A Short Review on Panel Data Econometrics

Patrick GAGLIARDINI
Università della Svizzera Italiana

May 2019

OUTLINE

1. Fixed and random effects in linear panel data models

2. IV estimation of linear dynamic panel data models

3. Nonlinear panel data models

1. FIXED AND RANDOM EFFECTS IN LINEAR PANEL DATA MODELS

1.1 LINEAR PANEL DATA MODELS

Panel data are doubly indexed by individual and time:

$$y_{i,t}, \quad i = 1, ..., n, \quad t = 1, ..., T$$

Pooling cross-sectional and time series information allows to

- **i)** avoid bias from unobservable individual heterogeneity
- **ii)** distinguish common parameters from individual specific effects and time specific effects
- **iii)** study the dynamics at the micro-level
- **iv)** ...

Econometric analysis of panel data started with Mundlak (1961), (1978), Hoch (1962), Balestra, Nerlove (1966)

Two linear specifications that are extreme approaches to address **unobservable individual heterogeneity**:

i) Full homogeneity: Pooled regression

$$y_{i,t} = \alpha + x'_{i,t}\beta + \varepsilon_{i,t}$$

ii) Full heterogeneity: A system of equations

$$y_{i,t} = \alpha_i + x'_{i,t}\beta_i + \varepsilon_{i,t}$$

The linear panel data literature has mostly focused on the intermediate specification:

$$y_{i,t} = \alpha_i + x'_{i,t}\beta + \varepsilon_{i,t}$$

where β is a common parameter and the α_i are the **individual effects**

 Source: HSI

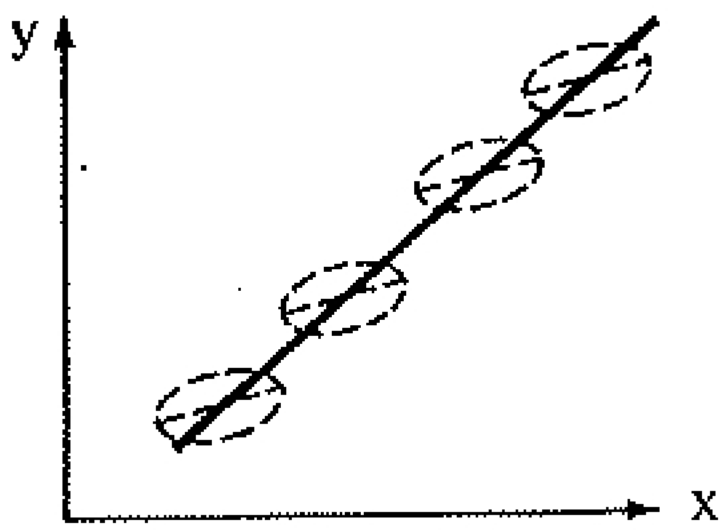


Fig. 1.1

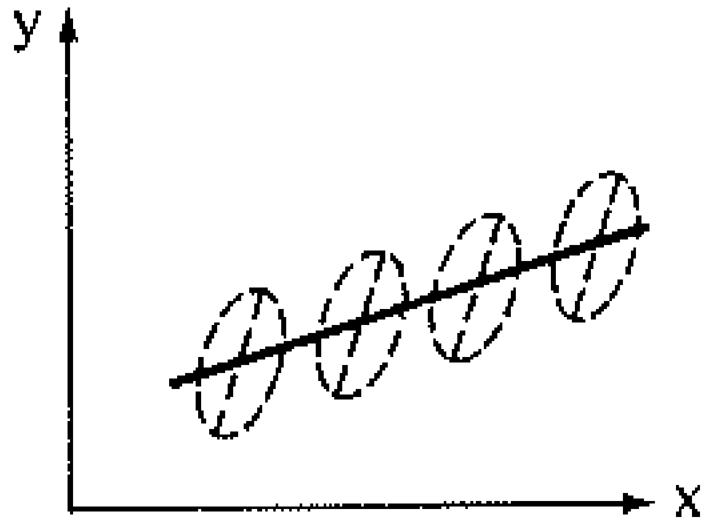


Fig. 1.2

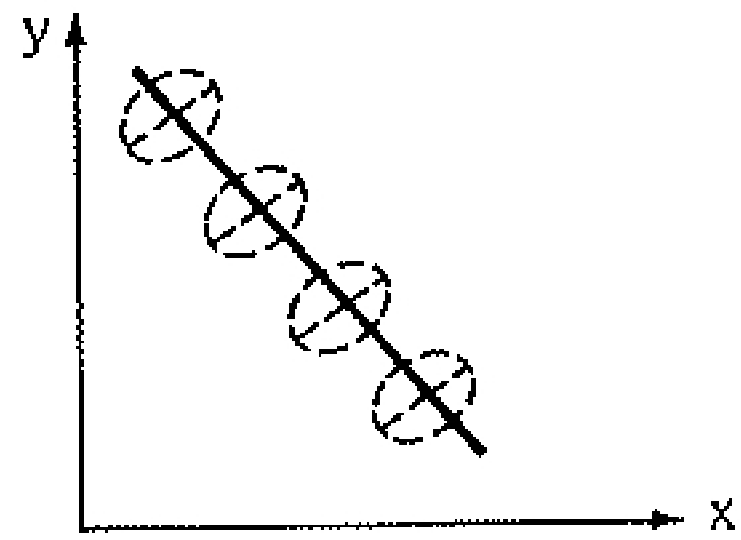


Fig. 1.3

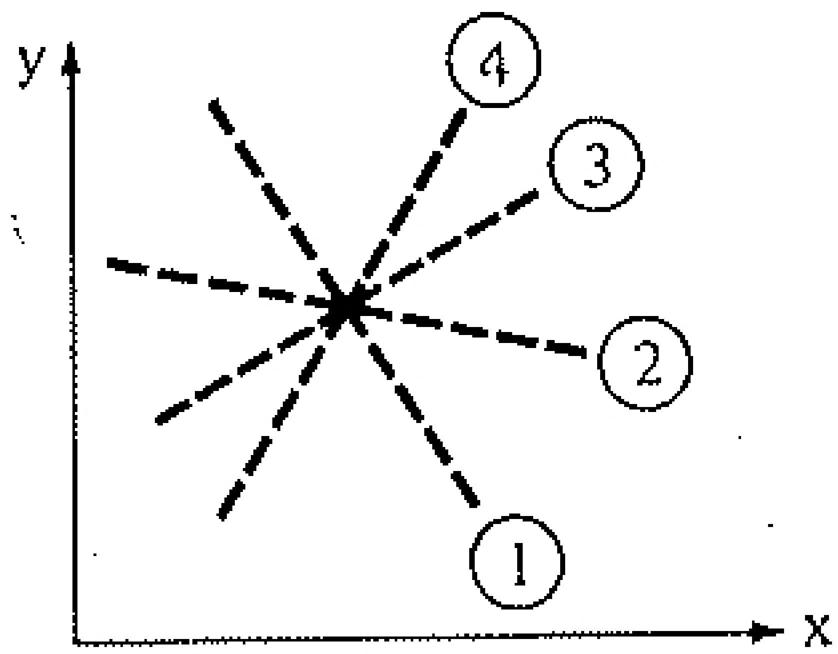


Fig. 1.4

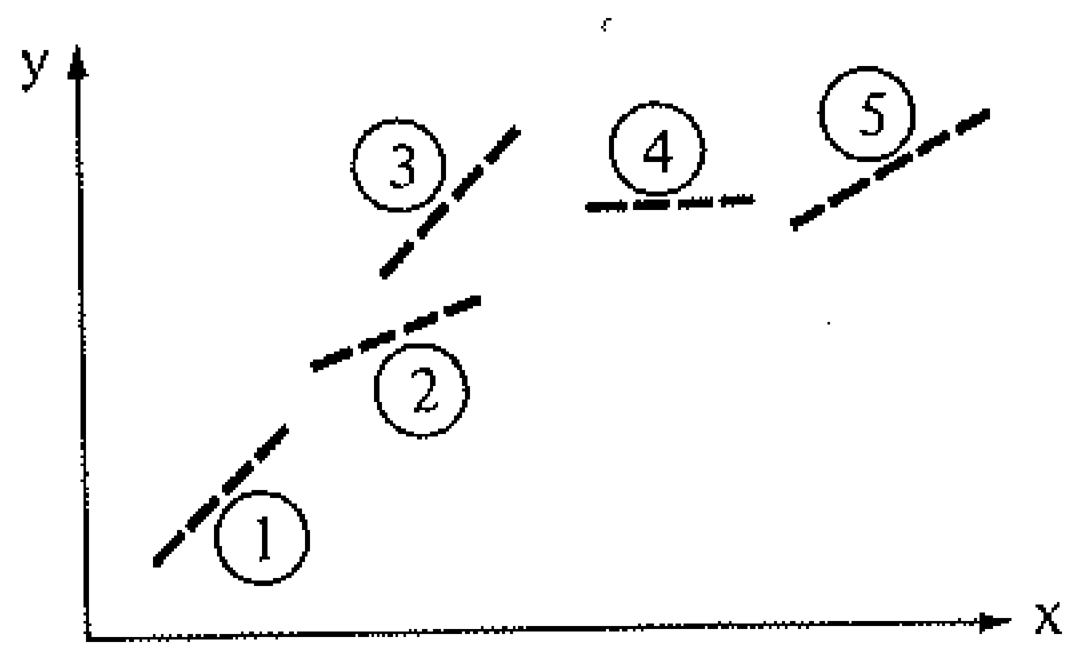


Fig. 1.5

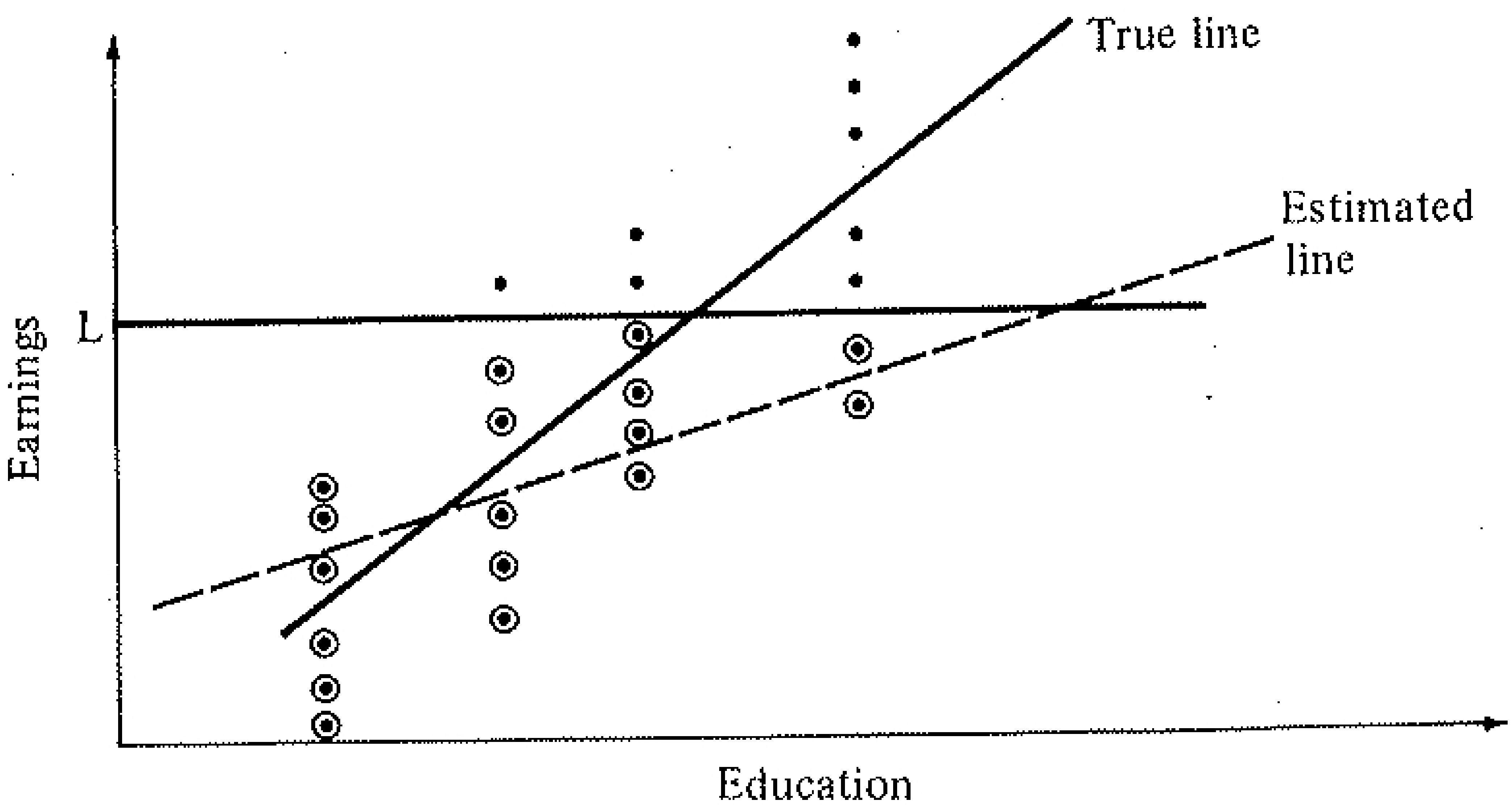


Fig. 1.6

The individual effects vary across individuals but are constant over time; they capture (parsimoniously) the **specificities of individual behaviours**

We distinguish two approaches:

i) **Fixed effects:** The α_i, for $i = 1, ..., n$, are unknown constants (to be included among the - nuisance - parameters)

ii) **Random effects:** The α_i, for $i = 1, ..., n$, are unobservable random variables (to be incorporated in the error terms)

Fixed and random effects yield different model interpretations and different estimators

1.2 FIXED EFFECTS

i) The model:

$$y_{i,t} = \alpha_i + \underset{1\times K}{x'_{i,t}} \beta + \varepsilon_{i,t}$$

where:

A.1: The errors $\varepsilon_{i,t}$ are i.i.d. across individuals and time dates, with $E[\varepsilon_{i,t}] = 0$ and $V[\varepsilon_{i,t}] = \sigma^2$, for all i and t

A.2: The regressor $x_{i,t}$ is independent of the error term $\varepsilon_{j,s}$, for all i, j and t, s

Compact form:

$$\underset{T\times 1}{y_i} = S_T \alpha_i + \underset{T\times K}{X_i} \beta + \varepsilon_i$$

where $y_i = (y_{i,1}, ..., y_{i,T})'$ and $S_T = (1, ..., 1)'$ is a T-dimensional vector of ones, and

$$\underset{nT\times 1}{y} = \underbrace{(I_n \otimes S_T)}_{\equiv D} \alpha + \underset{nT\times K}{X} \beta + \varepsilon \equiv W\gamma + \varepsilon \quad (1)$$

where $y = (y'_1, ..., y'_n)'$, $\alpha = (\alpha_1, ..., \alpha_n)'$, $\gamma = (\alpha', \beta')'$ and $W = [D \;\; X]$

Equation (1) written explicitly:

$$
\underbrace{\begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,T} \\ y_{2,1} \\ \vdots \\ y_{2,T} \\ \vdots \\ \vdots \\ y_{n,1} \\ \vdots \\ y_{n,T} \end{pmatrix}}_{=y}
=
\underbrace{\begin{pmatrix} 1 & & & & \\ \vdots & & & & \\ 1 & & & & \\ & 1 & & & \\ & \vdots & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & \vdots \\ & & & & 1 \end{pmatrix}}_{=D}
\underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_n \end{pmatrix}}_{=\alpha}
+
\underbrace{\begin{pmatrix} x'_{1,1} \\ \vdots \\ x'_{1,T} \\ x'_{2,1} \\ \vdots \\ x'_{2,T} \\ \vdots \\ \vdots \\ x'_{n,1} \\ \vdots \\ x'_{n,T} \end{pmatrix}}_{=X}
\beta +
\underbrace{\begin{pmatrix} \varepsilon_{1,1} \\ \vdots \\ \varepsilon_{1,T} \\ \varepsilon_{2,1} \\ \vdots \\ \varepsilon_{2,T} \\ \vdots \\ \vdots \\ \varepsilon_{n,1} \\ \vdots \\ \varepsilon_{n,T} \end{pmatrix}}_{=\varepsilon}
$$

ii) Least Squares Dummy Variables (LSDV) estimator: OLS on model (1)

By partitioned regression we get

$$\hat{\beta} = (X'M_D X)^{-1} X' M_D y = \left(\sum_i X_i' M_T X_i \right)^{-1} \sum_i X_i' M_T y_i \quad (2)$$

where

$$M_D = I_{nT} - D(D'D)^{-1}D' = I_n \otimes M_T, \quad M_T = I_T - \frac{1}{T} S_T S_T'$$

Interpretation of LSDV estimator: bring data in **difference from time means**

$$y_{i,t} - \bar{y}_{i\cdot} = (x_{i,t} - \bar{x}_{i\cdot})'\beta + \varepsilon_{i,t} - \bar{\varepsilon}_{i\cdot} \quad (3)$$

where $\bar{y}_{i\cdot} \equiv \frac{1}{T} \sum_t y_{i,t}$ and $\bar{x}_{i\cdot} \equiv \frac{1}{T} \sum_t x_{i,t}$, and apply pooled OLS to get

$$\hat{\beta} = \left[\sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(x_{i,t} - \bar{x}_{i\cdot})' \right]^{-1} \sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(y_{i,t} - \bar{y}_{i\cdot}) \quad (4)$$

The estimators of the fixed effects are $\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{x}_{i\cdot}'\hat{\beta}$

iii) Finite-sample properties of the LSDV estimator: follow from standard results on OLS

From A.1 and A.2 the regressors are **strictly exogenous** and the errors are spherical:

$$E[\varepsilon|X] = 0, \quad V[\varepsilon|X] = \sigma^2 I_{nT}$$

A.3: The matrix $W = [D \;\; X]$ has full column rank

For A.3 to be satisfied it is necessary that **the regressors X do not include time-invariant variables!**

Under A.1, A.2 and A.3 the LSDV estimator $\hat{\beta}$ is BLUE with variance

$$V[\hat{\beta}|X] = \sigma^2 (X' M_D X)^{-1} = \sigma^2 \left[\sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(x_{i,t} - \bar{x}_{i\cdot})' \right]^{-1}$$

Unbiased estimator of σ^2 based on the residuals $\hat{\varepsilon}_{i,t} = y_{i,t} - \hat{\alpha}_i - x'_{i,t}\hat{\beta}$:

$$\hat{\sigma}^2 = \frac{1}{n(T-1) - K} \sum_i \sum_t \hat{\varepsilon}^2_{i,t}$$

iv) Large sample properties of the LSDV estimator: depend on the asymptotic scheme

a) $n \to \infty$ and T fixed

b) n fixed and $T \to \infty$

c) $n, T \to \infty$ jointly

Consider asymptotic scheme a). Under standard regularity conditions:

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_i X_i' M_T X_i \equiv Q \quad \text{positive definite,}$$

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_i X_i' M_T \varepsilon_i = E[X_i' M_T \varepsilon_i] = 0 \quad \text{(from A.1 and A.2),}$$

$$\frac{1}{\sqrt{n}} \sum_i X_i' M_T \varepsilon_i \xrightarrow{d} N\left(0, \sigma^2 Q\right)$$

Then, by writing:

$$\hat{\beta} - \beta = \left(\frac{1}{n} \sum_i X_i' M_T X_i \right)^{-1} \frac{1}{n} \sum_i X_i' M_T \varepsilon_i$$

and:

$$\sqrt{n} \left(\hat{\beta} - \beta \right) = \left(\frac{1}{n} \sum_i X_i' M_T X_i \right)^{-1} \frac{1}{\sqrt{n}} \sum_i X_i' M_T \varepsilon_i$$

we deduce that the LSDV estimator is consistent and asymptotically normal:

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N \left(0, \sigma^2 Q^{-1} \right)$$

when $n \to \infty$ and T is fixed

Asymptotically valid standard errors and confidence intervals from

$$\widehat{AsVar}(\hat{\beta}) = \hat{\sigma}^2 \left[\sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(x_{i,t} - \bar{x}_{i\cdot})' \right]^{-1}$$

(without assuming normality of the errors!)

v) Individual and time fixed effects

The model can be extended to include both individual and time fixed effects:

$$y_{i,t} = c + \alpha_i + f_t + x'_{i,t}\beta + \varepsilon_{i,t}$$

where, to avoid the dummy variables trap, we exclude the effects for one individual and one time date (e.g. $\alpha_1 = f_1 = 0$) and include the constant c

The transformation that eliminates both individual and time effects is:

$$y_{i,t} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot} = (x_{i,t} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t} + \bar{x}_{\cdot\cdot})'\beta + \varepsilon_{i,t} - \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot t} + \bar{\varepsilon}_{\cdot\cdot} \quad (5)$$

where $\bar{y}_{\cdot t} = \frac{1}{n}\sum_i y_{i,t}$ and $\bar{y}_{\cdot\cdot} = \frac{1}{nT}\sum_i \sum_t y_{i,t}$ (see [MAS], Chapter 3)

The LSDV estimator of β is obtained by pooled OLS regression on (5)

1.3 RANDOM EFFECTS

i) The model:

$$y_{i,t} = \alpha + x'_{i,t}\beta + \varepsilon_{i,t} \equiv z'_{i,t}\gamma + \varepsilon_{i,t}$$

with the **error component** structure:

$$\varepsilon_{i,t} = u_i + v_{i,t}$$

where:

A.1: The **individual specific errors** u_i are i.i.d. across individuals with $E[u_i] = 0$ and $V[u_i] = \sigma_u^2$, for all i

A.2: The errors $v_{i,t}$ are i.i.d. across individuals and time dates with $E[v_{i,t}] = 0$ and $V[v_{i,t}] = \sigma_v^2$, for all i and t

A.3: The errors u_i and $v_{j,t}$ are independent, for all i, j and t

A.4: The regressor $x_{i,t}$ is independent of the errors u_j and $v_{j,s}$, for all i, j and t, s

Compact form: for individual i

$$\underset{T\times 1}{y_i} = S_T\alpha + X_i\beta + \varepsilon_i \equiv Z_i\gamma + \varepsilon_i,$$

with

$$E[\varepsilon_i] = 0, \quad V[\varepsilon_i] = \sigma_v^2 I_T + \sigma_u^2 S_T S_T' = \begin{pmatrix} \sigma_v^2+\sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_v^2+\sigma_u^2 & \sigma_u^2 & \\ \vdots & & \ddots & \\ \sigma_u^2 & \cdots & \sigma_u^2 & \sigma_v^2+\sigma_u^2 \end{pmatrix} \equiv \Omega,$$

and for the full sample

$$\underset{nT\times 1}{y} = S_{nT}\alpha + X\beta + \varepsilon \equiv Z\gamma + \varepsilon \qquad (6)$$

with

$$E[\varepsilon] = 0, \quad V[\varepsilon] = I_n \otimes \Omega \equiv V$$

ii) Random effects estimator: GLS estimator on model (6)

$$\hat{\gamma} = (Z'V^{-1}Z)^{-1}Z'V^{-1}y = \left(\sum_i Z_i'\Omega^{-1}Z_i\right)^{-1} \sum_i Z_i'\Omega^{-1}y_i \qquad (7)$$

where $V^{-1} = I_n \otimes \Omega^{-1}$

To invert Ω we use its **spectral decomposition**

$$\Omega = \sigma_v^2 I_T + (T\sigma_u^2)\frac{1}{T}S_T S_T' = \sigma_v^2(I_T - \frac{1}{T}S_T S_T') + (T\sigma_u^2 + \sigma_v^2)\frac{1}{T}S_T S_T' = \lambda_1 M_T + \lambda_2(I_T - M_T)$$

where $\lambda_1 \equiv \sigma_v^2$ and $\lambda_2 \equiv T\sigma_u^2 + \sigma_v^2$

Thus:

$$\Omega^{-1} = \lambda_1^{-1} M_T + \lambda_2^{-1}(I_T - M_T)$$

(use that M_T is idempotent) and

$$V^{-1} = \lambda_1^{-1} I_n \otimes M_T + \lambda_2^{-1} I_n \otimes (I_T - M_T) = \lambda_1^{-1} M_D + \lambda_2^{-1}(I_{nT} - M_D)$$

iii) Interpretation: Within and between (co-)variation

By partitioned GLS regression, the random effects estimator of parameter subvector β is

$$\hat{\beta} = (X'M_V X)^{-1} X' M_V y$$

where

$$M_V = V^{-1} - V^{-1} S_{nT} (S'_{nT} V^{-1} S_{nT})^{-1} S'_{nT} V^{-1}$$

By using $V^{-1} S_{nT} = \lambda_2^{-1} S_{nT}$ we can write

$$M_V = \lambda_1^{-1} M_D + \lambda_2^{-1} M_B$$

where

$$\begin{aligned} M_D &= I_n \otimes (I_T - \frac{1}{T} S_T S'_T) \\ M_B &= I_n \otimes \frac{1}{T} S_T S'_T - \frac{1}{nT} S_{nT} S'_{nT} \end{aligned}$$

Then:

$$\hat{\beta} = \left(\lambda_1^{-1} X' M_D X + \lambda_2^{-1} X' M_B X\right)^{-1} \left(\lambda_1^{-1} X' M_D y + \lambda_2^{-1} X' M_B y\right) \tag{8}$$

where

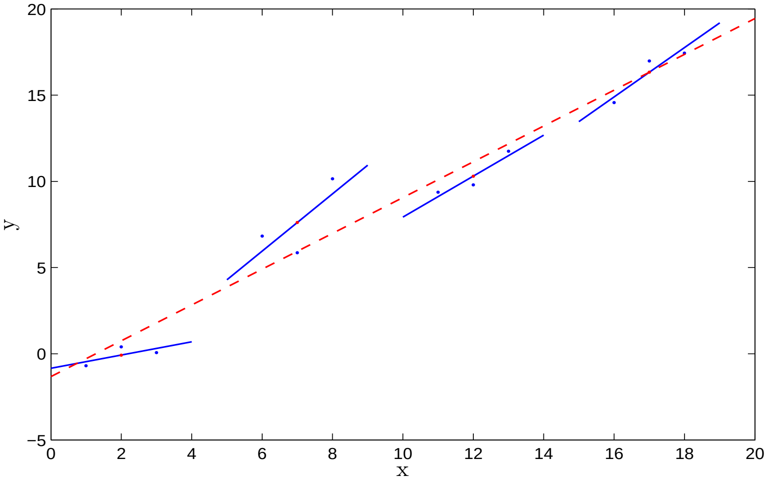
$$X' M_D X = \sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(x_{i,t} - \bar{x}_{i\cdot})', \quad X' M_D y = \sum_i \sum_t (x_{i,t} - \bar{x}_{i\cdot})(y_{i,t} - \bar{y}_{i\cdot})$$

are the **within variation** and **within covariation**, and

$$\frac{1}{T} X' M_B X = \sum_i (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})', \quad \frac{1}{T} X' M_B y = \sum_i (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})$$

are the **between variation** and **between covariation**

Compare (8) with formula (2) of the LSDV estimator!



The blue solid lines are the **within regressions** and the red dashed line is the **between regression**

iv) Properties of the random effects estimator

The random effects estimator is BLUE under assumptions A.1-A.4, with variance

$$V[\hat{\gamma}|X] = (Z'V^{-1}Z)^{-1}$$

It is consistent and asymptotically normal under regularity conditions that depend on the asymptotic scheme

These properties crucially depend on **strict exogeneity of the regressors**

$$E[\varepsilon_i|X_i] = 0 \tag{9}$$

(implied by assumptions A.1-A.4)

Strict exogeneity (9) is a strong assumption and **fails if the regressors** $x_{i,t}$ **are correlated with the individual specific effects** u_i

v) FGLS estimator

We need consistent estimators of σ_u^2 and σ_v^2 or, equivalently, of λ_1 and λ_2

We have:

$$E[\varepsilon_i' M_T \varepsilon_i] = tr\left(M_T E[\varepsilon_i \varepsilon_i']\right) = tr\left(M_T \Omega\right) = \lambda_1 tr(M_T) = \lambda_1(T-1)$$

and similarly $E[\varepsilon_i'(I_T - M_T)\varepsilon_i] = \lambda_2 tr(I_T - M_T) = \lambda_2$.

We deduce estimators based on the GLS residuals $\hat{\varepsilon}_i = y_i - Z_i\hat{\gamma}$:

$$\begin{aligned}
\hat{\lambda}_1 &= \frac{1}{T-1}\frac{1}{n}\sum_i \hat{\varepsilon}_i' M_T \hat{\varepsilon}_i = \frac{1}{n(T-1)}\sum_i\sum_t (\hat{\varepsilon}_{i,t} - \bar{\hat{\varepsilon}}_{i\cdot})^2 \\
\hat{\lambda}_2 &= \frac{1}{n}\sum_i \hat{\varepsilon}_i'(I_T - M_T)\hat{\varepsilon}_i = T\frac{1}{n}\sum_i \bar{\hat{\varepsilon}}_{i\cdot}^2
\end{aligned}$$

These estimators are consistent when $n \to \infty$ and T is fixed (and when $n, T \to \infty$ as well, but are biased in finite sample)

The FGLS estimator is defined as in (7) by replacing Ω^{-1} with $\hat{\Omega}^{-1} = \hat{\lambda}_1^{-1} M_T + \hat{\lambda}_2^{-1}(I_T - M_T)$

The estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ can be used to get estimators of the error variances $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ by solving the equations $\hat{\lambda}_1 = \hat{\sigma}_v^2$ and $\hat{\lambda}_2 = T\hat{\sigma}_u^2 + \hat{\sigma}_v^2$, namely:

$$\hat{\sigma}_v^2 = \hat{\lambda}_1, \quad \hat{\sigma}_u^2 = \frac{1}{T}\left(\hat{\lambda}_2 - \hat{\lambda}_1\right)$$

A drawback of estimator $\hat{\sigma}_u^2$ is that it can provide a **negative** estimated variance of the individual random effect

Gourieroux, Holly, Monfort (1981, 1982) provide an estimator of σ_u^2 that ensure positivity

1.4 FIXED EFFECTS OR RANDOM EFFECTS?

A delicate question!

The choice depends on the characteristics of the sample, the interpretation of the individual effects, the purpose of the model, etc. For instance:

- If the sample contains all the individuals of the underlying population (e.g. the 26 Swiss Cantons), the fixed effects specification appears more natural
- If we are interested in the coefficient of a time-invariant explanatory variable, the fixed effects estimator cannot be used

- If the model is used to **predict** the behaviour of an individual with given observable characteristics and average unobservable characteristics

$$E[g(y_{i,t})|x_{i,t}],$$

  the individual effect has to be integrated out and a random effects specification can be used:

$$E[g(y_{i,t})|x_{i,t}] = \int E[g(y_{i,t})|x_{i,t}, u_i]h(u_i)du_i$$

  where h is the pdf of the individual effect

- ...

In many economic applications, the sampling scheme is such that the individual effects can be considered as genuinely random

Then, it is customary to distinguish:

i) **Fixed effects approach**: no assumptions on distributional properties of the individual effects, in particular on the link with explanatory variables

ii) **Random effects approach:** explicit modeling of the distribution of the individual effects and their link with regressors

Main issue concerns the exogeneity of the regressors w.r.t the individual effects

If strict exogeneity holds ($E[\varepsilon_i|X_i] = 0$), the random effects estimator is BLUE and is more efficient than the fixed effects LSDV estimator

If strict exogeneity does not hold ($E[\varepsilon_i|X_i] \neq 0$), the random effects estimator is biased and inconsistent

The LSDV estimator of the coefficients of the **time-varying regressors** $\tilde{x}_{i,t}$

$$\hat{\beta}_{LSDW} = (\tilde{X}'M_D\tilde{X})^{-1}\tilde{X}'M_Dy = \beta + \left(\frac{1}{n}\sum_i \tilde{X}_i'M_T\tilde{X}_i\right)^{-1}\frac{1}{n}\sum_i \tilde{X}_i'M_Tv_i \qquad (10)$$

is unbiased and consistent as long as:

$$E[v_i|X_i] = 0 \qquad (11)$$

independently whether the individual effect u_i is correlated with $x_{i,t}$ or not!

A **specification test** for the null hypothesis of strict exogeneity of the regressors:

$$H_0 \quad : \quad E[\varepsilon_i | X_i] = 0$$

can be based on the difference

$$\hat{\beta}_{LSDV} - \hat{\beta}_{RE}$$

between the LSDV estimator $\hat{\beta}_{LSDV}$ and the random effects estimator $\hat{\beta}_{RE}$ of the coefficients of the time-varying explanatory variables (Hausman (1978))

Since the random effects estimator $\hat{\beta}_{RE}$ is efficient under the null hypothesis H_0 of strict exogeneity, we have $V[\hat{\beta}_{LSDV} - \hat{\beta}_{RE}] = V[\hat{\beta}_{LSDV}] - V[\hat{\beta}_{RE}]$

The **Hausman test statistic**

$$\xi^H = (\hat{\beta}_{LSDV} - \hat{\beta}_{RE})' \left(\hat{V}[\hat{\beta}_{LSDV}] - \hat{V}[\hat{\beta}_{RE}] \right)^{-1} (\hat{\beta}_{LSDV} - \hat{\beta}_{RE})$$

is distributed asymptotically as χ^2_m under H_0, where m is the number of time-varying explanatory variables

2. IV ESTIMATION OF LINEAR DYNAMIC PANEL DATA MODELS

2.1 INCONSISTENCY OF THE LSDV ESTIMATOR IN LINEAR DYNAMIC PANELS

i) The dynamic linear panel data model:

$$y_{i,t} = \alpha y_{i,t-1} + u_i + v_{i,t}$$

where:

A.1: The errors $v_{i,t}$ are i.i.d. across individuals and time dates with $E[v_{i,t}] = 0$ and $V[v_{i,t}] = \sigma_v^2$, for all i and t

A.2: The u_i are individual fixed effects

A.3: The **initial observations** $y_{i,0}$ are i.i.d. across individuals and independent of the errors $v_{i,t}$, for all i and $t \geq 1$

Compact form: for individual i

$$\underset{T\times 1}{y_i} = \alpha \underset{T\times 1}{y_{i,-1}} + u_i \underset{T\times 1}{S_T} + \underset{T\times 1}{v_i}$$

where $y_{i,-1} \equiv (y_{i,0}, y_{i,1}, ..., y_{i,T-1})'$

The regressor $y_{i,-1}$ is **not strictly exogenous** and $E[v_i|y_{i,-1}] \neq 0$, hence condition (11) (with $X_i = y_{i,-1}$) is not satisfied!

ii) The Nickel bias (Nickel (1981))

The LSDV estimator of parameter α is

$$\hat{\alpha} = \frac{\sum_i y'_{i,-1} M_T y_i}{\sum_i y'_{i,-1} M_T y_{i,-1}} = \alpha + \frac{\frac{1}{n}\sum_i y'_{i,-1} M_T v_i}{\frac{1}{n}\sum_i y'_{i,-1} M_T y_{i,-1}}$$

It is possible to show that, when $|\alpha| \neq 1$:

$$E[y'_{i,-1} M_T v_i] = \sum_t E[(y_{i,t-1} - \bar{y}_{i,-1})(v_{i,t} - \bar{v}_{i,.})] = -\sigma_v^2 h_T(\alpha)$$

and:

$$E[y'_{i,-1} M_T y_{i,-1}] = \frac{\sigma_v^2 (T-1)}{1-\alpha^2} \left(1 - \frac{2\alpha h_T(\alpha)}{T-1}\right)$$

where:

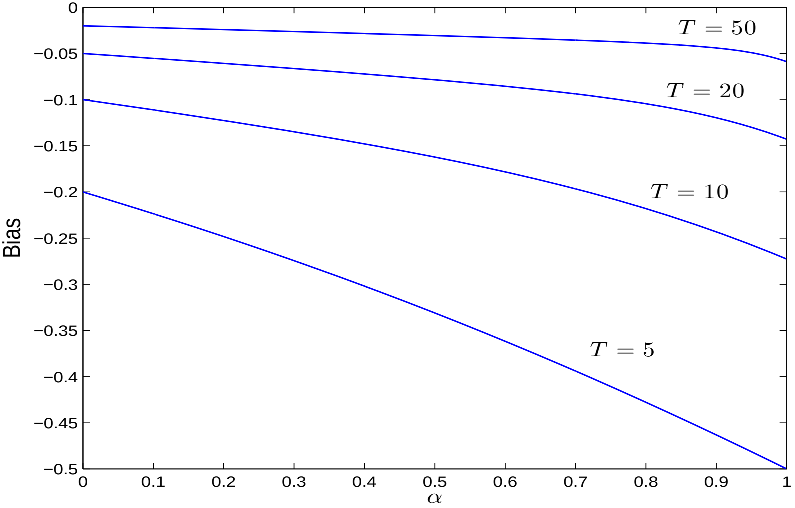
$$h_T(\alpha) \equiv \frac{1}{1-\alpha} \left[1 - \frac{1}{T} \left(\frac{1-\alpha^T}{1-\alpha}\right)\right]$$

(see [ARE], Chapter 6)

Then, the asymptotic bias of the LSDV estimator

$$\operatorname*{plim}_{n\to\infty} \hat{\alpha} - \alpha = -\frac{1}{T-1} \frac{(1-\alpha^2) h_T(\alpha)}{1 - \frac{2\alpha h_T(\alpha)}{T-1}} = O(1/T)$$

is non-zero (negative) for fixed T


T = 50
T = 20
T = 10
T = 5
Bias
α
0
−0.05
−0.1
−0.15
−0.2
−0.25
−0.3
−0.35
−0.4
−0.45
−0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

2.2 INSTRUMENTAL VARIABLE (IV) ESTIMATION

i) The orthogonality restrictions

Write the model in **first-differences** to eliminate the individual effects

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \Delta v_{i,t}, \quad t = 2, ..., T$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ and $\Delta v_{i,t} = v_{i,t} - v_{i,t-1}$

We have a set of $m = T(T-1)/2$ orthogonality restrictions for each individual

$$E[y_i^{t-2}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1})] = 0, \quad t = 2, ..., T$$

where $y_i^{t-2} \equiv (y_{i,0}, y_{i,1}, ..., y_{i,t-2})'$

The orthogonality restrictions can be written as:

$$E[Z_i'(\Delta y_i - \alpha \Delta y_{i,-1})] = 0$$

where $\Delta y_i = (\Delta y_{i,2}, ..., \Delta y_{i,T})'$, $\Delta y_{i,-1} = (\Delta y_{i,1}, ..., \Delta y_{i,T-1})'$ and Z_i is the $(T-1) \times m$ matrix such that

$$Z_i = \begin{pmatrix} y_{i,0} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i,0} & y_{i,1} & & 0 & \cdots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,0} & \cdots & y_{i,T-2} \end{pmatrix}$$

ii) The GMM (IV) estimator:

$$\begin{aligned}\hat{\alpha}_{GMM} &= \arg\min_{\alpha}(\Delta y - \alpha\Delta y_{-1})' Z\hat{\Omega}Z'(\Delta y - \alpha\Delta y_{-1}) \\ &= (\Delta y_{-1}' Z\hat{\Omega}Z'\Delta y_{-1})^{-1}\Delta y_{-1}' Z\hat{\Omega}Z'\Delta y\end{aligned}$$

where $Z'\Delta y \equiv \sum_i Z_i'\Delta y_i$ and $\hat{\Omega}$ is a positive definite $m \times m$ weighting matrix (see [ARE])

iii) The optimal weighting matrix: is $\Omega = V^{-1}$, where

$$V = E[Z_i'(\Delta y_i - \alpha\Delta y_{i,-1})(\Delta y_i - \alpha\Delta y_{i,-1})'Z_i] = E[Z_i'\Delta v_i \Delta v_i' Z_i]$$

iv) The two-step efficient GMM estimator: uses the weighting matrix $\hat{\Omega} = \hat{V}^{-1}$ where

$$\hat{V} = \frac{1}{n} \sum_i Z_i' \widehat{\Delta v_i} \widehat{\Delta v_i}' Z_i$$

with $\widehat{\Delta v_i} = \Delta y_i - \tilde{\alpha} \Delta y_{i,-1}$ and $\tilde{\alpha}$ is a consistent first-step GMM estimator of α

The two-step efficient GMM estimator is asymptotically normal (see [ARE]):

$$\sqrt{n} \left(\hat{\alpha}_{GMM} - \alpha \right) \xrightarrow{d} N(0, \Sigma)$$

when $n \to \infty$ and T is fixed, where

$$\Sigma = \left(E[\Delta y_{i,-1}' Z_i] V^{-1} E[Z_i' \Delta y_{i,-1}] \right)^{-1}$$

Asymptotically valid standard errors and confidence intervals from

$$\widehat{AsVar}(\hat{\alpha}_{GMM}) = \frac{1}{n} \hat{\Sigma} = \left[\left(\sum_i \Delta y_{i,-1}' Z_i \right) \left(\sum_i Z_i' \widehat{\Delta v_i} \widehat{\Delta v_i}' Z_i \right)^{-1} \left(\sum_i Z_i' \Delta y_{i,-1} \right) \right]^{-1}$$

2.3 PANEL MODELS WITH BOTH STRICTLY EXOGENOUS AND LAGGED DEPENDENT VARIABLES

Consider the model

$$
\begin{aligned}
y_{i,t} &= \alpha y_{i,t-1} + x'_{i,t}\beta + u_i + v_{i,t} \\
&= w'_{i,t}\gamma + u_i + v_{i,t}
\end{aligned}
$$

where $w_{i,t} = (y_{i,t-1}, x'_{i,t})'$ and $\gamma = (\alpha, \beta')'$

$$
\Longleftrightarrow \quad y_i = \alpha y_{i,-1} + X_i\beta + u_i S_T + v_i = W_i\gamma + u_i S_T + v_i
$$

We add to assumptions A.1-A.3

A.4: The regressor $x_{i,t}$ and the idiosyncratic error $v_{j,s}$ are independent, for any i, j, t and s (which implies **strict exogeneity** $E[v_i|X_i] = 0$)

Write the model in first-differences to eliminate the individual effects

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \Delta x'_{i,t}\beta + \Delta v_{i,t}$$

We have the orthogonality restrictions

$$E[z_{i,t}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta x'_{i,t}\beta)] = 0, \quad t = 2, ..., T, \qquad (12)$$

where the vector of instruments is:

- $z_{i,t} = (y_{i,t-2}, \Delta x'_{i,t})'$ in Anderson, Hsiao (1981, 1982)
- $z_{i,t} = (y_{i,0}, y_{i,1}, ..., y_{i,t-2}, \Delta x'_{i,t})'$ in Holtz-Eakin, Newey, Rosen (1988), Arellano, Bond (1991)
  (may include also lags or leads of $\Delta x_{i,t}$)

The Arellano-Bond (1991) GMM estimator:

$$\hat{\gamma}_{GMM} = \left(\Delta W' Z \hat{V}^{-1} Z' \Delta W \right)^{-1} \Delta W' Z \hat{V}^{-1} Z' \Delta y$$

where $\Delta W' Z \equiv \sum_i \Delta W_i' Z_i$ and $Z' \Delta y \equiv \sum_i Z_i' \Delta y_i$, with $\Delta W_i = (\Delta y_{i,-1}, \Delta X_i)$

and

$$Z_i = \begin{pmatrix} y_{i,0} & \Delta x_{i,2} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & y_{i,0} & y_{i,1} & \Delta x_{i,3} & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & y_{i,0} & \cdots & y_{i,T-2} & \Delta x_{i,T} \end{pmatrix}$$

the optimal weighting matrix $\hat{V}^{-1}$ is such that $\hat{V} = \frac{1}{n} \sum_i Z_i' \widehat{\Delta v}_i \widehat{\Delta v}_i' Z_i$, with

$\widehat{\Delta v}_i = \Delta y_i - \tilde{\alpha} \Delta y_{i,-1} - \Delta X_i \tilde{\beta}$ and $(\tilde{\alpha}, \tilde{\beta}')'$ is a first-step GMM estimator obtained with an identity weighting matrix

Remark 1: When the errors $v_{i,t}$ are serially correlated, in general

$$E[y_{i,s}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta x'_{i,t}\beta)] \neq 0$$

$\Rightarrow$ **The lagged endogenous variables up to $t-2$ cannot be used as instruments for $\Delta y_{i,t-1}$!**

However, the strictly exogenous regressors can still be used as instruments to get the orthogonality restrictions

$$E[\Delta x_{i,s}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta x'_{i,t}\beta)] = 0, \quad s = 1, ..., T, \quad t = 2, ..., T$$

Remark 2: If we only assume that $x_{i,s}$ and $v_{i,t}$ are uncorrelated for $s \leq t$, the orthogonality restrictions are:

$$E[\Delta x_{i,s}(\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta x'_{i,t}\beta)] = 0, \quad s = 1, ..., t-1, \quad t = 2, ..., T$$

Remark 3: When instruments z_i^* are available, which are uncorrelated with the individual effects, the set of orthogonality restrictions (12) can be extended to include **moment restrictions in level**

$$E[z_i^*(\bar{y}_{i\cdot} - \alpha\bar{y}_{i,-1} - \bar{x}_{i\cdot}'\beta] = 0$$

$\Rightarrow$ Arellano, Bover (1995) GMM estimator

Remark 4: Blundell, Bond (1998) suggest the use of additional moment restrictions that are very informative when individual histories are persistent (α close to 1)

Consider the equations:

$$E[\Delta y_{i,t-1}\varepsilon_{i,t}] = 0, \quad t = 3, 4, ..., T, \tag{13}$$

where $\varepsilon_{i,t} = u_i + v_{i,t}$, and for $t = 2$:

$$E[\Delta y_{i,1}\varepsilon_{i,2}] = 0 \tag{14}$$

Equation (13) is valid under assumptions A.1-A.4, equation (14) is valid under assumptions A.1-A.4 if in addition $Cov(y_{i,0} - \frac{u_i}{1-\alpha}, u_i) = 0$

Equations (13) and (14) imply moment conditions with instruments in difference for data in level:

$$E[\Delta y_{i,t-1}(y_{i,t} - \alpha y_{i,t-1} - x'_{i,t}\beta)] = 0, \quad t = 2, 3, ..., T$$

Remark 5: a) The number of orthogonality restrictions can be rather large even for a short panel and grows with sample size as $O(T^2)$

For instance, with $T = 5$ periods and $K = 5$ regressors, in Arellano, Bond (1991) we have $T(T-1)/2 + K(T-1) = 30$ orthogonality restrictions!

b) Some instruments might be "weak", i.e. weakly correlated with the explanatory variables

a) & b) cause poor finite-sample performance of standard GMM inference procedures; see e.g. Newey, Windmeijer (2009) for the use of modified GMM estimators

3. NONLINEAR PANEL DATA MODELS

i) Random utility model

Consider panel data for a **binary variable**, for instance transportation mode

$$y_{i,t} = \begin{cases} 0, & \text{private (car, motorcycle, ...)} \\ 1, & \text{public (train, bus, ...)} \end{cases}$$

How to model this discrete choice?

Unobservable random utilities from the two alternatives

$$\begin{aligned} U_{0,it} &= 0, \quad \text{(by normalization)} \\ U_{1,it} &= x'_{i,t}\beta + \varepsilon_{i,t} \end{aligned}$$

The observed choice is

$$y_{i,t} = \begin{cases} 1, & \text{if } U_{1,it} \geq U_{0,it} = 0 \\ 0, & \text{otherwise} \end{cases}$$

Assume the **error component** structure

$$\varepsilon_{i,t} = u_i + v_{i,t}$$

where **individual effects** $u_i \sim IIN(0, \sigma_u^2)$ and shocks $v_{i,t} \sim IIN(0, \sigma_v^2)$ are independent of each other and of regressors $x_{i,t}$

Identification constraint: By using $\varepsilon_{i,t} \sim N(0, \sigma_u^2 + \sigma_v^2)$

$$P\left[y_{i,t} = 1 | x_{i,t}\right] = P\left[\varepsilon_{i,t} \geq -x_{i,t}'\beta | x_{i,t}\right] = \Phi\left(\frac{x_{i,t}'\beta}{\sqrt{\sigma_u^2 + \sigma_v^2}}\right)$$

$\Rightarrow$ parameters are identifiable up to a scale and we can set $\sigma_v^2 = 1$

Probability of alternative $y_{i,t} = 1$ given observable characteristics and individual effect:

$$\begin{aligned} P\left[y_{i,t} = 1 | x_{i,t}, u_i\right] &= P\left[v_{i,t} \geq -x_{i,t}'\beta - u_i | x_{i,t}, u_i\right] \\ &= \Phi\left(x_{i,t}'\beta + u_i\right) \end{aligned}$$

ii) General specification

$$y_{i,t} = 1\left\{x'_{i,t}\beta + u_i + v_{i,t} \geq 0\right\}$$

A.1: The idiosyncractic errors $v_{i,t}$ are i.i.d. across i and t, independent from regressors $x_i = (x'_{i,1}, ..., x'_{i,T})'$, and $-v_{i,t}$ has c.d.f. F

Probability of alternative $y_{i,t} = 1$ given observable characteristics and individual effects:

$$P\left[y_{i,t} = 1|x_i, u_i\right] = P[-v_{i,t} \leq x'_{i,t}\beta + u_i|x_i, u_i] = F\left(x'_{i,t}\beta + u_i\right)$$

Probit model when $F(v) = \Phi(v)$ and **logit model** when $F(v) = \dfrac{1}{1+e^{-v}}$

A.2: The **random individual effects** u_i are assumed **independent of the regressors** x_i (strict exogeneity) and of errors v_i, are i.i.d. across individuals and distributed according to a p.d.f. $h(u_i;\delta)$ indexed by unknown parameter δ

Parameters of interest are β and δ

iii) The likelihood function

The pdf of observation $y_{i,t}$ given $x_i = (x'_{i,1}, x'_{i,2}, ..., x'_{i,T})'$ and the random effect u_i

$$f(y_{i,t}|x_i, u_i; \beta) = F\left(x'_{i,t}\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x'_{i,t}\beta + u_i\right)\right]^{1-y_{i,t}}$$

By the independence of the individual observations over time conditional on regressors x_i and random effect u_i, the pdf of vector $y_i = (y_{i,1}, y_{i,2}, ..., y_{i,T})'$ given x_i and u_i is

$$f(y_i|x_i, u_i; \beta) = \prod_{t=1}^{T} F\left(x'_{i,t}\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x'_{i,t}\beta + u_i\right)\right]^{1-y_{i,t}} \tag{15}$$

We integrate out the random effects and get the pdf of y_i given x_i

$$f(y_i|x_i; \beta, \delta) = \int \left(\prod_{t=1}^{T} F\left(x'_{i,t}\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x'_{i,t}\beta + u_i\right)\right]^{1-y_{i,t}}\right) h(u_i; \delta) du_i$$

By the independence across individuals, the pdf of vector $y = (y_1', y_2', ..., y_n')'$ given $x = (x_1', x_2', ..., x_n')'$ is

$$f(y|x; \beta, \delta) = \prod_{i=1}^{n} \int \left(\prod_{t=1}^{T} F\left(x_{i,t}'\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x_{i,t}'\beta + u_i\right)\right]^{1-y_{i,t}} \right) h(u_i; \delta) du_i$$

We get the log-likelihood function:

$$\begin{aligned} \mathcal{L}_{n,T}(\beta, \delta) &= \log f(y|x; \beta; \delta) \\ &= \sum_{i=1}^{n} \log \int \left(\prod_{t=1}^{T} F\left(x_{i,t}'\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x_{i,t}'\beta + u_i\right)\right]^{1-y_{i,t}} \right) h(u_i; \delta) du_i \end{aligned}$$

iv) The Maximum Likelihood (ML) estimator: is defined by

$$(\hat{\beta}', \hat{\delta}')' = \arg\max_{\beta,\delta} \mathcal{L}_{n,T}(\beta, \delta)$$

The ML estimators $\hat{\beta}$ and $\hat{\delta}$ are root-n consistent and asymptotically normal when $n \to \infty$ and T is fixed:

$$\sqrt{n}[(\hat{\beta}', \hat{\delta}')' - (\beta', \delta')'] \xrightarrow{d} N(0, \mathcal{I}^{-1}),$$

where $\mathcal{I}$ is the Fisher information matrix:

$$\mathcal{I} = E\left[-\frac{\partial^2 \log f(y_i|x_i; \beta, \delta)}{\partial(\beta', \delta')'\partial(\beta', \delta')}\right].$$

v) Numerical computation of the ML estimator

Evaluation of the likelihood function for given parameter values requires **numerical computation** of the n one-dimensional integrals

$$\int \left(\prod_{t=1}^{T} F\left(x'_{i,t}\beta + u_i\right)^{y_{i,t}} \left[1 - F\left(x'_{i,t}\beta + u_i\right)\right]^{1-y_{i,t}} \right) h(u_i;\delta) du_i$$

for $i = 1, ..., n$

Such a numerical approximation is required at each step of the iterative optimization algorithm used to get the ML estimates

In a probit model with Gaussian distributed random effects, the integral w.r.t. the Gaussian pdf $h(u_i;\sigma_u^2) = \frac{1}{\sigma_u}\phi\left(\frac{u_i}{\sigma_u}\right)$ can be well approximated by **Gaussian quadrature** (Butler, Moffitt (1982))

For more complicated distributions of the random effects, an alternative is to approximate the integral by **Monte-Carlo simulation**, which gives a **Simulated Maximum Likelihood (SML)** estimator (see Gourieroux, Monfort (1993))

3.2 DISCRETE CHOICE MODELS WITH FIXED EFFECTS

i) The model

$$y_{i,t} = 1\left\{x'_{i,t}\beta + u_i + v_{i,t} \geq 0\right\}$$

with assumption A.1 of Section 3.1 ii)

The conditional probability of alternative $y_{i,t} = 1$ is given by:

$$P[y_{i,t} = 1|x_i, u_i] = F(x'_{i,t}\beta + u_i)$$

where F is the c.d.f. of $-v_{i,t}$

On the contrary of Section 3.1, here the individual effect u_i **is not necessarily independent** of the regressor x_i, i.e. we don't impose assumption A.2

We treat the individual effects u_i, for $i = 1, ..., n$, as parameters to estimate

By taking the log in (15) and summing over i, the log-likelihood function is:

$$\mathcal{L}_{nT}(\beta, u_1, ..., u_n) = \sum_i \sum_t \left\{y_{i,t} \log F(x'_{i,t}\beta + u_i) + (1 - y_{i,t}) \log[1 - F(x'_{i,t}\beta + u_i)]\right\} \quad (16)$$

ii) Inconsistency of the ML estimator and incidental parameters problem

It turns out that the ML estimator of parameter β obtained by maximizing the log-likelihood function (16) is **inconsistent** for $n \to \infty$ and T fixed!

We illustrate this fact in a "simple" example: $F(v) = 1/(1+e^{-v})$ (logit), $T = 2$ periods, $x_{i,1} = 0$ and $x_{i,2} = 1$ for all i (see [HSI], Section 7.3.1)

The log-likelihood function (16) becomes:

$$\mathcal{L}_{nT}(\beta, u_1, ..., u_n) = \sum_i \left[y_{i,1}u_i - \log(1+e^{u_i}) + y_{i,2}(\beta + u_i) - \log(1+e^{\beta+u_i}) \right]$$

By maximizing the log-likelihood w.r.t. the u_i for given β, and plugging in $\mathcal{L}_{nT}$ the optimal values $\hat{u}_i(\beta)$, we get the **concentrated log-likelihood** of β:

$$\mathcal{L}_{nT}^c(\beta) = \sum_i 1\{y_{i,1} + y_{i,2} = 1\} \left[\beta y_{i,2} - \frac{\beta}{2} - \log(1+e^{-\beta/2}) - \log(1+e^{\beta/2}) \right]$$

By maximizing this concentrated log-likelihood w.r.t. β, we get the ML estimator $\hat{\beta}_{ML}$ of parameter β

$$\hat{\beta}_{ML} = 2\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right), \quad \hat{\pi} = \frac{\sum_i 1\{y_{i,1}=0, y_{i,2}=1\}}{\sum_i 1\{y_{i,1}+y_{i,2}=1\}}$$

The Law of Large Numbers (LLN) implies that

$$\underset{n\to\infty}{\text{plim}}\,\hat{\pi} = \frac{P(y_{i,1}=0, y_{i,2}=1)}{P(y_{i,1}+y_{i,2}=1)} = \frac{1}{1+e^{-\beta}}$$

We deduce that the ML estimator of β is such that

$$\underset{n\to\infty}{\text{plim}}\,\hat{\beta}_{ML} = 2\beta$$

that is, the ML estimator is inconsistent!

The inconsistency of $\hat{\beta}_{ML}$ is a manifestation of a more general fact called the **incidental parameters problem** (Neyman, Scott (1948))

The ML estimator of a parameter can be inconsistent when **the number of parameters in the likelihood function grows with the sample size**

The individual fixed effects are incidental parameters!

In linear panel models, the incidental parameter problem does not cause inconsistency of the ML estimator of the common parameter (i.e. the LSDV estimator)

In **nonlinear panel models**, fixed-T inconsistency of the fixed-effects ML estimator due to the incidental parameter problem is pervasive!

iii) The Conditional ML estimator in the panel logit model

For the logit specification, i.e. $F(v) = 1/(1+e^{-v})$, a consistent estimator of β is obtained by maximizing a **conditional likelihood function** given a well-chosen sufficient statistic (Chamberlain (1980))

Let us first consider the case $T = 2$ to get the intuition

For individual i, let us condition on the event $y_{i,1} + y_{i,2} = 1$. Then:

$$\begin{aligned} P[y_{i,2} = 1 | x_i, y_{i,1} + y_{i,2} = 1] &= \frac{P[y_{i,1} = 0, y_{i,2} = 1 | x_i]}{P[y_{i,1} = 0, y_{i,2} = 1 | x_i] + P[y_{i,1} = 1, y_{i,2} = 0 | x_i]} \\ &= \frac{e^{x'_{i,2}\beta}}{e^{x'_{i,1}\beta} + e^{x'_{i,2}\beta}} = F(\Delta x'_{i,2}\beta) \end{aligned}$$

with $\Delta x_{i,2} = x_{i,2} - x_{i,1}$, is independent of u_i!

Parameter β can be consistently estimated by running a logit regression on the subsample of individuals with $y_{i,1} + y_{i,2} = 1$

Extension to $T > 2$: The p.d.f. of $y_i = (y_{i,1}, ..., y_{i,T})'$ conditional on $x_i = (x'_{i,1}, ..., x'_{i,T})'$ and $\sum_t y_{i,t}$ is

$$f\left(y_i | x_i, \sum_t y_{i,t}\right) \propto \frac{\exp\left(\sum_t y_{i,t} x'_{i,t} \beta\right)}{\sum_{d \in B_i} \exp\left(\sum_t d_t x'_{i,t} \beta\right)}$$

where B_i denotes the set of vectors $d = (d_1, ..., d_T) \in \{0, 1\}^T$ such that $\sum_t d_t = \sum_t y_{i,t}$

The Conditional Maximum Likelihood (CML) estimator:

$$\hat{\beta}_{CML} = \arg\max_{\beta} \sum_i \log\left(\frac{\exp\left(\sum_t y_{i,t} x'_{i,t} \beta\right)}{\sum_{d \in B_i} \exp\left(\sum_t d_t x'_{i,t} \beta\right)}\right)$$

is consistent when $n \to \infty$ and T is fixed

iv) Fixed effects estimators with other error distributions

For the fixed effects probit model, a consistent CML estimator is not available

The Manski (1987) **maximum score estimator** is a semi-parametric estimator that can be used with a generic distribution of the idiosyncratic error terms

$$\hat{\beta} = \arg\max_{\beta} \sum_i \sum_{s<t} sign(y_{it} - y_{is}) sign\left((x_{it} - x_{is})'\beta\right)$$

The intuition is to get the β such that the association

$$y_{it} > y_{is} \quad \Leftrightarrow \quad x_{it}'\beta > x_{is}'\beta$$

is maximized across pairs of dates $t > s$

3.3 SAMPLE SELECTION IN PANEL DATA

i) Unbalanced panels

The panel is **unbalanced** if the data of some individuals at some dates are not observed

Unobservability may be due to exit from the sample, non-response / non-participation, rotation of individuals in the sample, attrition, etc

The underlying model is a linear regression with individual specific effects

$$y^*_{1,it} = x'_{i,t}\beta + u_i + v_{i,t}$$

the **observability indicator** is

$$y_{2,it} = \begin{cases} 1, & \text{if } y^*_{1,it} \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

and we define $y_{1,it} \equiv y^*_{1,it} y_{2,it}$

Available data are $y_{1,it}$, $y_{2,it}$, $x_{i,t}$

ii) The missing-at-random assumption

When does partial unobservability create a bias? When not?

Consider the LSDV estimator computed with the available observations

$$\hat{\beta} = \left[\sum_i \sum_t y_{2,it}(x_{i,t} - \bar{x}_{i,\cdot})(x_{i,t} - \bar{x}_{i,\cdot})'\right]^{-1} \sum_i \sum_t y_{2,it}(x_{i,t} - \bar{x}_{i,\cdot})(y_{1,it} - \bar{y}_{1,i\cdot})$$

$$= \beta + \left[\frac{1}{n}\sum_i \sum_t y_{2,it}(x_{i,t} - \bar{x}_{i,\cdot})(x_{i,t} - \bar{x}_{i,\cdot})'\right]^{-1} \frac{1}{n}\sum_i \sum_t y_{2,it}(x_{i,t} - \bar{x}_{i,\cdot})(v_{i,t} - \bar{v}_{i,\cdot})$$

where $\bar{x}_{i,\cdot} \equiv \frac{1}{T_i}\sum_t y_{2,it}x_{i,t}$, $\bar{y}_{1,i\cdot} \equiv \frac{1}{T_i}\sum_t y_{2,it}y_{1,it}$ and $T_i \equiv \sum_t y_{2,it}$

The LSDV estimator is unbiased, and consistent for $n \to \infty$ and T fixed, if:

$$E[v_i | X_i, y_{2,i}] = 0 \qquad (17)$$

Condition (17) is satisfied if $E[v_i|X_i] = 0$ (see (11)) and the **mixing-at-random assumption** holds: $y_{2,i}$ and v_i are independent, conditional on X_i

iii) Sample selection

Sample selection bias: When individuals elect to drop in/out of the sample, condition (17) is typically not satisfied and the LSDV estimator is inconsistent!

Type 2 Tobit panel model: The selection/observability mechanism is modeled via a discrete choice

$$\begin{aligned} y^*_{1,it} &= x'_{i,t}\beta + u_i + v_{i,t} \\ y_{2,it} &= 1\left\{z'_{i,t}\gamma + \eta_i + w_{i,t} \geq 0\right\} \\ y_{1,it} &= y^*_{1,it} y_{2,it} \end{aligned}$$

where η_i is the individual specific effect in the selection equation

The econometrician observes $y_{1,it}$, $y_{2,it}$, $x_{i,t}$

iv) Random effects estimator

Model the joint distribution of the individual random effects, idiosyncratic error terms and regressors

A.1: (u_i, η_i) is jointly distributed according to a p.d.f. $h(u_i, \eta_i; \delta)$ with unknown parameter δ, e.g. a bivariate Gaussian

A.2: $(v_{i,t}, w_{i,t})'$ are i.i.d. across time dates with Gaussian distribution $N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_v^2 & \rho\sigma_v \\ \rho\sigma_v & 1 \end{bmatrix}\right)$

A.3: (u_i, η_i), (v_i, w_i), X_i and Z_i are mutually independent and i.i.d. across individuals

Assumption A.3 implies **strict exogeneity** of the regressors!

Let us construct the likelihood function. From standard Tobit 2 model:

$$f(y_{1,it}, y_{2,it}|X_i, Z_i, u_i, \eta_i) = \left[1 - \Phi(z'_{i,t}\gamma + \eta_i)\right]^{1-y_{2,it}} \cdot \left[\frac{1}{\sigma_v}\phi\left(\frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v}\right)\Phi\left(\frac{z'_{i,t}\gamma + \eta_i + \rho\frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v}}{\sqrt{1-\rho^2}}\right)\right]^{y_{2,it}}$$

By aggregating over time and integrating out the individual effects we get:

$$f(y_{1,i}, y_{2,i}|X_i, Z_i) = \int \prod_t \left(\left[1 - \Phi(z'_{i,t}\gamma + \eta_i)\right]^{1-y_{2,it}} \left[\frac{1}{\sigma_v} \phi \left(\frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v} \right) \Phi \left(\frac{z'_{i,t}\gamma + \eta_i + \rho \frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v}}{\sqrt{1 - \rho^2}} \right) \right]^{y_{2,it}} \right) h(u_i, \eta_i; \delta) du_i d\eta_i$$

We get the log-likelihood function:

$$\mathcal{L}_{nT}(\beta, \gamma, \delta, \sigma_v^2) = \sum_i \log \left\{ \int \prod_t \left(\left[1 - \Phi(z'_{i,t}\gamma + \eta_i)\right]^{1-y_{2,it}} \left[\frac{1}{\sigma_v} \phi \left(\frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v} \right) \Phi \left(\frac{z'_{i,t}\gamma + \eta_i + \rho \frac{y_{1,it} - x'_{i,t}\beta - u_i}{\sigma_v}}{\sqrt{1 - \rho^2}} \right) \right]^{y_{2,it}} \right) h(u_i, \eta_i; \delta) du_i d\eta_i \right\}$$

The **random effects ML estimator**:

$$(\hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\sigma}_v^2) = \arg\max_{(\beta,\gamma,\delta,\sigma_v^2)} \mathcal{L}_{nT}(\beta, \gamma, \delta, \sigma_v^2)$$

is root-n consistent and asymptotically normal when $n \to \infty$ and T is fixed

The evaluation of the log-likelihood for given parameter values requires the numerical computation of the bivariate integral w.r.t. the individual random effects

v) Heckman-type approach for panel sample selection model (Wooldridge (1995))

Relax some of the distributional assumptions used for random effects estimation and implement a two-step procedure à la Heckman

B.1: $E[u_i|X_i, Z_i, \eta_i, w_i] = x_i'\pi + \theta\eta_i$ and $E[v_{it}|X_i, Z_i, \eta_i, w_i] = \rho w_{it}$

B.2: $\eta_i \sim N(0, \sigma_\eta^2)$ and $w_{i,t} \sim N(0,1)$ are independent, and independent of X_i, Z_i

Assumption B.1 allows for correlation between individual effect u_i and regressor x_i. In practice $x_i = \bar{x}_{i,\cdot}$

From assumption B.1 and the Law of Iterated Expectation, we have:

$$
\begin{aligned}
E[y_{1,it}|X_i, Z_i, y_{2,it} = 1] &= x_{i,t}'\beta + E\left[E(u_i + v_{i,t}|X_i, Z_i, \eta_i, w_i, y_{2,it} = 1)|X_i, Z_i, y_{2,it} = 1\right] \\
&= x_{i,t}'\beta + x_i'\pi + \theta E[\eta_i|X_i, Z_i, \eta_i + w_{i,t} > -z_{i,t}'\gamma] \\
&\quad +\rho E[w_{it}|X_i, Z_i, \eta_i + w_{i,t} > -z_{i,t}'\gamma]
\end{aligned}
$$

From assumption B.2 we get:

$$
\begin{aligned}
E[\eta_i | X_i, Z_i, \eta_i + w_{i,t} > -z'_{i,t}\gamma] &= \frac{\sigma_\eta^2}{1+\sigma_\eta^2} E[\eta_i + w_{i,t} | X_i, Z_i, \eta_i + w_{i,t} > -z'_{i,t}\gamma] \\
&= \frac{\sigma_\eta^2}{1+\sigma_\eta^2} \lambda \left(-\frac{1}{\sqrt{1+\sigma_\eta^2}} z'_{i,t}\gamma \right)
\end{aligned}
$$

where $\lambda(v) = \dfrac{\phi(v)}{1-\Phi(v)}$ is the inverse Mill's ratio

Similarly

$$
E[w_{it} | X_i, Z_i, \eta_i + w_{i,t} > -z'_{i,t}\gamma] = \frac{1}{1+\sigma_\eta^2} \lambda \left(-\frac{1}{\sqrt{1+\sigma_\eta^2}} z'_{i,t}\gamma \right)
$$

We get:

$$
E[y_{1,it} | X_i, Z_i, y_{2,it} = 1] = x'_{i,t}\beta + x'_i\pi + \frac{\theta\sigma_\eta^2 + \rho}{1+\sigma_\eta^2} \lambda \left(-\frac{1}{\sqrt{1+\sigma_\eta^2}} z'_{i,t}\gamma \right)
$$

Parameter β can be estimated by a two-step procedure:

1. Estimate parameters γ and σ_η^2 in the selection equation by probit random effects approach. Construct estimates of the inverse Mill's ratios

$$\hat{\lambda}_{i,t} = \lambda\left(-\frac{1}{\sqrt{1+\hat{\sigma}_\eta^2}} z_{i,t}'\hat{\gamma}\right)$$

2. Regress $y_{1,it}$ on $x_{i,t}$, x_i and $\hat{\lambda}_{i,t}$ by pooled OLS on the selected sample with $y_{2,it} = 1$

The estimator of β is root-n consistent and asymptotically normal when $n \to \infty$ and T is fixed

Correct standard errors are derived in Wooldridge (1995)

vi) Fixed effects estimator (Kyriazidou (1997))

A mild "stationarity" assumption on the idiosyncratic error terms

C.1: For any two dates $t \neq s$, vectors $(v_{i,t}, w_{i,t}, v_{i,s}, w_{i,s})$ and $(v_{i,s}, w_{i,s}, v_{i,t}, w_{i,t})$ are identically distributed conditionally on $\xi_i \equiv (x_{i,t}, z_{i,t}, x_{i,s}, z_{i,s}, u_i, \eta_i)$

Under C.1, for any two dates $t \neq s$ and an individual i such that $z_{i,t}'\gamma = z_{i,s}'\gamma$

$$\begin{aligned}
\tilde{\lambda}_{i,t} &\equiv E[v_{i,t}|w_{i,t} > -z_{i,t}'\gamma - \eta_i, w_{i,s} > -z_{i,s}'\gamma - \eta_i, \xi_i] \\
&= E[v_{i,s}|w_{i,s} > -z_{i,t}'\gamma - \eta_i, w_{i,t} > -z_{i,s}'\gamma - \eta_i, \xi_i] \\
&= E[v_{i,s}|w_{i,s} > -z_{i,s}'\gamma - \eta_i, w_{i,t} > -z_{i,t}'\gamma - \eta_i, \xi_i] = \tilde{\lambda}_{i,s}
\end{aligned}$$

$\Rightarrow$ Considering two dates $t \neq s$ for which data of individual i are observed and $z_{i,t}'\gamma = z_{i,s}'\gamma$, first-differenting eliminates both the fixed effect and the sample selection effect!

A two-step estimation procedure:

1. Get an estimator $\hat{\gamma}$ of the parameter in the selection equation by a fixed effects methodology
2. The **fixed effects estimator** of parameter β is:

$$\hat{\beta} = \left[\sum_i \sum_{t>s} y_{2,it} y_{2,is} (x_{i,t} - x_{i,s})(x_{i,t} - x_{i,s})' K\left(\frac{(x_{i,t} - x_{i,s})'\hat{\gamma}}{h_n}\right)\right]^{-1} \cdot \left[\sum_i \sum_{t>s} y_{2,it} y_{2,is} (x_{i,t} - x_{i,s})(y_{1,it} - y_{1,is}) K\left(\frac{(x_{i,t} - x_{i,s})'\hat{\gamma}}{h_n}\right)\right]$$

   where K is a kernel and $h_n > 0$ is a bandwidth that shrinks to zero as n increases

The estimator $\hat{\beta}$ is consistent and asymptotically normal when $n \to \infty$ and T is fixed, with **non-parametric convergence rate** $\sqrt{nh_n}$

The low convergence rate of this estimator is the cost for the weak assumptions imposed on the error terms

3.4 OUTLOOK

Random effects and fixed effects estimators have been developed for a variety of nonlinear panel models

Random effects: See e.g. Gourieroux, Monfort (1993), Keane (1994), Hajivassiliou, McFadden (1998)

- General applicability of the method based on the ML principle
- Computational difficulty due to the integral w.r.t. the random effects distribution appearing in the likelihood function
- Requires (sometimes quite restrictive) assumptions on the link between random effects and explanatory variables

Fixed effects: See Table 1 and the review in [AHO]

- Conceptual difficulty from the incidental parameters problem, insights behind the estimators are often model-specific
- Allows for general links between individual effects and explanatory variables
- Some fixed effects estimators have sub-parametric convergence rates

Table 1: Some fixed effects estimators for nonlinear panel models

Discrete choice	
Logit, multinomial logit	CML, Chamberlain (1980)
Semi-parametric	Maximum score, Manski (1987)
Tobit models	
Censored regression (Type 1)	Honore (1992)
Selection model (Type 2)	Kyriazidou (1997)
Type 3	Honore, Kyriazidou (2000a)
Dynamic models	
Dynamic logit	Honore, Kyriazidou (2000b)
Dynamic censored regression	Honore, Hu (2004)
Dynamic sample selection	Kyriazidou (2001)

Recently, much research considering the fixed-T inconsistency of the fixed effects ML estimator from the view point of a large-T bias expansion:

$$E[\hat{\beta}] = \beta + \frac{1}{T}B + O\left(\frac{1}{T^2}\right)$$

Bias adjusted estimators are obtained by eliminating the bias at order $O(1/T)$, e.g. by analytical methods:

$$\hat{\beta}_{adj} = \hat{\beta} - \frac{1}{T}\hat{B}$$

or by jackknife (see e.g. Hahn, Kuersteiner (2002), Hahn, Newey (2004), Dhaene, Jochmans, Thuysbaert (2006), Arellano, Bonhomme (2009), Fernandez-Val (2009))

While fixed-T consistency in not achieved, bias reduction can substantially improve the small-T properties of the fixed effects ML estimator. Moreover, the methodology has general applicability

REFERENCES

A) Books and review chapters

ARE Arellano, M. (2003): *Panel Data Econometrics*, Advanced Texts in Econometrics, Oxford University Press.

AHO Arellano, M., and B., Honore (2001): Panel Data Models: Some Recent Developments, in *Handbook of Econometrics*, Vol. 5, J. Heckman and E. Leamer (eds), Elsevier, Amsterdam.

CHA Chamberlain, G. (1984): Panel Data, in *Handbook of Econometrics*, Vol. 2, Z. Grilliches and M. Intriligator (eds), Elsevier, Amsterdam.

HSI Hsiao, C. (1986): *Analysis of Panel Data*, Econometric Society Monographs, Cambridge University Press.

MAS Matyas, L., and P., Sevestre (Eds.) (1992): *The Econometrics of Panel Data. Handbook of Theory and Applications*, Advanced Studies in Theoretical and Applied Econometrics, Kluwer Academic Publishers.

B) Articles

Anderson, T. W., and C., Hsiao (1981): ”Estimation of Dynamic Models with Error Components”, Journal of the American Statistical Association, 76, 598-606.

Anderson, T. W., and C., Hsiao (1982): ”Formulation and Estimation of Dynamic Models Using Panel Data”, Journal of Econometrics, 18, 47-82.

Arellano, M., and S., Bond (1991): ”Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”, Review of Economic Studies, 58, 277-297.

Arellano, M., and S., Bonhomme (2009): “Robust Priors in Nonlinear Panel Data Models”, Econometrica, 77, 489-536.

Arellano, M., and O., Bover (1995): ”Another Look at the Instrumental Variable Estimation of Error-Components Models”, Journal of Econometrics, 68, 29-51.

Balestra, P., and M., Nerlove (1966): ”Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas”, Econometrica, 34, 585-612.

Blundell, R., and S., Bond (1998): ”Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”, Journal of Econometrics, 87, 115-143.

Butler, I., and R., Moffitt (1982): ”A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model”, Econometrica, 50, 761-764.

Chamberlain, G. (1980): ”Analysis of Covariance with Qualitative Data”, Review of Economic Studies, 47, 225-238.

Dhaene, G., Jochmans, K., and B., Thuysbaert (2006): ”Jackknife Bias Reduction for Nonlinear Dynamic Panel Data Models with Fixed Effects”, Working Paper, Leuven University.

Fernandez-Val, I. (2009): ”Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Model”, Journal of Econometrics, 150, 71-85.

Gourieroux, C., Holly, A., and A. Monfort (1981): ”Kuhn-Tucker, Likelihood Ratio and Wald Tests for Nonlinear Models with Inequality Constraints on the Parameters”, Journal of Econometrics, 16, 166.

Gourieroux, C., Holly, A., and A. Monfort (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters", Econometrica, 50, 63-80.

Gourieroux, C., and A. Monfort (1993): "Simulation Based Inference: A Survey with Special Reference to Panel Data Models", Journal of Econometrics, 59, 5-33.

Hahn, J., and G., Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", Econometrica, 70, 1639-1657.

Hahn, J., and W., Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", Econometrica, 72, 1295-1319.

Hajivassiliou, V., and D., McFadden (1998): "The Method of Simulated Scores for the Estimation of LDV Models", Econometrica, 66, 863-896.

Hausman, J. (1978): "Specification Tests in Econometrics", Econometrica, 46, 1251-1271.

Hoch, I. (1962): "Estimation of Production Function Parameters Combining Time Series and Cross-Section Data", Econometrica, 30, 34-53.

Honore, B. (1992): ”Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects”, Econometrica, 60, 533-565.

Honore, B., and L., Hu (2004): ”Estimation of Cross-Sectional and Panel Data Censored Regression Models with Endogeneity”, Journal of Econometrics, 122, 293-316.

Honore, B., and E., Kyriazidou (2000a): ”Estimation of Tobit-Type Models with Individual Specific Effects”, Econometric Reviews, 19, 341-366.

Honore, B., and E., Kyriazidou (2000b): ”Panel Data Discrete Choice Models with Lagged Dependent Variables”, Econometrica, 68, 839-874.

Holtz-Eakin, D., Newey, W., and H., Rosen (1988): ”Estimating Vector Autoregressions with Panel Data”, Econometrica, 56, 1371-1395.

Keane, M. (1994): ”A Computationally Practical Simulation Estimator for Panel Data”, Econometrica, 62, 95-116.

Kyriazidou, E. (1997): ”Estimation of a Panel Data Sample Selection Model”, Econometrica, 65, 1335-1364.

Kyriazidou, E. (2001): "Estimation of Dynamic Panel Data Sample Selection Models", Review of Economic Studies, 68, 543-572.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data", Econometrica, 55, 357-362.

Mundlak, Y. (1961): "Empirical Production Functions Free of Management Bias", Journal of Farm Economics, 43, 69-85.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross-Section Data", Econometrica, 46, 69-85.

Neyman, J., and E., Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", Econometrica, 16, 1-31.

Newey, W., and F., Windmeijer (2009): "Generalized Method of Moments with Many Weak Moment Conditions", Econometrica, 77, 687-719.

Nickel, S. (1981): "Biases in Dynamic Models with Fixed Effects", Econometrica, 49, 1417-1426.

Wooldridge, J. (1995): "Selection Correction for Panel Data Models under Conditional Mean Independence Assumptions", Journal of Econometrics, 68, 115-132.