# Panel Data Estimation for Correlated Random Coefficients Models

**Cheng Hsiao [1,2,\*,†], Qi Li [3,†], Zhongwen Liang [4,†] and Wei Xie [1,†]**

[1]    Department of Economics, University of Southern California, Los Angeles, CA 90089, USA;
     best.xw@gmail.com
[2]    Department of Quantitative Finance, NTHU and WISE, Xiamen University, Xiamen 361005, China
[3]    Department of Economics, Texas A&M University, College Station, TX 77843, USA; qi-li@tamu.edu
[4]    Department of Economics, University at Albany, SUNY, Albany, NY 12222, USA; zliang3@albany.edu
\*    Correspondence: chsiao@usc.edu; Tel.: +1-213-740-2103
†    These authors contributed equally to this work.

**Abstract:** This paper considers methods of estimating a static correlated random coefficient model with panel data. We mainly focus on comparing two approaches of estimating unconditional mean of the coefficients for the correlated random coefficients models, the group mean estimator and the generalized least squares estimator. For the group mean estimator, we show that it achieves Chamberlain (1992) semi-parametric efficiency bound asymptotically. For the generalized least squares estimator, we show that when $T$ is large, a generalized least squares estimator that ignores the correlation between the individual coefficients and regressors is asymptotically equivalent to the group mean estimator. In addition, we give conditions where the standard within estimator of the mean of the coefficients is consistent. Moreover, with additional assumptions on the known correlation pattern, we derive the asymptotic properties of panel least squares estimators. Simulations are used to examine the finite sample performances of different estimators.

## 1. Introduction

One useful tool for reducing real-world details for econometric modeling is through "suitable" aggregations of micro data. For aggregation not to distort the fundamental behavioral relationships between the micro data and aggregate data, certain "homogeneity" conditions must hold between the micro units (e.g., Hsiao et al. 2005; Pesaran 2003; Stoker 1993; Theil 1954). However, the "homogeneity" assumption is often rejected by empirical investigators (e.g., Kuh 1963; Hsiao and Tahmiscioglu 1997). On the other hand, most policy makers are only interested in the average relationships of the population, not the individual relationship. Random coefficients formulation can be a useful tool to accommodate the "heterogeneity" among micro units and policy makers' desire to find the average relationship (e.g., Hsiao et al. 1993).

Standard random coefficients models assume the variation of coefficients are independent of the variation of regressors (e.g., Hsiao 1996; Hsiao and Pesaran 2008). In recent years, a great deal of attention has been devoted to the correlated random coefficients model. For instance, in the human capital literature, let the dependent variable $y$ denote the logarithm of earnings and the explanatory variable $x$ denote the years of schooling; the coefficient $\beta$ denotes the rate of return. It is possible that the return to schooling declines with the level of schooling. It is also plausible that

there are unmeasured ability or motivational factors that affect the return to schooling and are also correlated with the level of schooling (e.g., Card 1995; Heckman and Vytlacil 1998; Heckman et al. 2006; Heckman et al. 2010). Particularly, Heckman and Vytlacil (1998) propose an instrumental variable method for the population mean of slope coefficients but not the intercept in the cross sectional correlated random coefficients model. They require the existence of both instrumental variables for the regressors and random coefficients.

Many people have worked on correlated random coefficients panel data models. For instance, Chamberlain (1992) showed how to apply his general result on the semiparametric efficiency bound to a random coefficients panel model as an example. The model considered in Chamberlain (1992) also allows for time varying parameters, which is more general than our model. However, the expression of efficient bound obtained using Chamberlain (1992)'s formulas are different from the expression obtained using the direct derivation. We show, in this paper, that they are, indeed, exactly the same. Due to the inclusion of the time varying parameters, Chamberlain (1992) requires the number of time periods $T$ is greater than the number of random coefficients $K$. Otherwise, the information matrix of the time varying coefficients is singular. Graham and Powell (2012) further considered the situation when $T = K$, and proposed a novel "irregular" method that leads to consistent estimation. Their approach assumes the existence of panel data with two subpopulations, where one corresponds to units whose regressor values do not change across periods and the other changes across periods. Arellano and Bonhomme (2012) discuss the identification of the distribution of random coefficients conditional on the values of the regressors, extending the idea from Chamberlain (1992). Chernozhukov et al. (2013) consider more general nonseparable panel models that allow for correlated random coefficients model as a special case.

In this paper, we consider the parametric identification and estimation of the unconditional mean of the random coefficients using panel data when the regularity conditions hold. Two approaches are considered; the approach of ignoring the correlations between the coefficients and regressors and the approach of explicitly modeling the correlations between the coefficients and regressors.

The rest of the paper is organized as follows. We discuss the estimation of the unconditional mean of the random coefficients with panel data in Sections 2 and 3. Section 2 considers the approach without explicitly modeling the pattern of correlations. Section 3 considers the approach with explicit assumption about the correlations between the coefficients and regressors. Section 4 provides Monte Carlo results of the different estimators in a finite sample. Concluding remarks are in Section 5.

## 2. Panel Parametric Approaches without Explicit Assumption about the Correlations between Coefficients and Regressors

When only cross-sectional data are available, the identification conditions of average effects for a correlated random coefficients model require the existence of instrumental variables, which are very stringent and may not be satisfied for many data sets. However, when panel data are available, it is possible to obtain a consistent estimator of the population mean of random coefficients without the existence of instrumental variables.

Suppose there are $T$ time series observations of $(y_{it}, x_{it})_{t=1,\dots,T}$ for each individual $i$. Let $y_i$ and $x_i$ be $T \times 1$ vector and $T \times K$ matrix with typical row elements given by $y_{it}$ and $x_{it}^\top = (x_{it,1}, \dots, x_{it,K})$, respectively, for $i = 1, 2, \dots, N$. Also, let $\beta_i = (\beta_{i1}, \dots, \beta_{iK})^\top$. We have

$$y_i = x_i \beta_i + u_i, \qquad i = 1, \dots, N. \tag{1}$$

Let $u_i = (u_{i1}, \dots, u_{iT})^\top$, and we assume $u_i$ is iid across $i$, with $E(u_i|x_i) = 0$ and $E(u_i u_i^\top | x_i) = \Sigma_{x_i}$ (a $T \times T$ matrix). We assume that $\beta_i$ is iid with mean $\beta$ and variance $Var(\beta_i) = \Delta$. Then we can write

$$\beta_i = \beta + \alpha_i, \tag{2}$$

where

$$E(\alpha_i) = E(\beta_i - \beta) = 0,$$

and

$$Cov(\beta_i, \beta_j) = E(\alpha_i \alpha_j^\top) = \begin{cases} \Delta, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{3}$$

Substituting $\beta_i = \beta + \alpha_i$ into (1) yields

$$y_i = x_i \beta + x_i \alpha_i + u_i = x_i \beta + v_i, \tag{4}$$

where $v_i = u_i + x_i \alpha_i$.

The standard random coefficients model assumes that $\alpha_i$ is a random draw from a population with $E(\alpha_i | x_i) = 0$. Then

$$E(v_i | x_i) = E(x_i \alpha_i + u_i | x_i) = 0, \tag{5}$$

and

$$E(v_i v_i^\top | x_i) = x_i \Delta x_i^\top + \sigma_u^2. \tag{6}$$

Therefore, a consistent estimator of $\beta$ can be obtained by simply regressing $Y$ on $X$, where $Y$ and $X$ are of dimensions $N \times 1$ and $N \times K$, respectively. An efficient estimator of $\beta$ can be obtained by applying the generalized least squares estimator (GLS) (or feasible GLS) (e.g., Hsiao 2003, chp. 6; Swamy 1970).

When $E(\alpha_i | x_i) = 0$ is violated, which is very common in practice, there exist the correlations between the coefficients and regressors, which is our main focus in the paper. We discuss different conditions and estimations in the following subsections.

*2.1. Group Mean Estimator*

In this subsection we impose the following mild conditional moment restriction:

$$E(u_i | x_i) = 0. \tag{7}$$

Note that (7) is weaker than $E(u_i | x_i, \beta_i) = 0$ as we do not require that $\alpha_i$ and $u_i$ are orthogonal with each other. Equation (7) implies the following unconditional moment condition

$$E((x_i^\top x_i)^{-1} x_i^\top u_i) = 0. \tag{8}$$

When $(x_i^\top x_i)$ is invertible (which requires that $T \geq K$), then from (1) one obtains $(x_i^\top x_i)^{-1} x_i^\top u_i = (x_i^\top x_i)^{-1} x_i^\top y_i - \beta_i$. Taking expectation yields the unconditional moment condition,

$$E[(x_i^\top x_i)^{-1} x_i^\top y_i - \beta] = 0. \tag{9}$$

Moment condition (9) leads to the estimator of $\beta$ given by

$$\hat{\beta}_{GM} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i. \tag{10}$$

where $\hat{\beta}_i = (x_i^\top x_i)^{-1} x_i^\top y_i$. Estimator (10) is the group mean (GM) estimator of Pesaran and Smith (1995) or Hsiao et al. (1999).

Under certain regularity conditions, we show that the GM estimator achieves the semiparametric efficiency bound derived in Chamberlain (1992). Note that $(\alpha_i = \beta_i - \beta)$

$$\hat{\beta}_i - \beta = \alpha_i + (x_i^\top x_i)^{-1} x_i^\top u_i. \tag{11}$$

Then

$$
\begin{aligned}
Var(\hat{\beta}_i) &= Var(\alpha_i) + E[(x_i^\top x_i)^{-1}x_i^\top \Sigma_{x_i} x_i (x_i^\top x_i)^{-1}] + E[\alpha_i u_i^\top x_i (x_i^\top x_i)^{-1}] + E[(x_i^\top x_i)^{-1} x_i^\top u_i \alpha_i^\top] \\
&\equiv \Omega.
\end{aligned}
\tag{12}
$$

Particularly, in the uncorrelated case, we impose the restriction that

$$
E(u_i|x_i, \alpha_i) = 0.
\tag{13}
$$

Then the covariance term in (12) drops out. Moreover, if we further impose the conditional homoskedastic error assumption:

$$
Var(u_i|x_i) = Var(u_i) = \sigma_u^2 I_T.
\tag{14}
$$

Then $Var(\hat{\beta}_i)$ is simplified to

$$
\Delta + \sigma_u^2 E[(x_i^\top x_i)^{-1}],
\tag{15}
$$

where $\Delta = Var(\alpha_i)$.

The following proposition describes the asymptotic behavior of $\hat{\beta}_{GM}$.

**Proposition 1.** *If $E(u_i|x_i) = 0$ and $T \geq K$, then*

(i)　*The group mean estimator defined in (10) is $\sqrt{N}$-consistent and asymptotically normally distributed, specifically, we have*

$$
\sqrt{N}(\hat{\beta}_{GM} - \beta) \xrightarrow{d} N(0, \Omega),
\tag{16}
$$

　　　*where $\Omega$ is defined in (12).*

(ii)　*$\hat{\beta}_{GM}$ is semiparametrically efficient.*

(iii)　*If conditions (13) and (14) also hold, then the asymptotic variance $\Omega$ can be simplified to $\Omega = \Delta + \sigma_u^2 E[(x_i^\top x_i)^{-1}]$.*

**Proof.** (i) $\hat{\beta}_{GM} = \beta + \frac{1}{N}\sum_{i=1}^{N}[\alpha_i + (x_i^\top x_i)^{-1}x_i^\top u_i]$. Hence $\sqrt{N}(\hat{\beta}_{GM} - \beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} w_i$, where $w_i = \alpha_i + (x_i^\top x_i)^{-1}x_i^\top u_i$ is i.i.d. with mean zero and finite variance $\Omega$. Proposition 1(i) follows from the Lindeberg's central limit theorem. (iii) follows from (i), (13) and (14) directly. We postpone the proof for (ii) to the Appendix A.　□

**Remark 1.** *Note that (16) holds without imposing any restriction on the correlations between $x_i$ and $\alpha_i$, and $u_i$ and $\alpha_i$. The random coefficient $\alpha_i$ can be correlated with both $x_i$ and $u_i$ with arbitrary correlation patterns. Also, since $x_i$ can contain a constant (an intercept), the conventional fixed effects model is included in the correlated random coefficient model as a special case.*

*2.2. Generalized Least Squares Estimator*

In this subsection we consider a generalized least squares (GLS) estimator of $\beta$ under the assumption that $Cov(\beta_i, x_i) = 0$ and compare the relative efficiency of the group mean estimator and the GLS estimator. Under the assumption that

$$
E(\alpha_i|x_i) = 0.
\tag{17}
$$

and the assumptions of (13) and (14), i.e., $E(u_i|x_i, \alpha_i) = 0$ and $Var(u_i|x_i) = \sigma_u^2 I_T$, then the best linear unbiased estimator (BLUE) of $\beta$ is the generalized least squares estimator (e.g., Hsiao 2003, chp. 6):

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} x_i \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} y_i \right)$$

$$= \sum_{i=1}^{N} W_i \hat{\beta}_i, \tag{18}$$

where

$$W_i = \left\{ \sum_{i=1}^{N} [\Delta + \sigma_u^2 (x_i^\top x_i)^{-1}]^{-1} \right\}^{-1} [\Delta + \sigma_u^2 (x_i^\top x_i)^{-1}]^{-1}$$

is a positive definite weight matrix satisfying $\sum_{i=1}^{N} W_i = 1$.

Contrary to the group mean estimator (10) that takes the simple average of the individual least squares estimator, $\hat{\beta}_i$, the GLS estimator takes the weighted average of $\hat{\beta}_i$.

By noting that $y_i = x_i \beta_i + u_i = x_i \beta + \epsilon_i$, where $\epsilon_i = x_i \alpha_i + u_i$, we define $Y_{(NT) \times 1} = (y_1^\top, ..., y_N^\top)^\top$, $X_{(NT) \times K} = (x_1^\top, ..., x_N^\top)^\top$ and $\epsilon_{(NT) \times 1} = (\epsilon_1^\top, ..., \epsilon_N^\top)^\top$. Then we have $Var(\epsilon) = \Omega_{(NT) \times (NT)} = Block - diag(\Sigma_i)$ is a block-diagonal matrix with the *i*th block diagonal element $\Sigma_i = x_i \Delta x_i^\top + \sigma_u^2 I_T$. Hence,

$$\hat{\beta}_{GLS} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} Y$$

$$= \left( \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} x_i \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} y_i \right)$$

$$= \beta + \left( N^{-1} \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} x_i \right)^{-1} \left( N^{-1} \sum_{i=1}^{N} x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} \epsilon_i \right). \tag{19}$$

Then by the law of large numbers and the central limit theorem arguments and by noting that $Var(\epsilon_i | x_i) = x_i \Delta x_i^\top + \sigma_u^2 I_T$, we obtain

$$\sqrt{N}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} AN(0, A^{-1}) = N(0, A), \tag{20}$$

where

$$A = \{ E[x_i^\top (x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1} x_i] \}^{-1}. \tag{21}$$

Clearly, $\hat{\beta}_{GLS}$ is not feasible. A feasible GLS estimator is given with $\Delta$ and $\sigma_u^2$ replaced by

$$\hat{\Delta} = N^{-1} \sum_{i=1}^{N} (\hat{\beta}_i - \hat{\beta}_{GM})(\hat{\beta}_i - \hat{\beta}_{GM})^\top - \frac{\hat{\sigma}_u^2}{N} \sum_{i=1}^{N} (x_i^\top x_i)^{-1},$$

$$\hat{\sigma}_u^2 = N^{-1}(T - K)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}^\top \hat{\beta}_i)^2,$$

where $\hat{\beta}_i$ is given in (10). The consistency of $\hat{\Delta}$ and $\hat{\sigma}_u^2$ can be proved similarly as that of $\hat{\Delta}^*$ and $\tilde{\sigma}_u^2$ in the Appendix A.

**Remark 2.** *Note that without condition that $E(\alpha_i | x_i) = 0$ of (17), $\hat{\beta}_{GM}$ is still a root-N consistent estimator for $\beta$ as shown in Proposition 1 while $\hat{\beta}_{GLS}$ becomes inconsistent when T is finite due to $E(\epsilon_i | x_i) = x_i E(\alpha_i | x_i) \neq 0$. However, when T is large, by noting that $x_i^\top x_i / T = E(x_{it} x_{it}^\top) + O_p(1/\sqrt{T})$ under the strong mixing condition and the conditions given in Theorem 24.6 in Davidson (1994), the weight matrix $W_i$ is close to a constant matrix:*

$$
\begin{aligned}
W_i &= \left\{ \frac{1}{N} \sum_{i=1}^{N} [\Delta + \frac{1}{T}\sigma_u^2 (x_i^\top x_i/T)^{-1}]^{-1} \right\}^{-1} \frac{1}{N} [\Delta + \frac{1}{T}\sigma_u^2 (x_i^\top x_i/T)^{-1}]^{-1} \\
&= \left\{ \frac{1}{N} \sum_{i=1}^{N} [\Delta + \frac{1}{T}\sigma_u^2 [E(x_{it}x_{it}^\top)]^{-1}]^{-1} \right\}^{-1} \frac{1}{N}[\Delta + \frac{1}{T}\sigma_u^2 [E(x_{it}x_{it}^\top)]^{-1}]^{-1} + O_p\left(\frac{1}{NT^{3/2}}\right) \\
&\equiv \frac{1}{N} + O_p\left(\frac{1}{NT}\right).
\end{aligned}
\tag{22}
$$

*It is easy to see that $\hat\beta_{GLS} = N^{-1}\sum_{i=1}^N \hat\beta_i + O_p((NT)^{-1}\sum_{i=1}^N \hat\beta_i)$ is a consistent estimate for $\beta = E(\beta_i)$.*

The next proposition compares the relative efficiency of $\hat\beta_{GLS}$ and $\hat\beta_{GM}$ by comparing their asymptotic variances: $Avar(\sqrt{N}\hat\beta_{GLS}) = \{E[x_i^\top(x_i\Delta x_i^\top + \sigma_u^2 I_T)^{-1}x_i]\}^{-1}$ and $Avar(\sqrt{N}\hat\beta_{GM}) = \Omega = \Delta + \sigma_u^2 E[(x_i^\top x_i)^{-1}]$.

**Proposition 2.** *Assuming that T is small (but still $T \geq K$) and that conditions (13), (14) and (17) hold. Then $Avar(\sqrt{N}\hat\beta_{GLS}) \leq Avar(\sqrt{N}\hat\beta_{GM})$.*

The proof of Proposition 2 is given in the Appendix A.

Proposition 2 says that, under some additional assumptions, $\hat\beta_{GLS}$ is asymptotically more efficient than $\hat\beta_{GM}$. This is in no contradiction with Proposition 1(ii) because the result of Proposition 1 does not require any of the conditions (13), (14) and (17) to hold. With additional conditions, $\hat\beta_{GM}$ is no longer a semiparametric efficient estimator of $\beta$. However, these additional conditions, especially condition (17), are quite restrictive.

It was shown by Hsiao et al. (1999) that when $T$ is large, $\hat\beta_{GLS}$ becomes a consistent estimator for $\beta$ without needing the restrictive condition (17).

**Proposition 3.** *Under conditions (13) and (14), if both N, $T \to \infty$ and $N^{1/2}/T \to 0$, $\sqrt{N}(\hat\beta_{GLS} - \hat\beta_{GM}) = o_p(1)$.*

In other words, if both $N$ and $T$ are large and if $\lim_{N,T\to\infty}(N^{1/2}/T) \to 0$, contrary to the case of only cross-sectional data are available, one can ignore the issue of possible correlations between $\alpha_i$ and $x_i$ (i.e., we allow for $E(\alpha_i|x_i) \neq 0$) and simply treat the model as if $\beta_i$ and $x_i$ are uncorrelated and apply the conventional GLS (e.g., Hsiao 2003, eq. (6.2.6)).

*2.3. Within Estimator*

If $T < K$, neither the GM, nor the GLS can be implemented. However, the standard within estimator can still yield a consistent estimator of $\beta$ in certain cases. Let $\bar y_{i\cdot} = \frac{1}{T}\sum_t y_{it}$ and $\bar x_{i\cdot} = \frac{1}{T}\sum_t x_{it}$. The within estimator (or fixed effects estimator) first takes the deviation of each observation from its time series mean, then regress $(y_{it} - \bar y_{i\cdot})$ on $(x_{it} - \bar x_{i\cdot})$ (e.g., Hsiao 2003, chp. 3)). Model (1) leads to

$$
(y_{it} - \bar y_{i\cdot}) = (x_{it} - \bar x_{i\cdot})^\top \beta + (x_{it} - \bar x_{i\cdot})^\top \alpha_i + (u_{it} - \bar u_{i\cdot}), \quad i = 1,\dots,N,\ t = 1,\dots,T, \tag{23}
$$

where $\bar u_{i\cdot} = T^{-1}\sum_t u_{it}$. The fixed effects (FE) estimator of $\beta$ is the least squares estimator of (23).

$$
\begin{aligned}
\hat{\beta}_{FE} &= \left[\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top\right]^{-1}\left[\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})\right] \\
&= \beta + \left[\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top\right]^{-1} \\
&\quad \times \left[\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top \alpha_i + \sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(u_{it} - \bar{u}_{i\cdot})\right].
\end{aligned}
\tag{24}
$$

In general, (24) is inconsistent. However, if the data generating process of $x_{it}$ takes the form:

$$
x_{it} = \mu_i + \sum_j B_j \epsilon_{i,t-j}, \quad \sum \|B_j\| < \infty,
\tag{25}
$$

where $\mu_i$ is iid with mean $a$ and variance $\Sigma_\mu$ and $\epsilon_{it}$ is iid across both $i$ and over $t$ with

$$
E(\epsilon_{it}|\alpha_i) = E(\epsilon_{is}|\alpha_i) \equiv d_i \quad \text{for all } t, s = 1, \ldots, T.
\tag{26}
$$

Then (23) is consistent. To see this, note that under (25) and (26), we have

$$
E(x_{it}|\alpha_i) = E(\mu_i|\alpha_i) + \sum_j B_j E(\epsilon_{i,t-j}|\alpha_i) = \delta_i + d_i \sum_j B_j \equiv \mu_i^*,
\tag{27}
$$

where $\delta_i = E(\mu_i|\alpha_i)$ and $d_i = E(\epsilon_{i,t-j}|\alpha_i)$.

Let

$$
x_{it} = E(x_{it}|\alpha_i) + \eta_{it} \equiv \mu_i^* + \eta_{it},
\tag{28}
$$

where $\eta_{it} = x_{it} - \mu_i^* = (\mu_i - E(\mu_i|\alpha_i)) + \sum_j B_j(\epsilon_{i,t-j} - E(\epsilon_{i,t-j}|\alpha_i))$. Then $x_{it} - \bar{x}_{i\cdot} = \eta_{it} - \bar{\eta}_{i\cdot}$, where $\bar{m}_{i\cdot} = \frac{1}{T}\sum_{t=1}^T m_{it}$ ($m$ can be $x$ or $\eta$). Also, from $E(\eta_{it} - \bar{\eta}_{i\cdot}|\alpha_i) = 0$, we know that $E(x_{it} - \bar{x}_{i\cdot}|\alpha_i) \equiv E(\eta_{it} - \bar{\eta}_{i\cdot}|\alpha_i) = 0$. If in addition the following conditional homoskedastic error assumption holds:

$$
E[(x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top|\alpha_i] = E[(\eta_{it} - \bar{\eta}_{i\cdot})(\eta_{it} - \bar{\eta}_{i\cdot})^\top|\alpha_i] = C,
\tag{29}
$$

where $C = E[(\eta_{it} - \bar{\eta}_{i\cdot})(\eta_{it} - \bar{\eta}_{i\cdot})^\top]$ is a $K \times K$ nonsingular constant matrix. Then

$$
\frac{1}{NT}\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top \alpha_i \overset{p}{\to} C\, E[\alpha_i] = 0
\tag{30}
$$

as $N \to \infty$. Therefore

**Proposition 4.** *Under (25) and (29), the conventional fixed effects estimator is $\sqrt{N}$-consistent and asymptotically normally distributed as $N \to \infty$. The asymptotic covariance matrix of (24) can be approximated using Newey-West heteroscedasticity-autocorrelation consistent formula.*

When $(x_{it}, \alpha_i)$ has a joint elliptical distribution, the conditional homoscedasticity of $E[(x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top|\alpha_i] = C$ also holds (e.g., Fang and Zhang 1990; Gupta et al. 1993). Therefore,

**Proposition 5.** *When $(x_{it}, \alpha_i)$ are jointly elliptically distributed, or conditional homoscedasticity of $(x_{it} - \bar{x}_{i\cdot})$ of (29) holds, the FE estimator (24) is $\sqrt{N}$-consistent and asymptotically normally distributed.*

Another case where the fixed effects estimator can be consistent is that $(x_{it}, \alpha_i)$ are jointly symmetrically distributed. Since $x_{it} - \bar{x}_{i\cdot}$ has mean equal to zero, $(x_{it} - \bar{x}_{i\cdot}, \alpha_i)$ will be symmetrically distributed around $(0, 0)$, then $\frac{1}{NT}\sum_i \sum_t (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top \alpha_i \overset{p}{\to} 0$ even though $x_{it}$ has mean different from zero. We have

**Proposition 6.** *Under* (25), (26) *and if* $(x_{it}, \alpha_i)$ *are symmetrically distributed, the fixed effects estimator* (24) *is* $\sqrt{N}$ *consistent and asymptotically normally distributed.*

Wooldridge (2005) also discussed conditions for the validity of the fixed effects estimator. Although the conventional FE estimator (24) can yield a consistent estimator of $\beta$, if $x_{it}$ contains time-invariant variables, the mean effects of time-invariant variables cannot be identified by the conventional fixed effects estimator. Moreover, the FE estimator only makes use of the within (group) variation. Since, in general, the between group variation is much larger than within group variation, the FE estimator could also mean a loss of efficiency.

## 3. Panel Least Squares or Generalized Least Squares Estimator

If $\alpha_i$ is correlated with $x_i$, i.e., $E(\alpha_i|x_i) \neq 0$, we can re-write (1) as

$$y_{it} = x_{it}^\top \beta + x_{it}^\top E(\alpha_i|x_i) + v_{it}, \tag{31}$$

where $v_{it} = u_{it} + x_{it}^\top w_i$, $w_i = \alpha_i - E(\alpha_i|x_i)$. Equation (31) is no longer a linear function of $x_{it}$. For instance, suppose

$$E(\alpha_i|x_i) = a + B vec(x_i^\top), \tag{32}$$

as assumed by Mundlak (1978). Noting that

$$E(\alpha_i) = E[E(\alpha_i|x_i)] = a + BE(vec(x_i^\top)) = 0, \tag{33}$$

which implies that $a = -BE(vec(x_i^\top))$. Hence, (32) can be written as

$$E(\alpha_i|x_i) = B(vec(x_i^\top) - E(vec(x_i^\top))). \tag{34}$$

Equation (31) then becomes

$$y_{it} = x_{it}^\top \beta + x_{it}^\top B vec(x_i^\top - E(x_i^\top)) + v_{it}, \; i = 1, \ldots, N, t = 1, \ldots, T. \tag{35}$$

Let $v_i = (v_{i1}, \ldots, v_{iT})^\top$. Then $E(v_i|x_i) = 0$ and $E(v_i v_i^\top) = E(x_i \Delta^* x_i^\top) + \sigma_u^2 I_T$ where $\Delta^* = E(w_i w_i^\top) = E(w_i w_i^\top|x_i)$. Therefore, the least squares or the generalized least squares estimator of $\beta$ is $\sqrt{N}$ consistent provided

$$\frac{1}{N} \sum_{i=1}^N E \begin{pmatrix} x_i^\top x_i & x_i^\top (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \\ (x_i^\top \otimes vec((x_i - \bar{x}.)^\top))x_i & (x_i^\top \otimes vec((x_i - \bar{x}.)^\top))(x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \end{pmatrix} \tag{36}$$

is a full rank matrix, where $\bar{x}. = N^{-1} \sum_{i=1}^N x_i$.

Similar reasoning can be applied if $E(\alpha_i|x_i)$ is a higher order polynomial of $x_i$, say,

$$E(\alpha_i|x_i) = a + B vec(x_i^\top) + C vec(x_i^\top \otimes x_i^\top). \tag{37}$$

Then from $E[E(\alpha_i|x_i)] = 0$, we get $a = -BE(vec(x_i^\top)) - CE[vec(x_i^\top \otimes x_i^\top)]$, it follows that

$$E(\alpha_i|x_i) = B[vec(x_i^\top - E(x_i^\top))] + C[vec(x_i^\top \otimes x_i^\top - E(x_i^\top \otimes x_i^\top))]. \tag{38}$$

Substituting (38) into (31) we have

$$y_{it} = x_{it}^\top \beta + (x_{it}^\top \otimes (vec((x_i - E(x_i))^\top))^\top) vec(B^\top) + (x_{it}^\top \otimes (vec([(x_i \otimes x_i) - E(x_i \otimes x_i)]^\top))^\top) vec(C^\top) + v_{it}, \tag{39}$$

where $v_{it} = x_{it}^\top \omega_i + u_{it}$ and $\omega_i = \alpha_i - E(\alpha_i|x_i)$. By construction, $v_{it}$ is uncorrelated with the regressor. Therefore, the least squares (LS) or the feasible generalized least squares (FGLS) estimator of (39) yields a $\sqrt{N}$ consistent and asymptotically normally distributed estimator of $\beta$ when $\bar{x}.$ and $N^{-1}\sum_{i=1}^N x_i \otimes x_i$ are substituted in lieu of $E(x_i)$ and $E(x_i \otimes x_i)$ in (39).

Next, we derive the asymptotic distribution of $\hat{\beta}_{M,PLS}$ which is the estimator for $\beta$ in (35). The feasible GLS type estimator $\hat{\beta}_{M,PLS}$ of $\beta$ in (35) can be constructed as

$$\hat{\beta}_{M,PLS}$$
$$= \left(X^\top \hat{\Omega}^{-1} X - X^\top \hat{\Omega}^{-1} X_1 (X_1 \hat{\Omega}^{-1} X_1)^{-1} X_1^\top \hat{\Omega}^{-1} X\right)^{-1} \left(X^\top \hat{\Omega}^{-1} Y - X^\top \hat{\Omega}^{-1} X_1 (X_1 \hat{\Omega}^{-1} X_1)^{-1} X_1^\top \hat{\Omega}^{-1} Y\right),$$

$$X_1 = (x_1^\top \otimes vec((x_1 - \bar{x}.)^\top), \dots, x_N^\top \otimes vec((x_N - \bar{x}.)^\top))^\top, \quad \bar{x}. = N^{-1} \sum_{i=1}^N x_i,$$

$$\hat{\Omega} = diag\{\hat{\Sigma}_1, \dots, \hat{\Sigma}_N\}, \quad \hat{\Sigma}_i = x_i \hat{\Delta}^* x_i^\top + \tilde{\sigma}_u^2 I_T,$$

where $X$ and $Y$ are defined in (19), and

$$\tilde{\sigma}_u^2 = \left[\tilde{\sigma}_{GM}^2 - (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T x_{it}^\top \tilde{V}_{GM} x_{it}\right] / \left\{1 - (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T x_{it}^\top \hat{E}[(x_i^\top x_i)^{-1}] x_{it}\right\},$$

$$\hat{E}[(x_i^\top x_i)^{-1}] = N^{-1} \sum_{i=1}^N (x_i^\top x_i)^{-1},$$

$$\tilde{\sigma}_{GM}^2 = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - x_{it}^\top \hat{\beta}_{GM})^2, \quad \tilde{V}_{GM} = N^{-1} \sum_{i=1}^N (\hat{\beta}_i - \hat{\beta}_{GM})(\hat{\beta}_i - \hat{\beta}_{GM})^\top,$$

$$\hat{\Delta}^* = \tilde{V}_{GM} - \frac{\tilde{\sigma}_u^2}{N} \sum_{i=1}^N (x_i^\top x_i)^{-1} - \hat{B}^2 \frac{1}{N} \sum_{i=1}^N \left(vec\left(x_i^\top - \frac{1}{N} \sum_{i=1}^N x_i^\top\right)\right)^2,$$

$\hat{\beta}_i$ is given in (10), and $\hat{B}$ is the OLS estimator of $B$ in (35).

We have the following proposition

**Proposition 7.** *Under conditions (13), (14) and (34), we obtain*

$$\sqrt{N}(\hat{\beta}_{M,PLS} - \beta) \xrightarrow{d} N(0, V_{M,PLS}), \tag{40}$$

*where*

$$V_{M,PLS} = Var[E(\alpha_i|x_i)] + \left(E[x_i^\top \Sigma_i^{-1} x_i] - E[x_i^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)]\right.$$

$$\times \left(E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)]\right)^{-1} E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1} x_i]\Big)^{-1},$$

$$\Sigma_i = x_i \Delta^* x_i^\top + \sigma_u^2 I_T, \quad \Delta^* = E(w_i w_i^\top), \quad w_i = \alpha_i - E(\alpha_i|x_i).$$

The proof is given in the Appendix A.

**Remark 3.** *From the total variance decomposition,*

$$\Delta = Var(\alpha_i) = Var[E(\alpha_i|x_i)] + E[Var(\alpha_i|x_i)].$$

*Furthermore,*

$$Var[E(\alpha_i|x_i)] + E[Var(\alpha_i|x_i)] = Var[E(\alpha_i|x_i)] + E[\Delta^*]$$

*under the assumption that $E(\alpha_i|x_i)$ has a known functional form, for example as given by (34) or (38). Therefore, the asymptotic variance in Proposition 1(iii) can be rewritten as*

$$\Omega = \Delta + \sigma_u^2 E[(x_i^\top x_i)^{-1}] = Var[E(\alpha_i|x_i)] + E[\Delta^*] + \sigma_u^2 E[(x_i^\top x_i)^{-1}].$$

*By contrast, from Proposition 7 with some algebraic operations,*

$$V_{M,PLS} = Var[E(\alpha_i|x_i)] + \left( E[x_i^\top \Sigma_i^{-1} x_i] - E[x_i^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \right.$$

$$\left. \times \left( E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \right)^{-1} E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1} x_i] \right)^{-1}$$

$$\geq Var[E(\alpha_i|x_i)] + \left( E[x_i^\top \Sigma_i^{-1} x_i] \right)^{-1} = Var[E(\alpha_i|x_i)] + \{E[\Delta^* + \sigma_u^2(x_i^\top x_i)^{-1}]^{-1}\}^{-1},$$

*since $E[x_i^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \left( E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1}(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \right)^{-1}$ $E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1} x_i]$ is positive definite. Compared with the group mean estimator, it is not clear which estimator is more efficient, even though $\left( E[x_i^\top \Sigma_i^{-1} x_i] \right)^{-1} = \{E[\Delta^* + \sigma_u^2(x_i^\top x_i)^{-1}]^{-1}\}^{-1} \leq E[\Delta^*] + \sigma_u^2 E[(x_i^\top x_i)^{-1}].$*

## 4. Monte Carlo Studies

We consider several data generating designs for the following correlated random coefficients model

$$y_{it} = \beta_i x_{it} + u_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{41}$$

$$\beta_i = \beta + \alpha_i, \ \beta = 1, \tag{42}$$

and $u_{it}$ is a random draw from the standard normal distribution, and is independent with $(x_{it}, \alpha_i)$ for all simulation designs. The regressor $x_{it}$ and the random coefficient $\alpha_i$ are correlated with each other and are generated according to the following sample designs:

Design 1: Randomly draw

$$\begin{pmatrix} \alpha_i \\ v_{i0} \\ v_{i1} \\ v_{i2} \\ v_{i3} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0 & 0 & 0 \\ 0.2 & 0 & 0.5 & 0 & 0 \\ 0.2 & 0 & 0 & 0.5 & 0 \\ 0.2 & 0 & 0 & 0 & 0.5 \end{pmatrix} \right), \tag{43}$$

then generate $x_{it} = v_{it} + 0.3 v_{i,t-1}$.

Design 2: Randomly draw $(\alpha_i, v_{i0}, v_{i1}, v_{i2}, v_{i3})$ from independent multivariate normal as in (43), then generate $x_{it} = 2 + v_{it} + v_{i,t-1}$.

Design 3: Randomly draw $\alpha_i$ from a uniform distribution $(-0.75, 0.75)$. Then generate

$$x_{it} = \alpha_i + w_{it}, \tag{44}$$

where $w_{it} = 1 + \chi^2(5)$, where $\chi^2(5)$ is a random draw from a chi-square distribution with five degrees of freedom.

Design 4: Randomly draw $\alpha_i$ from uniform distribution $(-0.75, 0.75)$ and $w_{it}$ from $1 + \chi^2(5)$ where $\chi^2(5)$ is chi-square distribution with five degrees of freedom. Then generate

$$x_{it} = \alpha_i + w_{it} + 0.3 w_{i,t-1}. \tag{45}$$

Designs 5 and 6: Generate $\alpha_i$ from Gamma(1,1), then generate $x_{it}$ according to (44) and (45), respectively.[1]

Designs 7 and 8: Generate $\alpha_i$ from Beta(1,3), then generate $x_{it}$ according to (44) and (45), respectively.

Designs 9 and 10: Generate $x_{it} = 1 + \alpha_i w_{it}$, where $\alpha_i$ is from Gamma(1,1) and Beta(1,3), respectively.

Designs 1 and 2 generate $\beta_i$ and $x_{it}$ from uniformly jointly symmetric distribution with mean (1,0), and mean (1,2), respectively. Designs 3–8 generate $\beta_i$ and $x_{it}$ from correlated but not symmetrical distributions. Designs 9 and 10 yield nonlinearly correlated $\beta_i$ and $x_{it}$.

We examine the finite sample performances of the least squares estimator (LS), the conventional fixed effects estimator, (24), (FE), the panel least squares estimator, (35), (PLS1), (39), (PLS2), the group mean estimator, (10), (GM), and the generalized least squares estimator, (18), (GLS).

We consider a case where $T = 3$ and 20, and $N = 50$, 100, and 200, respectively. We replicate the experiments two thousand times. The simulation results are consistent with the theoretical results. Because the results for $T = 3$ or $T = 20$ are similar, we only report the results for $T = 3$. The results for $T = 20$ are available upon request. Tables 1 and 2 provide the bias and mean squared errors of our estimators. As expected, when ($\alpha_i$ and $x_{it}$) are generated from symmetric distribution with mean (0,0) (Design 1), the LS estimator yields unbiased estimator. However, if ($\alpha_i$ and $x_{it}$) are generated from symmetric distribution with mean (0,2) (Design 2) or nonsymmetric distribution (Design 3), the LS estimator yields biased estimates of $\beta(=1)$. Most other estimators will work well if there exhibits linear correlations between the coefficients and regressors except the LS estimator. Performances under nonlinear correlations (Designs 9 and 10) will tell them apart. The GM always enjoys the highest efficiency, which is consistent with our theories. The panel least squares estimators PLS1 and PLS2 could pick up the linear correlations well but fail to do so in the nonlinear cases. Also, they create larger biases to achieve smaller mean squared errors compared with the FE estimator. The GLS ignoring the correlations between the coefficients and regressors would work when there are no correlations or linear correlations, but fails when the correlation is nonlinear and $T$ is small which can be seen from the results of Designs 9 and 10. The conventional FE estimator is nearly unbiased if $\alpha_i$ and $x_{it}$ are symmetric (Designs 1 to 8), but is not consistent when the condition is not satisfied, and yields larger mean squared errors than that of GM or GLS. From Table 2 we see that for design 9, only GM estimator exhibits consistency result, while all other estimators' estimation MSEs do not decrease as sample size $N$ increases.

**Table 1.** The Bias of the LS, FE, PLS1, PLS2, GM and GLS.

|  | N | LS | FE | PLS1 | PLS2 | GM | GLS |
|---|---|---|---|---|---|---|---|
| | 50 | −0.0114 | −0.0115 | −0.0095 | −0.0090 | −0.0076 | −0.0115 |
| Design 1 | 100 | 0.0048 | 0.0029 | 0.0033 | 0.0032 | 0.0057 | 0.0043 |
| | 200 | −0.0021 | −0.0005 | −0.0015 | −0.0015 | 0.0004 | −0.0015 |
| | 50 | 0.3084 | −0.0087 | −0.0031 | −0.0054 | −0.0031 | 0.1613 |
| Design 2 | 100 | 0.3213 | 0.0034 | 0.0054 | 0.0039 | 0.0043 | 0.1835 |
| | 200 | 0.3198 | −0.0011 | 0.0016 | −0.0002 | −0.0003 | 0.1845 |
| | 50 | 0.0452 | −0.0043 | −0.0006 | −0.0006 | −0.0020 | 0.0185 |
| Design 3 | 100 | 0.0485 | −0.0004 | 0.0038 | 0.0036 | −0.0001 | 0.0235 |
| | 200 | 0.0477 | −0.0016 | 0.0039 | 0.0036 | −0.0007 | 0.0222 |

---

[1]　Note that when $\alpha_i$ is generated from Gamma(1,1) and Beta(1,3), the mean of $\alpha_i$ is 1 and 0.25, respectively. Therefore, $\beta$ in these cases will be 2 and 1.25, respectively.

