ANALYSIS OF PANEL DATA

Fixed-Effect and Random-Effect Models

Random-Effect Models

Random Effect Models

Recall the general linear regression models with unobserved individual effect:

$$y_{it} = \beta_0 + \underbrace{\beta_1 x_{1it} + \cdots + \beta_k x_{kit}} + \underbrace{a_i} + \epsilon_{it}$$

where $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$

- If the unobserved individual effect a_i is uncorrelated with the explanatory variables, the techniques described above are not needed to produce a consistent estimator.

$$\boxed{Cov(x_{itj}, a_i) = 0}$$

for $t = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, k$

- What is means is that the random effect assumptions include all of the fixed effect assumptions plus the additional (strong) requirement that a_i is independent of all explanatory variables in all time periods in the model.

- How should we estimate β_j in the above unobserved effect model?
- Note that under these assumptions, we can use the *random effect* models to obtain consistent OLS estimators using only a single cross section: there is no need to the panel data at all if the objective is to obtain consistent estimators for β_j.
- Of course, using a single cross section means we throw away potentially valuable information offered by panel data.
- Let's rewrite the model using a composite error term:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \nu_{it}$$

where $\nu_{it} = a_i + \epsilon_{it}$ $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$

- Because a_i is contained in the composite error term in each time period, ν_{it} is serially correlated. Under Random Effect assumptions, we have

$$\boxed{Corr(\nu_{it}, \nu_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}}$$

- In other words, when $E(X_{it}a_i) \neq 0$, panel data provides a valuable tool for eliminating omitted variables bias. We use Fixed Effects to gain the benefits of panel data.

- When $E(X_{it}a_i) = 0$, panel data does not offer special benefits. We use Random Effects to overcome the serial correlation of panel data.

- The correlation in the error term can be substantial. Because pooled OLS standard errors ignore this correlation, they will be in correct, as will the usual test statistics.

- A solution to this problem is to use generalized least square (GLS).

- For this procedure to come with good properties, we need large N and small T. This is, a short panel.

- Deriving the GLS transformation to eliminate serial correlation requires quite a bit of matrix algebra. However, the transformation itself is pretty simple:

$$\boxed{\lambda = 1 - \left[\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T\sigma_a^2}\right]^{1/2}}$$

which falls between 0 and 1.

- The transformed model becomes

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1\left(x_{1it} - \lambda\bar{x}_{1i}\right) + \cdots + \beta_k\left(x_{kit} - \lambda\bar{x}_{ki}\right) + \left(\nu_{it} - \lambda\bar{\nu}_i\right)$$

where the *overbar* denotes the time averages.

- While the FE estimators subtracts the time averages from the corresponding variable, the random effect transformation subtracts a fraction of that time average with the fraction being a function of σ_ϵ^2, σ_a^2 and T.

- The GLS estimator is simply the pooled estimator of the above model.

- One of the advantage of random effect model (relative to fixed effect model) is that it allows for time invariant explanatory variables to be included in the model; a fixed effect models eliminates all the time invariant (observed and unobserved) variables.

- In practice, λ needs to be estimated:

$$1 - \left[\frac{1}{1 + T(\hat{\sigma}_a^2 / \hat{\sigma}_\epsilon^2)} \right]^{1/2}$$

where $\hat{\sigma}_a^2$ and $\hat{\sigma}_\epsilon^2$ are consistent estimators of σ_a^2 and σ_ϵ^2.

- These estimators can be based on pooled OLS or fixed effect residuals.

- In practice, random effect models can be implemented easily by modern econometric packages, such as *plm*, and λ can be automated computed as well.

- The *feasible* GLS estimator that uses $\hat{\lambda}$ in place of λ is called the **random effect estimator**.

- Under the random effect assumptions, the estimator is *consistent* and *asymptotically normally distributed* for large N and fixed T.

- In applications of FE and RE, it is usually informative also to compute the pooled OLS estimates as well because comparing the three sets of estimates can help determine the nature of the biases caused by leaving the unobserved effect, a_i, entirely in the error term (as does pooled OLS) or partially in the error term (as does the RE transformation).

- But we must remember that, even if ai is uncorrelated with all explanatory variables in all time periods, the pooled OLS standard errors and test statistics are generally invalid: they ignore the often substantial serial correlation in the composite errors, $\nu_{it} = a_i + \epsilon_{it}$

Example: A Wage Equation Using Panel Data

- Let's use use the data in *wagepan.RData* to estimate a wage equation for men.

- Specifically, we use three methods: pooled OLS, random effects, and fixed effects. The first two methods include *educ* and *race dummies (black and hispan)*, but these drop out of the fixed effects model. The time- varying variables are *exper, exper2, union*, and *married*.

Three Different Estimators of a Wage Equation

Dependent Variable: log(*wage*)			
Independent Variables	**Pooled OLS**	**Random Effects**	**Fixed Effects**
educ	.091 (.005)	.092 (.011)	———
black	−.139 (.024)	−.139 (.048)	———
hispan	.016 (.021)	.022 (.043)	———
exper	.067 (.014)	.106 (.015)	———
$exper^2$	−.0024 (.0008)	−.0047 (.0007)	−.0052 (.0007)
married	.108 (.016)	.064 (.017)	.047 (.018)
union	.182 (.017)	.106 (.018)	.080 (.019)

- The coefficients on educ, black, and hispan are similar for the pooled OLS and random effects ~~esti- mations.~~
- The pooled OLS standard errors are the usual OLS standard errors, and these underestimate the true standard errors because they ignore the positive serial correlation, but we report them here for comparison only.
- The experience profile is somewhat different, and both the marriage and union *premiums* fall notably in the random effects estimation.
- Note that when we eliminate the unobserved effect entirely by using fixed effects, the marriage premium falls to about 4.7, although it is still statistically significant.

A partial list of the dataset

```
'data.frame':   4360 obs. of  44 variables:
 $ nr      : int  13 13 13 13 13 13 13 13 17 17 ...
 $ year    : int  1980 1981 1982 1983 1984 1985 1986 1987 1980 1981 ...
 $ agric   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ black   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ bus     : int  1 0 1 1 0 1 1 1 0 0 ...
 $ construc: int  0 0 0 0 0 0 0 0 0 0 ...
 $ ent     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ exper   : int  1 2 3 4 5 6 7 8 4 5 ...
 $ fin     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ hisp    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ poorhlth: int  0 0 0 0 0 0 0 0 0 0 ...
 $ hours   : int  2672 2320 2940 2960 3071 2864 2994 2640 2484 2804 ...
 $ manuf   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ married : int  0 0 0 0 0 0 0 0 0 0 ...
 $ min     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nrthcen : int  0 0 0 0 0 0 0 0 0 0 ...
 $ nrtheast: int  1 1 1 1 1 1 1 1 1 1 ...
 $ occ1    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ2    : int  0 0 0 0 0 1 1 1 1 1 ...
 $ occ3    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ4    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ5    : int  0 0 0 0 1 0 0 0 0 0 ...
 $ occ6    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ7    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ8    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ occ9    : int  1 1 1 1 0 0 0 0 0 0 ...
 $ per     : int  0 1 0 0 1 0 0 0 0 0 ...
 $ pro     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ pub     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ rur     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ south   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ educ    : int  14 14 14 14 14 14 14 14 13 13 ...
```

```
> head(cbind(wagepan$nr,wagepan$year),50)
      [,1] [,2]
 [1,]   13 1980
 [2,]   13 1981
 [3,]   13 1982
 [4,]   13 1983
 [5,]   13 1984
 [6,]   13 1985
 [7,]   13 1986
 [8,]   13 1987
 [9,]   17 1980
[10,]   17 1981
[11,]   17 1982
[12,]   17 1983
[13,]   17 1984
[14,]   17 1985
[15,]   17 1986
[16,]   17 1987
[17,]   18 1980
[18,]   18 1981
[19,]   18 1982
[20,]   18 1983
[21,]   18 1984
[22,]   18 1985
[23,]   18 1986
[24,]   18 1987
```

Convert the panel data into a structure suitable for the plm() function.

```
> wagepan.panel<-plm.data(wagepan, c("nr","year"))
> summary(wagepan.panel)
       nr                year            agric              black              bus         
 13     :   8    1980     : 545   Min.    :0.00000   Min.    :0.0000   Min.    :0.00000  
 17     :   8    1981     : 545   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000  
 18     :   8    1982     : 545   Median :0.00000   Median :0.0000   Median :0.00000  
 45     :   8    1983     : 545   Mean   :0.03211   Mean   :0.1156   Mean   :0.07592  
 110    :   8    1984     : 545   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000  
 120    :   8    1985     : 545   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000  
 (Other):4312    (Other):1090                                                          
    construc           ent                 exper             fin                 hisp      
 Min.    :0.000   Min.    :0.00000   Min.    : 0.000   Min.    :0.00000   Min.    :0.000  
 1st Qu.:0.000   1st Qu.:0.00000   1st Qu.: 4.000   1st Qu.:0.00000   1st Qu.:0.000  
 Median :0.000   Median :0.00000   Median : 6.000   Median :0.00000   Median :0.000  
 Mean   :0.075   Mean   :0.01514   Mean   : 6.515   Mean   :0.03693   Mean   :0.156  
 3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.: 9.000   3rd Qu.:0.00000   3rd Qu.:0.000  
 Max.    :1.000   Max.    :1.00000   Max.    :18.000   Max.    :1.00000   Max.    :1.000  
```

Random-Effect Estimation

```
# Setup the data
wagepan.panel<-plm.data(wagepan, c("nr","year"))
summary(wagepan.panel)
str(wagepan.panel)

wagepan.re <- plm(lwage ~
educ+black+hisp+exper+exper^2+married+union,data=wagepan.panel,
model="random")
summary(wagepan.re)
```

```
> summary(wagepan.re)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = lwage ~ educ + black + hisp + exper + exper^2 +
    married + union, data = wagepan.panel, model = "random")

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

Effects:
                 var std.dev share
idiosyncratic 0.1251  0.3537 0.543
individual    0.1055  0.3248 0.457
theta:  0.6407

Residuals :
   Min.  1st Qu.  Median 3rd Qu.    Max.
-4.5500 -0.1460  0.0253  0.1920  1.5500

Coefficients :
              Estimate Std. Error t-value  Pr(>|t|)
(Intercept) -0.0477025  0.1104704 -0.4318  0.665899
educ         0.1081869  0.0088615 12.2087 < 2.2e-16 ***
black       -0.1409950  0.0476417 -2.9595  0.003098 **
hisp         0.0160861  0.0426212  0.3774  0.705880
exper        0.0579448  0.0025026 23.1537 < 2.2e-16 ***
married      0.0757793  0.0167533  4.5232 6.252e-06 ***
union        0.1100202  0.0179187  6.1400 8.991e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    657.83
Residual Sum of Squares: 546.44
R-Squared:      0.16934
Adj. R-Squared: 0.16907
F-statistic: 147.903 on 6 and 4353 DF, p-value: < 2.22e-16
```

Berkeley SCHOOL OF INFORMATION