# L12: Panel Data Regression Models

**Feng Li**
feng.li@cufe.edu.cn

**School of Statistics and Mathematics**
**Central University of Finance and Economics**

Revision: December 15, 2014

# Today we are going to learn...

1. **Panel data**
2. **The fixed effect model**
3. **The random effect model**
4. **Mixed model**

# Types of data

- time series data
- cross sectional data
- panel data
  - also known as **pooled data**, **longitudinal data**, **micropanel data**.
  - pooling of time series and cross sectional observations.
  - e.g. survey of a company's financial status over years.

# The advantage of pooled data

- Panel data can enrich empirical analysis that not be possible if only cross-sectional or time series data we used.
- Bear in mind that panel data analysis is not the all-purpose tool for data modeling.

# The fixed effect model

- The airlines example
  - The usual regression model is

$$C_{it} = \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it}$$

    where $i = 1, ..., 6$ is the airline id, and $t = 1, ..., 15$ is the year id, $C$ is the total cost, $Q$ is output, $PF$ is fuel price, and $LF$ is load factor.
  - Let's assume the error term follows Gaussian Markov assumption at the moment.
  - If we pool all the airline id together, we can estimate the model in the usual way. **Table 16.2**.
- Now let's change the model to

$$C_{it} = \beta_i + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it}$$

  which allows the intercept to be different among different airline id.
- The previous model is called **fixed effect** model
  - The intercept is different from subjects.
  - But the intercept is invariant over time (the effect over time is fixed).
- How do we estimate the fixed effect?

# The fixed effect model

## ↪ Use dummy variable to estimate the fixed effect

- The fixed effect can be estimated by the dummy variables.
- We can write the fixed effect model in terms of dummies

$$C_{it} = \alpha_1 + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \alpha_5 D_{5t} + \alpha_6 D_{6t} + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u$$

where

$$D_{it} = \begin{cases} 1, & \text{airline i} \\ 0, & \text{otherwise} \end{cases}$$

- The old questions
  - How to interpret the dummies?
  - Why only five dummies.
- Now you can estimate the fixed effects by using OLS. **Table 16.3**

# The random effect model

- Recall the airline example.

$$C_{it} = \beta_i + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it}$$

- We now assume that the intercept $\beta_i$ is a random variable with a mean value of $\beta_1$ with a random error term $\epsilon_i$

$$\beta_i = \beta_1 + \epsilon_i$$

where $\epsilon_i$ has mean zero and variance $\sigma_\epsilon^2$ (this variance is different from the variance in error term $u_{it}$)

- This indicates the previous model can be re-expressed as

$$\begin{aligned} C_{it} =& (\beta_1 + \epsilon_i) + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \\ =& \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + (\epsilon_i + u_{it}) \\ =& \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + w_{it} \end{aligned}$$

- This is called the **random effect model**.

# The random effect model

## ↪ The assumption

- The assumption

$$
\begin{aligned}
&\epsilon_i \sim N(0, \sigma_\epsilon^2) \\
&u_{it} \sim N(0, \sigma_u^2) \\
&E(\epsilon_i u_{it}) = 0 \\
&E(\epsilon_i \epsilon_j) = 0, \ i \neq j \\
&E(u_{it} u_{is}) = E(u_{ij} u_{ij}) = E(u_{it} u_{js}) = 0 \ (i \neq j, t \neq s)
\end{aligned}
$$

- This yields

$$
\begin{aligned}
E(w_{it}) &= 0 \\
var(w_{it}) &= \sigma_\epsilon^2 + \sigma_u^2
\end{aligned}
$$

- And $w_{it}$ and $w_{is}$ are correlated

$$
\rho = cor(w_{it}, w_{is}) = \frac{\sigma_\epsilon^2}{\sqrt{\sigma_\epsilon^2 + \sigma_u^2}}
$$

## The random effect model

## ↪ The estimation method

- Because the new error terms are correlated, OLS will be inefficient. GLS can be used.
- Example **Table 16.5**

# Which model is better, fixed/random effect?

- There is no simple rule.
- We assume that the error variance is the same for all cross- section units in fixed effect model.
- The assumptions underlying random effect model is that the $\epsilon_i$ are a random drawing from a much larger population.
- If it is assumed that $\epsilon_i$ and the $X$ are uncorrelated, random effect model may be appropriate, whereas if $\epsilon_i$ and the $X$ are correlated, fixed effect model may be appropriate.

# Practical guide

- If $T$ (the number of time series data) is large and $N$ (the number of cross-sectional units) is small, there is likely to be little difference in the values of the parameters estimated by fixed effect model and random effect model. Hence the choice here is based on computational convenience. On this score, fixed effect model may be preferable.
- When $N$ is large and $T$ is small, the estimates obtained by the two methods can differ significantly.
- If the individual error component $\epsilon_i$ and one or more regressors are correlated, then the random effect model estimators are biased, whereas those obtained from fixed effect model are unbiased.
- If $N$ is large and $T$ is small, and if the assumptions underlying random effect model hold, random effect model estimators are more efficient than fixed effect model estimators.

# Mixed model

- A mixed model is a statistical model containing both fixed effects and random effects.
- In matrix notation a mixed model can be represented as

$$y = X\beta + Zu + \epsilon$$

  where $y$ is a vector of observations, with mean $E(y) = X\beta$, $\beta$ is a vector of fixed effects, $u$ is a vector of random effects with mean $E(u) = 0$ and variance-covariance matrix $\text{var}(u) = G$, $\epsilon$ is a vector of IID random error terms with mean $E(\epsilon) = 0$ and variance $\text{var}(\epsilon) = R$, $X$ and $Z$ are matrices of regressors relating the observations $y$ to $\beta$ and $u$.
- The solution to the mixed model is usually through a maximum likelihood estimate.

# Take home questions

- **16.7, 16.8, 16.9**
- Further read for an brief introduction to linear mixed model `http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf`