

PSet 6

Spencer Papay - ssp2170

3/30/2017

Question 1

a)

Panel data is comprised of observations on multiple entities, such as individuals, states, companies, etc. but each entity is observed at least two times with the same criteria/questions.

Repeated cross sections are the same criteria/topics each time, but the experimenter gets different samples of individuals at each observation. Repeated cross sections do not track the same individuals at each observation, but rather a random subset. One would use cross section data when not concerned about tracking specific individuals' changes over time.

b)

Omitted variable bias, measurement error, and simultaneity bias all lead to bias in OLS. In each case, the missing data is causal concerning our variables and is correlated with our regression error, (μ_i) as the regressor is biased and inconsistent. With OVB, we are missing an independent variable that is both a determinant of Y and correlated with at least one included regressor. In measurement error, the result is errors-in-variables bias and leads to correlation between the measured variable and the regression error (μ_i). In simultaneous causality bias we have not only that X causes Y but that Y in return causes X.

$\therefore \text{corr}(X_i, \mu_i) \neq 0$ in all three cases, or in other words, $E[\mu_i|X_i] \neq 0$, which is LSA #1.

OLS with a single cross section may encounter bias when sampling average student weight during finals season, whereas if we used panel data from before, during, and after finals season, it would eliminate the bias of the unhealthy eating habits of students during this season and perhaps might demonstrate that there is an increase in average weight during the stressful time.

c)

Estimator for β_1 is the OLS fixed effect estimator $\hat{\beta}_1$.

Assumptions: 1. $E(\mu_{it}|X_{i1}, \dots, X_{iT}, \alpha_i) = 0$. 2. $(X_{i1}, \dots, X_{iT}, u_{i1}, \dots, u_{iT}), i = 1, \dots, n$ are i.i.d. in joint distribution 3. (X_{it}, μ_{it}) have finite fourth moments. 4. There is no perfect multicollinearity (multiple Xs)

We also need to add the “clustered” standard error formula because observations for the same entity are not independent, regardless that observations across entities are independent if entities are drawn by simple random sampling.

Yes, however each entity has a different intercept and its possible to estimate β_0 per entity. It is possible to estimate 0 as the mean change in the dependent variable between periods 1 and 2 may be nonzero.

Question 2

a) γ_1 is an entity fixed effect, and changes between entities but not over time. In the regression line, each entity (i) has a different intercept. Changes in it change the intercept of the regression line.

b) γ_t is a time fixed effect, and changes over time but not between entities. In the regression line, each time period (t) can have a different intercept. Changes in it change the intercept of the regression line.

c) β_1 does not vary with either i or t , rather only the intercepts vary. Variation in x_1 lets us identify β_1 with each entity (i), but only after we control for the average of each time period (t). It is the variation within each entity that defines β_1 .

d) $y_{it} = \beta_0 + x_{it}\beta_1 + \gamma_i + \gamma_t + \epsilon_{it} \rightarrow$

$$y_{i,t} - y_{i,t-1} = \beta_0 - \beta_0 + (x_{i,t} \times \beta_1) - (x_{i,t-1} \times \beta_1) + \gamma_i - \gamma_i + (\gamma_t) - (\gamma_{t-1}) + (\epsilon_{i,t}) - (\epsilon_{i,t-1}) \rightarrow$$

$$\Delta(y_{i,t}) = \beta_1 \times \Delta(x_{i,t}) + \Delta(\gamma_t) + \Delta(\epsilon_{i,t})$$

e) γ_i gets removed from our equation through difference from itself. This implies that entity fixed effects aren't estimated after the first difference is constructed. The common entity mean was already controlled for after looking at change within the entity (i) over time.

f) γ_t remains in the equation after constructing the first difference but no longer measures the average, instead measuring the change in average between the two time periods γ_t and γ_{t-1} .

g) Introduce series of entity i dummy variables to the first differenced equation: dummy variables represent entity (i) specific trend, rather than a change in between period trend, as the remainder of the difference equation variables represent. Each period i, t generates a new dummy variables: resulting in $(t * \text{dummy})$ on right side. Their coefficients represent the change in the intercept resulting from the differences in y across entities. Thus interpret the coefficients on dummy variables as the effect on y that results purely from observed differences between entities.

Question 3

a)

```
m1 <- plm(lrent~y90+lpop+lavginc+pctstu, data=rental,
          index=c("city","year"),model="pooling")
se1 <- sqrt(diag(vcovHC(m1)))
```

The coefficient on $y90$ indicates that other things held equal, nominal rents grew by over 26% over the 10 year period.

```
stargazer(m1, type="latex", header=FALSE,
          title="Effects on Rent Price",
          omit.stat = c("f"), dep.var.caption = "Price on Log Scale",
          intercept.bottom = FALSE, se=list(se1),
          notes="Errors are heteroskedastic")
```

b) The $pctstu$ coefficient indicates that one percentage point increase in $pctstu$ is correlated to an increase in rent by half a percentage point (.5%). The errors show this to be statistically significant.

c) No, as α_i does not appear in the equation. If α_i is in the error term, the errors between the two periods for each city are positively correlated, and this renders the usual OLS standard errors and t-statistics invalid.

d)

```
m2 <- plm(clrent~clpop+clavginc+cpctstu, data=rental,
          index=c("city","year"),model="pooling")
```

The effect of $pctstu$ is more than twice as large as we estimated in the pooled OLS equation. In this regression, a one percentage point increase in $pctstu$ is estimated to increase rental rates by about 1.1%. However, this estimate is less precise when we difference the equations, as we may have expected. While α_i was differenced away, there may be other unobserved variables that change over time and are correlated with $\Delta pctstu$.

Table 1: Effects on Rent Price

	Price on Log Scale
	lrent
Constant	-0.569 (1.015)
y90	0.262*** (0.063)
lpop	0.041 (0.028)
lavginc	0.571*** (0.120)
pctstu	0.005*** (0.001)
Observations	128
R ²	0.861
Adjusted R ²	0.857
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01 Errors are heteroskedastic	

```
stargazer(m2, type="latex", header=FALSE,
  title="Effects on Rent Price",
  omit.stat = c("f"), dep.var.caption = "Change in Price on Log Scale",
  intercept.bottom = FALSE, notes="Errors are not heteroskedastic")
```

e)

```
se2 <- sqrt(diag(vcovHC(m2)))
```

```
stargazer(m2, type="latex", header=FALSE,
  title="Effects on Rent Price",
  omit.stat = c("f"), dep.var.caption = "Change in Price on Log Scale",
  intercept.bottom = FALSE, se=list(se2), notes="Errors are heteroskedastic")
```

f)

```
m3 <- plm(lrent~y90+lpop+lavginc+pctstu, data=rental,
  index=c("city", "year"), model="within")
se3 <- sqrt(diag(vcovHC(m3)))
```

```
stargazer(m3, type="latex", header=FALSE,
  title="Effects on Rent Price",
  omit.stat = c("f"), dep.var.caption = "Price on Log Scale",
  intercept.bottom = FALSE, se=list(se3), notes="Errors are heteroskedastic")
```

Table 2: Effects on Rent Price

	Change in Price on Log Scale
	clrent
Constant	0.386*** (0.037)
clpop	0.072 (0.088)
clavginc	0.310*** (0.066)
cpctstu	0.011*** (0.004)
Observations	64
R ²	0.322
Adjusted R ²	0.288
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01 Errors are not heteroskedastic

Table 3: Effects on Rent Price

	Change in Price on Log Scale
	clrent
Constant	0.386*** (0.047)
clpop	0.072 (0.067)
clavginc	0.310*** (0.086)
cpctstu	0.011*** (0.003)
Observations	64
R ²	0.322
Adjusted R ²	0.288
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01 Errors are heteroskedastic

Table 4: Effects on Rent Price

	Price on Log Scale
	lrent
y90	0.386*** (0.047)
lpop	0.072 (0.067)
lavginc	0.310*** (0.086)
pctstu	0.011*** (0.003)
Observations	128
R ²	0.977
Adjusted R ²	0.950
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01 Errors are heteroskedastic	