**Table 1.** *Cont*.

|  | N | LS | FE | PLS1 | PLS2 | GM | GLS |
|---|---|---|---|---|---|---|---|
| | 50 | 0.0376 | −0.0012 | 0.0028 | 0.0022 | −0.0021 | 0.0150 |
| Design 4 | 100 | 0.0396 | −0.0017 | 0.0054 | 0.0051 | −0.0002 | 0.0173 |
| | 200 | 0.0397 | −0.0007 | 0.0059 | 0.0054 | −0.0007 | 0.0182 |
| | 50 | 0.2600 | 0.0040 | −0.0482 | −0.0459 | −0.0004 | −0.0205 |
| Design 5 | 100 | 0.2610 | −0.0020 | −0.0518 | −0.0516 | −0.0022 | 0.0103 |
| | 200 | 0.2669 | 0.0054 | −0.0524 | −0.0523 | 0.0002 | 0.0417 |
| | 50 | 0.2143 | −0.0023 | −0.0240 | −0.0243 | −0.0005 | −0.0315 |
| Design 6 | 100 | 0.2149 | −0.0008 | −0.0252 | −0.0265 | −0.0022 | 0.0059 |
| | 200 | 0.2191 | 0.0015 | −0.0233 | −0.0249 | 0.0002 | 0.0310 |
| | 50 | 0.0094 | −0.0007 | 0.0011 | 0.0010 | 0.0005 | 0.0025 |
| Design 7 | 100 | 0.0102 | 0.0002 | 0.0013 | 0.0012 | 0.0006 | 0.0050 |
| | 200 | 0.0097 | 0.0001 | 0.0011 | 0.0010 | 0.0001 | 0.0057 |
| | 50 | 0.0075 | −0.0021 | 0.0014 | 0.0013 | 0.0005 | 0.0016 |
| Design 8 | 100 | 0.0082 | −0.0003 | 0.0016 | 0.0015 | 0.0006 | 0.0038 |
| | 200 | 0.0080 | 0.0001 | 0.0013 | 0.0012 | 0.0002 | 0.0046 |
| | 50 | 1.6667 | 1.6889 | 0.1461 | 0.1086 | 0.0004 | 0.0252 |
| Design 9 | 100 | 1.7484 | 1.7963 | 0.1763 | 0.1490 | −0.0025 | 0.0262 |
| | 200 | 1.8039 | 1.8913 | 0.2183 | 0.1856 | 0.0006 | 0.0237 |
| | 50 | 0.1782 | 0.2316 | 0.0263 | 0.0218 | 0.0011 | 0.0899 |
| Design 10 | 100 | 0.1830 | 0.2446 | 0.0296 | 0.0254 | −0.0002 | 0.0918 |
| | 200 | 0.1846 | 0.2472 | 0.0329 | 0.0296 | −0.0002 | 0.0923 |

**Table 2.** The Mean Squared Errors of the LS, FE, PLS1, PLS2, GM and GLS.

|  | N | LS | FE | PLS1 | PLS2 | GM | GLS |
|---|---|---|---|---|---|---|---|
| | 50 | 0.0622 | 0.0636 | 0.0423 | 0.0420 | 0.0564 | 0.0542 |
| Design 1 | 100 | 0.0322 | 0.0304 | 0.0208 | 0.0208 | 0.0281 | 0.0283 |
| | 200 | 0.0168 | 0.0162 | 0.0108 | 0.0107 | 0.0148 | 0.0148 |
| | 50 | 0.1231 | 0.0586 | 0.0235 | 0.0233 | 0.0215 | 0.0550 |
| Design 2 | 100 | 0.1178 | 0.0287 | 0.0119 | 0.0118 | 0.0113 | 0.0482 |
| | 200 | 0.1097 | 0.0154 | 0.0061 | 0.0061 | 0.0055 | 0.0416 |
| | 50 | 0.0073 | 0.0109 | 0.0045 | 0.0044 | 0.0040 | 0.0071 |
| Design 3 | 100 | 0.0051 | 0.0054 | 0.0024 | 0.0024 | 0.0021 | 0.0041 |
| | 200 | 0.0036 | 0.0028 | 0.0011 | 0.0011 | 0.0010 | 0.0023 |
| | 50 | 0.0063 | 0.0106 | 0.0043 | 0.0042 | 0.0039 | 0.0066 |
| Design 4 | 100 | 0.0042 | 0.0056 | 0.0023 | 0.0022 | 0.0020 | 0.0038 |
| | 200 | 0.0029 | 0.0029 | 0.0011 | 0.0011 | 0.0009 | 0.0021 |
| | 50 | 0.1344 | 0.0579 | 0.0236 | 0.0228 | 0.0208 | 0.0360 |
| Design 5 | 100 | 0.1004 | 0.0282 | 0.0150 | 0.0144 | 0.0105 | 0.0210 |
| | 200 | 0.0877 | 0.0143 | 0.0087 | 0.0085 | 0.0050 | 0.0131 |
| | 50 | 0.1036 | 0.0549 | 0.0220 | 0.0213 | 0.0207 | 0.0340 |
| Design 6 | 100 | 0.0739 | 0.0283 | 0.0123 | 0.0120 | 0.0105 | 0.0198 |
| | 200 | 0.0621 | 0.0147 | 0.0062 | 0.0061 | 0.0050 | 0.0121 |
| | 50 | 0.0014 | 0.0029 | 0.0011 | 0.0011 | 0.0010 | 0.0014 |
| Design 7 | 100 | 0.0008 | 0.0016 | 0.0005 | 0.0005 | 0.0005 | 0.0008 |
| | 200 | 0.0004 | 0.0007 | 0.0003 | 0.0003 | 0.0002 | 0.0004 |
| | 50 | 0.0012 | 0.0032 | 0.0010 | 0.0010 | 0.0009 | 0.0013 |
| Design 8 | 100 | 0.0006 | 0.0015 | 0.0005 | 0.0005 | 0.0004 | 0.0006 |
| | 200 | 0.0003 | 0.0008 | 0.0002 | 0.0002 | 0.0002 | 0.0003 |

**Table 2.** *Cont.*

|  | N | LS | FE | PLS1 | PLS2 | GM | GLS |
|---|---|---|---|---|---|---|---|
| | 50 | 3.4034 | 3.6486 | 0.1102 | 0.0695 | 0.0214 | 0.0223 |
| Design 9 | 100 | 3.4720 | 3.7850 | 0.0943 | 0.0693 | 0.0108 | 0.0113 |
| | 200 | 3.4851 | 3.9348 | 0.0983 | 0.0713 | 0.0051 | 0.0059 |
| | 50 | 0.0350 | 0.0699 | 0.0034 | 0.0031 | 0.0026 | 0.0100 |
| Design 10 | 100 | 0.0353 | 0.0678 | 0.0024 | 0.0021 | 0.0014 | 0.0093 |
| | 200 | 0.0349 | 0.0654 | 0.0018 | 0.0016 | 0.0007 | 0.0090 |

## 5. Concluding Remarks

Parameter heterogeneity among micro units is quite common and a random coefficients model is a convenient way to take into account unobserved heterogeneity in pooling the panel data (e.g., Hsiao and Tahmiscioglu 1997; Hsiao et al. 2005). However, as demonstrated by Card (1995), Heckman and Vytlacil (1998), etc. the parameter variation could often be correlated. When only cross-sectional data are available, it was shown by Heckman and Vytlacil (1998) that the consistent estimate of the mean of the coefficients requires very stringent conditions. In this paper, we show that when panel data are available, there is no need to find separate instruments for $x_{it}$ and $\beta_i$. As long as the time series dimension $T$ is no smaller than the number of regressors $K$, we can accommodate the correlations between the random coefficients ($\alpha_i$) and regressors ($x_i$). Particularly, the group mean estimator is consistent and achieves Chamberlain's semiparametric efficiency bound. We also give conditions under which the conventional fixed effects estimator and the generalized least squares estimator lead to consistent estimates of the mean coefficient vector $\beta = E(\beta_i)$. Simulation results strongly support our theoretical analysis. In particular, our Monte Carlo studies show that the group mean estimate is, indeed, robust to a variety of patterns of correlations between the coefficients and regressors.

**Author Contributions:** The authors contributed equally to this work.

## Appendix A

**Proof of Proposition 1(ii).** Applying Chamberlain (1992)'s result to our case, we have that the semiparametric lower bound for the asymptotic variance of a regular semiparametric estimator of $\beta$ is given by

$$V_\beta = Var[E(\beta_i|x_i)] + E[(x_i^\top Var(y_i|x_i)x_i)^{-1}]$$

Since $E(\beta_i|x_i) = \beta + E(\alpha_i|x_i)$, it implies $Var[E(\beta_i|x_i)] = Var[E(\alpha_i|x_i)]$. Also, $Var(y_i|x_i) = Var(x_i\alpha_i + u_i|x_i) = x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i\alpha_i^\top|x_i)x_i^\top$ because $Cov(\alpha_i, u_i|x_i) = E(\alpha_i u_i^\top|x_i)$. Thus, we have

$$
\begin{aligned}
V_\beta &= Var[E(\alpha_i|x_i)] + E\{(x_i^\top [x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i\alpha_i^\top|x_i)x_i^\top] x_i)^{-1}\} \\
&= Var[E(\alpha_i|x_i)] + E(M_i),
\end{aligned}
\tag{A1}
$$

where

$$M_i = \left( x_i^\top [x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top]^{-1}x_i \right)^{-1}.$$

Then from

$$M_i M_i^{-1} = M_i x_i^\top [x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top]^{-1}x_i = I_T,$$

post-multiplying both sides of the above equation by $x_i^\top (x_i x_i^\top)^{-1}$, we get

$$M_i x_i^\top [x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top]^{-1} = x_i^\top (x_i x_i^\top)^{-1}.$$

Pre-multiplying both sides of the above equation by $x_i$ gives

$$x_i M_i x_i^\top [x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top]^{-1} = I_T.$$

Then post-multiplying both sides of it by $[x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top]$ leads to

$$x_i M_i x_i^\top = x_i Var(\alpha_i|x_i)x_i^\top + Var(u_i|x_i) + x_i E(\alpha_i u_i^\top|x_i) + E(u_i \alpha_i^\top|x_i)x_i^\top.$$

Pre-multiplying both sides by $(x_i^\top x_i)^{-1}x_i^\top$ leads to

$$M_i x_i^\top = Var(\alpha_i|x_i)x_i^\top + (x_i^\top x_i)^{-1}x_i^\top Var(u_i|x_i) + E(\alpha_i u_i^\top|x_i) + (x_i^\top x_i)^{-1}x_i^\top E(u_i \alpha_i^\top|x_i)x_i^\top.$$

Post-multiplying both sides by $x_i(x_i^\top x_i)^{-1}$ gives

$$M_i = Var(\alpha_i|x_i) + (x_i^\top x_i)^{-1}x_i^\top Var(u_i|x_i)x_i(x_i^\top x_i)^{-1} + E(\alpha_i u_i^\top|x_i)x_i(x_i^\top x_i)^{-1} + (x_i^\top x_i)^{-1}x_i^\top E(u_i \alpha_i^\top|x_i). \quad (A2)$$

Combining (A1) and (A2) we obtain

$$\begin{aligned}
V_\beta &= Var[E(\alpha_i|x_i)] + E(M_i) \\
&= Var[E(\alpha_i|x_i)] + E[Var(\alpha_i|x_i)] + E[(x_i^\top x_i)^{-1}x_i^\top Var(u_i|x_i)x_i(x_i^\top x_i)^{-1}] \\
&\quad + E[E(\alpha_i u_i^\top|x_i)x_i(x_i^\top x_i)^{-1} + (x_i^\top x_i)^{-1}x_i^\top E(u_i \alpha_i^\top|x_i)] \\
&= Var(\alpha_i) + E[(x_i^\top x_i)^{-1}x_i^\top Var(u_i|x_i)x_i(x_i^\top x_i)^{-1}] + E[\alpha_i u_i^\top x_i(x_i^\top x_i)^{-1}] + E[(x_i^\top x_i)^{-1}x_i^\top u_i \alpha_i^\top],
\end{aligned} \quad (A3)$$

which is the same as $Avar(\sqrt{N}\hat{\beta}_{GM}) = \Omega$ given in (12). This completes the proof of Proposition 1(ii). $\square$

**Proof of Proposition 2.** Under the condition of Proposition 2, we have $M_i = (x_i^\top(x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1}x_i)^{-1}$. Then from $M_i M_i^{-1} = I_T$, one can show that $M_i = \Delta + \sigma_u^2(x_i^\top x_i)^{-1}$ (e.g., Hsiao 2003, p. 325)). Hence, $E(M_i) \equiv E\{[x_i^\top(x_i \Delta x_i^\top + \sigma_u^2 I_T)^{-1}x_i]^{-1}\} = \Delta + \sigma_u^2 E[(x_i x_i^\top)^{-1}]$.

Therefore, our asymptotic variance is $Avar(\sqrt{N}\hat{\beta}_{GLS}) = A = [E(M_i^{-1})]^{-1} \leq E[M_i] = Avar(\sqrt{N}\hat{\beta}_{GM})$ by applying diagonalization and utilizing Jansen's inequality to each diagonal element. That is, $\hat{\beta}_{GLS}$ is asymptotically more efficient than $\hat{\beta}_{GM}$. This is because that $\hat{\beta}_{GLS}$ uses additional information (17) while $\hat{\beta}_{GM}$ does not utilize this additional information.

However, if $T$ is large, we have $(x_i^\top x_i)/T = T^{-1}\sum_{t=1}^T x_{it}x_{it}^\top \xrightarrow{p} E(x_{it}x_{it}^\top)$. Hence, $[(x_i^\top x_i)/T]^{-1} = [E(x_{it}x_{it}^\top)]^{-1} + o_p(1)$. We have $\{E[(x_i^\top x_i)/T]^{-1}\} = [E(x_{it}x_{it}^\top) + o(1)]^{-1} = [E(x_{it}x_{it}^\top)]^{-1} + o(1) = E\{[(x_i^\top x_i)/T]^{-1}\} + o(1)$. So that we have $E(C_i^{-1}) = [E(C_i)]^{-1} + o(1)$ (for large $T$), where $C_i = x_i^\top x_i/T$. This explains why when $T \to \infty$, we have $Avar(\sqrt{N}\hat{\beta}_{GLS}) = Avar(\sqrt{N}\hat{\beta}_{GM})$. This gives an intuitive proof for Proposition 3. $\square$

**Proof of Proposition 7.** First, we show that $\tilde{\sigma}_u^2$ and $\hat{\Delta}^*$ are consistent estimators of $\sigma_u^2$ and $E(\Delta^*)$, respectively. Let $\tilde{u}_{it} = y_{it} - x_{it}^\top \hat{\beta}_{GM}$, then it can be shown that

$$
\begin{aligned}
\tilde{\sigma}_{GM}^2 &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \tilde{u}_{it}^2 \\
&= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [x_{it}^\top (\beta_i - \hat{\beta}_{GM}) + u_{it}]^2 \\
&= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [x_{it}^\top (\beta_i - \hat{\beta}_{GM})(\beta_i - \hat{\beta}_{GM})^\top x_{it} + u_{it}^2] + o_p(1) \\
&= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [x_{it}^\top (\beta_i - \beta)(\beta_i - \beta)^\top x_{it} + u_{it}^2] + o_p(1) \quad\quad\text{(A4)} \\
&= (T)^{-1} \sum_{t=1}^T E[x_{it}^\top (\beta_i - \beta)(\beta_i - \beta)^\top x_{it}] + \sigma_u^2 + o_p(1) \\
&= (T)^{-1} \sum_{t=1}^T E\{x_{it}^\top E[(\beta_i - \beta)(\beta_i - \beta)^\top | X] x_{it}\} + \sigma_u^2 + o_p(1) \\
&= (T)^{-1} \sum_{t=1}^T E\{x_{it}^\top \Delta x_{it}\} + \sigma_u^2 + o_p(1)
\end{aligned}
$$

where the first $o_p(1)$ term comes from $E(u_i|x_i) = 0$, and others used $\hat{\beta}_{GM} = \beta + O_p(N^{-1/2})$ and also the law of large numbers.

Furthermore, we have

$$
\begin{aligned}
\tilde{V}_{GM} &= N^{-1} \sum_{i=1}^N (\hat{\beta}_i - \hat{\beta}_{GM})(\hat{\beta}_i - \hat{\beta}_{GM})^\top \\
&= N^{-1} \sum_{i=1}^N [\beta_i + (x_i^\top x_i)^{-1} x_i^\top u_i - \hat{\beta}_{GM}][\beta_i + (x_i^\top x_i)^{-1} x_i^\top u_i - \hat{\beta}_{GM}]^\top \\
&= N^{-1} \sum_{i=1}^N [\beta_i - \beta + (x_i' x_i)^{-1} x_i^\top u_i][\beta_i - \beta + (x_i^\top x_i)^{-1} x_i^\top u_i]^\top + o_p(1) \quad\quad\text{(A5)} \\
&= N^{-1} \sum_{i=1}^N \{(\beta_i - \beta)(\beta_i - \beta)^\top + (x_i^\top x_i)^{-1} x_i^\top u_i^2 x_i (x_i^\top x_i)^{-1}] + o_p(1) \\
&= E[(\beta_i - \beta)(\beta_i - \beta)^\top] + E[(x_i^\top x_i)^{-1} x_i^\top E(u_i^2|X) x_i (x_i^\top x_i)^{-1}] + o_p(1) \\
&= \Delta + \sigma_u^2 E[(x_i^\top x_i)^{-1}] + o_p(1) = Var[E(\alpha_i|x_i)] + E[\Delta^*] + \sigma_u^2 E[(x_i^\top x_i)^{-1}] + o_p(1).
\end{aligned}
$$

Combining (A4) and (A5), we have

$$
\tilde{\sigma}_u^2 \xrightarrow{p} \sigma_u^2, \quad \hat{\Delta}^* \xrightarrow{p} E(\Delta^*).
$$

Next, we look at $\hat{\beta}_{M,PLS}$. We have

$\hat{\beta}_{M,PLS}$

$$= \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} x_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \right.$$

$$\times \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} x_i \left. \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} y_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right.$$

$$\times \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} y_i \left. \right)$$

$$= \beta + \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} x_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \right.$$

$$\times \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} x_i \left. \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} [(x_i \otimes (vec((x_i - E(x_i))^\top))^\top) vec(B^\top)] \right.$$

$$- \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \tag{A6}$$

$$\times \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} [(x_i \otimes (vec((x_i - E(x_i))^\top))^\top) vec(B^\top)] \left. \right)$$

$$+ \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} x_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \right.$$

$$\times \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} x_i \left. \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} v_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right.$$

$$\times \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} v_i \left. \right)$$

$$= \beta + B \, vec(\bar{x}.^\top - E(x_i^\top))$$

$$+ \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} x_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \right.$$

$$\times \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} x_i \left. \right)^{-1} \left( \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} v_i - \sum_{i=1}^{N} x_i^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right.$$

$$\times \left( \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top) \right)^{-1} \sum_{i=1}^{N} (x_i \otimes (vec((x_i - \bar{x}.)^\top))^\top)^\top \hat{\Sigma}_i^{-1} v_i \left. \right).$$

Therefore,

$V_{M,PLS}$

$$= \lim_{N \to \infty} Var(\sqrt{N} \hat{\beta}_{M,GLS}) = Var[E(\alpha_i | x_i)] + \left( E[x_i^\top \Sigma_i^{-1} x_i] - E[x_i^\top \Sigma_i^{-1} (x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \right. \tag{A7}$$

$$\times \left( E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1} (x_i \otimes (vec((x_i - Ex_i)^\top))^\top)] \right)^{-1} E[(x_i \otimes (vec((x_i - Ex_i)^\top))^\top)^\top \Sigma_i^{-1} x_i] \left. \right)^{-1},$$

where $\Sigma_i = x_i \Delta^* x_i^\top + \sigma_u^2 I_T$, $\Delta^* = E(w_i w_i^\top)$ and $w_i = \alpha_i - E(\alpha_i | x_i)$, which implies

$$\sqrt{N}(\hat{\beta}_{M,PLS} - \beta) \xrightarrow{d} N(0, V_{M,PLS}).$$

This completes the proof of Proposition 7.

Moreover, let $M_i^* = (x_i^\top \Sigma_i^{-1} x_i)^{-1} = (x_i^\top (x_i \Delta^* x_i^\top + \sigma_u^2 I_T)^{-1} x_i)^{-1}$. Similar as that in the proof of Proposition 2(i) above, we have

$$M_i^* = \Delta^* + \sigma_u^2 (x_i^\top x_i)^{-1}.$$

□

## References

Arellano, Manuel, and Stéphane Bonhomme. 2012. Identifying distributional characteristics in random coefficient panel data models. *Review of Economic Studies* 79: 987–1020. [CrossRef]

Card, David. 1995. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. Edited by Christofides Loizos Nicolaou, E. Kenneth Grant and Robert Swidinsky. Toronto: University of Toronto Press, pp. 201–22.

Chamberlain, Gary. 1992. Efficiency bounds for semiparametric regression. *Econometrica* 60: 567–96. [CrossRef]

Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey. 2013. Average and quantile effects in nonseparable panel models. *Econometrica* 81: 535–80.

Davidson, James. 1994. *Stochastic Limit Theory: An Introduction for Econometricians*. New York: Oxford University Press.

Fang, Kai-Tai, and Yao-Ting Zhang. 1990. *Generalized Multivariate Analysis*. New York: Springer.

Graham, Bryan S., and James L. Powell. 2012. Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models. *Econometrica* 80: 2105–52.

Gupta, Arjun K., Tamas Varga, and Taras Bodnar. 1993. *Elliptically Contoured Models in Statistics*. Norwell: Kluwer Academic Publishers.

Heckman, James J., Daniel Schmierer, and Sergio Urzua. 2010. Testing the correlated random coefficient model. *Journal of Econometrics* 158: 177–203. [CrossRef] [PubMed]

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88: 389–432. [CrossRef]

Heckman, James, and Edward Vytlacil. 1998. Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources* 33: 974–87. [CrossRef]

Hsiao, Cheng. 1996. Random coefficients models. In *The Econometrics of Panel Data: A Handbook of the Theory with Applications*. Edited by Lśzl Mátyás and Patrick Sevestre. Berlin: Springer, vol. 33, pp. 77–99.

Hsiao, Cheng. 2003. *Analysis of Panel Data*, 2nd ed. Cambridge: Cambridge University Press.

Hsiao, Cheng, Trent W. Appelbe, and Christopher R. Dineen. 1993. A general framework for panel data analysis—With an application to Canadian customer dialed long distance service. *Journal of Econometrics* 59: 63–86. [CrossRef]

Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu. 1999. Bayes estimation of short-run coefficients in dynamic panel data models. In *Analysis of Panels and Limited Dependent Variables Models*. Edited by Cheng Hsiao, M. Hashem Pesaran, Kajal Lahiri and Lung-Fei Lee. Berlin: Springer, pp. 185–213.

Hsiao, Cheng, and M. Hashem Pesaran. 2008. Random coefficient models. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Edited by Lśzl Mátyás and Patrick Sevestre. Berlin: Springer, vol. 46, pp. 185–213.

Hsiao, Cheng, Yan Shen, and Hiroshi Fujiki. 2005. Aggregate versus disaggregate data analysis—A paradox in the estimation of money demand function of Japan under the low interest rate policy. *Journal of Applied Econometrics* 20: 579–601. [CrossRef]

Hsiao, Cheng, and A. Kamil Tahmiscioglu. 1997. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92: 455–65. [CrossRef]

Kuh, Edwin. 1963. *Capital Stock Growth: A Micro-Econometric Approach*. Amsterdam: North-Holland.

Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica* 46: 69–85. [CrossRef]

Pesaran, M. Hashem. 2003. Aggregation of linear dynamic models: An application of life-cycle consumption models under habit formation. *Economic Modelling* 20: 227–435.

Pesaran, M. Hashem, and Ron Smith. 1995. Estimation of long-run relationships from dynamic heterogenous panels. *Journal of Econometrics* 68: 79–114. [CrossRef]

Stoker, Thomas M. 1993. Empirical approaches to the problem of aggregation over individuals. *Journal of Economic Literature* 31: 1827–74.

Swamy, Paravastu AVB. 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38: 311–23. [CrossRef]

Theil, Henri. 1954. *Linear Aggregation of Economic Relations*. Amsterdam: North Holland.

Wooldridge, Jeffrey M. 2005. Fixed-effects related estimators for correlated random-coefficient and treatment-effect panel data models. *The Review of Economics and Statistics* 87: 385–90. [CrossRef]