Fixed and Random Effects Models

A. Introduction

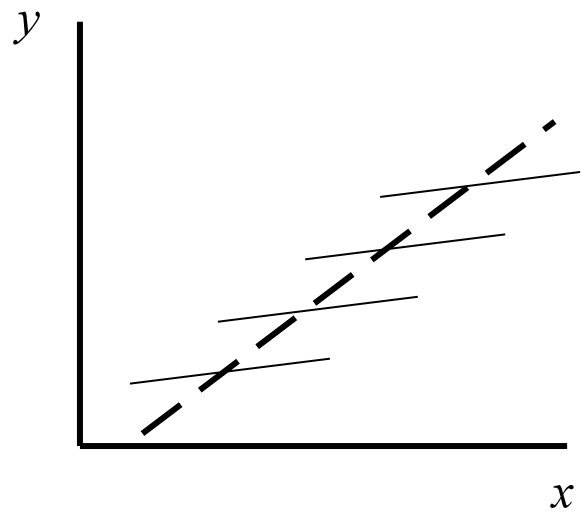
1. consider a model of the form

$$y_{it} = X_{it}\beta + u_{it} \qquad \text{where } u_{it} = \alpha_i + \varepsilon_{it}$$

   for $i = 1, N$ and $t = 1, T$. Let $E(\alpha_i) = E(\varepsilon_{it}) = 0$, $Var(\alpha_i) = \sigma_\alpha^2$, $Var(\varepsilon_{it}) = \sigma_\varepsilon^2$, and $E(\alpha_i\ \varepsilon_{it}) = 0$

2. the presence of α_i leads to serial correlation in the u_{it}, $E(u_{it}\ u_{is}) = \sigma_\alpha^2$ for $t \neq s$; thus, failure to account for α_i leads, at a minimum, to incorrect standard errors and inefficient estimation

3. if α_i is correlated with x_{it}, failure to account for α_i leads to heterogeneity (omitted variables) bias in the estimate of β; to see this consider the following illustration



Heterogeneity Bias

B. Fixed Effects Model

1. least squares dummy variable model
   a. note that in the model above, we could rewrite the α_i terms as coefficients on a set of dummy variables indicating membership in cross-sectional unit *i* and estimate the model simply by including the appropriate dummy variables
   b. this approach is straightforward; however, for large *N*, it may be impractical to specify so many dummy variables

2. mean-differenced model
   a. let $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$
   b. similarly, define $\bar{X}_i$ as the vector of unit *i* specific means for the explanatory variables
   c. then the mean-differenced estimator is

$$\hat{\beta}_W = W_{XX}^{-1} W_{Xy} \qquad \text{and} \qquad \hat{\alpha}_W = \bar{y}_i - \bar{X}_i \hat{\beta}_W$$

   where $W_{XX} = \sum_N \sum_T \left(X_{it} - \bar{X}_i\right)' \left(X_{it} - \bar{X}_i\right)$ and

$$W_{Xy} = \sum_N \sum_T \left(X_{it} - \bar{X}_i\right)' \left(y_{it} - \bar{y}_i\right)$$

   d. the mean square error in this model is

$$s_W^2 = \left(N(T-1) - K\right)^{-1}\left(W_{yy} - W_{Xy}' W_{XX} W_{Xy}\right)$$

   where $W_{yy} = \sum_N \sum_T \left(y_{it} - \bar{y}_i\right)^2$ and *K* is the number of columns in x_{it}

3. specification test
   a. test of H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_N$
   b. run OLS and fixed effects versions of model; test statistic is

$$F = \frac{\left(R_W^2 - R_{OLS}^2\right)/(N-1)}{\left(1-R_W^2\right)/\left(N(T-1)-K\right)} \sim F_{N-1,N(T-1)-K}$$

4. limitations of fixed effects approach
   a. cannot estimate effects of variables which vary across individuals but not over time
   b. "blunderbuss" approach to controlling for omitted variables – knocks out all cross-section variation in the dependent and independent variables
   c. cannot predict effects in levels outside of sample; prediction in levels requires prediction of the fixed effects
   d. use of fixed effects is inefficient if α_i is uncorrelated with x_{it} (i.e., if appropriate model is random effects)
   e. use of fixed effects can exacerbate biases from other types of specification problems, especially measurement error

C. Random Effects Model

1. specification of model
   a. consider a slightly respecified version of the model

$$y_{it} = \alpha_0 + X_{it}\beta + \alpha_i + \varepsilon_{it}$$

b. in addition, assume that the α_i are unobserved random variables which follow a probability distribution known up to some finite set of parameters
c. also assume

$$E(\alpha_i) = E(\varepsilon_{it}) = 0 \quad Var(\alpha_i) = \sigma_\alpha^2 \quad Var(\varepsilon_{it}) = \sigma_\varepsilon^2$$
$$E(\alpha_i\,\varepsilon_{it}) = 0 \quad E(\varepsilon_{it}\,\varepsilon_{js}) = 0 \quad E(\alpha_i\,\alpha_j) = 0$$

d. can write the covariance matrix as

$$E\left(u_i u_i'\right) = \Omega = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

e. a generalized least squares procedure is possible if we can transform the dependent and independent variables by $\Omega^{-1/2} = I - \frac{\theta}{T} ii'$ where $\theta = 1 - \sigma_\varepsilon \big/ \sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}$ i.e. run a regression with $\tilde{y}_{it} = y_{it} - \theta \bar{y}_i$ and $\tilde{X}_{it} = X_{it} - \theta \bar{X}_i$ as the dependent and independent variables
f. sometimes described as the "quasi-differenced estimator"

2. feasible generalized least squares (FGLS) estimation
   a. run fixed effects regression to obtain $\hat{\sigma}_\varepsilon^2 = s_W^2$

b. use slope coefficient from any consistent regression (e.g. OLS) to form $e_i^* = \bar{y}_i - \hat{\alpha}_0 - \bar{X}_i\hat{\beta}$ and $s^{*2} = (N-K)^{-1}\sum_N e_i^{*2}$

c. then $\hat{\sigma}_\alpha^2 = s^{*2} - \frac{\hat{\sigma}_\varepsilon^2}{T}$; note that it is possible for this estimator to be negative

3. Breusch and Pagan (1980) specification test
   a. Lagrange multiplier test based on OLS residuals
   b. test statistic is

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_N\left(\sum_T \varepsilon_{it}\right)^2}{\sum_N\sum_T \varepsilon_{it}^2} - 1\right]^2 \sim \chi_1^2$$

D. Fixed or Random Effects

1. key consideration is the orthogonality of α_i
   a. if α_i is uncorrelated with the variables in x_{it}, then random effects is the appropriate estimator
   b. if α_i is correlated with the variables in x_{it}, then the fixed effects model is appropriate

2. can be examined using a Hausman-Wu test
   a. run both FE and RE models
   b. test statistic is

$$\left[\hat{\beta}_{FE} - \hat{\beta}_{RE}\right]'\left[\mathrm{Var}(\hat{\beta}_{FE}) - \mathrm{Var}(\hat{\beta}_{RE})\right]^{-1}\left[\hat{\beta}_{FE} - \hat{\beta}_{RE}\right] \sim \chi_K^2$$

E. Two-way Fixed Effects Model

1. specification of model:

$$y_{it} = \alpha_0 + X_{it}\beta + \alpha_i + \gamma_t + \varepsilon_{it}$$

here the model includes both individual-specific effects α_i and period-specific effects γ_t

2. fixed effects estimator
   a. let $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$, $\bar{y} = NT^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$, and $y_{it}^{*} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$
   b. similarly, construct X_{it}^{*}
   c. and regress y_{it}^{*} on x_{it}^{*} and adjust for the appropriate degrees of freedom
   d. note that if either N or T is small, it may be easier to run one-way fixed effects and add dummy variables

F. Stata code

1. one-way fixed effects models – the model specification is

xtreg dependent_variable
list of independent variables**, fe i**(index_var**)**

where the index_var is a variable indicating membership in a group (either *i* for individual or *t* for time fixed effects); alternative the index could be set earlier in the program using an **iis** command

2. one-way random effects model – the model specification is

   xtreg dependent_variable
   list of independent variables**, re i(**index_var**)**

3. specification tests:
   a. **xttest0** command used after random effects specification conducts the Breusch-Pagan specification test
   b. **hausman** command used after random effects specification conducts the Hausman specification test
      – after the fixed effects regression, include the command **est store fixed**
      – after the random effects regression, include the command **hausman fixed .**

References

Greene, William. *Econometric Analysis*, 3rd Edition, Upper Saddle River, N.J.: Prentice-Hall, 1997, chapter 14.

Homework:

1. Using the program *afdc_fere.do* from the class website, run an OLS, random effects and fixed effects regression of the determinants of the natural log of maximum AFDC benefits (lafdc) using 1982 and later data from the Ribar-Wilhelm RESTAT paper. The explanatory variables for the model should be
   - the log of the price of transferring income (lprice)
   - the log of the average state income (linc)
   - an index of liberal voting sentiments (xada)
   - the percent of the state population that is black (xpcbl)
   - the percent of the state population that is over age 65 (xage65)
   - the percent of the state population that is under age 14 (xpc14un)
   - the percent of the state population with a high school education (xpchs)
   - the percent of the state population with a college education (xpcco)

2. Run and interpret the Breusch-Pagan and Hausman specification tests.

3. Modify the program to add dummy variable controls for the years 1983-1992. Re-run each of the models and each of the tests. Re-interpret the results.