CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Fixed vs. Random Effects Panel Data Models: Revisiting the Omitted Latent Variables and Individual heterogeneity Arguments

Aris Spanos

CEA@Cass Working Paper Series

WP–CEA–07-2008

Fixed vs. Random Effects Panel Data Models: Revisiting the Omitted Latent Variables and Individual Heterogeneity Arguments

Aris Spanos*
Department of Economics
Virginia Tech
Blacksburg, VA 24061
USA

March 2008

Abstract

One of the most crucial questions in panel data modeling concerns the choice between Fixed (FEP) and Random Effects Panel (REP) data models. A variety of arguments have been proposed in the literature on choosing between the two formulations, but none of them makes a clear case for the circumstances under which each of these models will be appropriate. The primary objective of this paper is to put forward such an argument based on assumptions which are ascertainable vis-a-vis the data. This is achieved by relating the error assumptions to the probabilistic structure of the stochastic process underlying the observed data. It is argued that the omitted latent variables argument, when properly interpreted, provides an elucidating interpretation for the REP model and the individual heterogeneity argument does the same for the FEP model. By proposing complete specifications for both models in terms of testable probabilistic assumptions concerning the observable vector stochastic process underlying the data, the choice between the two models is metamorphosed into a statistical adequacy issue.

*This is a preliminary draft; please do not quote without the permission of the author.

1 Introduction

One of the most crucial questions in panel data modeling concerns the appropriateness of the *fixed* vs. *random effects* formulations in modeling individual heterogeneity in such data. It turns out that the choice between them has important implications for the consistency and/or efficiency of the estimators for the parameters of interest; see Arellano (2003), Baltagi (2005), Hsiao (2003). Despite the importance of this choice, there is a dearth of convincing arguments in the literature concerning the circumstances under which each formulation is appropriate. As a result of this the two formulations are often used as interchangeable, with the final choice often made using a Hausman-type specification test. However, as argued by Baltagi (2005), relating the result of the latter test to the choice between the two formulations raises its own problems, rendering the issue anything but clear cut.

It is argued that the source of the problem of choosing between fixed vs. random effects models lies with the specification of these panel data models in terms of unobservable error terms:

$$u_{it} = c_i + \epsilon_{it}, \qquad i \in \mathbb{N},\ t \in \mathbb{T}, \tag{1}$$

where c_i denotes unobserved individual-specific effects (fixed or stochastic), and ϵ_{it} denotes the remaining non-systematic effects. The two sets of probabilistic assumptions comprising the fixed and random effects models seem arbitrary because they concern unobservables whose validity cannot be ascertained at the specification stage (when choosing the model). Hence, the key to addressing this problem is to render the choice between the two formulations ascertainable vis-a-vis the data. This can be achieved by recasting the fixed and random effects models in terms of probabilistic assumptions about the observable stochastic processes involved, say:

$$\{\mathbf{Z}_{it}{:=}(y_{it}, \mathbf{X}_{it}),\ i \in \mathbb{N},\ t \in \mathbb{T}\},$$

and not the error term process $\{u_{it},\ i \in \mathbb{N},\ t \in \mathbb{T}\}$. This recasting transforms the problem of fixed vs. random effects models into one of *statistical adequacy* [are the model assumptions pertaining to $\{\mathbf{Z}_{it},\ i \in \mathbb{N},\ t \in \mathbb{T}\}$ valid for the particular data?] and elucidates several outstanding issues in this literature.

By providing a complete set of probabilistic assumptions in terms of the structure of the process $\{\mathbf{Z}_{it},\ i \in \mathbb{N},\ t \in \mathbb{T}\}$, and specifying the parameterizations of the unknown parameters for each model, one can discuss the similarities and differences between the different panel data models, including when the various estimators are consistent/inconsistent and efficient/inefficient. In addition, this recasting can elucidate the alternative interpretations of these models based on the omitted latent variables and/or the individual heterogeneity renditions. In particular, it is shown that the omitted latent variables interpretation is inappropriate for the fixed effects model, but it can provide an elucidating perspective for the random effects model. On the other hand, the individual heterogeneity interpretation can provide a coherent interpretation for the fixed effects model.

2 Textbook perspective on Panel Data models

It is widely appreciated that panel data models offer two major advantages for empirical modeling in economics because for reasonable values of N and T the sample size NT is quite large, which creates an opportunity for:

(a) less *restrictive* statistical models defining the premises of inference, and

(b) enhanced reliability and precision of statistical inference.

What is *not* widely appreciated is that statistical misspecification, the probabilistic assumptions comprising these models being invalid for the data in question, will abnegate both of these potential advantages.

2.1 Specification, estimation and testing

The basic two statistical models for panel data are given in tables 1-2 below.

Table 1: Fixed Effects Panel (FEP) data model
$y_{it} = c_i + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}, \quad i{\in}\mathbb{N},\ t{\in}\mathbb{T},$
[i] $E(u_{it}){=}0$, [ii] $E(u_{it}^2){=}\sigma_u^2$, [iii] $E(u_{it}u_{js}){=}0$, for $t{\neq}s$, $i{\neq}j$, $i,j{\in}\mathbb{N}$, $t,s{\in}\mathbb{T}$,

Table 2: Random Effects Panel (FEP) data model
$y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta}{+}\eta_i + \epsilon_{it}, \quad i{\in}\mathbb{N},\ t{\in}\mathbb{T},$
[i] $E(\epsilon_{it}){=}0$, [ii] $E(\epsilon_{it}^2){=}\sigma_\epsilon^2$, [iii] $E(\epsilon_{it}\epsilon_{js}){=}0$, for $t{\neq}s$, $i{\neq}j$, [iv] $E(\eta_i){=}0$,
[v] $E(\eta_i^2){=}\sigma_\eta^2$, [vi] $E(\eta_i\eta_j){=}0$, for $i{\neq}j$, [vii] $E(\epsilon_{it}\eta_j){=}0$, for all $i,j{\in}\mathbb{N}$, $t{\in}\mathbb{T}$,

where $i{\in}\mathbb{N}{:=}\{1,2,...,N,...\}$ and $i{\in}\mathbb{T}{:=}\{1,2,...,T,...\}$ denote the cross-section and time dimensions.

These two models are usually viewed by the literature as interchangeable, with components c_i and η_i representing generic terms which aim to capture heterogeneity in the data relating to the cross-section dimension $i = 1,2,...,N$. On the nature of the heterogeneity terms, the prevailing view is that they should both be treated as random variables, with the crucial difference being that the REP model raises the possibility that $\mathbf{X}_{it}$ might be correlated with η_i; see Wooldridge (2002). As they stand, the probabilistic assumptions of the two models are not ascertainable vis-a-vis the data $\mathbf{Z}^* := (\mathbf{y}^* : \mathbf{X}^*)$ because both (c_i, η_i) and the error terms (u_{it}, ϵ_{it}) are unobservable.

Using an obvious notation for all T observations, the FEP data model, as specified in table 1, takes the form:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}c_i{+}\mathbf{u}_i,\ i = 1,..,N,$$

where $\mathbf{y}_i$ is $(T \times 1)$, $\mathbf{X}_i$ is $(T \times k)$, $\mathbf{1} :=(1,1,...,1)^\top$, $\mathbf{u}_i$ is $(T \times 1)$, with the covariance structure is of the form:

$$E(\mathbf{u}_i^\top \mathbf{u}_i) = \mathbf{\Sigma} = \begin{pmatrix} \sigma_\epsilon^2+\sigma_\eta^2 & \sigma_\eta^2 & \cdots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\epsilon^2+\sigma_\eta^2 & & \sigma_\eta^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \sigma_\eta^2 & \cdots & \sigma_\epsilon^2+\sigma_\eta^2 \end{pmatrix},$$

where $\mathbf{\Sigma}$ is a $T \times T$ matrix. One can express this model for all TN observations in matrix notation as:

$$\underbrace{\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}}_{\mathbf{y}^*} = \underbrace{\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{pmatrix}}_{\mathbf{X}^*}\boldsymbol{\beta}+\underbrace{\begin{pmatrix} \mathbf{1} & 0 & \cdots & 0 \\ 0 & \mathbf{1} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1} \end{pmatrix}}_{\mathbf{D}}\underbrace{\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}}_{\mathbf{c}} + \underbrace{\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{pmatrix}}_{\mathbf{u}^*}$$

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Dc} + \mathbf{u}^*.$$

The OLS (Dummy Variable) estimators of $(\boldsymbol{\beta}, \mathbf{c})$ are:

$$\widehat{\boldsymbol{\beta}}= \left(\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{y}^*, \qquad \widehat{\mathbf{c}}= \left(\mathbf{D}^\top\mathbf{M}_X\mathbf{D}\right)^{-1}\mathbf{D}^\top\mathbf{M}_X\mathbf{y}^*,$$

where the projection matrices $(\mathbf{M}_D, \mathbf{M}_X)$ are: $\mathbf{M}_D= \mathbf{I} - \mathbf{D}\left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{D}$, $\mathbf{M}_X= \mathbf{I} - \mathbf{X}^*\left(\mathbf{X}^{*\top}\mathbf{X}^*\right)^{-\mathbf{1}}\mathbf{X}^{*\top}$. Expressing the REP model for all NT observations takes the form:

$$\underbrace{\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}}_{\mathbf{y}^*} = \underbrace{\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{pmatrix}}_{\mathbf{X}^*}\boldsymbol{\beta} + \underbrace{\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{pmatrix}}_{\mathbf{u}^*} \qquad \Omega= \left[E(\mathbf{u}_i^\top\mathbf{u}_j)\right]_{i,j}^N = \underbrace{\begin{pmatrix} \mathbf{\Sigma} & 0 & \cdots & 0 \\ 0 & \mathbf{\Sigma} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{\Sigma} \end{pmatrix}}_{(\mathbf{I}_N \otimes \mathbf{\Sigma})},$$

where $\Omega=(\mathbf{I}_N \otimes \mathbf{\Sigma})$ is a $TN \times TN$ matrix. Hence, the GLS estimator of $\boldsymbol{\beta}$ takes the form:

$$\widetilde{\boldsymbol{\beta}} = \left(\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^*.$$

In practice, Ω is unknown and needs to be estimated using a feasible GLS procedure to yield $\widetilde{\boldsymbol{\beta}}_F = (\mathbf{X}^{*\top}\widehat{\Omega}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\widehat{\Omega}^{-1}\mathbf{y}^*$; see Baltagi (2005), Hsiao (2002).

The choice between the FEP and REP models in tables 1 and 2 is often based on a Hausman-type specification test whose test statistic relies on the 'distance' between the two estimators:

$$H = (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_F)^\top[Cov(\widehat{\boldsymbol{\beta}}) - Cov(\widetilde{\boldsymbol{\beta}}_F)]^{-1}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_F).$$

In addition to the above OLS and GLS estimators, several other estimators have been proposed in the literature, including the *pooled OLS*, the *between*, the *within* and the *first difference estimators*, with the discussion focusing on the circumstances under which each of these estimators is consistent and/or efficient; see Cameron and Trivedi (2005). The focus of this discussion revolves around whether the error term in each model is orthogonal to the explanatory variables $\mathbf{X}_{it}$ or not – orthogonality ensures the consistency of the estimator – and the form of the covariance matrix.

Consistent estimator of what? What is *not* clearly spelled out in this literature is the explicit form of the coefficients $\boldsymbol{\beta}$ that $(\widehat{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\beta}})$ are consistent estimators of. To see the problem consider the simple linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \ \mathbf{u} \backsim \mathsf{NIID}(\mathbf{0}, \sigma_u^2 \mathbf{I}_n),$$

where $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ but also $E(\mathbf{u} \mid \mathbf{Z}) = \mathbf{0}$; $Cov(\mathbf{Z}_t, \mathbf{X}_t) = \Sigma_{32} \neq \mathbf{0}$, $Cov(\mathbf{Z}_t) = \Sigma_{33} > 0$; $\mathbf{y} : (T \times 1)$, $\mathbf{X} : (T \times k)$, $\mathbf{Z} : (T \times k)$, rank$(\mathbf{X}^\top \mathbf{X}) = k$ =rank$(\mathbf{Z}^\top \mathbf{X}) = k$. The conventional wisdom suggests that both the OLS estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ and the Instrumental Variable (IV) estimator $\widetilde{\boldsymbol{\beta}} = (\mathbf{Z}^\top \mathbf{X})^{-1} \mathbf{Z}^\top \mathbf{y}$ are consistent estimators of $\boldsymbol{\beta}$. As the traditional argument goes, one can substitute $\mathbf{y}$ into the estimators and show that:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\tfrac{\mathbf{X}^\top \mathbf{X}}{n})^{-1} \tfrac{\mathbf{X}^\top \mathbf{u}}{n}, \qquad \widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\tfrac{\mathbf{Z}^\top \mathbf{X}}{n})^{-1} \tfrac{\mathbf{Z}^\top \mathbf{u}}{n}.$$

Consistency is then derived (mainly) from the fact that:

$$p\lim(\tfrac{\mathbf{X}^\top \mathbf{u}}{n}) = \mathbf{0}, \qquad p\lim(\tfrac{\mathbf{Z}^\top \mathbf{u}}{n}) = \mathbf{0}.$$

This argument, however, is highly misleading because $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ are consistent estimators of very different parameters. In particular,

$$\widehat{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{21}, \qquad \widetilde{\boldsymbol{\beta}} \overset{\mathbb{P}}{\to} \boldsymbol{\alpha} = \boldsymbol{\Sigma}_{32}^{-1} \boldsymbol{\sigma}_{31},$$

where $\boldsymbol{\Sigma}_{22} = Cov(\mathbf{X}_t)$, $\boldsymbol{\sigma}_{21} = Cov(\mathbf{X}_t, y_t)$, $\boldsymbol{\Sigma}_{32} = Cov(\mathbf{Z}_t, \mathbf{X}_t)$, and $\boldsymbol{\sigma}_{31} = Cov(\mathbf{Z}_t, y_t)$, with $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$ in general.

The proper way to view the consistency argument is to focus on the observables (not the errors), with the argument taking the form:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = (\tfrac{\mathbf{X}^\top \mathbf{X}}{n})^{-1} \tfrac{\mathbf{X}^\top \mathbf{y}}{n} \overset{\mathbb{P}}{\to} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\sigma}_{21}, \qquad \widetilde{\boldsymbol{\beta}} = (\tfrac{\mathbf{Z}^\top \mathbf{X}}{n})^{-1} \tfrac{\mathbf{Z}^\top \mathbf{y}}{n} \overset{\mathbb{P}}{\to} \boldsymbol{\Sigma}_{32}^{-1} \boldsymbol{\sigma}_{31}.$$

These results follow from $p\lim(\frac{\mathbf{X}^\top \mathbf{X}}{n}) = \boldsymbol{\Sigma}_{22}$, $p\lim(\frac{\mathbf{X}^\top \mathbf{y}}{n}) = \boldsymbol{\sigma}_{21}$, $p\lim(\frac{\mathbf{Z}^\top \mathbf{X}}{n}) = \boldsymbol{\Sigma}_{32}$ and $p\lim(\frac{\mathbf{Z}^\top \mathbf{y}}{n}) = \boldsymbol{\sigma}_{31}$, in conjunction with Slutsky's theorem; see White (2001).

In general, claims about consistency in panel data models should be treated with caution when they rely on certain orthogonality with the error term assumptions, but are not accompanied by the underlying parameterizations.

3 The Probabilistic Reduction (PR) perspective

The Probabilistic Reduction (PR) perspective views statistical models as *parameterizations* of the probabilistic structure of the underlying observable stochastic process $\{\mathbf{Z}_{it}{:=}(y_{it}, \mathbf{X}_{it}),\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$, which is defined on a probability space $(S, \mathcal{F}, \mathbb{P}(.))$. In particular, given that this probabilistic structure can be fully described using the joint distribution $D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi})$, one can view statistical models arising as reductions from this joint distribution; see Spanos (1986, 1999). To illustrate this perspective consider the simplest statistical model for panel data.

3.1 The Pooled Panel Data (PPD) model

Let us assume that the vector stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ is *Normal, Independent and Identically Distributed (NIID)*. The NIID *reduction assumptions* imply that the joint distribution of this process, say $D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi})$, can be simplified as follows:

$$\begin{aligned} D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi}) &\overset{\mathsf{I}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D_{it}(\mathbf{Z}_{it}; \boldsymbol{\varphi}(i,t)) \overset{\mathsf{IID}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{Z}_{it}; \boldsymbol{\varphi}) = \\ &\overset{\mathsf{IID}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{y}_{it}|\ \mathbf{x}_{it}; \boldsymbol{\varphi}_1) \cdot D(\mathbf{X}_{it}; \boldsymbol{\varphi}_2). \end{aligned} \tag{2}$$

In addition, Normality implies that $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi})$ is:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}\begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{21}^{\top} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right), \quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

Hence, the *regression* and *skedastic functions* take the form:

$$E(y_{it}|\ \mathbf{X}_{it} = \mathbf{x}_{it}) = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta}, \qquad Var(y_{it}|\ \mathbf{X}_{it} = \mathbf{x}_{it}){=}\sigma_u^2, \tag{3}$$

where the model parameters are:

$$\beta_0 = \mu_1 - \boldsymbol{\mu}_2^{\top}\boldsymbol{\beta}, \quad \boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}, \quad \sigma_u^2 = \sigma_{11} - \boldsymbol{\sigma}_{21}^{\top}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}. \tag{4}$$

The reduction in (2) gives rise to a statistical model known as the **pooled panel data (PPD) model**:

$$\boxed{y_{it} = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it},\ (u_{it}|\ \mathbf{x}_{it}) \backsim \mathsf{NIID}(0, \sigma_u^2),\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}.} \tag{5}$$

The complete specification of this model in terms of the observable stochastic processes is given in table 3.

An important dimension of the specification of statistical models is the **statistical Generating Mechanism** (GM) (table 3). This is based on an *orthogonal decomposition* of the form:

$$y_{it} = \mu_{it} + u_{it}, \text{ for } i{\in}\mathbb{N},\ t{\in}\mathbb{T},$$

$$\mu_{it} = E(y_{it} \mid \mathcal{D}_{it}) \text{ and } u_{it} = y_{it} - E(y_{it} \mid \mathcal{D}_{it}),$$

where μ_{it} represents the **systematic** and u_{it} the **non-systematic** (error) component, with $\mathcal{D}_{it} \subset \mathcal{F}$ specifying the relevant *conditioning information set* that would render $\{(u_{it}, \mathcal{D}_{it})\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ a *Martingale Difference* (MD) process; see White (2001). In the case of the PPD model (table 3) the statistical GM was based on $\mathcal{D}_{it} = \{\mathbf{X}_{it} = \mathbf{x}_{it}\}$.

Table 3 - Pooled Panel Data (PPD) model		
Statistical GM:		$y_{it} = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}.$
[1]	Normality:	$(y_{it} \vert\ \mathbf{X}_{it} = \mathbf{x}_{it}) \backsim \mathsf{N}(.,.)$,
[2]	Linearity:	$E\left(y_{it} \vert\ \mathbf{X}_{it} = \mathbf{x}_{it}\right) = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta}$,
[3]	Homoskedasticity:	$Var\left(y_{it} \vert\ \mathbf{X}_{it} = \mathbf{x}_{it}\right) = \sigma_u^2$,
[4]	Independence:	$\{(y_{it} \vert\ \mathbf{X}_{it} = \mathbf{x}_{it}),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ independent,
[5]	(i,t)-invariance:	$(\beta_0, \boldsymbol{\beta}, \sigma_u^2)$ are (i,t)-invariant.

3.2 The FEP and REP models from the PR perspective

The question that naturally arises at this stage is whether there is a way one can view the FEP and REP models given in tables 1 and 2, respectively, in the context of the PR perspective. The answer is provided by seeking the appropriate probabilistic structure for the stochastic process $\{\mathbf{Z}_{it}{:=}(y_{it}, \mathbf{X}_{it}),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ which would give rise to the two models, with their respective statistical GMs taking the form:

$$\textbf{FEP: } \mu_{it} = c_i + \mathbf{x}_{it}^{\top}\boldsymbol{\beta},\ u_{it}, \quad \textbf{REP: } \mu_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta},\ u_{it} = \eta_i + \epsilon_{it}. \tag{6}$$

Equivalently, one needs to address the issue as to how the terms (c_i, η_i) pertain to the underlying probabilistic structure of the observable stochastic processes $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, by specifying the relevant conditioning information set $\mathcal{D}_{it}$ in each case that will yield the systematic and non-systematic components given in (6).

In the panel data literature there are several alternative arguments being used to explain how the terms (c_i, η_i) could be rationalized, but they are often equivocal in the sense that they do not provide convincing justifications which can be directly assessed vis-a-vis the data $\mathbf{Z}^*$; see Wooldridge (2002). Indeed, some of these arguments can be misleading, as the discussion in the next section explains.

In the next two sections we will consider the two most widely used arguments, the *omitted latent variables* and *individual heterogeneity* arguments, first articulated by Chamberlain (1982, 1984).

4 Revisiting the omitted latent variables argument

Consider now an extension of the Probabilistic Reduction specification based on the joint distribution $D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi})$, where there might be several omitted, but potentially relevant, **latent factors** Ξ_i. In order to relate the latent vector Ξ_i to $(y_{it}, \mathbf{X}_{it})$ we extend the relevant stochastic process to $\{\mathbf{Z}^*_{it}:=(y_{it}, \mathbf{X}_{it}, \Xi_i),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, assumed to be NIID, with the joint distribution $D(y_{it}, \mathbf{X}_{it}, \Xi_i; \boldsymbol{\phi})$ taking the form:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \\ \Xi_i \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_{13} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}\right), \quad i\in\mathbb{N},\ t\in\mathbb{T}. \tag{7}$$

4.1 The Fixed Effects Panel (FEP) data model

For the FEP data model in table 1, which treats the individual effects $\{c_i,\ i\in\mathbb{N}\}$ as constants, it makes sense to consider:

$$\mathcal{D}_{it} = (\mathbf{X}_{it}= \mathbf{x}_{it}, \Xi_i= \boldsymbol{\xi}_i)\,.$$

as the *relevant conditioning information set*. The resulting regression and skedastic functions take the form:

$$E(y_{it}|\ \mathbf{X}_{it}= \mathbf{x}_{it},\ \Xi_i= \boldsymbol{\xi}_i)=\alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \boldsymbol{\xi}_i^\top\boldsymbol{\gamma}, \qquad Var(y_{it}|\ \mathbf{X}_{it}= \mathbf{x}_{it},\ \Xi_i= \boldsymbol{\xi}_i)=\sigma_\varepsilon^2, \tag{8}$$

where the model parameters $\boldsymbol{\theta} := (\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma_\varepsilon^2)$ take the form:

$$\begin{aligned}
&\alpha_0 = \mu_1 - \boldsymbol{\mu}_2^\top\boldsymbol{\alpha} - \boldsymbol{\mu}_\mathbf{3}^\top\boldsymbol{\gamma},\\
&\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{2.3}^{-1}(\boldsymbol{\sigma}_{\mathbf{21}} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31})= \boldsymbol{\beta} - \boldsymbol{\Delta\gamma},\\
&\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{3.2}^{-1}(\boldsymbol{\sigma}_{31}-\boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21})= \boldsymbol{\delta} - \mathbf{D}\boldsymbol{\alpha},\\
&\sigma_\varepsilon^2=\sigma_u^2- \left[\left(\boldsymbol{\sigma}_{13}-\boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23}\right)\boldsymbol{\Sigma}_{3.2}^{-1}\left(\boldsymbol{\sigma}_{13}-\boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23}\right)^\top\right],
\end{aligned} \tag{9}$$

$$\begin{aligned}
&\boldsymbol{\delta} := \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31},\ \boldsymbol{\Delta} := \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23},\ \mathbf{D} := \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32},\\
&\boldsymbol{\Sigma}_{3.2}:= \boldsymbol{\Sigma}_{33}-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}_{23},\ \boldsymbol{\Sigma}_{2.3}:= \boldsymbol{\Sigma}_{22}-\mathbf{D}^\top\boldsymbol{\Sigma}_{32};
\end{aligned} \tag{10}$$

see Spanos (2006) for the details. It is interesting to note that without the omitted latent variables Ξ_i, the probabilistic structure of NIID for $\{\mathbf{Z}_{it}:=(y_{it}, \mathbf{X}_{it}),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ gives rise to the PPD model in table 3, where the model parameterizations are different, unless the latent variables Ξ_i is *uncorrelated* with both observable variables $(y_{it}, \mathbf{X}_{it})$:

$$\boldsymbol{\sigma}_{31}= \mathbf{0} \text{ and } \boldsymbol{\Sigma}_{23}= \mathbf{0}. \tag{11}$$

That is, $(\alpha_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \sigma_\varepsilon^2)|_{\boldsymbol{\sigma}_{31}=\mathbf{0}\&\boldsymbol{\Sigma}_{23}=\mathbf{0}} = (\beta_0, \mathbf{0}, \boldsymbol{\beta}, \sigma_u^2)$. Note that $(\alpha_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \sigma_\varepsilon^2)|_{\boldsymbol{\Sigma}_{23}=\mathbf{0}} = (\beta_0, \mathbf{0}, \boldsymbol{\beta}, \sigma_w^2)$, where:

$$\sigma_w^2 = \sigma_u^2-\boldsymbol{\sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{13} \neq \sigma_\varepsilon^2.$$

There are two interesting special cases arising from (8) depending on whether $\boldsymbol{\xi}_i$ is observed or not.

