EXCERSISES IN APPLIED PANEL DATA ANALYSIS #10

CHRISTOPHER F. PARMETER

1. Introduction

This exercises is designed to allow you to gain familiarity with the `plm` interface in the presence of unbalanced panel data. The exercise will also go through an example where one 'balances the panel'.

2. Unbalanced Panel Data

This example follows our dicusson of pooled estimation of Ram (2009) from exercise 1. Recall that Ram (2009) looks at the debate over the relationshp between how open a country is to international trade, the size of the country (as measured by popoulation) and the amount of money that the country's government spends. We can specify these individual relationships as:

$$\log(gov_{it}) = \alpha_0 + \alpha_1 \log(open_{it}) + \alpha_2 \log(GDP_{it}) + \varepsilon_{1,it} \tag{1}$$

$$\log(open_{it}) = \beta_0 + \beta_1 \log(size_{it}) + \beta_2 \log(GDP_{it}) + \varepsilon_{2,it} \tag{2}$$

$$\log(gov_{it}) = \gamma_0 + \gamma_1 \log(size_{it}) + \gamma_2 \log(GDP_{it}) + \varepsilon_{3,it} \tag{3}$$

where *gov* measures government spending (a proxy for government size), *open* measures the ratio of total imports and exports to GDP (a proxy for openness to trade), *size* measures the population of a country (a proxy for the size of the country), and *GDP* is the gross domestic product per capita in the country. Rodrik (1998) argues for the specification of government spending in (1) while Alesina & Wacziarg (1998), argues that because of the relationship in (2), the appropriate government spending relationship is that in (3).

We use the `pwt` package to construct the dataset used by Ram (2009) to adjudicate between these two alternative propositions for government spending. Ram's (2009) construction of the database results in an unbalanced panel.

```
> ##load pwt library
> library(pwt)
> library(plm)
> library(stargazer)
```

University of Miami
Date: September 12, 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.

```
> ## First replicate based on PWT6.1
> data("pwt6.1")
> ## Create data.frame
> data6.1<- data.frame(pwt6.1)
> #Subset of years 1960-2000, to match Ram
> data6.1.ram <- data.frame(subset(data6.1,data6.1$year>1959 & data6.1$year<2001))
> ## First see which countries only have a single observation over the entire period
> ## Ram has 154 countries, but we have 168 countries
> uni.country <- unique(data6.1.ram$country)
> length(uni.country)
[1] 168
> ## Now keep only pop, rgdpch and openk (Ram uses openc -- current prices)
> ## But rgdpch is in chain prices !!#$#@
> check.ram <- na.omit(data6.1.ram[,c(1:4,10,15,20,24)])
> ## Which countries have a single observation
> singleyear.store <- as.numeric()
> for (i in 1:length(uni.country)){
+
+   c.id <- uni.country[i]
+
+         singleyear.store[i] <- length(which(check.ram$country==c.id))
+
+ }
> num.single <- length(which(singleyear.store==1))
> ## Remove the single year countries
> if(num.single>0){
+
+         ram.omit <- uni.country[which(singleyear.store==1)]
+
+         id.omit <- which(check.ram$country%in%ram.omit)
+         check.ram <- check.ram[-id.omit,]
+
+ }
> ## Subset of years 1960-2000, to match Ram
> pdata6.1.ram <- pdata.frame(check.ram,index=c("country","year"))
> ## Estimate Pooled OLS
> ## Table 1, Column (1)
> model1.1 <- plm(log(cg)~log(pop)+log(rgdpch),
+                 model="pool",
```

```
+                 data=pdata6.1.ram)
> ## Table 2, Column (1)
> model2.1 <- plm(log(openc)~log(pop)+log(rgdpch),
+                 model="pool",
+                 data=pdata6.1.ram)
> ## Table 3, Column (1)
> model3.1 <- plm(log(cg)~log(openc)+log(rgdpch),
+                 model="pool",
+                 data=pdata6.1.ram)
> ## Estimate Within
> ## Table 1, Column (4)
> model1.4 <- plm(log(cg)~log(pop)+log(rgdpch),
+                 model="within",
+                 effect="twoways",
+                 data=pdata6.1.ram)
> ## Table 2, Column (4)
> model2.4 <- plm(log(openc)~log(pop)+log(rgdpch),
+                 model="within",
+                 effect="twoways",
+                 data=pdata6.1.ram)
> ## Table 3, Column (4)
> model3.4 <- plm(log(cg)~log(openc)+log(rgdpch),
+                 model="within",
+                 effect="twoways",
+                 data=pdata6.1.ram)
```

The results from these three models appear in Table 1. The results match with those in Ram (2009) for the pooled and fixed effects frameworks. Ram (2009) does not report random effects estimates given that in all of his specifications the Hausman test rejects the null that the random effects framework is appropriate.

Now we can 'balanced' the panel either py shortening the time frame (to maximize the number of countries), or we can keep the time frame but omit countries that do not appear throughout the sample. We will used the second option.

```
> ## Remove the countries with less than 41 time observations
> ram.omit <- uni.country[which(singleyear.store<41)]
> id.omit <- which(check.ram$country%in%ram.omit)
> check.ram1 <- check.ram[-id.omit,]
> ## Balanced panel from 1960-2000
> pdata6.1.ram.bal <- pdata.frame(check.ram1,index=c("country","year"))
> ## Estimate Pooled OLS
```

Table 1.

	Dependent variable:					
	log(cg)		log(openc)		log(cg)	
	Pool	FE	Pool	FE	Pool	FE
log(pop)	-0.080^{***}	0.475^{***}	-0.205^{***}	0.032		
	(0.004)	(0.037)	(0.004)	(0.032)		
log(openc)					0.244^{***}	0.081^{***}
					(0.013)	(0.017)
log(rgdpch)	-0.139^{***}	-0.160^{***}	0.158^{***}	0.142^{***}	-0.176^{***}	-0.245^{***}
	(0.008)	(0.017)	(0.006)	(0.015)	(0.008)	(0.017)
Constant	4.653^{***}		4.514^{***}		3.282^{***}	
	(0.074)		(0.062)		(0.071)	
Observations	5,117	5,117	5,117	5,117	5,117	5,117
R^2	0.115	0.070	0.436	0.019	0.116	0.044
Adjusted R^2	0.115	0.067	0.435	0.019	0.116	0.042
Note:						*p<0.1; **p<0.05; ***p<0.01

```
> ## Table 1, Column (1)
> model1.1 <- plm(log(cg)~log(pop)+log(rgdpch),
+                 model="pool",
+                 data=pdata6.1.ram.bal)
> ## Table 2, Column (1)
> model2.1 <- plm(log(openc)~log(pop)+log(rgdpch),
+                 model="pool",
+                 data=pdata6.1.ram.bal)
> ## Table 3, Column (1)
> model3.1 <- plm(log(cg)~log(openc)+log(rgdpch),
+                 model="pool",
+                 data=pdata6.1.ram.bal)
> ## Estimate Within
> ## Table 1, Column (4)
> model1.4 <- plm(log(cg)~log(pop)+log(rgdpch),
+                 model="within",
+                 effect="twoways",
+                 data=pdata6.1.ram.bal)
> ## Table 2, Column (4)
> model2.4 <- plm(log(openc)~log(pop)+log(rgdpch),
+                 model="within",
```

```
+                  effect="twoways",
+                  data=pdata6.1.ram.bal)
> ## Table 3, Column (4)
> model3.4 <- plm(log(cg)~log(openc)+log(rgdpch),
+                  model="within",
+                  effect="twoways",
+                  data=pdata6.1.ram.bal)
> ## Estimate Random Effects
> ## Table 1
> model1.rd <- plm(log(cg)~log(pop)+log(rgdpch),
+                   model="random",
+                   effect="twoways",
+                   random.method="walhus",
+                   data=pdata6.1.ram.bal)
> ## Table 2
> model2.rd <- plm(log(openc)~log(pop)+log(rgdpch),
+                   model="random",
+                   effect="twoways",
+                   random.method="walhus",
+                   data=pdata6.1.ram.bal)
> ## Table 3
> model3.rd <- plm(log(cg)~log(openc)+log(rgdpch),
+                   model="random",
+                   effect="twoways",
+                   random.method="walhus",
+                   data=pdata6.1.ram.bal)
```

The results from these nine models appear in Table 2.

The p-values from the Hausman tests for Models (1)-(3) are 0, 0 and 0, respectively. We note that 'balancing' our panel has effectively reduced our sample size by almost 20%. However, in this setting, the main conclusions of Ram (2009) have not changed.

TABLE 2.

	Dependent variable:								
		log(cg)			log(openc)			log(cg)	
	Pool	FE	RD	Pool	FE	RD	Pool	FE	RD
log(pop)	−0.028***	0.549***	0.139***	−0.206***	0.035	−0.135***			
	(0.005)	(0.040)	(0.021)	(0.004)	(0.035)	(0.018)			
log(openc)							0.130***	0.090***	0.096***
							(0.015)	(0.019)	(0.018)
log(rgdpch)	−0.142***	−0.218***	−0.259***	0.138***	0.190***	0.157***	−0.160***	−0.339***	−0.295***
	(0.008)	(0.021)	(0.018)	(0.007)	(0.018)	(0.015)	(0.008)	(0.020)	(0.018)
Constant	4.149***		3.590***	4.632***		3.847***	3.531***		4.766***
	(0.079)		(0.263)	(0.068)		(0.227)	(0.076)		(0.160)
Observations	4,018	4,018	4,018	4,018	4,018	4,018	4,018	4,018	4,018
R^2	0.082	0.107	0.069	0.402	0.030	0.046	0.093	0.070	0.064
Adjusted R^2	0.082	0.104	0.069	0.402	0.029	0.046	0.093	0.068	0.064

Note: *p<0.1; **p<0.05; ***p<0.01

References

Alesina, A. & Wacziarg, R. (1998), 'Openness, country size, and government size', *Journal of Public Economics* **69**, 305–321.

Ram, R. (2009), 'Openness, country size, and government size: Additional evidence from a large cross-country panel', *Journal of Public Economics* **93**, 213–218.

Rodrik, D. (1998), 'Why do more open economies have bigger governments?', *Journal of Political Economy* **106**, 997–1032.