Generalized Random Coefficient Estimators of Panel Data Models: Asymptotic and Small Sample Properties

Mohamed R. Abonazel

April 2016

Online at https://mpra.ub.uni-muenchen.de/72586/
MPRA Paper No. 72586, posted 17 July 2016 01:50 UTC
Generalized Random Coefficient Estimators of Panel Data Models: Asymptotic and Small Sample Properties

Mohamed Reda Abonazel
Department of Applied Statistics and Econometrics
Institute of Statistical Studies and Research, Cairo University, Egypt
mabonazel@hotmail.com; mabonazel@cu.edu.eg
April 2016

ABSTRACT

This paper provides a generalized model for the random-coefficients panel data model where the errors are cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors. Of course, the conventional estimators, which used in standard random-coefficients panel data model, are not suitable for the generalized model. Therefore, the suitable estimator for this model and other alternative estimators have been provided and examined in this paper. Moreover, the efficiency comparisons for these estimators have been carried out in small samples and also we examine the asymptotic distributions of them. The Monte Carlo simulation study indicates that the new estimators are more reliable (more efficient) than the conventional estimators in small samples.

Keywords Classical pooling estimation; Contemporaneous covariance; First-order autocorrelation; Heteroskedasticity; Mean group estimation; Monte Carlo simulation; Random coefficient regression.

1. Introduction

Statistical methods can be characterized according to the type of data to which they are applied. The field of survey statistics usually deals with cross-sectional data describing each of many different individuals or units at a single point in time. Econometrics commonly uses time series data describing a single entity, usually an economy or market. The econometrics literature reveals another type of data called “panel data”, which refers to the pooling of observations on a cross-section of households, countries, and firms over several time periods. Pooling this data achieves a deep analysis of the data and gives a richer source of variation which allows for more efficient estimation of the parameters. With additional, more informative data, we can get more reliable estimates and test more sophisticated behavioral models with less restrictive assumptions. Another advantage of panel data sets is their ability to control for individual heterogeneity.¹

¹ For more information about the benefits of using pooled cross-sectional and time series data analysis, see Dielman (1983, 1989).
Panel data sets are also more effective in identifying and estimating effects that are simply not detectable in pure cross-sectional or pure time series data. In particular, panel data sets are more effective in studying complex issues of dynamic behavior. For example, in a cross-sectional data set, we can estimate the rate of unemployment at a particular point in time. Repeated cross sections can show how this proportion changes over time. Only panel data sets can estimate what proportion of those who are unemployed in one period remain unemployed in another period. Some of the benefits and limitations of using panel data sets are listed in Baltagi (2013) and Hsiao (2014).

In pooled cross-sectional and time series data (panel data) models, the pooled least squares (classical pooling) estimator is the best linear unbiased estimator (BLUE) under the classical assumptions as in the general linear regression model. An important assumption for panel data models is that the individuals in our database are drawn from a population with a common regression coefficient vector. In other words, the coefficients of a panel data model must be fixed. In fact, this assumption is not satisfied in most economic models, see, e.g., Livingston et al. (2010) and Alcacer et al. (2013). In this paper, the panel data models are studied when this assumption is relaxed. In this case, the model is called “random-coefficients panel data (RCPD) model”. The RCPD model has been examined by Swamy in several publications (Swamy 1970, 1973, and 1974), Rao (1982), Dielman (1992a, b), Beck and Katz (2007), Youssef and Abonazel (2009), and Mousa et al. (2011). Some statistical and econometric publications refer to this model as Swamy’s model or as the random coefficient regression (RCR) model, see, e.g., Poi (2003), Abonazel (2009), and Elhorst (2014, ch.3). In RCR model, Swamy assumes that the individuals in our panel data are drawn from a population with a common regression parameter, which is a fixed component, and a random component, that will allow the coefficients to differ from unit to unit. This model has been developed by many researchers, see, e.g., Beran and Millar (1994), Chelliah (1998), Anh and Chelliah (1999), Murtazashvili and Wooldridge (2008), Cheng et al. (2013), Fu and Fu (2015), Horváth and Trapani (2016), and Elster and Wübbeler (2016).

Depending on the type of assumption about the coefficient variation, Dziechciarz (1989) and Hsiao and Pesaran (2008) classified the random-coefficients models into two categories: stationary and non-stationary random-coefficients models. Stationary random-coefficients models regard the coefficients as having constant means and variance-covariances, like Swamy’s (1970) model. On the other hand, the coefficients in non-stationary random-coefficients models do not have a constant mean and/or variance and can vary systematically; these models are relevant mainly for modeling the systematic structural variation in time, like the Cooley-Prescott (1973) model.

In general, the random-coefficients models have been applied in different fields and they constitute a unifying setup for many statistical problems. Moreover, several applications of Swamy’s model have appeared in the literature of finance and economics. Boot and Frankfurter (1972) used the RCR model to examine the optimal mix of short and long-term debt for firms. Feige and Swamy (1974) applied this model to estimate demand equations for liquid assets, while Boness and Frankfurter (1977) used it to examine the concept of risk-classes in finance. Recently, Westerlund and Narayan (2015) used the random-coefficients approach to predict the stock returns at the New York Stock Exchange. Swamy et al. (2015) applied a random-coefficient framework to deal with two

---

2 These assumptions are discussed in Dielman (1983, 1989). In the next section in this paper, we will discuss different types of classical pooling estimators under different assumptions.

3 Cooley and Prescott (1973) suggested a model where coefficients vary from one time period to another on the basis of a non-stationary process. Similar models have been considered by Sant (1977) and Rausser et al. (1982).

4 The RCR model has been applied also in different sciences fields, see, e.g., Bodhlyera et al. (2014).
problems frequently encountered in applied work; these problems are correcting for misspecifications in a small area level model and resolving Simpson's paradox.

The main objective of this paper is to provide the researchers with general and efficient estimators for the stationary RCPD modes. To achieve this objective, we examine the conventional estimators of stationary RCPD models in small and moderate samples; we also propose alternative consistent estimators of these models under an assumption that the errors are cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors.

This paper is organized as follows. Section 2 presents the classical pooling estimations for panel data models when the coefficients are fixed. Section 3 provides generalized least squares (GLS) estimators for the different random-coefficients models. In section 4, we discuss the alternative estimators for these models, while section 5 examines the efficiency of these estimators. The Monte Carlo comparisons between various estimators have been carried out in section 6. Finally, section 7 offers the concluding remarks.

2. Fixed-Coefficients Models and the Pooled Estimations

Let there be observations for $N$ cross-sectional units over $T$ time periods. Suppose the variable $y$ for the $i$th unit at time $t$ is specified as a linear function of $K$ strictly exogenous variables, $x_{kit}$, in the following form:

$$y_{it} = \sum_{k=1}^{K} y_{ki} x_{kit} + u_{it} = x_{it}' y_i + u_{it}, \quad i = 1, 2, ..., N; \ t = 1, 2, ..., T,$$

where $u_{it}$ denotes the random error term, $x_{it}$ is a $1 \times K$ vector of exogenous variables, and $y_i$ is the $K \times 1$ vector of coefficients. Stacking equation (1) over time, we obtain:

$$y_t = X_t y_i + u_t,$$

where $y_t = (y_{i1}, ..., y_{iT})'$, $X_t = (x'_{i1}, ..., x'_{iT})'$, $y_i = (y_{i1}, ..., y_{iK})'$, and $u_t = (u_{i1}, ..., u_{iT})'$.

When the performance of one individual from the database is of interest, separate equation regressions can be estimated for each individual unit. If each relationship is written as in equation (2), the ordinary least squares (OLS) estimator of $y_i$, is given by:

$$\hat{y}_i = (X_i'X_i)^{-1}X_i'y_i.$$

In order for $\hat{y}_i$ to be a BLUE of $y_i$, the following assumptions must hold:

Assumption 1: The errors have zero mean, i.e., $E(u_i) = 0$ for every $i = 1, 2, ..., N$.

Assumption 2: The errors have a constant variance for each individual:

$$E(u_iu_j') = \begin{cases} \sigma_{ij}I_T & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad i,j = 1,2, ..., N.$$

Assumption 3: The exogenous variables are non-stochastic and the rank($X_i'X_i$) = $K$ for every $i = 1,2, ..., N$, where $K < T$.

Assumption 4: The exogenous variables and the errors are independent, i.e., $E(u_iX_j) = 0 \ \forall \ i, j$.

These conditions are sufficient but not necessary for the optimality of the OLS estimator.\textsuperscript{5} When OLS is not optimal, estimation can still proceed equation by equation in many cases. For

\textsuperscript{5} For more information about the optimality of the OLS estimators, see, e.g., Rao and Mitra (1971, ch. 8) and Srivastava and Giles (1987, pp. 17-21).
example, if variance of $u_i$ is not constant, the errors are either serially correlated and/or heteroskedastic, and the GLS method will provide relatively more efficient estimates than OLS, even if GLS was applied to each equation separately as in OLS.

If the covariances between $u_i$ and $u_j$ (for every $i, j = 1, 2, ..., N$) do not equal to zero, then contemporaneous correlation is present, and we have what Zellner (1962) termed as seemingly unrelated regression (SUR) equations, where the equations are related through cross-equation correlation of errors. If the $X_i$ ($i = 1, 2, ..., N$) matrices do not span the same column space and contemporaneous correlation exists, a relatively more efficient estimator of $\gamma_i$ than equation by equation OLS is the GLS estimator applied to the entire equation system as shown in Zellner (1962).

With either separate equation estimation or the SUR methodology, we obtain parameter estimates for each individual unit in the database. Now suppose it is necessary to summarize individual relationships and to draw inferences about certain population parameters. Alternatively, the process may be viewed as building a single model to describe the entire group of individuals rather than building a separate model for each. Again, assume that assumptions 1-4 are satisfied and add the following assumption:

**Assumption 5**: The individuals in our database are drawn from a population with a common regression parameter vector $\tilde{\gamma}$, i.e., $\gamma_1 = \gamma_2 = \cdots = \gamma_N = \tilde{\gamma}$.

Under assumption 5, the observations for each individual can be pooled, and a single regression performed to obtain an efficient estimator of $\tilde{\gamma}$. The equation system is now written as:

$$Y = X\tilde{\gamma} + u,$$

where $Y = (y_1', ..., y_N')'$, $X = (X_1', ..., X_N')'$, $u = (u_1', ..., u_N')'$, and $\tilde{\gamma} = (\tilde{\gamma}_1', ..., \tilde{\gamma}_K')'$ is a vector of fixed coefficients which to be estimated. Here we will differentiate between three cases based on the variance-covariance structure of $u$. In the first case, the errors have the same variance for each individual as given in the following assumption:

**Assumption 6**: $E(u_iu_j') = \begin{cases} \sigma_i^2I_T & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} i, j = 1, 2, ..., N.$

The efficient and unbiased estimator of $\tilde{\gamma}$ under assumptions 1 and 3-6 is:

$$\hat{\gamma}_{CP1} = (X'X)^{-1}X'Y.$$  \hfill (5)

This estimator has been termed the classical pooling (CP) estimator. In the second case, the errors have different variances for each individual, as given in assumption 2, in this case, the efficient and unbiased CP estimator of $\tilde{\gamma}$ under assumptions 1-5 is:

$$\hat{\gamma}_{CP2} = \left[X'(\Sigma_H \otimes I_T)^{-1}X\right]^{-1}\left[X'(\Sigma_H \otimes I_T)^{-1}Y\right],$$ \hfill (6)

where $\Sigma_H = \text{diag}(\sigma_i^2)$; for $i = 1, 2, ..., N$. The third case, if the errors have different variances for each individual and contemporaneously correlated as in the SUR model:

**Assumption 7**: $E(u_iu_j') = \begin{cases} \sigma_{ij}I_T & \text{if } i = j \\ \sigma_{ij}I_T & \text{if } i \neq j \end{cases} i, j = 1, 2, ..., N.$

Under assumptions 1, 3, 4, 5, and 7, the efficient and unbiased CP estimator of $\tilde{\gamma}$ is

---

6 In case of $X_i$ involves exactly the same elements and/or no cross-equation correlation of the errors, then no gain in efficiency is achieved by using Zellner’s SUR estimator and OLS can be applied equation by equation. Dwivedi and Srivastava (1978) showed further that whenever $X_i$ spans the same column space, OLS can be applied equation by equation without a loss in efficiency.
\( \hat{y}_{CP3} = [X'(\Sigma_{HC} \otimes I_T)^{-1}X]^{-1}[X'(\Sigma_{HC} \otimes I_T)^{-1}Y], \)

where

\[
\Sigma_{HC} = \begin{pmatrix}
\sigma_{11} & \sigma_{12} & \ldots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \ldots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \ldots & \sigma_{NN}
\end{pmatrix}.
\]

To make the above estimators (\( \hat{y}_{CP2} \) and \( \hat{y}_{CP3} \)) feasible, the \( \sigma_{ij} \) can be replaced with the following unbiased and consistent estimator:

\[
\hat{\sigma}_{ij} = \frac{\hat{u}_{ij}}{T-K}; \quad \forall \; i,j = 1,2,...,N,
\]

where \( \hat{u}_{ij} \) is the residuals vector obtained from applying OLS to equation number \( i \):

\[
\hat{u}_{ij} = y_{ij} - X_{ij}\hat{\gamma}_i,
\]

where \( \hat{\gamma}_i \) is defined in (3).

3. Random-Coefficients Models

In this section, we review the standard random-coefficients model, proposed by Swamy (1970). Moreover, we present the random-coefficients model in the general case; when the errors are cross-sectional heteroskedastic and contemporaneously correlated as well as with the first-order autocorrelation of the time series errors.

3.1. Swamy’s (RCR) Model

Suppose that each regression coefficient in equation (2) is now viewed as a random variable; that is the coefficients, \( \gamma_{ij} \), are viewed as invariant over time, but varying from one unit to another:

**Assumption 8:** According to the stationary random coefficient approach, we assume that the coefficient vector \( \gamma_i \) is specified as:

\[
\gamma_i = \bar{\gamma} + \mu_i,
\]

where \( \bar{\gamma} \) is a \( K \times 1 \) vector of constants, and \( \mu_i \) is a \( K \times 1 \) vector of stationary random variables with zero means and constant variance-covariances:

\[
E(\mu_i) = 0, \quad \text{and} \quad E(\mu_i\mu_j') = \Psi_{ij} = \begin{cases} \psi_k^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} i,j = 1,2,...,N,
\]

where \( \Psi = \text{diag}(\{\psi_k^2\}) \); for \( k = 1,2,...,K \), where \( K < N \). Also, we assume that \( E(\mu_iX_{jt}) = 0 \) and \( E(\mu_iu_{jt}) = 0 \) \( \forall \; i \) and \( j \).

Under the assumption 8, the model in equation (2) can be rewritten as:

\[
Y = X\bar{\gamma} + \mu + u,
\]

where \( Y, X, u \), and \( \bar{\gamma} \) are defined in (4), while \( \mu = (\mu'_1,...,\mu'_N)' \), and \( D = \text{diag}(X_{i}') \); for \( i = 1,2,...,N \).

---

7 The \( \hat{\sigma}_{ij} \) in (8) is unbiased estimator, because we assume, in the first, that the number of exogenous variables of each equation is equal, i.e., \( K_i = K \) for \( i = 1,2,...,N \). However, in the general case, \( K_i \neq K_j \), the unbiased estimator is \( \hat{\sigma}_{ij}/[T-K_i - K_j + \text{tr}(P_{xx})] \), where \( P_{xx} = X_i'(X'_iX_i)^{-1}X'_iX_j(X'jX_j)^{-1}X'_j \). See Srivastava and Giles (1987, pp. 13-17) and Baltagi (2011, pp. 243-244).

8 This means that the individuals in our database are drowning from a population with a common regression parameter \( \bar{\gamma} \), which is fixed component, and a random component \( \mu_i \), which will allow the coefficients to differ from unit to unit.

5
The model in (11), under assumptions 1-4 and 8, is called the “RCR model”, which was examined by Swamy (1970, 1971, 1973, and 1974), Youssef and Abonazel (2009), and Mousa et al. (2011). We will refer to assumptions 1-4 and 8 as RCR assumptions. Under these assumptions, the BLUE of $\bar{y}$ in equation (11) is:

$$\hat{\bar{y}}_{RCR} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y,$$

where $\Omega$ is the variance-covariance matrix of $e$:

$$\Omega = (\Sigma_H \otimes I_T) + D(I_N \otimes \Psi')D'.$$

Swamy (1970) showed that the $\hat{\bar{y}}_{RCR}$ estimator can be rewritten as:

$$\hat{\bar{y}}_{RCR} = \left[\sum_{i=1}^{N} X_i'(X_i'\Psi' + \sigma_i I_T)^{-1}X_i\right]^{-1}\sum_{i=1}^{N} X_i'(X_i'\Psi' + \sigma_i I_T)^{-1}y_i = \sum_{i=1}^{N} W_i \hat{y}_i,$$

where $\hat{y}_i$ is defined in (3), and

$$W_i = \left[\sum_{i=1}^{N} [\Psi + \sigma_i (X_i'X_i)^{-1}]^{-1}\right]^{-1}\left[\sum_{i=1}^{N} [\Psi + \sigma_i (X_i'X_i)^{-1}]^{-1}\right].$$

It shows that the $\hat{\bar{y}}_{RCR}$ is a weighted average of the least squares estimator for each cross-sectional unit, $\hat{y}_i$, and with the weights inversely proportional to their covariance matrices. It also shows that the $\hat{\bar{y}}_{RCR}$ requires only a matrix inversion of order $K$, and so it is not much more complicated to compute than the sample least squares estimator.

The variance-covariance matrix of $\hat{\bar{y}}_{RCR}$ under RCR assumptions is:

$$\text{var}(\hat{\bar{y}}_{RCR}) = (X'\Omega^{-1}X)^{-1} = \left[\sum_{i=1}^{N} [\Psi + \sigma_i (X_i'X_i)^{-1}]^{-1}\right]^{-1}.$$

To make the $\hat{\bar{y}}_{RCR}$ estimator feasible, Swamy (1971) suggested using the estimator in (8) as an unbiased and consistent estimator of $\sigma_H$, and the following unbiased estimator for $\Psi$:

$$\varphi = \left[\frac{1}{N-1} \left(\sum_{i=1}^{N} \hat{y}_i \hat{y}_i' - \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \sum_{i=1}^{N} \hat{y}_i'\right) - \left(\frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{ii}(X_i'X_i)^{-1}\right)\right].$$

Swamy (1973, 1974) showed that the estimator $\hat{\varphi}_{RCR}$ is consistent as both $N,T \to \infty$ and is asymptotically efficient as $T \to \infty$. It is worth noting that, just as in the error-components model, the estimator (17) is not necessarily non-negative definite. Mousa et al. (2011) explained that it is possible to obtain negative estimates of Swamy’s estimator in (17) in case of small samples and if some/all coefficients are fixed.

But in medium and large samples, the negative variance estimates does not appear even if all coefficients are fixed. To solve this problem, Swamy has suggested replacing (17) by:

$$\hat{\varphi}^+ = \frac{1}{N-1} \left(\sum_{i=1}^{N} \hat{y}_i \hat{y}_i' - \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \sum_{i=1}^{N} \hat{y}_i'\right),$$

this estimator, although biased, is non-negative definite and consistent when $T \to \infty$. See Judge et al. (1985, p. 542).

---

9. The final equality in (14) is obtained by using the fact that: $(A + BC'B)^{-1} = A^{-1} - A^{-1}BEB'A^{-1} + A^{-1}BE(E + C)^{-1}EB'A^{-1}$, where $E = (B'A^{-1}B)^{-1}$. See Rao (1973, p. 33).

10. The statistical properties of $\hat{\bar{y}}_{RCR}$ have been examined by Swamy (1971), of course, under RCR assumptions.

11. This suggestion was been used by Stata program, specifically in xtrchh and xtrchhh2 Stata’s commands. See Poi (2003).
It is worth mentioning here that if both \( u_{it} \) and \( \mu_i \) are normally distributed, the GLS estimator of \( \bar{y} \) is the maximum likelihood estimator of \( \bar{y} \) conditional on \( \Psi \) and \( \sigma_{ii} \). Without knowledge of \( \Psi \) and \( \sigma_{ii} \), we can estimate \( \bar{y} \), \( \Psi \) and \( \sigma_{ii} \) \((i = 1, 2, \ldots, N)\) simultaneously by the maximum likelihood method. However, computationally it can be tedious. A natural alternative is to first estimate \( \Omega \), then substitute the estimated \( \Omega \) into (12). See Hsiao and Pesaran (2008).

### 3.2. Generalized RCR Model

To generalize RCR model so that it would be more suitable for most economic models, we assume that the errors are cross-sectional heteroskedastic and contemporaneously correlated, as in assumption 7, as well as with the first-order autocorrelation of the time series errors. Therefore, we add the following assumption to assumption 7:

**Assumption 9:** \( u_{it} = \rho_t u_{i,t-1} + \varepsilon_{it}; \ |\rho_t| < 1 \), where \( \rho_t \) \((i = 1, 2, \ldots, N)\) are first-order autocorrelation coefficients and are fixed. Assume that: \( E(\varepsilon_{it}) = 0 \), \( E(u_{i,t-1}\varepsilon_{jt}) = 0 \); \( \forall \ i \) and \( j \), and

\[
E(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma_{ii} & \text{if } i = j \\ \sigma_{ij} & \text{if } i \neq j \end{cases} \quad i, j = 1, 2, \ldots, N,
\]

it is assumed that in the initial time period the errors have the same properties as in subsequent periods. So, we assume that: \( E(u_{i0}^2) = \sigma_{ii}/1 - \rho_i^2 \) and \( E(u_{i0}u_{j0}) = \sigma_{ij}/1 - \rho_i\rho_j \); \( \forall \ i \) and \( j \).

We will refer to assumptions 1, 3, 4, and 7-9 as the general RCR assumptions. Under these assumptions, the BLUE of \( \bar{y} \) is:

\[
\hat{\bar{y}}_{gCRR} = (X'\Omega^*^{-1}X)^{-1}X'\Omega^{*-1}Y,
\]

where

\[
\Omega^* = \begin{pmatrix}
X_1\Psi X_1' + \sigma_{e11}\omega_{11} & \sigma_{e12}\omega_{12} & \cdots & \sigma_{e1N}\omega_{1N} \\
\sigma_{e21}\omega_{21} & X_2\Psi X_2' + \sigma_{e22}\omega_{22} & \cdots & \sigma_{e2N}\omega_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{eN1}\omega_{N1} & \sigma_{eN2}\omega_{N2} & \cdots & X_N\Psi X_N' + \sigma_{eNN}\omega_{NN}
\end{pmatrix},
\]

with

\[
\omega_{ij} = \frac{1}{1 - \rho_i\rho_j} \begin{pmatrix} 1 & \rho_i & \rho_i^2 & \cdots & \rho_i^{T-1} \\ \rho_j & 1 & \rho_i & \cdots & \rho_i^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_j^{T-2} & \rho_j^{T-3} & \cdots & 1 \end{pmatrix}.
\]

Since the elements of \( \Omega^* \) are usually unknown, we develop a feasible Aitken estimator of \( \bar{y} \) based on consistent estimators of the elements of \( \Omega^* \):

\[
\hat{\rho}_t = \frac{\sum_{t=2}^T \tilde{u}_{it}\tilde{u}_{i,t-1}}{\sum_{t=2}^T \tilde{u}_{i,t-1}^2},
\]

where \( \tilde{u}_i = (\tilde{u}_{i1}, ..., \tilde{u}_{iT})' \) is given in (9).

\[
\hat{\delta}_{ij} = \frac{\tilde{\varepsilon}_{ij}}{T - K},
\]

where \( \tilde{\varepsilon}_t = (\tilde{\varepsilon}_{t1}, \tilde{\varepsilon}_{t2}, ..., \tilde{\varepsilon}_{tT})' \); \( \tilde{u}_{i1} = \tilde{u}_{i1}/(1 - \hat{\rho}_t^2) \), and \( \hat{\delta}_{it} = \tilde{\varepsilon}_{it} - \hat{\rho}_t\tilde{u}_{i,t-1} \) for \( t = 2, ..., T \).
By replacing $\rho_i$ by $\hat{\rho}_i$ in (21), we get consistent estimators of $\omega_{ij}$, say $\hat{\omega}_{ij}$. And then we will use $\hat{\delta}_{ij}$ and $\hat{\omega}_{ij}$ to get a consistent estimator of $\Psi$:\(^{12}\)

$$
\varphi^* = \left[ \frac{1}{N} - 1 \right] \left( \frac{N}{N} \sum_{i=1}^{N} \hat{y}_i^* \hat{y}_i^{**} - \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i^* \sum_{i=1}^{N} \hat{y}_i^{**} \right) - \frac{1}{N} \sum_{i=1}^{N} \hat{\delta}_{ii} \left( X_i^\prime \hat{\omega}_{ii}^{-1} X_i \right)^{-1}
+ \frac{1}{N(N-1)} \sum_{i \neq j} \hat{\delta}_{ij} \left( X_i^\prime \hat{\omega}_{ii}^{-1} X_i \right)^{-1} X_i^\prime \hat{\omega}_{ij}^{-1} \hat{\omega}_{ji} \left( X_j^\prime \hat{\omega}_{jj}^{-1} X_j \right)^{-1},
$$

(24)

where

$$
\hat{y}_i^* = \left( X_i^\prime \hat{\omega}_{ii}^{-1} X_i \right)^{-1} X_i^\prime \hat{\omega}_{ii}^{-1} X_i^\prime \hat{\omega}_{ii}^{-1} X_i^\prime \hat{\omega}_{ii}^{-1} y_i.
$$

By using the consistent estimators ($\hat{\delta}_{ij}$, $\hat{\omega}_{ij}$, and $\hat{\Psi}^*$) in (20), we have a consistent estimator of $\Omega^*$, say $\hat{\Omega}^*$. Then we use $\hat{\Omega}^*$ to get the generalized RCR (GRCR) estimator of $\hat{y}$:

$$
\hat{y}_{GRCR} = \left( X^\prime \hat{\Omega}^*^{-1} X \right)^{-1} X^\prime \hat{\Omega}^*^{-1} Y.
$$

(26)

The estimated variance-covariance matrix of $\hat{y}_{GRCR}$ is:

$$
\text{var} (\hat{y}_{GRCR}) = \left( X^\prime \hat{\Omega}^*^{-1} X \right)^{-1}.
$$

(27)

4. Mean Group Estimation

A consistent estimator of $\bar{y}$ can also be obtained under more general assumptions concerning $\gamma_i$ and the regressors. One such possible estimator is the mean group (MG) estimator, proposed by Pesaran and Smith (1995) for estimation of dynamic panel data (DPD) models with random coefficients.$^{13}$ The MG estimator is defined as the simple average of the OLS estimators:

$$
\hat{y}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i.
$$

(28)

Even though the MG estimator has been used in DPD models with random coefficients, it will be used here as one of the alternative estimators of static panel data models with random coefficients. Moreover, the efficiency of MG estimator in the two random-coefficients models (RCR and GRCR) will be studied. Note that the simple MG estimator in (28) is more suitable for the RCR Model. But to make it suitable for the GRCR model, we suggest a general mean group (GMG) estimator as:

$$
\hat{y}_{GMG} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i^*.
$$

(29)

where $\hat{y}_i^*$ is defined in (25).

Lemma 1.

If the general RCR assumptions are satisfied, then the $\hat{y}_{MG}$ and $\hat{y}_{GMG}$ are unbiased estimators of $\bar{y}$ and the estimated variance-covariance matrices of $\hat{y}_{MG}$ and $\hat{y}_{GMG}$ are:

---

\(^{12}\) The estimator of $\rho_i$ in (22) is consistent, but it is not unbiased. See Srivastava and Giles (1987, p. 211) for other suitable consistent estimators of $\rho_i$ that are often used in practice.

\(^{13}\) For more information about the estimation methods for DPD models, see, e.g., Baltagi (2013), Abonazel (2014), Youssef et al. (2014a,b), and Youssef and Abonazel (2015).
\[
\text{var}(\hat{\beta}_{MG}) = \frac{1}{N} \hat{\sigma}^2 + \frac{1}{N^2} \sum_{i=1}^{N} \hat{\sigma}_{ei} (X'_{i}X_{i})^{-1} X'_{i} \hat{\omega}_{ii} X_{i}(X'_{i}X_{i})^{-1} \\
+ \frac{1}{N^2} \sum_{i \neq j}^{N} \hat{\sigma}_{eij} (X'_{i}X_{i})^{-1} X'_{i} \hat{\omega}_{ij} X_{j}(X'_{j}X_{j})^{-1},
\]

(30)

\[
\text{var}(\hat{\beta}_{GMG}) = \frac{1}{N(N-1)} \left[ \left( \sum_{i=1}^{N} \hat{\gamma}_{i} \hat{\gamma}_{i}' - \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_{i} \sum_{i=1}^{N} \hat{\gamma}_{i}' \right) \right.
+ \sum_{i \neq j}^{N} \hat{\sigma}_{eij} (X'_{i} \hat{\omega}_{ii}^{-1} X_{i})^{-1} X'_{i} \hat{\omega}_{ij}^{-1} \hat{\omega}_{ij}^{-1} X_{j}(X'_{j} \hat{\omega}_{jj}^{-1} X_{j})^{-1}. \]

(31)

It is noted from lemma 1 that the variance of GMG estimator is less than the variance of MG estimator when the general RCR assumptions are satisfied. In other words, the GMG estimator is more efficient than the MG estimator. But under RCR assumptions, we have:

\[
\text{var}(\hat{\beta}_{MG}) = \frac{1}{N(N-1)} \left( \sum_{i=1}^{N} \hat{\gamma}_{i} \hat{\gamma}_{i}' \right) \frac{1}{N} \text{ var}(\hat{\beta}_{GMG}) = \frac{1}{N} \theta^*.
\]

(32)

5. Efficiency Comparisons

In this section, we examine the efficiency gains from the use of GRCR estimator. Moreover, the asymptotic variances (as \(T \to \infty\) with \(N\) fixed) of GRCR, RCR, GMG, and MG estimators have been derived.

Under the general RCR assumptions, It is easy to verify that the classical pooling estimators \((\hat{\beta}_{CP1}, \hat{\beta}_{CP2}, \text{ and } \hat{\beta}_{CP3})\) and Swamy’s estimator \((\hat{\beta}_{RCR})\) are unbiased for \(\beta\) and with variance-covariance matrices:

\[
\text{var}(\hat{\beta}_{CP1}) = G_1 \Omega^* G_1', \quad \text{var}(\hat{\beta}_{CP2}) = G_2 \Omega^* G_2', \quad \text{var}(\hat{\beta}_{CP3}) = G_3 \Omega^* G_3', \quad \text{var}(\hat{\beta}_{RCR}) = G_4 \Omega^* G_4',
\]

(33)

(34)

Where \(G_1 = (X'X)^{-1}X'\), \(G_2 = [X' (\Sigma_{H}^{-1} \otimes I_T) X]^{-1} X' (\Sigma_{H}^{-1} \otimes I_T)\), \(G_3 = [X' (\Sigma_{HC}^{-1} \otimes I_T) X]^{-1} X' (\Sigma_{HC}^{-1} \otimes I_T)\), and \(G_4 = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1}\). The efficiency gains, from the use of GRCR estimator, can be summarized in the following equation:

\[
EG_\alpha = \text{var}(\hat{\beta}_\alpha) - \text{var}(\hat{\beta}_{GRCR}) = (G_m - G_0) \Omega^* (G_m - G_0)', \quad h = 1, \ldots, 4,
\]

(35)

where the subscript \(\alpha\) indicates the estimator that is used (CP1, CP2, CP3, or RCR), \(G_m\) matrices are defined in (33) and (34), and \(G_0 = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1}.\) Since \(\Omega^*, \Sigma_{H}, \Sigma_{HC}\), and \(\Omega\) are positive definite matrices, then \(EG_\alpha\) matrices are positive semi-definite matrices. In other words, the GRCR estimator is more efficient than CP1, CP2, CP3, and RCR estimators. These efficiency gains are increasing when \([\rho_i], \sigma_{eij}, \text{ and } \psi_k^2\) are increasing. However, it is not clear to what extent these efficiency gains hold in small samples. Therefore, this will be examined in a simulation study.

The next lemma explains the asymptotic variances (as \(T \to \infty\) with \(N\) fixed) properties of GRCR, RCR, GMG, and MG estimators. In order to the derivation of the asymptotic variances, we must assume the following:
**Assumption 10:** \( \lim_{T \to \infty} T^{-1} X_i' X_i \) and \( \lim_{T \to \infty} T^{-1} X_i' \hat{\omega}_{ii}^{-1} X_i \) are finite and positive definite for all \( i \) and for \( 0 < |\rho_i| < 1 \).

**Lemma 2.**

If the general RCR assumptions and assumption 10 are satisfied, then the estimated asymptotic variance-covariance matrices of GRCR, RCR, GMG, and MG estimators are equal:

\[
\lim_{T \to \infty} \text{var}(\hat{\gamma}_{GRCR}) = \lim_{T \to \infty} \text{var}(\hat{\gamma}_{RCR}) = \lim_{T \to \infty} \text{var}(\hat{\gamma}_{GMG}) = \lim_{T \to \infty} \text{var}(\hat{\gamma}_{MG}) = \frac{1}{N} \Psi^+. 
\]

We can conclude from lemma 2 that the means and the variance-covariance matrices of the limiting distributions of \( \hat{\gamma}_{GRCR}, \hat{\gamma}_{RCR}, \hat{\gamma}_{GMG}, \) and \( \hat{\gamma}_{MG} \) estimators are the same and are equal to \( \bar{\gamma} \) and \( \frac{1}{N} \Psi \) respectively even if the errors are correlated as in assumption 9. Therefore, it is not expected to increase the asymptotic efficiency of \( \hat{\gamma}_{GRCR} \) about \( \hat{\gamma}_{RCR}, \hat{\gamma}_{GMG}, \) and \( \hat{\gamma}_{MG} \). This does not mean that the GRCR estimator cannot be more efficient than RCR, GMG, and MG in small samples when the errors are correlated as in assumption 9, this will be examined in a simulation study.

**6. The Simulation Study**

In this section, the Monte Carlo simulation has been used for making comparisons between the behavior of the classical pooling estimators (\( \hat{\gamma}_{CP1}, \hat{\gamma}_{CP2}, \) and \( \hat{\gamma}_{CP3} \)), random-coefficients estimators (\( \hat{\gamma}_{RCR} \) and \( \hat{\gamma}_{GRCR} \)), and mean group estimators (\( \hat{\gamma}_{MG} \) and \( \hat{\gamma}_{GMG} \)) in small and moderate samples. We use R language to create our program to set up the Monte Carlo simulation and this program is available if requested.

**6.1. Design of the Simulation**

Monte Carlo experiments were carried out based on the following data generating process:

\[
y_{it} = \sum_{k=1}^{3} y_{kt} x_{kit} + u_{it} = x_{it} \bar{\gamma} + x_{it} \mu_i + u_{it}, \quad i = 1, 2, ..., N; \quad t = 1, 2, ..., T. \tag{36}
\]

To perform the simulation under the general RCR assumptions, the model in (36) was generated as follows:

1. The values of the independent variables, \( x_{kit}; k = 1, 2, 3, \) were generated as independent normally distributed random variables with constant mean zero and also constant standard deviation one. The values of \( x_{kit} \) were allowed to differ for each cross-sectional unit. However, once generated for all \( N \) cross-sectional units the values were held fixed over all Monte Carlo trials.

2. The coefficients, \( y_{kt} \), were generated as in assumption 8: \( y_i = \bar{\gamma} + \mu_i \), where the vector of \( \bar{\gamma} = (1,1,1)' \), and \( \mu_i \) were generated as multivariate normal distributed with means zeros and a variance-covariance matrix \( \Psi = \text{diag}(\psi_k^2); k = 1, 2, 3 \). The values of \( \psi_k^2 \) were chosen to be fixed for all \( k \) and equal to 0, 5, or 25. Note that when \( \psi_k^2 = 0 \), the coefficients are fixed.

3. The errors, \( u_{it} \), were generated as in assumption 9: \( u_{it} = \rho u_{i,t-1} + \varepsilon_{it} \), where the values of \( \varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})' \forall i = 1, 2, ..., N \) were generated as multivariate normal distributed with means zeros and a variance-covariance matrix.
The values of $\sigma_{e_{ii}}$, $\sigma_{e_{ij}}$, and $\rho$ were chosen to be: $\sqrt{\sigma_{e_{ii}}} = 5 \text{ or } 15$; $\sigma_{e_{ij}} = 0$, 0.75, or 0.95; and $\rho = 0$, 0.55, or 0.85, where the values of $\sigma_{e_{ii}}$, $\sigma_{e_{ij}}$, and $\rho$ are constants for all $i, j = 1, 2, ..., N$ in each Monte Carlo trial. The initial values of $u_{it}$ are generated as $u_{it} = \varepsilon_{it} / \sqrt{1 - \rho^2} \forall i = 1, 2, ..., N$. The values of errors were allowed to differ for each cross-sectional unit on a given Monte Carlo trial and were allowed to differ between trials. The errors are independent with all independent variables.

4. The values of $N$ and $T$ were chosen to be 5, 8, 10, 12, 15, and 20 to represent small and moderate samples for the number of individuals and the time dimension. To compare the small and moderate samples performance for the different estimators, the three different samplings have been designed in our simulation where each design of them contain $s$ four pairs of $N$ and $T$; the first two of them represent the small samples while the moderate samples are represented by the second two pairs. These designs have been created as follows: First, case of $\sigma_{e_{ii}}$, the different pairs of $N$ and $T$ were chosen to be $(5, 8)$, $(5, 12)$, $(10, 15)$, or $(10, 20)$. Second, case of $\sigma_{e_{ij}}$, the different pairs are $(5, 5)$, $(10, 10)$, $(15, 15)$, or $(20, 20)$. Third, case of $\rho$, the different pairs are $(8, 5)$, $(12, 5)$, $(15, 10)$, or $(20, 10)$.

5. In all Monte Carlo experiments, we ran 1000 replications and all the results of all separate experiments are obtained by precisely the same series of random numbers.

To raise the efficiency of the comparison between these estimators, we calculate the total standard errors (TSE) for each estimator by:

$$TSE = trace \left\{ \frac{1}{1000} \sum_{t=1}^{1000} \left[ \text{var}(\hat{\gamma}_t) \right]^{0.5} \right\},$$

where $\hat{\gamma}_t$ is the estimated vector of the true vector of coefficients mean ($\gamma$) in (36), and $\text{var}(\hat{\gamma}_t)$ is the estimated variance-covariance matrix of the estimator. More detailed, to calculate TSE for $\hat{\gamma}_{GCR}$, $\hat{\gamma}_{CP}$, $\hat{\gamma}_{CP}$, $\hat{\gamma}_{CR}$, $\hat{\gamma}_{MG}$, and $\hat{\gamma}_{GMG}$, equations (27), (33), (34), (30), and (31) should be used, respectively.

6.2. Monte Carlo Results

The results are given in Tables 1-6. Specifically, Tables 1-3 present the TSE values of the estimators when $\sqrt{\sigma_{e_{ii}}} = 5$, and in cases of $N < T$, $N = T$, and $N > T$, respectively. While case of $\sqrt{\sigma_{e_{ii}}} = 15$ is presented in Tables 4-6 in the same cases of $N$ and $T$. In our simulation study, the main factors that have an effect on the TSE values of the estimators are $N, T, \sigma_{e_{ii}}, \sigma_{e_{ij}}, \rho$, and $\psi_k^2$. From Tables 1-6, we can summarize some effects for all estimators (classical pooling, random-coefficients, and mean group estimators) in the following points:

- When the value of $\psi_k^2$ is increased, the values of TSE are increasing for all simulation situations.
- When the values of $N$ and $T$ are increased, the values of TSE are decreasing for all situations.
- When the value of $\sigma_{e_{ii}}$ is increased, the values of TSE are increasing in most situations.
- When the values of $(\sigma_{e_{ij}}, \rho)$ are increased, the values of TSE are increasing in most situations.

For more deeps in simulation results, we can conclude the following results:
1. In general, when $\sigma_{e_{ij}} = \rho = \psi_k^2 = 0$, the TSE values of classical pooling estimators (CP1, CP2, and CP3) are similar (approximately equivalent), especially when the sample size is moderate and/or $N \leq T$. However, the TSE values of GMG and GRCR estimators are smaller than the classical pooling estimators in this situation ($\sigma_{e_{ij}} = \rho = \psi_k^2 = 0$) and other simulation situations (case of $\sigma_{e_{ij}}, \sigma_{e_{ij}'}$ and $\psi_k^2$ are increasing). In other words, the GMG and GRCR estimators are more efficient than CP1, CP2, and CP3 estimators whether the regression coefficients are fixed ($\psi_k^2 = 0$) or random ($\psi_k^2 > 0$).

2. Also, when the coefficients are random (when $\psi_k^2 > 0$), the values of TSE for GMG and GRCR estimators are smaller than MG and RCR estimators in all simulation situations (for any $N, T, \sigma_{e_{ij}}, \sigma_{e_{ij}'}$ and $\rho$). However, the TSE values of GRCR estimator are smaller than the values of TSE for GMG estimator in most situations, especially when the sample size is moderate. In other words, the GRCR estimator performs well than all other estimators as long as the sample size is moderate regardless of other simulation factors.

3. If $T \geq 15$, the values of TSE for MG and GMG estimators are approximately equivalent. This result is consistent with Lemma 2. According our study, the case of $T \geq 15$ is achieved when the sample size is moderate in Tables 1, 2, 4 and 5. Moreover, that convergence is slowing down if $\sigma_{e_{ij}}, \sigma_{e_{ij}'}$ and $\rho$ are increasing. But the situation for RCR and GRCR estimators is different; the convergence between them is very slow even if $T = 20$. So the MG and GMG estimators are more efficient than RCR estimator in all simulation situations.

4. Generally, the performance of all estimators in cases of $N < T$ and $N = T$ is better than their performance in case of $N > T$. Similarly, Their performance in cases of $\sqrt{\overline{\sigma_{e_{ij}}}} = 5$ is better than the performance in case of $\sqrt{\overline{\sigma_{e_{ij}}}} = 15$, but it is not significantly as in $N$ and $T$.

7. Conclusion

In this paper, the classical pooling (CP1, CP2, and CP3), random-coefficients (RCR and GRCR), and alternative (MG and GMG) estimators of stationary RCPD models were examined in different sample sizes in case the errors are cross-sectionally and serially correlated. Efficiency comparisons for these estimators indicate that the mean group and random-coefficients estimators are equivalent when $T$ sufficiently large. Moreover, we carried out Monte Carlo simulations to investigate the small samples performance for all estimators given above.

The Monte Carlo results show that the classical pooling estimators are not suitable for random-coefficients models absolutely. Also, the MG and GMG estimators are more efficient than RCR estimator in random- and fixed-coefficients models especially when $T$ is small ($T \leq 12$). Moreover, the GMG and GRCR estimators perform well in small samples if the coefficients are random or fixed. The MG, GMG, and GRCR estimators are approximately equivalent when $T \geq 20$. However, the GRCR estimator performs well than the GMG estimator in most situations especially in moderate samples. Therefore, we conclude that the GRCR estimator is suitable to stationary RCPD models whether the coefficients are random or fixed.
Appendix

A.1 Proof of Lemma 1

a. Show that $E(\hat{\mathcal{P}}_{GMG}) = E(\bar{\mathcal{P}}_{MG}) = \bar{y}$:

By substituting (25) into (29), we can get

$$\hat{\mathcal{P}}_{GMG} = \frac{1}{N} \sum_{i=1}^{N} (X'_i\omega^{-1}_{ii}X_i)^{-1}X'_i\omega^{-1}_{ii}y_i,$$

(A.1)

by substituting $y_i = X_iy_i + u_i$ into (A.1), then

$$\hat{\mathcal{P}}_{GMG} = \frac{1}{N} \sum_{i=1}^{N} [y_i + (X'_i\omega^{-1}_{ii}X_i)^{-1}X'_i\omega^{-1}_{ii}u_i].$$

(A.2)

Similarly, we can rewrite $\hat{\mathcal{P}}_{MG}$ in (28) as:

$$\hat{\mathcal{P}}_{MG} = \frac{1}{N} \sum_{i=1}^{N} [y_i + (X'_iX_i)^{-1}X'_iu_i].$$

(A.3)

Taking the expectation for (A.2) and (A.3), and using assumption 1, we get

$$E(\hat{\mathcal{P}}_{GMG}) = E(\hat{\mathcal{P}}_{MG}) = \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{y}.$$

b. Derive the variance-covariance matrix of $\hat{\mathcal{P}}_{GMG}$:

Beginning, note that under assumption 8, we have $y_i = \bar{y} + \mu_i$. Let us add $\hat{\mu}_i^*$ to the both sides:

$$y_i + \hat{\mu}_i^* = \bar{y} + \mu_i + \hat{\mu}_i^*,$$

$$\hat{\mu}_i^* = \bar{y} + \mu_i + (\hat{\mu}_i - y_i),$$

(A.4)

let $\hat{\mu}_i - y_i = \tau_i$, then we can rewrite the equation (A.4) as follows:

$$\hat{\mu}_i = \bar{y} + \mu_i + \tau_i,$$

(A.5)

where $\tau_i = (X'_i\omega^{-1}_{ii}X_i)^{-1}X'_i\omega^{-1}_{ii}u_i$. From (A.5), we can get

$$\frac{1}{N} \sum_{i=1}^{N} \hat{\mu}_i = \bar{y} + \frac{1}{N} \sum_{i=1}^{N} \mu_i + \frac{1}{N} \sum_{i=1}^{N} \tau_i,$$

which means that

$$\hat{\mathcal{P}}_{GMG} = \bar{y} + \bar{\mu} + \bar{\tau},$$

(A.6)

where $\bar{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mu_i$ and $\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i$. From (A.6) and using the general RCR assumptions, we get

$$\text{var}(\hat{\mathcal{P}}_{GMG}) = \text{var}(\bar{\mu}) + \text{var}(\bar{\tau})$$

$$= \frac{1}{N} \text{var}(\Psi) + \frac{1}{N^2} \sum_{i=1}^{N} \sigma_{\epsilon_{ii}}(X'_i\omega^{-1}_{ii}X_i)^{-1}$$

$$+ \frac{1}{N^2} \sum_{i \neq j}^{N} \sigma_{\epsilon_{ij}}(X'_i\omega^{-1}_{ii}X_i)^{-1}X'_i\omega^{-1}_{ii}X'_i\omega^{-1}_{ij}X_i(X'_i\omega^{-1}_{ii}X_i)^{-1}.$$

(A.7)

Using the consistent estimators of $\Psi, \sigma_{\epsilon_{ij}},$ and $\omega_{ij}$ that defined in above, we get

$$\text{var}(\hat{\mathcal{P}}_{GMG}) = \frac{1}{N(N-1)} \left[ \left( \sum_{i=1}^{N} \hat{\mu}_i \hat{\mu}_i' - \frac{1}{N} \sum_{i=1}^{N} \hat{\mu}_i \sum_{i=1}^{N} \hat{\mu}_i' \right)$$

$$+ \sum_{i \neq j}^{N} \sigma_{\epsilon_{ij}}(X'_i\omega^{-1}_{ii}X_i)^{-1}X'_i\omega^{-1}_{ii}X'_i\omega^{-1}_{ij}X_i(X'_i\omega^{-1}_{ii}X_i)^{-1} \right].$$
c. Derive the variance-covariance matrix of $\hat{\Psi}_{MG}$:

As above, we can rewrite the equation (3) as follows:

$$\hat{y}_i = \bar{y} + \mu_i + \lambda_i,$$

where $\lambda_i = \hat{y}_i - y_i = (X'_i X_i)^{-1}X'_i u_i$. From (A.8), we can get

$$\frac{1}{N} \sum_{i=1}^{N} \hat{y}_i = \bar{y} + \frac{1}{N} \sum_{i=1}^{N} \mu_i + \frac{1}{N} \sum_{i=1}^{N} \lambda_i,$$

which means that

$$\hat{\Psi}_{MG} = \bar{\mu} + \bar{\lambda},$$

where $\bar{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mu_i$, and $\bar{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i$. From (A.9) and using the general RCR assumptions, we get

$$\text{var}(\hat{\Psi}_{MG}) = \text{var}(\bar{\mu}) + \text{var}(\bar{\lambda}) = \frac{1}{N} \text{var} + \frac{1}{N^2} \sum_{i=1}^{N} \sigma_{\epsilon_i}(X'_i X_i)^{-1}X'_i \omega_i X_i (X'_i X_i)^{-1} +$$

$$\frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{\epsilon_i}(X'_i X_i)^{-1}X'_i \omega_i X_i (X'_i X_i)^{-1}.$$

(A.10)

As in MG estimator, by using the consistent estimators of $\Psi, \sigma_{\epsilon_i}$ and $\omega_i$, we get

$$\text{var}(\hat{\Psi}_{MG}) = \frac{1}{N} \bar{\Psi}^* + \frac{1}{N^2} \sum_{i=1}^{N} \sigma_{\epsilon_i}(X'_i X_i)^{-1}X'_i \omega_i X_i (X'_i X_i)^{-1} + \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{\epsilon_i}(X'_i X_i)^{-1}X'_i \omega_i X_i (X'_i X_i)^{-1}.$$

(A.11)

A.2 Proof of Lemma 2:

Following the same argument as in Parks (1967) and utilizing assumption 10, we can show that

$$\text{plim}_{T \to \infty} \hat{y}_i = \text{plim}_{T \to \infty} \hat{y}_i' = y_i, \text{ plim}_{T \to \infty} \hat{\sigma}_{ij} = \sigma_{\epsilon_i}, \text{ and } \text{plim}_{T \to \infty} \hat{\omega}_{ij} = \omega_{ij},$$

and then,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} T (X'_i \omega_i^{-1} X_i)^{-1} = \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} T (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) =$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} T (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) =$$

$$= \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} T (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) = 0.$$

(A.12)

Substituting (A.11) and (A.12) in (24), we get

$$\lim_{T \to \infty} \hat{\Psi}^* = \frac{1}{N-1} \left( \sum_{i=1}^{N} y_i - \frac{1}{N} \sum_{i=1}^{N} y_i \right) = \Psi^*.$$

(A.13)

By substitute (A.11)-(A.13) into (30), (31), and (27), we get

$$\lim_{T \to \infty} \text{var}(\hat{\Psi}_{MG}) = \frac{1}{N} \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} (X'_i \omega_i^{-1} X_i)^{-1} + \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{\epsilon_i} (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) = \frac{1}{N} \Psi^*,$$

(A.14)

$$\lim_{T \to \infty} \text{var}(\hat{\Psi}_{GRCR}) =$$

$$= \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{\epsilon_i} (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) = \frac{1}{N} \Psi^*,$$

(A.15)

$$\lim_{T \to \infty} \text{var}(\hat{\Psi}_{GM}) = \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \sigma_{\epsilon_i} (X'_i \omega_i^{-1} X_i (X'_i X_i)^{-1}) = \frac{1}{N} \Psi^*.$$
Similarly, we will use the results in (A.11)-(A.13) in case of RCR estimator:

\[
\lim_{T \to \infty} \text{var}(\hat{\beta}_{RCR}) = \lim_{T \to \infty} \left[ (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \right] = \frac{1}{N} \Psi^+.
\] (A.17)

From (A.14)-(A.17), we can conclude that:

\[
\lim_{T \to \infty} \text{var}(\hat{\beta}_{GCR}) = \lim_{T \to \infty} \text{var}(\hat{\beta}_{RCR}) = \lim_{T \to \infty} \text{var}(\hat{\beta}_{GMG}) = \lim_{T \to \infty} \text{var}(\hat{\beta}_{MG}) = \frac{1}{N} \Psi^+.
\]

References


Table 1: TSE for various estimators when $\sqrt{\sigma_{\epsilon_i}} = 5$ and $N < T$

<table>
<thead>
<tr>
<th>$(\sigma_{\epsilon_i}, \rho)$</th>
<th>$(0, 0)$</th>
<th>$(0.75, 0.55)$</th>
<th>$(0.95, 0.85)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, T)$</td>
<td>(5, 8)</td>
<td>(5, 12)</td>
<td>(10, 15)</td>
</tr>
<tr>
<td>$\psi_k^2 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>2.579 1.812 0.965 0.765</td>
<td>2.970 1.764 1.071 0.893</td>
<td>5.016 2.881 1.473 1.337</td>
</tr>
<tr>
<td>CP2</td>
<td>2.739 1.819 0.950 0.746</td>
<td>3.087 1.773 1.052 0.882</td>
<td>5.483 2.875 1.493 1.324</td>
</tr>
<tr>
<td>CP3</td>
<td>2.875 1.795 0.904 0.657</td>
<td>3.235 1.723 0.955 0.785</td>
<td>5.796 2.756 1.344 1.144</td>
</tr>
<tr>
<td>MG</td>
<td>2.793 1.912 1.068 0.813</td>
<td>2.925 1.917 1.165 0.960</td>
<td>5.337 2.935 1.594 1.267</td>
</tr>
<tr>
<td>GMG</td>
<td>2.055 1.479 0.904 0.701</td>
<td>2.207 1.218 0.846 0.684</td>
<td>3.441 1.531 0.785 0.613</td>
</tr>
<tr>
<td>GRCR</td>
<td>2.394 1.728 0.839 0.672</td>
<td>2.527 1.623 0.812 0.714</td>
<td>4.165 2.255 0.992 0.810</td>
</tr>
<tr>
<td>$\psi_k^2 = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi_k^2 = 25$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: TSE for various estimators when $\sqrt{\sigma_{e_i}} = 5$ and $N = T$

<table>
<thead>
<tr>
<th>$(\sigma_{e_i}, \rho)$</th>
<th>(0, 0)</th>
<th>(0.75, 0.55)</th>
<th>(0.95, 0.85)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, T)$</td>
<td>(5, 5)</td>
<td>(10, 10)</td>
<td>(15, 15)</td>
</tr>
<tr>
<td>$\psi^2 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>4.015</td>
<td>1.398</td>
<td>0.704</td>
</tr>
<tr>
<td>CP2</td>
<td>5.107</td>
<td>1.451</td>
<td>0.682</td>
</tr>
<tr>
<td>CP3</td>
<td>6.626</td>
<td>2.038</td>
<td>0.858</td>
</tr>
<tr>
<td>MG</td>
<td>4.078</td>
<td>1.573</td>
<td>0.791</td>
</tr>
<tr>
<td>GMG</td>
<td>2.848</td>
<td>1.302</td>
<td>0.701</td>
</tr>
<tr>
<td>GRCR</td>
<td>3.376</td>
<td>1.152</td>
<td>0.541</td>
</tr>
<tr>
<td>$\psi^2 = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRCR</td>
<td>4.762</td>
<td>2.823</td>
<td>1.809</td>
</tr>
<tr>
<td>$\psi^2 = 25$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: TSE for various estimators when $\sqrt{\sigma_{e_{ij}}} = 5$ and $N > T$

<table>
<thead>
<tr>
<th>$(\sigma_{e_{ij}}, \rho)$</th>
<th>$(N, T)$</th>
<th>(0, 0)</th>
<th>(0.75, 0.55)</th>
<th>(0.95, 0.85)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, T)$</td>
<td>(8, 5)</td>
<td>(12, 5)</td>
<td>(15, 10)</td>
<td>(20, 10)</td>
</tr>
<tr>
<td>$\psi^2 = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>8.059</td>
<td>5.111</td>
<td>0.915</td>
<td>1.286</td>
</tr>
<tr>
<td>CP2</td>
<td>12.611</td>
<td>9.223</td>
<td>0.914</td>
<td>1.474</td>
</tr>
<tr>
<td>CP3</td>
<td>12.098</td>
<td>8.479</td>
<td>1.037</td>
<td>1.790</td>
</tr>
<tr>
<td>MG</td>
<td>7.346</td>
<td>4.968</td>
<td>1.048</td>
<td>1.497</td>
</tr>
<tr>
<td>GMG</td>
<td>5.085</td>
<td>3.780</td>
<td>0.912</td>
<td>1.161</td>
</tr>
<tr>
<td>RCR</td>
<td>7.583</td>
<td>6.827</td>
<td>1.963</td>
<td>3.424</td>
</tr>
<tr>
<td>GRCR</td>
<td>6.269</td>
<td>3.781</td>
<td>0.594</td>
<td>0.984</td>
</tr>
<tr>
<td>$\psi^2 = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>7.211</td>
<td>4.939</td>
<td>2.659</td>
<td>2.498</td>
</tr>
<tr>
<td>CP2</td>
<td>11.436</td>
<td>9.220</td>
<td>3.138</td>
<td>2.956</td>
</tr>
<tr>
<td>CP3</td>
<td>10.724</td>
<td>8.292</td>
<td>3.822</td>
<td>3.592</td>
</tr>
<tr>
<td>MG</td>
<td>6.429</td>
<td>4.963</td>
<td>2.360</td>
<td>2.346</td>
</tr>
<tr>
<td>GMG</td>
<td>6.011</td>
<td>4.623</td>
<td>2.359</td>
<td>2.343</td>
</tr>
<tr>
<td>RCR</td>
<td>7.966</td>
<td>7.216</td>
<td>2.363</td>
<td>2.801</td>
</tr>
<tr>
<td>GRCR</td>
<td>5.929</td>
<td>3.838</td>
<td>2.173</td>
<td>1.938</td>
</tr>
<tr>
<td>$\psi^2 = 25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>8.409</td>
<td>7.200</td>
<td>5.316</td>
<td>5.907</td>
</tr>
<tr>
<td>CP2</td>
<td>13.196</td>
<td>13.419</td>
<td>6.278</td>
<td>7.128</td>
</tr>
<tr>
<td>CP3</td>
<td>12.464</td>
<td>12.334</td>
<td>7.654</td>
<td>8.452</td>
</tr>
<tr>
<td>MG</td>
<td>7.703</td>
<td>6.546</td>
<td>4.555</td>
<td>4.956</td>
</tr>
<tr>
<td>GMG</td>
<td>7.762</td>
<td>6.554</td>
<td>4.554</td>
<td>4.954</td>
</tr>
<tr>
<td>RCR</td>
<td>11.761</td>
<td>7.170</td>
<td>4.547</td>
<td>4.882</td>
</tr>
<tr>
<td>GRCR</td>
<td>6.661</td>
<td>5.629</td>
<td>4.462</td>
<td>4.782</td>
</tr>
</tbody>
</table>
Table 4: TSE for various estimators when $\sqrt{\sigma_{\epsilon_{ij}}}$ = 15 and $N < T$

<table>
<thead>
<tr>
<th>$(\sigma_{\epsilon_{ij}}, \rho)$</th>
<th>$(0, 0)$</th>
<th>$(0.75, 0.55)$</th>
<th>$(0.95, 0.85)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, T)$</td>
<td>(5, 8)</td>
<td>(5, 12)</td>
<td>(10, 15)</td>
</tr>
<tr>
<td>$\psi_k^2 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP3</td>
<td>5.109</td>
<td>2.822</td>
<td>1.505</td>
</tr>
<tr>
<td>MG</td>
<td>4.823</td>
<td>3.074</td>
<td>1.747</td>
</tr>
<tr>
<td>GMG</td>
<td>3.652</td>
<td>2.410</td>
<td>1.480</td>
</tr>
<tr>
<td>RCR</td>
<td>7.652</td>
<td>10.706</td>
<td>2.723</td>
</tr>
<tr>
<td>$\psi_k^2 = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi_k^2 = 25$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5: TSE for various estimators when $\sqrt{\sigma_{e_i}} = 15$ and $N = T$

<table>
<thead>
<tr>
<th>$(\sigma_{e_i}, \rho)$</th>
<th>$(N, T)$</th>
<th>$(0, 0)$</th>
<th>$(0.75, 0.55)$</th>
<th>$(0.95, 0.85)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(5, 5)$</td>
<td>$(10, 10)$</td>
<td>$(15, 15)$</td>
<td>$(20, 20)$</td>
</tr>
<tr>
<td>$\psi_i^2 = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP1</td>
<td>25.198</td>
<td>2.054</td>
<td>1.214</td>
<td>0.882</td>
</tr>
<tr>
<td>CP2</td>
<td>31.269</td>
<td>2.081</td>
<td>1.172</td>
<td>0.852</td>
</tr>
<tr>
<td>CP3</td>
<td>41.189</td>
<td>2.802</td>
<td>1.463</td>
<td>0.992</td>
</tr>
<tr>
<td>MG</td>
<td>20.301</td>
<td>2.302</td>
<td>1.336</td>
<td>0.966</td>
</tr>
<tr>
<td>GMG</td>
<td>12.441</td>
<td>1.946</td>
<td>1.184</td>
<td>0.872</td>
</tr>
<tr>
<td>GRCR</td>
<td>18.106</td>
<td>1.687</td>
<td>0.876</td>
<td>0.592</td>
</tr>
<tr>
<td>$\psi_i^2 = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRCR</td>
<td>17.694</td>
<td>3.031</td>
<td>2.305</td>
<td>2.323</td>
</tr>
<tr>
<td>$\psi_i^2 = 25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: TSE for various estimators when $\sqrt{\sigma_{\varepsilon_{ij}}} = 15$ and $N > T$

<table>
<thead>
<tr>
<th>$(\sigma_{\varepsilon_{ij}}, \rho)$</th>
<th>$(0, 0)$</th>
<th>$(0.75, 0.55)$</th>
<th>$(0.95, 0.85)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, T)$</td>
<td>(8, 5)</td>
<td>(12, 5) (15, 10)</td>
<td>(20, 10)</td>
</tr>
<tr>
<td>CP1</td>
<td>8.099</td>
<td>17.393</td>
<td>1.731</td>
</tr>
<tr>
<td>CP3</td>
<td>12.232</td>
<td>29.385</td>
<td>2.033</td>
</tr>
<tr>
<td>GMG</td>
<td>5.447</td>
<td>9.402</td>
<td>1.768</td>
</tr>
<tr>
<td>GRCR</td>
<td>6.386</td>
<td>12.973</td>
<td>1.153</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\psi_{k} = 5$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GMG</td>
<td>6.193</td>
<td>9.803</td>
<td>2.975</td>
<td>2.520</td>
<td>7.059</td>
<td>7.670</td>
</tr>
<tr>
<td>GRCR</td>
<td>6.490</td>
<td>11.975</td>
<td>2.384</td>
<td>1.995</td>
<td>8.071</td>
<td>6.935</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\psi_{k} = 25$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CP3</td>
<td>15.672</td>
<td>23.191</td>
<td>9.144</td>
<td>8.111</td>
<td>17.441</td>
<td>17.000</td>
</tr>
</tbody>
</table>