

Dual Generalized Maximum Entropy Estimation for Panel Data Regression Models

Jaejun Lee^a, Sooyoung Cheon^{1,b}

^aCenter for Military Planning, Korea Institute for Defense Analyses, Korea

^bDepartment of Applied Statistics, Korea University, Korea

Abstract

Data limited, partial, or incomplete are known as an ill-posed problem. If the data with ill-posed problems are analyzed by traditional statistical methods, the results obviously are not reliable and lead to erroneous interpretations. To overcome these problems, we propose a dual generalized maximum entropy (dual GME) estimator for panel data regression models based on an unconstrained dual Lagrange multiplier method. Monte Carlo simulations for panel data regression models with exogeneity, endogeneity, or/and collinearity show that the dual GME estimator outperforms several other estimators such as using least squares and instruments even in small samples. We believe that our dual GME procedure developed for the panel data regression framework will be useful to analyze ill-posed and endogenous data sets.

Keywords: Collinearity, endogeneity, exogeneity, generalized maximum entropy, ill-posed problems, panel data.

1. Introduction

The amount of recorded data in the world has exploded. Managing and analyzing data (especially non-experimental data) have always offered the greatest challenges for statisticians and econometricians. One of the reasons is that data are limited, partial, or incomplete in most of the cases, which is called an ill-posed problem (Golan *et al.*, 1996a). For example, in regression models, ill-posed problems can result from a variety of causes, including non-stationarity, model misspecification, and collinearity. If the data with ill-posed problems are analyzed by traditional estimation and inference procedures, the results obviously are not reliable and lead to erroneous interpretations. Even it is highly unlikely to find feasible solutions.

Shannon (1948) initially applied the entropy concept, which is the quantitative measure of disorder in physics, to the theory of communication and transmission of information. Since then, Jaynes (1957a, b) developed the maximum entropy methodology that can be applied to predicting probability distribution in situations, where he only knew what the possible outcome values are and that the sum of corresponding probabilities is unity. Golan *et al.* (1994), Golan and Judge (1996), and Golan *et al.* (1996b, 1997) coped with the ill-posed and underdetermined problems in economics.

For the classical maximum entropy formalism, Golan *et al.* (1996a) considered a finite and discrete linear model

$$y = X\beta = Xp, \quad (1.1)$$

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2062740).

¹ Corresponding author: Department of Applied Statistics, Korea University, 2511 Sejong-ro, Sejong-city, 339-700, South Korea. E-mail: scheon@korea.ac.kr

Published 30 September 2014 / journal homepage: <http://csam.or.kr>

© 2014 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

where \mathbf{y} is a T -dimensional vector and the linear operator X is a $T \times K$ ($K > T$) non-invertible matrix. Here the unknown and unobservable frequencies $\mathbf{p} = (p_1, \dots, p_K)'$ satisfy the conditions $\sum_{k=1}^K p_k = 1$ and $p_k \geq 0$. Given this specification, determining the unknown probabilities \mathbf{p} is often difficult because the number of data points are less than the number of unknowns. To solve these undetermined problems, Shannon (1948) used an axiomatic method of probability to define a unique function to measure the uncertainty of a collection of events.

Suppose that X be a random variable with possible outcome values $x_k, k = 1, \dots, K$ and probabilities p_k such that $\sum_k p_k = 1$. Shannon defined the *entropy* of the distribution of probabilities, $\mathbf{p} = (p_1, \dots, p_K)'$ as the measure $H(\mathbf{p}) = -\sum_{k=1}^K p_k \ln(p_k)$ where $0 \cdot \ln(0) = 0$. This measure H obtains a maximum when $p_1 = \dots = p_K = 1/K$; i.e., the probabilities are uniform. Jaynes (1957a, b) proposed making use of the entropy concept in choosing the unknown distribution of probabilities in (1.1). Under the above maximum entropy principle studied by Jaynes, we choose the distribution for which the information(data) is sufficient to determine the probability assignment.

The definition of the principle of maximum entropy (ME; Jaynes, 1957a, b; Golan *et al.*, 1996a) about \mathbf{p} is given by

$$H(\mathbf{p}) = -\sum_{k=1}^K p_k \ln(p_k) = -\mathbf{p}' \ln \mathbf{p} \quad (1.2)$$

subject to data consistency and normalized-additivity requirements $\mathbf{y} = X\mathbf{p}$ and $\mathbf{p}'\mathbf{1} = \mathbf{1}$ where $\mathbf{1}$ is a $K \times 1$ vector of ones and $\ln \mathbf{p}$ is a $K \times 1$ vector. This maximum entropy formulation is based on the work of Shannon (1948) and Jaynes (1957a, b, 1984) and extended by Kullback (1959), Levine (1980) and Levine and Tribus (1979). The solution to the maximization problem (1.2) can be obtained from the Lagrangian function

$$L = -\mathbf{p}' \ln \mathbf{p} + \lambda'(\mathbf{y} - X\mathbf{p}) + \mu'(\mathbf{1} - \mathbf{p}'\mathbf{1}), \quad (1.3)$$

where λ and μ are Lagrangian multipliers.

For panel data regression models with strict exogeneity of the regressors, Song and Cheon (2006) proposed a generalized maximum entropy(GME) approach and showed that their approach provides a better alternative to classical least squares approaches such as ordinary least squares(OLS), generalized least squares(GLS), and feasible generalized least squares(FGLS).

However, their approach has several disadvantages since they did not use a constrained or unconstrained numerical method as in Golan *et al.* (1996a) even though the objective function has no closed-form solution. Their estimates computed by a least squares method do not satisfy one among the optimal conditions. In their simulation studies, their new design matrix using the modified singular value decomposition is not appropriate in panel data regression models. Finally, their iterative method is too sensitive to the choice of starting values compared with our new methodology.

In this paper, we overcome the drawback of Song and Cheon (2006) by adopting Golan *et al.* (1996a)'s unconstrained dual GME procedures so that all the assumptions stated in Golan *et al.* (1996a) are satisfied. We consider panel data regression models with strict exogeneity as well as collinearity and endogeneity.

The rest of this paper is organized as follows. Section 2 gives the definition of a panel data regression model and introduces existing estimation methodologies. In Section 3, we describe the dual GME methodology for a panel data regression model. Section 4 provides computational details and considers the performance of several different estimation methods throughout Monte Carlo simulations, and the last section provides a final remark of the main results.

2. Panel Data Regression Models

As a point of departure, consider the following panel data regression model (Hsiao, 1986; Baltagi, 2001),

$$y_{nt} = \mathbf{x}'_{nt}\boldsymbol{\beta} + u_{nt}, \quad n = 1, \dots, N; t = 1, \dots, T, \quad (2.1)$$

where y_{nt} is the observation on a dependent variable for the n th cross sectional unit (*e.g.*, individuals, firms, countries) at the t^{th} time period, \mathbf{x}_{nt} and $\boldsymbol{\beta}$ denote the $K \times 1$ vectors of independent variables and regression coefficients, respectively, and u_{nt} is the regression disturbance. The prime symbol also indicates the transpose of a matrix or vector.

The regression disturbance u_{nt} follows an error components structure

$$u_{nt} = \mu_n + \varepsilon_{nt},$$

where μ_n denotes the n^{th} individual specific effect assumed to be *i.i.d.*(0, σ_μ^2) and ε_{nt} is the remainder disturbance assumed to be *i.i.d.*(0, σ_ε^2) which is independent of μ_n .

The model (2.1) can be rewritten in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (2.2)$$

where \mathbf{y} is now of dimension $NT \times 1$, \mathbf{X} is $NT \times K$ where the constant is absorbed into \mathbf{X} , $\boldsymbol{\beta}$ is $K \times 1$, and \mathbf{u} is $NT \times 1$. The error term \mathbf{u} can be written in vector form as

$$\mathbf{u} = (\mathbf{I}_N \otimes \mathbf{i}_T)\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad (2.3)$$

with $\mathbf{u}' = (u_{11}, \dots, u_{1T}, \dots, u_{N1}, \dots, u_{NT})$, $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_N)$, and $\boldsymbol{\varepsilon}' = (\varepsilon_{11}, \dots, \varepsilon_{1T}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{NT})$, where \mathbf{I}_N is a $N \times N$ identity matrix, \mathbf{i}_T is a $T \times 1$ vector of ones, and \otimes denotes the Kronecker product.

Under these assumptions, the variance-covariance matrix for \mathbf{u} is

$$\begin{aligned} \boldsymbol{\Sigma} \equiv \text{Cov}(\mathbf{u}) &= E(\mathbf{u}\mathbf{u}') = \sigma_\mu^2 (\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\varepsilon^2 \mathbf{I}_{NT} \\ &= (T\sigma_\mu^2 + \sigma_\varepsilon^2) (\mathbf{I}_N \otimes \bar{\mathbf{J}}_T) + \sigma_\varepsilon^2 (\mathbf{I}_{NT} - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T) \\ &= \sigma_\alpha^2 \mathbf{A} + \sigma_\varepsilon^2 \mathbf{B}, \end{aligned} \quad (2.4)$$

where \mathbf{J}_T is a $T \times T$ matrix of all ones, $\bar{\mathbf{J}}_T = \mathbf{J}_T/T$, $\sigma_\alpha^2 = (T\sigma_\mu^2 + \sigma_\varepsilon^2)$, $\mathbf{A} = (\mathbf{I}_N \otimes \bar{\mathbf{J}}_T)$, and $\mathbf{B} = (\mathbf{I}_{NT} - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T)$. In (2.4), two matrices \mathbf{A} and \mathbf{B} are orthogonal to each other and idempotent, and they sum up to the identity matrix. Therefore, the inverse of $\boldsymbol{\Sigma}$ is, as devised by Wansbeek and Kapteyn (1989),

$$\boldsymbol{\Sigma}^{-1} = \sigma_\alpha^{-2} \mathbf{A} + \sigma_\varepsilon^{-2} \mathbf{B}. \quad (2.5)$$

As classical least squares estimators for the panel data regression model, an OLS estimator is $\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, which is consistent when N goes to infinity but T remains finite as well as asymptotically normally distributed under the above assumptions. Knowing the value of $\boldsymbol{\Sigma}$ in (2.4), a GLS estimator is given by $\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$. Given the assumptions stated above the GLS estimator is unbiased and efficient; moreover, it is consistent, asymptotically efficient, and normally distributed for N going to infinity with finite T .

In practice, Σ is unknown and has to be estimated. A FGLS estimator is obtained by a consistent estimator of Σ , say $\hat{\Sigma}$. That is, $\hat{\beta}_{\text{FGLS}} = (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} y$, which is asymptotically equivalent to the GLS estimator as long as N tends to infinity. Amemiya(AM) (1971) proposed an estimator of Σ using two estimators of the variance components using the least squares dummy variable residuals, say \tilde{u} ; i.e., $\hat{\Sigma}^{-1} = \hat{\sigma}_\alpha^{-2} A + \hat{\sigma}_\varepsilon^{-2} B$ where $\hat{\sigma}_\alpha^2 = \tilde{u}' A \tilde{u} / \text{tr}(A)$ and $\hat{\sigma}_\varepsilon^2 = \tilde{u}' B \tilde{u} / \text{tr}(B)$. Amemiya showed that the proposed estimators have the same asymptotic distribution as that knowing the true disturbances. Swamy and Arora(SA) (1972) proposed two regressions (within and between) to estimate σ_α^2 and σ_ε^2 from respective mean square errors, which are $\hat{\sigma}_\alpha^2 = 1/(N-K)[y' A y - y' A X (X' A X)^{-1} X' A y]$ and $\hat{\sigma}_\varepsilon^2 = 1/(N(T-1)-K+1)[y' B y - y' B X_{-1} (X_{-1}' B X_{-1})^{-1} X_{-1}' B y]$, where X_{-1} indicates a design matrix which does not include a column of ones for the intercept.

So far, we assume that $E(\mu_n | x_n) = 0$ for every n as well as $E(\varepsilon_{nt} | x_n) = 0$ for every n and t ; that is, we consider a panel data model with strictly exogenous regressors. However, the endogeneity can be caused by omitted variables, measurement error, and simultaneous equations. Since the effect of endogeneity is biased in estimates and two types of errors (Type I and Type II) occur in testing hypotheses, the endogeneity is one of the most major challenges in econometric analysis, especially in panel data analysis. All of the solutions require finding an instrumental variable; for example, a two-stage least squares(2SLS) or a generalized method of moments(GMM) estimation as appropriate.

As estimators using instruments, Im *et al.* (1999) uses the instruments $Z_n \equiv I_T \otimes x_n^o$ in order to estimate β under the assumption $E(X_n \otimes u_n) = 0$, where X_n is the $T \times K$ matrix of explanatory variables for the n^{th} individual; u_n is the $T \times 1$ vector of the disturbances for this individual; and x_n^o is the row vector containing every row of X_n . Then the three-stage least squares(3SLS) estimator of β is defined as

$$\hat{\beta}_{\text{3SLS}} = \left[X' Z (Z' \hat{\Sigma} Z)^{-1} Z' X \right]^{-1} X' Z (Z' \hat{\Sigma} Z)^{-1} Z' y, \quad (2.6)$$

where $\hat{\Sigma} = I_N \otimes \hat{\Sigma}^*$, $\hat{\Sigma}^* = \sum_n \hat{u}_n \hat{u}_n'$, and $\hat{u}_n = y_n - X_n \hat{\beta}$, and $\hat{\beta}$ is the two-stage least squares estimator of β . Wooldridge (2003) stated that the 3SLS estimator above is a bit different from the usual 3SLS estimator that appeared in most textbooks and called it a traditional 3SLS estimator(T3SLS). T3SLS is defined as

$$\hat{\beta}_{\text{T3SLS}} = \left(\sum_{n=1}^N \dot{X}_n' \hat{\Sigma}^* \dot{X}_n \right)^{-1} \left(\sum_{n=1}^N \dot{X}_n' \hat{\Sigma}^* y_n \right), \quad (2.7)$$

where $\dot{X}_n = X_n (Z' Z)^{-1} Z' X$. He also noted that the 3SLS estimator is consistent under three assumptions, which are (i) $E(Z_n' u_n) = 0$, (ii) $\text{rank}[E(Z_n' X_n)] = K$, and (iii) $\hat{\Sigma}$ is consistent, but the traditional 3SLS estimator is not.

Since a generalized methods of moments(GMM) estimation was formalized by Hansen (1982), a GMM has become the most recently popularized estimation techniques in many fields such as economics and finance. A GMM estimator, which also leads to consistency and efficiency under some regularity conditions (Ahn and Schmidt, 1999), is

$$\hat{\beta}_{\text{GMM}} = \left[X' Z \hat{V}^{-1} Z' X \right]^{-1} X' Z \hat{V}^{-1} Z' y, \quad (2.8)$$

where $\hat{V} = \sum_n Z_n' \hat{u}_n \hat{u}_n' Z_n$ and \hat{u}_n is the two-stage least squares estimator as before. Ahn and Schmidt (1999) mentioned that this GMM estimator is asymptotically identical to the 3SLS estimator under

the condition $E(Z_n' u_n u_n' Z_n) = E(Z_n' \Sigma^* Z_n)$. However, the GMM estimator is more efficient than the 3SLS estimator when the condition is not satisfied.

Other instrumental variables estimators, the filtered instrumental variables(FIV) estimator and the generalized instrumental variables(GIV) estimator, are popular in the panel data analysis as well. FIV and GIV estimators are defined by, respectively. Refer to Ahn and Schmidt (1999) and Im *et al.* (1999) for more details.

$$\hat{\beta}_{\text{FIV}} = \left[X' \hat{\Sigma}^{-\frac{1}{2}} Z (Z' Z)^{-1} Z' \hat{\Sigma}^{-\frac{1}{2}} X \right]^{-1} X' \hat{\Sigma}^{-\frac{1}{2}} Z (Z' Z)^{-1} Z' \hat{\Sigma}^{-\frac{1}{2}} y, \quad (2.9)$$

$$\hat{\beta}_{\text{GIV}} = \left[X' \hat{\Sigma}^{-1} Z (Z' \hat{\Sigma}^{-1} Z)^{-1} Z' \hat{\Sigma}^{-1} X \right]^{-1} X' \hat{\Sigma}^{-1} Z (Z' \hat{\Sigma}^{-1} Z)^{-1} Z' \hat{\Sigma}^{-1} y. \quad (2.10)$$

Baltagi (2001) pointed out that panel data usually give a large number of data points, increasing the degrees of freedom and reducing the collinearity among explanatory variables because the cross-sectional dimension adds lots of variability. However, we cannot avoid collinearity even in panel data analysis if there are several categorical or dummy variables as explanatory variables. When we use a statistical package such as SAS, STATA, or SPSS to analyze such panel data, an error may occur with no surprise.

3. A Dual Generalized Maximum Entropy Estimator

3.1. A review of the GME problem

We first describe the GME methodology for a panel data regression model, introduced by Song and Cheon (2006). First of all, let reparameterize β and u in (2.2) as in Judge and Golan (1992).

$$\beta = Zp = \begin{bmatrix} z_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & z_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & z_K' \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}, \quad (3.1)$$

where Z is a $K \times KM$ matrix, p is a $KM \times 1$ vector, $\beta_k = \sum_m z_{km} p_{km}$ for every k , $z_k' = (z_{k1}, \dots, z_{kM})$, and $p_k' = (p_{k1}, \dots, p_{kM})$. In (2.3),

$$\mu = Fg = \begin{bmatrix} f_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & f_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & f_N' \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{bmatrix}, \quad (3.2)$$

where F is a $N \times NR$ matrix, g is a $NR \times 1$ vector, $\mu_n = \sum_r f_{nr} g_{nr}$ for every n , $f_n' = (f_{n1}, \dots, f_{nR})$, and $g_n' = (g_{n1}, \dots, g_{nR})$. Moreover,

$$\varepsilon = Vw = \begin{bmatrix} v_{11}' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & v_{12}' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & v_{NT}' \end{bmatrix} \begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{NT} \end{bmatrix}, \quad (3.3)$$

where V is a $NT \times NTJ$ matrix, w is a $NTJ \times 1$ vector, $\varepsilon_{nt} = \sum_j v_{ntj} w_{ntj}$ for every n and t , $v'_{nt} = (v_{nt1}, \dots, v_{ntJ})$, and $w'_{nt} = (w_{nt1}, \dots, w_{ntJ})$.

Then, we can rewrite (2.2) as a generic linear model

$$y = X\beta + u = XZp + (I_N \otimes i_T)Fg + Vw.$$

Therefore, the generic GME problem selects $p, g, w \in (0, 1)$ to maximize

$$H(p, g, w) = -p' \log(p) - g' \log(g) - w' \log(w),$$

subject to $i_K = (I_K \otimes i'_M)p$, $i_N = (I_N \otimes i'_R)g$ and $i_{NT} = (I_{NT} \otimes i'_J)w$. Then, the Lagrangian function is

$$\begin{aligned} L = & -p' \log(p) - g' \log(g) - w' \log(w) + \lambda' [y - XZp - (I_N \otimes i_T)Fg - Vw] \\ & + \theta' [i_K - (I_K \otimes i'_M)p] + \gamma' [i_N - (I_N \otimes i'_R)g] + \tau' [i_{NT} - (I_{NT} \otimes i'_J)w]. \end{aligned} \quad (3.4)$$

Taking the gradient of L to derive the first-order conditions, we have

$$\begin{aligned} \nabla_p L &= -\log(p) - i_{KM} - Z'X'\lambda - (I_K \otimes i_M)\theta = 0, \\ \nabla_g L &= -\log(g) - i_{NR} - F'(I_N \otimes i'_T)\lambda - (I_N \otimes i_R)\gamma = 0, \\ \nabla_w L &= -\log(w) - i_{NTJ} - V'\lambda - (I_{NT} \otimes i_J)\tau = 0, \\ \nabla_\lambda L &= y - XZp - (I_N \otimes i_T)Fg - Vw = 0, \\ \nabla_\theta L &= i_K - (I_K \otimes i'_M)p = 0, \\ \nabla_\gamma L &= i_N - (I_N \otimes i'_R)g = 0, \\ \nabla_\tau L &= i_{NT} - (I_{NT} \otimes i'_J)w = 0. \end{aligned}$$

After some algebra, we obtain

$$p = \exp(-Z'X'\lambda) \odot \exp[-i_{KM} - (I_K \otimes i_M)\theta], \quad (3.5)$$

$$g = \exp(-F'(I_N \otimes i'_T)\lambda) \odot \exp[-i_{NR} - (I_N \otimes i_R)\gamma], \quad (3.6)$$

$$w = \exp(-V'\lambda) \odot \exp[-i_{NTJ} - (I_{NT} \otimes i_J)\tau], \quad (3.7)$$

where \odot denotes the Hadamard product.

Furthermore, since $\exp[-i_{KM} - (I_K \otimes i_M)\theta] = \{(I_K \otimes J_M) \exp(-Z'X'\lambda)\}^{\odot(-1)}$, (3.5) can be rewritten as

$$p = \exp(-Z'X'\lambda) \odot \{(I_K \otimes J_M) \exp(-Z'X'\lambda)\}^{\odot(-1)}, \quad (3.8)$$

where $\odot(-1)$ means the Hadamard inverse; that is, the elementwise reciprocation. In a similar way, (3.6) and (3.7) can be rewritten, respectively, as

$$g = \exp(-F'(I_N \otimes i'_T)\lambda) \odot \{(I_N \otimes J_R) \exp(-F'(I_N \otimes i'_T)\lambda)\}^{\odot(-1)}, \quad (3.9)$$

$$w = \exp(-V'\lambda) \odot \{(I_{NT} \otimes J_J) \exp(-V'\lambda)\}^{\odot(-1)}. \quad (3.10)$$

3.2. A dual version of the GME problem

Now we describe a dual version of the GME problem which may be solved with simpler and more widely available unconstrained numerical methods, and then propose a dual generalized maximum entropy (dual GME, D-GME) estimator.

Table 1: Comparison of the regression coefficient estimates in Model 1

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	0.142	0.142	0.145	0.145	0.152	0.158	0.166	0.154	0.158	0.122
0.2	0.180	0.164	0.166	0.166	0.188	0.187	0.209	0.184	0.187	0.148
0.4	0.216	0.159	0.160	0.161	0.204	0.183	0.234	0.188	0.183	0.145
0.6	0.243	0.143	0.144	0.144	0.206	0.169	0.257	0.179	0.169	0.126
0.8	0.289	0.127	0.127	0.127	0.208	0.154	0.273	0.170	0.154	0.108

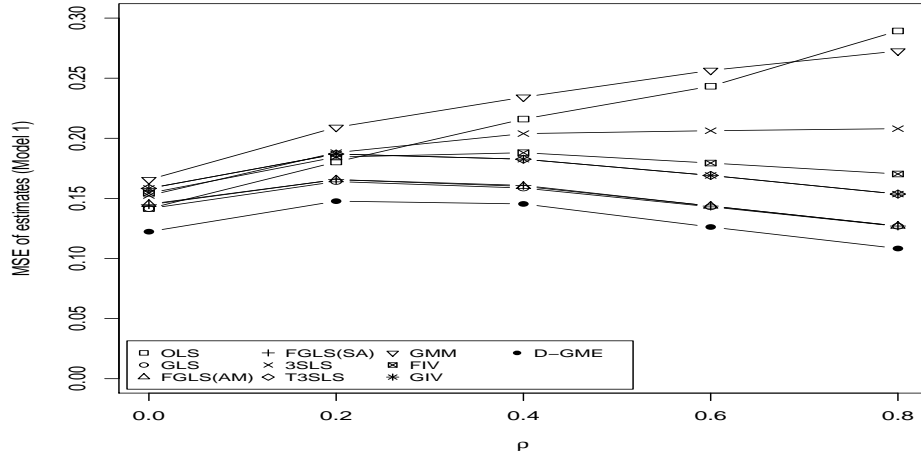


Figure 1: Comparison of the regression coefficient estimates in Model 1

Building on the Lagrangian function (3.4) and the solutions (3.8) through (3.10), the dual objective is

$$\begin{aligned}
\mathcal{L} &= -\mathbf{p}' \log(\mathbf{p}) - \mathbf{g}' \log(\mathbf{g}) - \mathbf{w}' \log(\mathbf{w}) + \lambda' [\mathbf{y} - \mathbf{XZp} - (\mathbf{I}_N \otimes \mathbf{i}_T) \mathbf{Fg} - \mathbf{Vw}] \\
&= -\mathbf{p}' [-\mathbf{Z}' \mathbf{X}' \lambda - \log(\mathbf{\Omega}(\lambda))] - \mathbf{g}' [-\mathbf{F}' (\mathbf{I}_N \otimes \mathbf{i}_T') \lambda - \log(\mathbf{\Phi}(\lambda))] \\
&\quad - \mathbf{w}' [-\mathbf{V}' \lambda - \log(\mathbf{\Psi}(\lambda))] + [\mathbf{y}' - \mathbf{p}' \mathbf{Z}' \mathbf{X}' - \mathbf{g}' \mathbf{F}' (\mathbf{I}_N \otimes \mathbf{i}_T') - \mathbf{w}' \mathbf{V}'] \lambda \\
&= \mathbf{y}' \lambda + \mathbf{p}' \log(\mathbf{\Omega}(\lambda)) + \mathbf{g}' \log(\mathbf{\Phi}(\lambda)) + \mathbf{w}' \log(\mathbf{\Psi}(\lambda)),
\end{aligned} \tag{3.11}$$

where $\mathbf{\Omega}(\lambda) = (\mathbf{I}_K \otimes \mathbf{J}_M) \exp(-\mathbf{Z}' \mathbf{X}' \lambda)$, $\mathbf{\Phi}(\lambda) = (\mathbf{I}_N \otimes \mathbf{J}_R) \exp(-\mathbf{F}' (\mathbf{I}_N \otimes \mathbf{i}_T') \lambda)$ and $\mathbf{\Psi}(\lambda) = (\mathbf{I}_{NT} \otimes \mathbf{J}_J) \exp(-\mathbf{V}' \lambda)$. More specifically, \mathcal{L} can be written by

$$\mathcal{L} = \mathbf{y}' \lambda + \sum_{k=1}^K \log(\Omega_k(\lambda)) + \sum_{n=1}^N \log(\Phi_n(\lambda)) + \sum_{n=1}^N \sum_{t=1}^T \log(\Psi_{nt}(\lambda)), \tag{3.12}$$

where $\Omega_k(\lambda) = \sum_m \exp(-z_{km} \sum_n \sum_t x_{knt} \lambda_{nt})$, $\Phi_n(\lambda) = \sum_r \exp(-f_{nr} \sum_t \lambda_{nt})$ and $\Psi_{nt}(\lambda) = \sum_j \exp(-v_{ntj} \lambda_{nt})$. Then, the dual GME estimator of β is defined as

$$\hat{\beta}_{\text{D-GME}} = \mathbf{Z} \hat{\mathbf{p}}, \tag{3.13}$$

where $\hat{\mathbf{p}} = \exp(-\mathbf{Z}' \mathbf{X}' \hat{\lambda}) \odot \{(\mathbf{I}_K \otimes \mathbf{J}_M) \exp(-\mathbf{Z}' \mathbf{X}' \hat{\lambda})\}^{\odot(-1)}$ from (3.8), and $\hat{\lambda} = \min_{\lambda} \mathcal{L}$ from (3.12).

Table 2: Comparison of the regression coefficient estimates in Model 2

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	1.036	1.036	1.032	0.978	1.369	1.465	1.084	1.411	1.465	0.296
0.2	0.936	1.028	0.948	0.892	1.243	1.414	0.992	1.357	1.414	0.278
0.4	0.809	0.931	0.829	0.797	1.033	1.277	0.857	1.217	1.277	0.282
0.6	0.625	0.739	0.642	0.629	0.742	1.031	0.652	0.964	1.031	0.292
0.8	0.449	0.444	0.406	0.410	0.406	0.653	0.446	0.592	0.653	0.272

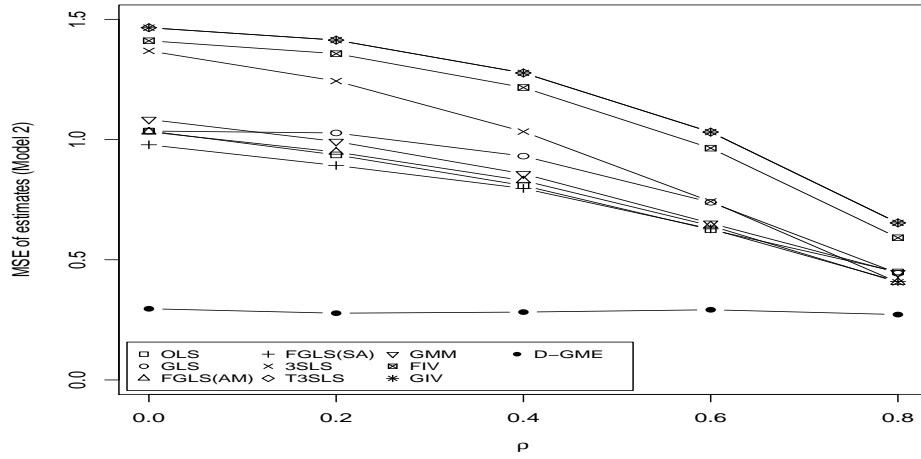


Figure 2: Comparison of the regression coefficient estimates in Model 2

3.3. Large sample properties of a dual GME estimator

Let $H_N(\lambda)$ (henceforth denoted as $H(\lambda)$) be the dual objective function based on new normed-moment formulation $X'y/N = X'XZp/N + (I_N \otimes i_T)Fg + Vw$ where $V = X'u/T$. Now we expand $\nabla_\lambda H(\lambda)$ about the true value λ_0 to get

$$\nabla_\lambda H(\lambda) = \nabla_\lambda H(\lambda_0) + \nabla_{\lambda\lambda'} H(\lambda^*)(\lambda - \lambda_0), \quad (3.14)$$

where λ^* is between λ and λ_0 . Replacing λ by $\hat{\lambda}$ in (3.14) and noting that $\nabla_\lambda H(\hat{\lambda}) = 0$, we obtain

$$\sqrt{N}(\hat{\lambda} - \lambda_0) = -[\nabla_{\lambda\lambda'} H(\lambda^*)]^{-1} [\sqrt{N}\nabla_\lambda H(\lambda_0)], \quad (3.15)$$

where $\sqrt{N}\nabla_\lambda H(\lambda_0) = X'u/\sqrt{N} + O(1/\sqrt{N})$ which converges in law to $N(\mathbf{0}, \Xi)$ by the following assumption; $E(u) = \mathbf{0}$, $Var(u) = \Sigma_u$, and $F(u)$ satisfies the Lindeberg condition and $\lim_N X'\Sigma_u X/N = \Xi$. It leads to $\lim_N \nabla_{\lambda\lambda'} H(\lambda_0) = Q\Sigma_Z Q'$, where $Q = \lim_N (X'X)/N$ and Σ_Z is the covariance matrix of p . Therefore the limiting distribution for $\hat{\lambda}$ is

$$\sqrt{N}(\hat{\lambda} - \lambda_0) \longrightarrow N(\mathbf{0}, (Q\Sigma_Z Q')^{-1} \Xi (Q\Sigma_Z Q')^{-1}). \quad (3.16)$$

Now, by the continuity of β and the delta-method, the limiting distribution for $\hat{\beta}_{D-GME}$ is

$$\sqrt{N}(\hat{\beta}_{D-GME} - \beta_0) \longrightarrow N(\mathbf{0}, Q^{-1} \Xi Q^{-1}). \quad (3.17)$$

Refer to Golan *et al.* (1996a) for large sample properties of more general GME estimators.

Table 3: Comparison of the prediction risks in Model 1

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	0.534	0.534	0.541	0.539	0.557	0.566	0.576	0.560	0.566	0.523
0.2	0.624	0.600	0.603	0.602	0.639	0.636	0.670	0.633	0.636	0.589
0.4	0.687	0.601	0.604	0.605	0.675	0.644	0.721	0.651	0.644	0.592
0.6	0.744	0.599	0.600	0.600	0.697	0.643	0.767	0.657	0.643	0.580
0.8	0.817	0.572	0.572	0.572	0.712	0.624	0.808	0.653	0.624	0.540

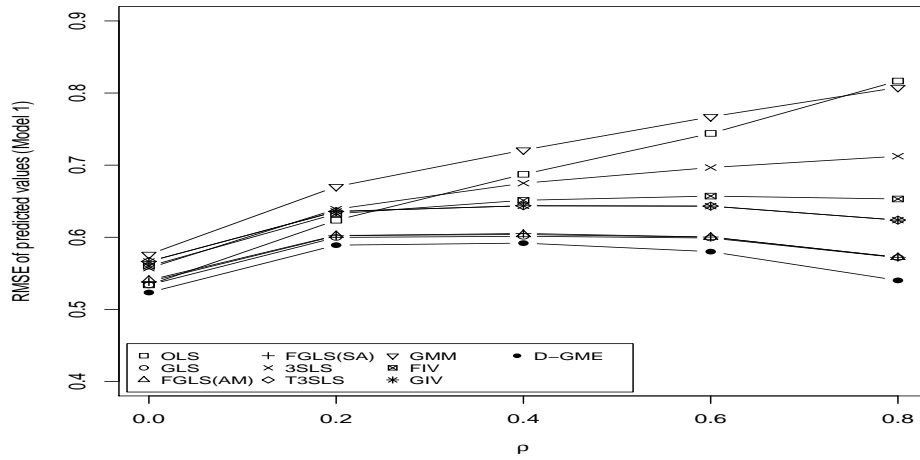


Figure 3: Comparison of the prediction risks in Model 1

4. Simulated Examples

In this study, we carried out four different simulation studies to assess the performance of our dual GME estimator.

In the first study, data came from a classical panel data regression model (Model 1). As introduced by Nerlove (1971) and used by many authors (Arora, 1973; Baltagi, 1981; Heckman, 1981; Wansbeek and Kapteyn, 1989), the exogenous variables were generated by $x_{nt} = 0.1t + 0.5x_{n,t-1} + w_{nt}$, where w_{nt} is uniformly distributed on the interval $[-0.5, 0.5]$ and $x_{n0} = 5 + 10w_{n0}$. Furthermore, x_{11}, \dots, x_{NT} were independently drawn $K - 1$ times and then we constructed a design matrix \mathbf{X} .

In the error term, μ_n and ε_{nt} were independently generated from $N(0, \sigma_\mu^2)$ and $N(0, \sigma_\varepsilon^2)$, respectively, where $\sigma_\mu^2 = \rho\sigma^2$ and $\sigma_\varepsilon^2 = (1 - \rho)\sigma^2$ with $\sigma^2 = \sigma_\mu^2 + \sigma_\varepsilon^2$. We set $N = 25$, $T = 5$, and $\sigma^2 = 10$ in all simulations of this paper. However, ρ varied over the set $(0, 0.2, 0.4, 0.6, 0.8)$ such that $(1 - \rho)$ is always positive.

As for the formation of support matrices (\mathbf{Z} , \mathbf{F} , and \mathbf{V} in Section 3) for parameter β , individual effect μ , and error ε , we assumed that the diagonal elements of each support matrix are the same; that is, $z_1 = \dots = z_K$ in \mathbf{Z} of (3.1), $f_1 = \dots = f_N$ in \mathbf{F} of (3.2), and $v_{11} = \dots = v_{NT}$ in \mathbf{V} of (3.3).

We first computed $\hat{\beta}_{OLS}$ and then the plausible upper and lower bounds of z were computed by $\pm 3 \times [\max |\hat{\beta}_{OLS}|]$, where $[\cdot]$ indicates $\min\{c \in \mathbb{Z} | c \geq \cdot\}$. Likewise, to find the reasonable upper and lower bounds of f and v , we used $\check{\sigma}_\mu^2$ and $\check{\sigma}_\varepsilon^2$ of Swamy and Arora (1972). We also assumed that all support vectors, z , f , and v , consist of equally distanced discrete points and are centered at zero. We

Table 4: Comparison of the prediction risks in Model 2

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	2.580	2.580	2.574	2.083	3.127	3.259	2.676	3.187	3.259	1.623
0.2	2.251	2.412	2.281	1.977	2.697	2.942	2.345	2.856	2.942	1.326
0.4	1.884	2.111	1.945	1.787	2.221	2.539	1.963	2.453	2.539	1.165
0.6	1.492	1.718	1.565	1.505	1.681	2.051	1.545	1.963	2.051	1.063
0.8	1.101	1.182	1.116	1.112	1.074	1.413	1.112	1.329	1.413	0.920

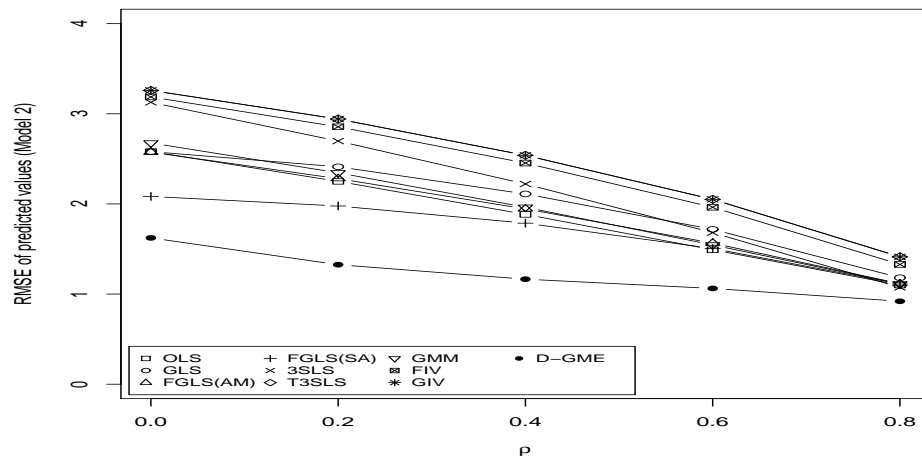


Figure 4: Comparison of the prediction risks in Model 2

fixed $M = 5$ for the number of support points since Golan *et al.* (1996a) showed that the variance of the point estimator decreases as M increases but the empirical prediction risk as well as computational burden also increase; additionally, they argued that using $M = 5$ gave the greatest improvement in precision through empirical studies; see Section 8.7 in Golan *et al.* (1996a) for more details.

The dual GME estimation can be performed using any standard nonlinear optimization methods. Here, we used a Newton-type algorithm for our computation. We used R software, invoking the nonlinear minimization (nlm; Dennis and Schnabel, 1983; Schnabel *et al.*, 1985) routine to calculate the dual GME estimates. The nlm function carries out a minimization of an objective function using a Newton-type algorithm. For numerical optimizations involving Newton-type algorithms, the choice of initial values is critical. For our dual GME procedure, extensive preliminary simulations indicated that the estimation is not sensitive to the choice of initial values at all. In fact, we conducted simulations where the initial values were randomly chosen from a normal distribution.

For the second study we considered a panel data regression model with endogenous explanatory variables (Model 2). We followed the simulation scheme given by Everaert and Pozzi (2014) of setting endogeneity which ε_{nt} is expected to be contemporaneously correlated with some regressors. Especially, in our design matrix, the fourth regressor was computed by $x_{4nt} + 0.4\varepsilon_{nt}$, which results in an inconsistent estimator. Other settings are the same as in Model 1.

The third simulation study examined the collinearity problem in a panel data regression model (Model 3). Song and Cheon (2006) proposed the GME estimator to handle collinearity in a panel regression model with strictly exogenous explanatory variables. In order to form a design matrix in

Table 5: Comparison of the regression coefficient estimates in Model 3

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	48.158	48.158	47.409	48.482	51.720	55.183	59.331	52.346	55.183	4.524
0.2	47.178	44.597	44.009	44.820	51.662	51.166	54.419	49.630	51.166	5.248
0.4	47.581	33.797	33.842	34.213	48.020	39.308	56.032	39.911	39.308	5.727
0.6	46.064	23.308	23.213	23.392	44.541	29.188	50.234	30.557	29.188	5.721
0.8	44.430	12.121	12.118	12.135	36.609	15.380	45.007	17.125	15.380	4.633

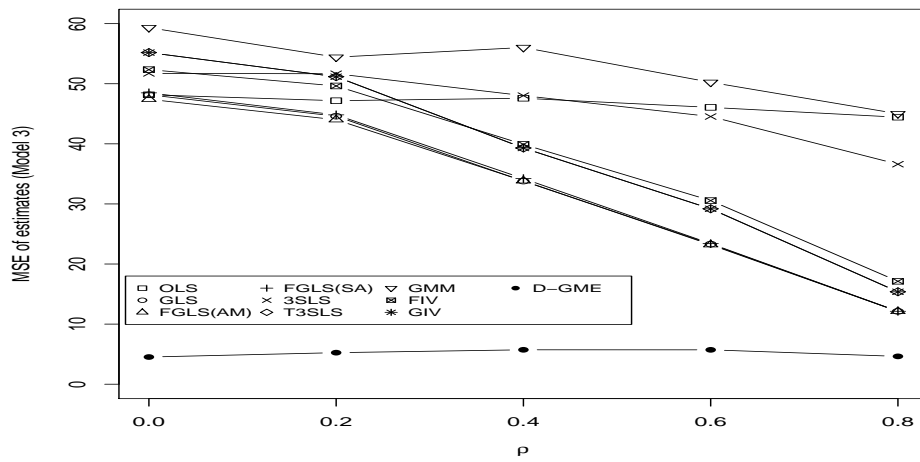


Figure 5: Comparison of the regression coefficient estimates in Model 3

which there is a linear relationship among one or more of the independent variables, as in Golan *et al.* (1996a) they created a new design matrix by replacing the singular values of X with $(\sqrt{2/(1+\kappa)}, 1, 1, \sqrt{2\kappa/(1+\kappa)})$ where κ is a condition number of $X'X$.

However, their methodology has some imitations. First, they used a least squares method to compute the GME estimates, instead of a dual Lagrange multiplier method suggested by Golan *et al.* (1996a). Consequently, their estimates computed by a least squares method do not satisfy one among the optimal conditions of Lagrangian function in (3.4) of the Section 3; *i.e.*, $\nabla_{\lambda} L = y - XZp - (I_N \otimes I_T)Fg - Vw = 0$. Second, their new design matrix using the modified singular value decomposition is appropriate only in classical regression models. That is, in panel data regression models their formation of a design matrix with a desired condition number causes the nature of time series to be lost. Finally, their iterative method for solving the least squares problem is highly sensitive to the choice of initial values.

In contrast, we solved an unconstrained dual GME problem and found more efficient estimates. In order to construct collinear panel data series, as in Belsley (1991, Chapter 4), the fourth regressor (X_4) was constructed as $X_4 = 0.2X_2 + \varepsilon$, where $\varepsilon_{nt} \sim N(0, 10^{-2}\text{Var}(0.2X_2)I)$. Once again, we used normal-distributed random numbers for starting points to compute our dual GME estimates.

In the last study, we considered an integrated model with Model 2 and 3; a collinear panel data regression model with endogenous explanatory variables (Model 4). As in Model 2 and 3, we replaced the fourth regressor with $x_{4nt} + 0.4\varepsilon_{nt}$ and the third regressor with $0.2x_{2nt} + \varepsilon_{nt}$, where $\varepsilon_{nt} \sim N(0, 10^{-2}\text{Var}(0.2x_{2nt})I)$.

Table 6: Comparison of the prediction risks in Model 3

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	0.533	0.533	0.535	0.536	0.553	0.562	0.583	0.555	0.562	0.453
0.2	0.589	0.566	0.567	0.569	0.609	0.603	0.632	0.601	0.603	0.506
0.4	0.652	0.579	0.581	0.582	0.649	0.619	0.687	0.625	0.619	0.533
0.6	0.728	0.592	0.593	0.594	0.689	0.630	0.750	0.645	0.630	0.554
0.8	0.764	0.563	0.564	0.564	0.676	0.601	0.760	0.619	0.601	0.528

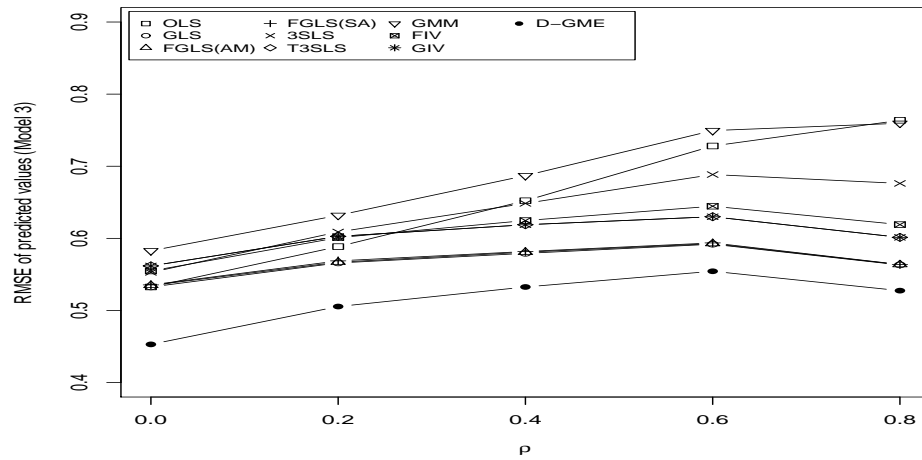


Figure 6: Comparison of the prediction risks in Model 3

Table 7: Comparison of the regression coefficient estimates in Model 4

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	18.652	18.652	18.514	86.871	10.952	8.541	18.037	9.180	8.541	1.991
0.2	26.759	17.455	23.605	93.081	20.468	9.393	26.396	11.569	9.393	2.017
0.4	32.847	15.112	23.567	63.579	29.327	10.290	33.981	13.176	10.287	2.011
0.6	37.001	12.348	18.274	32.860	35.446	10.690	38.037	13.932	10.694	1.983
0.8	43.397	8.5141	11.153	14.060	37.022	9.614	43.482	12.889	9.614	1.889

In order to assess the performance of $\hat{\beta}_{\text{D-GME}}$, we considered $\beta' = (\beta_1, \beta_2, \beta_3, \beta_4) = (1, 2, 3, 4)$. We implemented our dual GME computational algorithm described above with 1000 Monte Carlo replications. In order to avoid any sampling bias and ensure fair comparison between our estimate and others, we computed each of the ten estimates (OLS, GLS, FGLS(AM), FGLS(SA), 3SLS, T3SLS, GMM, FIV, GIV, and D-GME) using the same generated data set in each replication. We reported the mean squared error of each estimator($\hat{\beta}$) and the root mean squared error(RMSE) of predicted values(\hat{y}) for each model. MSE and RMSE were estimated by $(\sum_{i=1}^K \text{MSE}_i)/K$ and $(\sum_{i=1}^K \text{RMSE}_i)/K$ respectively, where $\text{MSE}_i(\hat{\beta}) = \text{Var}(\hat{\beta}) + [E(\hat{\beta}) - \beta]^2$ and $\text{RMSE}_i = \sqrt{\text{MSE}_i(\hat{\beta})}$ for i, \dots, K . In this simulation, we set $K = 10$. For ease of comparison, the smallest MSE and RMSE were identified in bold.

Tables and Figures 1 and 3 show that for Model 1, the dual GME is more efficient than the other

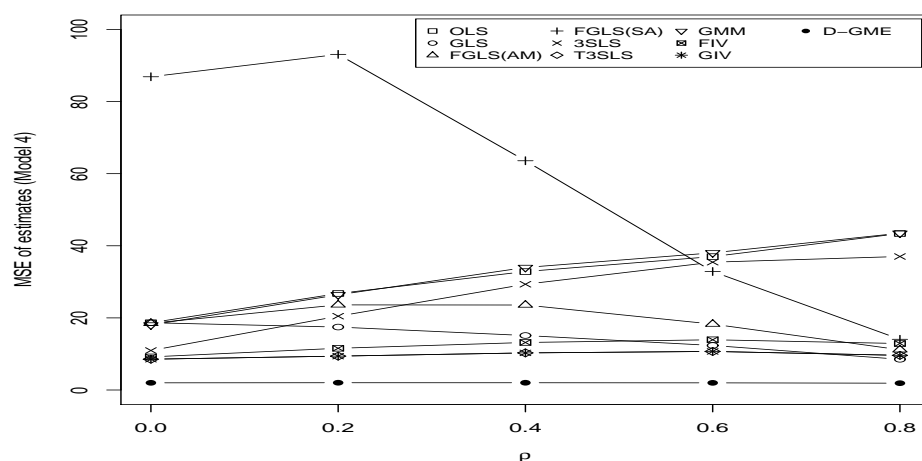


Figure 7: Comparison of the regression coefficient estimates in Model 4

Table 8: Comparison of the prediction risks in Model 4

ρ	OLS	GLS	FGLS (AM)	FGLS (SA)	3SLS	T3SLS	GMM	FIV	GIV	D-GME
0.0	2.532	2.532	2.530	2.118	3.091	3.248	2.618	3.169	3.248	1.548
0.2	2.216	2.360	2.232	2.005	2.651	2.944	2.304	2.848	2.944	1.449
0.4	1.818	2.034	1.871	1.739	2.123	2.503	1.876	2.404	2.503	1.310
0.6	1.433	1.644	1.499	1.449	1.567	2.002	1.466	1.902	2.002	1.152
0.8	1.015	1.118	1.042	1.035	0.962	1.322	1.027	1.226	1.322	0.909

estimators and the dual GME prediction risk is smaller than the others. For Models 2, 3, and 4, as expected, the performance of all estimators worsen compared to Model 1 (Tables and Figures 2 through 8). Interestingly, all estimates except for the dual GME are worse, and the dual GME once again is more efficient than the others. When there are severe problems with collinearity and endogeneity like Models 3 and 4, which their condition numbers are more than 100, the dual GME is superior in comparison with the others (Tables and Figures 5 and 7). In summary, these simulation results show that the dual GME is a preferred choice over the OLS, GLS, FGLS, 3SLS, GMM, FIV, and GIV in simple, endogenous, as well as collinear panel models.

5. Conclusion

In this paper we have introduced a new dual GME estimator for panel data regression models under the assumption of exogeneity, endogeneity, or/and collinearity of the regressors. A GME estimator was initially introduced by Song and Cheon (2006) for the panel data model with strict exogeneity assumption, but we have modified their critical limitations. As compared to Song and Cheon's GME estimator, our dual GME estimator is easy to compute because we applied unconstrained numerical method rather than constrained one and turned out to be less sensitive to the choice of initial values through the simulation studies. Our estimates were obtained using truly simulated panel data; however, Song and Cheon's estimates were not.

It is known that the dual GME methodology is applicable when the underlying sampling model

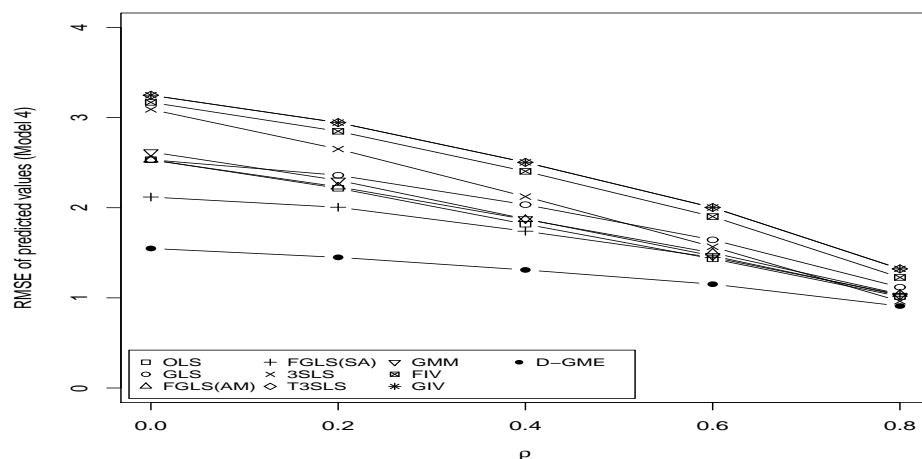


Figure 8: Comparison of the prediction risks in Model 4

is incompletely or incorrectly known and the data are limited, partial or incomplete; however, such situations are not entirely avoidable in econometrics. With Monte Carlo simulations, we have shown that the dual GME estimator is a better alternative to classical least squares estimators, instrumental estimators and GME estimator, especially in a small set of data, to handle ill-posed problems in panel data and some violations of panel data regression assumptions. Hence, the entropy metric and the maximum entropy formalism can be extended to more complicated econometric models.

References

- Ahn, S. C. and Schmidt, P. (1999). Estimation of linear panel data models using GMM, in *Generalized method of Moments Estimation*, L. Mátyás (Eds.), Cambridge University Press, Cambridge.
- Amemiya, T. (1971). The estimation of the variances in a variance-components models, *International Economic Review*, **12**, 1–13.
- Arora, S. (1973). Error components regression models and their applications, *Annals of Economic and Social Measurement*, **2**, 451–462.
- Baltagi, B. H. (1981). Pooling: An experimental study of alternative testing and estimation procedures in a two-way error component model, *Journal of Econometrics*, **17**, 21–49.
- Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*, John Wiley, New York.
- Belsley, D. A. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression*, John Wiley & Sons, New York.
- Dennis, J. E. and Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Everaert, G. and Pozzi, L. (2014). The predictability of aggregate consumption growth in OECD countries: A panel data analysis, *Journal of Applied Econometrics*, **29**, 431–453.
- Golan, A. and Judge, G. (1996). Recovering information in the case of underdetermined problems and incomplete data, *Journal of Statistical Planning and Inference*, **49**, 127–136.
- Golan, A., Judge, G. and Miller, D. (1996a). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, Chichester.

- Golan, A., Judge, G. and Perloff, J. (1996b). A maximum entropy approach to recovering information from multinomial response data, *Journal of the American Statistical Association*, **91**, 841–853.
- Golan, A., Judge, G. and Perloff, J. (1997). Estimation and inference with censored and ordered multinomial response data, *Journal of Econometrics*, **79**, 23–51.
- Golan, A., Judge, G. and Robinson, S. (1994). Recovering information in the case of partial multisectorial economic data, *Review of Economics and Statistics*, **76**, 541–549.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029–1054.
- Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimation a discrete time-discrete data stochastic process, in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Application*, MIT Press, Cambridge, Massachusetts.
- Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge University Press, Cambridge.
- Im, K. S., Ahn, S. C., Schmidt, P. and Wooldridge, M. J. (1999). Efficient estimation of panel data models with strictly exogenous explanatory variables, *Journal of Econometrics*, **93**, 177–201.
- Jaynes, E. T. (1957a). Information theory and statistical mechanics I, *Physics Review*, **106**, 620–630.
- Jaynes, E. T. (1957b). Information theory and statistical mechanics II, *Physics Review*, **108**, 171–190.
- Jaynes, E. T. (1984). Prior information and ambiguity in inverse problems, In McLaughlin D. W. (Ed.) *Inverse Problems*, 151–166, SIAM Proceedings, American Mathematical Society, Providence, Rhode Island.
- Judge, G. and Golan, A. (1992). Recovering information in the case of ill-posed inverse problems with noise, University of California at Berkeley, *Unpublished Paper*.
- Kullback, J. (1959). *Information Theory and Statistics*, John Wiley, New York.
- Levine, R. D. (1980). An Information theoretical approach to inverse problems, *Journal of Physics A*, **13**, 91–108.
- Levine, R. D. and Tribus, M. (1979). *The Maximum Entropy Formalism*, MIT Press, Cambridge, MA.
- Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections, *Econometrica*, **39**, 359–382.
- Schnabel, R. B., Koontz, J. E. and Weiss, B. E. (1985). A modular system of algorithms for unconstrained minimization, *ACM Transactions on Mathematical Software*, **11**, 419–440.
- Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423.
- Song, S. H. and Cheon, S. Y. (2006). A study of generalized maximum entropy estimator for the panel regression model, *The Korean Journal of Applied Statistics*, **19**, 521–534.
- Swamy, P. A. V. B. and Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models, *Econometrica*, **40**, 261–275.
- Wansbeek, T. and Kapteyn, A. (1989). Estimation of the error-components model with incomplete panels, *Journal of Econometrics*, **41**, 341–361.
- Wooldridge, J. M. (2003). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge.