Revisiting the Statistical Foundations of Panel Data Modeling

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

Last revision: December 2011

Abstract

Despite the impressive developments in panel data modeling, the statistical foundations of such models are rather weak in so far that they are inadequate for securing the reliability and precision of inference. In statistical induction we learn from data about phenomena of interest when we employ reliable and incisive inference procedures. When one invokes either untested probabilistic assumptions, or/and broad (including nonparametric) premises, one has to rely on crude approximations (asymptotic) for evaluating the relevant error probabilities. This invariably leads to imprecise inference of unknown reliability because one does not know how closely the *actual* error probabilities approximate the assumed *nominal* ones, or how *effective* the inference procedures are! The primary objective of the paper is to revisit these probabilistic foundations with a view to: (a) recast the error assumptions in terms of the probabilistic structure of the observable processes underlying the data, (b) provide a complete and internally consistent set of testable probabilistic assumptions for several statistical models for panel data, and (c) propose pertinent interpretations for the individual-specific (fixed or random) and time-specific effects. It is shown that the current interpretations of the individual-specific effects (fixed or random) need to be reconsidered in light of the implicit statistical parameterizations in terms of the observable stochastic processes involved. This provides a more appropriate framework for securing learning from panel data by bringing out the neglected facets of empirical modeling which include specification, misspecification testing and respecification.

1 Introduction

The literature on statistical models for panel data has experienced a enormous growth over the last 20 years with several recent textbooks focusing on that subfield of econometrics; see Arellano (2003), Baltagi (2005), Cameron and Trivedi (2005), Hsiao (2003), Wooldridge (2010). The growth is mostly focused on developing asymptotically justifiable estimation techniques by making probabilistic assumptions for error terms that allow for certain forms of heterogeneity and dependence.

Panel data, expressed in the form:

$$\mathbf{Z}_0:=\{\mathbf{z}_{it},\ i=1,2,...,N,\ t=1,2,...,T\}, \tag{1}$$

where $\mathbf{z}_{it}:=\left(\begin{array}{cc} y_{it} & \mathbf{x}_{it}^{\top}\end{array}\right)^{\top}$, $(k+1)\times 1$, combine cross-section $(i=1,2,...,N)$ and time series $(t=1,2,...,T)$ data in so far as they vary both over *individual cross-section units* $(i\in\mathbb{N})$ *and* over *time* $(t\in\mathbb{T})$. The primary advantage of panel data is that for reasonable values of N and T the sample size NT is quite large; for $N=100$, $T=25$, $NT=2500$. The great advantage of panel data is that there is more than one realization of the observable processes, making it possible to allow for various forms of heterogeneity/dependence that are not possible for cross-section or time-series data separately. This creates the opportunity for:
(a) less *restrictive* statistical models defining the inductive premises of inference, and
(b) enhanced *reliability* and *precision* of statistical inferences based on such models.

For instance, the most basic error structure for panel data models is:

$$y_{it}=\mathbf{x}_{it}^{\top}\boldsymbol{\beta}+u_{it},\ u_{it}=c_i+\epsilon_{it},\qquad i\in\mathbb{N},\ t\in\mathbb{T}, \tag{2}$$

where c_i denotes the unobserved individual-specific effects (fixed or stochastic), and ϵ_{it} denotes the remaining non-systematic effects. The large sample size enhances the precision of estimators (point or interval) and increases the power of tests.

What is often ignored in this literature is that panel data can only give rise to enhanced learning from data relating to (a)-(b) when the assumed model is *statistically adequate*; the probabilistic assumptions underlying the statistical model are valid for data $\mathbf{Z}_0$. Any departures from these assumptions will often undermine the reliability as well as the precision of inference, foiling any learning from data. Precision in inference results from employing the most optimal (effective) inference procedures stemming from statistically adequate models. The precision of inference is derailed when one uses broad statistical premises that rely exclusively on asymptotic inference procedures whose inductive premises have not been validated. The unreliability of inference due to statistical misspecification comes in the form of discrepancies between the actual and nominal error probabilities associated with particular inferences. The surest way to lead an inference astray is to use .05 significance level test, when the actual type I error probability is closer to .9. As demonstrated in Spanos and McGuirk (2001) and Spanos (2009) such a big discrepancy can easily arise in cases where modelers might consider as minor departures from the model assumptions.

Despite the impressive development of statistical techniques in analyzing panel data (Wooldridge, 2010), the statistical foundations of panel data modeling are rather weak in so far as the current textbook perspective is inadequate for securing the reliability and precision of inference based on panel data models.

In statistical induction we learn from data about phenomena of interest when we employ reliable and incisive inference procedures; see Spanos (2007a). To be able to ensure the reliability and precision of inference one needs to strengthen the following facets of modeling:

(i) **Specification:** a complete, internally consistent and testable set of assumptions that specify the statistical premises of inference,

(ii) **Mis-Specification (M-S) testing:** strategies for thorough probing of potential departures from the model assumptions, and

(iii) **Respecification:** effective ways to respecify the initial premises when any of the assumptions are found to be invalid.

The weaknesses in the foundations arise primarily from the Pre-Eminence of Theory (PET) perspective that dominates current model specification in econometrics; see Spanos (2006a, 2010a-c). Typically, a traditional textbook modeler begins with a *theory model*, and turns it into a *statistical (econometric) model* by:

[i] viewing the theory model as the *systematic component* and

[ii] attaching random error terms to define the *non-systematic* component.

The PET perspective offers a way to use statistical inference techniques by introducing the inductive premises via probabilistic assumptions pertaining to error terms. These assumptions are chosen to justify the 'quantification' of the theory model by ensuring the existence of Consistent and Asymptotically Normal (CAN) estimators for unknown parameters of the model. In the Fixed Effects Panel (FEP) data model, the probabilistic assumptions about the error term are:

$$E(u_{it})=0,\ E(u_{it}^2)=\sigma_u^2,\ E(u_{it}u_{js})=0,\ \text{for } t\neq s,\ i\neq j,\ i=1,2,...,N,\ t=1,2,...,T.$$

For the Random Effects Panel (REP) data model, the error term $u_{it}=\eta_i+\epsilon_{it}$ assumptions are:

$$E(\epsilon_{it})=0,\ E(\epsilon_{it}^2)=\sigma_\epsilon^2,\ E(\epsilon_{it}\epsilon_{js})=0, \text{ for } t\neq s,\ i\neq j,$$
$$E(\eta_i)=0,\ E(\eta_i^2)=\sigma_\eta^2,\ E\left(\eta_i\eta_j\right)=0, \text{ for } i\neq j,\ E\left(\epsilon_{it}\eta_j\right)=0, \text{ for all } i,j \text{ and } t.$$

Could one assess the validity of these probabilistic assumptions vis-a-vis data $\mathbf{Z}_0$? The answer is clearly not! Indeed, it is obvious that the connection between the error assumptions and the probabilistic structure of the observable vector stochastic process:

$$\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}, \tag{3}$$

underlying data $\mathbf{Z}_0$, is neither obvious nor direct. As a result, the connection between the individual effects (fixed c_i or stochastic η_i) and the probabilistic structure of (3) is rather tenuous. In addition, there is little emphasis on validating the error assumptions, and almost no guidance as to what to do next when these assumptions are invalid for data $\mathbf{Z}_0$, apart from 'adjusting' the error term structure. When one invokes either untested probabilistic assumptions, or/and broad (including non-parametric) premises, one usually has to rely on *crude approximations* (asymptotic results or upper bounds) for evaluating the relevant error probabilities. This invariably leads to imprecise inference of unknown reliability because one does not know how closely the *actual* error probabilities approximate the *nominal* (assumed) ones, or how *effective* the inference procedures are!

A major hurdle in doing a good job with the specification, M-S testing and respecification in panel data modeling has been the difficulty to relate the error assumptions to the probabilistic structure of the observable process (3). The primary focus of this paper is to use the Probabilistic Reduction perspective to address this problem directly and shed light on a number of issues bedeviling panel data models, including the interpretation of fixed and random effects terms. From the PR perspective a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{x})$ is viewed as parameterizations of the observable stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ giving rise to data $\mathbf{Z}_0$. That is, data $\mathbf{Z}_0$ is viewed as a realization of a sample $\mathbf{Z}:=\{\mathbf{Z}_{it},\ i=1,2,...,N,\ t=1,2,...,T\}$ from a prespecified statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$, generically specified by:

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})=\{f(\mathbf{z};\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta\},\ \mathbf{z}\in\mathbb{R}_X^{NT},\ \text{for }\boldsymbol{\theta}\in\Theta\subset\mathbb{R}^m,\ m\ll NT, \tag{4}$$

where $f(\mathbf{z};\boldsymbol{\theta})$ denotes the (joint) *distribution of the sample* $\mathbf{Z}$; see Spanos (2006b).

Section 2 gives a brief summary of the current textbook discussion of the two most basic panel data models as a prelude to the discussion that follows. Section 3 motivates the PR perspective by relating the fixed and random effects terms, c_i and η_i, to a parameterization of $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, with a view to assess the traditional textbook latent omitted variables interpretation. It is shown that this interpretation is inappropriate for the fixed effects term c_i, but can be rendered appropriate for the random effects term η_i, when its proper interpretation is changed from the traditional that views η_i as capturing the effect of the omitted latent variables; it does not. A detailed discussion of the PR perspective is given in section 4, where it is shown that an way to interpret statistical models as parameterizations of the stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ is to explicitly derive the parameterization in question by imposing three types of probabilistic assumptions, distribution, dependence and heterogeneity, on its joint distribution. Several well-known panel data models are explicitly derived via the PR perspective with the view to provide a list of complete, internally consistent and testable probabilistic assumptions. It is shown how panel data provide an opportunity to relax the homogeneity assumptions traditionally imposed on statistical models for time series and cross-section data. A complete set of probabilistic assumptions is given for several statistical models of interest in modeling panel data. Section 5 extends the PR specification to dynamic models for panel data. Section 6 concludes the discussion by reflecting on how to further enhance the precision of inference when modeling with panel data.

2 Textbook perspective on Panel Data models

In the traditional econometric textbook perspective the emphasis is placed on the *least restrictive assumptions* that would justify (asymptotically) the 'quantification' technique for the particular model, irrespective of whether they are testable or not.

The two basic models for panel data are given in tables A and B.

Table A: Fixed Effects Panel (FEP) model

$$\begin{gathered} y_{it}=\mathbf{x}_{it}^{\top}\boldsymbol{\beta}+c_i+u_{it},\quad i\in\mathbb{N},\ t\in\mathbb{T},\\ E\left(u_{it}\right)=0,\ E\left(u_{it}^2\right)=\sigma_u^2,\ E\left(u_{it}u_{js}\right)=0,\text{ for } t\neq s,\ i\neq j.\\ c_i,\ i\in\mathbb{N},\text{ fixed in repeated samples.} \end{gathered} \tag{5}$$

Table B: Random Effects Panel (REP) model

$$\begin{gathered} y_{it}=\mathbf{x}_{it}^{\top}\boldsymbol{\alpha}+\eta_i+\epsilon_{it},\qquad i\in\mathbb{N},\ t\in\mathbb{T},\\ E\left(\epsilon_{it}\right)=0,\ E\left(\epsilon_{it}^2\right)=\sigma_\epsilon^2,\ E\left(\epsilon_{it}\epsilon_{js}\right)=0,\text{ for } t\neq s,\ i\neq j,\\ E\left(\eta_i\right)=0,\ E\left(\eta_i^2\right)=\sigma_\eta^2,\ E\left(\eta_i\eta_j\right)=0,\text{ for} i\neq j,\\ E\left(\epsilon_{it}\eta_j\right)=0,\text{ for all } i,j \text{ and } t. \end{gathered} \tag{6}$$

Apart from the apparent arbitrariness of the error probabilistic assumptions, the key issue that the specification of these two models raises is:

where do the terms c_i in (5) and η_i in (6) come from?

The textbook discussion on this issue is equivocal and several justifications are being discussed but none is fully satisfactory in the sense that it relates the terms (c_i,η_i) to the underlying distribution of the stochastic processes $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$; this is a question of *specification*. Related to this issue are the questions of assessing the validity of the probabilistic assumptions constituting these models (*misspecification testing*) and choosing different models when some of the assumptions are found to be invalid (*respecification*). In the panel data literature these issues are neglected, and the emphasis is placed on deriving CAN estimators.

2.1 Traditional textbook estimation

2.1.1 The Fixed Effects Panel data model

Using an obvious notation for all T observations the FEP data model, as specified in (5), takes the form:

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

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 $(T\times 1)$. One can express this model for all TN observations in matrix notation as:

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

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

The Ordinary Least Squares (OLS) estimators of $(\boldsymbol{\alpha},\mathbf{c})$ are:

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

where the projection matrices $(\mathbf{M}_D,\mathbf{M}_X)$ are defined by:

$$\mathbf{M}_D=\mathbf{I}-\mathbf{D}\left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{D},\qquad \mathbf{M}_X=\mathbf{I}-\mathbf{X}^*\left(\mathbf{X}^{*\top}\mathbf{X}^*\right)^{-\mathbf{1}}\mathbf{X}^{*\top}.$$

2.1.2 The Random Effects Panel data model

The the T observations of the REP data model as given in (6), with $u_{it}=\eta_i+\epsilon_{it}$, can be written in matrix notation as:

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

where the covariance structure is exchangeable, taking the form:

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

Expressing the REP model for all NT observations takes the form:

$$\underbrace{\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \vdots\\ \mathbf{y}_N\end{pmatrix}}_{\mathbf{y}^*}=\underbrace{\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\\ \vdots\\ \mathbf{X}_N\end{pmatrix}}_{\mathbf{X}^*}\boldsymbol{\alpha}+\underbrace{\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_N\end{pmatrix}}_{\mathbf{u}^*}$$

$$\Omega=\left[E(\mathbf{u}_i^\top\mathbf{u}_j)\right]_{i,j}^N=\begin{pmatrix}\boldsymbol{\Sigma} & 0 & \cdots & 0\\ 0 & \boldsymbol{\Sigma} & & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \boldsymbol{\Sigma}\end{pmatrix}=(\mathbf{I}_N\otimes\boldsymbol{\Sigma}):TN\times TN$$

Hence, the Generalized Least Squares (GLS) estimator of $\boldsymbol{\alpha}$ takes the form:

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

In practice, Ω is unknown and needs to be estimated; see Arellano (2003), Baltagi (2005), Hsiao (2002).

2.2 Where do the fixed and random effects come from?

The main weakness of the specification of these models is that the specification of the fixed and random effects $(c_i, \eta_i,\ i{\in}\mathbb{N})$ is not properly integrated into the specification of these panel data models in a way which allows one to assess the appropriateness of these assumptions vis-a-vis the data.

In an attempt to provide a more informative framework for the statistical analysis of panel data models, the Probabilistic Reduction (PR) perspective (Spanos, 1986, 1999, 2006a) is used to view statistical models for panel data as reductions from the joint distribution of the observable (vector) stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$. In the next two sections we consider first the *latent omitted variables* and then the *neglected heterogeneity* interpretations of the individual effects term in an attempt to shed light on the appropriateness of these textbook interpretations first articulated by Chamberlain (1982, 1984).

3 The latent variables perspective on specification

In this section we revisit the specification of the Fixed Effects and Random Effects models using the omitted latent variables argument in the context of the PR approach by relating them directly to the distribution of the vector stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$, defined on a probability space $(S, \mathfrak{F}, \mathbb{P}(.))$. The connection between the terms c_i and η_i and the underlying distribution of the observables can shed light on the merits and demerits of the latent variable interpretation.

3.1 Linear Regression for Panel data: pooled model

Consider the vector stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$, defined on $(S, \mathfrak{F}, \mathbb{P}(.))$, where we separate these observables into $\mathbf{Z}_{it} := \begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \end{pmatrix}$, assumed to be *Normal, Independent and Identically Distributed (NIID).* These probabilistic assumptions imply that the joint distribution of this process, say $D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi})$, can be reduced as follows:

$$\begin{aligned} D(\mathbf{Z}_{11}, \mathbf{Z}_{21}, ..., \mathbf{Z}_{NT}; \boldsymbol{\phi}) &\overset{\mathsf{I}}{=} \textstyle\prod_{i=1}^{N} \prod_{t=1}^{T} D(\mathbf{Z}_{it}; \boldsymbol{\varphi}(i,t)) = \\ &\overset{\mathsf{IID}}{=} \textstyle\prod_{i=1}^{N} \prod_{t=1}^{T} D(\mathbf{Z}_{it}; \boldsymbol{\varphi}) = \\ &\overset{\mathsf{IID}}{=} \textstyle\prod_{i=1}^{N} \prod_{t=1}^{T} D(\mathbf{y}_{it}|\mathbf{x}_{it}; \boldsymbol{\varphi}_1) \cdot D(\mathbf{X}_{it}; \boldsymbol{\varphi}_2), \end{aligned} \tag{7}$$

where Normality implies that $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi})$ is of the form:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{21}^{\top} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

One can show that the *regression* and *skedastic functions* take the form:

$$E(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}) = \mathbf{x}_{it}^{\top}\boldsymbol{\beta}, \qquad Var(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}) = \sigma_u^2, \tag{8}$$

where the model parameters are:

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}, \qquad \sigma_u^2{=}\sigma_{11}{-}\boldsymbol{\sigma}_{21}^{\top}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}. \tag{9}$$

This gives rise to a statistical model known as the **pooled panel data model**:

$$\begin{gathered} y_{it}{=}\ \mathbf{x}_{it}^{\top}\boldsymbol{\beta}{+}u_{it}, \qquad i{\in}\mathbb{N},\ t{\in}\mathbb{T}, \\ (u_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}) \backsim \mathsf{NIID}(0,\sigma_u^2). \end{gathered} \tag{10}$$

As argued below, this is the most restrictive panel data model which will be used as a benchmark for comparison.

3.2 Fixed Individual Effects Panel Data Model

Let us consider an extension of the above specification when the above regression model is *substantively misspecified* because a potentially relevant set of factors Ξ_i has been omitted; Ξ_i is a **latent (unobserved) variable**. In what follows it is shown that interpreting the fixed effects term c_i as capturing the effect of omitted latent variables is highly questionable.

3.2.1 Specification

In order to relate the latent vector Ξ_i to $(y_{it},\mathbf{X}_{it})$ we extend the joint distribution of the relevant stochastic process and assume to $D(y_{it},\mathbf{X}_{it},\Xi_i;\boldsymbol{\phi})$:

$$\begin{pmatrix} y_{it} \\ \mathbf{X}_{it} \\ \Xi_i \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_{13} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}\right), \quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}, \tag{11}$$

The regression function of y_{it} conditional on $(\mathbf{X}_{it}{=}\mathbf{x}_{it},\Xi_i{=}\ \boldsymbol{\xi}_i)$ is now:

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

where the model parameters take the form (Spanos, 2006c):

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

$$\begin{gathered} \boldsymbol{\delta} := \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31},\ \boldsymbol{\Delta} := \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23},\ \mathbf{D} := \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32}, \\ \boldsymbol{\Sigma}_{3.2}{:=}\ \boldsymbol{\Sigma}_{33}{-}\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_{23},\ \boldsymbol{\Sigma}_{2.3}{:=}\ \boldsymbol{\Sigma}_{22}{-}\mathbf{D}^{\top}\boldsymbol{\Sigma}_{32}; \end{gathered} \tag{13}$$

see Spanos (2006c) for the details. The above parameterization is directly estimable only when $\boldsymbol{\xi}_i$ is observed. When $\boldsymbol{\Xi}_i$ is unobservable $\boldsymbol{\xi}_i^\top\boldsymbol{\gamma}$ is also latent and can be written in the form (see Greene, 2003):

$$c_i = \boldsymbol{\xi}_i^\top\boldsymbol{\gamma}, \qquad i\in\mathbb{N},$$

yielding a regression function with latent individual effects:

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

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

$$\begin{gathered} y_{it} = \mathbf{x}_{it}^\top\boldsymbol{\alpha}{+}c_i{+}\varepsilon_{it}, \qquad i\in\mathbb{N},\ t\in\mathbb{T}, \\ (\varepsilon_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}, \boldsymbol{\Xi}_i{=}\ \boldsymbol{\xi}_i) \backsim \mathsf{NIID}(0, \sigma_\varepsilon^2). \end{gathered} \tag{14}$$

In view of the above model parameterizations, it's clear that $\boldsymbol{\beta}$ in (10) and $\boldsymbol{\alpha}$ in (14) coincide only when:

$$\boldsymbol{\sigma}_{31}{=}\ \mathbf{0} \text{ and } \boldsymbol{\Sigma}_{23}{=}\ \mathbf{0}, \tag{15}$$

i.e. $\boldsymbol{\Xi}_i$ is *uncorrelated* with both observable variables $(y_{it}, \mathbf{X}_{it})$. Under the restrictions in (15):

$$\boldsymbol{\gamma} = \mathbf{0}, \quad \boldsymbol{\alpha} = \boldsymbol{\beta}, \quad \sigma_\varepsilon^2 = \sigma_u^2.$$

In general, when the restrictions in (15) do *not* hold, one needs to estimate the fixed effects model (14) in some other way because $\boldsymbol{\alpha} \neq \boldsymbol{\beta}, \quad \sigma_\varepsilon^2 \neq \sigma_u^2.$

3.2.2 Consistent estimator of what?

There are two basic problems with the above latent variable specification argument.

The *first* is that conditioning on the *observed value* of an unobservable variable $(\boldsymbol{\Xi}_i{=}\ \boldsymbol{\xi}_i)$ is conceptually problematic. If a variable is unobservable, conditioning on *its observed value* makes no probabilistic sense. In this sense, the latent variable interpretation for the fixed effects c_i formulation is questionable on probabilistic grounds.

The *second*, and more practical problem, concerns the question 'what is $\widehat{\boldsymbol{\alpha}}$, where:

$$\widehat{\boldsymbol{\alpha}}{=}\left(\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{y}^*$$

a consistent estimator of?' It is obvious that it's *not* a consistent estimator of $\boldsymbol{\alpha}$ with the implicit parameterization:

$$\boldsymbol{\alpha} = \left(\boldsymbol{\Sigma}_{22}{-}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32}\right)^{-1}(\boldsymbol{\sigma}_{21}{-}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\sigma}_{31}). \tag{16}$$

The *third* question is: in what sense can c_i be viewed as providing a proxy for the omitted variables $\boldsymbol{\Xi}_i$? These are questions that need to be addressed.

3.3 Random Individual Effects Panel Data Model

In light of the fact that conditioning on the observed value of a latent variable is problematic on probabilistic grounds, the issue that arises is whether one can rectify that by conditioning on the σ-field generated by the latent variable Ξ_i, say $\sigma(\Xi_i)$, i.e. the conditioning information set is now $\mathcal{D}_{it}=(\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))$.

Conditioning on the σ-field generated by Ξ_i makes more sense than conditioning on its observed value because the σ-field simply acknowledges the events associated with Ξ_i, by restricting the original σ-field $\mathfrak{F}$. This is a restriction because $\sigma(\Xi_i)\subset\mathfrak{F}$ and $E(y_{it}|\mathfrak{F})=y_{it}$, but in contrast,

$$E(y_{it}|\sigma(\Xi_i))=g(\Xi_i)\neq y_{it}.$$

This also explains why conditioning on $\{\Xi_i=\boldsymbol{\xi}_i\}$ is *problematic.* For any random variable W defined on the same probability space $(S,\mathfrak{F},\mathbb{P}(.))$, 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 (Renyi, 1970):

$$h(\boldsymbol{\xi}_i)\neq h(\boldsymbol{\xi}_j)\text{ when }\boldsymbol{\xi}_i\neq\boldsymbol{\xi}_j,\text{ for all }i,j\in\mathbb{N},$$

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

The conditioning on $(\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))$ yields the stochastic linear regression function:

$$E(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))=\mathbf{x}_{it}^{\top}\boldsymbol{\alpha}+\Xi_i^{\top}\boldsymbol{\gamma},\tag{17}$$

where Ξ_i denotes the random vector itself, rendering the term $\Xi_i^{\top}\boldsymbol{\gamma}$ stochastic; see Spanos (1986), p. 413. This can be written in the form:

$$E(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))=\mathbf{x}_{it}^{\top}\boldsymbol{\alpha}+\eta_i,$$

where:

$$\eta_i=\Xi_i^{\top}\boldsymbol{\gamma}\backsim\mathsf{N}(0,\sigma_{\eta}^2),\qquad\sigma_{\eta}^2=\boldsymbol{\gamma}^{\top}\boldsymbol{\Sigma}_{33}\boldsymbol{\gamma}.$$

This gives rise to the **statistical model**:

$$\begin{gathered}y_{it}=\mathbf{x}_{it}^{\top}\boldsymbol{\alpha}+\eta_i+\epsilon_{it},\qquad i\in\mathbb{N},\ t\in\mathbb{T},\\(\epsilon_{it}|\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))\backsim\mathsf{NIID}(0,\sigma_{\epsilon}^2),\ \eta_i=\left(\Xi_i^{\top}\boldsymbol{\gamma}\right)\backsim\mathsf{NIID}(0,\sigma_{\eta}^2),\end{gathered}\tag{18}$$

which can potentially be viewed as providing a particular interpretation for the REP model.

As in the case of the fixed effects model, the important question is whether the estimator:

$$\widetilde{\boldsymbol{\alpha}}=\left(\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^{*},\tag{19}$$

is consistent for the unknown parameter $\boldsymbol{\alpha}$ in (16). The answer is clearly not! What is $\widetilde{\boldsymbol{\alpha}}$ a consistent estimator of? In what sense $\widetilde{\boldsymbol{\alpha}}$ 'captures' the potential effect of the omitted variables Ξ_i? These are questions that need to be addressed.

3.3.1 Shedding light on the consistency question

The guiding principle in investigating the consistency of an estimator is that, although substantive models can be specified in terms of latent variables, statistical models need to be specified exclusively in terms of the observable (vector) stochastic process involved. That is, any latent variables in the substantive model should be, somehow, eliminated when the statistical model is specified. The elimination can take a number of different forms, but the most obvious is to substitute the latent variables out by using functionals of the observable variables.

With that in mind, let us return to the joint distribution (11) and consider the case when $\boldsymbol{\Sigma}_{23}\neq\mathbf{0}$; note that Ξ_i is irrelevant in the case where $\boldsymbol{\sigma}_{31}=0$ and $\boldsymbol{\Sigma}_{23}=\mathbf{0}$. In such a case there exists a relationship between Ξ_i and $\mathbf{X}_{it}$ of the form:

$$\begin{aligned}&\Xi_i=\boldsymbol{\delta}_0+\boldsymbol{\Delta}^{\top}\mathbf{x}_{it}+\mathbf{v}_i,\quad (\mathbf{v}_i|\mathbf{x}_{it})\backsim \mathsf{NIID}(\mathbf{0},\boldsymbol{\Sigma}_{3.2}),\\ &\boldsymbol{\delta}_0=\boldsymbol{\mu}_3-\boldsymbol{\Delta}^{\top}\boldsymbol{\mu}_2,\ \ \boldsymbol{\Delta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{23},\ \ \boldsymbol{\Sigma}_{3.2}=\boldsymbol{\Sigma}_{33}-\boldsymbol{\Delta}^{\top}\boldsymbol{\Sigma}_{23},\end{aligned}\tag{20}$$

where $E(\Xi_i|\mathbf{X}_{it}=\mathbf{x}_{it})=\boldsymbol{\delta}_{\mathbf{0}}+\mathbf{x}_{it}^{\top}\boldsymbol{\Delta}$, and $\mathbf{v}_i$ is orthogonal to $\mathbf{X}_{it}$, denoted by $\mathbf{X}_{it}\perp\mathbf{v}_i$. An obvious way to eliminate Ξ_i is to substitute (20) into (18) yielding:

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

In view of the parameterizations in (12), one can deduce that the coefficient of $\mathbf{x}_{it}$ is no longer $\boldsymbol{\alpha}$ but:

$$\boldsymbol{\beta}=\boldsymbol{\alpha}+\boldsymbol{\Delta}\boldsymbol{\gamma}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}.$$

This give rise to a coherent interpretation of the Random Effects Panel (REP) data model, based on:

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

where (i) the implicit parameterization of the coefficient of $\mathbf{x}_{it}$ is $\boldsymbol{\beta}$, not $\boldsymbol{\alpha}$, and (ii) the omitted effect is $\left(\mathbf{v}_i^{\top}\boldsymbol{\gamma}\right)$ not $\Xi_i^{\top}\boldsymbol{\gamma}$. The primary difference between (18) and (21) is that in the latter case $\mathbf{X}_{it}\perp(\mathbf{v}_i,\epsilon_{it})$ by construction, eliminating the original (potential) correlation between $\mathbf{X}_{it}$ and $\Xi_i^{\top}\boldsymbol{\gamma}$. If pertinent, the interpretation in (21) will render the original interpretation in (18) inappropriate.

The question that naturally arises at this stage is the extent to which (21), interpreted as:

$$\begin{aligned}&y_{it}=\mathbf{x}_{it}^{\top}\boldsymbol{\beta}+v_i+\epsilon_{it},\quad i\in\mathbb{N},\ t\in\mathbb{T},\\ &(\epsilon_{it}|\mathbf{X}_{it}=\mathbf{x}_{it},\sigma(\Xi_i))\backsim\mathsf{NIID}(0,\sigma_{\epsilon}^2),\ v_i=\left(\mathbf{v}_i^{\top}\boldsymbol{\gamma}\right)\backsim\mathsf{NIID}(0,\sigma_v^2),\end{aligned}\tag{22}$$

provides a more pertinent specification for the Random Effects Panel (REP) data model. What renders the interpretation in (22) superior to that in (18) is that the former provides a clear answer to the consistency question in so far as $\widetilde{\boldsymbol{\alpha}}$ in (19) is not a consistent estimator of $\boldsymbol{\alpha}$ in (16), but a consistent estimator of $\boldsymbol{\beta}$ in (9), i.e. $\widetilde{\boldsymbol{\alpha}}=\left(\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^{*}$ provides a GLS-type estimator for $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$.

3.3.2 Revisiting the Mundlak interpretation

Motivated by the fact that in general one would expect $E(\eta_i|\mathbf{X}_{it}=\mathbf{x}_{it})\neq 0$, for $i\in\mathbb{N}$, 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 (2010). To capture the correlation $Corr(\eta_i,\mathbf{X}_{it})\neq 0$, Mundlak proposed the *auxiliary regression*:

$$\eta_i=\textstyle\sum_{t=1}^{T}\boldsymbol{\delta}_1^\top\mathbf{x}_{it}+\omega_i=\boldsymbol{\delta}_1^\top\overline{\mathbf{x}}_i+\omega_i,\quad \omega_i\backsim\mathsf{NIID}(0,\sigma_\omega^2), \tag{23}$$

where $\overline{\mathbf{x}}_i=\frac{1}{n}\sum_{t=1}^{T}\mathbf{x}_{it},\ i=1,2,...,n,$ and:

$$E(\eta_i|\mathbf{X}_{it}=\mathbf{x}_{it})=\textstyle\sum_{t=1}^{T}\boldsymbol{\delta}_1^\top\mathbf{x}_{it}=\boldsymbol{\delta}_1^\top\overline{\mathbf{x}}_i,\ i=1,2,...,n.$$

When (23) is substituted back into the original formulation, it gives rise to the **Mundlak formulation**:

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

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 (Hsiao, 2003):

(i) the GLS estimator $\widetilde{\boldsymbol{\alpha}}$ a consistent estimator of $\boldsymbol{\alpha}$ in (16).

(ii) the GLS estimator $\widetilde{\boldsymbol{\delta}}_1$ is not a consistent estimator of $\boldsymbol{\gamma}$ in (12).

(iii) what are $(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\delta}}_1)$ consistent estimators of?

(iv) what is Mundlak's formulation a solution of?

The answers to (iii)-(iv) are rather surprising!

Taking a closer look at the underlying parameterization of $v_i=(\mathbf{v}_i^\top\boldsymbol{\gamma})$ in (22) reveals that**:**

$$\sigma_\nu^2=\boldsymbol{\gamma}^\top\boldsymbol{\Sigma}_{3.2}\boldsymbol{\gamma}=(\boldsymbol{\sigma}_{31}-\boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21})^\top\boldsymbol{\Sigma}_{3.2}^{-1}(\boldsymbol{\sigma}_{31}-\boldsymbol{\Sigma}_{32}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21})\neq\sigma_\eta^2=\boldsymbol{\gamma}^\top\boldsymbol{\Sigma}_{33}\boldsymbol{\gamma}.$$

Hence, the end result is that one can estimate σ_ϵ^2 as well as σ_ν^2 by orthogonally decomposing σ_u^2 :

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

That is, the latent random effects term v_i does *not* represent a linear combination of the omitted latent variables ($v_i\neq\eta_i=\boldsymbol{\Xi}_i^\top\boldsymbol{\gamma}$), but a linear combination of the omitted *error effects* from the auxiliary regression of $\boldsymbol{\Xi}_i$ on $\mathbf{X}_{it}$, i.e.

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

Moreover, the above interpretation renders the parameterization $(\beta_0,\boldsymbol{\beta},\sigma_\nu^2,\sigma_\epsilon^2)$ estimable**,** as opposed to the parameterization $(\alpha_0,\boldsymbol{\alpha},\sigma_\epsilon^2,\sigma_\eta^2)$ which is *not* estimable unless $\boldsymbol{\Xi}_i$ is observable!

The general principle is that statistical models are given proper probabilistic interpretation when the observed systematic component is orthogonal to the unobserved non-systematic component. Indeed, its orthogonality with the non-systematic component ensures the consistent estimation of its parameterization. In the above case, due to the fact that Ξ_i is a latent vector, the only appropriate interpretation is one which ensures that $(\mathbf{x}_{it}^{\top}\boldsymbol{\beta}) \perp (v_i{+}\epsilon_{it})$. This principle calls into question the practice of specifying statistical models in terms of unobservables and then choose the interpretation(s) one would like to attribute to the unknown parameters, ignoring the structure of the underlying observable processes involved. An interpretation makes statistical sense only to the extent that it can be related to the probabilistic structure of the observable process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$.

3.3.3 An unambiguous interpretation of the REP data model

Viewed from the Probabilistic Reduction (PR) perspective, the complete set of probabilistic assumptions specifying the model in (22) is given in table 1.

Table 1 - Random Effects Panel (REP) data model

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

In summary, viewing the REP model as specified table 1, provides:

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

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

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

Let us look at the details more closely.

(i) The statistical parameterization $\boldsymbol{\theta}{:=}\left(\beta_0,\boldsymbol{\beta},\sigma_{\nu}^2,\sigma_{\epsilon}^2\right):$

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

is the relevant and estimable one.

(ii) $\widetilde{\boldsymbol{\beta}}{=}(\mathbf{X}^{*\top}\Omega^{-1}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\top}\Omega^{-1}\mathbf{y}^{*}$ is *consistent* for $\boldsymbol{\beta}{=}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$ (not $\boldsymbol{\alpha}$ in (16)),

(iii) $\widetilde{\boldsymbol{\beta}}$ is also *efficient* because it takes account of the Ω covariance structure.

(iv) What about the estimators $(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\delta}}_1)$ associated with the Mundlak formulation?

$\widetilde{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\beta}$, and $\widetilde{\boldsymbol{\delta}}_1=\widehat{\boldsymbol{\beta}}_b-\widehat{\boldsymbol{\beta}}_w$, where $\widehat{\boldsymbol{\beta}}_b$ and $\widehat{\boldsymbol{\beta}}_w$ are the between and within estimators of $\boldsymbol{\beta}$; see Hsiao (2003). Indeed, since $p\lim(\widetilde{\boldsymbol{\delta}}_1)=\boldsymbol{\delta}_1\neq\boldsymbol{\gamma}$, $\widetilde{\boldsymbol{\delta}}_1$ estimates no parameter of interest!

(v) The interpretation of the random effects (REP) model given in (22) addresses another awkward problem associated with the interpretation in (18); a problem often dismissed as innocuous by the current literature; see Wooldridge (2010). This is the problem of assuming that $E(\Xi_i)=\mathbf{0}$. If, instead, $E(\Xi_i)=\boldsymbol{\mu}_3$, it follows that:

$$\eta_i=\left(\Xi_i^\top\boldsymbol{\gamma}\right)\backsim\mathsf{NIID}(d,\sigma_\eta^2),\quad d=\boldsymbol{\mu}_3^\top\boldsymbol{\gamma}.$$

A closer look at $E(\Xi_i)=\mathbf{0}$ indicates that this is a non-operational restriction. How can one ensure that the latent variable has mean zero? In the context of (22) not such restriction is needed to ensure the coherence of the specification.

4 The Probabilistic Reduction (PR) perspective

In this section we propose a systematic account of the specification of the various statistical models for panel data from the Probabilistic Reduction perspective. This account enables one to relate the fixed effects directly to the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ defined on the probability space $(S,\mathfrak{F},\mathbb{P}(.))$, and opens the way for generalizing the basic models in a general but coherent way that avoids the various layers of ad hoc assumptions about unobservable errors and latent variables.

In sub-section 3 below it is shown that the 'neglected heterogeneity' interpretation (Wooldridge, 2010), when properly framed in the context of the PR approach, provides an appropriate interpretation of the fixed effects term c_i, $i\in\mathbb{N}$.

4.1 A General Panel Data Model

Consider the vector stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$, $\mathbf{Z}_{it}:=(y_{it},\mathbf{X}_{it}^\top)^\top$ assumed to be *Normal, Independent (NI) but* ***non-Identically Distributed.*** In particular, we allow this stochastic process to be heterogeneous with respect to both (i,t). These probabilistic assumptions imply that the joint distribution of this process $D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi})$, can be reduced as follows:

$$\begin{aligned}D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{I}}{=}\textstyle\prod_{i=1}^N\prod_{t=1}^T D(\mathbf{Z}_{it};\boldsymbol{\varphi}(i,t))=\\ &=\prod_{i=1}^N\prod_{t=1}^T D(\mathbf{y}_{it}|\mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))\cdot D(\mathbf{X}_{it};\boldsymbol{\varphi}_2(i,t)),\end{aligned}\tag{24}$$

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:

$$E(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it})=\beta_0(i,t)+\mathbf{x}_{it}^\top\boldsymbol{\beta}(i,t),\qquad Var(y_{it}|\mathbf{X}_{it}=\mathbf{x}_{it})=\sigma_u^2(i,t),\tag{25}$$

where the model parameters are:

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

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

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

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 $m{=}(K{+}2)NT{=}(6)(100)(25){=}15000$!

4.2 Pooled Panel Data Model

Consider the most extreme case where the $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ process is assumed to be *Normal, Independent and* ***Identically Distributed.*** These probabilistic assumptions give rise to the reduction in (7) where $D(y_{it},\mathbf{X}_{it};\boldsymbol{\varphi})$ is:

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

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

$$E(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}) = \beta_0{+}\mathbf{x}_{it}^\top\boldsymbol{\beta},\qquad Var(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}) = \sigma_u^2, \tag{26}$$

where the model parameters are:

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

This is clearly an operational model because for, say $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the number of unknown model parameters are $m{=}k{+}2{=}(4{+}2){=}6$.

This gives rise to the *operational* statistical model:

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

The complete specification of this model in terms of the observables is given in table 2.

Table 2 - Pooled Panel Data 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:	$(y_{it}\|\mathbf{X}_{it}{=}\mathbf{x}_{it}) \backsim \mathsf{N}(.,.)$,
[2]	Linearity:	$E\left(y_{it}\|\mathbf{X}_{it}{=}\mathbf{x}_{it}\right){=}\beta_0{+}\mathbf{x}_{it}^\top\boldsymbol{\beta}$,
[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]	(i,t)-invariance:	$(\beta_0,\boldsymbol{\beta},\sigma_u^2)$ are (i,t)-invariant.

4.3 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* with respect to $i{\in}\mathbb{N}$, but *completely homogeneous* with respect to $t{\in}\mathbb{T}$. These probabilistic assumptions imply that the joint distribution of this process can be reduced as follows:

$$\begin{aligned} D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\text{I}}{=} \textstyle\prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{Z}_{it};\boldsymbol{\varphi}(i)){=} \\ &= \prod_{i=1}^{N}\prod_{t=1}^{T} D(\mathbf{y}_{it}|\mathbf{x}_{it};\boldsymbol{\varphi}_1(i)){\cdot}D(\mathbf{X}_{it};\boldsymbol{\varphi}_2(i)), \end{aligned} \tag{27}$$

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:

$$E(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}){=}\beta_0(i){+}\mathbf{x}_{it}^{\top}\boldsymbol{\beta},\qquad Var(y_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it}){=}\sigma_u^2, \tag{28}$$

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 *operational* **statistical model**:

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

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

Table 3 - Fixed Individual Effects Panel (FEP) 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:	$(y_{it}\vert\mathbf{X}_{it}{=}\mathbf{x}_{it}) \backsim \mathsf{N}(.,.),$
[2]	Linearity:	$E\left(y_{it}\vert\mathbf{X}_{it}{=}\mathbf{x}_{it}\right){=}\beta_0(i){+}\mathbf{x}_{it}^{\top}\boldsymbol{\beta},$
[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]	(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 operational 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 that the OLS estimator of $(\boldsymbol{\beta}, \mathbf{c}, \sigma_u^2)$:

$$\widehat{\boldsymbol{\beta}}{=}\left(\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\top}\mathbf{M}_D\mathbf{y}^*,\quad \widehat{\mathbf{c}}{=}\left(\mathbf{D}^\top\mathbf{M}_X\mathbf{D}\right)^{-1}\mathbf{D}^\top\mathbf{M}_X\mathbf{y}^*,\quad s^2{=}\tfrac{\widehat{\mathbf{u}}^{*\top}\widehat{\mathbf{u}}^*}{(NT-k-N)}$$

(see notation in section 1), are *consistent estimators* of the underlying parameterizations:

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

This was *not* the case when the omitted latent variables perspective was used to specify this model.

The above PR perspective provides a strong support for the 'neglected heterogeneity' interpretation of the fixed effects term $c_i,\ i{\in}\mathbb{N}$. In particular, what this term captures is a linear combination of the mean heterogeneity in the observable process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$.

4.3.1 Fixed vs. Random effects? Is that the question?

The textbook perspective on the two basic panel data models, FEP and REP, views them as alternative models for accounting for some vague form of heterogeneity. The above discussion has shed light on the form of heterogeneity the individual effects terms could account for.

The random effects term v_i of the REP model (table 1) represents a linear combination of the *errors* from the auxiliary regression of the latent omitted variables $\boldsymbol{\Xi}_i$ on the included observable variables $\mathbf{X}_{it}$, i.e. $v_i=\boldsymbol{\gamma}^\top(\boldsymbol{\Xi}_i{-}\boldsymbol{\delta}_0{-}\boldsymbol{\Delta}^\top\mathbf{x}_{it})$. The fixed effects term c_i of the FEP model (table 3) represents a linear combination of the mean-heterogeneity of the observables, i.e. $c_i{=}\mu_1(i){-}\boldsymbol{\mu}_2^\top(i)\boldsymbol{\beta}$. This makes it clear that the two individual effects terms are totally different, rendering the traditional textbook hypothesis testing for choosing between the two formulations, using Hausman-type tests (Greene, 2008), irrelevant at best, and potentially highly misleading, at worst; see Spanos (2006a).

4.4 Fixed Individual and Time Effects Panel 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* with respect to both $i{\in}\mathbb{N}$ and $t{\in}\mathbb{T}$. These probabilistic assumptions imply a reduction as in (24) where $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi}(i,t))$ takes the form:

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

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

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

where the model parameters are:

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

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

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

This is because, for say $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the unknown model parameters are:

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

whose total number is $m{=}NT{+}k{+}2{=}((100)(25){+}4{+}2){=}2506$.

This model can be rendered operational if one is prepared to assume that the (i,t) *−heterogeneity* **is separable** in the sense that:

$$\beta_0(i,t){=}\gamma_0(i){+}\delta_0(t).$$

This assumption implies that the unknown model parameters now are:

$$\gamma_0(i),\ \delta(t)\quad \boldsymbol{\beta},\quad \sigma_u^2,\quad i{=}1,..,N,\ t{=}1,..,T,$$

whose total number is $m{=}N{+}T{+}k{+}2{=}(100{+}25{+}4{+}2){=}131$.

This gives rise to a (potentially) *operational* **statistical model**:

$$y_{it}{=}\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^\top\boldsymbol{\beta}{+}u_{it},\ (u_{it}|\mathbf{X}_{it}{=}\mathbf{x}_{it})\backsim\mathsf{NIID}(0,\sigma_u^2),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The complete specification of this model in terms of the observable stochastic processes is given in table 4.

Table 4 - Fixed Individual and Time Effects Panel Data Model

Statistical GM: $y_{it}=\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^\top\boldsymbol{\beta}{+}u_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$

[1]	Normality:	$(y_{it}\vert\mathbf{X}_{it}{=}\mathbf{x}_{it})\backsim\mathsf{N}(.,.)$,
[2]	Linearity:	$E\left(y_{it}\vert\mathbf{X}_{it}{=}\mathbf{x}_{it}\right)=\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^\top\boldsymbol{\beta}$
[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 process,
[5]	(a) (i,t)-invar.:	$(\boldsymbol{\beta},\sigma_u^2)$ - (i,t)-invariant,
	(b) t-invariance:	$\{\gamma_0(i),\ i{\in}\mathbb{N}\}$ - t-invariant but i-heterogeneous
	(c) i-invariance:	$\{\delta_0(t),\ t{\in}\mathbb{T}\}$ - i-invariant but t-heterogeneous.

4.5 Fixed Individual and Slope Effects Panel 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), Homogeneous with respect to* $t{\in}\mathbb{T}$, *but heterogeneous* with respect to $i\in\mathbb{N}$. The reduction implied by these assumption is of the form (27) where $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi}(i))$ is:

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

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

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

where the model parameters are:

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

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

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

For $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the unknown model parameters are:

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

which amounts to $m{=}N{+}kN{+}N{=}(100{+}400{+}100){=}600$; rather problematic! By the same token, in the case where $k{=}4$, $N{=}25$, $T{=}100$, $m{=}N{+}kN{+}N{=}150$; which is potentially operational.

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

Table 5 - Fixed Individual and Slope Effects Panel Model

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

4.6 Time Effects Panel (TEP) 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), completely homogeneous* with respect to $i\in\mathbb{N}$, *but heterogeneous* with respect to $t\in\mathbb{T}$. These probabilistic assumptions imply that the joint distribution of this process can be reduced as follows:

$$D(\mathbf{Z}_{11},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) \overset{\text{I}}{=} \textstyle\prod_{i=1}^N\prod_{t=1}^T D(\mathbf{Z}_{it};\boldsymbol{\varphi}(i,t)){=}$$

$$= \textstyle\prod_{i=1}^N\prod_{t=1}^T D(\mathbf{y}_{it}|\mathbf{x}_{it};\boldsymbol{\varphi}_1(t)){\cdot}D(\mathbf{X}_{it};\boldsymbol{\varphi}_2(t)),$$

where $D(y_{it}, \mathbf{X}_{it}; \boldsymbol{\varphi}(t))$ takes the form:

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

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

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

where the model parameters are:

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

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

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

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

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

whose total number is $m{=}(T{+}Tk{+}T){=}(25{+}100{+}25){=}150$.

The complete specification of this model in terms of the observable stochastic processes is given in table 6.

Table 6 - Fixed Time Effects Panel Data Model

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

5 Dynamic Panel Data Models

In this section we consider extending the static models specified in tables 1-5 to dynamic specifications which allow for temporal dependence. To simplify the exposition in what follows the reduction assumption of *independence* used for the reduction that underlies the static models will be replaced with *Markov dependence*. The latter can be easily extended to Markov(ℓ) dependence or even asymptotic weak dependence (see Spanos, 1999, ch. 8) but that would only complicate the notation unnecessarily.

Consider the vector stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$, $\mathbf{Z}_{it}{:=}(y_{it},\mathbf{X}_{it}^{\top})^{\top}$ assumed to be *Normal, Markov (NM) but **non-stationary**.* In particular, we allow this stochastic process to be (i,t)-heterogeneous. These probabilistic assumptions imply that the joint distribution of this process $D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi})$, can be reduced as follows:

$$\begin{aligned} D(\mathbf{Z}_{11},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{I}}{=}\textstyle\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t)){=}\\ &=\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{y}_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))D(\mathbf{X}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}_2(i,t)) \end{aligned} \tag{29}$$

where $D(\mathbf{Z}_{it}|\ \mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t))$ takes the form:

$$\begin{pmatrix}\mathbf{Z}_{it}\\ \mathbf{Z}_{it-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\boldsymbol{\mu}(i,t)\\ \boldsymbol{\mu}(i,t{-}1)\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}(i,t) & \boldsymbol{\Sigma}(i,t{-}1)\\ \boldsymbol{\Sigma}(i,t{-}1)^{\top} & \boldsymbol{\Sigma}(i,t{-}1)\end{pmatrix}\right),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The regression and skedastic functions associated with $D(\mathbf{y}_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))$ are:

$$E(y_{it}|\mathfrak{D}_{it}){=}\alpha_0(i,t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i,t){+}\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i,t){+}\gamma_2(i,t)y_{it-1},\quad Var(y_{it}|\mathfrak{D}_{it}){=}\sigma_v^2(i,t),$$

where the conditioning information set is: $\mathfrak{D}_{it}{=}\{\mathbf{X}_{it}{=}\mathbf{x}_{it},\mathbf{X}_{it-1}{=}\ \mathbf{x}_{it-1},\sigma(y_{it-1})\}$. This gives rise to the *non-operational* dynamic statistical model:

$$\left.\begin{array}{c} y_{it}{=}\alpha_0(i,t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i,t){+}\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i,t){+}\gamma_2(i,t)y_{it-1}{+}v_{it},\\ \\ (v_{it}|\mathfrak{D}_{it})\backsim\mathsf{NIID}(0,\sigma_v^2(i,t)), \end{array}\right\}i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

For the derivations see Spanos (1986), pp. 523-4.

For $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the number of unknown parameters is:
$m{=}NT{+}k(NT){+}(k{+}1)(NT){+}NT{=}NT\,(2k{+}3)\,{=}2500(11){=}27500$.

Can this model be rendered operational? The short answer is yes, under certain restrictions on the non-stationarity (heterogeneity) of the process.

5.1 Dynamic Pooled Panel Data Model

Let us begin with the most restrictive case where the vector stochastic process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ is Normal, Markov and $(i,t){-}$stationary. These probabilistic assumptions imply that:

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

where $D(\mathbf{Z}_{it},\mathbf{Z}_{it-1};\boldsymbol{\phi})$ is:

$$\begin{pmatrix}\mathbf{Z}_{it}\\ \mathbf{Z}_{it-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1)\\ \boldsymbol{\Sigma}(1)^{\top} & \boldsymbol{\Sigma}(0)\end{pmatrix}\right),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The regression and skedastic functions associated with $D(\mathbf{y}_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1)$ are:

$$E(y_{it}|\mathfrak{D}_{it})=\alpha_0+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+y_{it-1}\gamma_2, \quad Var(y_{it}|\mathfrak{D}_{it})=\sigma_v^2,$$

where the conditioning information set is: $\mathfrak{D}_{it}=\{\mathbf{X}_{it}=\mathbf{x}_{it}, \mathbf{X}_{it-1}=\mathbf{x}_{it-1}, \sigma(y_{it-1})\}$. This gives rise to the *operational* dynamic statistical model:

$$y_{it}=\alpha_0+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+\gamma_2 y_{it-1}+v_{it}, \ (v_{it}|\mathfrak{D}_{it})\backsim\mathsf{NIID}(0,\sigma_v^2), \quad i\in\mathbb{N}, \ t\in\mathbb{T}.$$

The complete specification of this model in terms of the observable stochastic processes is given in table 7.

Table 7 - Dynamic Pooled Panel Data Model

Statistical GM: $y_{it}=\alpha_0+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+\gamma_2 y_{it-1}+v_{it}, \ i\in\mathbb{N}, \ t\in\mathbb{T}.$

[1]	Normality:	$(y_{it} \mid \mathfrak{D}_{it}) \backsim \mathsf{N}(.,.)$,
[2]	Linearity:	$E\left(y_{it}|\mathfrak{D}_{it}\right)=\alpha_0+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+\gamma_2 y_{it-1}$,
[3]	Homo/sticity:	$Var\left(y_t \mid \mathfrak{D}_{it}\right)=\sigma_v^2$,
[4]	Markov :	$\{(y_{it} \mid \mathfrak{D}_{it}), \ i\in\mathbb{N}, \ t\in\mathbb{T}\}$ is a Markov process,
[5]	(i,t)-invariance:	$(\alpha_0,\boldsymbol{\gamma},\boldsymbol{\gamma}_1,\gamma_2,\sigma_v^2)$ are (i,t)-invariant.

This is a particularly parsimonious statistical model because for k=4, N=100, T=25, the number of unknown parameters is: m=1+k+(k+1)+1=2k+3=11.

5.2 Dynamic Fixed Individual Effects Panel Data Model

Consider the case where $\mathbf{Z}_{it}:=(y_{it},\mathbf{X}_{it}^{\top})^{\top}$ is Normal, Markov and t-stationary, but i mean non-stationary process. These probabilistic assumptions imply that:

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

where $D(\mathbf{Z}_{it},\mathbf{Z}_{it-1};\boldsymbol{\phi}(i))$ takes the form:

$$\begin{pmatrix}\mathbf{Z}_{it}\\ \mathbf{Z}_{it-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\boldsymbol{\mu}(i)\\ \boldsymbol{\mu}(i)\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1)\\ \boldsymbol{\Sigma}(1)^{\top} & \boldsymbol{\Sigma}(0)\end{pmatrix}\right), \quad i\in\mathbb{N}, \ t\in\mathbb{T}.$$

The regression and skedastic functions associated with $D(\mathbf{y}_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(i))$ are:

$$E(y_{it}|\mathfrak{D}_{it})=\alpha_0(i)+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+\gamma_2 y_{it-1}, \quad Var(y_{it}|\mathfrak{D}_{it})=\sigma_v^2,$$

where the conditioning information set is: $\mathfrak{D}_{it}=\{\mathbf{X}_{it}=\mathbf{x}_{it}, \mathbf{X}_{it-1}=\mathbf{x}_{it-1}, \sigma(y_{it-1})\}$. This gives rise to the *operational* dynamic statistical model:

$$y_{it}=\alpha_0(i)+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1+\gamma_2 y_{it-1}+v_{it}, \ (v_{it}|\mathfrak{D}_{it})\backsim\mathsf{NIID}(0,\sigma_v^2), \quad i\in\mathbb{N}, \ t\in\mathbb{T},$$

For $k{=}4$, $N{=}100$, $T{=}25$, the number of unknown parameters is $n{=}N{+}k{+}(k{+}1){+}1{=}110$.

The complete specification of this model in terms of the observable stochastic processes is given in table 8.

Table 8 - Dynamic Fixed Individual Effects Panel Data Model

Statistical GM: $y_{it}{=}\alpha_0(i){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$

[1]	Normality:	$(y_{it} \mid \mathfrak{D}_{it}) \backsim \mathsf{N}(.,.)$,
[2]	Linearity:	$E\,(y_{it}\vert\mathfrak{D}_{it})={\alpha_0(i)}{+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}$,
[3]	Homosk/sticity:	$Var\,(y_t \mid \mathfrak{D}_{it})=\sigma_v^2$,
[4]	Markov:	$\{(y_{it} \mid \mathfrak{D}_{it})\,,\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ is a Markov process,
[5]	(a) (i,t)-invar.:	$(\boldsymbol{\gamma},\boldsymbol{\gamma}_1,\gamma_2,\sigma_v^2)$ are (i,t)-invariant,
	(b) t-invariance:	$\{\alpha_0(i),i{\in}\mathbb{N}\}$ are t-invariant, but i-heterogeneous.

5.3 Dynamic Fixed Individual and Time Effects Panel Data Model

Consider the case where $\mathbf{Z}_{it}{:=}(y_{it},\mathbf{X}_{it}^{\top})^{\top}$ is Normal, Markov and mean (i,t) non-stationary process. These probabilistic assumptions imply that:

$$\begin{aligned} D(\mathbf{Z}_{11},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{M}}{=}\textstyle\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t))= \\ &\overset{\mathsf{M\&S}}{=}\textstyle\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t))= \\ &=\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{y}_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))D(\mathbf{X}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}_2(i,t)) \end{aligned}$$

where $D(\mathbf{Z}_{it},\mathbf{Z}_{it-1};\boldsymbol{\phi}(i,t))$ takes the form:

$$\begin{pmatrix} \mathbf{Z}_{it} \\ \mathbf{Z}_{it-1} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \boldsymbol{\mu}(i,t) \\ \boldsymbol{\mu}(i,t) \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1) \\ \boldsymbol{\Sigma}(1)^{\top} & \boldsymbol{\Sigma}(0) \end{pmatrix}\right),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The regression and and skedastic functions associated with $D(\mathbf{y}_{it}|\ \mathbf{Z}_{it-1}{,}\mathbf{x}_{it};\boldsymbol{\varphi}_1(i,t))$ are:

$$E(y_{it}|\mathfrak{D}_{it}){=}\alpha_0(i,t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1},\quad Var(y_{it}|\mathfrak{D}_{it}){=}\sigma_v^2,$$

where the conditioning information set is: $\mathfrak{D}_{it}{=}\{\mathbf{X}_{it}{=}\mathbf{x}_{it},\mathbf{X}_{it-1}{=}\ \mathbf{x}_{it-1},\sigma(y_{it-1})\}$. This gives rise to the non-*operational* dynamic statistical model:

$$y_{it}{=}\alpha_0(i,t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ (v_{it}|\mathfrak{D}_{it}){\backsim}\mathsf{NIID}(0,\sigma_v^2),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

For $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the number of unknown parameters is: $m{=}NT{+}k{+}(k{+}1){+}1{=}2k{+}NT{+}2{=}2510$.

As in the static case, when the *heterogeneity is separable* in the sense that:

$$\alpha_0(i,t)=\gamma_0(i){+}\delta_0(t),\ i{=}1,2,...,N,\ t{=}1,2,...T,$$

this gives rise to a potentially operational dynamic statistical model:

$$y_{it}{=}\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ (v_{it}|\mathfrak{D}_{it}){\backsim}\mathsf{NIID}(0,\sigma_v^2),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The complete specification of this model in terms of the observables is given in table 9.

Table 9 - Dynamic Fixed Individual and Time Effects Panel Model

Statistical GM:		$y_{it}{=}\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$
[1]	Normality:	$(y_{it}|\mathfrak{D}_{it})\backsim\mathsf{N}(.,.),$
[2]	Linearity:	$E\left(y_{it}|\mathfrak{D}_{it}\right){=}\gamma_0(i){+}\delta_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1},$
[3]	Homosk/sticity:	$Var\left(y_t|\mathfrak{D}_{it}\right){=}\sigma_v^2,$
[4]	Markov:	$\{(y_{it}|\mathfrak{D}_{it})\,,\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ is a Markov process,
[5]	(a) (i,t)-invar.:	$(\boldsymbol{\gamma},\boldsymbol{\gamma}_1,\gamma_2,\sigma_v^2)$ are (i,t)-invariant,
	(b) t-invariance:	$\{\gamma_0(i),i{\in}\mathbb{N}\}$ are t-invariant, but i-heterogeneous
	(c) i-invariance:	$\{\delta_0(t),t{\in}\mathbb{T}\}$ are i-invariant, but t-heterogeneous

For $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the number of unknown parameters is: $m{=}N{+}T{+}k{+}(k{+}1){+}1{=}N{+}T{+}2k{+}2{=}135.$

5.4 Dynamic Time Effects Panel Data Model

Consider the case where $\mathbf{Z}_{it}{:=}(y_{it},\mathbf{X}_{it}^{\top})^{\top}$ is Normal, Markov and mean t non-stationary process. These probabilistic assumptions imply that:

$$\begin{aligned}
D(\mathbf{Z}_{11},\mathbf{Z}_{21},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{M}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t)){=}\\
&\overset{\mathsf{M\&S}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}(t)){=}\\
&=\prod_{i=1}^{N}\prod_{t=1}^{T}D(y_{it}|\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(t))D(\mathbf{X}_{it}|\mathbf{Z}_{it-1};\boldsymbol{\varphi}_2(t))
\end{aligned}$$

where $D(\mathbf{Z}_{it},\mathbf{Z}_{it-1};\boldsymbol{\phi}(t))$ takes the form:

$$\begin{pmatrix}\mathbf{Z}_{it}\\ \mathbf{Z}_{it-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\boldsymbol{\mu}(t)\\ \boldsymbol{\mu}(t{-}1)\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1)\\ \boldsymbol{\Sigma}(1)^{\top} & \boldsymbol{\Sigma}(0)\end{pmatrix}\right),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The regression and and skedastic functions associated with $D(\mathbf{y}_{it}|\ \mathbf{Z}_{it-1}{,}\mathbf{x}_{it};\boldsymbol{\varphi}_1(t))$ are:

$$E(y_{it}|\mathfrak{D}_{it}){=}\alpha_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1},\quad Var(y_{it}|\mathfrak{D}_{it}){=}\sigma_v^2,$$

where the conditioning information set is: $\mathfrak{D}_{it}{=}\{\mathbf{X}_{it}{=}\mathbf{x}_{it},\mathbf{X}_{it-1}{=}\ \mathbf{x}_{it-1},\sigma(y_{it-1})\}$. This gives rise to the *operational* dynamic statistical model:

$$y_{it}{=}\alpha_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ (v_{it}|\mathfrak{D}_{it}){\backsim}\mathsf{NIID}(0,\sigma_v^2),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The complete specification of this model in terms of the observables is given in table 10.

Table 10 - Dynamic Time Effects Panel Data Model

Statistical GM: $y_{it}{=}\alpha_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1}{+}v_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$

[1]	Normality:	$(y_{it}\vert\mathfrak{D}_{it})\backsim\mathsf{N}(.,.),$
[2]	Linearity:	$E\left(y_{it}\vert\mathfrak{D}_{it}\right)=\alpha_0(t){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1{+}\gamma_2 y_{it-1},$
[3]	Homosk/sticity:	$Var\left(y_t\vert\mathfrak{D}_{it}\right)=\sigma_v^2,$
[4]	Markov:	$\{(y_{it}\vert\mathfrak{D}_{it}),\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ is a Markov process,
[5]	(a) (i,t)-invar.:	$(\boldsymbol{\gamma},\boldsymbol{\gamma}_1,\gamma_2,\sigma_v^2)$ are (i,t)-invariant,
	(b) i-invariance:	$\{\alpha_0(t),\ t{\in}\mathbb{T}\}$ are i-invariant, but t-heterogeneous

For $k{=}4$, $N{=}100$, $T{=}25$, $NT{=}2500$, the number of unknown parameters is: $m{=}T{+}k{+}(k{+}1){+}1{=}T{+}2k{+}2{=}35$.

5.5 Dynamic Fixed Individual and Slope Effects Panel Data Model

Consider the case where $\mathbf{Z}_{it}:=(y_{it},\mathbf{X}_{it}^{\top})^{\top}$ is Normal, Markov and t-stationary, but i-non-stationary. These probabilistic assumptions imply that:

$$\begin{aligned}D(\mathbf{Z}_{11},...,\mathbf{Z}_{NT};\boldsymbol{\phi}) &\overset{\mathsf{M}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}\vert\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i,t))=\\ &\overset{\mathsf{M\&S}}{=}\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{Z}_{it}\vert\mathbf{Z}_{it-1};\boldsymbol{\varphi}(i))=\\ &=\prod_{i=1}^{N}\prod_{t=1}^{T}D(\mathbf{y}_{it}\vert\mathbf{Z}_{it-1},\mathbf{x}_{it};\boldsymbol{\varphi}_1(i))\cdot D(\mathbf{X}_{it}\vert\mathbf{Z}_{it-1};\boldsymbol{\varphi}_2(i))\end{aligned}$$

where $D(\mathbf{Z}_{it},\mathbf{Z}_{it-1};\boldsymbol{\phi}(i))$ takes the form:

$$\begin{pmatrix}\mathbf{Z}_{it}\\ \mathbf{Z}_{it-1}\end{pmatrix}\backsim\mathsf{N}\left(\begin{pmatrix}\boldsymbol{\mu}(i)\\ \boldsymbol{\mu}(i)\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}(i,0) & \boldsymbol{\Sigma}(i,1)\\ \boldsymbol{\Sigma}(i,1)^{\top} & \boldsymbol{\Sigma}(i,0)\end{pmatrix}\right),\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T}.$$

The regression and and skedastic functions associated with $D(\mathbf{y}_{it}\vert\mathbf{Z}_{it-1}{,}\mathbf{x}_{it};\boldsymbol{\varphi}_1(i))$ are:

$$E(y_{it}\vert\mathfrak{D}_{it}){=}\alpha_0(i){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i){+}\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i){+}\gamma_2(i)y_{it-1},\quad Var(y_{it}\vert\mathfrak{D}_{it}){=}\sigma_v^2(i),$$

where the conditioning information set is: $\mathfrak{D}_{it}{=}\{\mathbf{X}_{it}{=}\mathbf{x}_{it},\mathbf{X}_{it-1}{=}\ \mathbf{x}_{it-1},\sigma(y_{it-1})\}$. This gives rise to the non-*operational* **dynamic statistical model**:

$$\begin{gathered}y_{it}{=}\alpha_0(i){+}\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i){+}\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i){+}\gamma_2(i)y_{it-1}{+}v_{it},\quad i{\in}\mathbb{N},\ t{\in}\mathbb{T},\\ (v_{it}\vert\mathfrak{D}_{it})\backsim\mathsf{NIID}(0,\sigma_v^2(i)).\end{gathered}$$

The complete specification of this model in terms of the observables is given in table 11.

Table 11 - Dynamic Fixed Individual and Slope Effects Panel Model

Statistical GM: $y_{it}=\alpha_0(i)+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i)+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i)+\gamma_2(i)y_{it-1}+v_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}.$

[1]	Normality:	$(y_{it} \mid \mathfrak{D}_{it}) \backsim \mathsf{N}(.,.),$
[2]	Linearity:	$E\left(y_{it}\vert\mathfrak{D}_{it}\right)=\alpha_0(i)+\mathbf{x}_{it}^{\top}\boldsymbol{\gamma}(i)+\mathbf{x}_{it-1}^{\top}\boldsymbol{\gamma}_1(i)+\gamma_2(i)y_{it-1},$
[3]	Homosk/sticity:	$Var\left(y_t \mid \mathfrak{D}_{it}\right)=\sigma_v^2,$
[4]	Markov:	$\{(y_{it} \mid \mathfrak{D}_{it}),\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ is a Markov process,
[5]	t-invariance:	$\{(\alpha_0(i),\boldsymbol{\gamma}(i),\boldsymbol{\gamma}_1(i),\gamma_2(i),\sigma_v^2(i)),\ i\in\mathbb{N}\}$ are t-invariant, but i-heterogeneous

For $k{=}4$, $N{=}100$, $T{=}25$, the number of unknown parameters is: $m{=}N{+}kN{+}(k{+}1)N{+}N{=}N\left(2k{+}3\right){=}1100$. On the other hand for $k{=}4$, $N{=}25$, $T{=}100$, the number of unknown parameters is: $m{=}N{+}kN{+}(k{+}1)N{+}N{=}N\left(2k{+}3\right){=}275$, which can be manageable under certain conditions.

6 Enhancing the precision of panel data modeling

The rationale underlying the specification of statistical models in the context of the PR approach is that the more *specific* and *complete* the probabilistic structure imposed on the vector stochastic process $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$ is, the more precise the inference based on the related model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$, as long as statistical adequacy has been secured - hence the important role ascribed to statistical model validation. The *weaker* (less specific) the postulated probabilistic structure the *less precise and incisive* the inference. Moreover, weaker probabilistic assumptions do not necessarily render the inference more reliable, just because they are less specific. Indeed, the trade-off between specificity of probabilistic assumptions and statistical adequacy is a false dilemma. What renders inference more reliable is *valid* probabilistic assumptions, not *weak*; the latter only contribute to the imprecision of the resulting inferences even when the premises are valid; see Spanos (2000).

This goes against the econometric textbook conventional wisdom that *weaker assumptions* are less vulnerable to *misspecification*; hence the current popularity of the Instrumental Variables (IV), Generalized Method of Moments (GMM) as well as Nonparametric methods; see Spanos (2007b). This is a flawed argument for several reasons. **First**, *weaker* assumptions indirectly imposed on the stochastic process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ underlying the data $\mathbf{Z}_0$, that largely ignore the probabilistic structure of the data, are *more* not less vulnerable to statistical misspecification, and they will invariably give rise to *less precise inferences.* **Second**, *non-testable assumptions, often incomplete,* made for mathematical convenience, render the substantiation of *statistical adequacy impossible.* **Third**, the fact that one can demonstrate the existence of CAN estimators, under certain assumptions about the underlying process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$, does not secure the reliability of the resulting inference results based on a given sample n, even when the invoked assumptions have been validated vis-a-vis data $\mathbf{Z}_0$; it is hopeless in cases where the validation facet has been ignored. What renders an inference reliable in practice is the approximate equality between actual and nominal error probabilities.

As the statistical models given in tables 1-11 stand, do not always take full advantage of the large sample size available in the case of panel data in order to enhance the precision of inference. This is primarily because of the way the heterogeneity is modeled raising the incidental parameter problems when any of the parameters are changing with i or/and t. This renders asymptotic results such as consistency and asymptotic Normality problematic since the number of unknown parameters increases with N or/and T. To take full advantage of the additional data information contained in panel data one needs to 'invent' or 'borrow' heterogeneity structures that would address the incidental parameter problem. This is exactly what fields like geography, geoscience and environmental science had to do that to take into account the spatial heterogeneity and dependence exhibited by their data; see Gressie, and Wikle (2011).

The best way forward in order to make the best of what panel data can offer is to construct particular forms of $(i,t)-$heterogeneity that would capture the observed regularities in panel data, in an analogous way that Markov and related forms of *dependence* capture the temporal/spatial structure of panel series data in a great variety of circumstances. The equal-correlation structure imposed by the stochastic effects model, and modeling spatial dependence by emulating the time series error autocorrelation using contiguity weight matrices (Anselin, 2001), are very restrictive.

In Spanos (1999), ch. 8, it is argued that heterogeneity restrictions, more than the other the two broad categories of probabilistic assumptions (distributional and dependence), requires further enrichment in order to account for such systematic information in panel data. These forms of heterogeneity could be as simple as groupings of the individual observation units into homogeneous sub-groups. As long as these groupings do not change with the index i, the incidental parameter problem can be addressed and the precision of inference is substantially enhanced.

7 Summary and conclusions

Despite the impressive developments in panel data modeling, the statistical foundations of such models are shown to be rather weak in so far that they are inadequate for securing the reliability and precision of inference. Defining the statistical premises of inference by making sufficient assumptions in terms of unobservable error terms to ensure the existence of CAN estimators does not secure the reliability or precision of inference. The paper proposes a recasting of the statistical foundations of panel data models using the Probabilistic Reduction perspective where statistical models are viewed as parameterizations of the observable stochastic (vector) process $\{\mathbf{Z}_{it},\ i{\in}\mathbb{N},\ t{\in}\mathbb{T}\}$ underlying the data $\mathbf{Z}_0$.

The *first* advantage of this recasting is that it offers a complete, internally consistent and testable list of probabilistic assumptions for the most popular statistical models for panel data. If one does not even know what probabilistic assumptions are invoked by the inference procedures in question, one cannot even begin to address the issue of statistical adequacy.

The *second* advantage is that it provides a more appropriate framework for all three facets of modeling: specification, M-S testing and respecification. The PR perspective advances a coherent methodology on how to specify and validate statistical models by probing model assumptions using thorough M-S testing, isolate the sources of departures, and account for them in a respecified model with a view to secure statistical adequacy; see Spanos (1986, 2010c).

The *third* advantage is that it proposes pertinent interpretations for the individual-specific (fixed or random) and time-specific effects by relating such effects to the probabilistic structure of $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$. It is shown that the current interpretations of the individual-specific effects (fixed or random) need to be reconsidered. It is shown that the 'omitted latent variables' interpretation is not pertinent for the fixed effects term, but it can be rendered appropriate for the random effects term when properly interpreted in terms of the implicit statistical parameterizations relating to $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$. The proposed interpretation sheds additional light on the Mundlak formulation. Similarly, the 'neglected heterogeneity' is pertinent for the fixed effects term when it is related to the mean-heterogeneity of $\{\mathbf{Z}_{it},\ i\in\mathbb{N},\ t\in\mathbb{T}\}$. This provides a more appropriate framework for securing learning from panel data by bringing out the neglected facets of empirical modeling which include specification, misspecification testing and respecification.

Using the PR perspective several statistical models for panel data are given a complete list of assumptions in terms of the probabilistic structure of the observable processes in tables 1-11. These specifications bring out certain weaknesses in the probabilistic structure of current panel data models, including the inefficient way such models account for the heterogeneity (individual or time) and/or dependence in panel data. This opens the door to 'inventing', or 'borrowing' from other disciplines, probabilistic assumptions that can capture such regularities more efficiently, and thus enhancing the reliability and precision of inference based on panel data.

References

[1] Anselin, L. (2001), "Spatial Econometrics," pp. 310-330 in Baltagi (2001).

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

[3] Baltagi, B. H. (2001) (editor), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

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

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

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

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

[8] Greene, W. H. (2008), *Econometric Analysis*, 6th ed., Prentice Hall, NJ.

[9] Gressie, N. and C. K. Wikle (2011), Statistics for Spatio-Temporal Data, Wiley, NY.

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

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

[12] Renyi, A. (1970), *Probability Theory*, North-Holland, Amsterdam.

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

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

[15] Spanos, A. (2000), "Revisiting Data Mining: 'hunting' with or without a license," *The Journal of Economic Methodology*, **7**: 231-264.

[16] Spanos, A. (2006a) "Econometrics in Retrospect and Prospect," pp. 3-58 in Mills, T.C. and K. Patterson, *New Palgrave Handbook of Econometrics*, vol. 1, MacMillan, London.

[17] Spanos, A. (2006b), "Where Do Statistical Models Come From? Revisiting the Problem of Specification," pp. 98-119 in *Optimality: The Second Erich L. Lehmann Symposium*, edited by J. Rojo, Lecture Notes-Monograph Series, vol. 49, Institute of Mathematical Statistics.

[18] Spanos, A. (2006c), "Revisiting the omitted variables argument: substantive vs. statistical adequacy," *Journal of Economic Methodology*, **13**: 179–218.

[19] Spanos, A. (2007a), "Curve-Fitting, the Reliability of Inductive Inference and the Error-Statistical Approach," *Philosophy of Science*, **74**: 1046–1066.

[20] Spanos, A. (2007b), “The Instrumental Variables Method revisited: On the Nature and Choice of Optimal Instruments,” pp. 34-59 in *Refinement of Econometric Estimation and Test Procedures*, ed. by G. D. A. Phillips and E. Tzavalis, Cambridge University Press, Cambridge.

[21] Spanos, A. (2009), “Statistical Misspecification and the Reliability of Inference: the simple t-test in the presence of Markov dependence,” *Korean Economic Review,* **25**: 165-213.

[22] Spanos, A. (2010a), “Theory Testing in Economics and the Error Statistical Perspective,” pp. 202-246 in *Error and Inference*, edited by Mayo, D. G. and A. Spanos (2010).

[23] Spanos, A. (2010b), “Akaike-type Criteria and the Reliability of Inference: Model Selection vs. Statistical Model Specification,” *Journal of Econometrics,* **158**: 204-220.

[24] Spanos, A. (2010c), “Statistical Adequacy and the Trustworthiness of Empirical Evidence: Statistical vs. Substantive Information,” *Economic Modelling*, **27**: 1436–1452.

[25] Spanos, A. and A. McGuirk (2001), “The Model Specification Problem from a Probabilistic Reduction Perspective,” *Journal of the American Agricultural Association*, **83**: 1168-1176.

[26] Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, The MIT Press, Cambridge.