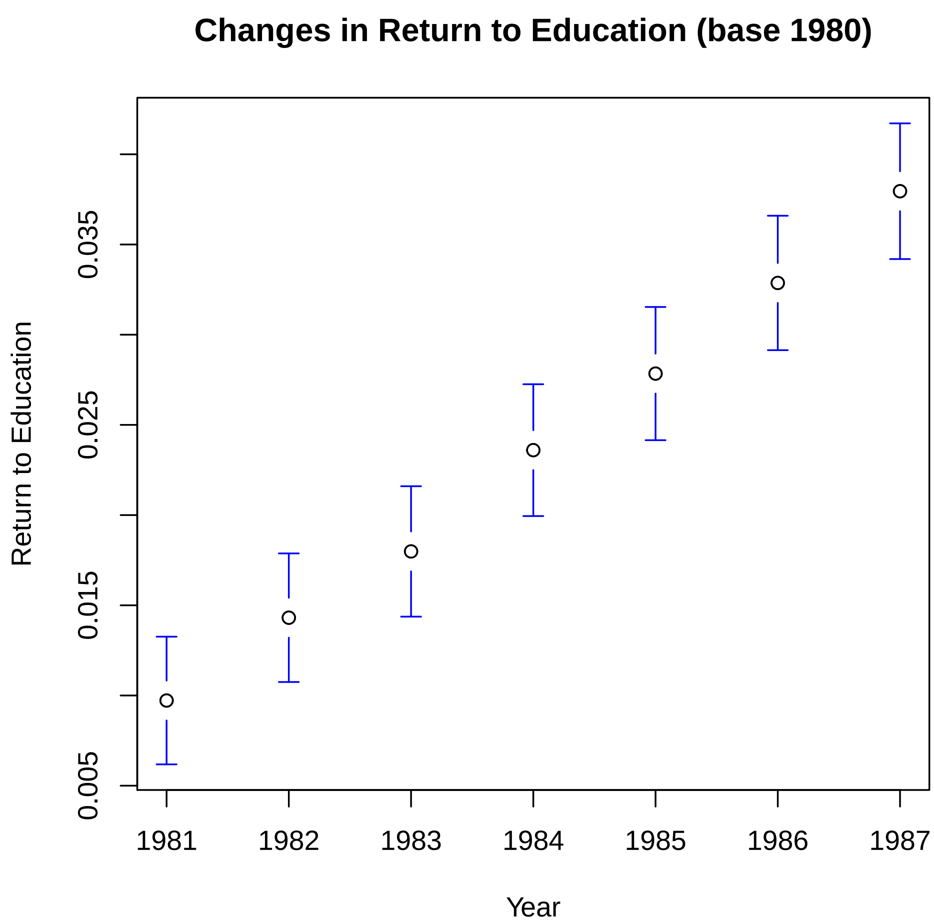
We can also investigate the data in a different manner than **?**. For example, we can investigate if the return to education for the young men in the dataset has changed over time. Having panel data offers access to these types of queries. Note that education is fixed across time for each individual so the within estimator cannot recover an effect. However, by intereacting the level of schooling with the given year, we can effectively make schooling time varying.

```
> # Estimate model
>
> pdata$D81 <- ifelse(pdata$YEAR==1981,1,0)
> pdata$D82 <- ifelse(pdata$YEAR==1982,1,0)
> pdata$D83 <- ifelse(pdata$YEAR==1983,1,0)
> pdata$D84 <- ifelse(pdata$YEAR==1984,1,0)
> pdata$D85 <- ifelse(pdata$YEAR==1985,1,0)
> pdata$D86 <- ifelse(pdata$YEAR==1986,1,0)
> pdata$D87 <- ifelse(pdata$YEAR==1987,1,0)
> vvmodel <- plm(WAGE~UNION+MAR+I(SCHOOL*D81)+
+                  I(SCHOOL*D82)+I(SCHOOL*D83)+
+                  I(SCHOOL*D84)+I(SCHOOL*D85)+
```

```
+                  I(SCHOOL*D86)+I(SCHOOL*D87),
+                data=pdata,
+                effect="individual",model="within")
> psumvv <- summary(vvmodel)
> psumvv

Oneway (individual) effect Within Model

Call:
plm(formula = WAGE ~ UNION + MAR + I(SCHOOL * D81) + I(SCHOOL *
    D82) + I(SCHOOL * D83) + I(SCHOOL * D84) + I(SCHOOL * D85) +
    I(SCHOOL * D86) + I(SCHOOL * D87), data = pdata, effect = "individual",
    model = "within")

Balanced Panel: n=545, T=8, N=4360

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-4.1600 -0.1260  0.0117  0.1600  1.4800

Coefficients :
                 Estimate Std. Error t-value  Pr(>|t|)
UNION           0.0832173  0.0194171  4.2858 1.866e-05 ***
MAR             0.0548293  0.0183944  2.9808  0.002894 **
I(SCHOOL * D81) 0.0097214  0.0018049  5.3861 7.635e-08 ***
I(SCHOOL * D82) 0.0143111  0.0018181  7.8715 4.532e-15 ***
I(SCHOOL * D83) 0.0179854  0.0018437  9.7553 < 2.2e-16 ***
I(SCHOOL * D84) 0.0235997  0.0018638 12.6620 < 2.2e-16 ***
I(SCHOOL * D85) 0.0278436  0.0018831 14.7862 < 2.2e-16 ***
I(SCHOOL * D86) 0.0328681  0.0019000 17.2987 < 2.2e-16 ***
I(SCHOOL * D87) 0.0379534  0.0019182 19.7854 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    572.05
Residual Sum of Squares: 474.51
R-Squared      :  0.17052
      Adj. R-Squared :  0.14885
F-statistic: 86.9361 on 9 and 3806 DF, p-value: < 2.22e-16
```

Now let plot out just what these effects look like over time. To do this we will use the `gregmisc` library and the command `CIplot`. The corresponding figure appears as Figure 1. We can see that the impact of education is increasing over time and the effects after 1984 are statistically significantly diffrent from 0.

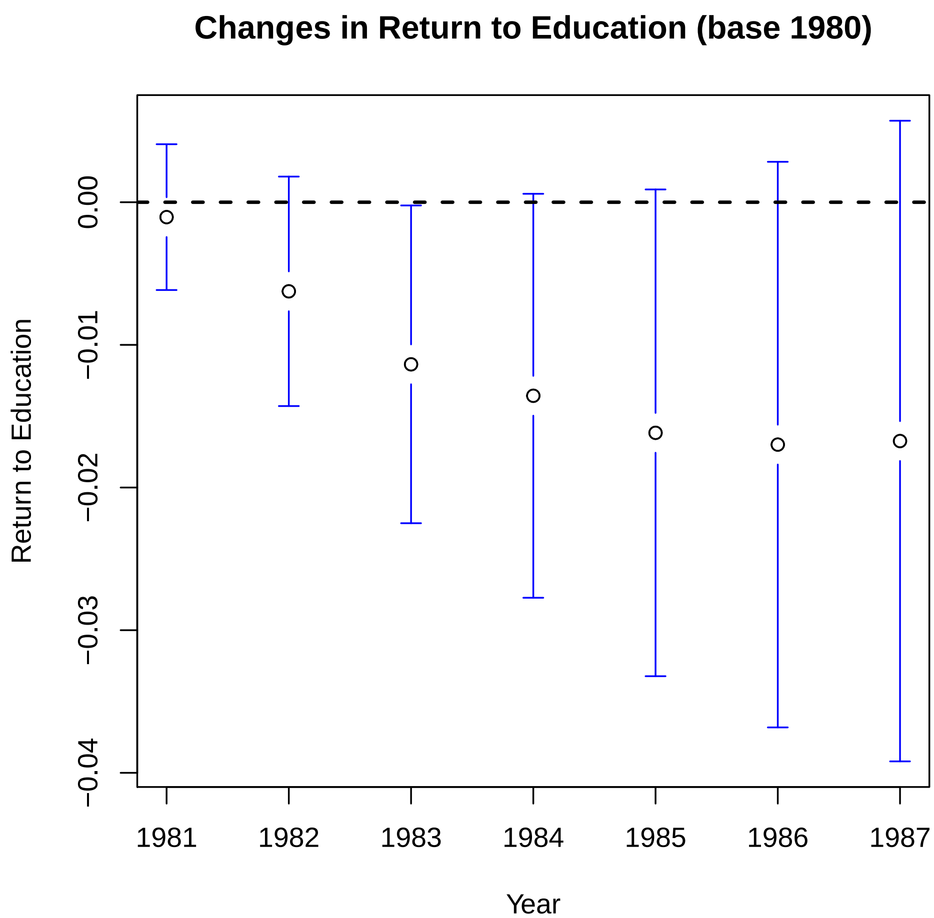
FIGURE 1. 95% Confidence interval plots over time for the return to education.



Notice that what we learn from this example is heavily dependent upon our specification. If we add in experience and experience squared, we see that the impact of education over time has much different effect. Why is this? The omission of experience, which is heavily correlated with education in each year leads to the education-year effects also picking up the impact of overall experience.

Next, we can test if these joint education-time effects are statistically significant. This is just a simple F-test that these 7 coefficients are jointly zero. To perform this test we will use the `linearHypothesis` command in the `car` package. We do this test for our initial unobserved effects model that omits experience and experience squared.

FIGURE 2. 95% Confidence interval plots over time for the return to education with an alternative specification.




```
> library(car)
> ## Test for joint significance of education*time effects,
> hypothesis.matrix <- c(0,0,1,0,0,0,0,0,0,
+                        0,0,0,1,0,0,0,0,0,
+                        0,0,0,0,1,0,0,0,0,
+                        0,0,0,0,0,1,0,0,0,
+                        0,0,0,0,0,0,1,0,0,
+                        0,0,0,0,0,0,0,1,0,
+                        0,0,0,0,0,0,0,0,1)
> hypothesis.matrix <- matrix(hypothesis.matrix,7,9,byrow=TRUE)
> ## The 6 corresponds to the number of restrictions (# of
> ## monthly dummies), and the 9 to the
```

```
> ## number of parameters in the unrestricted model.
>
> ## Here we have a matrix since we have a multiple hypothesis.
> rhs <- c(0,0,0,0,0,0,0)
> ## This is a vector since we have a multiple hypothesis. The 0s
> ## correspond to the values assumed in the null hypothesis.
>
> test <- linearHypothesis(vvmodel,
+                                                  hypothesis.matrix=hypothesis.matrix,
+                                                  rhs=rhs)
> test
Linear hypothesis test

Hypothesis:
I(SCHOOL * D81) = 0
I(SCHOOL * D82) = 0
I(SCHOOL * D83) = 0
I(SCHOOL * D84) = 0
I(SCHOOL * D85) = 0
I(SCHOOL * D86) = 0
I(SCHOOL * D87) = 0

Model 1: restricted model
Model 2: WAGE ~ UNION + MAR + I(SCHOOL * D81) + I(SCHOOL * D82) + I(SCHOOL *
    D83) + I(SCHOOL * D84) + I(SCHOOL * D85) + I(SCHOOL * D86) +
    I(SCHOOL * D87)

  Res.Df Df  Chisq Pr(>Chisq)
1   3813
2   3806  7 553.75  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We see that we round reject the null hypothesis that the joint time-education effects are statistically significant. This example of interacting a time constant variable with a time dummy helps to illustrate one way in which we can incorporate time constant variables in a fixed effects framework for the unobserved effects model.

References

Baltagi, B. H. & Pinnoi, N. (1995), 'Public capital stock and state productivity growth: Further evidence from an error components model', *Empirical Economics* **20**, 351–359.

Munnell, A. H. (1990), 'How does public infrastructure affect regional economic performance?', *New England Economic Review* **September**, 11–32.

Vella, F. & Verbeek, M. (1998), 'Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men', *Journal of Applied Econometrics* **13**, 163–183.