Case 1: $\boldsymbol{\xi}_i$ **is observed**. When all the variables are observed the least-squares estimators $\widehat{\boldsymbol{\theta}}:=(\widehat{\alpha}_0, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, s_\varepsilon^2)$:

$$\widehat{\boldsymbol{\alpha}}=\left(\mathbf{X}^{*\top}\mathbf{M}_{\boldsymbol{\Xi}}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_{\boldsymbol{\Xi}}\mathbf{y}^*, \qquad \widehat{\boldsymbol{\gamma}}=\left(\boldsymbol{\Xi}^{*\top}\mathbf{M}_X\boldsymbol{\Xi}^*\right)^{-1}\boldsymbol{\Xi}^{*\top}\mathbf{M}_X\mathbf{y}^*,$$

$$\widehat{\alpha}_0=\overline{y}-\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\alpha}}-\overline{\boldsymbol{\xi}}^\top\widehat{\boldsymbol{\gamma}}, \qquad s_\varepsilon^2=\tfrac{1}{(NT-k-m-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-\widehat{\alpha}_0-\mathbf{x}_{it}^\top\widehat{\boldsymbol{\alpha}}-\boldsymbol{\xi}_i^\top\widehat{\boldsymbol{\gamma}}\right)^2$$

converge in probability ($\widehat{\boldsymbol{\theta}}\overset{\mathbb{P}}{\to}\boldsymbol{\theta}$) to their respective parameterizations in (9). This case is of interest only in so far as it sheds light on the latent case discussed next.

Case 2: $\boldsymbol{\xi}_i$ **is unobserved.** When $\boldsymbol{\Xi}_i$ is *latent* $c_i=\boldsymbol{\xi}_i^\top\boldsymbol{\gamma}$, $i\in\mathbb{N}$, cannot be observed giving rise to a regression function with latent individual fixed effects:

$$E(y_{it}|\ \mathbf{X}_{it}=\mathbf{x}_{it}, \boldsymbol{\Xi}_i=\boldsymbol{\xi}_i)=\alpha_0+\mathbf{x}_{it}^\top\boldsymbol{\alpha}+\boldsymbol{\xi}_i^\top\boldsymbol{\gamma}=\alpha_0+\mathbf{x}_{it}^\top\boldsymbol{\alpha}+c_i.$$

This gives rise to the **statistical model** with unobserved fixed individual effects:

$$\begin{gathered} y_{it}=\alpha_0+\mathbf{x}_{it}^\top\boldsymbol{\alpha}+c_i+\varepsilon_{it},\ (\varepsilon_{it}|\mathcal{D}_{it})\backsim\mathsf{NIID}(0,\sigma_\varepsilon^2),\ i\in\mathbb{N},\ t\in\mathbb{T}, \\ \mathcal{D}_{it}:=\{\mathbf{X}_{it}=\mathbf{x}_{it}, \boldsymbol{\Xi}_i=\boldsymbol{\xi}_i\}\ i\in\mathbb{N},\ t\in\mathbb{T}. \end{gathered} \tag{12}$$

The question that needs to be posed is whether the FEP model given in table 1 can be interpreted in the context of (12) which views the fixed effects factor c_i as a linear combination of the observed values of the omitted latent variables $\boldsymbol{\Xi}_i$, i.e. $c_i=\boldsymbol{\xi}_i^\top\boldsymbol{\gamma}$, $i\in\mathbb{N}$. There are two basic problems with this interpretation argument. The *first* is that conditioning on the *observed value* of an unobservable variable ($\boldsymbol{\Xi}_i=\boldsymbol{\xi}_i$) is *conceptually problematic*; it constitutes an oxymoron even when contemplated as a hypothetical scenario. The *second*, and more practical problem, concerns the question 'what is the fixed effects OLS (Dummy Variable) estimator $\widehat{\boldsymbol{\beta}}=\left(\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{y}^*$ a consistent estimator of?' It is obvious that it's *not* a consistent estimator of $\boldsymbol{\alpha}=\left(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32}\right)^{-1}(\boldsymbol{\sigma}_{21}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31})$, given in (9), because $\mathbf{M}_D\neq\mathbf{M}_{\boldsymbol{\Xi}}$. Hence, as it stands, the omitted latent factors interpretation of the FEP given in (12) seems inappropriate.

4.2 The Random Effects Panel (REP) Data Model

For the REP model in table 2, which treats the individual effects $\{\eta_i,\ i\in\mathbb{N}\}$ as random variables, it makes sense to replace conditioning on $\boldsymbol{\Xi}_i=\boldsymbol{\xi}_i$ with conditioning on the σ-*field* generated by the latent variable $\boldsymbol{\Xi}_i$, say $\sigma(\boldsymbol{\Xi}_i)$. That is, the relevant conditioning information set is now:

$$\mathcal{D}_{it}^*=(\mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i)).$$

This conditioning makes more sense than conditioning on $\{\boldsymbol{\Xi}_i=\boldsymbol{\xi}_i\}$ because the σ-field simply acknowledges the events associated with $\boldsymbol{\Xi}_i$ by restricting the σ-field underlying the universal probability space $(S, \mathcal{F}, \mathbb{P}(.))$ by conditioning on $\sigma(\boldsymbol{\Xi}_i)$ since

$\sigma(\Xi_i) \subset \mathcal{F}$. It also explains why conditioning on $\{\Xi_i = \boldsymbol{\xi}_i\}$ is problematic, since for any random variable W defined on the same probability space $(S, \mathcal{F}, \mathbb{P}(.))$, the random variable $E(W|\sigma(\Xi_i))$ does not depend on the actual values $\boldsymbol{\xi}_i$ of Ξ_i, $i\in\mathbb{N}$. This is because for any Borel function $h()$ such that $h(\boldsymbol{\xi}_i) \neq h(\boldsymbol{\xi}_j)$ when $\boldsymbol{\xi}_i \neq \boldsymbol{\xi}_j$, for all $i, j\in\mathbb{N}$, i.e. it keeps the values of Ξ_i distinct:

$$\sigma(\Xi_i)=\sigma(h(\Xi_i)) \Rightarrow E(W|\sigma(\Xi_i))=E(W|\sigma(h(\Xi_i)));$$

see Renyi (1970), p. 259. As shown in Spanos (1986), p. 413, conditioning on $\mathcal{D}^*_{it}$ yields the *stochastic* linear regression and skedastic functions:

$$E(y_{it}|\mathcal{D}^*_{it})=\alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \Xi_i^\top\boldsymbol{\gamma} = \boldsymbol{\alpha}_0+\mathbf{x}_{it}^\top\boldsymbol{\alpha}+\eta_i, \qquad Var(y_{it}|\mathcal{D}^*_{it})=\sigma^2_\epsilon, \tag{13}$$

where Ξ_i denotes the random vector itself. In contrast to $c_i = \boldsymbol{\xi}_i^\top\boldsymbol{\gamma}$ in (8), the term $\eta_i= \Xi_i^\top\boldsymbol{\gamma}$ in (13) is stochastic with a distribution:

$$(\eta_i|\mathcal{D}^*_{it}) \backsim \mathsf{N}(\eta,\sigma^2_\eta), \text{ where } \eta= \boldsymbol{\mu}_3^\top\boldsymbol{\gamma},\ \sigma^2_\eta= \boldsymbol{\gamma}^\top\boldsymbol{\Sigma}_{33}\boldsymbol{\gamma}, \tag{14}$$

where the parameterization of $(\alpha_0, \boldsymbol{\alpha})$ coincides with that in (8); see (9). This gives rise to the **statistical model with a latent random effect**:

$$y_{it} = \alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \eta_i + \epsilon_{it},\ (\epsilon_{it}|\mathcal{D}^*_{it}) \backsim \mathsf{NIID}(0, \sigma^2_\epsilon),\ (\eta_i|\mathcal{D}^*_{it}) \backsim \mathsf{NIID}(\eta,\sigma^2_\eta),\ i\in\mathbb{N},\ t\in\mathbb{T},$$

$$\mathcal{D}^*_{it} = (\mathbf{X}_{it}= \mathbf{x}_{it}, \sigma(\Xi_i)),\ i\in\mathbb{N},\ t\in\mathbb{T}. \tag{15}$$

Taking the mean deviation $\eta_i^*=\eta_i-\boldsymbol{\mu}_3^\top\boldsymbol{\gamma}$ form, one can re-defines the constant to $\alpha_0^*=\mu_1-\boldsymbol{\mu}_2^\top\boldsymbol{\alpha}$, to yield the modified **statistical model**:

$$\begin{gathered} y_{it} = \alpha_0^* + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \eta_i^* + \epsilon_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}, \\ (\epsilon_{it}|\mathcal{D}^*_{it}) \backsim \mathsf{NIID}(0, \sigma^2_\epsilon),\ (\eta_i^*|\mathcal{D}^*_{it}) \backsim \mathsf{NIID}(0,\sigma^2_\eta),\ i\in\mathbb{N},\ t\in\mathbb{T}. \end{gathered} \tag{16}$$

The question that naturally arises at this stage, is whether the specifications (15) or (16) can be viewed as providing a meaningful interpretation for the REP model in table 2. The surprising answer is no, because as in the case of the fixed effects model, the estimator $\widetilde{\boldsymbol{\beta}} = \left(\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^*$ is *not* a consistent estimator of $\boldsymbol{\alpha} = \left(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32}\right)^{-1}(\boldsymbol{\sigma}_{21}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31})$. So, what is $\widetilde{\boldsymbol{\beta}}$ a consistent estimator of? In what sense $\widetilde{\boldsymbol{\beta}}$ 'captures' the potential effect of the omitted variables Ξ_i? Does this imply that the omitted latent variables specification in (15) is also inappropriate for the REP model in table 2? These are crucial questions that need to be addressed.

An interesting attempt to answer these questions was made by Mundlak (1978), which is considered next.

4.2.1 Mundlak's random effects formulation

Motivated by the fact that there is no reason to assume that $E(\eta_i|\,\mathbf{X}_{it}=\mathbf{x}_{it})=0$, in general, Mundlak (1978) argued that the crucial difference between the fixed and random effects formulation is *not* the randomness of the latter, but the potential correlation between η_i and $\mathbf{X}_{it}$; see also Wooldridge (2002). To capture that correlation he introduced the *auxiliary regression*:

$$\eta_i = \sum_{t=1}^{T} \boldsymbol{\delta}_1^\top \mathbf{x}_{it} + \omega_i = \boldsymbol{\delta}_1^\top \overline{\mathbf{x}}_i + \omega_i, \quad \omega_i \backsim \mathsf{NIID}(0, \sigma_\omega^2), \tag{17}$$

where $\overline{\mathbf{x}}_i=\frac{1}{n}\sum_{t=1}^{T}\mathbf{x}_{it},\ i=1,2,...,n,\ E(\eta_i|\,\mathbf{X}_{it}=\mathbf{x}_{it})=\sum_{t=1}^{T}\boldsymbol{\delta}_1^\top\mathbf{x}_{it}$. When (17) is substituted back into the original formulation it gives rise to the **Mundlak formulation**:

$$\begin{array}{c} y_{it} = \alpha_0^\dagger + \mathbf{x}_{it}^\top \boldsymbol{\alpha} + \boldsymbol{\delta}_1^\top \overline{\mathbf{x}}_i + \omega_i + \epsilon_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}, \\ (\epsilon_{it}|\,\mathbf{x}_{it}) \backsim \mathsf{NIID}(0,\sigma_\epsilon^2),\ (\omega_i|\,\mathbf{x}_{it}) \backsim \mathsf{NIID}(0,\sigma_\omega^2),\ i\in\mathbb{N},\ t\in\mathbb{T}. \end{array} \tag{18}$$

As argued by Hsiao (2003), this formulation captures the correlation between η_i and $\mathbf{X}_{it}$, but raises other questions concerning the consistency and efficiency of the resulting estimators under different scenarios. Of particular interest for the present discussion is the question whether the GLS estimator $\widehat{\boldsymbol{\alpha}}$ in (18) (see Hsiao, 2003) is a consistent estimator of $\boldsymbol{\alpha}$ as specified in (9). As shown next, the surprising answer is no! So, what is $\widehat{\boldsymbol{\alpha}}$ a consistent estimator of?

4.2.2 A PR latent random effects specification

To answer the latter question one needs to take a more systematic way in defining Mundlak's auxiliary regression. Let us return to the joint distribution $D(y_{it}, \mathbf{X}_{it}, \boldsymbol{\Xi}_i; \boldsymbol{\phi})$ in (7) and consider the relationship between $\mathbf{X}_{it}$ and $\boldsymbol{\Xi}_i$, as it pertains to their correlation in the context of the model in (15). When $\boldsymbol{\Sigma}_{23}\neq\mathbf{0}$ there exists an auxiliary regression between $\boldsymbol{\Xi}_i$ and $\mathbf{X}_{it}$ of the form:

$$\boldsymbol{\Xi}_i = \boldsymbol{\delta}_0 + \boldsymbol{\Delta}^\top \mathbf{x}_{it} + \mathbf{v}_i, \quad (\mathbf{v}_i|\,\mathbf{x}_{it}) \backsim \mathsf{NIID}(\mathbf{0}, \boldsymbol{\Sigma}_{3.2}), \tag{19}$$

$$E(\boldsymbol{\Xi}_i \mid \mathbf{X}_{it}=\mathbf{x}_{it})=\delta_0 + \mathbf{x}_{it}^\top\boldsymbol{\Delta},\quad \boldsymbol{\delta}_0=\boldsymbol{\mu}_3-\boldsymbol{\Delta}^\top\boldsymbol{\mu}_2,\quad \boldsymbol{\Delta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23},\quad \boldsymbol{\Sigma}_{3.2}:=\boldsymbol{\Sigma}_{33}-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}_{23}.$$

Hence, a convenient way to eliminate $\boldsymbol{\Xi}_i$ from (13) is to substitute (19) into (16) yielding:

$$\begin{aligned} y_{it} = \alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \boldsymbol{\Xi}_i^\top\boldsymbol{\gamma} + \epsilon_{it} &= \alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + (\boldsymbol{\delta}_0 + \boldsymbol{\Delta}^\top\mathbf{x}_{it} + \mathbf{v}_i)^\top\boldsymbol{\gamma} + \epsilon_{it} = \\ &= \left[\alpha_0 + \boldsymbol{\delta}_0^\top\boldsymbol{\gamma}\right] + \mathbf{x}_{it}^\top\left[\boldsymbol{\alpha} + \boldsymbol{\Delta}\boldsymbol{\gamma}\right] + \mathbf{v}_i^\top\boldsymbol{\gamma} + \epsilon_{it}. \end{aligned} \tag{20}$$

Given $\alpha_0=\mu_1-\boldsymbol{\mu}_2^\top\boldsymbol{\alpha}-\boldsymbol{\mu}_3^\top\boldsymbol{\gamma}$, and $\boldsymbol{\delta}_0^\top\boldsymbol{\gamma}=\boldsymbol{\mu}_3^\top\boldsymbol{\gamma}-\boldsymbol{\mu}_2^\top\boldsymbol{\Delta}\boldsymbol{\gamma}$:

$$\left[\alpha_0 + \boldsymbol{\delta}_0^\top\boldsymbol{\gamma}\right] = \mu_1 - \boldsymbol{\mu}_2^\top\left[\boldsymbol{\Delta}\boldsymbol{\gamma} + \boldsymbol{\alpha}\right] = \mu_1 - \boldsymbol{\mu}_2^\top\boldsymbol{\beta} = \beta_0, \quad \left[\boldsymbol{\alpha} + \boldsymbol{\Delta}\boldsymbol{\gamma}\right] = \boldsymbol{\beta},$$

from (9). These simplifications imply that:

$$y_{it} = \alpha_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\alpha} + \boldsymbol{\Xi}_i^{\top}\boldsymbol{\gamma} + \epsilon_{it} \quad = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \mathbf{v}_i^{\top}\boldsymbol{\gamma} + \epsilon_{it}, \tag{21}$$

and give rise to a **statistical model with latent random effects** of the form**:**

$$\begin{gathered} y_{it} = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \upsilon_i + \epsilon_{it}, \qquad i\in\mathbb{N},\ t\in\mathbb{T}, \\ (\epsilon_{it}|\mathcal{D}_{it}^{*}) \backsim \mathsf{NIID}(0, \sigma_{\epsilon}^2),\ (\upsilon_i|\ \mathbf{x}_{it}) \backsim \mathsf{NIID}(0,\sigma_{\nu}^2),\ i\in\mathbb{N},\ t\in\mathbb{T}. \end{gathered} \tag{22}$$

where in view of the fact that $\upsilon_i = \mathbf{v}_i^{\top}\boldsymbol{\gamma}$:

$$\sigma_{\nu}^2 = \boldsymbol{\gamma}^{\top}\boldsymbol{\Sigma}_{3.2}\boldsymbol{\gamma} = (\boldsymbol{\sigma}_{31} - \boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21})^{\top}\boldsymbol{\Sigma}_{3.2}^{-1}(\boldsymbol{\sigma}_{31} - \boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}). \tag{23}$$

Note that in light of (9), σ_{ν}^2 is related to the variances of u_{it} and ϵ_{it} via the orthogonal decomposition:

$$\sigma_u^2 = \sigma_{\epsilon}^2 + \sigma_{\nu}^2 = \sigma_{11} - \boldsymbol{\sigma}_{21}^{\top}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}. \tag{24}$$

The complete set of assumptions comprising the model in (22) is given in table 4.

Table 4 - Random Effects Panel (REP) data model

Statistical GM:		$y_{it} = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \upsilon_i + \epsilon_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}.$
[1]	Normality:	$\begin{cases} (y_{it}\|\ \mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i)) \backsim \mathsf{N}(.,.), \\ (\upsilon_i\|\ \mathbf{X}_{it}=\mathbf{x}_{it}) \backsim \mathsf{N}(.,.), \end{cases}$
[2]	Linearity:	$E\left(y_{it}\|\ \mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i)\right) = \beta_0 + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \upsilon_i,$
[3]	Homoskedasticity:	$\begin{cases} Var\left(y_{it}\|\ \mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i)\right) = \sigma_{\epsilon}^2, \\ Var\left(\upsilon_i\|\ \mathbf{X}_{it}=\mathbf{x}_{it}\right) = \sigma_{\nu}^2, \end{cases}$
[4]	Independence:	$\{(y_{it}\|\left(\mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i)\right)),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ independent,
[5]	(i,t)-invariance:	$(\beta_0, \boldsymbol{\beta}, \sigma_{\nu}^2, \sigma_{\epsilon}^2)$ are (i,t)-invariant.

The question now is whether the statistical model in table 4 provides a pertinent interpretation for the REP model with $\mathcal{D}_{it}^{*} = (\mathbf{X}_{it}=\mathbf{x}_{it}, \sigma(\boldsymbol{\Xi}_i))$ being the relevant conditioning information set.

To begin with the statistical parameterization of this model, $\boldsymbol{\theta} := (\beta_0, \boldsymbol{\beta}, \sigma_{\nu}^2, \sigma_{\epsilon}^2)$:

$$\beta_0 = \mu_1 - \boldsymbol{\mu}_2^{\top}\boldsymbol{\beta}, \quad \boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}, \quad \sigma_{\varepsilon}^2 = \sigma_u^2 - \sigma_{\nu}^2, \quad \sigma_{\nu}^2 = (\boldsymbol{\sigma}_{31} - \boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21})^{\top}\boldsymbol{\Sigma}_{3.2}^{-1}(\boldsymbol{\sigma}_{31} - \boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}),$$

is clearly the relevant one. The estimator $\widetilde{\boldsymbol{\beta}} = (\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^{*}$ is consistent for $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$ (not $\boldsymbol{\alpha} = (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32})^{-1}(\boldsymbol{\sigma}_{21} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31})$), and it's also efficient because it takes account of the variance-covariance structure in Ω; the same applies to the estimators of $(\beta_0, \sigma_{\nu}^2, \sigma_{\epsilon}^2)$.

The primary difference between (16) and (22) is that in the latter case $\mathbf{X}_{it}$ is orthogonal ($\perp$) to $(\upsilon_i, \epsilon_{it})$ by construction, eliminating the original (potential) correlation between $\mathbf{X}_{it}$ and Ξ_i. Indeed, this orthogonality is crucial in achieving the proper parameterization and the resulting consistency of the relevant estimators of $(\beta_0, \boldsymbol{\beta})$. Moreover, the same orthogonality is instrumental in being able to estimate consistently the relevant conditional variance σ_ε^2. In relation to the latter, it is important to emphasize that although one *cannot* estimate $\boldsymbol{\alpha}$ consistently in the regression $E\left(y_{it}|\ \mathbf{X}_{it}{=}\ \mathbf{x}_{it}, \sigma(\Xi_i)\right) {=}\alpha_0 + \mathbf{x}_{it}^\top\boldsymbol{\alpha} + \boldsymbol{\Xi}_i^\top\boldsymbol{\gamma}$ when the omitted variables Ξ_i are latent, one *can* estimate the relevant conditional variance $Var\left(y_{it}|\ \mathbf{X}_{it}{=}\ \mathbf{x}_{it}, \sigma(\Xi_i)\right) {=}\sigma_\epsilon^2$, by decomposing (orthogonally) σ_u^2, the conditional variance associated with $(\beta_0, \boldsymbol{\beta})$, into: $\sigma_u^2 = \sigma_\epsilon^2 + \sigma_\nu^2$; see (24).

In summary, viewing the REP model table 2 in terms of the underlying stochastic process $\{\mathbf{Z}_{it}^*{:=}(y_{it}, \mathbf{X}_{it}, \Xi_i),\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$, provides a coherent interpretation where:

(a) an unambiguous interpretation of the various terms in the statistical GM, including the errors,

(b) an explicit statistical parameterization for all the unknown parameters, and

(c) a complete set of testable assumptions (table 4) which pertain to the probabilistic structure of the observable process $\{\mathbf{Z}_{it}{:=}(y_{it}, \mathbf{X}_{it}),\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$.

4.2.3 Revisiting the Mundlak formulation

Given the above discussion, it is interesting to revisit the Mundlak (1978) formulation where he introduced the auxiliary regression (17) as follows:

"The properties of the various estimators to be considered depend on the existence and extent of the relations between the X's and the effects.
In order to take an explicit account of such relationships we introduce the auxiliary regression (2.3) $\eta_i = \mathbf{x}_{it}^\top\boldsymbol{\pi} + \omega_{it}$; averaging over t for a given i :
(2.4) $\eta_i = \overline{\mathbf{x}}_i^\top\boldsymbol{\pi} + \omega_i$, (2.5) $\omega_i \backsim (0, \sigma_\omega^2)$." (pp. 71-72) [the notation is changed]

He went on to note that:

"$E(\eta_i|\ \mathbf{x}_{it})$ need not be linear. However, only the linear expression is pertinent for the present analysis." (p. 71)

Viewing the Mundlak auxiliary regression in light of the specification in (19), three issues arise. *First*, the linearity and homoskedasticity he imposed are (together) equivalent to assuming that $\mathbf{X}_{it}$ and Ξ_i are jointly Normal (see Spanos, 1995). This renders the original assumption that $D(y_{it}, \mathbf{X}_{it}, \Xi_i; \boldsymbol{\phi})$ is multivariate Normal a very reasonable working assumption. *Second*, the averaging is clearly unnecessary to secure the invariance of the error with respect to t, since, by definition the error term is the non-systematic component of η_i, not accounted for by conditioning on $\mathbf{X}_{it}$, $\omega_i = \eta_i - E(\eta_i|\ \mathbf{x}_{it})$, which does not change with t. *Third*, a minor point is that there is a missing constant term in (2.3) and (2.4) which captures the linear combination of the means of η_i and $\mathbf{X}_{it}$.

Putting all these pieces together, and noting that $\eta_i{=}\ \boldsymbol{\Xi}_i^\top\boldsymbol{\gamma}$, one can relate Mundlak's (2.3) to the auxiliary regression in (19) via:

$$\eta_i = \boldsymbol{\Xi}_i^\top \boldsymbol{\gamma} = \boldsymbol{\delta}_0^\top \boldsymbol{\gamma} + \mathbf{x}_{it}^\top \boldsymbol{\Delta} \boldsymbol{\gamma} + \mathbf{v}_i^\top \boldsymbol{\gamma}. \tag{25}$$

Substituting (25) into (15) yields the model in (22) since:

$$\begin{aligned} y_{it} = \alpha_0 + \mathbf{x}_{it}^\top \boldsymbol{\alpha} + \eta_i + \epsilon_{it} &= \alpha_0 + \mathbf{x}_{it}^\top \boldsymbol{\alpha} + \left[\boldsymbol{\delta}_0^\top \boldsymbol{\gamma} + \mathbf{x}_{it}^\top \boldsymbol{\Delta}\boldsymbol{\gamma} + \mathbf{v}_i^\top \boldsymbol{\gamma}\right] + \epsilon_{it} \\ &= \left[\alpha_0 + \boldsymbol{\delta}_0^\top \boldsymbol{\gamma}\right] + \mathbf{x}_{it}^\top \left[\boldsymbol{\alpha} + \boldsymbol{\Delta}\boldsymbol{\gamma}\right] + \mathbf{v}_i^\top \boldsymbol{\gamma} + \epsilon_{it} \\ &= \beta_0 + \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{v}_i^\top \boldsymbol{\gamma} + \epsilon_{it}, \end{aligned}$$

where the simplification follows directly from (20). That is, when the Mundlak auxiliary regression is properly specified, it gives rise to the same formulation as in (22).

This suggests that the specification of the REP model proposed in table 4 provides a most pertinent interpretation of the original specification in table 2, shedding light on several issues raised by the original specification in terms of unobservable error terms. In particular, contrary to conventional wisdom, the random effects term c_i should be interpreted, *not* as representing the omitted latent variables $\boldsymbol{\Xi}_i$, but as a linear combination of the errors from the auxiliary regression of $\mathbf{X}_{it}$ on $\boldsymbol{\Xi}_i$, i.e.

$$\nu_i = \mathbf{v}_i^\top \boldsymbol{\gamma} = \left(\boldsymbol{\Xi}_i - \boldsymbol{\delta}_0 - \boldsymbol{\Delta}^\top \mathbf{x}_{it}\right)^\top \boldsymbol{\gamma}.$$

This brings out the danger in choosing arbitrarily the interpretation one would like to attribute to unobserved terms and the unknown parameters of a statistical model, while ignoring the structure of the underlying observable processes involved and the associated parameterizations.

5 Revisiting the heterogeneity argument

Having rejected the omitted latent variables interpretation for the fixed effects (FEP) model as given in table 1, we proceed to explore the heterogeneity argument from the Probabilistic Reduction (PR) perspective in an attempt to find out whether one can relate the fixed effects to the heterogeneity of the process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$.

5.1 Unrestricted heterogeneity and panel data modeling

Assume that the vector stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, $\mathbf{Z}_{it}:=(y_{it}, \mathbf{X}_{it}^\top)^\top$ be *Normal, Independent (NI) but* ***non-Identically Distributed,*** i.e. heterogeneous with respect to both (i,t). These probabilistic assumptions imply the reduction:

$$\begin{aligned} D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi}) &\overset{\mathsf{I}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D_{it}(\mathbf{Z}_{it}; \boldsymbol{\varphi}(i,t)) \overset{\mathsf{NI}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{Z}_{it}; \boldsymbol{\varphi}(i,t)) = \\ &\overset{\mathsf{NI}}{=} \prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{y}_{it} \mid \mathbf{x}_{it}; \boldsymbol{\varphi}_1(i,t)) \cdot D(\mathbf{X}_{it}; \boldsymbol{\varphi}_2(i,t)) \end{aligned} \tag{26}$$

where $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi}_1(i,t))$ is multivariate Normal of the form:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1(i,t) \\ \boldsymbol{\mu}_2(i,t) \end{pmatrix} \begin{pmatrix} \sigma_{11}(i,t) & \boldsymbol{\sigma}_{21}^\top(i,t) \\ \boldsymbol{\sigma}_{21}(i,t) & \boldsymbol{\Sigma}_{22}(i,t) \end{pmatrix}\right), \quad i\in\mathbb{N},\ t\in\mathbb{T}.$$

The regression and skedastic functions associated with $D(\mathbf{y}_{it}|\ \mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))$ are:

$$E(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it})=\beta_0(i,t)+\mathbf{x}_{it}^{\top}\boldsymbol{\beta}(i,t), \quad Var(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it})=\sigma_u^2(i,t), \tag{27}$$

where the model parameters take the form:

$$\beta_0(i,t)=\mu_1(i,t)-\boldsymbol{\mu}_2^{\top}(i,t)\boldsymbol{\beta}(i,t), \quad \boldsymbol{\beta}(i,t)=\boldsymbol{\Sigma}_{22}^{-1}(i,t)\boldsymbol{\sigma}_{21}(i,t),$$

$$\sigma_u^2(i,t)=\sigma_{11}(i,t)-\boldsymbol{\sigma}_{12}(i,t)\boldsymbol{\Sigma}_{\mathbf{22}}^{\mathbf{-1}}(i,t)\boldsymbol{\sigma}_{21}(i,t).$$

This gives rise to the *non-estimable* **statistical model**:

$$\boxed{y_{it}=\beta_0(i,t)+\mathbf{x}_{it}^{\top}\boldsymbol{\beta}(i,t)+u_{it},\ (u_{it}|\mathbf{x}_{it})\backsim\mathsf{NIID}(0,\sigma_u^2(i,t)),\ i\in\mathbb{N},\ t\in\mathbb{T}.}$$

For $k=4$, $N=100$, $T=25$, the unknown model parameters are:

$$\beta_0(i,t), \quad \boldsymbol{\beta}(i,t), \quad \sigma_u^2(i,t), \quad i=1,..,N,\ t=1,..,T,$$

whose total number is $m=(K+2)NT=(6)(100)(25)=15000$! To render this model estimable one needs to impose certain restrictions on the form of the heterogeneity of $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$.

5.2 Fixed Individual Effects Panel (FEP) data model

Consider the case where vector stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ is assumed to be *Normal, Independent (NI) but mean heterogeneous* (but covariance homogeneous) with respect to $i\in\mathbb{N}$, and *completely homogeneous* with respect to $t\in\mathbb{T}$. These probabilistic assumptions imply the reduction:

$$\begin{aligned} D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{I}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D_{it}(\mathbf{Z}_{it};\boldsymbol{\varphi}(i))\overset{\mathsf{NI}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it};\boldsymbol{\varphi}(i))= \\ &\overset{\mathsf{NI}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{y}_{it}|\ \mathbf{x}_{it};\boldsymbol{\varphi}_1(i))\cdot D(\mathbf{X}_{it};\boldsymbol{\varphi}_2(i)), \end{aligned} \tag{28}$$

where $D(y_{it},\mathbf{X}_{it};\boldsymbol{\varphi}(i))$ is multivariate Normal of the form:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1(i) \\ \boldsymbol{\mu}_2(i) \end{pmatrix}\begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right), \quad i\in\mathbb{N},\ t\in\mathbb{T}.$$

The regression and skedastic functions associated with $D(\mathbf{y}_{it}|\ \mathbf{x}_{it};\boldsymbol{\varphi}_1(i))$ are:

$$E(y_{it}|\ \mathbf{X}_{it}=\mathbf{x}_{it})=\beta_0(i)+\mathbf{x}_{it}^{\top}\boldsymbol{\beta}, \quad Var(y_{it}|\ \mathbf{X}_{it}=\mathbf{x}_{it})=\sigma_u^2, \tag{29}$$

where the model parameters are:

$$\beta_0(i)=\mu_1(i)-\boldsymbol{\mu}_2^{\top}(i)\boldsymbol{\beta}, \quad \boldsymbol{\beta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}, \quad \sigma_u^2=\sigma_{11}-\boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{\mathbf{22}}^{\mathbf{-1}}\boldsymbol{\sigma}_{21}.$$

This gives rise to the (potentially) *estimable* **statistical model**:

$$y_{it} = \beta_0(i) + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it},\ i\in\mathbb{N},\ (u_{it}|\mathbf{x}_{it}) \backsim \mathsf{NIID}(0, \sigma_u^2),\ i\in\mathbb{N},\ t\in\mathbb{T}.$$

The complete specification of this model in terms of the observable stochastic processes is given in table 5.

Table 5 - Fixed Individual Effects Panel data model

Statistical GM: $y_{it} = \beta_0(i) + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}.$

[1]	Normality:	$(y_{it}\|\ \mathbf{X}_{it} = \mathbf{x}_{it}) \backsim \mathsf{N}(.,.)$,
[2]	Linearity:	$E\left(y_{it}\|\ \mathbf{X}_{it} = \mathbf{x}_{it}\right) = \beta_0(i) + \mathbf{x}_{it}^{\top}\boldsymbol{\beta}$,
[3]	Homoskedasticity:	$Var\left(y_{it}\|\ \mathbf{X}_{it} = \mathbf{x}_{it}\right) = \sigma_u^2$,
[4]	Independence:	$\{(y_{it}\|\ \mathbf{X}_{it} = \mathbf{x}_{it}),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ independent,
[5]	(a) (i,t)-invariance:	$(\boldsymbol{\beta}, \sigma_u^2)$ are (i,t)-invariant,
	(b) t-invariance:	$\{\beta_0(i),\ i\in\mathbb{N}\}$ are t-invariant, but $i-$heterogeneous

This can be an estimable model because for, say $k = 4$, $N = 100$, $T = 25$, $NT = 2500$, the unknown model parameters are:

$$\beta_0(i) = c_i,\quad \boldsymbol{\beta},\quad \sigma_u^2,\quad i{=}1,..,N,\ t{=}1,..,T,$$

whose total number is $m = N + k + 2 = (100 + 4 + 2) = 106$.

The important thing to emphasize about the above (heterogeneity motivated) specification is that one can state unequivocally is that the OLS estimators, $\widehat{\boldsymbol{\beta}}{=}\left(\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{y}^{*}$ and $\widehat{\mathbf{c}}{=}\left(\mathbf{D}^{\top}\mathbf{M}_X\mathbf{D}\right)^{-1}\mathbf{D}^{\top}\mathbf{M}_X\mathbf{y}^{*}$ of $(\boldsymbol{\beta}, \mathbf{c})$, are *consistent estimators* of the relevant parameterizations. In particular,

(a) $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, where:

$$\boldsymbol{\Sigma}_{22}{=}E\left([\mathbf{X}_{it}{-}\boldsymbol{\mu}_2(i)][\mathbf{X}_{it}{-}\boldsymbol{\mu}_2(i)]^{\top}\right),\qquad \boldsymbol{\sigma}_{21}{=}E\left([\mathbf{X}_{it}{-}\boldsymbol{\mu}_2(i)][y_{it}{-}\mu_1(i)]\right),$$

(b) $\widehat{c}_i$ is a consistent estimator of $\beta_0(i){=}\mu_1(i){-}\boldsymbol{\mu}_2^{\top}(i)\boldsymbol{\beta}$, for all $i{=}1,2,...,N$,

(c) $s_u^2 = \frac{1}{NT-N-k-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \widehat{c}_i - \mathbf{x}_{it}^{\top}\widehat{\boldsymbol{\beta}})^2$ is a consistent estimator of σ_u^2.

6 Revisiting the Fixed vs. Random Effects models

Having proposed two different interpretations for the fixed and random effects models in tables 1 and 2, respectively, based on specifying the assumptions comprising the two models (tables 4-5) in terms of the probabilistic structure of the observable stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, it is important to bring out certain crucial advantages of the latter specification.

At the **specification** stage (when choosing the appropriate statistical model), the specifications of the FEP and REP models given in tables 1 and 2 provide no basis for choosing between these models because assumptions [i]-[iii] and [i]-[vii] cannot be related to the structure of the data. In contrast, the specifications given in tables 3 and 4 in terms of probabilistic assumptions concerning $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, are ascertainable vis-a-vis data $\mathbf{Z}^* := (\mathbf{y}^*: \mathbf{X}^*)$ using a variety of graphical techniques. This is because reduction assumptions such as NIID, or NI but non-ID are easy to assess directly using data plots. In assessing dependence and/or heterogeneity one needs to decide on a particular ordering of the data. In the case of the time dimension $t\in\mathbb{T}$, it is naturally considered to provide *the* ordering of interest, but in the case of the cross-section dimension $i\in\mathbb{N}$, it is sometimes argued that the ordering does not matter; it does! In practice, there are several natural orderings that one needs to explore with respect to $i\in\mathbb{N}$, such as the size of the firm, geographical position of the city, etc., etc. For each of those possible orderings of interest one could assess the presence or absence of i-dependence and/or i-heterogeneity by a variety of graphical techniques, the simplest of which is the t-plot; see Spanos (1999), ch. 5-6. This suggests that the choice between the FEP and REP models, as specified in tables 3 and 4 can be made on the basis of whether the process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ exhibits i-heterogeneity or not.

At the **Mis-Specification (M-S) testing** stage the probabilistic assumptions [1]-[5] for both models in tables 3 and 4 are all testable vis-a-vis data $\mathbf{Z}^*$, and should be tested thoroughly for possible departures in order to secure the reliability of inference. Applying Hausman-type specification tests is no substitute for thorough M-S testing because any departures from assumptions [1]-[4] of either model will undermine the reliability of this test in the sense that the *nominal* size and power will be very different from the *actual*, giving rise to misleading inferences; see Spanos (2006). The original preliminary data analysis based on graphical techniques that guided the specification is also useful in deciding the type os M-S tests one should apply in order to probe for potential misspecifications in a thorough manner. Ensuring **statistical adequacy** (the validity of the model assumptions vis-a-vis data $\mathbf{Z}^*$) is of paramount importance for any form of statistical inference, because without it the reliability of inference is questionable. Hence, without statistical adequacy the primary advantages of panel data, including enhanced precision of inference and less restrictive models, are abnegated.

A potential criticism of the PR specifications of the FEP and REP models is that the probabilistic assumptions [1]-[5] are rather restrictive and unrealistic for many data sets in practice. As argued in Spanos (2008) such arguments are misplaced and highly misleading on a number of counts. In particular, these are not the only assumptions one can impose on the underlying process; the PR approach makes explicit the model assumptions that can give rise to reliable and precise inferences and one should modify these assumptions to ensure their validity vis-a-vis the particular data. The traditional approach's relactance to make explicit distributional assumptions carries a heavy price in terms of the precision of inference without any guarantees for reliability. Similiarly, nonparametric models foresake reliability by invoking non-testable assumptions and relying on asymptotic results whose validity is not established; see Spanos (2001). For reliable and precise inference there is no substitute to fully specified parametric statistical models whose statistical adequacy is secured by thorough M-S testing and respecification before any inferences are drawn. To paraphrase Peirce (1878), 'there is no royal road to *learning from data* about phenomena of interest, and really valuable ideas can only be had at the price of securing statistical adequacy'.[1]

Where the Probabilistic Reduction (PR) perspective shines is at the **respecification** stage where one needs to respecify the original model when it is found to be wanting. Again the graphical analysis can be very useful in suggesting alternative specifications that take account of systematic information the original model could not account for. Spanos (2007) provides a more comprehensive discussion that includes several additional statistical models of interest in panel data modeling that arise as respecifications of the FEP and REP models given in tables 4 and 5.

7 Conclusion

The paper has proposed specifications of the Fixed (FEP) and Random Effects Panel (REP) models (tables 4-5) in terms of the probabilistic structure of the observable stochastic processes $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ underlying the observed data. The results bring out the danger in choosing arbitrarily the interpretation one would like to attribute to unobserved effect terms (c_i, η_i) . The proposed specifications show that the interpretation of these terms is inextricably bound up with the statistical parameterizations of these models. Moreover, such specifications offer certain distinct advantages over the traditional specifications based on error terms (tables 1-2), in relation to the problems of specification, Mis-Specification (M-S) testing and respecification.

Specification: choosing an appropriate statistical model. The choice is now informed since the probabilistic assumptions concerning $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, are ascertainable vis-a-vis data $\mathbf{Z}^* := (\mathbf{y}^*\!: \mathbf{X}^*)$ using a variety of graphical techniques. One needs to decide on *the* particular ordering(s) of interest over the cross-section dimension $i\in\mathbb{N}$, (size, geographical position, etc.), and then assess the presence/absence of i-dependence and/or i-heterogeneity using a variety of graphical techniques; see Spanos (1999), ch. 5-6. The choice between the FEP and REP models can be (initially) made on the basis of whether the process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ exhibits i-heterogeneity or not.

Mis-Specification (M-S) testing. The probabilistic assumptions [1]-[5] for both models in tables 3 and 4 are all testable vis-a-vis data $\mathbf{Z}^*$. The original Preliminary Data Analysis, based on graphical techniques that guided the specification, is also useful in deciding the type os M-S tests. Securing *statistical adequacy* is par-

[1] The original phrase "there is no royal road to geometry" is attributed to Euclid in reply to King Ptolemy's request for an easier way of learning mathematics. Charles S. Peirce (1878) paraphrasing was: "There is no royal road to logic, and really valuable ideas can only be had at the price of close attention."

ticularly important in panel data models in order to take advantage of the primary benefits of panel data, including enhanced precision of inference and less restrictive models. Indeed, without *statistical adequacy* no learning from data can take place.

Applying *Hausman-type tests* is no substitute for thorough M-S testing because any departures from assumptions [1]-[4] of either model will undermine the reliability of the Hausman test in the sense that the *nominal* size and power will be very different from the *actual*, giving rise to misleading inferences; see Spanos (2006).

A potential criticism of the PR specifications of the FEP and REP models is that the probabilistic assumptions [1]-[5] are *rather restrictive and unrealistic* for many data sets in practice. As argued in Spanos (2001, 2008), however, for reliable and precise inference that can give rise to learning from data about economic phenomena, there is no substitute to fully specified parametric statistical models whose statistical adequacy is secured before any inferences are drawn.

Where the Probabilistic Reduction (PR) perspective shines is at the **respecification** stage when respecifying the above models to account for the presence of i-heterogeneity/dependence or/and t-heterogeneity/dependence, as well as non-Normality; see Spanos (2007).

References

[1] Arellano, M. (2003), *Panel Data Econometrics*, Cambridge University Press, Cambridge.

[2] Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, 3rd ed., Wiley, NY.

[3] Cameron, A. C. and P. K. Trivedi (2005), *Microeconometrics : Methods and Applications*, Cambridge University Press, Cambridge.

[4] Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, **18**, 5-46.

[5] Chamberlain, G. (1984), "Panel Data," in Griliches, Z. and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[6] Hsiao, C. (2002), *Analysis of Panel Data*, 2nd ed., , Cambridge University Press, Cambridge.

[7] Mundlak, Y. (1978), "On the pooling of time series and cross-sectional data," *Econometrica*, **46**, pp. 69-86.

[8] Peirce, C. S. (1878), "'How to Make Our Ideas Clear," *Popular Science Monthly*, pp. 286-302. Reprinted in *Chance, Love, and Logic: Philosophical Essays*, edited by M. R. Cohen, Bison Books, 1998.

[9] Renyi, A. (1970), *Foundations of Probability*, Holden-Day, San Francisco.

[10] Spanos, A., (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.

[11] Spanos, A. (1995), "On Normality and the Linear Regression model", *Econometric Reviews*, **14**, 195-203.

[12] Spanos, A. (1999), *Probability Theory and Statistical Inference: econometric modeling with observational data*, Cambridge University Press, Cambridge.

[13] Spanos, A. (2001), "Parametric versus Non-parametric Inference: Statistical Models and Simplicity," pp. 181-206 in *Simplicity, Inference and Modelling*, edited by A. Zellner, H. A. Keuzenkamp and M. McAleer, Cambridge University Press.

[14] Spanos, A. (2006), "Revisiting the Omitted Variables Argument: Substantive vs. Statistical Reliability of Inference," *Journal of Economic Methodology*, **13**: 179-218.

[15] Spanos, A. (2007), "Revisiting the Statistical Foundations of Panel Data Models," Working Paper, Virginia Tech.

[16] Spanos, A. (2008), "Philosophy of Econometrics," forthcoming in *Philosophy of Economics*, the *Handbook of Philosophy of Science*, Elsevier (editors) D. Gabbay, P. Thagard, and J. Woods.

[17] White, H. (2001), *Asymptotic Theory for Econometricians*, Academic Press, London.

[18] Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge.