EXCERSISES IN APPLIED PANEL DATA ANALYSIS #5

CHRISTOPHER F. PARMETER

1. Introduction

This R example will introduce you to conduct inference for the unobserved effects model. Notably, we will consider tests of poolability and for comparing across the fixed and random effects frameworks. The plm library offers a range of tests for testing these important hypotheses.

2. Tests for Poolability and Hausman Specification

2.1. **Estimating the Demand for Gasoline.** The first of our two examples will use the classic study of Baltagi & Griffin (1983), which we discussed in PanelExercise4. Recall that Baltagi & Griffin (1983) estimated a demand equation for gasoline at the country level. Their balanced panel constituted 18 OECD countries over the period 1960-1978. Their baseline econometric model is

$$\ln(Gas/Car)_{it} = \beta_0 + \beta_1 \ln(GDP/Pop)_{it} + \beta_2 \ln(P_{Gas}/P_{GDP})_{it} + \beta_3 \ln(Car/Pop)_{it} + c_i + \varepsilon_{it}, \tag{1}$$

where Gas/Car is gasoline consumption per car, GDP/Pop is per capita income, P_{Gas}/P_{GDP} is the price of gasoline and Car/Pop is the stock of cars per capita. The key coefficient of interest in β_2, which identifies the price elasticity of gasoline.

```
> library(plm)
> ## Load in dataset of Baltagi and Griffin (1983)
> data("Gasoline")
```

Our primary concern in this exercise is to consider aspects of poolability and the appropriate framework for the unobserved effects. Given that we have four coefficients of interest and 19 time periods per country, we could essentially estimate a single demand equation for each country. To do this we can use the subset option within the lm() call.

```
> gas.USA <- lm(lgaspcar~lincomep+lrpmg+
+                     lcarpcap,
+               data=Gasoline,subset=(country=="U.S.A."))
```

University of Miami
Date: September 11, 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..

Of course this is an inefficient way to estimate the model across individual countries. I could store all my results in a matrix and then analyze them separately.

```
> country.coef <- matrix(0,18,4)
> country.id <- unique(Gasoline$country)
> for (j in 1:length(country.id)){
+
+ country.coef[j,] <- coefficients(lm(lgaspcar~lincomep+
+                                      lrpmg+
+                                      lcarpcap,
+                                     data=Gasoline,
+                                     subset=(country==country.id[j])))
+ }
> summary(country.coef)
```

```
       V1                 V2                 V3                  V4
 Min.   :-2.8861   Min.   :-0.8302   Min.   :-0.79320   Min.   :-0.84469
 1st Qu.: 0.5156   1st Qu.: 0.1101   1st Qu.:-0.36909   1st Qu.:-0.61400
 Median : 3.0839   Median : 0.3776   Median :-0.24557   Median :-0.49504
 Mean   : 2.1928   Mean   : 0.3504   Mean   :-0.27696   Mean   :-0.43179
 3rd Qu.: 4.1293   3rd Qu.: 0.7191   3rd Qu.:-0.13899   3rd Qu.:-0.24993
 Max.   : 4.9260   Max.   : 1.1193   Max.   :-0.04165   Max.   : 0.03992
```

```
> ## A quicker way to do this is to use the pvcm()
> ## command in plm
> gas.vc <-  pvcm(lgaspcar~lincomep+
+                   lrpmg+lcarpcap,
+                 data = Gasoline)
> summary(coefficients(gas.vc))
```

```
  (Intercept)         lincomep           lrpmg             lcarpcap
 Min.   :-2.8861   Min.   :-0.8302   Min.   :-0.79320   Min.   :-0.84469
 1st Qu.: 0.5156   1st Qu.: 0.1101   1st Qu.:-0.36909   1st Qu.:-0.61400
 Median : 3.0839   Median : 0.3776   Median :-0.24557   Median :-0.49504
 Mean   : 2.1928   Mean   : 0.3504   Mean   :-0.27696   Mean   :-0.43179
 3rd Qu.: 4.1293   3rd Qu.: 0.7191   3rd Qu.:-0.13899   3rd Qu.:-0.24993
 Max.   : 4.9260   Max.   : 1.1193   Max.   :-0.04165   Max.   : 0.03992
```

Clearly there is some degree of heterogeneity across the coefficient estimates for each country. So we can test for pooling and for fixed effects. We first need to estimate th e pooled model and the within model.

```
> gas.pool <- plm(lgaspcar~lincomep+lrpmg+
+                     lcarpcap,
```

```
+                 model="pooling",
+                 data=Gasoline)
> gas.wn <- plm(lgaspcar~lincomep+lrpmg+
+                     lcarpcap,
+                 model="within",
+                 effect="individual",
+                 data=Gasoline)
> ## Now test for poolability and for presence of
> ## unobserved additive heterogeneity
>
> ## Test for unobserved heterogeneity
> ## NOTE, the first model must be the
> ## within model!
> unobs.test <- pFtest(gas.wn,gas.pool)
> unobs.test

        F test for individual effects

data:  lgaspcar ~ lincomep + lrpmg + lcarpcap
F = 83.9608, df1 = 17, df2 = 321, p-value < 2.2e-16
alternative hypothesis: significant effects

> ## Test for pooling
> ##
> pool.test <- pooltest(gas.wn,gas.vc)
> pool.test

        F statistic

data:  lgaspcar ~ lincomep + lrpmg + lcarpcap
F = 27.3352, df1 = 51, df2 = 270, p-value < 2.2e-16
alternative hypothesis: unstability
```

We can see from these tests that it appears for this data there is substantial heterogeneity in gasoline demand across OECD countries. Next we turn our attention to the appropriate framework for the unobserved effects to enter the model. Here we will deploy the Hausman test. To do this we use `phtest()`.

```
> ## Estimate Random Effects Model
> gas.rd.nerl <- plm(lgaspcar~lincomep+lrpmg+
+                       lcarpcap,
+                     model="random",
+                     effect="individual",
```

```
+                       random.method="nerl",
+                       data=Gasoline)
[1] 342  18
> gas.rd.swar <- plm(lgaspcar~lincomep+lrpmg+
+                         lcarpcap,
+                       model="random",
+                       effect="individual",
+                       random.method="swar",
+                       data=Gasoline)
> gas.rd.wh <- plm(lgaspcar~lincomep+lrpmg+
+                         lcarpcap,
+                       model="random",
+                       effect="individual",
+                       random.method="walhus",
+                       data=Gasoline)
> gas.rd.amem <- plm(lgaspcar~lincomep+lrpmg+
+                         lcarpcap,
+                       model="random",
+                       effect="individual",
+                       random.method="amemiya",
+                       data=Gasoline)
> ## Now perform the Hausman test.
> phtest.amem <- phtest(gas.wn,gas.rd.amem)
> phtest.nerl <- phtest(gas.wn,gas.rd.nerl)
> phtest.swar <- phtest(gas.wn,gas.rd.swar)
> phtest.wh   <- phtest(gas.wn,gas.rd.wh)
> ## Print out p values
> phtest.amem$p.value
           [,1]
[1,] 0.01567889
> phtest.nerl$p.value
           [,1]
[1,] 0.03703636
> phtest.swar$p.value
            [,1]
[1,] 2.46008e-65
> phtest.wh$p.value
```

```
            [,1]
[1,] 3.78466e-59
```

Across all four random effects estimators we reject the null hypothesis that the random effects framework is appropriate for our cross country gasoline deman equation. Taken collectively these tests suggest there is substantial heterogeneity across gasonline demand and an panel setup is necessary to help capture this heterogeneity.

2.2. **The Effect of Health Aid on Life Expectancy.**

```
> ## Load the data. Need the foreign package since this is
> ## a STATA dataset.
> library(foreign)
> health.data <- read.dta(file="Thierry.dta")
> ## First lets condense names of variables to make things easier
> names(health.data) <- c("country","code","year","demo","gdp.gro",
+                         "life.ex","docs","urbanpop","health")
> ## Now convert to a panel data.frame
> phealth.data <- pdata.frame(health.data,index=c("country","year"))
> ## Now lets estimate a pooled model, the within model, and a random
> ## effects model. Then we will test for unobserved effects and
> ## appropriate framework
>
> ## Pooled model
> health.pool <- plm(life.ex~log(health)+demo+gdp.gro+
+                      docs+urbanpop,
+                    data=phealth.data,
+                    model="pooling")
> ## Within model
> health.wn <- plm(life.ex~log(health)+demo+gdp.gro+
+                      docs+urbanpop,
+                    data=phealth.data,
+                    model="within",effect="individual")
> ## Random Effects model
> health.rd <- plm(life.ex~log(health)+demo+gdp.gro+
+                      docs+urbanpop,
+                    data=phealth.data,
+                    model="random",
+                  effect="individual",
+                  random.method="swar")
```

```
> ## Now test for unobserved effects
> pFtest(health.pool,health.wn)

        F test for individual effects

data:  life.ex ~ log(health) + demo + gdp.gro + docs + urbanpop
F = 1.8921, df1 = -37, df2 = 71, p-value = NA
alternative hypothesis: significant effects
> plmtest(health.wn,effect="individual")

        Lagrange Multiplier Test - (Honda)

data:  life.ex ~ log(health) + demo + gdp.gro + docs + urbanpop
normal = 138.1908, p-value < 2.2e-16
alternative hypothesis: significant effects
> ## Now Hausman test
> phtest(health.wn,health.rd)

        Hausman Test

data:  life.ex ~ log(health) + demo + gdp.gro + docs + urbanpop
chisq = 9.0088, df = 5, p-value = 0.1087
alternative hypothesis: one model is inconsistent
```

References

Baltagi, B. H. & Griffin, J. M. (1983), 'Gasoline demand in the OECD: An application of pooling and testing procedures', *European Economic Review* **22**, 117–137.