Estimators Under Correlated Random Effects Models with Cluster Data

Tung Liu*

Professor of Economics

Department of Economics

Ball State University

* Contact Information: Department of Economics, Ball State University, Muncie, IN 47306.
Phone: 765-285-3405. Fax: 765-285-4313. Email: tliu@bsu.edu.

Estimators Under Correlated Random Effects Models with Cluster Data

Abstract

This paper provides a complete analysis on the properties of six basic estimators under correlated random effects models. We examine the relationships of these estimators as in panel data models and extend the results to models with group variables and cluster data. In our analysis, we consider the between and within regressions as two fundamental regressions and assume there are two partial effects of an explanatory variable. Our theoretic results demonstrate some different discussions and new findings. A noteworthy new finding is that the inference of the coefficients of group variables can be made directly with the between regression.

JEL Subject Code: C18, C23, C40

Keywords: Correlated Random Effects Models, Random Effects Model, Within Effects, Fixed Effects, Between Effects, Panel data

Estimators Under Correlated Random Effects Models with Cluster Data

1. Introduction

Correlated random effects (CRE) models are extensions of the random effects model with additional group mean variables and other group variables in the regression (Wooldridge, 2010). One simplified version of CRE models is the well-known Mundlak model (Mundlak, 1978), where only group mean variables are added to the random effects model. In CRE models, there are three sets of key parameters: the coefficients of the explanatory variables, group means of the explanatory variables, and other group variables. To make inferences of these parameters, several estimators are commonly considered, and their relationships have been examined. For models with balanced panel data, two examples of the relationship are commonly known: (1) The pooled ordinary least squares (OLS) estimator of the pooled regression and the random effects estimator are matrix weighted averages of the between and within estimators (Maddala, 1971), and (2) The generalized least squares (GLS) estimator of the coefficients of the explanatory variables in the Mundlak model is the same as the within estimator (Mundlak, 1978). Wooldridge (2019) extended Mundlak's (1978) result to CRE models with unbalanced panel data. But his extension is only limited to the equivalence of the GLS estimator of the coefficients of the explanatory variables and the within estimator. Currently, it is still unknown how the GLS estimators of group mean variables and other group variables are related to the between and within estimators of CRE models.

The choice between the fixed effects and random effects estimators has long been a debate in panel data analysis. Mundlak (1978) used a CRE model with balanced panel data to unify the fixed effects model and the random effects model. His extension of the random effects model to CRE models is an important step in understanding the relationship between the two estimators (Greene, 2018, p. 415; Wooldridge, 2010, p. 286). Furthermore, based on the Mundlak model and CRE models, an alternative to the Hausman test (Hausman, 1978) can be derived for determining between the fixed effects and random effects estimators (Baltagi, 2021; Wooldridge, 2010 & 2019; Greene, 2018). This paper follows Mundlak (1978) and provides a complete analysis on the properties of six estimators under CRE models with cluster data. The six estimators are: the between estimator, the within estimator, the pooled OLS estimator, the random effects estimator, the GLS estimator of CRE models, and the GLS estimator of panel data regression with group variables (Moulton, 1986, 1990). Our analysis of these estimators begins and focuses on the between and within regressions derived from CRE models. Based on these two essential regressions, we use the variances of their random errors and estimators to analyze the properties of different estimators under CRE models. Then we examine the relationships of six estimators. Since classical panel data models are similar to restricted CRE models, our results can be applied to models with balanced and unbalanced panel data.

We consider two types of CRE models. The first type of CRE models is the basic CRE model where the explanatory variables and their group means are included in the regression. This basic model with balanced panel data is the specification in Mundlak (1978). The second type of CRE models extends the basic CRE model with the inclusion of other group variables. The use of the basic CRE model allows us to reexamine the previous results derived from models with balanced panel data (Maddala, 1971; Mundlak, 1978). Mundlak's (1978) main results on the relationships of different estimators are: the GLS estimator of the coefficients of the explanatory variables in the basic CRE model is the same as the within estimator, and the GLS estimator of the coefficients of group mean variables is equal to the difference of the between estimator and the within estimator. His analysis of the results focused on the single partial effect of an explanatory variable. For example, his comparison of the mean squared error (MSE) for different estimators is based on this single partial effect. In his modeling and interpretation, an additional parameter associated with the group mean of the explanatory variable is added only to reflect the statistical correlation between the explanatory variable and unobserved group characteristics, and this correlation is not considered as a partial effect. We, instead, argue that there are two partial effects of an explanatory variable. The first partial effect is the direct impact of the explanatory variable, and the second partial effect is the impact of the group mean of the explanatory variable. Treating the impact of the group mean variable of the explanatory variable as the second partial effect or as a "structural" parameter is well known in the literature of social interactions model (Manski, 1993, 2000; Blume, et al. 2015), where the second partial effect is a measure of network or peer effects. With this different interpretation, we examine the unbiasedness, efficiency, asymptotic and small sample properties of each estimator. We also compare our interpretation of minimum mean squared error (MMSE) estimators with those in Mundlak (1978). Our observations provide a different viewpoint in selecting a better estimator between the fixed effects estimator and other estimators.

For the extended CRE model, we expand the results on the relationship between two estimators (Wooldridge, 2019) to the relationships among six estimators. We check if Mundlak's main results on the relationships of different estimators continue to hold when other group variables are included in the model. In addition, we show how the analysis of the extended CRE model is related to studies in Moulton (1986, 1990), Amemiya (1978), and Donald and Lang (2007). Moulton (1986, 1990) examined the biased issue of OLS standard errors when group variables are included in panel data models. Using a random coefficients model, Amemiya (1978) showed that the coefficients of group variables in a restricted CRE model can be estimated by a two-step procedure. Donald and Lang (2007) used this two-step procedure to address the degrees of freedom issue in inference with panel data models when the number of groups is small. Their study challenged some conclusions and policy implications derived from inference with difference-in-differences models. Using our result on the coefficients of group variables, we revisit Moulton's concern, and the two-step procedure and its implications.

The rest paper is organized as follows. Section 2 reviews basic data matrices for cluster data. Section 3 focuses on the basic CRE model with cluster data. We derive the properties of between and within estimators and examine how the GLS estimator of the basic CRE model, the pooled OLS estimator, and random effects estimator are related to the between and within estimators under the basic CRE model. Then, we compare these five estimators in terms of bias, efficiency, and MSE. We also address the issues in choosing between the fixed effects/within estimator and other estimators. In Section 4, we extend the basic CRE model in Section 3 with additional group variables. In addition to the five estimators under the basic CRE model, we consider an additional estimator: the GLS estimator of the Moulton model (Moulton, 1986, 1990). We show the relationships among six different estimators and discuss the implications with additional group variables. The last section concludes our results and observations.

2. Review of Basic Data Matrices for CRE Models with Cluster Data

Consider regressions with cluster data. Suppose there are n groups and the g^{th} group contains m_g individuals, $g = 1,2,\cdots,n$. The variables for the regression include y, x and z, where y is a scalar dependent variable, x represents k explanatory variables observable at both group and individual levels, and z represents l explanatory group variables observable only at the group level. Both x and z are exogenous variables in the model. With these variables, we consider two types of CRE models. The CRE1 model includes x and their group means; the CRE2 model extends the CRE1 model to include group variables z. The CRE1 model with balanced panel data is the model used in Mundlak (1978). Instead of using the term "the Mundlak model," we use "the CRE1 model" for convenience and for the generalized Mundlak model with cluster data. The two models are

$$\text{CRE1: } y_{gi} = x_{gi}\beta + \bar{x}_g\gamma + \alpha_g + \varepsilon_{gi} \tag{1}$$

$$\text{CRE2: } y_{gi} = x_{gi}\beta + \bar{x}_g\gamma + z_g\xi + \alpha_g + \varepsilon_{gi} \tag{2}$$

where β and γ are $k \times 1$ parameter vectors and ξ is an $l \times 1$ parameter vector. Both α_g and ε_{gi} are unobserved random errors and we denote $u_{gi} = \alpha_g + \varepsilon_{gi}$ as the composite random error. The subscript gi is the index for the ith individual in the gth group, where $g = 1,2,\cdots,n$ and $i = 1,2,\cdots,m_g$. In Equations (1) and (2), y_{gi} is a scalar value of y, x_{gi} is a $1 \times k$ row vector of k explanatory variables, and z_g is a $1 \times l$ row vector of explanatory variables observable only at the group level. The row vector of group mean variables of x is $\bar{x}_g = 1/m_g \sum_i x_{gi}$. This model representation with cluster data is similar to models for panel data with g as the index for the cross-sectional domain and i as the index for the time domain. The CRE1 model with $\gamma = 0$ for panel data is the random effects model. For balanced panel data, $m_g = m$ for all g and m is a constant.

Before we examine the properties of different estimators under CRE1 and CRE2 models, this section reviews basic data matrices for analysis. These matrices for balanced panel data are introduced in Maddala (1978) and most econometrics textbooks (i.e., Hsiao, 2014, pp. 41 - 43; Baltagi, 2021, pp. 15 – 28; Greene, 2018, p. 391, 404 – 408). Here, we construct data matrices for cluster data. The data matrices are constructed by stacking the data of individuals in the same group together like panel data. For example, the data matrix of x_{gi} is

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \sim N \times k, \text{ with } X_g = \begin{pmatrix} x_{g1} \\ x_{g2} \\ \vdots \\ x_{gm_g} \end{pmatrix} \sim m_g \times k, g = 1,2,\cdots n$$

where N is the total number of data values, i.e., $N = \sum_g m_g$. Based on the data matrix X, we introduce the cross-product matrices T_{XX}, B_{XX}, and W_{XX} for the total, between, and within variations of x, respectively. Let I be an N-dimensional identity matrix and e_{m_g} be an $m_g \times 1$

column vector of ones. We follow Mundlak (1978) to define the following two basic matrix operators, $\bar{J}$ and M.

$$\bar{J} = \begin{pmatrix} \bar{J}_1 & 0 & \cdots & 0 \\ 0 & \bar{J}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{J}_n \end{pmatrix}, \bar{J}_g = \frac{1}{m_g} e_{m_g} e'_{m_g}, g = 1,2,\cdots,n$$

$$M = I - \bar{J}$$

Both $\bar{J}$ and M are symmetric and idempotent matrices, and they satisfy

$$\bar{J}' = \bar{J}, \bar{J}\bar{J} = \bar{J}, M' = M, MM = M, \bar{J}M = M\bar{J} = 0$$

The operator $\bar{J}$ is used to define group mean variables and M is used to define variables deviated from their group means. Denote T_{XX} as the cross-product of x_{gi} for the total variation:

$$T_{XX} = X'X = \sum_{g=1}^{n} X'_g X_g = \sum_g \sum_i x'_{gi} x_{gi}$$

For the between variation, we introduce two different data matrices for group mean variables $\bar{x}_g$:

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{pmatrix} \sim n \times k$$

$$\bar{X} = \bar{J}X = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_n \end{pmatrix} \sim N \times k, \text{with } \bar{X}_g = \begin{pmatrix} \bar{x}_g \\ \bar{x}_g \\ \vdots \\ \bar{x}_g \end{pmatrix} \sim m_g \times k, g = 1,2,\cdots,n$$

The row elements of $\bar{X}_g$ are the same; $\bar{x}$ and $\bar{X}$ have different row dimensions. These two data matrices give two different cross-product matrices of $\bar{x}_g$ for the between variation:

$$b_{XX} = \bar{x}'\bar{x} = \sum_{g=1}^{n} \bar{x}'_g \bar{x}_g$$

$$B_{XX} = \bar{X}'\bar{X} = X'\bar{J}X = \bar{X}'\bar{J}X = \bar{X}'\bar{J}\bar{X} = \sum_{g=1}^{n} m_g \bar{x}'_g \bar{x}_g$$

For balanced panel data, $m_g = m$ for all g. Then $B_{XX} = m b_{XX}$.

The deviation of x_{gi} from its group mean $\bar{x}_g$ is defined as $\tilde{x}_{gi} = x_{gi} - \bar{x}_g$. The data matrix of $\tilde{x}_{gi}$ is

$$\tilde{X} = MX = X - \bar{X}; \text{or} \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ . \\ \tilde{X}_n \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ . \\ X_n \end{pmatrix} - \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ . \\ \bar{X}_n \end{pmatrix}, \text{with } \tilde{X}_g = X_g - \bar{X}_g, g = 1,2 \cdots n$$

Then the cross-product of $\tilde{x}_{gi}$ for the within variation is

$$W_{XX} = \tilde{X}'\tilde{X} = X'MX = \sum_{g=1}^{n} \tilde{X}_g'\tilde{X}_g = \sum_g \sum_i \tilde{x}_{gi}'\tilde{x}_{gi}$$

Using $M = I - \bar{J}$,

$$X'MX = X'(I - \bar{J})X = X'X - X'\bar{J}X$$

Rearranging the terms in the equation, it gives the following relationship among T_{XX}, W_{XX} and B_{XX}.

$$T_{XX} = W_{XX} + B_{XX} = W_{XX} + \sum_{g=1}^{n} m_g \bar{x}_g' \bar{x}_g$$

For balanced panel data, $T_{XX} = W_{XX} + mb_{XX}$.

The cross-product matrices with y and the decomposition of total variation of y can be constructed and derived similarly. Denote $\bar{y}_g = 1/m_g \sum_i y_{gi}$ and $\tilde{y}_{gi} = y_{gi} - \bar{y}_g$. We define data matrices: Y, $\bar{y}$, $\bar{Y}$ and $\tilde{Y}$, which have the same data structure as X, $\bar{x}$, $\bar{X}$, and $\tilde{X}$, respectively. Then, $\bar{Y} = \bar{J}Y$, $\tilde{Y} = MY$, $T_{YY} = W_{YY} + B_{YY}$ and $T_{XY} = W_{XY} + B_{XY}$. More specifically, $Y'Y = \tilde{Y}'\tilde{Y} + \bar{Y}'\bar{Y} = \tilde{Y}'\tilde{Y} + \sum_{g=1}^{n} m_g \bar{y}_g' \bar{y}_g$ and $X'Y = \tilde{X}'\tilde{Y} + \bar{X}'\bar{Y} = \tilde{X}'\tilde{Y} + \sum_{g=1}^{n} m_g \bar{x}_g' \bar{y}_g$.

The above introduced data matrices are similar to those in panel data analysis. However, we note two things. First, the individuals are not repeatedly observed in different groups in cluster samples. For panel data, the same individuals are observed in different time periods. This difference cannot be specified and addressed with above data matrices. Second, it is common to use balanced data for panel data analysis. For cluster data, it is common that sample size for each group varies across different groups. In the following sections, the data matrices are assumed to be for cluster data. These matrices can be applied to both balanced and unbalanced panel data.

3. The CRE1 Model with Cluster Data

We focus on the CRE1 model in this section. Mundlak (1978) provided an important analysis on the properties of the GLS estimator and other estimators under the CRE1 model with balanced panel data. The analysis in this section follows Mundalk's (1978). We generalize his analysis to cluster data, which can be applied to unbalanced panel data. In addition to the generalization, we adopt some different interpretations. One major difference is that we consider there are two partial effects of an explanatory variable x. One is the direct effect β and the other is the effect from its group mean, γ. Because of these two partial effects, there are two fundamental regressions to consider: the between and within regressions. We begin our analysis with these two regressions and explain why they are fundamental to analyze different estimators.

3.1 The CRE1 Model: Three Main Regressions and the Variances of Their Random Errors

For the CRE1 model in Equation (1), we consider three main regressions and examine the variances of their random errors. The three main regressions are Equation (1) and the between and within regressions. Here, the fundamental assumption is that the between and within regressions are derived from Equation (1). Hence, our analysis begins with the relationship among these three regressions. Taking the average of the data in each group in Equation (1), it gives the between regression:

$$\bar{y}_g = \bar{x}_g \beta_B + \bar{u}_g \tag{3}$$

where $\beta_B = \beta + \gamma$, $\bar{u}_g = \alpha_g + \bar{\varepsilon}_g$, and $\bar{\varepsilon}_g = 1/m_g \sum_i \varepsilon_{gi}$. Subtracting Equation (3) from Equation (1), it gives the within regression:

$$\tilde{y}_{gi} = \tilde{x}_{gi} \beta + \tilde{\varepsilon}_{gi} \tag{4}$$

where $\tilde{\varepsilon}_{gi} = \varepsilon_{gi} - \bar{\varepsilon}_g$. Note that the between and within regressions have different parameters such that $\beta_B \neq \beta$ when $\gamma \neq 0$. Traditional analysis of partial effects of x focuses only on β. However, we consider γ as the second partial effect of x caused by group means of x. The recognition of the two partial effects β and γ is important when we compare different estimators.

The variance of the random errors in these three main regressions are essential to the properties of different estimators of the key parameters β and γ. We assume that the two basic unobserved random errors ε_{gi} and α_g have zero means and constant variances. Also, random errors for different individuals and groups are uncorrelated. We have

$$E(\varepsilon_{gi}^2) = \sigma_\varepsilon^2; E(\alpha_g^2) = \sigma_\alpha^2 \tag{5}$$

$$E(\varepsilon_{gi}\alpha_{g'}) = 0 \text{ for any } g, g', \text{and } i;$$

$$E(\alpha_g \alpha_{g'}) = 0, \text{for } g \neq g'; E(\varepsilon_{gi}\varepsilon_{gj}) = 0, \text{for } i \neq j \tag{6}$$

Then the variance and covariance of the composite random errors $u_{gi} = \alpha_g + \varepsilon_{gi}$ of Equation (1) are

$$E(u_{gi}\, u_{gj}) = \sigma_\alpha^2 \text{ for } i \neq j \tag{7}$$

$$E(u_{gi}^2) = \sigma_\alpha^2 + \sigma_\varepsilon^2 \tag{8}$$

$$E(u_{gi}\, u_{g'j}) = 0 \text{ for } g \neq g' \tag{9}$$

Write Equation (1) in data matrix form as

$$Y = X\beta + \bar{X}\gamma + u \tag{10}$$

where

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \text{ with } u_g = \begin{pmatrix} u_{g1} \\ u_{g2} \\ \vdots \\ u_{gm_g} \end{pmatrix}, g = 1,2,\cdots,n$$

From Equations (7) – (9), the covariance of u is a block diagonal matrix:

$$\Omega = E(uu') = \begin{pmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & 0\cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \Omega_n \end{pmatrix} = diag(\Omega_1, \Omega_2, \cdots, \Omega_n),$$

$$\text{with } \Omega_g = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdot & \cdot \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \end{pmatrix} = \sigma_\varepsilon^2 I_{m_g} + \sigma_\alpha^2 \, e_{m_g} e'_{m_g} = \sigma_\varepsilon^2 I_{m_g} + m_g \sigma_\alpha^2 \, \bar{J}_g,$$

$$g = 1,2,\cdots n \tag{11}$$

where I_{m_g} is an m_g-dimensional identity matrix. We write Ω as

$$\Omega = \sigma_\varepsilon^2 I + D_{m\sigma_\alpha^2} \bar{J} \tag{12}$$

where $D_{m\sigma_\alpha^2}$ is a diagonal matrix with block feature:

$$D_{m\sigma_\alpha^2} = diag\left(m_1 \sigma_\alpha^2 I_{m_1}, m_2 \sigma_\alpha^2 I_{m_2}, \cdots, m_n \sigma_\alpha^2 I_{m_n}\right)$$

Note that any matrix denoted as D_c in this paper is a diagonal matrix with a specific form. The diagonal elements form in blocks with the form of $c_g I_{m_g}$, where c_g is a constant on diagonals; the constant c_g may vary from one block to another block. Also, D_c satisfies $D_c \bar{J} = \bar{J} D_c = \bar{J} D_c \bar{J}$. When the covariance matrix of random errors is not equal to cI, where c is a constant, the inverse of the covariance matrix of random errors is usually used for the GLS estimator. Using the inverse of a partition matrix, the inverse of Ω_g is (Graybill, 1983, p. 189)

$$\Omega_g^{-1} = \left(\sigma_\varepsilon^2 I_{m_g} + \sigma_\alpha^2 \, e_{m_g} e'_{m_g}\right)^{-1}$$

$$= \pi_2 I_{m_g} + \pi_{1g} e_{m_g} e'_{m_g} = \pi_2 I_{m_g} + m_g \pi_{1g} \bar{J}_g \tag{13}$$

where

$$\pi_2 = \frac{1}{\sigma_\varepsilon^2}, \pi_{1g} = \frac{-\sigma_\alpha^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + m_g \sigma_\alpha^2)}, g = 1,2,\cdots,n \tag{14}$$

Then

$$\Omega^{-1} = \pi_2 I + D_{m\pi_1}\bar{J} \tag{15}$$

with $D_{m\pi_1} = diag(m_1\pi_{11}I_{m_1}, m_2\pi_{12}I_{m_2}, \cdots, m_n\pi_{1n}I_{m_n})$. For balanced panel data, $m_g = m$ and $\pi_{1g} = \pi_1 = \frac{-\sigma_\alpha^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2+m\sigma_\alpha^2)}$ are constants for all g. Then

$$\Omega^{-1} = \pi_2 I + m\pi_1\bar{J} \tag{16}$$

This specification is given in Mundlak (1978).

For the between and within regressions, we derive the variances of their random errors based on the assumptions on the random errors of the CRE1 model. The variance and covariance of the random errors $\bar{u}_g$ of the between regression in Equation (3) are

$$E(\bar{u}_g\bar{u}_g) = \sigma_{\bar{u}_g}^2, \sigma_{\bar{u}_g}^2 = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g} \tag{17}$$

$$E(\bar{u}_g\bar{u}_{g'}) = 0, \text{ for } g \neq g' \tag{18}$$

The matrix form of Equation (3) with $\bar{y}$ and $\bar{x}$ is

$$\bar{y} = \bar{x}\beta_B + \bar{u} \tag{19}$$

where $\bar{u} = (\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_n)'$. The covariance matrix of the random error $\bar{u}$ is

$$\Omega_{\bar{u}} = E(\bar{u}\bar{u}') = d_{\sigma_{\bar{u}}^2}, d_{\sigma_{\bar{u}}^2} = diag\left(\sigma_{\bar{u}_1}^2, \sigma_{\bar{u}_2}^2, \cdots, \sigma_{\bar{u}_n}^2\right) \tag{20}$$

This shows that $\bar{u}$ is heteroscedastic. Then the inverse of $\Omega_{\bar{u}}$ is

$$\Omega_{\bar{u}}^{-1} = d_{\sigma_{\bar{u}}^{2}{}^{-1}} = diag\left({\sigma_{\bar{u}_1}^2}^{-1}, {\sigma_{\bar{u}_2}^2}^{-1}, \cdots, {\sigma_{\bar{u}_n}^2}^{-1}\right) \tag{21}$$

Because the data matrix $\bar{X}$ instead of $\bar{x}$ is involved in the matrix form of the CRE1 model in Equation (10), we consider the matrix form of the between regression in Equation (3) with $\bar{Y}$ and $\bar{X}$ as follows.

$$\bar{Y} = \bar{X}\beta_B + \bar{U} \tag{22}$$

where $\bar{U} = (\bar{u}_1 e_{m_1}', \bar{u}_2 e_{m_2}', \cdots, \bar{u}_n e_{m_n}')'$. From Equations (17) and (18), the covariance matrix of the random errors $\bar{U}$ is

$$\Omega_{\bar{U}} = E(\bar{U}\bar{U}') = diag\left(\Omega_{\bar{U}_1}, \Omega_{\bar{U}_2}, \cdots, \Omega_{\bar{U}_n}\right), \tag{23}$$

with $\Omega_{\bar{U}_g} = \begin{pmatrix} \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g} & \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g} & . & . \\ \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g} & \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g} & & \\ . & & . & \\ . & & & . \end{pmatrix} = \left(\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g}\right) e_{m_g} e_{m_g}' = m_g \sigma_{\bar{u}_g}^2 \bar{J}_g, g = 1,2,\cdots n$

Or

$$\Omega_{\bar{U}} = D_{m\sigma_{\bar{u}}^2}\bar{J} \tag{24}$$

with $D_{m\sigma_{\bar{u}}^2} = diag(m_1\sigma_{\bar{u}_1}^2 I_{m_1}, m_2\sigma_{\bar{u}_2}^2 I_{m_2}, \cdots, m_n\sigma_{\bar{u}_n}^2 I_{m_n})$. Since the rank of $\Omega_{\bar{U}_g}$ is 1, the rank of $\Omega_{\bar{U}}$ is n. $\Omega_{\bar{U}_g}$ is not full rank and is singular. $\Omega_{\bar{U}}^{-1}$ does not exist. However, we can define the inverse of $\Omega_{\bar{U}}$ as

$$\Omega_{\bar{U}}^{-1} = D_{(m\sigma_{\bar{u}}^2)^{-1}}\bar{J} \tag{25}$$

with $D_{(m\sigma_{\bar{u}}^2)^{-1}} = diag\left(\left(m_1\sigma_{\bar{u}_1}^2\right)^{-1} I_{m_1}, \left(m_2\sigma_{\bar{u}_2}^2\right)^{-1} I_{m_2}, \cdots, \left(m_n\sigma_{\bar{u}_n}^2\right)^{-1} I_{m_n}\right)$. Using the formulas of π_1 and π_2 in Equation (14), it gives

$$m_g\pi_{1g} = \frac{-m_g\sigma_\alpha^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + m_g\sigma_\alpha^2)} = \frac{-\sigma_\alpha^2}{\sigma_\varepsilon^2\left(\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g}\right)} \tag{26}$$

$$\pi_2 + m_g\pi_{1g} = \frac{1}{\sigma_\varepsilon^2} + \frac{-\sigma_\alpha^2}{\sigma_\varepsilon^2\left(\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g}\right)} = \frac{1}{m_g}\left(\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m_g}\right)^{-1} = \left(m_g\sigma_{\bar{u}}^2\right)^{-1} \tag{27}$$

Then $\Omega_{\bar{U}}^{-1}$ can be rewritten as

$$\Omega_{\bar{U}}^{-1} = D_{\pi_2 + m\pi_1}\bar{J} \tag{28}$$

with $D_{\pi_2+m\pi_1} = diag\left((\pi_2 + m_1\pi_{11})I_{m_1}, (\pi_2 + m_2\pi_{12})I_{m_2}, \cdots, (\pi_2 + m_n\pi_{1n})I_{m_n}\right)$. This alternative form of $\Omega_{\bar{U}}^{-1}$ is useful when examining different estimators under the CRE1 model.

For balanced panel data, $m_g = m$, $\pi_{1g} = \pi_1$, and $\sigma_{\bar{u}_g}^2 = \sigma_{\bar{u}}^2 = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m}$ are constants. Then

$$\Omega_{\bar{u}} = \sigma_{\bar{u}}^2 I \text{ and } \Omega_{\bar{u}}^{-1} = \sigma_{\bar{u}}^{2^{-1}} I \tag{29}$$

$$\Omega_{\bar{U}} = m\sigma_{\bar{u}}^2\bar{J} \text{ and } \Omega_{\bar{U}}^{-1} = (m\sigma_{\bar{u}}^2)^{-1}\bar{J} = (\pi_2 + m\pi_1)\bar{J} \tag{30}$$

Equation (29) shows the random errors $\bar{u}_g$ in $\bar{u}$ are homoscedastic and serial uncorrelated. However, the random errors $\bar{u}_g$ in $\bar{U}$ are still serial correlated due to the repetition of $\bar{u}_g$ in the same group of $\bar{U}$.

For the within regression in Equation (4), the variance and covariance of $\tilde{\varepsilon}_{gi}$ are

$$E\left(\tilde{\varepsilon}_{gi}^2\right) = \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{m_g} \tag{31}$$

$$E\left(\tilde{\varepsilon}_{gi}\tilde{\varepsilon}_{gj}\right) = -\frac{\sigma_\varepsilon^2}{m_g}, i \neq j \tag{32}$$

$$E\left(\tilde{\varepsilon}_{gi}\tilde{\varepsilon}_{g'j}\right) = 0, g \neq g' \tag{33}$$

The matrix form of the within regression is

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon} \tag{34}$$

where

$$\tilde{\varepsilon} = \begin{pmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{pmatrix}, \text{with } \tilde{\varepsilon}_g = \begin{pmatrix} \tilde{\varepsilon}_{g1} \\ \tilde{\varepsilon}_{g2} \\ \vdots \\ \tilde{\varepsilon}_{gm_g} \end{pmatrix} = \begin{pmatrix} \varepsilon_{g1} - \bar{\varepsilon}_g \\ \varepsilon_{g2} - \bar{\varepsilon}_g \\ \vdots \\ \varepsilon_{gm_g} - \bar{\varepsilon}_g \end{pmatrix} = \varepsilon_g - \bar{\varepsilon}_g e_{m_g}, g = 1,2,\cdots,n$$

From Equations (31) – (33), the covariance matrix of $\tilde{\varepsilon}$ is

$$\Omega_{\tilde{\varepsilon}} = E(\tilde{\varepsilon}\tilde{\varepsilon}') = diag\left(\Omega_{\tilde{\varepsilon}_1}, \Omega_{\tilde{\varepsilon}_2}, \cdots, \Omega_{\tilde{\varepsilon}_n}\right),$$

$$\text{with } \Omega_{\tilde{\varepsilon}_g} = \begin{pmatrix} \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{m_g} & -\frac{\sigma_\varepsilon^2}{m_g} & . & . \\ -\frac{\sigma_\varepsilon^2}{m_g} & \sigma_\varepsilon^2 - \frac{\sigma_\varepsilon^2}{m_g} & & \\ . & & . & \\ . & & & . \end{pmatrix} = \sigma_\varepsilon^2\left(I_{m_g} - \frac{1}{m_g} e_{m_g} e'_{m_g}\right) = \sigma_\varepsilon^2\left(I_{m_g} - \bar{J}_g\right),$$

$$g = 1,2,\cdots n \tag{35}$$

Or,

$$\Omega_{\tilde{\varepsilon}} = \sigma_\varepsilon^2(I - \bar{J}) = \sigma_\varepsilon^2 M \tag{36}$$

The rank of $\Omega_{\tilde{\varepsilon}}$ and M is $N - n$. M is singular and $\Omega_{\tilde{\varepsilon}}^{-1}$ does not exist. Also, $\Omega_{\tilde{\varepsilon}_g}$ reflects that $\tilde{\varepsilon}_{gi}$ is heteroscedastic and serial correlated. However, we can define the inverse of $\Omega_{\tilde{\varepsilon}}$ as

$$\Omega_{\tilde{\varepsilon}}^{-1} = \sigma_\varepsilon^{2^{-1}} M = \pi_2 M \tag{37}$$

It is still a singular matrix with a rank of $N - n$.

After we introduce the covariance matrices of the random errors of the three main regressions, we check the relationship of these covariance matrices. Without using the above derived covariance matrices, a simple way to check the relationship is to rewrite the composite random errors u_g in Equation (10) as

$$u_g = u_g - \overline{U}_g + \overline{U}_g = \tilde{\varepsilon}_g + \overline{U}_g \tag{38}$$

Denote $C\left(\tilde{\varepsilon}_{gi}, \bar{u}_g\right)$ as the covariance of $\tilde{\varepsilon}_{gi}$ and $\bar{u}_g$. Then $C\left(\tilde{\varepsilon}_{gi}, \bar{u}_g\right) = E\left(\left(\varepsilon_{gi} - \bar{\varepsilon}_g\right)\left(\alpha_g + \bar{\varepsilon}_g\right)\right) = E\left(\left(\varepsilon_{gi} - \bar{\varepsilon}_g\right)\left(\bar{\varepsilon}_g\right)\right) = 0$; the random errors of the within and between regressions are uncorrelated. Hence, the variance of u_g in Equation (38) is

$$V\left(u_g\right) = V\left(\tilde{\varepsilon}_g\right) + V(\overline{U}_g) \tag{39}$$

The variance of the random errors of the CRE1 model is the sum of the variances of the random errors of the within and between regressions. This relationship can be verified with Equations (11), (35), and (23), the covariance matrices Ω_g, $\Omega_{\tilde{\varepsilon}_g}$, and $\Omega_{\bar{U}_g}$, respectively, as follows.

$$\Omega_g = \Omega_{\tilde{\varepsilon}_g} + \Omega_{\bar{U}_g} \text{ and } \Omega = \Omega_{\tilde{\varepsilon}} + \Omega_{\bar{U}} \qquad (40)$$

The definitions of the inverses of $\Omega_{\tilde{\varepsilon}}$ and $\Omega_{\bar{U}}$ in Equations (37) and (28), respectively, give

$$\Omega_{\tilde{\varepsilon}}^{-1} = \pi_2 I - \pi_2 \bar{J}$$

$$\Omega_{\bar{U}}^{-1} = \pi_2 \bar{J} + D_{m\pi_1} \bar{J}$$

Using Ω^{-1} in Equation (15) and the above equations, it shows

$$\Omega^{-1} = \Omega_{\tilde{\varepsilon}}^{-1} + \Omega_{\bar{U}}^{-1} \qquad (41)$$

Note that, instead of $\bar{u}$, $\bar{U}$ is involved to derive Equations (39) – (41). This shows the need for the between regression $\bar{Y} = \bar{X}\beta + \bar{U}$ in describing the relationship among the three main regressions.

Now we discuss the properties of the four covariance matrices of random errors when $m_g \to \infty$. Comparing these covariance matrices, the elements of Ω in Equation (11) does not contain m_g while other three matrices $\Omega_{\bar{u}}$, $\Omega_{\bar{U}}$ and $\Omega_{\tilde{\varepsilon}}$ in Equations (20), (23), and (35), respectively, involve m_g. When $m_g \to \infty$, both $\Omega_{\bar{u}}$ and $\Omega_{\tilde{\varepsilon}}$ become diagonal matrices, and random errors $\bar{u}_g$ and $\tilde{\varepsilon}_{gi}$ are homoscedastic and uncorrelated. But Ω and $\Omega_{\bar{U}}$ are still nondiagonal with serial correlated random errors. Ω stays nondiagonal since $\Omega_{\bar{U}}$ is nondiagonal. We note two issues with the inverse of these two nondiagonal matrices when $m_g \to \infty$. The first issue is related to Ω^{-1}. When $m_g \to \infty$, $\pi_{1g} = 0$ in Equation (14) and $m_g \pi_{1g} = -\pi_2$ in Equation (26). Consider the first part of Equation (13) for Ω_g^{-1}. Given $\pi_{1g} = 0$,

$$\Omega_g^{-1} = \pi_2 I_{m_g} + \pi_{1g} e_{m_g} e'_{m_g} = \pi_2 I_{mg} \text{ and } \Omega^{-1} = \pi_2 I \qquad (42)$$

If $\Omega^{-1} = \pi_2 I$, then $\Omega = \sigma_\varepsilon^2 I$. This is inconsistent with that Ω is nondiagonal and u_{gi} are serial correlated. Consider the second part of Equation (13) for Ω_g^{-1}. Given $m_g \pi_{1g} = -\pi_2$,

$$\Omega_g^{-1} = \pi_2 I_{m_g} + m_g \pi_{1g} \bar{J}_g = \pi_2 I_{mg} - \pi_2 \bar{J}_g \text{ and } \Omega^{-1} = \pi_2 M \qquad (43)$$

The rank of Ω^{-1} becomes $N - n$ and is nonsingular. This is inconsistent with that Ω and Ω^{-1} are nonsingular.

The second issue is related to $\Omega_{\bar{U}}^{-1}$. When $m_g \to \infty$, $\Omega_{\bar{U}}^{-1} = D_{(m\sigma_{\bar{u}}^2)^{-1}} \bar{J} = 0$ in Equation (25). It appears that the definition of $\Omega_{\bar{U}}^{-1}$ is problematic. These two issues are related to $\bar{J}$ in Ω^{-1}

and $\Omega_{\bar{U}}^{-1}$. These issues will be explained and resolved when we discuss Ω^{-1} and $\Omega_{\bar{U}}^{-1}$ in the GLS estimation of the CRE1 model and the between regression.

We also need to pay special attention to the estimation of the variances of the random errors when $m_g \to \infty$. Since σ_α^2 and σ_ε^2 in the covariance matrices Ω, $\Omega_{\bar{u}}$, and $\Omega_{\tilde{\varepsilon}}$ are unknown, the estimation of these two parameters involves the use of regression residuals and sample sizes m_g and n. When we consider asymptotic properties of the estimators of σ_α^2 and σ_ε^2, we should separate $m_g \to \infty$ from $n \to \infty$. The estimators of $\Omega_{\bar{u}}$ and $\Omega_{\tilde{\varepsilon}}$ become simple when $m_g \to \infty$ since these covariance matrices are diagonal when $m_g \to \infty$. In using consistent estimators, we conveniently simplify heteroscedasticity and serial correlation issues by assuming $m_g \to \infty$. However, the inference of these estimators is more complicated when m_g is finite. We will address this issue when we discuss the estimation of Ω, $\Omega_{\bar{u}}$, and $\Omega_{\tilde{\varepsilon}}$ in the next subsection.

We conclude this subsection with two observations. First, the serial correlations of u_{gi} in the CRE1 model, $\bar{u}_g$ in the between regression $\bar{Y} = \bar{X}\beta_B + \bar{U}$, and $\tilde{\varepsilon}_{gi}$ in the within regression are caused by the repetition of the random errors α_g and ε_{gi} in the same group, even though each of these two key random errors are assumed to be uncorrelated, such that $E(\alpha_g \alpha_{g'}) = E(\varepsilon_{gi}\varepsilon_{g'j}) = 0$ for $g \neq g'$. The use of GLS estimation for CRE models is necessary, mainly because of the repetition of ε_{gi} and α_g in the same group. This implies that the GLS estimation of CRE models can be replaced by the between and within estimations. Second, the above analysis of random errors of the three main regressions under the CRE1 model can be applied to the CRE2 model and classic panel models. The properties and fundamental structures of random errors of the three main regressions are useful and important for panel data analysis as well.

3.2. The CRE1 Model: Five Different Estimators and Their Mean and Variances

Based on the above analysis of the between and within regressions and the random errors of the three main regressions, we discuss five different estimators in estimating the parameters of the CRE1 model in this subsection. We begin with two essential estimators: the between estimator $\hat{\beta}_B$ and the within estimator $\hat{\beta}_w$. Then we examine the rest three estimators – the GLS estimator of the CRE1 model, the pooled OLS estimator of the pooled regression, $\hat{\beta}_{OLS}$, and the random effects estimator $\hat{\beta}_{RE}$. We show how these three estimators are related to the two essential estimators under the CRE1 model.

3.2.1 The Between and Within Estimators

The between and within estimators are the two essential estimators in panel data analysis. These estimators are derived from the between and within regressions. Two different matrix forms of the between regression, Equations (19) and (22), give two forms of between estimators. The OLS estimator of β_B in $\bar{y} = \bar{x}\beta_B + \bar{u}$ is denoted as $\hat{\beta}_{B1O_n}$ and it is given as

$$\hat{\beta}_{B1O_n} = (\bar{x}'\bar{x})^{-1}\bar{x}'\bar{y} \quad (44)$$

Substituting $\bar{y} = \bar{x}\beta_B + \bar{u}$ into $\hat{\beta}_{B1O_n}$, the mean and the variance of $\hat{\beta}_{B1O_n}$ are

$$E(\hat{\beta}_{B1O_n}) = \beta_B = \beta + \gamma$$

$$V(\hat{\beta}_{B1O_n}) = (\bar{x}'\bar{x})^{-1}\bar{x}'d_{\sigma^2_{\bar{u}}}\bar{x}(\bar{x}'\bar{x})^{-1}$$

Since x is exogenous, such that $E(\varepsilon_{gi}|x) = E(\alpha_g|x) = 0$, $\hat{\beta}_{B1O_n}$ is an unbiased estimator of β_B. Because of heteroscedasticity of $\bar{u}_g$, $\hat{\beta}_{B1O_n}$ is not an efficient estimator. Given $\Omega_{\bar{u}}^{-1} = d_{(\sigma^2_{\bar{u}})^{-1}}$, the efficient estimator of β_B is the GLS estimator or the weighted least squared estimator:

$$\hat{\beta}_{B1G_n} = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{y} = \left(\bar{x}'d_{\sigma^{2}_{\bar{u}}{}^{-1}}\bar{x}\right)^{-1}\bar{x}'d_{\sigma^{2}_{\bar{u}}{}^{-1}}\bar{y} \quad (45)$$

Substituting $\bar{y} = \bar{x}\beta_B + \bar{u}$ into the above equation, the mean and variance of $\hat{\beta}_{B1G_n}$ are

$$E(\hat{\beta}_{B1G_n}) = \beta_B$$

$$V(\hat{\beta}_{B1G_n}) = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1} = \left(\bar{x}'d_{\sigma^{2}_{\bar{u}}{}^{-1}}\bar{x}\right)^{-1}$$

For the unbiasedness of $\hat{\beta}_{B1G_n}$, we assume x is exogenous and the group size m_g is uncorrelated with x (Green, 2018, p. 305). It gives $E((\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{u}_g) = 0$ and $E(\hat{\beta}_{B1G_n}) = \beta_B$.

With the data of group mean variables $\bar{y}_g$ and $\bar{x}_g$, it seems that there is no need to consider the between regression $\bar{Y} = \bar{X}\beta_B + \bar{U}$ empirically. However, the estimator of β_B in $\bar{Y} = \bar{X}\beta_B + \bar{U}$ is involved with other estimators under the CRE1 model. The OLS estimator of β_B in $\bar{Y} = \bar{X}\beta_B + \bar{U}$ is

$$\hat{\beta}_{B1O_N} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y} = \left(\sum m_g\bar{x}_g'\bar{x}_g\right)^{-1}\sum m_g\bar{x}_g'\bar{y}_g \quad (46)$$

When $m_1, m_2, \cdots$, and m_n are not the same, i.e., $m_g \neq m_{g'}$ for some $g \neq g'$, the between estimators for two forms of between regressions are different, i.e., $\hat{\beta}_{B1O_n} \neq \hat{\beta}_{B1O_N}$. Since u_g in $\bar{U}$ are heteroscedastic and serial correlated as shown by $\Omega_{\bar{U}} = D_{m\sigma^2_{\bar{u}}}\bar{J}$, $\hat{\beta}_{B1O_N}$ is not efficient. Although $\Omega_{\bar{U}}$ is singular, we apply $\Omega_{\bar{U}}^{-1} = D_{(m\sigma^2_{\bar{u}})^{-1}}\bar{J}$ to define the GLS estimator of β_B in $\bar{Y} = \bar{X}\beta_B + \bar{U}$ as:

$$\hat{\beta}_{B1G_N} = \left(\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X}\right)^{-1}\bar{X}'\Omega_{\bar{U}}^{-1}\bar{Y} \tag{47}$$

We show the matrix equivalence of $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ as follows.

$$\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X} = \bar{X}'D_{(m\sigma^2_{\bar{u}})^{-1}}\bar{J}\bar{X} = \textstyle\sum_g \left(m_g\sigma^2_{\bar{u}_g}\right)^{-1}\bar{X}_g'\bar{X}_g = \sum_g m_g^{-1}\sigma^2_{\bar{u}_g}{}^{-1}m_g\bar{x}_g'\bar{x}_g$$

$$= \textstyle\sum_g \sigma^2_{\bar{u}_g}{}^{-1}\bar{x}_g'\bar{x}_g = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} \tag{48}$$

Similarly, $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{Y} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$. Hence, $\beta_{B1G_n} = \beta_{B1G_N}$. This implies that the GLS estimators for two different matrix forms of the between regression are the same. We denote $\hat{\beta}_{B1G} = \beta_{B1G_n} = \beta_{B1G_N}$ as the GLS estimator of the between regression for the CRE1 model. We have

$$\hat{\beta}_{BIG} = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{y} \tag{49}$$

$$E\left(\hat{\beta}_{B1G}\right) = \beta_B \tag{50}$$

$$V(\hat{\beta}_{B1G}) = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1} \tag{51}$$

$\hat{\beta}_{B1G}$ is an MMSE estimator of β_B. It is important to consider the GLS between estimator when cluster data is used.

In the previous subsection, we noted the issue with the definition of $\Omega_{\bar{U}}^{-1} = D_{(m\sigma^2_{\bar{u}})^{-1}}\bar{J}$, where $\Omega_{\bar{U}}^{-1} = 0$ when $m_g \to \infty$. The above use of $\Omega_{\bar{U}}^{-1}$ in the GLS between estimator shows that this is not an issue. In the two matrix equivalences of $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X}$ and $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{Y}$, the term m_g^{-1} in $\Omega_{\bar{U}}^{-1}$ cancels out with the m_g generated from the conversion from $\bar{X}'\bar{J}\bar{X}$ to $\sum_g m_g\bar{x}_g'\bar{x}_g$ or from $\bar{X}'\bar{J}\bar{Y}$ to $\sum_g m_g\bar{x}_g'\bar{y}_g$ as shown in Equation (48). Hence, $m_g \to \infty$ has no impact in applying $\Omega_{\bar{U}}^{-1}$ to the GLS between estimator, except that $\sigma^2_{\bar{u}_g}{}^{-1} \to \sigma^2_{\alpha}{}^{-1}$ when $m_g \to \infty$.

For balanced panel data, $m_g = m$ for all g. The GLS between estimator is the same as the OLS between estimator, i.e., $\hat{\beta}_{B1O_n} = \hat{\beta}_{B1O_N} = \hat{\beta}_{B1G}$. Then $\hat{\beta}_{B1O_n}$ is an unbiased and efficient estimator of β_B. We consider asymptotic properties of the variance of the between estimator when $m \to \infty$ and $n \to \infty$. The variance of the between estimator for balanced panel data is

$$V(\hat{\beta}_{B1O_n}) = \sigma_{\bar{u}}^2(\bar{x}'\bar{x})^{-1} = \left(\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m}\right)(\bar{x}'\bar{x})^{-1}$$

When $m \to \infty$, $Var(\hat{\beta}_{B1O_n}) = \sigma_\alpha^2(\bar{x}'\bar{x})^{-1}$. To consider $n \to \infty$, we rewrite the variance of $\hat{\beta}_{B1O_n}$ as

$$V(\hat{\beta}_{B1O_n}) = \frac{\sigma_{\bar{u}}^2}{n}\left(\frac{\bar{x}'\bar{x}}{n}\right)^{-1}$$

where $\frac{\sigma_{\bar{u}}^2}{n}$ is the variance of the sample mean of $\bar{u}_g$ and $\frac{\bar{x}'\bar{x}}{n}$ is the covariance of $\bar{x}$. It shows that the variance of the OLS between estimator is the noise-to-signal ratio of regressing $\bar{y}$ on $\bar{x}$, where the noise is determined by $\sigma_{\bar{u}}^2$ and the signal is from $\bar{x}_g$. The larger the noise, the larger the variance of $\hat{\beta}_{B1O_n}$. Since the noise $\sigma_{\bar{u}}^2$ is unknown, it is usually estimated by $\hat{\sigma}_{\bar{u}}^2 = \frac{1}{n-k}\sum\hat{\bar{u}}_g^2$, where $\hat{\bar{u}}_g$ is the residual term from the between regression $\bar{y} = \bar{x}\beta_B + \bar{u}$. When $n \to \infty$, $\frac{\bar{x}'\bar{x}}{n}$ is assumed to converge to a positive definite matrix and the Central Limit Theory (CLT) is applied to the inference of β_B with $\hat{\beta}_{B1O_n}$. For small samples, the inference of β_B depends on the sample sizes m and n. When m is small, $\sigma_{\bar{u}}^2 = \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{m}$ is affected by σ_ε^2 more than the case when m is large. The number of the degrees of freedom in inference of β_B is $n - k$ and it is not related to m; the CLT cannot be applied even if $m \to \infty$.

For the within regression $\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$, the OLS within estimator is

$$\hat{\beta}_w = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \quad (52)$$

Substituting $\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$ into $\hat{\beta}_W$, and using $E(\tilde{\varepsilon}\tilde{\varepsilon}') = \sigma_\varepsilon^2 M$, the mean and variance of the within estimator $\hat{\beta}_W$ are:

$$E(\hat{\beta}_w) = \beta \quad (53)$$

$$V(\hat{\beta}_w) = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'E(\tilde{\varepsilon}\tilde{\varepsilon}')\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}$$
$$= \sigma_\varepsilon^2\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'M\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}$$
$$= \sigma_\varepsilon^2\left(\tilde{X}'\tilde{X}\right)^{-1} \quad (54)$$

Similar to the issue with $\Omega_{\bar{U}}$ in analyzing the between regression $\bar{Y} = \bar{X}\beta_B + \bar{U}$, $\Omega_{\tilde{\varepsilon}}$ is singular and $\tilde{\varepsilon}_{gi}$ are heteroscedastic and serial correlated. However, the variance of $\hat{\beta}_W$ shows that $\hat{\beta}_W$ is an efficient estimator of β despite heteroscedasticity and serial correlation of $\tilde{\varepsilon}$. It is interesting to examine the elements of $\Omega_{\tilde{\varepsilon}_g}$ in Equation (35). The term $-\frac{\sigma_\varepsilon^2}{m_g}$ is observed in all elements of $\Omega_{\tilde{\varepsilon}_g}$.

It seems that the negative correlation between $\tilde{\varepsilon}_{gi}$ and $\tilde{\varepsilon}_{gj}$ within the same group g, i.e., $E(\tilde{\varepsilon}_{gi}\tilde{\varepsilon}_{gj}) = -\frac{\sigma_{\varepsilon}^2}{m_g}, i \neq j$, causes serial correlation issue. In fact, this negative correlation is important since it ensures the existence of the M matrix in $\Omega_{\tilde{\varepsilon}}$ and then ensures $\hat{\beta}_W$ to be an efficient estimator. Hence, $\hat{\beta}_W$ is an MMSE estimator of β for balanced and unbalanced panel data, and for any finite or infinite values of m_g.

For asymptotic properties of the variance of $\hat{\beta}_w$, we consider $m_g \to \infty$ and $n \to \infty$. Rewrite the variance of $\hat{\beta}_W$ as

$$V(\hat{\beta}_W) = \frac{\sigma_{\varepsilon}^2}{N}\left(\frac{\tilde{X}'\tilde{X}}{N}\right)^{-1}$$

where $\frac{\sigma_{\varepsilon}^2}{N}$ is the variance of the sample mean of ε_{gi} and $\frac{\tilde{X}'\tilde{X}}{N}$ is the covariance of $\tilde{X}$. This form of variance is similar to that of the variance of the between estimator. The variance of the within estimator is the noise-to-signal ratio of the within regression, where the noise is determined by σ_{ε}^2 and the signal is from $\tilde{x}_{gi}$. The larger the noise, the larger the variance of $\hat{\beta}_W$. The unknown σ_{ε}^2 is estimated by $\hat{\sigma}_{\varepsilon}^2 = \frac{1}{N-n-k}\tilde{e}'\tilde{e}$, where $\tilde{e} = \tilde{Y} - \tilde{X}\hat{\beta}_W$ (Greene, 2018, p. 395). When $m_g \to \infty$, $\Omega_{\tilde{\varepsilon}}$ is diagonal and the serial correlation issue disappears (Wooldridge, 2010, p. 305). If either $m_g \to \infty$ or $n \to \infty$, then $N \to \infty$ and the CLT is applied to the inference of β with $\hat{\beta}_W$. Small sample inferences of β with $\hat{\beta}_W$ are applied only when both m_g and n are small, and the number of the degrees of freedom is $N - n - k$.

We show that the covariance of the GLS between estimator and the within estimator is zero, i.e., $C(\hat{\beta}_{B1G}, \hat{\beta}_W) = 0$. Using $\hat{\beta}_{B1G} = \beta_B + (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{u}$ and $\hat{\beta}_W = \beta + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\varepsilon}$, the covariance of $\hat{\beta}_{B1G}$ and $\hat{\beta}_W$ is

$$C(\hat{\beta}_{B1G}, \hat{\beta}_W) = E\left((\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{u}\tilde{\varepsilon}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\right)$$

$$= (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'d_{\sigma_{\bar{u}}^{2}{}^{-1}}E(\bar{u}\tilde{\varepsilon}')\tilde{X}(\tilde{X}'\tilde{X})^{-1} = 0 \quad (55)$$

since $E(\bar{u}_g\tilde{\varepsilon}_{gi}) = C(\bar{u}_g, \tilde{\varepsilon}_{gi}) = 0$ as shown in Equation (39).

We summarize three conclusions about the two essential estimators. First, the between and within estimators are unbiased estimators of two different parameters when $\gamma \neq 0$. Both GLS between estimator and within estimator are efficient estimators. Hence, these are two different MMSE estimators. Second, each between and within estimator has its own noise-to-signal ratio. Third, the between and within estimators have different numbers of degrees of freedom in small sample inferences. These differences in noise-to-signal ratios and degrees of freedom affect the properties of the three remaining estimators under the CRE1 model since these three estimators are related to the between and within estimators.

3.2.2 The GLS Estimator of the CRE1 Model

After introducing the two essential estimators – the between and within estimators – we show how other estimators are related to these two estimators. We begin with the GLS estimator of the CRE1 model. The relationship between this GLS estimator and ($\hat{\beta}_{B1O_n}$, $\hat{\beta}_W$) has been derived for balanced panel data in Mundlak (1978). This paper focuses on the relationship when cluster data is used. The GLS estimator of (β, γ) in the CRE1 model $Y = X\beta + \bar{X}\gamma + u$ is denoted as $\left(\hat{\beta}_{C1}, \hat{\gamma}_{C1}\right)$, and

$$\begin{pmatrix} \hat{\beta}_{C1} \\ \hat{\gamma}_{C1} \end{pmatrix} = \left(\begin{pmatrix} X' \\ \bar{X}' \end{pmatrix} \Omega^{-1} (X \quad \bar{X}) \right)^{-1} \begin{pmatrix} X' \\ \bar{X}' \end{pmatrix} \Omega^{-1} Y \tag{56}$$

We show the following theorem for the relationship between this GLS estimator and two essential estimators $\hat{\beta}_W$ and $\hat{\beta}_{B1G}$ of Equations (52) and (49), respectively.

Theorem 1. Let $\left(\hat{\beta}_{C1}, \hat{\gamma}_{C1}\right)$ be the GLS estimator of the parameters in the CRE1 model, and $\hat{\beta}_W$ and $\hat{\beta}_{B1G}$ be the within estimator and the GLS between estimator of the CRE1 model, respectively. Then

$$\hat{\beta}_{C1} = \hat{\beta}_W \text{ and } \hat{\gamma}_{C1} = \hat{\beta}_{B1G} - \hat{\beta}_W \tag{57}$$

Mundlak (1978) proved the above equalities for the CRE1 model with balanced panel data using the inverse of a partition matrix while the proof by Baltagi (2006) used system estimation with the between and within regressions. Chamberlian (1980, p. 234) and Wooldridge (2019) proved the first equality $\hat{\beta}_{C1} = \hat{\beta}_W$ using the Frisch-Waugh theorem (Frisch and Waugh, 1933); both Abrevaya (2013) and Wooldridge (2019) considered the equivalence of $\hat{\beta}_{C1} = \hat{\beta}_W$ for models with unbalanced panel data. We introduce an extended Frisch-Waugh theorem to prove $\hat{\beta}_{C1} = \hat{\beta}_W$ in Appendix A. The standard Frisch-Waugh theorem is for models with the OLS estimator; the extended theorem we developed is for the GLS estimator as in the case of the

CRE1 model. However, neither the standard nor extended Frisch-Waugh theorem can be used to prove the second equality in Theorem 1. Using the typical process to solve a system of equations and some matrix equivalences, such as Equation (48), the proof of both equalities is provided in Appendix B.

The theorem shows that the numerical solutions of $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$ can be derived from the between and within regressions, without estimating the CRE1 model. $\hat{\gamma}_{C1}$ is the estimator of the partial effect of $\bar{x}_g$. The equation $\hat{\gamma}_{C1} = \hat{\beta}_{B1G} - \hat{\beta}_W$ shows that this partial effect is equal to the difference between the GLS between estimator and the within estimator.

We discuss a possible issue with the use Ω^{-1} in the GLS estimator when $m_g \to \infty$. In the previous subsection, we showed that $\Omega^{-1} = \pi_2 I$ or $\Omega^{-1} = \pi_2 M$ (Equations (42) and (43)) when $m_g \to \infty$. These results on Ω^{-1} are inconsistent with the property of Ω since the components of Ω do not include m_g and the inverse of Ω should not be affected by m_g. These inconsistent results affect the use of the GLS estimator. If $\Omega^{-1} = \pi_2 I$, then there is no need to consider the GLS estimator of the CRE1 model when $m_g \to \infty$. This issue and puzzle can be solved since Ω^{-1} is always used with other data matrices in the GLS estimator. For example, consider $X'\Omega^{-1}X$ in Equation (56). Using $\Omega^{-1} = \pi_2 I + D_{m\pi_1}\bar{J}$, we show $X'\Omega^{-1}X = \pi_2 \tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ for any m_g, even if $m_g \to \infty$ (Equation (B.6) in Appendix B). This matrix equivalence includes two terms: $\pi_2 \tilde{X}'\tilde{X} = V(\hat{\beta}_W)^{-1}$ and $\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} = V(\hat{\beta}_{B1G})^{-1}$. Both terms are positive definite matrices, and $X'\Omega^{-1}X$ do not converge to $\pi_2 X'X$ nor $\pi_2 \tilde{X}'\tilde{X}$ when $m_g \to \infty$. Appendix B shows the same results for other matrix equivalences of matrices with Ω^{-1} when $m_g \to \infty$. The discussion of this issue with Ω^{-1} is similar to the discussion of $\Omega_{\bar{U}}^{-1}$ in $\bar{X}\Omega_{\bar{U}}^{-1}\bar{X}$ in Equation (48), where m_g^{-1} in $\Omega_{\bar{U}}^{-1}$ is cancelling with m_g from $\bar{X}_g'\bar{X}_g = m_g \bar{x}'\bar{x}$. Hence, the use of $\Omega^{-1} = \pi_2 I + D_{m\pi_1}\bar{J}$ in the GLS estimator of the CRE1 model is still valid when $m_g \to \infty$. It is inappropriate to reduce Ω^{-1} into $\Omega^{-1} = \pi_2 I$ or $\Omega^{-1} = \pi_2 M$ when $m \to \infty$.

Based on the means, variances, and covariance of $\hat{\beta}_{B1G}$ and $\hat{\beta}_W$ (Equations (50), (51), (53), (54), and (55), the means, variances, and covariance of $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$ are

$$E(\hat{\beta}_{C1}) = \beta$$

$$E(\hat{\gamma}_{C1}) = \beta_B - \beta = \gamma$$

$$V(\hat{\beta}_{C1}) = V(\hat{\beta}_W) = \sigma_\varepsilon^2 (\tilde{X}'\tilde{X})^{-1}$$

$$V(\hat{\gamma}_{C1}) = V(\hat{\beta}_{B1G}) + V(\hat{\beta}_W) = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1} + \sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1}$$

$$C(\hat{\beta}_{C1}, \hat{\gamma}_{C1}) = -V(\hat{\beta}_W) = -\sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1}$$

With these variances, the inference of β and γ can be made without the estimation of Ω and the CRE1 model.

The above results can be applied to balanced panel data. For balanced panel data, $\hat{\beta}_{B1O_n} = \hat{\beta}_{B1G}$. Then we have $\hat{\gamma}_{C1} = \hat{\beta}_{B1O_n} - \hat{\beta}_W$, which is the result in in Mundlak (1987). To check asymptotic and small sample properties of the GLS estimator, we rewrite the variances of $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$ as

$$V(\hat{\beta}_{C1}) = \frac{\sigma_\varepsilon^2}{N}\left(\frac{\tilde{X}'\tilde{X}}{N}\right)^{-1}$$

$$V(\hat{\gamma}_{C1}) = \frac{\sigma_{\bar{u}}^2}{n}\left(\frac{\bar{x}'\bar{x}}{n}\right)^{-1} + \frac{\sigma_\varepsilon^2}{N}\left(\frac{\tilde{X}'\tilde{X}}{N}\right)^{-1}$$

The variance of $\hat{\beta}_{C1}$ is the same as the variance of the within estimator and it is related to the noise-to-signal ratio of the within regression; the variance of $\hat{\gamma}_{C1}$ is the sum of the variances of the between and within estimators and it is related to the noise-to-signal ratios of the between and within regressions. In a large sample inference of β and γ with $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$, the CLT can be applied. However, we need to separate $m_g \to \infty$ from $n \to \infty$ in applying the CLT. When $n \to \infty$, the CLT can be applied to both the between and within estimators, and therefore the CLT can be applied to the inference of β and γ. When $m_g \to \infty$ and n is finite, the inference is complicate since the CLT cannot be applied to the between estimator.

In a small sample inference of β and γ when both m_g and n are finite, the issue of determining an appropriate number of degrees of freedom occurs. The number of degrees of freedom in inference of β_B in the between regression is $n - k$, and it is $N - n - k$ in inference of β in the within regression. The standard procedure in inference of all parameters in the CRE1 model is to use the same number of degrees of freedom $N - n - k$. This can be incorrect in estimating γ since the variance of $\hat{\gamma}$ is the sum of the variances of the between and within estimators. With $C(\hat{\beta}_{B1G}, \hat{\beta}_W) = 0$, a solution is to use the Welch–Satterthwaite equation to determine the number of degrees of freedom in estimating γ, which is similar to the inference of the difference between two population means with independent samples.

3.2.3 The Pooled OLS and Random Effects Estimators Under the CRE1 Model

Next, we examine the pooled OLS estimator $\hat{\beta}_{OLS}$ and the random effects estimator $\hat{\beta}_{RE}$. It has been showed that $\hat{\beta}_{OLS}$ and $\hat{\beta}_{RE}$ are matrix weighted averages of the between and within estimators. While Maddala's (1971) analysis is based on classical panel data models without group mean variables and Mundlak (1978) extended the analysis to the CRE1 model with balanced panel data, we demonstrate the relationships of these two estimators with the between and within estimators under the CRE1 model with cluster data.

Consider the pooled regression $Y = X\beta + u^*$, where u^* is the random error term with elements of u_{gi}^*. The pooled OLS estimator of β in this regression is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \tag{58}$$

Using $X'Y = \tilde{X}'\tilde{Y} + \bar{X}'\bar{Y}$, the decomposition of the pooled OLS estimator $\hat{\beta}_{OLS}$ is (Maddala, 1971; Mundlak, 1978)

$$\hat{\beta}_{OLS} = (X'X)^{-1}\tilde{X}'\tilde{X}\hat{\beta}_W + (X'X)^{-1}\bar{X}'\bar{X}\hat{\beta}_{B1O_N}$$

Define $\lambda_{OLS} = (X'X)^{-1}\bar{X}'\bar{X}$. Then $I - \lambda_{OLS} = (X'X)^{-1}\tilde{X}'\tilde{X}$ and

$$\begin{aligned}\hat{\beta}_{OLS} &= \lambda_{OLS}\hat{\beta}_{B1O_N} + (I - \lambda_{OLS})\hat{\beta}_W \\ &= \hat{\beta}_W + \lambda_{OLS}\left(\hat{\beta}_{B1O_N} - \hat{\beta}_W\right)\end{aligned} \tag{59}$$

The first equality shows that the pooled OLS estimator is a matrix weighed average of the between and within estimators; λ_{OLS} is the weighting matrix for the pooled OLS estimator. When the between variation is relatively larger than the within variation ($\bar{X}'\bar{X} > \tilde{X}'\tilde{X}$), more weight is on the between estimator; otherwise, more weight is on the within estimator. Under the CRE1 model, substituting $Y = X\beta + \bar{X}\gamma + u$ into $\hat{\beta}_{OLS}$, the mean and the variance of the OLS estimator $\hat{\beta}_{OLS}$ are

$$E\left(\hat{\beta}_{OLS}\right) = \beta + \lambda_{OLS}\gamma$$

$$\begin{aligned}V\left(\hat{\beta}_{OLS}\right) &= (X'X)^{-1}X'E(u'u)X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1}\end{aligned}$$

This shows that $\hat{\beta}_{OLS}$ is an unbiased estimator of $\beta + \lambda_{OLS}\gamma$. However, it is not an efficient estimator. Using $\Omega = \Omega_{\tilde{\varepsilon}} + \Omega_{\bar{U}} = \sigma_\varepsilon M + D_{m\sigma_{\bar{u}}^2}\bar{J}$ from Equations (24), (36), and (40), we have

$$\begin{aligned}X'\Omega X &= X'\left(\sigma_\varepsilon M + D_{m\sigma_{\bar{u}}^2}\bar{J}\right)X \\ &= \sigma_\varepsilon^2\tilde{X}'\tilde{X} + \bar{x}'D_{m^2\sigma_{\bar{u}}^2}\bar{x}\end{aligned} \tag{60}$$

The variance of $\hat{\beta}_{OLS}$ can be written as

$$V(\hat{\beta}_{OLS}) = (X'X)^{-1}\left(\sigma_{\varepsilon}^2 \tilde{X}'\tilde{X} + \bar{x}' D_{m^2\sigma_{\bar{u}}^2}\bar{x}\right)(X'X)^{-1}$$

To consider the asymptotic properties of the variance, we rewrite the variance as

$$V(\hat{\beta}_{OLS}) = \left(\frac{X'X}{N}\right)^{-1}\left(\frac{\sigma_{\varepsilon}^2}{N}\frac{\tilde{X}'\tilde{X}}{N} + \frac{\bar{x}' D_{m^2\sigma_{\bar{u}}^2}\bar{x}}{N^2}\right)\left(\frac{X'X}{N}\right)^{-1}$$

$$= \left(\frac{X'X}{N}\right)^{-1}\frac{\sigma_{\varepsilon}^2}{N}\frac{\tilde{X}'\tilde{X}}{N}\left(\frac{X'X}{N}\right)^{-1} + \left(\frac{X'X}{N}\right)^{-1}\frac{\bar{x}' D_{m^2\sigma_{\bar{u}}^2}\bar{x}}{N^2}\left(\frac{X'X}{N}\right)^{-1}$$

This form is complex. It can be simplified for balanced panel data. For balanced panel data, using $\hat{\beta}_{B1O_n} = \hat{\beta}_{B1G}$ and $\hat{\beta}_{B1O_n} - \hat{\beta}_W = \hat{\gamma}_{C1}$, Equation (59) becomes

$$\hat{\beta}_{OLS} = \hat{\beta}_W + \lambda_{OLS}\hat{\gamma}_{C1} \tag{61}$$

This is the result in Mundlak (1978), which can only be specified under the CRE1 model. The variance of $\hat{\beta}_{OLS}$ is simplified as

$$V(\hat{\beta}_{OLS}) = (X'X)^{-1}\left(\sigma_{\varepsilon}^2 \tilde{X}'\tilde{X} + m^2\sigma_{\bar{u}}^2\bar{x}'\bar{x}\right)(X'X)^{-1}$$

Using $N = n \cdot m$, rewrite the variance of $\hat{\beta}_{OLS}$ as

$$V(\hat{\beta}_{OLS}) = \left(\frac{X'X}{N}\right)^{-1}\left(\frac{\sigma_{\varepsilon}^2}{N}\frac{\tilde{X}'\tilde{X}}{N} + \frac{\sigma_{\bar{u}}^2}{n}\frac{\bar{x}'\bar{x}}{n}\right)\left(\frac{X'X}{N}\right)^{-1}$$

$$= \left(\frac{X'X}{N}\right)^{-1}\frac{\sigma_{\varepsilon}^2}{N}\cdot\frac{\tilde{X}'\tilde{X}}{N}\left(\frac{X'X}{N}\right)^{-1} + \left(\frac{X'X}{N}\right)^{-1}\frac{\sigma_{\bar{u}}^2}{n}\cdot\frac{\bar{x}'\bar{x}}{n}\left(\frac{X'X}{N}\right)^{-1}$$

This shows that the signal x_{gi} is affected by the noises from ε_{gi} and $\bar{u}_g$. Hence, there are two noise-to-signal ratios: $\left(\frac{X'X}{N}\right)^{-1}\frac{\sigma_{\varepsilon}^2}{N}$ and $\left(\frac{X'X}{N}\right)^{-1}\frac{\sigma_{\bar{u}}^2}{n}$, which are weighted by $\frac{\tilde{X}'\tilde{X}}{N}\left(\frac{X'X}{N}\right)^{-1}$ and $\frac{\bar{x}'\bar{x}}{n}\left(\frac{X'X}{N}\right)^{-1} = \frac{\bar{X}'\bar{X}}{N}\left(\frac{X'X}{N}\right)^{-1}$, respectively. These weighting matrices correspond to the ratios of the within and between variations to the total variation of x. Note that the weighting matrix λ_{OLS} does not include any noise-to-signal ratios.

The random effects estimator $\hat{\beta}_{RE}$ is the GLS estimator of β in the regression $Y = X\beta + u^*$, and

$$\hat{\beta}_{RE} = (X'\Omega_r^{-1}X)^{-1}X'\Omega_r^{-1}Y \tag{62}$$

where $\Omega_r = E(u^*u^{*\prime})$ is the covariance matrix of the random errors u_{gi}^*. Note that $u^* \neq u$ and $\Omega_r \neq \Omega$. If the CRE1 model is the true model, the random effects model is the restricted CRE1 model without including group mean variables $\bar{x}_g$. This implies that omitted variables $\bar{x}$ should be included in u_{gi}^*. For the random effects model, there is no fixed effects, and we can assume that $\bar{x}_g$ is a row vector of random variables with a mean vector of $\mu_{\bar{x}}$ and a variance matrix of $V(\bar{x}_g)$. To ensure that the random errors u_{gi}^* have a zero mean, we can write u_{gi}^* as $u_{gi}^* = \bar{x}_g\gamma - \mu_{\bar{x}}\gamma + u_{gi}$ given that the CRE1 model is the true model. Then the variance and covariance of u_{gi}^* are

$$E\left(u_{gi}^* u_{gj}^*\right) = \gamma' V\left(\bar{x}_g\right)\gamma + \sigma_\alpha^2 = \sigma_{\alpha_r}^2 \text{ for } i \neq j \tag{63}$$

$$E\left(u_{gi}^* u_{gi}^*\right) = \gamma' V\left(\bar{x}_g\right)\gamma + \sigma_\alpha^2 + \sigma_\varepsilon^2 = \sigma_{\alpha_r}^2 + \sigma_\varepsilon^2 \tag{64}$$

$$E\left(u_{gi}^* u_{g'j}^*\right) = 0 \text{ for } g \neq g' \tag{65}$$

For the last equation, we assume that $E\left(\bar{x}_g \bar{x}_{g'}\right) = 0$, for $g \neq g'$, i.e., $\bar{x}_g$ is selected from random sampling. Since $\sigma_{\alpha_r}^2 > \sigma_\alpha^2$, the covariance matrix of u_{gi}^* in the random effect models is different from the covariance matrix of u_{gi} in the CRE1 model. Hence, Ω_r derived from Equations (63) – (65) is different from Ω in Equation (11). With $\Omega_r \neq \Omega$, it is difficult to establish the relationship between $\hat{\beta}_{RE}$ and other estimators. This issue can be solved in the empirical estimation when the feasible GLS estimator is used. The feasible GLS estimators of β in the random effects model and (β, γ) in the CRE1 mdoel require the estimation of Ω_r and Ω. The estimators of $\sigma_{\alpha_r}^2$, σ_α^2 and σ_ε^2 in Ω_r and Ω are usually derived from the between and within regressions. The between and within regressions derived from the random effects model $y_{gi} = x_{gi}\beta + u_{gi}^*$ are

$$\bar{y}_g = \bar{x}_g \beta_B + \bar{u}_g$$

$$\tilde{y}_{gi} = \tilde{x}_{gi}\beta + \tilde{\varepsilon}_{gi}$$

which are the same as Equations (3) and (4) from the CRE1 model. Since the estimation of Ω_r and Ω are derived from the same between and within regressions, it gives $\widehat{\Omega}_r = \widehat{\Omega}$, where $\widehat{\Omega}_r$ and $\widehat{\Omega}$ are the estimators of Ω_r and Ω, respectively. With $\widehat{\Omega}_r = \widehat{\Omega}$ and the estimators of σ_ε^2 and $\sigma_{\bar{u}}^2$ derived from the same between and within regressions, the matrix equivalences of $X'\Omega^{-1}X = \pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ and $X'\Omega^{-1}Y = \pi_2\tilde{X}'\tilde{Y} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$ shown in Equation (B.6) in Appendix B can be applied to $\hat{\beta}_{RE}$. Then, we can show how the random effects estimator is related to the between and within estimators without introducing additional notations by simplifying $\widehat{\Omega}_r = \widehat{\Omega}$ into $\Omega_r = \Omega$. It gives

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

$$= \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\pi_2\tilde{X}'\tilde{Y} + \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$$

$$= \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\pi_2\tilde{X}'\tilde{X}\hat{\beta}_W + \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\hat{\beta}_{B1G}$$

Define the following weighting matrices.

$$\lambda_{RE} = \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$$

$$(I - \lambda_{RE}) = \left(\pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\right)^{-1}\pi_2\tilde{X}'\tilde{X}$$

Then

$$\hat{\beta}_{RE} = \lambda_{RE}\hat{\beta}_{B1G} + (I - \lambda_{RE})\hat{\beta}_W \tag{66}$$

This shows that the random effects estimator is a matrix weighted average of the GLS between estimator and the within estimator. Using $\hat{\gamma}_{C1} = \hat{\beta}_{B1G} - \hat{\beta}_W$ from Theorem 2,

$$\hat{\beta}_{RE} = \hat{\beta}_W + \lambda_{RE}\hat{\gamma}_{C1} \tag{67}$$

Substituting $Y = X\beta + \bar{X}\gamma + u$ into $\hat{\beta}_{RE}$ in Equation (62), the mean and the variance of $\hat{\beta}_{RE}$ are

$$E\left(\hat{\beta}_{RE}\right) = \beta + \lambda_{RE}\gamma$$

$$V(\hat{\beta}_{RE}) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(u'u)\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1}$$

The random effects estimator is an unbiased and efficient estimator of $\beta + \lambda_{RE}\gamma$. It is an MMSE estimator. Using $X'\Omega^{-1}X = \pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ and the variances of the between and within estimators, we can rewrite the variance of the random effects estimator as

$$V(\hat{\beta}_{RE}) = \left(V\left(\hat{\beta}_W\right)^{-1} + V\left(\hat{\beta}_{B1G}\right)^{-1}\right)^{-1}$$

This implies that we can derive the estimate of the variance of the random effects estimator using the estimates of the variances of the between and within estimators. There is no need to estimate Ω for the inference of β in the random effects regression.

Now we consider the asymptotic properties of the variance of $\hat{\beta}_{RE}$ for balanced panel data with $m_g = m$ and $\pi_{1g} = \pi_1$ for all g. Rewrite the variance of $\hat{\beta}_{RE}$ as

$$V(\hat{\beta}_{RE}) = (X'\Omega^{-1}X)^{-1}$$

$$= \left(\sigma_\varepsilon^{2\,-1}\tilde{X}'\tilde{X} + \sigma_{\bar{u}}^{2\,-1}\bar{x}\bar{x}\right)^{-1}$$

$$= \left(\left(\frac{\sigma_\varepsilon^2}{N}\right)^{-1}\frac{\tilde{X}'\tilde{X}}{N} + \left(\frac{\sigma_{\bar{u}}^2}{n}\right)^{-1}\frac{\bar{x}'\bar{x}}{n}\right)^{-1}$$

It contains two signal-to-noise ratios: $\left(\frac{\sigma_{\tilde{\varepsilon}}^2}{N}\right)^{-1} \frac{\tilde{X}'\tilde{X}}{N}$ and $\left(\frac{\sigma_{\bar{u}}^2}{n}\right)^{-1} \frac{\bar{x}'\bar{x}}{n}$, which are from the within and between regressions, respectively. These two ratios are related to N and n, separately. The CLT can be applied to the inference of $\beta + \lambda_{RE}\gamma$ with $\hat{\beta}_{RE}$ when $n \to \infty$. For $m \to \infty$ and finite n, the CLT cannot be applied to the estimation of the between estimator, and small sample inferences of $\beta + \lambda_{RE}\gamma$ with $\hat{\beta}_{RE}$ may be complicate.

The weighting matrices of the random effects estimator are also based on $X'\Omega^{-1}X$. The weights can be rewritten as

$$\lambda_{RE} = \left(V(\hat{\beta}_W)^{-1} + V(\hat{\beta}_{B1G})^{-1}\right)^{-1} V(\hat{\beta}_{B1G})^{-1}$$

$$I - \lambda_{RE} = \left(V(\hat{\beta}_W)^{-1} + V(\hat{\beta}_{B1G})^{-1}\right)^{-1} V(\hat{\beta}_W)^{-1}$$

The weights are related to the inverses of the variances of the between and within estimator or the signal-to-noise ratios of the between and within regressions. The random effects estimator is related to the signal-to-noise ratios of the between and within regressions. If the signal-to-noise ratio $\left(\frac{\sigma_{\tilde{\varepsilon}}^2}{N}\right)^{-1} \frac{\tilde{X}'\tilde{X}}{N}$ of the within regression is bigger than the signal-to-noise ratio $\left(\frac{\sigma_{\bar{u}}^2}{n}\right)^{-1} \frac{\bar{x}'\bar{x}}{n}$ of the between regression, then more weight is placed on the within estimator. Otherwise, more weight is on the between estimator.

3.2.4 Comments on the Analyses of the Random Effects Estimator in Maddala and Mundlak

For models with balanced panel data, Maddala (1971) and Mundlak (1978) showed $\hat{\beta}_{RE}$ is the matrix weighted average of the between and within estimators. Assuming $\Omega = \Omega_r$, Mundlak (1978) showed further that $\hat{\beta}_{RE}$ is a linear function of $\hat{\beta}_W$ and $\hat{\gamma}$ under the CRE1 model. Our results for cluster data should be the same as their results if we apply our results to balanced panel data. Our results show that the weighting matrix λ_{RE} in Equation (66) under the CRE1 model is the same as the weighting matrix under the standard random effects model in Maddala (1971) using $\widehat{\Omega}_r = \widehat{\Omega}$. However, we observe two differences between our results on λ_{RE} and those found in Mundlak (1978) and Maddala (1971). To show the differences, we rewrite λ_{RE} for balanced panel data. For balanced panel data, $m_g = m$ and $\pi_{1g} = \pi_1$ are constants for all g. Using $\Omega^{-1} = \Omega_{\tilde{\varepsilon}}^{-1} + \Omega_{\bar{U}}^{-1} = \pi_2 M + (\pi_2 + m\pi_1)\bar{J} = \pi_2 M + (m\sigma_{\bar{u}}^2)^{-1}\bar{J}$ from Equations (25), (28), (37), and (41), and $\bar{X}'\bar{X} = m\bar{x}'\bar{x}$, rewrite $\lambda_{RE} = (X'\Omega^{-1}X)^{-1}\bar{X}'\Omega^{-1}\bar{X}$ as

$$\lambda_{RE} = \left(\pi_2 \tilde{X}'\tilde{X} + (\pi_2 + m\pi_1)\bar{X}'\bar{X}\right)^{-1}(\pi_2 + m\pi_1)\bar{X}'\bar{X} \tag{68.1}$$

$$= \left({\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X} + (m\sigma_{\bar{u}}^2)^{-1}\bar{X}'\bar{X}\right)^{-1}(m\sigma_{\bar{u}}^2)^{-1}\bar{X}'\bar{X} \tag{68.2}$$

$$= \left({\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X} + {\sigma_{\bar{u}}^2}^{-1}\bar{x}'\bar{x}\right)^{-1}{\sigma_{\bar{u}}^2}^{-1}\bar{x}'\bar{x} \tag{68.3}$$

First, we note that our specification of λ_{RE} is different from that in Mundlak (1978, Equation (3.5), p.73). His λ_{RE} based on Chamberlain (1975) is shown as

$$\left(\pi_2 \tilde{X}'\tilde{X} + (\pi_2 + \pi_1)\bar{X}'\bar{X}\right)^{-1}(\pi_2 + \pi_1)\bar{X}'\bar{X}$$

$$= \left(\bar{X}'\bar{X} + {\sigma_\varepsilon^2}^{-1}(\pi_2 + \pi_1)^{-1}\tilde{X}'\tilde{X}\right)^{-1}\bar{X}'\bar{X}$$

Comparing this formula with λ_{RE} in Equation (68.1), his formula of λ_{RE} is missing an "m" in $(\pi_2 + m\pi_1)$ and he did not show $V(\hat{\beta}_{RE}) = (X'\Omega^{-1}X)^{-1}$ under the Mundlak model. With the correct formula, λ_{RE} is the same under both Mundlak model and random effects model. $\hat{\beta}_{RE}$ is an unbiased and efficient estimator of $\beta + \lambda_{RE}\gamma$; it is the MMSE estimator for $\beta + \lambda_{RE}\gamma$.

Second, Maddala (1971) notes that, from Equation (68.2),

$$\lambda_{RE} = \left({\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X} + (m\sigma_{\bar{u}}^2)^{-1}\bar{X}'\bar{X}\right)^{-1}(m\sigma_{\bar{u}}^2)^{-1}\bar{X}'\bar{X} = 0 \tag{69}$$

when $m \to \infty$. He concluded that the random effects estimator is equal to the within estimator when the sample size in each group is infinite. Mundlak (1978, p. 79), Ahn and Moon (2014), and some econometrics textbooks (i.e., Hsiao, 2014, p. 43; Baltagi, 2021, p. 27; Wooldridge, 2010, p. 327) follow this conclusion. This conclusion is incorrect since Maddala (1971) did not consider $\bar{X}'\bar{X} = m\bar{x}\bar{x}$. Using $\bar{X}'\bar{X} = m\bar{x}'\bar{x}$, we can rewrite $(m\sigma_{\bar{u}}^2)^{-1}\bar{X}'\bar{X}$ in Equation (69) as $(m\sigma_{\bar{u}}^2)^{-1}m\bar{x}'\bar{x} = (\sigma_{\bar{u}}^2)^{-1}\bar{x}'\bar{x}$. Then λ_{RE} in Equation (69) becomes Equation (68.3). When $m \to \infty$, $\sigma_{\bar{u}}^2 = \sigma_\alpha^2$ and $\lambda_{RE} = \left({\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X} + {\sigma_\alpha^2}^{-1}\bar{x}'\bar{x}\right)^{-1}{\sigma_\alpha^2}^{-1}\bar{x}'\bar{x}$. λ_{RE} is nonzero. The incorrect conclusion of $\lambda_{RE} \to 0$ is the same issue as the conclusion of $X'\Omega^{-1}X = {\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X}$ based on $\Omega^{-1} = \pi_2 M$ in the earlier discussion of Ω^{-1} in the GLS estimator of the CRE1 model, and the conclusion of $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X} = 0$ based on $\Omega_{\bar{U}}^{-1} = 0$ in the discussion of the GLS between estimator when $m_g \to \infty$. Hence, the random effects estimator is always a matrix weighted average of the between and within estimators for any group size, even if the group size m is infinite.

3.3 The CRE1 Model: Comparison of Five Estimators

After examining the properties of five different estimators and deriving their relationships in the above subsection, we compare these estimators and discuss which estimator is the "best" in this subsection. Based on the result of $\hat{\beta}_{C1} = \hat{\beta}_W$, Mundlak (1978, pp. 69-70) concluded that there is only one estimator to estimate β and any matrix combination of the within and between estimates is generally biased. His conclusion is derived by assuming that there is only one partial effect of the explanatory variables, β. This paper opts for a different interpretation. Instead of focusing on a single partial effect β, we consider two partial effects of x and each estimator is to estimate a linear combination of these two partial effects. The first partial effect is the direct impact of x and it is measured by β; the second partial effect is the impact of group mean variables and it is measured by γ. Based on the within regression, the direct effect of x or the within estimator of β only measures the within effect. We list five estimators of the parameters β and γ as follows to explain our different interpretation from Mundlak (1978).

(i) GLS between estimator: $\hat{\beta}_{B1G}$, with $E(\hat{\beta}_{B1G}) = \beta + \gamma$

(ii) Within estimator: $\hat{\beta}_W$, with $E(\hat{\beta}_W) = \beta$

(iii) GLS estimator of the CRE1 model: $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$, with $E(\hat{\beta}_{C1}) = \beta$ and $E(\hat{\gamma}_{C1}) = \gamma$

(iv) Pooled OLS estimator: $\hat{\beta}_{OLS} = \hat{\beta}_W + \lambda_{OLS}\hat{\gamma}_{C1}$, with $E(\hat{\beta}_{OLS}) = \beta + \lambda_{OLS}\gamma$

(v) Random effects estimator: $\hat{\beta}_{RE} = \hat{\beta}_W + \lambda_{RE}\hat{\gamma}_{C1}$, with $E(\hat{\beta}_{RE}) = \beta + \lambda_{RE}\gamma$

Table 1 summarizes the weighting matrices, means, and variances of these five estimators. We compare these estimators in terms of unbiasedness, efficiency, and modeling structure with respect to β and γ. First, we discuss unbiasedness. Traditionally, the pooled OLS estimator and the random effects estimator are considered as matrix weighted averages of the between and within estimators, $\hat{\beta}_B$ and $\hat{\beta}_W$. The CRE1 model shows that $\hat{\beta}_B$ estimate $\beta + \gamma$ and $\hat{\beta}_W$ estimate β. A linear combination of $\hat{\beta}_B$ and $\hat{\beta}_W$ assigns the weight to β twice. Hence, we consider each estimator as a linear function of $\hat{\beta}_W$ and $\hat{\gamma}$, instead of $\hat{\beta}_B$ and $\hat{\beta}_W$. The between estimator in (i) estimates $\beta + \gamma$, which assign a full weight of one to γ; the within estimator in (ii) only estimates β and assigns a zero weight to γ. The GLS estimator of the CRE1 model in (iii) estimates β and γ, separately. The pooled OLS estimator in (iv) and the random effects estimator in (v) estimate the sum of β and a different weight of γ. These specifications of estimators in terms of $\hat{\beta}_W$ and $\hat{\gamma}$ show that a key difference between the within estimator and other four estimators is the role of γ.

When $\gamma = 0$, all estimators, except $\hat{\gamma}_{C1}$, are unbiased estimators for β. When $\gamma \neq 0$, $\hat{\beta}_W$ and $\hat{\beta}_{C1}$ are still unbiased estimators for β; the rest three estimators are biased. However, if we consider both β and γ as partial effects of x, each of the five estimators is an unbiased estimator of the linear combination of β and γ. The unbiasedness issue is related to whether γ is zero or not and whether we treat γ as a partial effect of x or not.

Second, we compare the efficiency of these five estimators by examining the variances of these estimators shown in Table 1. We note that the between and within regressions are two different regressions with different dependent and independent variables. Furthermore, these two regressions estimate different parameters if $\gamma \neq 0$. Therefore, we cannot compare the variances of the between and within estimators for efficiency and significance. For example, the significance (insignificance) of $\hat{\beta}_W$ does not imply the significance (insignificance) of $\hat{\beta}_B$, and vice versa. For these two estimators, there is no dominance of one estimator over the other in terms of efficiency. The GLS estimator of the CRE1 model estimates each partial effect individually. The estimator is unbiased and efficient. When $\gamma = 0$, both pooled OLS estimator and random effects estimator are unbiased but less efficient than $\hat{\beta}_W$ in estimating β. The Hausman test (1978) is based on inefficiency of the random effects estimator. If $\gamma \neq 0$ and we consider the parameter estimation of a linear combination of β and γ, then the pooled OLS estimator is unbiased but inefficient in estimating $\beta + \lambda_{OLS}\gamma$ while the random effects estimator is unbiased and efficient in estimating $\beta + \lambda_{RE}\gamma$. In addition, the GLS between estimator is unbiased and efficient in estimating $\beta + \gamma$. Hence, there are multiple MMSE estimators.

With multiple MMSE estimators, it raises two issues: how to compare different multiple MMSE estimators and are there more MMSE estimators? For the first issue, note that we only compare MSE of different estimators that estimate the same parameter. One of the results in Mundlak (1978) is driving the best MMSE estimator among different linear combinations of the between and within estimators. The target parameter for his comparison of different MMSE estimators is the single key parameter β. Our analysis of different estimators gives multiple MMSE estimators since we consider the linear function of two partial effects of x, β and γ. $\hat{\beta}_W$ and $\hat{\beta}_{C1}$ are the MMSE estimators of β; $\hat{\gamma}_{C1}$ is the MMSE estimator of γ; $\hat{\beta}_{B1G}$ is the MMSE estimator of $\beta + \gamma$; and $\hat{\beta}_{RE}$ is the MMSE estimator of $\beta + \lambda_{RE}\gamma$. Different linear functions of β and γ represent different characteristics of the aggregates of the two partial effects. Hence, we don't compare MMSE estimators of different parameters with different characteristics, and we cannot conclude which MMSE estimator is the best. However, we can compare the characteristics of different aggregates of two partial effects based on their weighting matrices. The random effects estimator considers both signal-to-noise ratios of the between and within regressions in its weighting matrix λ_{RE}, and it can be better than the GLS between estimator, which ignores signal-to-noise ratios in its weighting matrices. For the second issue, we examine if there are other MMSE estimators. Suppose $\hat{\beta}_\lambda$ is an estimator of a linear combination of β and γ, and it is a matrix weighted average of the between and within estimators. We can define $\hat{\beta}_\lambda$ as $\hat{\beta}_\lambda = (A+B)^{-1}A\hat{\beta}_{B1G} + (A+B)^{-1}B\hat{\beta}_w$, where A and B are symmetric positive definite matrices derived from the model and data. The weighting matrices assigned to the between and within estimators are $\lambda = (A+B)^{-1}A$ and $I-\lambda = (A+B)^{-1}B$, respectively. Then the mean and variance of $\hat{\beta}_\lambda$ are

$$E(\hat{\beta}_\lambda) = E(\lambda\hat{\beta}_{B1G} + (I-\lambda)\hat{\beta}_W) = \beta + \lambda\gamma \tag{70}$$

$$V(\hat{\beta}_\lambda) = (A+B)^{-1}\left(A(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}A + B\sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1}B\right)(A+B)^{-1} \tag{71}$$

This shows that $\hat{\beta}_\lambda$ is an unbiased estimator of $\beta+\lambda\gamma$ as long as λ and $I-\lambda$ are the weighting matrices assigned to the between and within estimators. This implies that we can derive different unbiased estimator of linear combination of β and γ with different λ. For example, four estimators discussed in this subsection have different λs and weighting matrices (See Table 1). For efficiency, $\hat{\beta}_\lambda$ is efficient and it is an MMSE estimator of $\beta+\lambda\gamma$ if

$$V(\hat{\beta}_\lambda) = (A+B)^{-1}$$

This implies that the condition of efficiency is

$$A(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}A + B\sigma_\varepsilon^2(\tilde{X}'\tilde{X})^{-1}B = A+B \tag{72}$$

This condition includes both signals and noises of the between and within regressions. One obvious solution of A and B that satisfy this condition is $A = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ and $B = {\sigma_\varepsilon^2}^{-1}\tilde{X}'\tilde{X} = \pi_2\tilde{X}'\tilde{X}$, with $A+B = X'\Omega^{-1}X$. This solution gives $\lambda = \lambda_{RE}$ and the random effects estimator. Note that the pooled OLS estimator is inefficient since $A = \bar{X}'\bar{X}$, $B = \tilde{X}'\tilde{X}$, and $A+B = X'X$ for the pooled OLS estimator, and these A and B do not satisfy Equation (72). If there is another MMSE estimator with a different λ, then the weighting matrices, A and B, should satisfy Equation (72).

Third, we discuss modeling structures and strategies of these five estimators. In our previous discussion of unbiasedness and efficiency, we show that a major difference between the within estimator and other four estimators is related to γ. The specification of γ in modeling is related to an underlying assumption of the model: whether the unobserved group component α_g is fixed or random. Although the within regression can be derived from the random effects model and CRE models, the within regression is identical to the fixed effects model in terms of modeling structure. Both fixed effect model and within regression assume α_g is fixed. The fixed effect model explicitly assumes that α_g is a fixed parameter; the within regression implicitly assumes that α_g is fixed by eliminating all random group components in its estimation when the data of deviations from group means are used. All four other estimators explicitly or implicitly assume that α_g is random. In estimating the GLS estimator of the CRE1 model, α_g is assumed to be random such that both $\bar{x}_g$ and α_g can be included in the model. The estimations of the three remaining estimators implicitly assume that α_g is random as in the CRE1 model. Hence, the comparison of modeling structures of these five estimators is related to an important debate in studies of panel data models: whether we should use the fixed effects model or the random effects model? Suppose α_g is fixed, the estimation of γ is irrelevant and there is no need to consider the random effects model or the CRE1 model; then the fixed effects model is sufficient. However, whether γ is relevant or not should be based on empirical evidence. If $\gamma \neq 0$, all estimators, except the within estimator, capture the second partial effect γ by considering a linear function of the two partial effects. In an extreme case, $\beta = 0$ and $\gamma \neq 0$, then the within estimator cannot detect any partial effects and four other estimators are able to estimate the partial effect γ. Although these four estimators consider both partial effects β and γ, a significant cost in this flexibility is that these estimators may suffer from the omitted variable bias, which occurs if group mean variables are correlated to omitted group variables. The advantage of the fixed effects model and the within regression is avoiding this omitted variable bias by either controlling all group characteristics in the fixed parameter α_g or eliminating α_g in formulating the model. The cost of using the fixed effects/within estimator is the omission of γ.

The fixed effects model and the within regression focus solely on single partial effect β, and ignores the second partial effect γ, which may be important. The presumption of a fixed effects parameter may suggest that the fixed effects model is a “restricted” model. We note that it is not a restricted version of the CRE1 model by restricting $\gamma = 0$. In general, a restricted model has a larger prediction error. We demonstrate the impacts of presuming a fixed effects parameter by comparing predictions and prediction errors from the fixed effects model and the CRE1 model. Instead of treating α_g as random as in CRE models, suppose α_g is a fixed parameter for the g-th group. Then the fixed effects model is

$$y_{gi} = x_{gi}\beta + \alpha_g + \varepsilon_{gi}^*. \tag{73}$$

where α_g is a parameter and ε_{gi}^* are homoscedastic and serial uncorrelated random errors. Note that $\bar{x}_g$ and z_g are excluded from the model to avoid the multicollinearity issue. The estimation of the parameters in the fixed effects model is based on the least squares dummy variable (LSDV) regression:

$$Y = G\alpha + X\beta + \varepsilon^* \tag{74}$$

where α is a $n \times 1$ parameter vector and G is a $N \times n$ matrix of dummy variables such that $G = (G_1, G_2, \cdots, G_n)$ and G_g is a $N \times 1$ column vector of the dummy variable for the g-th group with $G_g = \left(G_{g1}', G_{g2}', \cdots, G_{gn}'{}'\right)'$, $G_{gi} = 0e_{m_i}$ for $i \neq g$ and $G_{gg} = e_{m_g}$. The OLS estimator of α and β in the LSDV regression is the fixed effects estimator and denoted as $\hat{\alpha}_{FE}$ and $\hat{\beta}_{FE}$. Using $G(G'G)^{-1}G' = \bar{J}$, $I - G(G'G)^{-1}G' = I - \bar{J} = M$, it can be shown that (Greene, 2018, p. 393)

$$\hat{\alpha}_{FE} = (G'G)^{-1}G'\left(Y - X\hat{\beta}_W\right) = \bar{y} - \bar{x}\hat{\beta}_W$$

$$\hat{\beta}_{FE} = (X'MX)^{-1}X'MY = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} = \hat{\beta}_W$$

Using $G\hat{\alpha}_{FE} = \bar{Y} - \bar{X}\hat{\beta}_W$, the fitted equation of the fixed effects model is

$$\hat{Y}_{FE} = G\hat{\alpha}_{FE} + X\hat{\beta}_W = \left(\bar{Y} - \bar{X}\hat{\beta}_W\right) + X\hat{\beta}_W = \bar{Y} + \tilde{X}\hat{\beta}_W \tag{75}$$

The prediction of y_{gi} contains two components. The first component is directly measured by the group mean $\bar{y}_g$ and the second is related to within deviations. Note that the fitted equation of the within regression is $\hat{\tilde{Y}} = \tilde{X}\hat{\beta}_W$. Let $\hat{\tilde{Y}} = \hat{Y} - \bar{Y}$. Then $\hat{Y} - \bar{Y} = \tilde{X}\hat{\beta}_W$ and $\hat{Y} = \bar{Y} + \tilde{X}\hat{\beta}_W$, which is the same as Equation (75). The fixed effects model and the within regression have the same fitted equation and prediction interpretation.

The fitted equation of the CRE1 model is

$$\hat{Y}_{C1} = X\hat{\beta}_W + \bar{X}\hat{\gamma}_{C1} = X\hat{\beta}_W + \bar{X}\left(\hat{\beta}_{B1G} - \hat{\beta}_W\right) = \bar{X}\hat{\beta}_{B1G} + \tilde{X}\hat{\beta}_W \tag{76}$$

This shows the prediction of y_{gi} also contains two components. The within deviation component $\tilde{X}\hat{\beta}_W$ is the same as that in the fitted equation of the fixed effects model. However, the group mean component is predicted by $\hat{\bar{Y}} = \bar{X}\hat{\beta}_{B1G}$, which is fitted equation of the between regression $\bar{Y} = \bar{X}\beta_B + \bar{U}$. The comparison of the two prediction equations $\hat{Y}_{FE}$ and $\hat{Y}_{C1}$ in Equations (75) and (76) shows that the fixed effects model ignores the predictivity of $\bar{y}_g$ and it assumes that the population group mean of y in the g-th group is estimated by a fixed value $\bar{y}_g$. We explain why the predictivity of $\bar{y}_g$ is omitted in the fixed effects model. Pre-multiplying the LSDV regression in Equation (74) by $(G'G)^{-1}G$ gives the between regression of the fixed effects model:

$$\bar{y} = \alpha + \bar{x}\beta + \bar{\varepsilon}^*$$

where $\bar{\varepsilon}^* = (G'G)^{-1}G\varepsilon^*$. The parameters α and β in this between regression cannot be identified since both $\bar{y}$ and α have the same row dimension. Once we use the fixed effects model, we cannot predict $\bar{y}$. The prediction of $\bar{y}$ must be omitted. This omission of the predictivity of $\bar{y}_g$ is not an issue if the main purpose of the estimation is to estimate the partial effect β or when $\gamma = 0$. The lack of predictivity of $\bar{y}$ implied by the presumption of the fixed effects parameter in the fixed effects model is usually overlooked in the empirical analysis.

Different modeling strategies give different prediction powers. We compare the prediction powers of the fixed effects model and the CRE1 model. From Equation (75), the fitted error equation of the LSDV regression is

$$Y - \hat{Y}_{FE} = Y - \bar{Y} - \tilde{X}\hat{\beta}_W = \tilde{Y} - \tilde{X}\hat{\beta}_W = \tilde{e}$$

which is the same as the residual term in the within regression. This again shows that the fixed effects model and the within regression are fundamentally the same and they only capture the impact and the prediction of within variations. The fitted error equation of the CRE1 model is

$$Y - \hat{Y}_{C1} = Y - \bar{X}\hat{\beta}_{B1G} - \tilde{X}\hat{\beta}_W = Y - \bar{Y} + \bar{Y} - \bar{X}\hat{\beta}_{B1G} - \tilde{X}\hat{\beta}_W = \hat{\bar{U}} + \tilde{e}$$

where $\hat{\bar{U}} = \bar{Y} - \bar{X}\hat{\beta}_{B1G}$ is the residual term of the between regression. The sum of squared prediction errors from the CRE1 model is $\hat{\bar{U}}'\hat{\bar{U}} + \tilde{e}'\tilde{e}$. This is larger than the sum of squared prediction errors from the fixed effects model, $\tilde{e}'\tilde{e}$. However, we cannot conclude that the "restricted" fixed effects model is better than the CRE1 model since it has a smaller sum of squared prediction errors. The fixed effects model has a better prediction power simply because this model is a prediction model only focusing on the partial effect β due to within deviations $\tilde{x}$. It does not consider the predictivity of $\bar{y}_g$, group means of y. The CRE1 model is a prediction model focusing on two partial effects due to both $\bar{x}_g$ and within deviations. Another simple way to demonstrate the difference is to use the equation $Y = \bar{Y} + \tilde{Y}$. The fixed effect model and the within regression only concern the estimation and prediction of $\tilde{Y}$ while the CRE1 model estimates Y and concerns both $\bar{Y}$ and $\tilde{Y}$. In summary, the choice between the fixed effects/within estimator and other estimators depends on the importance of the partial effect γ. If we are concerned with both β and γ and it is possible that $\gamma \neq 0$, then other estimators, such as the GLS estimator of the CRE1 model and the random effects estimator, are worth considering and may provide additional information on the partial effects of x.

4. The CRE2 Models with Cluster Data

When group variables z_g are added to the CRE1 model, we have the CRE2 model, $y_{gi} = x_{gi}\beta + \bar{x}_g\gamma + z_g\xi + u_{gi}$. Similar to $\bar{x}_g$, z_g is a $1 \times l$ row vector and has only one value for each z variable in each group. Since z_g are only observable at the group level, z variables are not included in x variables. We denote two data matrices of z_g as $\ddot{z}$ and Z, with the data structures similar to $\bar{x}$ and $\bar{X}$, respectively. The matrix form of the CRE2 model is

$$Y = X\beta + \bar{X}\gamma + Z\xi + u \tag{77}$$

where u is the column vector contains elements of composite random errors $u_{gi} = \alpha_g + \varepsilon_{gi}$. Assume the random errors α_g and ε_{gi} have the same properties as the random errors of the CRE1 model as shown in Equations (5) and (6). Then $Var(u) = \Omega$ as in Equation (11). We use the same notation Ω for the covariance matrix of the random errors in both CRE1 model and CRE2 model. If the CRE2 model is true, the CRE1 model is a misspecified model. The CRE1 and CRE2 models should have different covariance matrices of the random errors. This implies the CRE2 model requires a different set of notations related to Ω. To avoid creating excessive notations in differentiating between the CRE2 and CRE1 model, we keep the same notations, such as u_g, $\bar{u}_g$, Ω, and $\Omega_{\bar{u}}$, used in the CRE1 model for the CRE2 model; but we use these notations with caution.

4.1 The Between and Within Estimators and the GLS Estimator of the CRE2 Model

To analyze the CRE2 model, we begin with the between and within regressions derived from the CRE2 model. The between regression is

$$\bar{y} = \bar{x}\beta_B + \ddot{z}\xi + \bar{u} \tag{78}$$

The covariance of $\bar{u}$ is $V(\bar{u}) = \Omega_{\bar{u}}$ as in Equation (20). The inverse of $\Omega_{\bar{u}}$ is $\Omega_{\bar{u}}^{-1}$ as in Equation (21). Because of heteroscedasticity of $\bar{u}_g$, we consider the following GLS estimator of β_B and ξ in the between regression.

$$\begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \left(\begin{pmatrix} \bar{x}' \\ \ddot{z}' \end{pmatrix} \Omega_{\bar{u}}^{-1} (\bar{x} \quad \ddot{z}) \right)^{-1} \begin{pmatrix} \bar{x}' \\ \ddot{z}' \end{pmatrix} \Omega^{-1} \bar{y} \tag{79}$$

Substituting $\bar{y} = \bar{x}\beta_B + \ddot{z}\xi + \bar{u}$ into the above equation, the mean and variance of the GLS between estimator are

$$E \begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \begin{pmatrix} \beta_B \\ \xi \end{pmatrix} \tag{80}$$

$$V \begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix}^{-1} \tag{81}$$

The unbiasedness is derived from the exogeneity of x and z, and they are uncorrelated with group size m_g. This shows that the GLS between estimator is an unbiased and efficient estimator of (β_B, ξ).

The within regression from the CRE2 model is $\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$. All properties associated with the within regression in the CRE2 model are the same those in the CRE1 model. The covariance of the random errors is $V(\tilde{\varepsilon}) = \Omega_{\tilde{\varepsilon}}$ as in Equation (35); the within estimator is $\hat{\beta}_W = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y}$; the mean and variance of the within estimator are Equations (53) and (54), respectively; the between and within estimators are uncorrelated, i.e., $C\left(\hat{\beta}_W, \hat{\beta}_{B2G}\right) = C\left(\hat{\beta}_W, \hat{\xi}_{B2G}\right) = 0$, which can be derived from $E\left(\bar{u}_g \tilde{\varepsilon}_{gi}\right) = 0$ as in Equation (55).

The GLS estimator of the CRE2 model $Y = X\beta + \bar{X}\gamma + Z\xi + u$ is

$$\begin{pmatrix} \hat{\beta}_{C2} \\ \hat{\gamma}_{C2} \\ \hat{\xi}_{C2} \end{pmatrix} = \left(\begin{pmatrix} X' \\ \bar{X}' \\ Z' \end{pmatrix} \Omega^{-1} (X \quad \bar{X} \quad Z) \right)^{-1} \begin{pmatrix} X' \\ \bar{X}' \\ Z' \end{pmatrix} \Omega^{-1} Y \tag{82}$$

We show the following theorem for the relationship between the GLS estimator of the CRE2 model and the within and between estimators of the CRE2 model.

Theorem 2. Let $\left(\hat{\beta}_{C2}, \hat{\gamma}_{C2}, \hat{\xi}_{C2}\right)$ be the GLS estimator of the CRE2 model, and $\hat{\beta}_W$ and $\left(\hat{\beta}_{B2G}, \hat{\xi}_{B2G}\right)$ be the within estimator and the GLS between estimator of the CRE2 model, respectively. Then

$$\hat{\beta}_{C2} = \hat{\beta}_W, \hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W, \text{ and } \hat{\xi}_{C2} = \hat{\xi}_{B2G} \quad (83)$$

Wooldridge (2019) proved the first equality $\hat{\beta}_{C2} = \hat{\beta}_W$ for the CRE2 model with unbalanced panel data. Appendix A demonstrates an alternative proof of this equality using the extended Frisch-Waugh theorem. We show all equalities for cluster data in Appendix B. The equalities $\hat{\beta}_{C2} = \hat{\beta}_W$ and $\hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W$ are basically the same as Theorem 1 for the CRE1 model. The equality $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$ shows that the GLS estimator of the coefficients of group variables for the CRE2 model is numerically the same as its GLS between estimator of the between regression.

Using equalities in Equation (83) in Theorem 2 and the means, variances, and covariances of $\hat{\beta}_W$, $\hat{\beta}_{B2G}$, and $\hat{\xi}_{B2G}$ (Equations (53), (54), (80), and (81)), the means, variances, and covariances of the GLS estimator of the CRE2 model can be derived as shown in Appendix C. This implies that the inference of β, γ, and ξ can be made using the results from the between and within regressions; the GLS estimation of the CRE2 model is not necessary. Appendix C also shows the derivation of the means, variances, and covariances of the GLS estimator without using Theorem 2. This provides additional evidence of Theorem 2.

There are three implications from $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$. These three implications are related to the studies in Moulton (1986, 1990), Amemiya's (1978), and Donald and Long (2007). First, the use of the between regression based on the result of $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$ implies that the impact of z on y is only related to the data at the group level, not at the individual level. The estimation of ξ is important for panel data analysis when time-invariant variables are included in the regression (Moulton, 1986). Group variables z_g in the CRE2 model are time-invariant variables in the panel data regression. With z_g included in the regression, Moulton (1986, 1990) argued that the GLS standard errors, instead of the OLS standard errors, should be used. The results of $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$ and $V(\hat{\xi}_{C2}) = V(\hat{\xi}_{B2G})$ show that the GLS estimation of ξ in the CRE2 model with the data at the individual level is not necessary and it can be replaced by the GLS estimation of the between regression. The need for robust standard errors due to cluster data in the Moulton's model is reflected on the use of the GLS between estimator. The between regression is sufficient if the main concern of inference is on the impact of z_g. In addition, the number of degrees of freedom in inference of ξ in this regression is $n - l$ instead of $N - k - l$.

Second, based on Amemiya's (1978) random coefficient model, $\hat{\xi}_{C2}$ can be derived using a two-step procedure. The first step is to derive $\hat{\alpha}$, the fixed effects estimate of α in the LSDV regression, Equation (74); the second step is to regress $\hat{\alpha}$ on z. The estimated coefficient of z in the second step regression is the same as $\hat{\xi}_{C2}$. Our result of $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$ provides a new procedure to replace the two-step procedure. A direct estimation with the between regression gives the same numerical estimate of $\hat{\xi}_{C2}$. The variance of $\hat{\xi}_{C2}$ for inferences is also directly available from the between regression.

Third, the use of group variables z_g in the between regression implies that $(z_1, z_2, \cdots, z_G)$ is a random sample from G groups. This random sample of z_g may help us to explain the conclusions in Donald and Lang (2007). Donald and Lang (2007) used Amemiya's (1978) two-step procedure to address an issue in inference with difference-in-differences models. When time-invariant policy changes are considered, the second step regression in the two-step procedure can be used to estimate ξ. Because the number of degrees of freedom in the second step regression is $n - l$, they argued that policy implications derived from the inference of ξ with the degrees of freedom of $N - k - l$ can be misleading. They reexamined some studies on the impact of policy changes using difference-in-differences models. In these studies, data is classified into different groups due to policy changes. In this case, the number of groups created by policy changes is usually limited. In Donald and Lang's (2007) example of the two-by-two case, there are two groups: control and treatment groups. With only two groups, they concluded that the degrees of freedom in inference of ξ in the two-step procedure can be zero.

In using the two-step procedure, it is unclear if z_g has to be a random sample of groups from G groups in the second step's regression. For the between regression, the observations of z_g are considered a random sample from G groups. If a policy change can be randomly observed multiple times, i.e., n times, or it randomly affects n different groups, then the between group regression can be applied. In this case, the number of degrees of freedom is $n - l$. In Donald and Lang's example of the two-by-two case, the policy change only happened once and there is only one pair of control and treatment groups. Hence, the number of degrees of freedom issue occurs. If we consider a policy change as an experimental design and we can treat different individuals as random observations in an experimental design as in ANOVA such that there are many random observations in the designed groups, then the number of the degrees of freedom is not limited to $n - l$. Hence, the number of degrees of freedom issue is related to whether we should treat a policy change as a random event or an experimental design with many random observations in control and treatment groups. If we treat policy changes as experimental designs, then it is inappropriate to apply the between regression. The remaining issues are whether we should treat a policy change as an experimental design and what appropriate econometric methods to use if a policy change is considered as an experimental design. There are many discussions on these issues in the treatment effects literature (Imbens & Wooldridge, 2009).

4.2 Other Three Estimators Under the CRE2 Model

We continue to examine the properties of the pooled OLS estimator $\hat{\beta}_{OLS}$ and the random effects estimator $\hat{\beta}_{RE}$, and derive their relationship with the between and within estimators under the CRE2 model. Both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{RE}$ are the estimators of β assuming no group mean variables and other group variables in the model. If the CRE2 model is the true model, these two estimators are derived from a misspecified model. Both estimators may involve the omitted variable bias issue. After we show how these estimators are related to the between and within estimators, we derive the sizes of their biases under the CRE2 model.

Consider the pooled OLS estimator $\hat{\beta}_{OLS} = \lambda_{OLS}\hat{\beta}_{B1O_N} + (I - \lambda_{OLS})\hat{\beta}_W$ in Equation (59). Note that $\hat{\beta}_{B1O_N} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}$ is the OLS estimator of β_B for the regression of $\bar{Y}$ on $\bar{X}$. Under the CRE2 model, the OLS estimator of β_B and ξ in the between regression $\bar{Y} = \bar{X}\beta_B + Z\xi + \bar{U}$ is

$$\begin{pmatrix} \hat{\beta}_{B2O_N} \\ \hat{\xi}_{B2O_N} \end{pmatrix} = \begin{pmatrix} \bar{X}'\bar{X} & \bar{X}'Z \\ Z'\bar{X} & Z'Z \end{pmatrix}^{-1} \begin{pmatrix} \bar{X}'\bar{Y} \\ Z'\bar{Y} \end{pmatrix}$$

The matrix system equations can be rewritten as

$$\begin{pmatrix} \bar{X}'\bar{X} & \bar{X}'Z \\ Z'\bar{X} & Z'Z \end{pmatrix} \begin{pmatrix} \hat{\beta}_{B2O_N} \\ \hat{\xi}_{B2O_N} \end{pmatrix} = \begin{pmatrix} \bar{X}'\bar{Y} \\ Z'\bar{Y} \end{pmatrix}$$

We rewrite the first normal equation as

$$\hat{\beta}_{B2O_N} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y} - (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2O_N}$$

$$= \hat{\beta}_{B1O_N} - (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2O_N}$$

$$\hat{\beta}_{B1O_N} = \hat{\beta}_{B2O_N} + (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2O_N} \quad (84)$$

Substituting the above equation into $\hat{\beta}_{OLS}$ in Equation (59),

$$\hat{\beta}_{OLS} = \lambda_{OLS}\left(\hat{\beta}_{B2O_N} + (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2O_N}\right) + (I - \lambda_{OLS})\hat{\beta}_W$$

This shows that $\hat{\beta}_{OLS}$ is a weighted average of the within and between estimators from the CRE2 model, where the between estimators include $\hat{\beta}_{B2O_N}$ and $\hat{\xi}_{B2O_N}$. Substituting $Y = X\beta + \bar{X}\gamma + Z\xi + u$ into $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$, the mean and the variance of $\hat{\beta}_{OLS}$ are

$$E(\hat{\beta}_{OLS}) = \beta + \lambda_{OLS}\gamma + (\bar{X}'\bar{X})^{-1}\bar{X}'Z\xi$$

$$V(\hat{\beta}_{OLS}) = (X'X)^{-1}X'E(u'u)X(X'X)^{-1}$$

$$= (X'X)^{-1}\left(\sigma_\varepsilon^2\tilde{X}'\tilde{X} + \bar{X}'D_{\sigma_\varepsilon^2+m\sigma_\alpha^2}\bar{X}\right)(X'X)^{-1}$$

Because the mean of $\hat{\beta}_{OLS}$ involves ξ, $\hat{\beta}_{OLS}$ is a biased estimator of $\beta + \lambda_{OLS}\gamma$ if $\bar{x}_g$ and z_g are correlated. The size of the bias is $(\bar{X}'\bar{X})^{-1}\bar{X}'Z\xi$. As in the CRE1 model, $\hat{\beta}_{OLS}$ is inefficient. Note that $V(\hat{\beta}_{OLS})$ under the CRE2 model is different from $V(\hat{\beta}_{OLS})$ under the CRE1 model since the covariance of the random errors in the CRE2 model is different from the covariance of the random errors in the CRE1 model given that the CRE2 model is true.

For balanced panel data, we can simplify the weighting function of $\hat{\beta}_{OLS}$. Since the GLS between estimator is the same as the OLS between estimator for balanced panel data and using Theorem 2, we have $\hat{\beta}_{B2O_N} = \hat{\beta}_{B2G}$, $\hat{\xi}_{B2O_N} = \hat{\xi}_{B2G} = \hat{\xi}_{C2}$ and $\hat{\beta}_{B2O_N} - \hat{\beta}_W = \hat{\beta}_{B2G} - \hat{\beta}_W = \hat{\gamma}_{C2}$. Then

$$\hat{\beta}_{OLS} = \hat{\beta}_W + \lambda_{OLS}\hat{\gamma}_{C2} + (X'X)^{-1}X'Z\hat{\xi}_{C2}$$

This shows $\hat{\beta}_{OLS}$ is a linear function of the GLS estimator of the CRE2 model. $\hat{\beta}_{OLS}$ is still biased and inefficient estimator of $\beta + \lambda_{OLS}\gamma$ when $\bar{x}_g$ and z_g are correlated.

Next, we discuss the properties of the random effects estimator under the CRE2 model. Let Ω_r be the variance matrix of the random errors u_{gi}^* in the random effects model $Y = X\beta + u^*$. The random effects estimator is $\hat{\beta}_{RE} = (X'\Omega_r^{-1}X)^{-1}X'\Omega_r^{-1}Y$. When the CRE2 model is true, we can write u_{gi}^* as $u_{gi}^* = \bar{x}_g\gamma + z\xi - \mu_{\bar{x}}\gamma - \mu_z\xi + u_{gi}$. Assume that the variance matrix of z is $V(z)$. Then the variance and covariance of u_{gi}^* are

$$E(u_{gi}^*u_{gj}^*) = \gamma'V(\bar{x}_g)\gamma + \xi'V(z)\xi + \sigma_\alpha^2 \text{ for } i \neq j$$

$$E(u_{gi}^*u_{gi}^*) = \gamma'V(\bar{x}_g)\gamma + \xi'V(z)\xi + \sigma_\alpha^2 + \sigma_\varepsilon^2$$

$$E\left(u_{gi}^*u_{g'j}^*\right) = 0 \text{ for } g \neq g'$$

As in the CRE1 model, $\Omega_r \neq \Omega$. We then check if $\hat{\Omega}_r = \hat{\Omega}$, where $\hat{\Omega}_r$ is the estimator of Ω_r under the CRE2 model. Both random effects model and CRE2 model have the same within regression

$\tilde{y}_{gi} = \tilde{x}_{gi}\beta + \tilde{\varepsilon}_{gi}$, but have different between regressions. The between regression for the CRE2 model is Equation (78); the between regression for the random effects model is

$$\bar{y}_g = \bar{x}\beta_B + \bar{u}_g^*$$

where $\bar{u}_g^* = z_g\xi - \mu_z\xi + \bar{u}_g$. Since the estimators of Ω_r and Ω are derived from two different between regressions with $\bar{u}_g^* \neq \bar{u}_g$, it gives $\widehat{\Omega}_r \neq \widehat{\Omega}$. With $\Omega_r \neq \Omega$ and $\widehat{\Omega}_r \neq \widehat{\Omega}$, it is difficult to find the relationship between the random effects estimator and other estimators under the CRE2 model. We can still find the bias of $\hat{\beta}_{RE}$ under the CRE2 model. Substituting $Y = X\beta + \bar{X}\gamma + Z\xi + u$ into $\hat{\beta}_{RE} = (X'\Omega_r^{-1}X)^{-1}X'\Omega_r^{-1}Y$, the mean and the variance of the random effects estimator are

$$E(\hat{\beta}_{RE}) = \beta + \lambda_{RE}\gamma + (X'\Omega_r^{-1}X)^{-1}\bar{X}'\Omega_r^{-1}Z\xi$$

$$V(\hat{\beta}_{RE}) = (X'\Omega_r^{-1}X)^{-1}X'\Omega_r^{-1}E(u'u)\Omega_r^{-1}X(X'\Omega_r^{-1}X)^{-1}$$

$$= (X'\Omega_r^{-1}X)^{-1}X'\Omega_r^{-1}\Omega\,\Omega_r^{-1}X(X'\Omega_r^{-1}X)^{-1}$$

$\hat{\beta}_{RE}$ is a biased estimator of $\beta + \lambda_{RE}\gamma$ when $\bar{x}$ and z are correlated. The size of the bias under the CRE2 model is $(X'\Omega_r^{-1}X)^{-1}\bar{Z}'\Omega_r^{-1}Z\xi$. The variance of $\hat{\beta}_{RE}$ cannot be simplified into $V(\hat{\beta}_{RE}) = (X'\Omega_r^{-1}X)^{-1}$ since $\Omega_r \neq \Omega$.

Suppose the random effects model uses the same between and within regressions (Equations (78) and (34)) for the CRE2 model to estimate $\widehat{\Omega}_r$. Then $\widehat{\Omega}_r = \widehat{\Omega}$, and we can examine the relationship between $\hat{\beta}_{RE}$ and other estimators under the CRE2 model by simplifying $\widehat{\Omega}_r = \widehat{\Omega}$ into $\Omega_r = \Omega$. Using the first normal equation for the GLS estimator of the CRE2 model in Equation (B.12) in Appendix B,

$$\hat{\beta}_{RE} = \hat{\beta}_{C2} + (X'\Omega^{-1}X)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}\hat{\gamma}_{C2} + (X'\Omega^{-1}X)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z}\hat{\xi}_{C2}$$

Let $\lambda_{RE} = (X'\Omega^{-1}X)^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$. It gives

$$\hat{\beta}_{RE} = \hat{\beta}_{C2} + \lambda_{RE}\hat{\gamma}_{C2} + \lambda_{RE}(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z}\hat{\xi}_{C2} \quad (85)$$

This shows that the random effects estimator is a linear function of the GLS estimator of the CRE2 model. Using Theorem 2,

$$\hat{\beta}_{RE} = \lambda_{RE}\left(\hat{\beta}_{B2G} + (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z}\hat{\xi}_{B2G}\right) + (I - \lambda_{RE})\hat{\beta}_W$$

The random effects estimator is a matrix weighted average of the GLS between and the within estimators of the CRE2 model. Given $\Omega_r^{-1} = \Omega^{-1}$, the variance of $\hat{\beta}_{RE}$ becomes

$$V(\hat{\beta}_{RE}) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(u'u)\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= (X'\Omega^{-1}X)^{-1}$$

With $X'\Omega^{-1}X = \pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$, the variance of $\hat{\beta}_{RE}$ is related to the variances of the within estimator and the GLS between estimator for the CRE2 model as in the CRE1 model.

For balanced panel data, the relationship between $\hat{\beta}_{RE}$ and other estimators under the CRE2 model can be simplified and easily derived with $\Omega_r = \Omega$. In this case, $\hat{\beta}_{B1O_N} = \hat{\beta}_{B1G}$ and $\hat{\beta}_{B2O_N} = \hat{\beta}_{B2G}$. Substituting $\hat{\beta}_{B1O_N} = \hat{\beta}_{B2G} + (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2G}$ (Equation (84)) into $\hat{\beta}_{RE} = \lambda_{RE}\hat{\beta}_{B1O_N} + (I - \lambda_{RE})\hat{\beta}_W$ in Equation (66),

$$\hat{\beta}_{RE} = \lambda_{RE}\left(\hat{\beta}_{B2G} + (\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2G}\right) + (I - \lambda_{RE})\hat{\beta}_W$$
$$= \hat{\beta}_W + \lambda_{RE}\hat{\gamma}_{C2} + \lambda_{RE}(\bar{X}'\bar{X})^{-1}\bar{X}'Z\hat{\xi}_{B2G}$$

which is similar to Equation (85).

The random effects estimator is defined as the GLS estimator of the regression model $y_{gi} = x_{gi}\beta + u_{gi}^*$, which is a restricted CRE2 model by excluding group mean variables $\bar{x}_g$ and group variables z_g. We consider another restricted CRE2 model where only group mean variables $\bar{x}_g$ are excluded. This model is the standard Moulton model (Moulton, 1986, 1990):

$$y_{gi} = x_{gi}\beta + z_g\xi + \alpha_g^* + \varepsilon_{gi}$$

If the CRE2 model is the true model, then the Moulton model is a misspecified model because of the omission of $\bar{x}_g$. Since the omitted $\bar{x}_g$ should be included in α_g^*, let $\alpha_g^* = \bar{x}_g\gamma - \mu_{\bar{x}}\gamma + \alpha_g$. Denote the composite error as $u_{gi}^{**} = \alpha_g^* + \varepsilon_{gi}$. The variance and covariance of u_{gi}^{**} are

$$E\left(u_{gi}^{**}u_{gj}^{**}\right) = \gamma'V\left(\bar{x}_g\right)\gamma + \sigma_\alpha^2 \text{ for } i \neq j$$
$$E\left(u_{gi}^{**}u_{gi}^{**}\right) = \gamma'V\left(\bar{x}_g\right)\gamma + \sigma_\alpha^2 + \sigma_\varepsilon^2$$
$$E\left(u_{gi}^{*}u_{g'j}^{*}\right) = 0 \text{ for } g \neq g'$$

which are the same as Equations (63) – (65) for the random effects model under the CRE1 model. This shows Ω_{r2}, the covariance matrix of u_{gi}^{**}, is different from Ω for the CRE2 model. However, the between regressions for the Moulton model and the CRE2 model are the same since the omitted variables $\bar{x}_g$ in the Moulton model are automatically included in the estimation of its between regression. With the same between and within regressions for both Moulton model and CRE2 model, the estimators of Ω_{r2} and Ω are the same, i.e., $\hat{\Omega}_{r2} = \hat{\Omega}$, where $\hat{\Omega}_{r2}$ is the estimator of Ω_{r2}. Because of $\hat{\Omega}_{r2} = \hat{\Omega}$, we simplify it into $\Omega_{r2} = \Omega$ in examining the properties of the GLS estimator of (β, ξ) in the Moulton model, which is the estimator considered by Kloek (1981) and Moulton (1986), when the CRE2 model is true. The data matrix form of the Moulton model is

$$Y = X\beta + Z\xi + u^{**}$$

Since the Moulton model is the random effects model with the added group variables z, we denote the GLS estimator of the Moulton model as $\hat{\beta}_{REZ}$ and $\hat{\xi}_{REZ}$, and

$$\begin{pmatrix} \hat{\beta}_{REZ} \\ \hat{\xi}_{REZ} \end{pmatrix} = \left(\begin{pmatrix} X' \\ Z' \end{pmatrix} \Omega_{\bar{u}}^{-1} (X \quad Z) \right)^{-1} \begin{pmatrix} X' \\ Z' \end{pmatrix} \Omega^{-1} Y$$

When $\gamma = 0$, the CRE2 is reduced to the Moulton model. Then $\hat{\beta}_{REZ} = \hat{\beta}_{C2}$ and $\hat{\xi}_{REZ} = \hat{\xi}_{C2}$ from Theorem 2 and other results from the CRE2 model can be applied to the Moulton model. When $\gamma \neq 0$, $\hat{\beta}_{REZ}$ and $\hat{\gamma}_{REZ}$ can be represented in terms of $\hat{\beta}_{C2}$, $\hat{\gamma}_{C2}$, and $\hat{\xi}_{C2}$ as follows (See the proof in Appendix D).

$$\hat{\beta}_{REZ} = \hat{\beta}_{C2} + \lambda_{REZ}\hat{\gamma}_{C2} \tag{86}$$

$$\hat{\xi}_{REZ} = \hat{\xi}_{C2} + (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X(I - \lambda_{REZ})\hat{\gamma}_{C2} \tag{87}$$

where

$$\lambda_{REZ} = (X_Z^{*\prime}\Omega^{-1}X_Z^*)^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^*$$

$$I - \lambda_{REZ} = \pi_2(X_Z^{*\prime}\Omega^{-1}X_Z^*)^{-1}\tilde{X}'\tilde{X}$$

$$X_Z^* = X - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X$$

$$\bar{X}_Z^* = \bar{X}' - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}\bar{X}$$

Using Theorem 2, rewrite Equation (86) as

$$\hat{\beta}_{REZ} = \lambda_{REZ}\hat{\beta}_{B2G} + (I - \lambda_{REZ})\hat{\beta}_W$$

$\hat{\beta}_{REZ}$ is a linear function of $\hat{\beta}_W$ and $\hat{\gamma}_{C2}$, and it is a matrix weighted average of $\hat{\beta}_{B2G}$ and $\hat{\beta}_W$. Based on Equations (70) and (71), $\hat{\beta}_{REZ}$ is the unbiased estimator of $\beta + \lambda_{REZ}\gamma$, but $\hat{\beta}_{REZ}$ is not efficient and not a MMSE estimator of $\beta + \lambda_{REZ}\gamma$.

Both $\hat{\beta}_{REZ}$ and $\hat{\xi}_{REZ}$ in Equations (86) and (87) include the component of $\hat{\gamma}_{C2}$, which is the estimate of the second partial effect of x. When $\bar{x}$ and z are uncorrelated, λ_{REZ} is reduced to λ_{RE} and the second term of $\hat{\xi}_{REZ}$ is reduced to zero in expectation. Then $E(\hat{\beta}_{REZ}) = E(\hat{\beta}_{RE})$ and $E(\hat{\xi}_{REZ}) = E(\hat{\xi}_{C2}) = \xi$. When $\bar{x}$ and z are correlated, $E(\hat{\beta}_{REZ}) \neq E(\hat{\beta}_{RE})$ and $E(\hat{\xi}_{REZ}) \neq \xi$. The second term of $\hat{\xi}_{REZ}$ in Equation (87) is the size of bias caused by the correlation between $\bar{x}$ and z. Hence, the use of the Mouton model may lead to an incorrect conclusion on the coefficient of ξ when $\gamma \neq 0$ and $\bar{x}$ and z are correlated.

Based on the results on the six estimators under the CRE2 model, we derive the same main conclusion as those from the CRE1 model; we conclude that the key difference between the within estimator and other estimators is whether we should treat α_g as a fixed parameter or a random component. The estimation of the within estimator assumes α_g to be fixed while the estimations of all other estimators assume α_g to be random. The advantage of assuming α_g to be random is to estimate γ and ξ, the partial effects associated with group mean variables and other group variables. If the main concern of the estimation is ξ, the partial effect of group variables, we only need the between regression. There is no need for the CRE2 model or the Moulton model since the GLS estimator of ξ in the CRE2 model is the same as its GLS between estimator. If the concern of the study is about both x and z, the explanatory variables observed at individual and group levels, then the issue is whether γ or the second partial effect is zero or not. When $\gamma = 0$, we use the within regression to estimate β and the between regression to estimate ξ. When $\gamma \neq 0$, we should consider the CRE2 model or its between and within regressions. Note that both pooled OLS and random effects estimators are biased for $\beta + \lambda\gamma$ and the GLS estimator of ξ in the Moulton model is biased for ξ when $\bar{x}$ is correlated to z. In addition, the estimation of any model considering α_g as random rather than a fixed parameter may suffer from omitted variable bias since the included variables $\bar{x}$ and z may be correlated with the unobserved group random component α_g.

5. Conclusion

Correlated random effects models extend classical panel data regressions with group mean variables and other group variables. Both correlated random effects models and panel data regressions involve some basic estimators. This paper provides a complete analysis on the properties of six basic estimators and examines the relationships among these estimators under correlated random effects models with cluster data. Our study follows Maddala (1971) and Mundlak (1978) to explore the properties of the estimators such as the pooled OLS, within, between, random effects, and GLS estimators. These estimators are linked by the assumptions of random errors of the models.

Our methodology in analyzing different estimators differs from the literature in three aspects. Firstly, we consider the between and within regressions as fundamental to the analysis of the relationships among different estimators. Previous studies primarily consider the between and within estimators as given, without examining the detailed properties of the two regressions generating these two estimators. Instead, by recognizing two matrix forms of the between regression, we establish the relationships of different estimators based on the variances of the random errors and the estimators of two different forms of the between regression and the within regression. Secondly, we adopt a new and simple approach to derive the theoretic relationship among different estimators. This approach involves matrix equivalences and the typical process in solving a system of equations as in the Frisch-Waugh theorem. It is worth noting that the derivation of matrix equivalences is based on the two forms of between regression and the within regression. Thirdly, in exploring the properties of the different estimators, we emphasize that there are two different partial effects of an explanatory variable. The first partial effect is the direct impact of the explanatory variable, and second partial effect is the impact from the group mean of the explanatory variable. The first partial effect is the main concern in most theoretic and empirical studies since the direct impact of the explanatory variable is equal to the fixed effect or the within effect of CRE models. Empirically, one interpretation of the second partial effect is the network or peer effect in social interactions models (Manski, 1993, 2000; Blume, et al. 2015). Theoretically, the second partial effect plays an important role in explaining the differences among different estimators once we accept the existence of the second partial effect.

Our two theoretic contributions are 1) We extend Mundlak's (1978) results to CRE models with cluster data. Mundlak's main results with balanced panel data show the relationships of the GLS estimator with the between and within estimators. Most balanced panel data analysis can be applied trivially to models with cluster data. It is still necessary to derive a formal theorem if Mundlak's results can be extended to models with cluster data. Wooldridge (2019) provided a partial extension of Mundlak's results. Our theorems complete the extension and verify that the relationships under CRE models with cluster data are the same as those under the Mundlak model. The only adjustment is the OLS between estimator is replaced by the GLS between estimator in the relationships. We show that the relationships continue to hold for models with the addition of other group variables, i.e., additional group variables do not affect how the GLS estimator relates to the two fundamental estimators. Furthermore, we found that the coefficients of group variables can be estimated by the between regression. 2) We show the properties of six different estimators, including their means and variances. These results help us to compare different estimators.

Based on our theoretical analysis, we summarize the results and empirical implications as follows. First, the analysis of CRE models allows us to revisit related previous studies. When we apply our analysis of the basic CRE model to balanced panel data, we derive some different conclusions from Maddala (1971) and Mundlak (1978). For the extended CRE model, our results on the GLS estimator of the coefficients of group variables are related to the studies in Moulton (1986, 1990), Amemiya (1978), and Donald and Lang (2007). One of the two methods in proving our theorems is related to Lovell (1963). The following is a summary of our discussions and results in these revisits.

a. We found that Maddala (1971) has an error in interpreting the random effects estimator. We show that the random effects estimator is not equal to the within estimator when the number of time periods is infinite.
b. We found that the weighting matrices of the random effects estimator under the Mundlak model (Mundlak, 1978) and the standard random effects regression (Maddala, 1971) are the same. Hence, the random effects estimator has the same properties under two different modeling approaches. This result is not observed in Mundlak (1978).
c. Mundlak (1978) used his model to show that there is only one estimator to estimate partial effects of explanatory variables. Instead, we argue that there are multiple MMSE estimators if the second partial effect is taken into the consideration.
d. Moulton's (1986) concern on the robust estimation of the coefficients of group variables in a panel data model can be resolved using the between regression. In this case, the number of degrees of freedom in inferences is related to the number of cross-sectional units, instead of the total number of data values in the sample.
e. Amemiya's (1978) two-step procedure to estimate the coefficients of group variables can be replaced by the between regression. The statistical inferences of these coefficients are directly applicable with the between regression.
f. Donald and Lang's (2007) critique on the use of difference-in-differences models is related to how we treat a policy change in modelling; is it an experimental design with control and treated groups or an experiment generating a random sample of

groups? If a policy change is associated with a random sample of groups, then the between regression can be applied.

g. Instead of using the standard Frisch-Waugh Theorem as described in Lovell (1963), we introduce an extended version of Frisch-Waugh Theorem for models with the GLS estimator. We show an application of this extended theorem.

The above summary of our discussions and results can be divided into three categories: new results (items d, e, g), different results (items a, b), and different interpretations of the same results (items c and f).

Second, we summarize the implications from different modeling and estimation strategies used by different estimators. In our comparison of different estimators, we found that the main difference between the within estimator and other estimators is the role of the second partial effect of an explanatory variable, and the main difference in modelling strategies for different estimators is the presumption of a fixed effects parameter. If the sole concern of the regression is the first partial effect or the within effect of the explanatory variables, then the use of the within estimator is sufficient, and there is no need for other estimators and no need for CRE models nor the random effects model. If the second partial effect is not zero, CRE models may provide additional information about the partial effects of explanatory variables. Furthermore, CRE models provide the prediction based on both group means and within group deviations of the dependent variable. The random effects estimator and the GLS estimator of the Moulton model are biased if the group mean variables are correlated to observed group variables. For the inference of the coefficients of group variables, the use of between regression is sufficient.

Although our analysis and results are for cluster data, the same can be applied to balanced and unbalanced panel data models with one-way fixed effects in the cross-sectional or time domain. Two improvements on this study can be considered in future research. First, our analysis assumes that basic random errors are homoscedastic and uncorrelated. This assumption can be relaxed. Second, the analysis with cluster data focuses only on the cross-sectional domain of panel data. A future study can analyze different estimators for two-way fixed effects models such that group characteristics for both cross-sectional and time domains are considered.

References:

Abrevaya, J. (2013). The Projection Approach for Unbalanced Panel Data. *Econometrics Journal* 16, 161 – 178.

Ahn, S.C., Moon, H.R. (2014). Large-N and Large-T Properties of Panel Data Estimators and the Hausman Test. In: Sickles R., Horrace W. (eds) *Festschrift in Honor of Peter Schmidt*. Springer, New York, NY. https://doi.org/10.1007/978-1-4899-8008-3_7

Amemiya, T. (1978). A Note on a Random Coefficients Model. *International Economic Review* 19(3), 793-796.

Baltagi, B. H. (2006). An Alternative Derivation of Mundlak's Fixed Effects Results Using System Estimation. *Econometric Theory* 22(6), 1191-1194.

Baltagi, B. H. (2021). *Economic Analysis of Panel Data*, 6th Edition. John Wiley & Sons Ltd. West Sussex, United Kingdom.

Blume, L.E., Brock, W.A., Durlauf, S.N., & Jayaraman, R. (2015). Linear Social Interactions Models. *Journal of Political Economy* 123(2), 444–496.

Chamberlain, G. (1975). Unobservables in Econometric Models, unpublished Ph.D. dissertation, Harvard University, Cambridge.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* 47(1), 225-238.

Donald, S. G., & Lang, K. (2007). Inference with Difference-In-Differences and Other Panel Data. *The Review of Economics and Statistics* 89(2), 221–233.

Fiebig, D. G., Bartels, R. & Kramer, W. (1996). The Frisch-Waugh Theorem and Generalized Least Squares. *Econometric Reviews*, 15(4), 431-443.

Frisch, R. & Waugh, F.V. (1933). Partial Time Regressions as Compared with Individual Trends *Econometrica* 1(4), 387-401.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*, 2nd Edition. Wardsworth Statistics/Probability Series.

Greene, W. H. (2018). *Econometric Analysis*, 8th Edition. Pearson Education, Inc. New York, NY.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica* 46(6), 1251-1271.

Hsiao, C. (2014). *Analysis of Panel Data*, 3nd ed. In: Econometric Society Monographs, vol. 54, Cambridge University Press, Cambridge.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* 47(1), 5-86.

Kloek, T. (1981). OLS Estimation in a Model where a Microvariable is Explained by Aggregates and Contemporaneous Disturbances are Equicorrelated. *Econometrica* 49(1), 205–207.

Lovell, M. C. (1963). Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis. *Journal of the American Statistical Association* 58(304), 993-1010.

Lovell, M. C. (2008). A Simple Proof of the FWL Theorem. *The Journal of Economic Education* 39(1), 88-91

Maddala, G. S. (1971). The Use of Variance Components Models in Pooling Cross Section and Time Series Data. *Econometrica* 39(2), 341–358.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60(3), 531–542.

Manski, C. F. (2000). Economic Analysis of Social Interactions. *Journal of Economic Perspective* 14(3), 115–136.

Moulton, B. R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics* 32, 385–397.

Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *The Review of Economics and Statistics* 72, 334–338.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data, *Econometrica* 46(1), 69–85.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2th Edition. The MIT Press: Cambridge, MA.

Wooldridge, J. M. (2019). Correlated Random Effects Models with Unbalanced Panels. *Journal of Econometrics* 211(1), 137-150.

Table 1. The Means and Variances of the Estimators Under the CRE1 Model

	Estimator $\hat{\beta}_\lambda = \lambda\hat{\beta}_{B1G} + (I-\lambda)\hat{\beta}_W$ $= \hat{\beta}_W + \lambda\hat{\gamma}_{C1}$	**Weights** $\lambda = (A+B)^{-1}A$ $I-\lambda = (A+B)^{-1}B$	**Mean**	**Variance**
	$\hat{\beta}_{B1O_n} = (\bar{x}'\bar{x})^{-1}\bar{x}'\bar{y}$	(NA)	$\beta+\gamma$	$(\bar{x}'\bar{x})^{-1}\bar{x}'d_{\sigma_{\bar{u}}}\bar{x}(\bar{x}'\bar{x})^{-1}$
	$\hat{\beta}_{B1O_N} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}$	(NA)	$\beta+\gamma$	$(\bar{X}'\bar{X})^{-1}\bar{X}'\Omega_{\bar{U}}\bar{X}(\bar{X}'\bar{X})^{-1}$
(i)	$\hat{\beta}_{B1G} = (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$	$\lambda = I$	$\beta+\gamma$	$(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}$
(ii)	$\hat{\beta}_w = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y}$	$\lambda = 0$	β	$\sigma_\varepsilon^2\left(\tilde{X}'\tilde{X}\right)^{-1}$
(iii)	$\hat{\beta}_{C1} = \hat{\beta}_W$ $\hat{\gamma}_{C1} = \hat{\beta}_{B1G} - \hat{\beta}_W$	(NA)	β γ	$\sigma_\varepsilon^2\left(\tilde{X}'\tilde{X}\right)^{-1}$ $(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1} + \sigma_\varepsilon^2\left(\tilde{X}'\tilde{X}\right)^{-1}$
(vi)	$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$ $= \hat{\beta}_W + \lambda_{OLS}\hat{\gamma}_{C1}$	$A = \bar{X}'\bar{X}$ $B = \tilde{X}'\tilde{X}$ $A+B = X'X$	$\beta + \lambda_{OLS}\gamma$	$(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$
(v)	$\hat{\beta}_{RE} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ $= \hat{\beta}_W + \lambda_{RE}\hat{\gamma}_{C1}$	$A = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ $B = \sigma_\varepsilon^{2^{-1}}\tilde{X}'\tilde{X}$ $A+B = X'\Omega^{-1}X$	$\beta + \lambda_{RE}\gamma$	$(X'\Omega^{-1}X\,)^{-1} =$ $\left(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} + \sigma_\varepsilon^{2^{-1}}\tilde{X}'\tilde{X}\right)^{-1}$

Notes: There are three different between estimators under the CRE1 model with cluster data. Both $\hat{\beta}_{B1O_n}$ and $\hat{\beta}_{B1O_n}$ are the OLS estimators, where $\hat{\beta}_{B1O_n}$is based on $\bar{y}$ and $\bar{x}$ and $\hat{\beta}_{B1O_N}$is based on $\bar{Y}$ and $\bar{X}$. $\hat{\beta}_{B1G}$ is the GLS between estimator. $\hat{\beta}_{B1G}$ and $\hat{\beta}_W$ in (i) and (ii) are two fundamental estimators. The remaining three estimators are related to these two estimators. The $\hat{\beta}_{C1}$ and $\hat{\gamma}_{C1}$ in (iii) are the GLS estimator of β and γ in the CRE1 model. λ is the weight assigned to γ, the second partial effect of x.

Appendices

Appendix A. The Extended Frisch-Waugh Theorem and An Application

Appendix A is organized as follows: First, we introduce an extended version of the Frisch-Waugh theorem (Frisch & Waugh, 1933). The standard theorem is for the OLS estimator while the extended theorem is for the GLS estimator. Second, we use the extended theorem to prove the GLS estimator of β for CRE models with cluster data is the same as the within estimator.

Part I. The Extended Frisch-Waugh Theorem

The standard Frisch-Waugh theorem is introduced in some econometrics textbooks (Lovell, 2008); Greene (2018, pp. 35 – 37) provides a good review of the theorem. Consider the following regression with two sets of variables X_1 and X_2 and parameters β_1 and β_2.

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{A.1}$$

Let $\hat{\beta}_{1,OLS}$ and $\hat{\beta}_{2,OLS}$ be the OLS estimators of β_1 and β_2, respectively. The theorem shows that

$$\hat{\beta}_{1,OLS} = (X_1^* X_1^*)^{-1} X_1^{*\prime} Y_{X_2}^*$$

$$\hat{\beta}_{2,OLS} = (X_2^* X_2^*)^{-1} X_2^{*\prime} Y_{X_1}^*$$

where X_i^* are the residuals from the regression of X_i on X_j, and $Y_{X_i}^*$ are the residuals from the regression of Y on X_i, with $i, j = 1,2$ and $i \neq j$. One constraint of the theorem is that it is designed for models with the OLS estimator. When the random errors ε are heteroscedastic and serial correlated, the OLS estimator is inefficient, and the process of applying the theorem needs to be modified. Suppose the covariance matrix of the random errors ε is

$$V(\varepsilon) = \Omega$$

A method in applying the theorem with any Ω is to convert the random errors and the covariance matrix of ε as

$$V(Q\varepsilon) = \sigma^2 I$$

where σ^2 is a constant and Q is a conversion matrix, such that $Q\Omega Q' = \sigma^2 I$ and $Q'Q = \Omega^{-1}$. We can then use Q to convert the variables Y, X_1 and X_2 and apply the standard Frisch-Waugh theorem (Lovell, 1963; Chamberlian, 1980, p. 234; Fiebig et al., 1996; Wooldridge, 2019). This appendix extends this standard theorem to general models where the GLS estimator is used for heteroscedastic and serial correlated random errors. With this extended theorem, there is no need for transformation of variables. Denote the GLS estimator of (β_1, β_2) in Equation (A.1) as $(\hat{\beta}_1, \hat{\beta}_2)$. Then

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \Omega^{-1} (X_1 \; X_2) \right)^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \Omega^{-1} Y \tag{A.2}$$

Or,

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'\Omega^{-1}X_1 & X_1'\Omega^{-1}X_2 \\ X_2'\Omega^{-1}X_1 & X_2\Omega^{-1}X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1'\Omega^{-1}Y \\ X_2'\Omega^{-1}Y \end{pmatrix} \tag{A.3}$$

The extended Frisch-Waugh theorem is that the partial coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are

$$\hat{\beta}_1 = (X_1^*\Omega^{-1}X_1^*)^{-1}X_1^{*\prime}\Omega^{-1}Y_{X_2}^* \tag{A.4}$$

$$\hat{\beta}_2 = (X_2^*\Omega^{-1}X_2^*)^{-1}X_2^{*\prime}\Omega^{-1}Y_{X_1}^* \tag{A.5}$$

where X_i^* are the residuals from the GLS estimation of the regression of X_i on X_j, $Y_{X_i}^*$ are the residuals from the GLS estimation of the regression of Y on X_i, with $i, j = 1,2$ and $i \neq j$. For example,

$$X_1^* = X_1 - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X_1, \text{ with } \hat{\beta}_{X_1X_2} = (X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X_1$$

$$Y_{X_2}^* = Y - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y, \text{ with } \hat{\beta}_{YX_2} = (X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y$$

The following is the proof of this extended theorem.

Rewrite the system of equations of the GLS estimator, Equation (A.3), as

$$\begin{pmatrix} X_1'\Omega^{-1}X_1 & X_1'\Omega^{-1}X_2 \\ X_2'\Omega^{-1}X_1 & X_2'\Omega^{-1}X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'\Omega^{-1}Y \\ X_2'\Omega^{-1}Y \end{pmatrix}$$

The normal equations are

$$(X_1'\Omega^{-1}X_1)\hat{\beta}_1 + (X_1'\Omega^{-1}X_2)\hat{\beta}_2 = X_1'\Omega^{-1}Y \tag{A.6}$$

$$(X_2'\Omega^{-1}X_1)\hat{\beta}_1 + (X_2'\Omega^{-1}X_2)\hat{\beta}_2 = X_2'\Omega^{-1}Y \tag{A.7}$$

Rewrite Equation (A.7) as

$$\hat{\beta}_2 = (X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y - (X_2'\Omega^{-1}X_2)^{-1}(X_2'\Omega^{-1}X_1)\hat{\beta}_1 \tag{A.8}$$

Substituting Equation (A.8) into Equation (A.6) and rearranging terms,

$$X_1'\Omega^{-1}(I - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1})X_1\hat{\beta}_1$$
$$= X_1'\Omega^{-1}(I - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1})Y \tag{A.9}$$

Let $M_{X_2} = I - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}$. Then $X_1^* = M_{X_2}X_1$, and $Y_{X_2}^* = M_{X_2}Y$. Equation (A.9) can be written as

$$\hat{\beta}_1 = (X_1'\Omega^{-1}X_1^*)^{-1}X_1'\Omega^{-1}Y_{X_2}^* \tag{A.10}$$

Note $X_1'\Omega^{-1}X_1^* = X_1'\Omega^{-1}X_1 - X_1'\Omega^{-1}X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X_1$ is nonsingular since $\begin{pmatrix} X_1'\Omega^{-1}X_1 & X_1'\Omega^{-1}X_2 \\ X_2'\Omega^{-1}X_1 & X_2\Omega^{-1}X_2 \end{pmatrix}$ in Equation (A.3) is nonsingular (Graybill, 1983, p. 184). Using $M_{X_2} = M_{X_2}M_{X_2}$, rewrite $X_1'\Omega^{-1}X_1^*$ and $X_1'\Omega^{-1}Y_{X_2}^*$ in Equation (A.10) as

$$X_1'\Omega^{-1}X_1^* = X_1'\Omega^{-1}M_{X_2}X_1^* \text{ and } X_1'\Omega^{-1}Y_{X_2}^* = X_1'\Omega^{-1}M_{X_2}Y_{X_2}^* \tag{A.11}$$

Consider $X_1'\Omega^{-1}M_{X_2}$ in the above equations.

$$\begin{aligned} X_1'\Omega^{-1}M_{X_2} &= X_1'\Omega^{-1}(I - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}) \\ &= X_1'\Omega^{-1} - X_1'\Omega^{-1}X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1} \\ &= (X_1' - X_1'\Omega^{-1}X_2(X_2'\Omega^{-1}X_2)^{-1}X_2')\Omega^{-1} \\ &= (X_1 - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X_1)'\Omega^{-1} \\ &= X_1^{*\prime}\Omega^{-1} \end{aligned} \tag{A.12}$$

Substituting Equation (A.12) into Equation (A.11),

$$X_1'\Omega^{-1}X_1^* = X_1^{*\prime}\Omega^{-1}X_1^* \text{ and } X_1'\Omega^{-1}Y_{X_2}^* = X_1^{*\prime}\Omega^{-1}Y_{X_2}^* \tag{A.13}$$

Substituting the equations in (A.13) into Equation (A.10), Equation (A.10) becomes Equation (A.4). This proves the partial coefficient $\hat{\beta}_1$. The partial coefficient $\hat{\beta}_2$ in Equation (A.5) can be derived similarly, and the extended theorem is proved.

There is a limitation in using this theorem empirically. Note that Ω is the covariance matrix of the random errors in $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$. The partial regression of Y on X_i and the regression of X_i on X_j are considered misspecified model under $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$. The estimator of Ω from partial regressions can be biased. Hence, Ω cannot be estimated by partial regressions. The extended theorem can still be applied to the models using the feasible GLS estimator if Ω is appropriately estimated by $\widehat{\Omega}$ and $\widehat{\Omega}$ is used in partial regressions.

Part II. Proof of $\widehat{\boldsymbol{\beta}}_{C1} = \widehat{\boldsymbol{\beta}}_W$ and $\widehat{\boldsymbol{\beta}}_{C2} = \widehat{\boldsymbol{\beta}}_W$

In part II, we apply the extended Frisch-Waugh theorem to prove that the GLS estimator of β for CRE models with cluster data is the same as the within estimator (the first equality in Theorem 1, Equation (57), and the first equality in Theorem 2, Equation (83)). The proof provided here is similar to the proof of Proposition 2.1 in Wooldridge (2019). While the proof by Wooldridge (2019) used the standard Frisch-Waugh theorem, we apply the extended Frisch-

Waugh theorem and matrix equivalences, and show the proof with details in the matrix algebra. In the following, matrix equivalences of $\bar{X}'\Omega^{-1}X = \bar{X}'\Omega^{-1}\bar{X}$, $\tilde{X}'\Omega^{-1}\tilde{X} = \pi_2\tilde{X}'\tilde{X}$, and $\tilde{X}'\Omega^{-1}Y = \tilde{X}'\Omega^{-1}\tilde{Y} = \pi_2\tilde{X}'\tilde{Y}$ are used in the derivation of some equations. Using $\Omega^{-1} = \pi_2 I + D_{m\pi_1}\bar{J}$, $\bar{X}'X = \bar{X}'\bar{J}X = \bar{X}'\bar{X}$, and $\bar{J}\tilde{X} = 0$, we derive the following matrix equivalences.

$$\bar{X}'\Omega^{-1}X = \bar{X}'\left(\pi_2 I + D_{m\pi_1}\bar{J}\right)X = \bar{X}'\left(\pi_2 I + D_{m\pi_1}\bar{J}\right)\bar{X} = \bar{X}'\Omega^{-1}\bar{X} \tag{A.14}$$

$$\tilde{X}'\Omega^{-1}\tilde{X} = \tilde{X}'\left(\pi_2 I + D_{m\pi_1}\bar{J}\right)\tilde{X} = \pi_2\tilde{X}'\tilde{X} + \tilde{X}'D_{m\pi_1}\bar{J}\tilde{X} = \pi_2\tilde{X}'\tilde{X} \tag{A.15}$$

Similarly, $\tilde{X}'\Omega^{-1}Y = \pi_2\tilde{X}'\tilde{Y}$ using $\tilde{X}'Y = \tilde{X}'MY = \tilde{X}'\tilde{Y}$.

Consider the CRE1 model $Y = X\beta + \bar{X}\gamma + u$ with $V(u) = \Omega$. Based on the extended Frisch-Waugh theorem, the GLS estimator of β is

$$\hat{\beta}_{C1} = (X_{\bar{X}}^{*\prime}\Omega^{-1}X_{\bar{X}}^{*})^{-1}X_{\bar{X}}^{*\prime}\Omega^{-1}Y_{\bar{X}}^{*} \tag{A.16}$$

where

$$X_{\bar{X}}^{*} = X - \bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}X \tag{A.17}$$

$$Y_{\bar{X}}^{*} = Y - \bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}Y \tag{A.18}$$

Using $\bar{X}'\Omega^{-1}X = \bar{X}'\Omega^{-1}\bar{X}$, Equation (A.17) is simplified as $X_{\bar{X}}^{*} = X - \bar{X} = \tilde{X}$. Substituting $X_{\bar{X}}^{*} = \tilde{X}$ into Equation (A.16),

$$\hat{\beta}_{C1} = \left(\tilde{X}'\Omega^{-1}\tilde{X}\right)^{-1}\tilde{X}'\Omega^{-1}Y_{\bar{X}}^{*} \tag{A.19}$$

Using $Y_{\bar{X}}^{*}$ defined in Equation (A.18), rewrite $\tilde{X}'\Omega^{-1}Y_{\bar{X}}^{*}$ in the above equation as

$$\begin{aligned}\tilde{X}'\Omega^{-1}Y_{\bar{X}}^{*} &= \tilde{X}'\Omega^{-1}(Y - \bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}Y)\\ &= \tilde{X}'\Omega^{-1}Y - \tilde{X}'\Omega^{-1}\bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}Y\\ &= \tilde{X}'\Omega^{-1}\tilde{Y}\end{aligned} \tag{A.20}$$

To derive the last equation, we use $\tilde{X}'\Omega^{-1}\bar{X} = 0$ since $\tilde{X}'\bar{J} = 0$ and $\tilde{X}'\bar{X} = \tilde{X}'J\bar{X} = 0$. Substituting Equation (A.20) into Equation (A.19) and using the matrix equivalences as in Equation (A.15),

$$\hat{\beta}_{C1} = \left(\tilde{X}'\Omega^{-1}\tilde{X}\right)^{-1}\tilde{X}'\Omega^{-1}\tilde{Y} = \left(\pi_2\tilde{X}'\tilde{X}\right)^{-1}\pi_2\tilde{X}'\tilde{Y} = \hat{\beta}_W \tag{A.21}$$

This proves the first equality of Theorem 1.

However, we cannot use the extended theorem to prove the second equality of Theorem 1. Based on the extended theorem,

$$\hat{\gamma}_{C1} = (\bar{X}_X^{*\prime}\Omega^{-1}\bar{X}_X^{*})^{-1}\bar{X}_X^{*\prime}\Omega^{-1}Y_X^{*}$$

where

$$\bar{X}_X^* = \bar{X} - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\bar{X}$$

$$Y_X^* = Y - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

With this definition of the partial coefficient of γ, we cannot derive $\hat{\gamma}_{C1} = \hat{\beta}_B - \hat{\beta}_W$ because $(X'\Omega^{-1}X)^{-1}$ in $\bar{X}_X^*$ and Y_X^* cannot be simplified.

For the CRE2 model $Y = X\beta + \bar{X}\gamma + Z\xi + u$ with $V(u) = \Omega$ and $\hat{\beta}_{C2}$ as the GLS estimator of β, we also use the extended theorem to prove $\hat{\beta}_{C2} = \hat{\beta}_W$. Define $X_2 = (\bar{X}\ Z)$ and $\beta_2 = (\gamma'\ \xi')'$. Then we apply the extended theorem to β in $Y = X\beta + X_2\beta_2 + u$. The GLS estimator of β is

$$\hat{\beta}_{C2} = \left(X_{X_2}^{*\prime}\Omega^{-1}X_{X_2}^*\right)^{-1}X_{X_2}^{*\prime}\Omega^{-1}Y_{X_2}^* \tag{A.22}$$

where

$$X_{X_2}^* = X - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X \tag{A.23}$$

$$Y_{X_2}^* = Y - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y \tag{A.24}$$

Equation (A.23) is the residual equation for the GLS estimation of the partial regression of X on X_2. We can apply the extended theorem again to this partial regression. Consider the regression of X on $X_2 = (\bar{X}\ Z)$ as

$$X = X_2(a'\ b')' + u_{X_2} = \bar{X}a + Zb + u_{X_2} \tag{A.25}$$

where a and b are parameters. In applying the extended theorem, we let $V\left(u_{X_2}\right) = \Omega$. The GLS estimator of (a, b) is $(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}X$, which can also be derived from the extended theorem as:

$$\hat{a} = (\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^*)^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}X_Z^* \tag{A.26}$$

$$\hat{b} = (Z_{\bar{X}}^{*\prime}\Omega^{-1}Z_{\bar{X}}^*)^{-1}Z_{\bar{X}}^{*\prime}\Omega^{-1}X_{\bar{X}}^* \tag{A.27}$$

where

$$\bar{X}_Z^* = \bar{X} - Z(Z'\Omega^{-1}Z')^{-1}Z'\Omega^{-1}\bar{X} \tag{A.28}$$

$$X_Z^* = X - Z(Z'\Omega^{-1}Z')^{-1}Z'\Omega^{-1}X \tag{A.29}$$

$$Z_{\bar{X}}^* = Z - \bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}Z \tag{A.30}$$

$$X_{\bar{X}}^* = X - \bar{X}(\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}X \tag{A.31}$$

We show that $\hat{a} = I$ and $\hat{b} = 0$. Consider $\bar{X}_Z^{*\prime}\Omega^{-1}X_Z^*$ in $\hat{a}$ (Equation (A.26)). Using $\bar{X}_Z^{*\prime}\Omega^{-1}X_Z^* = \bar{X}'\Omega^{-1}X_Z^*$ from Equation (A.13) and substituting X_Z^* in Equation (A.29) into $\bar{X}'\Omega^{-1}X_Z^*$,

$$\bar{X}_Z^{*\prime}\Omega^{-1}X_Z^* = \bar{X}'\Omega^{-1}X_Z^*$$

$$= \bar{X}'\Omega^{-1}X - \bar{X}'\Omega^{-1}(Z'\Omega^{-1}Z')^{-1}Z'\Omega^{-1}X$$
$$= \bar{X}'\Omega^{-1}\bar{X} - \bar{X}'\Omega^{-1}(Z'\Omega^{-1}Z')^{-1}Z'\Omega^{-1}\bar{X}$$
$$= \bar{X}'\Omega^{-1}\bar{X}_Z^* = \bar{X}_Z^{*'}\Omega^{-1}\bar{X}_Z^* \quad (A.32)$$

To derive the above equation, we use the matrix equivalence of $Z'\Omega^{-1}X = Z'\Omega^{-1}\bar{X}$ as in Equation (A.14). Hence, $\hat{a} = (\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^*)^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^* = I$. Using Equation (A.13), $X_{\bar{X}}^* = X - \bar{X} = \tilde{X}$ from Equation (A.31), and $Z'\tilde{X} = Z'\bar{J}\tilde{X} = 0$, rewrite $Z_{\bar{X}}^{*'}\Omega^{-1}X_{\bar{X}}^*$ in $\hat{b}$ (Equation (A.27)) as

$$Z_{\bar{X}}^{*'}\Omega^{-1}X_{\bar{X}}^* = Z'\Omega^{-1}\tilde{X} = 0$$

Hence, $\hat{b} = 0$. The fitted equation of $X = \bar{X}a + Zb + u_{X_2}$ is $\bar{X}\hat{a} + Z\hat{b}$, and the residual equation is

$$X_{X_2}^* = X - (\bar{X}\hat{a} + Z\hat{b}) = X - \bar{X} = \tilde{X} \quad (A.33)$$

Substituting $X_{X_2}^* = \tilde{X}$ into $\hat{\beta}_{C2}$ in Equation (A.22),

$$\hat{\beta}_{C2} = (\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}Y_{X_2}^* \quad (A.34)$$

Using $Y_{X_2}^*$ defined in Equation (A.24), rewrite $\tilde{X}'\Omega^{-1}Y_{X_2}^*$ in the above equation as

$$\tilde{X}'\Omega^{-1}Y_{X_2}^* = \tilde{X}'\Omega^{-1}(Y - X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y)$$
$$= \tilde{X}'\Omega^{-1}Y - \tilde{X}'\Omega^{-1}X_2(X_2'\Omega^{-1}X_2)^{-1}X_2'\Omega^{-1}Y$$
$$= \tilde{X}'\Omega^{-1}\tilde{Y}$$

To derive the last equation, we use $\tilde{X}'\Omega^{-1}X_2 = 0$ since $\tilde{X}'\Omega^{-1}\bar{X} = 0$ and $\tilde{X}'\Omega^{-1}Z = 0$. Using the matrix equivalence as in Equation (A.15), Equation (A.34) becomes

$$\hat{\beta}_{C2} = (\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}\tilde{Y} = (\pi_2\tilde{X}'\tilde{X})^{-1}\pi_2\tilde{X}'\tilde{Y} = \hat{\beta}_W \quad (A.35)$$

Q.E.D.

Appendix B. Proof of the Theorems

This appendix provides the proof of Theorems 1 and 2 without using the matrix inversion nor the Frisch-Waugh theorem. The method is based on the typical process in solving a system of equations and matrix equivalences.

Proof of Theorem 1:

Consider the CRE1 model

$$Y = X\beta + \bar{X}\gamma + u, V(u) = \Omega \quad (B.1)$$

Let $(\hat{\beta}_{C1}, \hat{\gamma}_{C1})$ be the GLS estimator of the parameters in the CRE1 model, and $\hat{\beta}_W$ and $\hat{\beta}_{B1G}$ be the within estimator and the GLS between estimator of the CRE1 model, respectively. Theorem 1 states that

$$\hat{\beta}_{C1} = \hat{\beta}_W \text{ and } \hat{\gamma}_{C1} = \hat{\beta}_{B1G} - \hat{\beta}_W \tag{B.2}$$

The following is the proof of this theorem.

The GLS estimator of (β, γ) in the CRE1 model, Equation (B.1), is

$$\begin{pmatrix} \hat{\beta}_{C1} \\ \hat{\gamma}_{C1} \end{pmatrix} = \begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} \\ \bar{X}'\Omega^{-1}X & \bar{X}\Omega^{-1}\bar{X} \end{pmatrix}^{-1} \begin{pmatrix} X'\Omega^{-1}Y \\ \bar{X}'\Omega^{-1}Y \end{pmatrix}$$

Rewrite this matrix system of equations as

$$\begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} \\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{C1} \\ \hat{\gamma}_{C1} \end{pmatrix} = \begin{pmatrix} X'\Omega^{-1}Y \\ \bar{X}'\Omega^{-1}Y \end{pmatrix}$$

The normal equations are

$$(X'\Omega^{-1}X)\hat{\beta}_{C1} + (X'\Omega^{-1}\bar{X})\hat{\gamma}_{C1} = X'\Omega^{-1}Y \tag{B.3}$$

$$(\bar{X}'\Omega^{-1}X)\hat{\beta}_{C1} + (\bar{X}'\Omega^{-1}\bar{X})\hat{\gamma}_{C1} = \bar{X}'\Omega^{-1}Y \tag{B.4}$$

Consider the matrix equivalences of all six matrix triplets with Ω^{-1} in the normal equations. Using $\Omega^{-1} = \pi_2 I + D_{m\pi_1}\bar{J}$, $\bar{X} = \bar{J}X = \bar{J}\bar{X}$, $\pi_2 + m\pi_1 = (m\sigma_{\bar{u}}^2)^{-1}$ (Equation (27)), and $\bar{X}'\Omega_{\bar{U}}^{-1}\bar{X} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ (Equation (48)), $\bar{X}'\Omega^{-1}\bar{X}$ is equivalent to

$$\bar{X}'\Omega^{-1}\bar{X} = \bar{X}'(\pi_2 I + D_{m\pi_1}\bar{J})\bar{X} = \bar{X}'D_{\pi_2+m\pi_1}\bar{X} = \bar{X}'D_{(m\sigma_{\bar{u}}^2)^{-1}}\bar{X} = \bar{X}'\Omega_{\bar{U}}^{-1}\bar{X} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} \tag{B.5}$$

Similarly, from Equations (A.14) and (B.5), we have $\bar{X}'\Omega^{-1}X = X'\Omega^{-1}\bar{X} = \bar{X}'\Omega^{-1}\bar{X} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ and $\bar{X}'\Omega^{-1}Y = \bar{X}'\Omega^{-1}\bar{Y} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$. For the matrix equivalence of $X'\Omega^{-1}X$, we have

$$\begin{aligned} X'\Omega^{-1}X &= X'(\pi_2 I + D_{m\pi_1}\bar{J})X \\ &= \pi_2 X'X + X'D_{m\pi_1}\bar{J}X \\ &= \pi_2(\tilde{X}'\tilde{X} + \bar{X}'\bar{X}) + \bar{X}'D_{m\pi_1}\bar{X} \\ &= \pi_2\tilde{X}'\tilde{X} + \bar{X}'D_{\pi_2+m\pi_1}\bar{X} \\ &= \pi_2\tilde{X}'\tilde{X} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} \end{aligned} \tag{B.6}$$

Similarly, $X'\Omega^{-1}Y = \pi_2\tilde{X}'\tilde{Y} + \bar{x}'\Omega_{\bar{u}}^{-1}\bar{y}$. Subtracting (B.5) from Equation (B.6), we have

$$X'\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X} = \pi_2\tilde{X}'\tilde{X} \tag{B.7}$$

Similarly, $X'\Omega^{-1}Y - \bar{X}'\Omega^{-1}\bar{Y} = \pi_2\tilde{X}'\tilde{Y}$.

Consider the normal equations, Equations (B.3) and (B.4). Subtract Equation (B.4) from Equation (B.3) and using the matrix equivalencies, such as Equations (A.14) and (B.7),

$$(X\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X})\hat{\beta}_{C1} = X'\Omega^{-1}Y - \bar{X}'\Omega^{-1}\bar{Y}$$

$$\left(\pi_2\tilde{X}'\tilde{X}\right)\hat{\beta}_{C1} = \pi_2\tilde{X}'\tilde{Y}$$

$$\hat{\beta}_{C1} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} = \hat{\beta}_W \quad \text{(B.8)}$$

Using the matrix equivalencies, such as Equations (A.14) and (B.5), rewrite the second normal equation (Equation (B.4)) as

$$\begin{aligned}\hat{\gamma}_{C1} &= (\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}Y - (\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}X\hat{\beta}_{C1}\\ &= (\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})^{-1}\bar{x}'\Omega_{\bar{u}}^{-1}\bar{y} - \hat{\beta}_W\\ &= \hat{\beta}_{B1G} - \hat{\beta}_W \quad \text{(B.9)}\end{aligned}$$

Q.E.D.

Proof of Theorem 2:

Consider the CRE2 model

$$Y = X\beta + \bar{X}\gamma + Z\xi + u, V(u) = \Omega \quad \text{(B.10)}$$

Let $\left(\hat{\beta}_{C2}, \hat{\gamma}_{C2}, \hat{\xi}_{C2}\right)$ be the GLS estimator of the parameters in the CRE2 model, $\hat{\beta}_W$ be the within estimator, and $\hat{\beta}_{B2G}$ and $\hat{\xi}_{B2G}$ be the GLS between estimator of β and ξ in the between regression $\bar{y} = \bar{x}\beta_B + \ddot{z}\xi + \bar{u}$. Theorem 2 states that

$$\hat{\beta}_{C2} = \hat{\beta}_W, \hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W, \text{ and } \hat{\xi}_{C2} = \hat{\xi}_{B2G} \quad \text{(B.11)}$$

The proof of this theorem is similar to the proof of Theorem 1. The GLS estimator of β, γ, and ξ in the CRE2 model is

$$\begin{pmatrix}\hat{\beta}_{C2}\\ \hat{\gamma}_{C2}\\ \hat{\xi}_{C2}\end{pmatrix} = \begin{pmatrix}X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} & X'\Omega^{-1}Z\\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} & \bar{X}'\Omega^{-1}Z\\ Z'\Omega^{-1}X & Z'\Omega^{-1}\bar{X} & Z'\Omega^{-1}Z\end{pmatrix}^{-1}\begin{pmatrix}X'\Omega^{-1}Y\\ \bar{X}'\Omega^{-1}Y\\ Z'\Omega^{-1}Y\end{pmatrix}$$

Rewrite this matrix system of equations as:

$$\begin{pmatrix}X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} & X'\Omega^{-1}Z\\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} & \bar{X}'\Omega^{-1}Z\\ Z'\Omega^{-1}X & Z'\Omega^{-1}\bar{X} & Z'\Omega^{-1}Z\end{pmatrix}\begin{pmatrix}\hat{\beta}_{C2}\\ \hat{\gamma}_{C2}\\ \hat{\xi}_{C2}\end{pmatrix} = \begin{pmatrix}X'\Omega^{-1}Y\\ \bar{X}'\Omega^{-1}Y\\ Z'\Omega^{-1}Y\end{pmatrix}$$

The normal equations are

$$(X'\Omega^{-1}X)\hat{\beta}_{C2} + (X'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} + (X'\Omega^{-1}Z)\hat{\xi}_{C2} = X'\Omega^{-1}Y \quad \text{(B.12)}$$

$$(\bar{X}'\Omega^{-1}X)\hat{\beta}_{C2} + (\bar{X}'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} + (\bar{X}'\Omega^{-1}Z)\hat{\xi}_{C2} = \bar{X}'\Omega^{-1}Y \quad \text{(B.13)}$$

$$(Z'\Omega^{-1}X)\hat{\beta}_{C2} + (Z'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} + (Z'\Omega^{-1}Z)\hat{\xi}_{C2} = Z'\Omega^{-1}Y \quad \text{(B.14)}$$

Similar to the proof of the matrix equivalences of $\bar{X}'\Omega^{-1}X = X'\Omega^{-1}\bar{X} = \bar{X}'\Omega^{-1}\bar{X} = \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x}$ in Equations (A14) and (B.5), we find matrix equivalences of all six matrix triplets with Z and Ω^{-1}. We have $X'\Omega^{-1}Z = \bar{X}'\Omega^{-1}Z = \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z}$, $Z'\Omega^{-1}X = Z'\Omega^{-1}\bar{X} = \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x}$, $Z'\Omega^{-1}Z = \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z}$, and $Z'\Omega^{-1}Y = Z'\Omega^{-1}\bar{Y} = \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{y}$.

Consider the first two normal equations. Subtracting Equation (B.13) from Equation (B.12) and using these matrix equivalences and those in Equations (A14) and (B.7),

$$(X\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X})\hat{\beta}_{C2} = X'\Omega^{-1}Y - \bar{X}'\Omega^{-1}\bar{Y}$$

$$\left(\pi_2\tilde{X}'\tilde{X}\right)\hat{\beta}_{C2} = \pi_2\tilde{X}'\tilde{Y}$$

$$\hat{\beta}_{C2} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} = \hat{\beta}_W \quad \text{(B.15)}$$

Next, we proceed to prove $\hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W$ and $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$. Using the derived matrix equivalences, rewrite the last two normal equations, Equations (B.13) and (B.14), as the following matrix system of equations.

$$\begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix}\begin{pmatrix} \hat{\beta}_{C2} + \hat{\gamma}_{C2} \\ \hat{\xi}_{C2} \end{pmatrix} = \begin{pmatrix} \bar{x}\Omega_{\bar{u}}^{-1}\bar{y} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{y} \end{pmatrix} \quad \text{(B.16)}$$

Consider the GLS estimator of β_B and ξ in the between regression $\bar{y} = \bar{x}\beta_B + \ddot{z}\xi + \bar{u}$ as follows.

$$\begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix}^{-1}\begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{y} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{y} \end{pmatrix}$$

Rewrite the above system of equations as

$$\begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix}\begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \begin{pmatrix} \bar{x}\Omega_{\bar{u}}^{-1}\bar{y} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{y} \end{pmatrix} \quad \text{(B.17)}$$

Compare the two sets of system of equations, Equations (B.16) and (B.17), which are the system of equations for the GLS estimator of the CRE2 model and the GLS between estimator, respectively. Both systems of equations have the same coefficients. It gives

$$\hat{\beta}_{C2} + \hat{\gamma}_{C2} = \hat{\beta}_{B2G}, \hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W$$

$$\hat{\xi}_{C2} = \hat{\xi}_{B2G}$$

Q.E.D.

Appendix C. **The Means and Variances of the GLS Estimator of the CRE2 Model**

There are two methods to derive the means and variance of the GLS estimator of the CRE2 model. The first method is to use Theorem 2 and the means, variances, and covariances of $\hat{\beta}_W$, $\hat{\beta}_{B2G}$, and $\hat{\xi}_{B2G}$ from the within and between regressions. Using $\hat{\beta}_{C2} = \hat{\beta}_W$, $\hat{\gamma}_{C2} = \hat{\beta}_{B2G} - \hat{\beta}_W$, $\hat{\xi}_{C2} = \hat{\xi}_{B2G}$, Equations (53), (54), (80), (81), and $C\left(\hat{\beta}_W, \hat{\beta}_{B2G}\right) = C\left(\hat{\beta}_W, \hat{\xi}_{B2G}\right) = 0$, we derive the following means, variances, and covariances.

$$E\left(\hat{\beta}_{C2}\right) = E\left(\hat{\beta}_W\right) = \beta \qquad \text{(C.1)}$$

$$E(\hat{\gamma}_{C2}) = E\left(\hat{\beta}_{B2G} - \hat{\beta}_W\right) = \beta_B - \beta = \gamma \qquad \text{(C.2)}$$

$$E\left(\hat{\xi}_{C2}\right) = E\left(\hat{\xi}_{B2G}\right) = \xi \qquad \text{(C.3)}$$

$$V\left(\hat{\beta}_{C2}\right) = V\left(\hat{\beta}_W\right) \qquad \text{(C.4)}$$

$$V(\hat{\gamma}_{C2}) = V\left(\hat{\beta}_{B2G} - \hat{\beta}_W\right) = V\left(\hat{\beta}_{B2G}\right) + V\left(\hat{\beta}_W\right) \qquad \text{(C.5)}$$

$$V\left(\hat{\xi}_{C2}\right) = V(\hat{\xi}_{B2G}) \qquad \text{(C.6)}$$

$$C\left(\hat{\beta}_{C2}, \hat{\gamma}_{C2}\right) = C\left(\hat{\beta}_W, \hat{\beta}_{B2G} - \hat{\beta}_W\right) = -V\left(\hat{\beta}_W\right) \qquad \text{(C.7)}$$

$$C\left(\hat{\beta}_{C2}, \hat{\xi}_{C2}\right) = C\left(\hat{\beta}_W, \hat{\xi}_{B2G}\right) = 0 \qquad \text{(C.8)}$$

$$C\left(\hat{\gamma}_{C2}, \hat{\xi}_{C2}\right) = C\left(\hat{\beta}_{B2G} - \hat{\beta}_W, \hat{\xi}_{B2G}\right) = C(\hat{\beta}_{B2G}, \hat{\xi}_{B2G}) \qquad \text{(C.9)}$$

The second method to derive the above equations is to follow the standard procedure without using Theorem 2. Substituting $Y = X\beta + \bar{X}\gamma + Z\xi + u$ into the GLS estimator of the CRE2 estimator,

$$\begin{pmatrix} \hat{\beta}_{C2} \\ \hat{\gamma}_{C2} \\ \hat{\xi}_{C2} \end{pmatrix} = \begin{pmatrix} \beta \\ \gamma \\ \xi \end{pmatrix} + \begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} & X'\Omega^{-1}Z \\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} & \bar{X}'\Omega^{-1}Z \\ Z'\Omega^{-1}X & Z'\Omega^{-1}\bar{X} & Z'\Omega^{-1}Z \end{pmatrix}^{-1} \begin{pmatrix} X'\Omega^{-1}u \\ \bar{X}'\Omega^{-1}u \\ Z'\Omega^{-1}u \end{pmatrix}$$

Then $E\left(\hat{\beta}_{C2}\right) = \beta, E(\hat{\gamma}_{C2}) = \gamma, E\left(\hat{\xi}_C\right) = \xi$. For the variances and covariances of the estimator, we begin with the variance of the GLS between estimator,

$$V\begin{pmatrix} \hat{\beta}_{B2G} \\ \hat{\xi}_{B2G} \end{pmatrix} = \begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix}^{-1}$$

Rewrite the variance matrix as:

$$\begin{pmatrix} \bar{x}'\Omega_{\bar{u}}^{-1}\bar{x} & \bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z} \\ \ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x} & \ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z} \end{pmatrix} \begin{pmatrix} V(\hat{\beta}_{B2G}) & C(\hat{\beta}_{B2G}, \hat{\xi}_{B2G}) \\ C(\hat{\xi}_{B2G}, \hat{\beta}_{B2G}) & V(\hat{\xi}_{B2G}) \end{pmatrix} = I$$

It gives four equations.

$$(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})V\left(\hat{\beta}_{B2G}\right) + (\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z})C\left(\hat{\xi}_{B2G}, \hat{\beta}_{B2G}\right) = I \qquad \text{(C.10)}$$

$$(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})C(\hat{\beta}_{B2G},\hat{\xi}_{BG}) + (\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z})V(\hat{\xi}_{B2G}) = 0 \tag{C.11}$$

$$(\ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x})V(\hat{\beta}_{B2G}) + (\ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z})C(\hat{\xi}_{B2G},\hat{\beta}_{B2G}) = 0 \tag{C.12}$$

$$(\ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x})C(\hat{\beta}_{B2G},\hat{\xi}_{B2G}) + (\ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z})V(\hat{\xi}_{B2G}) = I \tag{C.13}$$

The variance of the GLS estimator of the CRE2 model is

$$V\begin{pmatrix}\hat{\beta}_{C2}\\ \hat{\gamma}_{C2}\\ \hat{\xi}_{C2}\end{pmatrix} = \begin{pmatrix}X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} & X'\Omega^{-1}Z\\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} & \bar{X}'\Omega^{-1}Z\\ Z'\Omega^{-1}X & Z'\Omega^{-1}\bar{X} & Z'\Omega^{-1}Z\end{pmatrix}^{-1}$$

Rewrite the matrix as:

$$\begin{pmatrix}X'\Omega^{-1}X & X'\Omega^{-1}\bar{X} & X'\Omega^{-1}Z\\ \bar{X}'\Omega^{-1}X & \bar{X}'\Omega^{-1}\bar{X} & \bar{X}'\Omega^{-1}Z\\ Z'\Omega^{-1}X & Z'\Omega^{-1}\bar{X} & Z'\Omega^{-1}Z\end{pmatrix}\begin{pmatrix}V(\hat{\beta}_{C2}) & C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) & C(\hat{\beta}_{C2},\hat{\xi}_{C2})\\ C(\hat{\gamma}_{C2},\hat{\beta}_{C2}) & V(\hat{\gamma}_{C2}) & C(\hat{\gamma}_{C2},\hat{\xi}_{C2})\\ C(\hat{\xi}_{C2},\hat{\beta}_{C2}) & C(\hat{\xi}_{C2},\hat{\gamma}_{C2}) & V(\hat{\xi}_{C2})\end{pmatrix} = I$$

It gives nine equations.

$$(X'\Omega^{-1}X)V(\hat{\beta}_{C2}) + (X'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\beta}_{C2}) + (X'\Omega^{-1}Z)C(\hat{\xi}_{C2},\hat{\beta}_{C2}) = I \tag{C.14}$$

$$(X'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) + (X'\Omega^{-1}\bar{X})V(\hat{\gamma}_{C2}) + (X'\Omega^{-1}Z)C(\hat{\xi}_{C2},\hat{\gamma}_{C2}) = 0 \tag{C.15}$$

$$(X'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\xi}_{C2}) + (X'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\xi}_{C2}) + (\bar{X}'\Omega^{-1}Z)V(\hat{\xi}_{C2}) = 0 \tag{C.16}$$

$$(\bar{X}'\Omega^{-1}X)V(\hat{\beta}_{C2}) + (\bar{X}'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\beta}_{C2}) + (\bar{X}'\Omega^{-1}Z)C(\hat{\xi}_{C2},\hat{\beta}_{C2}) = 0 \tag{C.17}$$

$$(\bar{X}'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) + (\bar{X}'\Omega^{-1}\bar{X})V(\hat{\gamma}_{C2}) + (\bar{X}'\Omega^{-1}Z)C(\hat{\xi}_{C2},\hat{\gamma}_{C2}) = I \tag{C.18}$$

$$(\bar{X}'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\xi}_{C2}) + (\bar{X}'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\xi}_{C2}) + (\bar{X}'\Omega^{-1}Z)V(\hat{\xi}_{C2}) = 0 \tag{C.19}$$

$$(Z'\Omega^{-1}X)V(\hat{\beta}_{C2}) + (Z'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\beta}_{C2}) + (Z'\Omega^{-1}Z)C(\hat{\xi}_{C2},\hat{\beta}_{C2}) = 0 \tag{C.20}$$

$$(Z'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) + (Z'\Omega^{-1}\bar{X})V(\hat{\gamma}_{C2}) + (Z'\Omega^{-1}Z)C(\hat{\xi}_{2C},\hat{\gamma}_{C2}) = 0 \tag{C.21}$$

$$(Z'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\xi}_{C2}) + (Z'\Omega^{-1}\bar{X})C(\hat{\gamma}_{C2},\hat{\xi}_{C2}) + (Z'\Omega^{-1}Z)V(\hat{\xi}_{C2}) = I \tag{C.22}$$

(C.14) – (C.17):

$$(X'\Omega^{-1}X)V(\hat{\beta}_{C2}) - (\bar{X}'\Omega^{-1}\bar{X})V(\hat{\beta}_{C2}) = I$$

$$V(\hat{\beta}_{C2}) = (X'\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X})^{-1} = (\pi_2\tilde{X}'\tilde{X})^{-1} = V(\hat{\beta}_W) \tag{C.23}$$

(C.15) – (C.18):

$$(X'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) - (\bar{X}'\Omega^{-1}\bar{X})C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) = -I$$

$$C(\hat{\beta}_{C2},\hat{\gamma}_{C2}) = -(X'\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X})^{-1} = -V(\hat{\beta}_W) \tag{C.24}$$

(C.16) – (C.19):

$$(X'\Omega^{-1}X)C(\hat{\beta}_{C2},\hat{\xi}_{C2}) - (\bar{X}'\Omega^{-1}\bar{X})C(\hat{\beta}_{C2},\hat{\xi}_{C2}) = 0$$

$$\left((X'\Omega^{-1}X) - \bar{X}'\Omega^{-1}\bar{X}\right)C\left(\hat{\beta}_{C2}, \hat{\xi}_{C2}\right) = 0$$

$$V\left(\hat{\beta}_W\right)C\left(\hat{\beta}_{C2}, \hat{\xi}_{C2}\right) = 0$$

$$C\left(\hat{\beta}_{C2}, \hat{\xi}_{C2}\right) = 0 \tag{C.25}$$

Using $C\left(\hat{\beta}_{C2}, \hat{\xi}_C\right) = 0$ and the matrix equivalences, rewrite Equations (C.19) and (C.22) as

$$(\bar{x}'\Omega_{\bar{u}}^{-1}\bar{x})C\left(\hat{\gamma}_{C2}, \hat{\xi}_{C2}\right) + (\bar{x}'\Omega_{\bar{u}}^{-1}\ddot{z})V\left(\hat{\xi}_{C2}\right) = 0 \tag{C.26}$$

$$(\ddot{z}'\Omega_{\bar{u}}^{-1}\bar{x})C\left(\hat{\gamma}_{C2}, \hat{\xi}_{C2}\right) + (\ddot{z}'\Omega_{\bar{u}}^{-1}\ddot{z})V\left(\hat{\xi}_{C2}\right) = I \tag{C.27}$$

Comparing the above system of equations (C.26) and (C.27) with the system of equations (C.11) and (C.13) from the between regression, both systems have the same coefficients. Therefore,

$$V\left(\hat{\xi}_{C2}\right) = V\left(\hat{\xi}_{B2G}\right) \tag{C.28}$$

$$C\left(\hat{\gamma}_{C2}, \hat{\xi}_{C2}\right) = C(\hat{\beta}_{B2G}, \hat{\xi}_{B2G}) \tag{C.29}$$

Using the last equation and matrix equivalences, rewrite Equation (C.10) from the between regression as

$$(\bar{X}'\Omega^{-1}X)V(\hat{\beta}_{B2G}) = I - (\bar{X}'\Omega^{-1}Z)C\left(\hat{\xi}_{B2G}, \hat{\beta}_{B2G}\right)$$

$$(\bar{X}'\Omega^{-1}X)V\left(\hat{\beta}_{B2G}\right) = I - (\bar{X}'\Omega^{-1}Z)C\left(\hat{\xi}_{C2}, \hat{\gamma}_{C2}\right)$$

Using $C\left(\hat{\beta}_{C2}, \hat{\gamma}_{C2}\right) = -V\left(\hat{\beta}_W\right)$ from Equation (C.24) and the above equation, rewrite Equation (C.18) as

$$(\bar{X}'\Omega^{-1}X)C\left(\hat{\beta}_{C2}, \hat{\gamma}_{C2}\right) + (\bar{X}'\Omega^{-1}\bar{X})V(\hat{\gamma}_{C2}) = I - (\bar{X}'\Omega^{-1}Z)C\left(\hat{\xi}_{C2}, \hat{\gamma}_{C2}\right)$$

$$-(\bar{X}'\Omega^{-1}X)V\left(\hat{\beta}_W\right) + (\bar{X}'\Omega^{-1}\bar{X})V(\hat{\gamma}_{C2}) = (\bar{X}'\Omega^{-1}X)V(\hat{\beta}_{B2G})$$

$$V(\hat{\gamma}_{C2}) = V\left(\hat{\beta}_{B2G}\right) + V\left(\hat{\beta}_W\right) \tag{C.30}$$

In summary, the above equations show: (C.23) = (C.4), (C.24) = (C.7), (C.25) = (C.8), (C.28) = (C.6), (C.29) = (C.9), and (C.30) = (C.5). The second method gives the same results as those from the first method, but it does not use Theorem 2, $C\left(\hat{\beta}_{B2G}, \hat{\beta}_W\right) = 0$, and $C\left(\hat{\xi}_{B2G}, \hat{\beta}_W\right) = 0$.

Appendix D. The GLS Estimators of the Moulton Model and the CRE2 Model

In appendix D, we derive the relationship between the GLS estimator of the Moulton model (Moulton, 1986, 1990) and the GLS estimator of the CRE2 model. The Moulton model is

$$y_{gi} = x_{gi}\beta + z_g\xi + u_{gi}^{**}, V\left(u_{gi}^{**}\right) = \Omega_{r2} \tag{D.1}$$

Assume $\Omega_{r2} = \Omega$. The GLS estimator is

$$\begin{pmatrix} \hat{\beta}_{REZ} \\ \hat{\xi}_{REZ} \end{pmatrix} = \begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}Z \\ Z'\Omega^{-1}X & Z'\Omega^{-1}Z \end{pmatrix}^{-1} \begin{pmatrix} X'\Omega^{-1}Y \\ Z'\Omega^{-1}Y \end{pmatrix} \quad \text{(D.2)}$$

Rewrite the system of equations as:

$$\begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}Z \\ Z'\Omega^{-1}X & Z'\Omega^{-1}Z \end{pmatrix} \begin{pmatrix} \hat{\beta}_{REZ} \\ \hat{\xi}_{REZ} \end{pmatrix} = \begin{pmatrix} X'\Omega^{-1}Y \\ Z'\Omega^{-1}Y \end{pmatrix}$$

The normal equations are

$$(X'\Omega^{-1}X)\hat{\beta}_{REZ} + (X'\Omega^{-1}Z)\hat{\xi}_{REZ} = X'\Omega^{-1}Y \quad \text{(D.3)}$$

$$(Z'\Omega^{-1}X)\hat{\beta}_{REZ} + (Z'\Omega^{-1}Z)\hat{\xi}_{REZ} = Z'\Omega^{-1}Y \quad \text{(D.4)}$$

Consider the differences of the two normal equations from the CRE2 model and the above normal equations. Subtracting Equation (B.12) from Equation (D.3) and subtracting Equation (B.14) from Equation (D.4),

$$\begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}Z \\ Z'\Omega^{-1}X & Z'\Omega^{-1}Z \end{pmatrix} \begin{pmatrix} \hat{\beta}_{REZ} - \hat{\beta}_{C2} \\ \hat{\xi}_{REZ} - \hat{\xi}_{C2} \end{pmatrix} = \begin{pmatrix} (X'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} \\ (Z'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} \end{pmatrix}$$

We solve $\hat{\beta}_{REZ} - \hat{\beta}_{C2}$ and $\hat{\xi}_{REZ} - \hat{\xi}_{C2}$ in terms of $\hat{\gamma}_{C2}$. Rewrite the above system of equations as

$$(X'\Omega^{-1}X)\big(\hat{\beta}_{REZ} - \hat{\beta}_{C2}\big) = -X'\Omega^{-1}Z\big(\hat{\xi}_{REZ} - \hat{\xi}_{C2}\big) + (X'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} \quad \text{(D.5)}$$

$$\begin{aligned} \hat{\xi}_{REZ} - \hat{\xi}_{C2} &= -(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\big(\hat{\beta}_{REZ} - \hat{\beta}_{C2}\big) + (Z'\Omega^{-1}Z)^{-1}(Z'\Omega^{-1}\bar{X})\hat{\gamma}_{C2} \\ &= -(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\big(\hat{\beta}_{REZ} - \hat{\beta}_{C2} - \hat{\gamma}_{C2}\big) \quad \text{(D.6)} \end{aligned}$$

Substituting Equation (D.6) into (D.5),

$$\begin{aligned} (X'\Omega^{-1}X)\big(\hat{\beta}_{REZ} - \hat{\beta}_{C2}\big) &= X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\big(\hat{\beta}_{REZ} - \hat{\beta}_{C2}\big) \\ &\quad \big(X'\Omega^{-1}\bar{X} - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}(Z'\Omega^{-1}\bar{X})\big)\hat{\gamma}_{C2} \end{aligned}$$

Rearrange the terms in this equation, with $(X'\Omega^{-1}X - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X)$ as a nonsingular matrix since $\begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}Z \\ Z'\Omega^{-1}X & Z'\Omega^{-1}Z \end{pmatrix}$ in Equation (D.2) is nonsingular.

$$\begin{aligned} \hat{\beta}_{REZ} = \hat{\beta}_{C2} &+ (X'\Omega^{-1}X - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X)^{-1} \\ &(\bar{X}'\Omega^{-1}\bar{X} - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X)\hat{\gamma}_{C2} \quad \text{(D.7)} \end{aligned}$$

We can simplify the above equation using the notations in the extended Frisch-Waugh Theorem in Appendix A. Let $M_Z = I - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}$ and $M_Z X = X_Z^*$, which are the residuals from the GLS estimation of the regression of X on Z. Then, as in Equation (A.13),

$$\begin{aligned} X'\Omega^{-1}X - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X &= X'\Omega^{-1}M_Z X \\ &= X_Z^{*\prime}\Omega^{-1}X_Z^* \quad \text{(D.8)} \end{aligned}$$

Similarly, as in Equation (A.32),

$$\bar{X}'\Omega^{-1}\bar{X} - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X = \bar{X}'\Omega^{-1}M_Z\bar{X}$$
$$= \bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^{*} \quad \text{(D.9)}$$

Using Equations (D.8) and (D.9), $\hat{\beta}_{REZ}$ in Equation (D.7) can be rewritten as

$$\hat{\beta}_{REZ} = \hat{\beta}_{C2} + \left(X_Z^{*\prime}\Omega^{-1}X_Z^{*}\right)^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^{*}\hat{\gamma}_{C2} \quad \text{(D.10)}$$

Define

$$\lambda_{REZ} = (X_Z^{*\prime}\Omega^{-1}X_Z^{*})^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^{*}$$

Subtracting Equation (D.9) from (D.8) and using matrix equivalences of $X'\Omega^{-1}Z = \bar{X}'\Omega^{-1}Z$ and Equation (B.7),

$$X_Z^{*\prime}\Omega^{-1}X_Z^{*} - \bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^{*}$$
$$= (X'\Omega^{-1}X - X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X)$$
$$- (\bar{X}'\Omega^{-1}\bar{X} - \bar{X}'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}\bar{X})$$
$$= X'\Omega^{-1}X - \bar{X}'\Omega^{-1}\bar{X} = \pi_2\tilde{X}'\tilde{X}$$

Then

$$I - \lambda_{REZ} = \left(X_Z^{*\prime}\Omega^{-1}X_Z^{*}\right)^{-1}X_Z^{*\prime}\Omega^{-1}X_Z^{*} - \left(X_Z^{*\prime}\Omega^{-1}X_Z^{*}\right)^{-1}\bar{X}_Z^{*\prime}\Omega^{-1}\bar{X}_Z^{*}$$
$$= \pi_2(X_Z^{*\prime}\Omega^{-1}X_Z^{*})^{-1}\tilde{X}'\tilde{X}$$

Equation (D.10) can be written as

$$\hat{\beta}_{REZ} = \hat{\beta}_{C2} + \lambda_{REZ}\hat{\gamma}_{C2}. \quad \text{(D.11)}$$

Rewrite the above equation as

$$\hat{\beta}_{REZ} - \hat{\beta}_{C2} - \hat{\gamma}_{C2} = -(I - \lambda_{REZ})\hat{\gamma}_{C2}$$

Substituting the above equation into Equation (D.6),

$$\hat{\xi}_{REZ} = \hat{\xi}_{C2} + (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X(I - \lambda_{REZ})\hat{\gamma}_{C2} \quad \text{(D.12)}$$

Equations (D.11) and (D.12) show the relationship between the GLS estimator of the Moulton model and the GLS estimator of the CRE2 model.

Q.E.D.