Panel Data Econometrics I (PhD)
Static Linear Models

Yonas Alem

Department of Economics
University of Gothenburg

April 19, 2017

Panel Data Models: Lecture Outline

- Static Linear Models (Wooldridge 2010: ch. 10-11)
    - Introduction
    - OLS
    - Random Effects
    - Fixed Effects
    - First Difference
    - Comparison of models
    - Hausman-Taylor Model
- Dynamic Linear Models (Bond 2002)
    - Autoregressive Models
        - Anderson and Hsiao 2SLS Estimator
        - Arellano and Bond GMM Estimator
    - Autoregressive-Distributed Lag Models
        - Persistent Series
        - Blundell and Bond System GMM Estimator
    - Application: Estimation of a Dynamic Production Function

Panel Data Models: Lecture Outline Cont.

- Non-linear Static Models (Wooldridge 2010, Ch. 15)
  - Introduction
  - Random Effects Probit
  - Chamberlain's RE Probit
  - Fixed Effects Logit
- Dynamic Models and the Initial Conditions Problem (Arulampalam and Stewart, 2009; Stewart, 2006; Wooldridge 2005)
  - Heckman's Estimator
  - Wooldridge's Conditional Maximum Likelihood Estimator

Panel Data Models: Lecture Outline Cont.

References

- Arulampalam, W., and Stewart, MB. 2009. Simplified Implementation of the Heckman Estimator of Dynamic Probit Model and ..... *Oxf. Bulletin of Econ. and Stat.* 71: 659-681.
- Bond S. 2002. Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. Cemmap WP, No. CWP09/02.
- Stewart MB. 2006. -redprob- A Stata Program for the Heckman Estimator of the Random Effects Dynamic Probit Model, University of Warwick.
- Wooldridge JM. 2005. Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *J. of App. Econometrics* 20: 39-54.
- Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Introduction
Static Linear Models - Reference

- Wooldridge JM. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, Ch. 10

Introduction

Panel Data - Features and Advantages

- A panel data set contains repeated observations over the same units (individuals, households, firms, countries) collected over a number of periods
  - Can be micro or macro level
  - Balanced (all units observed in all periods), or unbalanced (not all units observed in all periods due to attrition)
  - Allows estimation of more complicated and more realistic models than a single cross-section or a single timeseries data
    - E.g., Dynamic models: require at least three rounds of data
- Some practical limitations:
  - Difficult to assume that different observations of an individual or a household are independent, which results in some complications in non-linear and dynamic models.
  - Very often, panel data suffer from missing observations (attrition).

Table: Panel Data - Example

id	t	y	x1	x2
1	2000	250	1	45
1	2001	275	1	46
1	2002	322	1	47
2	2000	500	0	29
2	2001	550	0	30
2	2002	600	0	31
3	2000	175	1	36
3	2001	225	1	37
3	2002	305	1	38

Introduction
The Omitted Variables Problem

- We focus on panel data where N is large and T is small often found in microeconometric data
- Panel Data can reduce the problem of **omitted variables**
- Let y and $\boldsymbol{x} \equiv (x_1, x_2, ..., x_K)$ be observable random variables and c be unobservable random variable
- The vector $(y, x_1, x_2, ..., x_K, c)$ represents the population of interest
- One is interested in the partial effects of x_j (holding c constant) in the population regression function

$$E(y|x_1, x_2, ..., x_K, c) \tag{1}$$

- Conventionally used alternative notations for c are α, η and ϕ

Introduction
The Omitted Variables Problem Cont.

- Thus, the complete linear model of interest:

$$E(y|\boldsymbol{x}, c) = \beta_0 + \boldsymbol{x}\boldsymbol{\beta} + c, \tag{2}$$

- where our interest is in $\boldsymbol{\beta}$, a $K \times 1$ vector
- If $Cov(x_j, c) \neq 0$ for some j, incorporating c in the error term would be problematic (i.e.,$\boldsymbol{\beta}$ would be inconsistent)
- With additional assumptions, one could address the problem
- In cross-sectional models: (i) a suitable proxy for c, (ii) Instrumental Variables, (iii) indicators for c
- With panel data, other possibilities arise.

Introduction

The Omitted Variables Problem Cont.

- Consider two periods $(t = 1, 2)$, and assume that c is time-invariant (a very useful assumption)

$$E(y_t | \boldsymbol{x}_t, c) = \beta_0 + \boldsymbol{x_t \beta} + c, \qquad t = 1, 2 \qquad (3)$$

- Where $\boldsymbol{x_t \beta} = \beta_1 x_{t1} + ... + \beta_K x_{tK}$ and x_{tj} indicates variable j at time t.
- Let's set a coefficient of 1 to c, b/c it doesn't have a unit of measurement and estimating its partial effect does not give senses
- The assumption that c is constant (time-invariant) is crucial to the following analysis
- Captures factors such cognitive ability, motivation, or early family upbringing (for individuals), and e.g., managerial quality (in the case of firms), which are assumed to be given.

Introduction

The Omitted Variables Problem Cont.

- Rewrite (3) in error components form:

$$y_t = \beta_0 + \boldsymbol{x_t}\beta + c + u_t \tag{4}$$

- Assume:

$$E(u_t|\boldsymbol{x_t}, c) = 0, \qquad t = 1, 2. \tag{5}$$

- One implication of (5) is:

$$E(\boldsymbol{x_t'}u_t) = 0, \qquad t = 1, 2. \tag{6}$$

- If $E(\boldsymbol{x_t'}c) = 0$, one could apply OLS otherwise, OLS would be biased and inconsistent

Introduction
The Omitted Variables Problem Cont.

- If $E(\boldsymbol{x}_t'c) \neq 0$, one can use fixed effects, first differences, or random effects to deal with c provided some conditions hold (we will see this shortly)

Assumptions about c

Random or Fixed Effects?

- The basic **unobserved effects model (UEM)** is given as:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \qquad t = 1, 2, ..., T, \tag{7}$$

- Where $\boldsymbol{x}_{it}$ is $1 \times K$ and contains both time-variant and time-invariant variables
- Other names for c_i: **unobserved component, latent variable, unobserved heterogeneity**
- The u_{it} are called the **idiosyncratic errors** or **idiosyncratic disturbances** (change across t as well as across i).
- c_i is called a "random effect" when treated as a random variable and a "fixed effect" when it is treated as a parameter to be estimated for each i

Assumptions about c

Random or Fixed Effects?

- With a large number of random draws from the cross section, it makes sense to treat c_i as random draws from the population along with y_{it} and $\boldsymbol{x}_{it}$.
- A key issue would be whether:
  - $Cov(\boldsymbol{x}'_{it}, c_i) = 0, \quad t = 1, 2, ..., T.$
- Statistical inference also requires the assumption:
  - $E(c_i|\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT}) = E(c_i)$
- Mundlak (1978): a RE framework is synonymous with $Cov(\boldsymbol{x}_{it}, c_i = 0), \quad t = 1, 2, ..., T.$
- Note: FE framework $\neq$ c_i is nonrandom! Rather it is allowed to be correlated arbitrarily with $\boldsymbol{x}_{it}$

Assumptions about c

Random or Fixed Effects?

- More recently, the Mundlak's **Correlated Random Effects (CRE)** framework has become popular (especially for non-linear models)
  - Allows correlation between c_i and $\boldsymbol{x}_i$ with some restricted specification
- Key difference with the Fixed Effects (FE) approach
  - FE: the relationship between c_i and x_i is left entirely unspecified
  - CRE: restricts the dependence in some way (more often used in non-linear models)

Assumptions about c and x_{it}

Strict Exogeneity

- The strict exogeneity assumption

$$E(y_{it}|\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},...,\boldsymbol{x}_{iT},c_i) = E(y_{it}|\boldsymbol{x}_{it},c_i) = \boldsymbol{x}_{it}\boldsymbol{\beta}+c_i, \qquad t = 1,2,...,T. \tag{8}$$

- Eq. (8) implies that, once $\boldsymbol{x}_{it}$ and c_i are controlled for, $\boldsymbol{x}_{is}$ has no partial effect on y_{it} for $s \neq t$
  - Eq. (8) $\Longrightarrow \{\boldsymbol{x}_{it} : t = 1,2,...,T\}$ are **strictly exogenous conditional on** c_i
- Without conditioning on c_i, the strict exogeneity assumption is given as:

$$E(y_{it}|\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},...,\boldsymbol{x}_{iT}) = E(y_{it}|\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}\boldsymbol{\beta}, \qquad t = 1,2,...,T. \tag{9}$$

- But in a panel data context, (9) is unreasonable because of the presence of c_i!

Assumptions about c and x_{it}

Strict Exogeneity

- Given (7), the strict exogeneity assumption can also be stated as

$$E(u_{it}|\boldsymbol{x}_{i1},...,\boldsymbol{x}_{iT},c_i) = 0, \quad t = 1,2,...,T \Longrightarrow \tag{10}$$

$$E(\boldsymbol{x}'_{is}u_{it}) = 0, \quad s,t = 1,...,T. \tag{11}$$

- This assumption is much stronger than the *zero contemporaneous correlation* assumption given by

$$E(\boldsymbol{x}'_{it}u_{it}) = 0, \quad t = 1,...,T.$$

Estimation
OLS

- Rewrite (7)

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \qquad t = 1, 2, ..., T, \tag{12}$$

- Where $v_{it} \equiv c_i + u_{it}, t = 1, 2, ..., T$, called the **composite errors**

$$\hat{\boldsymbol{\beta}_{OLS}} = (\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i)^{-1} (\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{y}_i) \tag{13}$$

- $\hat{\boldsymbol{\beta}_{OLS}}$ would be consistent if $E(\boldsymbol{x}_{it}' v_{it}) = 0, t = 1, 2, ..., T. \Longrightarrow \quad E(\boldsymbol{x}_{it}' u_{it}) = 0, \quad and$

$$E(\boldsymbol{x}_{it}' c_i) = 0, \qquad t = 1, 2, ..., T. \tag{14}$$

- Even if (14) holds, $E(v_{it} v_{is}) \neq 0$ for $s \neq t$ why?
- One therefore needs robust variance matrix estimator and robust test statistics.

Estimation
Random Effects

- Key assumptions for consistency: Strict Exogeneity, and Orthogonality between c_i & $\boldsymbol{x}_{it}$

$$RE.1(a)E(u_{it}|\boldsymbol{x}_i, c_i) = 0, t = 1, ..., T\;;(b)E(c_i|\boldsymbol{x_i}) = E(c_i) = 0 \tag{15}$$

- Where $\boldsymbol{x}_i \equiv (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iT})$
- Exploits the serial correlation in v_{it} in a GLS framework (hence assumption (15) is slightly stronger than needed for OLS)
- Under RE.1, write:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \tag{16}$$

$$E(v_{it}|\boldsymbol{x}_i) = 0, \qquad t = 1, 2, ..., T, \tag{17}$$

- Where

$$v_{it} = c_i + u_{it}. \tag{18}$$

Estimation
Random Effects Cont.

- Write (16) for all T as:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{v}_i \tag{19}$$

- And $\boldsymbol{v}_i = c_i\boldsymbol{j}_T + \boldsymbol{u}_i$, where $\boldsymbol{j}_T$ is a $T \times 1$ vector of ones
- Define the variance matrix of $\boldsymbol{v}_i$

$$\boldsymbol{\Omega} \equiv E(\boldsymbol{v}_i\boldsymbol{v}_i') \tag{20}$$

- a $T \times T$ positive definite matrix (the same $\forall \quad i$) because of the random sampling assumption
- Consistency requires, RE.2: rank $E(\boldsymbol{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i) = K$ and more assumptions:

$$E(u_{it}^2) = \sigma_u^2, \quad t = 1, 2, ..., T. \tag{21}$$

$$E(u_{it}u_{is}) = 0, \quad all \;\; t \neq s \tag{22}$$

Estimation
Random Effects Cont.

- Under RE1 (a), $E(c_i u_{it}) = 0, t = 1, 2, ..., T \Longrightarrow$
- $E(v_{it}^2) = E(c_i^2) + 2E(c_i u_{it}) + E(u_{it}^2) = \sigma_c^2 + \sigma_u^2,$
- $E(v_{it} v_{is}) = E[(c_i + u_{it})(c_i + u_{is})] = E(c_i^2) = \sigma_c^2$
- RE.1, (21), & (22) $\Longrightarrow$

$$\boldsymbol{\Omega} = E(\boldsymbol{v}_i \boldsymbol{v}_i^{'}) = \tag{23}$$

$$\begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & ... & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & ... & \\ \sigma_c^2 & & & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$

$$\boldsymbol{\Omega} = \sigma_u^2 \boldsymbol{I}_T + \sigma_c^2 \boldsymbol{j}_T \boldsymbol{j}_T^{\prime} \tag{24}$$

Estimation
Random Effects Cont.

- In $(24), \mathbf{\Omega}$ has the **random effects structure**: depends on σ_c^2 & σ_u^2 regardless of T
- $Corr(v_{is}, v_{it}) = \sigma_c^2/(\sigma_c^2 + \sigma_u^2) \geq 0, s \neq t$ useful to measure the relative importance of c_i
- For efficiency of the FGLS, we assume that the variance matrix of $\boldsymbol{v}_i$ conditional on $\boldsymbol{x}_i$ is constant:

$$E(\boldsymbol{v}_i\boldsymbol{v}_i'|\boldsymbol{x}_i) = E(\boldsymbol{v}_i\boldsymbol{v}_i') \tag{25}$$

- Assumptions (21), (22) and (25) $\Longrightarrow$ RE.3: $(a)E(\boldsymbol{u}_i\boldsymbol{u}_i'|\boldsymbol{x}_i, c_i) = \sigma_u^2\boldsymbol{I}_T; (b)E(c_i^2|\boldsymbol{x}_i) = \sigma_c^2$
- To implement an FGLS model: define $\sigma_v^2 = \sigma_c^2 + \sigma_u^2$. Then

$$\hat{\mathbf{\Omega}} \equiv \hat{\sigma_u^2}\boldsymbol{I}_T + \hat{\sigma_c^2}\boldsymbol{j}_T\boldsymbol{j'}_T, \tag{26}$$

Estimation
Random Effects Cont.

- a positive definite $T \times T$ matrix. The FGLS estimator that uses (26) is what is called the **random effects estimator**.

$$\hat{\beta_{RE}} = (\sum_{i=1}^{N} \mathbf{X}_i' \hat{\Omega}^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^{N} \mathbf{X}_i' \hat{\Omega^{-1}} \mathbf{y}_i) \tag{27}$$

- RE.1-RE.3. $\implies RE \equiv GLS$ and is efficient
- One needs $\hat{\sigma}_c^2$ and $\hat{\sigma}_u^2$ to implement RE. Easy to start with $\hat{\sigma}_v^2 = \hat{\sigma}_c^2 + \hat{\sigma}_u^2$, and use the OLS β denoted as $\tilde{\beta}$. Thus, a consistent estimator of σ_v^2

$$\hat{\sigma}_v^2 = \frac{1}{(NT - K)} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{v}_{it}^2 \tag{28}$$

Estimation

Random Effects Cont. Robust Variance

- A consistent estimator of σ_c^2: Note $\sigma_c^2 = E(v_{it}v_{is}), \forall t \neq s$.

$$\hat{\sigma}_c^2 = \frac{1}{[NT(T-1)/2 - K]} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \tilde{v}_{it}\tilde{v}_{is} \qquad (29)$$

- Given $\hat{\sigma}_v^2$ and $\hat{\sigma}_c^2 \implies \hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$
- Other methods of estimation: FE (for σ_u), the between estimator (for σ_c^2)
- If (29) is negative, there is a negative serial correlation in u_{it}, violating RE.3(a). Not controlling for time dummies in estimation is one main reason!

Estimation

Random Effects Cont. Robust Variance

- RE.3 may fail for two reasons (but RE would still be consistent)
  1. $E(\boldsymbol{v_i}\boldsymbol{v_i}'|\boldsymbol{x_i})$ may not be constant. Possible in GLS
  2. $E(\boldsymbol{v_i}\boldsymbol{v_i}')$ may not have the RE structure. u_{it} my have variances that change over time, or they could be serially correlated
- One could use a robust variance analysis of the GLS with RE residuals under either case
- But a more general estimator of $\boldsymbol{\Omega}$ can be used in FGLS when the u_{it} are generally heteroskedastic and serially correlated

$$\hat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}_i}' \quad (30)$$

- Where $\hat{\boldsymbol{v}}_i$ is the pooled OLS residuals. Consistent under RE.1 and RE.2, and is asymptotically efficient if $E(\boldsymbol{v_i}\boldsymbol{v_i}'|\boldsymbol{x_i}) = \boldsymbol{\Omega}$

Estimation

Random Effects Cont. Robust Variance

- Eq. (30) is more general and as efficient as the RE under RE.1-RE.3, and even asymptotically more efficient if $E(v_i v_i' | x_i) = \mathbf{\Omega}$, but $\mathbf{\Omega}$ does not have the RE form
- But if N is not several times larger than T, an unrestricted FGLS analysis can have poor finite sample properties!

Estimation

Random Effects Cont. Testing for the presence of c_i

- If $\sigma_c^2 > 0$, and $E(\boldsymbol{x}'_{it} c_i) = 0$, RE is efficient
- If RE.1-RE.3 hold, but c_i does not exist, pooled OLS is efficient
- Testing for this is equivalent to testing $H_0 : \sigma_c^2 = 0$
- The conventional test for this is the Lagrange multiplier test of Breusch and Pagan (1980)
- One can use the OLS residuals to compute the test statistic

$$LM = \frac{NT}{2(T-1)} [\frac{\sum_i (\sum_t \hat{v}_{it})^2}{\sum_i \sum_t \hat{v}_{it}^2} - 1]^2 \quad (31)$$

- Under the null, LM is distributed χ^2 with one degree of freedom.

Estimation

Random Effects Cont. Testing for the presence of c_i

- If $H_0 : \sigma_c^2 = 0, \Longrightarrow (\sum_t \hat{v}_{it}^2) = \sum_t \hat{v}_{it}^2 \Longrightarrow$ the term in the square bracket will be zero
- If H_0 is not true, i.e., $\sigma_c^2 > 0$, then $(\sum_t \hat{v}_{it}^2) > \sum_t \hat{v}_{it}^2 \Longrightarrow LM > 0.$

Estimation
Fixed Effects Method

- Consider the linear unobserved effects model for T time periods

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \qquad t = 1, 2, ..., T, \tag{32}$$

- FE, unlike RE allows for correlation between c_i and $\boldsymbol{x}_{it}$. Two key assumptions:
  1. FE.1: c_i does not have to be orthogonal to $\boldsymbol{x}_{it}$
  2. FE.2: $E(u_{it}|\boldsymbol{x}_i, c_i) = 0, \;\; t = 1, 2, ..., T.$ **(Strict Exogeneity)**
- Average eq. (32) over $t = 1, ..., T$ to get:

$$\bar{y}_i = \bar{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\beta} + c_i + \bar{u}_i \tag{33}$$

- Where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}, \bar{\boldsymbol{x}}_{\boldsymbol{i}} = T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{it}$, and $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$
- Subtracting 33 from 32 gives

$$(y_{it} - \bar{y}_i) = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{\boldsymbol{i}})\boldsymbol{\beta} + (u_{it} - \bar{u}_i) \tag{34}$$

Estimation
Fixed Effects Method Cont.

- Or
- $\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{\boldsymbol{it}}\boldsymbol{\beta} + \ddot{u}_{it} \quad t = 1, 2, ..., T, i = 1, ..., N.$
- Where $\ddot{y}_{it} = (y_{it} - \bar{y}_i), \ddot{\boldsymbol{x}}_{\boldsymbol{it}} = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{\boldsymbol{i}})$, and $\ddot{u}_{it} = (u_{it} - \bar{u}_i)$
- The transformation in (34), which removes c_i is what is known as the **With-in or fixed effects transformation**
- It is possible to estimate $\boldsymbol{\beta}$ consistently using OLS on (**??**)
- Consistency requires $E(\ddot{\boldsymbol{x}}'_{\boldsymbol{it}}\ddot{u}_{it}) = 0$, which is of course fulfilled under assumption FE.2 above.
- Also, $E(\ddot{u}_{it}|\ddot{\boldsymbol{x}}_{\boldsymbol{it}}) = 0$ showing that $\ddot{\boldsymbol{x}}_{\boldsymbol{it}}$ satisfies the conditional expectation from the strict exog. assumption in the With-in transformation.

Estimation
Fixed Effects Method Cont.

- The standard rank condition is also required for FE model to be well behaved asymptotically

$$FE.3: \quad rank(\sum_{t=1}^{T} E(\ddot{x}_{it}'\ddot{x}_{it})) = rank[E(\ddot{X}_i'\ddot{X}_i] = K \tag{35}$$

$$\hat{\beta}_{FE} = (\sum_{i=1}^{N} \ddot{X}_i'\ddot{X}_i')^{-1}(\sum_{i=1}^{N} \ddot{X}_i'\ddot{y}_i) = (\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{x}_{it})^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{y}_{it}) \tag{36}$$

- You can see that FE.3 won't hold if there are time-invariant variables (unless interacted with time-varying variables)!
- This is the major limitation of the FE model. But if the interest is not in the time-invariant variables, the model works much robustly than most of the linear models.

Estimation
Fixed Effects Method Cont.

- To ensure that FE is efficient, we need one last assumption.
- FE.4: $E(\boldsymbol{u}_i\boldsymbol{u}_i'|\boldsymbol{x}_i, c_i = \sigma_u^2 \boldsymbol{I}_T)$
- I.e., the u_{it} have a constant variance across t and are serially uncorrelated. Similar with RE but difficult to achieve in reality
- Instead of demeaning the data, one could control for dummies for the individual *is*
- i.e., regress y_{it} on $d1_i, d2_i, ..., dN_i, \boldsymbol{x}_{it}$
- $\hat{c}_i$ is now estimated and the resulting estimator is called the **Least Square Dummy Variable Model.**
- The resulting model and the residuals would be identical to the FE which works through the With-in transformation
- When N is large, the approach is not very attractive though.

Estimation
The First Differencing Method

- Applies first differencing to eliminate c_i
- Assumptions:
  1. FD.1: c_i does not have to be orthogonal to $\boldsymbol{x}_{it}$, same as FE
  2. FD.2: $E(\boldsymbol{x}_{it}' u_{is}) = 0$ where $s = t, t-1$ Weak exogeneity!
  3. FD.3: rank $(\sum_{t=2}^{T} E(\Delta \boldsymbol{x}_{it}' \Delta \boldsymbol{x}_{it}) = K$

$$(y_{it} - y_{it-1}) = (\boldsymbol{x}_{it} - \boldsymbol{x}_{it-1})\boldsymbol{\beta} + (u_{it} - u_{it-1}) \quad (37)$$

$$\Delta y_{it} = \Delta \boldsymbol{x}_{it} \boldsymbol{\beta} + \Delta u_{it} \quad (38)$$

- OLS estimator of (38) would result in a consistent estimator of $\boldsymbol{\beta}$
- Assumption FD.3, like the FE, shows that time-invariant variables are dropped in the transformation!
- Perfect collinearity may also arise in the FD model. E.g., consider a wage equation: change in experience and year dummies when we have two consecutive years of data.

Comparison of Estimators
FE or FD?

- Both are appealing methods of eliminating c_i. Which one to choose?
- If $T = 2, FE = FD$
- When $T > 2$, and strict exog. holds, the choice depends on assumption on u_{it}
- Under FE.1-FE.4 (with no serial corr in u_{it}), the FE is asymptotically efficient.
- If u_{it} follows a random walk, the FD estimator is more efficient.
- But assumption FE.4 (no serial correlation in the u_{it}), is too strong
- Many cases are however somewhere in between the two
- If the FE and FD differ significantly and if that is not due to sampling error, then we need to question the validity of the strict exog. assumption.

Comparison of Estimators
RE or FE?

- The key consideration in choosing between an RE and an FE model is the condition $E(x'_{it}c_i) = 0$
- One can test this using the **Hausman Test** is based on the difference between the RE and FE estimates
- FE is consistent when $E(\boldsymbol{x}'_{it}c_i) \neq 0$, but RE is not
- A statistically significant difference (assuming $E(x'_{it}u_{it} = 0)$) is interpreted as evidence against RE.1b $E(c_i|x_i = 0)$

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[var\hat{\beta}_{FE} - var\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (39)$$

- Which has a χ^2_M distribution under the null hypothesis (M being the number of time-varying explanatory variables

Comparison of Estimators
RE or FE? Cont.

- If the interest is on a single coefficient of a time-varying variable, a t statistic from the Hausman test could be used
- Let δ denote a scalar on the time-varying variable of interest. Then

$$t^H = \frac{\hat{\delta}_{FE} - \hat{\delta}_{RE}}{[se(\hat{\delta}_{FE})^2 - se(\hat{\delta}_{RE})^2]^{1/2}} \quad (40)$$

- Under RE.1-RE.3, the t statistics has an asymptotic standard normal distribution
- It is possible to come across rejection of RE1.b with RE and FE coefficients differing slightly.
- It is also possible to come across with large differences between FE and RE but due to large differences in se, the Hausman statistic fails to reject
- Under such cases, it is better to prefer the RE (with a risk of Type II error)

Relationship between OLS RE FE

- A very appealing way of showing the RE - GLS transformation and its relationship with OLS and FE:

$$y_{it}^{RE} = x_{it}^{RE}\beta + v_{it}^{RE} \tag{41}$$

$$y_{it}^{RE} = y_{it} - (1-\phi)\bar{y}_i \tag{42}$$

$$x_{it}^{RE} = x_{it} - (1-\phi)\bar{x}_i \tag{43}$$

$$v_{it}^{RE} = v_{it} - (1-\phi)\bar{v}_i \tag{44}$$

$$\phi = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_c^2} \tag{45}$$

- As $T \longrightarrow \infty, \phi \longrightarrow 0$, RE=FE
- If $\sigma_c^2 = 0, \phi \longrightarrow 1, RE = OLS$

Other Linear Panel Data Models

- There are alternative unobserved effects models to estimate our linear panel data model specified in (7)
    1. GMM 3SLS which uses Instrumental Variables estimation technique
    2. Chamberlain's method: based on Mundlak (1978) while maintaining the strict exogeneity assumption, allowing for arbitrary correlation between the time-varying explanatory variables and the unobserved heterogeneity term, i.e., x_{it} and c_i
- Chamberlain's approach is a more recent approach being applied especially in non-linear models

Hausman and Taylor Model

- A convenient Instrumental Variable model used when the variable of interest is time-invariant and RE can not be used
- Rewrite the basic panel data model as follows:

$$y_{it} = w_{i1}\beta_{11} + w_{i2}\beta_{12} + x_{it1}\beta_{21} + x_{it2}\beta_{22} + c_i + u_{it} \quad (46)$$

- Assume:
- $E(\boldsymbol{w}_{i1}c_i) = 0$: uncorrelated
- $E(\boldsymbol{x}_{it1}c_i) = 0$: uncorrelated
- $E(\boldsymbol{w}_{i2}c_i) \neq 0$: correlated
- $E(\boldsymbol{x}_{it2}c_i) \neq 0$: correlated
- FE or FD on (46) will not identify β_{11} and β_{12} but will identify β_{21} and β_{22}
- Hausman and Taylor suggest estimating (46) using a two-stage Instrumental Variable approach from with in the model

Hausman and Taylor Model Cont.

- The exogenous variables w_{i1} and x_{it1} serve as their own instrument
- w_{i2} is instrumented by $\hat{x}_{i1}$ (averages of the time varying exog. variables)
- $x_{it2}s$ are instrumented by their own deviations from their means $(\ddot{x}_{it2} = x_{it2} - \bar{x}_i)$
- Identification requires that number of variables in $x_{it1} \geq w_{i2}$