J. Japan Statist. Soc.
Vol. 42 No. 2 2012 109–123

GMM ESTIMATION OF SHORT DYNAMIC PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

Kazuhiko Hayakawa*

In this paper, we propose GMM estimators for short dynamic panel data models with interactive fixed effects. Moment conditions are obtained for the model where the projection method is applied to remove the correlation between regressors and interactive fixed effects. Monte Carlo simulation shows that the proposed GMM estimators perform reasonably well in finite sample.

Key words and phrases: Factor structure, GMM, interactive fixed effects, panel data, projection method.

1. Introduction

Recently, the use of panel data has been increasing in empirical studies of economic problems. A basic model for panel data is given by

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x}_{it} + \lambda_i + \gamma_t + v_{it}, \qquad (i = 1, \dots, N; t = 1, \dots, T) \tag{1.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{x}_{it}$ are K dimensional vectors. The index i denotes cross-sectional unit and the index t denotes time period. λ_i and γ_t are unobserved individual and time effects, respectively, which are typical for panel models. v_{it} is an idiosyncratic error term. This model has been widely used in empirical studies[1]. If we regard λ_i as the parameter to be estimated, the model (1.1) is called the fixed effects model and the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ is known as the within-groups (WG) estimator[2]. On the other hand, if we regard λ_i as a part of the disturbance, the model (1.1) is called the random effects model, and the generalized least squares (GLS) estimator is usually used. The most important difference between these two estimators is that the WG estimator is consistent even if λ_i is correlated with $\boldsymbol{x}_{it}$, while it is not the case for the GLS estimator.

Although, these two estimators are widely used in empirical studies, the model (1.1) is somewhat restrictive. For instance, if the time effect γ_t denotes an economic shock, then, the model (1.1) indicates that the shock has identical effects on all cross-sectional units, which is unlikely to hold in practice. To relax the restriction of the model (1.1), let us consider a model with a factor structure as follows:

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x}_{it} + \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t + v_{it} \tag{1.2}$$

Received April 14, 2008. Revised February 18, 2012. Accepted December 1, 2012.

*Department of Economics, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8525, Japan.

1 For the review of the model (1.1), see Hsiao (2003), Baltagi (2008) and Arellano (2003).

2 The WG estimator is also called the least squares dummy variable (LSDV) estimator or fixed effects estimator.

where $\boldsymbol{\lambda}_i$ and $\boldsymbol{\gamma}_t$ are unobserved m dimensional vectors. $\boldsymbol{\lambda}_i$ denotes the factor loadings and $\boldsymbol{\gamma}_t$ denotes the common factors. Since $\boldsymbol{\lambda}_i$ and $\boldsymbol{\gamma}_t$ enter the model interactively, Bai (2009b) calls the model (1.2) the *interactive fixed effects model*. It is easy to see that in the model (1.2) shocks $\boldsymbol{\gamma}_t$ have heterogeneous effects on cross-sectional units through $\boldsymbol{\lambda}_i$. However, it should be noted that there are two interpretations for the interactive term $\boldsymbol{\lambda}_i'\gamma_t$, i.e., time varying individual effects and cross section dependence[3]. Some examples of economic applications that are relevant to (1.2) are provided by Ahn *et al.* (2001) and Bai (2009b). There are several papers that study the model (1.2). Ahn *et al.* (2001, 2010) propose the generalized method of moments (GMM) estimators in a large N and fixed T context, while Pesaran (2006) and Bai (2009b) propose least squares estimators in a large N and large T context.

Models (1.1) and (1.2) are static. However, as Nerlove (2002) says, the economic behavior is dynamic in its nature. Hence, it is important to account for the dynamics in the model. A common approach is to use the lagged dependent variable of y_{it} as an explanatory variable as follows:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}'\boldsymbol{x}_{it} + \lambda_i + \gamma_t + v_{it}.$$

This type of model was first studied by Balestra and Nerlove (1966) and is often called the dynamic panel data model. After this study, a lot of papers proposed several estimators and discussed their properties. These include Nickell (1981), Anderson and Hsiao (1981, 1982), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995, 1997), and Blundell and Bond (1998) to mention a few. Nickell (1981) shows that the WG estimator is inconsistent when T is fixed and N is large. Other studies consider the estimation with instrumental variables.

As in the static panel models, several papers consider a dynamic panel data model with interactive fixed effects. These include Holtz-Eakin *et al.* (1988), Meghir and Windmeijer (1999), Phillips and Sul (2003, 2007), Nauges and Thomas (2003), Sarafidis *et al.* (2009), Sarafidis and Robertson (2009), Bai (2009a), Sarafidis and Yamagata (2010) and Moon and Weidner (2010). They consider a variant of the following dynamic panel data model with interactive fixed effects[4]:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}'\boldsymbol{x}_{it} + \boldsymbol{\lambda}_i'\gamma_t + v_{it}.$$

Phillips and Sul (2003) investigated the effect of cross section dependence on the behavior of the WG estimator, and proposed the generalized median unbiased estimator. Phillips and Sul (2007) derives the large N and fixed T asymptotic properties of the WG estimator and proposed a panel feasible generalized mean unbiased estimator to reduce the bias of the WG estimator. Sarafidis and Yamagata (2010) and Moon and Weidner (2010) propose instrumental variables and quasi maximum likelihood estimators in a large N and large T model, respectively. Bai (2009a) proposes a likelihood based estimator in a small T and large

3 See Sarafidis and Wansbeek (2012) for a recent survey of the issue of cross section dependence.

4 In fact, Phillips and Sul (2003, 2007) and Sarafidis and Robertson (2009) consider an AR(1) model.

N model. Meghir and Windmeijer (1999) consider a GMM estimation of dynamic panel data models with interactive fixed effects and ARCH errors. While these studies mainly focus on the estimation, Sarafidis *et al.* (2009) propose a statistic to test the presence of cross section dependence. Sarafidis and Robertson (2009) show that including the time effects mitigates the effect of cross section dependence on the first-difference and system GMM estimators.

This paper tries to make a contribution to this literature of dynamic panel data models with interactive fixed effects. Specifically, we propose GMM estimators that are consistent and asymptotically normal for large N and small T by using the projection method of Chamberlain (1982, 1984). Monte Carlo simulations are carried out to investigate the finite sample performance of the GMM estimator. Simulation results show that the proposed GMM estimator performs reasonably well in finite sample.

The remainder of the paper is organized as follows. In Section 2, we introduce the model and assumptions. In Section 3, we rewrite the approach of Ahn *et al.* (2010) in the context of dynamic panel data models and propose new GMM estimators. In Section 4, we conduct a Monte Carlo simulation to investigate the finite sample performance of the proposed GMM estimator. Finally, we conclude in Section 5.

2. Model

We consider the following dynamic panel data model with interactive fixed effects:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \boldsymbol{\lambda}_i' \boldsymbol{\gamma}_t + v_{it}, \qquad (i = 1, \ldots, N; t = 1, \ldots, T) \tag{2.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{x}_{it}$ are $K \times 1$ vectors. $\boldsymbol{\lambda}_i$ and $\boldsymbol{\gamma}_t$ are $m \times 1$ unobserved factor loadings and factors. The regressors $\boldsymbol{x}_{it}$ can be correlated with $\boldsymbol{\lambda}_i$ as in the usual fixed effects model but assume strict exogeneity in the sense that $E(\boldsymbol{x}_{is} v_{it}) = \mathbf{0}$ for all t and s. We also assume that $|\alpha| < 1$ and the number of factors m is known. The idiosyncratic error term v_{it} is assumed to be $v_{it} \sim iid(0, \sigma_v^2)$. The panel data we consider consist of short time series and a large number of cross sectional units, therefore, asymptotics are taken with large N and fixed T. Since time effects are consistently estimated with large N, we assume that $\boldsymbol{\gamma}_t$ is a non-random parameters to be estimated, while $\boldsymbol{\lambda}_i$ are random variables that can be correlated with $\boldsymbol{x}_{it}$.

The model (2.1) is simplified in special case. If we let $\boldsymbol{\lambda}_i = (\eta_i, 1)'$ and $\boldsymbol{\gamma}_t = (1, \kappa_t)'$, we have a usual dynamic panel data model:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \eta_i + \kappa_t + v_{it}.$$

If we let $\boldsymbol{\gamma}_t = (1, t)'$, we have a dynamic panel data models with heterogeneous time trend:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \lambda_{1i} + \lambda_{2i} t + v_{it}.$$

This type of model was studied by Wansbeek and Knaap (1999). For this model, we can use a GMM estimator for models in second-differences because taking the second-difference removes both λ_{1i} and λ_{2i}. However, if $\boldsymbol{\gamma}_t$ is a nonlinear function of t, such a simple approach is not usable and we need to consider a more general approach, which is the main purpose of this paper.

For later use, we rewrite the model (2.1) in a compact form as follows:

$$y_{it} = \boldsymbol{\delta}'\boldsymbol{w}_{it} + \boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t + v_{it} \qquad (i = 1, \ldots, N; t = 1, \ldots, T) \tag{2.2}$$

where $\boldsymbol{\delta} = (\alpha, \boldsymbol{\beta}')'$ and $\boldsymbol{w}_{it} = (y_{i,t-1}, \boldsymbol{x}_{it}')'$. Stacking this model over time, we have

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\delta} + \boldsymbol{\Gamma}\boldsymbol{\lambda}_i + \boldsymbol{v}_i \qquad (i = 1, \ldots, N) \tag{2.3}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iT})'$, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_T)'$ and $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{iT})'$.

3. GMM estimators

In this section, we first review the approach proposed by Ahn *et al.* (2010), and then propose a new approach.

3.1. Quasi-difference approach

To obtain a consistent estimator of $\boldsymbol{\delta}$, Ahn *et al.* (2010) remove the interactive fixed effects term $\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t$ by quasi-difference[5]. Although Ahn *et al.* (2010) consider a static model, their approach is easily extended to a dynamic case[6].

For the general multiple factor case, we need to impose m^2 restrictions to identify the factor structure since $\boldsymbol{\lambda}_i'\boldsymbol{\gamma}_t = (\boldsymbol{\lambda}_i'\boldsymbol{C})(\boldsymbol{C}^{-1}\boldsymbol{\gamma}_t) = \boldsymbol{\lambda}_i^{*'}\boldsymbol{\gamma}_t^*$ for any invertible $m \times m$ matrix $\boldsymbol{C}$. Ahn *et al.* (2010) consider the following restriction:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Psi} \\ \boldsymbol{I}_m \end{bmatrix}$$

where $\boldsymbol{\Psi}$ is a $(T - m) \times m$ matrix. In this case, the assumption that the bottom $m \times m$ matrix is $\boldsymbol{I}_m$ imposes m^2 restrictions which are required for identification. Let us define

$$\boldsymbol{\Xi} = \begin{bmatrix} \boldsymbol{I}_{T-m} \\ -\boldsymbol{\Psi}' \end{bmatrix}, \qquad \boldsymbol{\Psi}' = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{T-m}].$$

Multiplying $\boldsymbol{\Xi}'$ by (2.3), we have

$$\boldsymbol{\Xi}'\boldsymbol{y}_i = \boldsymbol{\Xi}'\boldsymbol{W}_i\boldsymbol{\delta} + \boldsymbol{\Xi}'\boldsymbol{v}_i \tag{3.1}$$

where we used $\boldsymbol{\Xi}'\boldsymbol{\Gamma} = \boldsymbol{0}$. By letting $\dot{\boldsymbol{y}}_i = (y_{i1}, \ldots, y_{i,T-m})'$, $\ddot{\boldsymbol{y}}_i = (y_{i,T-m+1}, \ldots, y_{iT})'$, $\dot{\boldsymbol{W}} = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{i,T-m})'$, $\ddot{\boldsymbol{W}} = (\boldsymbol{w}_{i,T-m+1}, \ldots, \boldsymbol{w}_{iT})'$, $\dot{\boldsymbol{v}}_i = (v_{i1}, \ldots, v_{i,T-m})'$ and $\ddot{\boldsymbol{v}}_i = (v_{i,T-m+1}, \ldots, v_{iT})'$, (3.1) can be written as

$$\begin{aligned} \dot{\boldsymbol{y}}_i &= \dot{\boldsymbol{W}}_i\boldsymbol{\delta} + \boldsymbol{\Psi}\ddot{\boldsymbol{y}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{W}}_i\boldsymbol{\delta} + \dot{\boldsymbol{v}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{v}}_i \\ &= \dot{\boldsymbol{W}}_i\boldsymbol{\delta} + (\boldsymbol{I}_{T-m} \otimes \ddot{\boldsymbol{y}}_i')\boldsymbol{\psi} - (\mathrm{vec}(\ddot{\boldsymbol{W}}_i)' \otimes \boldsymbol{I}_{T-m})\,\mathrm{vec}(\boldsymbol{\delta}' \otimes \boldsymbol{\Psi}) + \dot{\boldsymbol{v}}_i - \boldsymbol{\Psi}\ddot{\boldsymbol{v}}_i \end{aligned} \tag{3.2}$$

5 Alternatively, we may use the method by Holtz-Eakin *et al.* (1988) to remove the interactive fixed effects.

6 See also Sarafidis and Wansbeek (2012).

where $\boldsymbol{\psi} = \mathrm{vec}(\boldsymbol{\Psi}')$. The tth equation is given by

$$y_{it} = \boldsymbol{\delta}'\boldsymbol{w}_{it} + \boldsymbol{\psi}_t'\ddot{\boldsymbol{y}}_i - \boldsymbol{\psi}_t'\ddot{\boldsymbol{W}}_i\boldsymbol{\delta} + \widetilde{u}_{it} \qquad (i = 1, \ldots, N; t = 1, \ldots, T - m) \tag{3.3}$$

where $\widetilde{u}_{it} = v_{it} - \boldsymbol{\psi}_t'\ddot{\boldsymbol{v}}_i$. Since $y_{i0}, \ldots, y_{i,t-1}$ and $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}$ are uncorrelated with $\widetilde{u}_{it}$, we have the following moment conditions:

$$E[\widetilde{\boldsymbol{z}}_{it}\widetilde{u}_{it}] = \mathbf{0} \qquad (t = 1, \ldots, T - m)$$

where $\widetilde{\boldsymbol{z}}_{it} = (y_{i0}, \ldots, y_{i,t-1}, \boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT}')'$, or in a matrix form,

$$E[\widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{u}}_i] = \mathbf{0} \tag{3.4}$$

where $\widetilde{\boldsymbol{Z}}_i = \mathrm{diag}(\widetilde{\boldsymbol{z}}_{i1}', \ldots, \widetilde{\boldsymbol{z}}_{i,T-m}')$ and $\widetilde{\boldsymbol{u}}_i = (\widetilde{u}_{i1}, \ldots, \widetilde{u}_{i,T-m})'$. Note that $\widetilde{\boldsymbol{u}}_i$ depends on unknown parameters $\boldsymbol{\theta}_{QD} = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$. Since the number of parameters is $K + 1 + m(T - m)$ and that of moment conditions is $(T - m)(T - m + 1)/2 + KT(T - m)$, we require $(T - m)(T - m + 1)/2 + KT(T - m) \geq K + 1 + m(T - m)$ for identification[7]. The GMM estimators based on moment conditions (3.4) are given as follows:

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{QD(1step)} &= \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{QD(1step)} \\ \widehat{\boldsymbol{\psi}}_{QD(1step)} \end{pmatrix} \\ &= \underset{\delta,\psi}{\mathrm{argmin}} \left(\sum_{i=1}^N \widetilde{\boldsymbol{u}}_i'\widetilde{\boldsymbol{Z}}_i\right)\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{Z}}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{u}}_i\right), \end{aligned} \tag{3.5}$$

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{QD(2step)} &= \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{QD(2step)} \\ \widehat{\boldsymbol{\psi}}_{QD(2step)} \end{pmatrix} \\ &= \underset{\delta,\psi}{\mathrm{argmin}} \left(\sum_{i=1}^N \widetilde{\boldsymbol{u}}_i'\widetilde{\boldsymbol{Z}}_i\right)\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widehat{\widetilde{\boldsymbol{u}}}_i\widehat{\widetilde{\boldsymbol{u}}}_i'\widetilde{\boldsymbol{Z}}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{u}}_i\right), \end{aligned} \tag{3.6}$$

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{QD(CUE)} &= \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{QD(CUE)} \\ \widehat{\boldsymbol{\psi}}_{QD(CUE)} \end{pmatrix} \\ &= \underset{\delta,\psi}{\mathrm{argmin}} \left(\sum_{i=1}^N \widetilde{\boldsymbol{u}}_i'\widetilde{\boldsymbol{Z}}_i\right)\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{u}}_i\widetilde{\boldsymbol{u}}_i'\widetilde{\boldsymbol{Z}}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{\boldsymbol{Z}}_i'\widetilde{\boldsymbol{u}}_i\right) \end{aligned} \tag{3.7}$$

where $\widehat{\widetilde{\boldsymbol{u}}}_i$ is a consistent estimate of $\widetilde{\boldsymbol{u}}_i$.

Under the identification assumption and regularity conditions, as $N \to \infty$, we have[8]

$$\widehat{\boldsymbol{\theta}}_{QD(1step)}, \widehat{\boldsymbol{\theta}}_{QD(2step)}, \widehat{\boldsymbol{\theta}}_{QD(CUE)} \xrightarrow{p} \boldsymbol{\theta}_{QD}^0$$

7 Note that the number of parameters can be very large for moderate values of m and T, which may be undesirable.

8 For the proofs, see Newey and McFadden (1994) and Hall (2005).

and

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\theta}}_{QD(1step)} - \boldsymbol{\theta}^0_{QD}) &\xrightarrow{d} \mathcal{N}(\mathbf{0},\, \boldsymbol{V}_{QD(1step)}),\\ \sqrt{N}(\widehat{\boldsymbol{\theta}}_{QD(2step)} - \boldsymbol{\theta}^0_{QD}) &\xrightarrow{d} \mathcal{N}(\mathbf{0},\, \boldsymbol{V}_{QD(2step)}),\\ \sqrt{N}(\widehat{\boldsymbol{\theta}}_{QD(CUE)} - \boldsymbol{\theta}^0_{QD}) &\xrightarrow{d} \mathcal{N}(\mathbf{0},\, \boldsymbol{V}_{QD(CUE)})\end{aligned}$$

where $\boldsymbol{\theta}^0_{QD}$ is the true value of $\boldsymbol{\theta}_{QD}$,

$$\begin{aligned}\boldsymbol{V}_{QD(1step)} &= (\boldsymbol{G}'_{QD}\boldsymbol{W}^{-1}_{QD}\boldsymbol{G}_{QD})^{-1}\boldsymbol{G}'_{QD}\boldsymbol{W}^{-1}_{QD}\boldsymbol{\Omega}_{QD}\boldsymbol{W}^{-1}_{QD}\boldsymbol{G}_{QD}\\ &\quad\times(\boldsymbol{G}'_{QD}\boldsymbol{W}^{-1}_{QD}\boldsymbol{G}_{QD})^{-1},\\ \boldsymbol{V}_{QD(2step)} &= \boldsymbol{V}_{QD(CUE)} = (\boldsymbol{G}'_{QD}\boldsymbol{\Omega}^{-1}_{QD}\boldsymbol{G}_{QD})^{-1},\\ \widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}_{QD}) &= \widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{u}}_i, \qquad \boldsymbol{\Omega}_{QD} = E[\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}^0_{QD})\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}^0_{QD})'],\\ \boldsymbol{G}_{QD} &= E\left(\left.\frac{\partial\widetilde{\boldsymbol{g}}_i(\boldsymbol{\theta}_{QD})}{\partial\boldsymbol{\theta}'_{QD}}\right|_{\boldsymbol{\theta}_{QD}=\boldsymbol{\theta}^0_{QD}}\right), \qquad \boldsymbol{W}_{QD} = E(\widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{Z}}_i).\end{aligned}$$

Since the model (3.2) is nonlinear in parameters, we need to use a numerical optimization procedure in practice. However, it is possible to estimate $\boldsymbol{\delta}$ by a linear estimator. If we let $\widetilde{\boldsymbol{X}}_{1i} = \dot{\boldsymbol{W}}_i$, $\widetilde{\boldsymbol{X}}_{2i} = (\boldsymbol{I}_{T-m}\otimes\ddot{\boldsymbol{y}}'_i)$, $\widetilde{\boldsymbol{X}}_{3i} = -(\mathrm{vec}(\ddot{\boldsymbol{W}}_i)'\otimes\boldsymbol{I}_{T-m})$, $\boldsymbol{\pi}_1 = \boldsymbol{\delta}$, $\boldsymbol{\pi}_2 = \boldsymbol{\psi}$, $\boldsymbol{\pi}_3 = \mathrm{vec}(\boldsymbol{\delta}'\otimes\boldsymbol{\Psi})$, $\widetilde{\boldsymbol{X}}_i = (\widetilde{\boldsymbol{X}}_{1i}, \widetilde{\boldsymbol{X}}_{2i}, \widetilde{\boldsymbol{X}}_{3i})$, $\boldsymbol{\pi}_{QD} = (\boldsymbol{\pi}'_1, \boldsymbol{\pi}'_2, \boldsymbol{\pi}'_3)'$, the model (3.2) can be written as

$$\dot{\boldsymbol{y}}_i = \widetilde{\boldsymbol{X}}_i\boldsymbol{\pi}_{QD} + \widetilde{\boldsymbol{u}}_i.$$

Since $E[\widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{u}}_i] = \mathbf{0}$ still holds for this model, we have a linear estimator:

$$\begin{aligned}\widehat{\boldsymbol{\pi}}_{QD(linear)} &= \left[\left(\sum_{i=1}^N\widetilde{\boldsymbol{X}}'_i\widetilde{\boldsymbol{Z}}_i\right)\left(\sum_{i=1}^N\widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{Z}}_i\right)^{-1}\left(\sum_{i=1}^N\widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{X}}_i\right)\right]^{-1}\\ &\quad\times\left[\left(\sum_{i=1}^N\widetilde{\boldsymbol{X}}'_i\widetilde{\boldsymbol{Z}}_i\right)\left(\sum_{i=1}^N\widetilde{\boldsymbol{Z}}'_i\widetilde{\boldsymbol{Z}}_i\right)^{-1}\left(\sum_{i=1}^N\widetilde{\boldsymbol{Z}}'_i\dot{\boldsymbol{y}}_i\right)\right].\end{aligned} \tag{3.8}$$

To the best of author's knowledge, this linear estimator has not been proposed in the literature. Also note that since the number of parameters is $K+1+m(T-m)+m(K+1)(T-m)$, we require $(T-m)(T-m+1)/2+KT(T-m) \geq K+1+m(T-m)+m(K+1)(T-m)$ for identification. This condition is different from the one for the nonlinear estimators.

It is easy to show that

$$\begin{aligned}&\widehat{\boldsymbol{\pi}}_{QD(linear)} \xrightarrow{p} \boldsymbol{\pi}^0_{QD},\\ &\sqrt{N}(\widehat{\boldsymbol{\pi}}_{QD(linear)} - \boldsymbol{\pi}^0_{QD}) \xrightarrow{d} \mathcal{N}(\mathbf{0},\, \boldsymbol{V}_{QD(linear)})\end{aligned}$$

where $\boldsymbol{\pi}_{QD}^0$ is the true value of $\boldsymbol{\pi}_{QD}$,

$$\boldsymbol{V}_{QD(linear)} = (\boldsymbol{Q}_{QD}' \boldsymbol{W}_{QD}^{-1} \boldsymbol{Q}_{QD})^{-1} \boldsymbol{Q}_{QD}' \boldsymbol{W}^{-1} \boldsymbol{\Omega}_{QD} \boldsymbol{W}_{QD}^{-1} \boldsymbol{Q}_{QD} (\boldsymbol{Q}_{QD}' \boldsymbol{W}_{QD}^{-1} \boldsymbol{Q}_{QD})^{-1}$$

and $\boldsymbol{Q}_{QD} = E(\widetilde{\boldsymbol{Z}}_i' \widetilde{\boldsymbol{X}}_i)$.

Since $\widehat{\boldsymbol{\pi}}_{QD(linear)}$ is consistent, its first $K+1$ elements are a consistent estimate of $\boldsymbol{\delta}$. While it is simple to compute, this estimator is considered to be less efficient than the above nonlinear GMM estimators since parameter restriction is not fully used. The difference between the nonlinear GMM estimators and linear estimator is investigated by Monte Carlo simulation in Section 4.

3.2. Projection approach

We now propose new GMM estimators. As shown above, Ahn *et al.* (2010) removed the interactive fixed effects term by quasi-difference, and that term does not appear in the resulting model (3.3). Here, instead of removing the interactive fixed effects term, we try to remove the correlation between the regressors and interactive fixed effects term by using a projection method by Mundlak (1978) and Chamberlain (1982, 1984), and then construct moment conditions that are required for GMM estimation[9,10]. For this, let us assume the following linear projection of $\boldsymbol{\lambda}_i$ onto constant, y_{i0} and $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}$:

$$\boldsymbol{\lambda}_i = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 y_{i0} + \sum_{s=1}^{T} \boldsymbol{\Phi}_{2s} \boldsymbol{x}_{is} + \widetilde{\boldsymbol{\lambda}}_i = \boldsymbol{\Phi} \boldsymbol{z}_i + \widetilde{\boldsymbol{\lambda}}_i \tag{3.9}$$

where $\boldsymbol{\Phi} = (\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \boldsymbol{\Phi}_{21}, \ldots, \boldsymbol{\Phi}_{2T})$ and $\boldsymbol{z}_i = (1, y_{i0}, \boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT}')'$. Projection onto initial conditions y_{i0} is used by Blundell and Bond (1998) or Semykina and Wooldridge (2012)[11]. Also, note that $E(y_{i0}\widetilde{\boldsymbol{\lambda}}_i) = \boldsymbol{0}$ and $E(\boldsymbol{x}_{is}\widetilde{\boldsymbol{\lambda}}_i') = \boldsymbol{0}$ for $s = 1, \ldots, T$ by definition. Substituting (3.9) into (2.2), we have

$$y_{it} = \boldsymbol{\delta}' \boldsymbol{w}_{it} + \boldsymbol{\gamma}_t' \boldsymbol{\Phi} \boldsymbol{z}_i + \boldsymbol{\gamma}_t' \widetilde{\boldsymbol{\lambda}}_i + v_{it}.$$

This is the model to be estimated. Compared with the model (3.3), we find that the interactive term $\boldsymbol{\gamma}_t'\widetilde{\boldsymbol{\lambda}}_i$ is still present. However, it is no longer correlated with regressors $\boldsymbol{x}_{is}$, $(s = 1, \ldots, T)$ and initial conditions y_{i0}. Hence, we can use this for constructing the moment conditions. Specifically, since y_{i0} and $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}$ are uncorrelated with $\boldsymbol{\gamma}_t'\widetilde{\boldsymbol{\lambda}}_i + v_{it}$ and correlated with $\boldsymbol{w}_{it}$ and $\boldsymbol{z}_i$, we have the following moment conditions:

$$E[\boldsymbol{z}_i u_{it}] = \boldsymbol{0}, \qquad (t = 1, \ldots, T) \tag{3.10}$$

where $u_{it} = \boldsymbol{\gamma}_t'\widetilde{\boldsymbol{\lambda}}_i + v_{it}$. Note that since the number of parameters is $K + 1 + Tm + m(KT + 2)$ and that of moment conditions are $T(KT + 2)$, we require $T(KT + 2) \geq K + 1 + Tm + m(KT + 2)$ for identification[12].

9 A similar approach is used by Bai (2009a) where the maximum likelihood estimator is proposed.

10 The quasi-difference approach can be seen as a fixed effects approach while the projection approach can be seen as a random effects approach.

11 It is also possible to project onto $y_{i0}, \ldots, y_{i,T-1}$ and $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}$ as in Bond and Windmeijer (2002).

12 Note that the number of parameters can be very large for moderate values of m, K and T, which may be undesirable.

In a matrix form, we have

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\delta} + \boldsymbol{\Gamma\Phi z}_i + \boldsymbol{\Gamma}\widetilde{\boldsymbol{\lambda}}_i + \boldsymbol{v}_i = \boldsymbol{W}_i\boldsymbol{\delta} + (\boldsymbol{z}_i' \otimes \boldsymbol{I}_T)\operatorname{vec}(\boldsymbol{\Gamma\Phi}) + \boldsymbol{u}_i$$

where $\boldsymbol{u}_i = \boldsymbol{\Gamma}\widetilde{\boldsymbol{\lambda}}_i + \boldsymbol{v}_i = (u_{i1}, \ldots, u_{iT})'$. The moment conditions can be written as

$$E[\boldsymbol{Z}_i'\boldsymbol{u}_i] = \boldsymbol{0} \tag{3.11}$$

where $\boldsymbol{Z}_i = \boldsymbol{I}_T \otimes \boldsymbol{z}_i'$. Note that $\boldsymbol{u}_i$ depends on unknown parameters $\boldsymbol{\theta}_{Pro} = (\boldsymbol{\delta}', \operatorname{vec}(\boldsymbol{\Gamma})', \operatorname{vec}(\boldsymbol{\Phi})')'$.

The GMM estimators based on moment conditions (3.11) are given as follows:

$$\widehat{\boldsymbol{\theta}}_{Pro(1step)} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{Pro(1step)} \\ \operatorname{vec}(\widehat{\boldsymbol{\Gamma}}_{Pro(1step)}) \\ \operatorname{vec}(\widehat{\boldsymbol{\Phi}}_{Pro(1step)}) \end{pmatrix} = \underset{\delta,\Gamma,\Phi}{\operatorname{argmin}} \left(\sum_{i=1}^N \boldsymbol{u}_i'\boldsymbol{Z}_i\right)\left(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\left(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{u}_i\right), \tag{3.12}$$

$$\widehat{\boldsymbol{\theta}}_{Pro(2step)} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{Pro(2step)} \\ \operatorname{vec}(\widehat{\boldsymbol{\Gamma}}_{Pro(2step)}) \\ \operatorname{vec}(\widehat{\boldsymbol{\Phi}}_{Pro(2step)}) \end{pmatrix} = \underset{\delta,\Gamma,\Phi}{\operatorname{argmin}} \left(\sum_{i=1}^N \boldsymbol{u}_i'\boldsymbol{Z}_i\right)\left(\sum_{i=1}^N \boldsymbol{Z}_i'\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i'\boldsymbol{Z}_i\right)^{-1}\left(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{u}_i\right), \tag{3.13}$$

$$\widehat{\boldsymbol{\theta}}_{Pro(CUE)} = \begin{pmatrix} \widehat{\boldsymbol{\delta}}_{Pro(CUE)} \\ \operatorname{vec}(\widehat{\boldsymbol{\Gamma}}_{Pro(CUE)}) \\ \operatorname{vec}(\widehat{\boldsymbol{\Phi}}_{Pro(CUE)}) \end{pmatrix} = \underset{\delta,\Gamma,\Phi}{\operatorname{argmin}} \left(\sum_{i=1}^N \boldsymbol{u}_i'\boldsymbol{Z}_i\right)\left(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{Z}_i\right)^{-1}\left(\sum_{i=1}^N \boldsymbol{Z}_i'\boldsymbol{u}_i\right) \tag{3.14}$$

where $\widehat{\boldsymbol{u}}_i$ is a consistent estimate of $\boldsymbol{u}_i$.

Under the identification assumption and standard regularity conditions, as $N \to \infty$, we have[13]

$$\widehat{\boldsymbol{\theta}}_{Pro(1step)}, \widehat{\boldsymbol{\theta}}_{Pro(2step)}, \widehat{\boldsymbol{\theta}}_{Pro(CUE)} \xrightarrow{p} \boldsymbol{\theta}_{Pro}^0$$

and

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{Pro(1step)} - \boldsymbol{\theta}_{Pro}^0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{V}_{Pro(1step)}),$$
$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{Pro(2step)} - \boldsymbol{\theta}_{Pro}^0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{V}_{Pro(2step)}),$$
$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{Pro(CUE)} - \boldsymbol{\theta}_{Pro}^0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\, \boldsymbol{V}_{Pro(CUE)})$$

13 For the proofs, see Newey and McFadden (1994) and Hall (2005).

where $\boldsymbol{\theta}^0_{Pro}$ is the true value of $\boldsymbol{\theta}_{Pro}$,

$$\begin{aligned}
\boldsymbol{V}_{Pro(1step)} &= (\boldsymbol{G}'_{Pro}\boldsymbol{W}^{-1}_{Pro}\boldsymbol{G}_{Pro})^{-1}\boldsymbol{G}'_{Pro}\boldsymbol{W}^{-1}_{Pro}\boldsymbol{\Omega}_{Pro}\boldsymbol{W}^{-1}_{Pro}\boldsymbol{G}_{Pro} \\
&\quad \times (\boldsymbol{G}'_{Pro}\boldsymbol{W}^{-1}_{Pro}\boldsymbol{G}_{Pro})^{-1}, \\
\boldsymbol{V}_{Pro(2step)} &= \boldsymbol{V}_{Pro(CUE)} = (\boldsymbol{G}'_{Pro}\boldsymbol{\Omega}^{-1}_{Pro}\boldsymbol{G}_{Pro})^{-1}, \\
\boldsymbol{g}_i(\boldsymbol{\theta}_{Pro}) &= \boldsymbol{Z}'_i\boldsymbol{u}_i, \qquad \boldsymbol{\Omega}_{Pro} = E[\boldsymbol{g}_i(\boldsymbol{\theta}^0_{Pro})\boldsymbol{g}_i(\boldsymbol{\theta}^0_{Pro})'], \\
\boldsymbol{G}_{Pro} &= E\left(\left.\frac{\partial \boldsymbol{g}_i(\boldsymbol{\theta}_{Pro})}{\partial \boldsymbol{\theta}'_{Pro}}\right|_{\boldsymbol{\theta}_{Pro}=\boldsymbol{\theta}^0_{Pro}}\right), \qquad \boldsymbol{W}_{Pro} = E(\boldsymbol{Z}'_i\boldsymbol{Z}_i).
\end{aligned}$$

4. Monte Carlo simulation

In this section, we conduct Monte Carlo simulations to investigate the finite sample properties of the nonlinear GMM and linear estimators introduced in the previous section.

4.1. Design

We consider the following data generating process:

$$\begin{aligned}
y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \boldsymbol{\lambda}'_i\boldsymbol{\gamma}_t + v_{it}, \qquad (i = 1,\dots,N; t = 1,\dots,T) \\
x_{it} &= \boldsymbol{\lambda}'_i\boldsymbol{\gamma}_t + \boldsymbol{\lambda}'_i\boldsymbol{\iota}_m + \boldsymbol{\gamma}'_t\boldsymbol{\iota}_m + w_{it}, \\
w_{it} &= \rho w_{i,t-1} + \varepsilon_{it}
\end{aligned}$$

where $v_{it}, \varepsilon_{it}, \lambda_{1i}, \dots, \lambda_{mi} \sim$ i.i.d.$\mathcal{N}(0,1)$. The first 50 periods are discarded to reduce the effect of initial conditions. Observations used in estimation are y_{it} and x_{it} for $i = 1,\dots,N$; $t = 0,\dots,T$[14]. For $\boldsymbol{\gamma}_t$, we consider the following three cases:

$$\begin{aligned}
&\text{Case 1}: \quad \gamma_{1t} = 1 - 0.6t + 0.1t^2, \qquad (t = 0,1,\dots,T), \\
&\text{Case 2}: \quad \gamma_{2t} = 1 - \sqrt{0.25t/T}, \qquad (t = 0,1,\dots,T), \\
&\text{Case 3}: \quad \boldsymbol{\gamma}_t = (\gamma_{1t}, \gamma_{2t})
\end{aligned}$$

with $\gamma_{jt} = 1$, $(j = 1,2)$ for $t = -50, -49, \dots, -1$. Note that $m = 1$ for the Cases 1 and 2 and $m = 2$ for Case 3. For the sample sizes and parameter values, we set $T = 5$, $N = 100$, 250, 500, $\alpha = 0.4$, 0.8, $\beta = 1$, and $\rho = 0.5$. The number of replication is 1000.

4.2. Results

Simulation results are summarized in Tables 1 to 3. In Table 1, we report the results for the case of γ_{1t}. We first focus on α. In terms of bias, we find that the projection estimators have smaller biases than the QD estimators in almost all cases. In terms of dispersion, it is observed that the QD estimators have smaller standard deviations when $N = 100$, while the projection estimators are less dispersed when $N = 500$. When $N = 250$, the results are mixed: the

14 x_{i0} is used as instruments as well as x_{it}, $(t = 1,\dots,T)$.

Table 1. Simulation results: Case 1.

Design	Estimator	α			β		
		Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
$T=5$	QD(linear)	0.420	0.031	0.037	0.992	0.066	0.067
$N=100$	QD(1step)	0.429	0.034	0.044	1.006	0.054	0.055
$m=1$	QD(2step)	0.428	0.037	0.046	1.006	0.057	0.058
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.409	0.048	0.049	1.010	0.070	0.070
$\alpha=0.4$	Pro(1step)	0.400	0.046	0.046	1.005	0.054	0.055
$\beta=1.0$	Pro(2step)	0.400	0.048	0.048	1.003	0.057	0.057
	Pro(CUE)	0.406	0.052	0.052	1.003	0.066	0.066
$T=5$	QD(linear)	0.418	0.025	0.031	0.999	0.041	0.041
$N=250$	QD(1step)	0.419	0.028	0.033	1.003	0.033	0.033
$m=1$	QD(2step)	0.417	0.028	0.033	1.003	0.033	0.033
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.405	0.030	0.031	1.002	0.036	0.036
$\alpha=0.4$	Pro(1step)	0.398	0.026	0.026	1.000	0.032	0.032
$\beta=1.0$	Pro(2step)	0.399	0.027	0.027	1.000	0.033	0.033
	Pro(CUE)	0.401	0.027	0.027	0.999	0.034	0.034
$T=5$	QD(linear)	0.413	0.020	0.024	0.999	0.030	0.030
$N=500$	QD(1step)	0.410	0.020	0.022	1.001	0.024	0.024
$m=1$	QD(2step)	0.409	0.020	0.022	1.001	0.023	0.023
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.402	0.020	0.020	1.000	0.023	0.023
$\alpha=0.4$	Pro(1step)	0.399	0.018	0.018	0.999	0.022	0.022
$\beta=1.0$	Pro(2step)	0.399	0.018	0.018	1.000	0.022	0.022
	Pro(CUE)	0.400	0.018	0.018	0.999	0.022	0.022
$T=5$	QD(linear)	0.798	0.025	0.025	0.978	0.068	0.071
$N=100$	QD(1step)	0.811	0.013	0.017	1.004	0.052	0.052
$m=1$	QD(2step)	0.811	0.014	0.018	1.003	0.056	0.056
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.800	0.026	0.026	1.007	0.064	0.064
$\alpha=0.8$	Pro(1step)	0.793	0.028	0.028	1.001	0.052	0.052
$\beta=1.0$	Pro(2step)	0.793	0.028	0.029	1.000	0.056	0.056
	Pro(CUE)	0.797	0.026	0.026	1.000	0.065	0.065
$T=5$	QD(linear)	0.805	0.018	0.018	0.989	0.045	0.046
$N=250$	QD(1step)	0.809	0.012	0.015	1.001	0.031	0.031
$m=1$	QD(2step)	0.808	0.013	0.015	1.001	0.032	0.032
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.800	0.020	0.020	1.000	0.035	0.035
$\alpha=0.8$	Pro(1step)	0.796	0.016	0.016	0.999	0.031	0.031
$\beta=1.0$	Pro(2step)	0.797	0.016	0.016	0.998	0.033	0.033
	Pro(CUE)	0.798	0.015	0.015	0.998	0.034	0.034
$T=5$	QD(linear)	0.806	0.013	0.015	0.991	0.030	0.032
$N=500$	QD(1step)	0.808	0.010	0.013	1.001	0.022	0.022
$m=1$	QD(2step)	0.807	0.011	0.013	1.001	0.021	0.021
$\boldsymbol{\gamma}=\gamma_{1t}$	QD(CUE)	0.801	0.015	0.015	0.999	0.023	0.023
$\alpha=0.8$	Pro(1step)	0.798	0.010	0.010	0.998	0.021	0.021
$\beta=1.0$	Pro(2step)	0.798	0.010	0.010	0.999	0.022	0.022
	Pro(CUE)	0.799	0.010	0.010	0.999	0.022	0.022

Note: "QD(linear)", "QD(1step)", "QD(2step)" and "QD(CUE)" denote GMM estimators defined as (3.8), (3.5), (3.6) and (3.7), respectively. "Pro(1step)", "Pro(2step)" and "Pro(CUE)" denote GMM estimators defined as (3.12), (3.13) and (3.14), respectively.

Table 2. Simulation results: Case 2.

Design	Estimator	α Mean	α Std. Dev.	α RMSE	β Mean	β Std. Dev.	β RMSE
$T = 5$	QD(linear)	0.456	0.039	0.068	0.973	0.068	0.073
$N = 100$	QD(1step)	0.436	0.078	0.086	1.000	0.066	0.066
$m = 1$	QD(2step)	0.438	0.075	0.084	1.005	0.063	0.064
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.419	0.068	0.071	1.004	0.071	0.071
$\alpha = 0.4$	Pro(1step)	0.395	0.053	0.053	1.003	0.055	0.055
$\beta = 1.0$	Pro(2step)	0.395	0.055	0.055	1.003	0.060	0.060
	Pro(CUE)	0.411	0.064	0.065	1.003	0.068	0.068
$T = 5$	QD(linear)	0.449	0.033	0.059	0.984	0.045	0.048
$N = 250$	QD(1step)	0.411	0.052	0.053	0.998	0.044	0.044
$m = 1$	QD(2step)	0.413	0.049	0.051	1.003	0.038	0.038
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.402	0.033	0.033	1.001	0.037	0.037
$\alpha = 0.4$	Pro(1step)	0.395	0.028	0.029	1.000	0.035	0.035
$\beta = 1.0$	Pro(2step)	0.395	0.030	0.031	1.001	0.036	0.036
	Pro(CUE)	0.401	0.032	0.032	1.000	0.036	0.036
$T = 5$	QD(linear)	0.439	0.028	0.048	0.988	0.034	0.036
$N = 500$	QD(1step)	0.401	0.030	0.030	0.998	0.033	0.033
$m = 1$	QD(2step)	0.401	0.025	0.025	0.999	0.024	0.024
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.400	0.021	0.021	1.000	0.024	0.024
$\alpha = 0.4$	Pro(1step)	0.398	0.020	0.020	0.999	0.023	0.023
$\beta = 1.0$	Pro(2step)	0.398	0.021	0.021	1.000	0.024	0.024
	Pro(CUE)	0.401	0.021	0.021	1.000	0.024	0.024
$T = 5$	QD(linear)	0.812	0.034	0.036	0.953	0.070	0.085
$N = 100$	QD(1step)	0.797	0.050	0.050	0.986	0.065	0.067
$m = 1$	QD(2step)	0.799	0.049	0.049	0.990	0.061	0.061
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.805	0.046	0.046	0.996	0.068	0.068
$\alpha = 0.8$	Pro(1step)	0.789	0.038	0.039	0.999	0.054	0.054
$\beta = 1.0$	Pro(2step)	0.789	0.039	0.041	0.999	0.059	0.059
	Pro(CUE)	0.799	0.042	0.042	0.997	0.067	0.067
$T = 5$	QD(linear)	0.820	0.025	0.032	0.968	0.047	0.057
$N = 250$	QD(1step)	0.796	0.033	0.033	0.992	0.042	0.043
$m = 1$	QD(2step)	0.800	0.031	0.031	0.994	0.037	0.037
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.802	0.026	0.026	1.000	0.037	0.037
$\alpha = 0.8$	Pro(1step)	0.795	0.024	0.024	0.998	0.035	0.035
$\beta = 1.0$	Pro(2step)	0.795	0.024	0.025	0.999	0.036	0.036
	Pro(CUE)	0.800	0.024	0.024	1.000	0.037	0.037
$T = 5$	QD(linear)	0.820	0.020	0.028	0.973	0.034	0.043
$N = 500$	QD(1step)	0.796	0.022	0.023	0.996	0.033	0.033
$m = 1$	QD(2step)	0.798	0.021	0.021	0.997	0.024	0.025
$\boldsymbol{\gamma} = \gamma_{2t}$	QD(CUE)	0.801	0.018	0.018	1.001	0.024	0.024
$\alpha = 0.8$	Pro(1step)	0.798	0.016	0.017	0.999	0.023	0.023
$\beta = 1.0$	Pro(2step)	0.798	0.017	0.017	1.000	0.024	0.024
	Pro(CUE)	0.801	0.017	0.017	1.000	0.024	0.024

Note: "QD(linear)", "QD(1step)", "QD(2step)" and "QD(CUE)" denote GMM estimators defined as (3.8), (3.5), (3.6) and (3.7), respectively. "Pro(1step)", "Pro(2step)" and "Pro(CUE)" denote GMM estimators defined as (3.12), (3.13) and (3.14), respectively.

Table 3. Simulation results: Case 3.

		α			β		
Design	Estimator	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
$T = 5$	QD(linear)	0.403	0.083	0.083	0.954	0.159	0.166
$N = 100$	QD(1step)	0.414	0.055	0.057	1.011	0.076	0.077
$m = 2$	QD(2step)	0.419	0.054	0.057	1.013	0.077	0.078
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.416	0.098	0.099	1.000	0.170	0.170
$\alpha = 0.4$	Pro(1step)	0.391	0.050	0.051	0.998	0.067	0.067
$\beta = 1.0$	Pro(2step)	0.390	0.053	0.053	0.997	0.068	0.068
	Pro(CUE)	0.401	0.055	0.055	0.998	0.072	0.072
$T = 5$	QD(linear)	0.406	0.048	0.049	0.973	0.082	0.087
$N = 250$	QD(1step)	0.406	0.037	0.037	1.009	0.047	0.047
$m = 2$	QD(2step)	0.417	0.036	0.040	1.013	0.046	0.048
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.415	0.049	0.051	1.011	0.082	0.082
$\alpha = 0.4$	Pro(1step)	0.395	0.030	0.030	0.998	0.038	0.038
$\beta = 1.0$	Pro(2step)	0.396	0.031	0.031	0.999	0.039	0.039
	Pro(CUE)	0.400	0.031	0.031	0.999	0.040	0.040
$T = 5$	QD(linear)	0.403	0.037	0.038	0.987	0.057	0.058
$N = 500$	QD(1step)	0.398	0.024	0.024	1.009	0.034	0.035
$m = 2$	QD(2step)	0.408	0.024	0.026	1.011	0.032	0.034
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.410	0.032	0.033	1.012	0.052	0.053
$\alpha = 0.4$	Pro(1step)	0.397	0.021	0.021	0.999	0.028	0.028
$\beta = 1.0$	Pro(2step)	0.397	0.022	0.022	1.000	0.029	0.029
	Pro(CUE)	0.400	0.022	0.022	0.999	0.029	0.029
$T = 5$	QD(linear)	0.754	0.101	0.111	0.960	0.172	0.177
$N = 100$	QD(1step)	0.791	0.037	0.038	1.002	0.071	0.071
$m = 2$	QD(2step)	0.791	0.040	0.041	1.002	0.070	0.070
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.806	0.100	0.100	1.016	0.130	0.131
$\alpha = 0.8$	Pro(1step)	0.792	0.032	0.033	0.992	0.060	0.061
$\beta = 1.0$	Pro(2step)	0.792	0.033	0.034	0.993	0.060	0.061
	Pro(CUE)	0.799	0.034	0.034	0.995	0.064	0.064
$T = 5$	QD(linear)	0.773	0.070	0.075	0.991	0.105	0.105
$N = 250$	QD(1step)	0.792	0.025	0.027	1.004	0.043	0.043
$m = 2$	QD(2step)	0.795	0.026	0.026	1.003	0.039	0.039
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.805	0.043	0.043	1.009	0.068	0.069
$\alpha = 0.8$	Pro(1step)	0.797	0.022	0.022	0.996	0.036	0.036
$\beta = 1.0$	Pro(2step)	0.797	0.021	0.021	0.996	0.036	0.036
	Pro(CUE)	0.801	0.021	0.021	0.996	0.037	0.037
$T = 5$	QD(linear)	0.783	0.052	0.055	1.002	0.081	0.081
$N = 500$	QD(1step)	0.795	0.018	0.019	1.009	0.032	0.033
$m = 2$	QD(2step)	0.797	0.018	0.018	1.006	0.028	0.029
$\boldsymbol{\gamma} = (\gamma_{1t}, \gamma_{2t})$	QD(CUE)	0.805	0.034	0.035	1.006	0.052	0.053
$\alpha = 0.8$	Pro(1step)	0.798	0.015	0.015	0.998	0.027	0.027
$\beta = 1.0$	Pro(2step)	0.798	0.015	0.015	0.998	0.028	0.028
	Pro(CUE)	0.799	0.015	0.015	0.998	0.028	0.028

Note: "QD(linear)", "QD(1step)", "QD(2step)" and "QD(CUE)" denote GMM estimators defined as (3.8), (3.5), (3.6) and (3.7), respectively. "Pro(1step)", "Pro(2step)" and "Pro(CUE)" denote GMM estimators defined as (3.12), (3.13) and (3.14), respectively.

superiority depends on α. When $\alpha = 0.4$ and $N = 250$, the projection estimators have smaller dispersion, while the result reverses when $\alpha = 0.8$. Also, in some cases, one-step estimators have the smaller standard deviations than the 2step and CU-GMM estimators, which is surprising since asymptotic theory predicts that 2step and CU-GMM estimators are more efficient than the one-step GMM estimators. It seems that this is a finite sample issue. In Section 3, we noted that the QD(linear) estimator is simple to compute but it may not be efficient. From the simulation results, we find that whether the QD(linear) estimator is less efficient or not depends on the situation. For instance, when $\alpha = 0.4$, we find that the QD(linear) estimator has smaller standard deviations than other nonlinear QD estimators. However, when $\alpha = 0.8$, the standard deviation of QD(linear) is larger than those of other QD estimators. In terms of RMSE, the superiority between the QD and projection estimators depends on α and N. For the estimation of β, the overall performance is reasonably good.

We now turn to consider Table 2. Compared with Table 1, we find that the biases of α are larger. In terms of bias, the projection estimators perform better than the QD estimators when $\alpha = 0.4$ and $N = 100$. For the dispersion, the projection estimators have smaller standard deviations than the QD estimators in almost all cases. This result is also reflected in RMSE. In terms of RMSE, the projection estimators perform better than the QD estimators in almost all cases.

Finally, we investigate the results of the two-factor case, which are given in Table 3. From the results, we find that the projection estimators are less biased than the QD estimators in many cases. In terms of standard deviations, the projection estimators have smaller dispersions than the QD estimators in almost all cases for α and β. In terms of RMSE, the projection estimators tend to perform better than the QD estimators in almost all cases.

Summarising the overall results, we find that the projection estimators are less biased than the QD estimators in almost all cases, and in term of RMSE, the projection estimators perform better than the QD estimators in many cases with a few exceptions.

5. Conclusion

In this paper, we proposed new GMM estimators for short dynamic panel data models with interactive fixed effects. We used the projection method of Chamberlain (1982, 1984) to remove the correlation between regressors and the interactive fixed effects term and constructed moment conditions required for GMM estimation. Monte Carlo simulation revealed that our new GMM estimator performs reasonably well in finite samples and has better or comparable performance with the GMM estimator by Ahn *et al.* (2010).

Finally, we mention one remaining issue of this paper. The important problem not addressed in this paper is the estimation of the number of factors. In large N and large T panels, several papers such as Bai and Ng (2002), Onatski (2009, 2010), Kapetanios (2010) propose methods to estimate the number of factors. However, for large N and small T panels, to the best of author's knowledge, the only approach available now is that of Ahn *et al.* (2010) where a sequential procedure based on overidentification restriction test is proposed. We expect that this procedure is applicable to our case although investigating the performance is left for a future topic.

Acknowledgements

The author is deeply grateful to two referees, Taku Yamamoto, Eiji Kurozumi, Hiroaki Chigira, Hosung Jung and Asuka Takeuchi for helpful comments. Part of this paper was written while the author was visiting the University of Cambridge as a JSPS Postdoctoral Fellow for Research Abroad. The author is thankful to Hashem Pesaran for his support. Financial support from the JSPS Fellowship and KAKENHI(22730178) is also gratefully acknowledged. All remaining errors are my own.

References

Ahn, S. C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data, *J. Econom.*, **68**, 5–27.

Ahn, S. C. and Schmidt, P. (1997). Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation, *J. Econom.*, **76**, 309–321.

Ahn, S. C., Lee, H. Y. and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects, *J. Econom.*, **101**, 219–255.

Ahn, S. C., Lee, Y. H. and Schmidt, P. (2010). Panel Data Models with Multiple Time-Varying Individual Effects, mimeo.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components, *J. Am. Stat. Assoc.*, **76**, 598–606.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data, *J. Econom.*, **18**, 47–82.

Arellano, M. (2003). *Panel Data Econometrics*, Oxford, Oxford University Press.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, **58**, 277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models, *J. Econom.*, **68**, 29–51.

Bai, J. (2009a). Likelihood approach to small T dynamic panel models with interactive effects, mimeo.

Bai, J. (2009b). Panel data models with interactive fixed effects, *Econometrica*, **77**, 1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica*, **70**, 191–221.

Balestra, P. and Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas, *Econometrica*, **34**, 585–612.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*, Chichester, John Wiley and Sons, 4th edition.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models, *J. Econom.*, **87**, 115–143.

Bond, S. and Windmeijer, F. (2002). Projection estimators for autoregressive panel data models, *Econometrics Journal*, **5**, 457–479.

Chamberlain, G. (1982). Multivariate regression models for panel data, *J. Econom.*, **18**, 5–46.

Chamberlain, G. (1984). Panel data, *Handbook of Econometrics*, 2, North-Holland, Chap. 22, 1248–1318.

Hall, A. (2005). *Generalized Method of Moments*, Oxford, Oxford University Press.

Holtz-Eakin, D., Newey, W. K. and Rosen, H. S. (1988). Estimating vector autoregressions with panel data, *Econometrica*, **56**, 1371–1395.

Hsiao, C. (2003). *Analysis of Panel Data*, Cambridge and New York, Cambridge University Press.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets, *J. Bus. Econ. Stat.*, **28**, 397–409.

Meghir, C. and Windmeijer, F. (1999). Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance, *Ann. Econ. Stat.*, **55–56**, 317–330.

Moon, H. R. and Weidner, M. (2010). Dynamic linear panel regression models with interactive fixed effects, mimeo.

Mundlak, Y. (1978). On the pooling of time series and cross section data, *Econometrica*, **46**, 69–85.

Nauges, C. and Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects, *Ann. Econ. Stat.*, **70**, 53–75.

Nerlove, M. (2002). *Essays in Panel Data Econometrics*, Cambridge, Cambridge University Press.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing, *Handbook of Econometrics* (eds. R. F. Engle and D. McFadden), 4, North-Holland, Chap. 36, 2113–2245.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects, *Econometrica*, **49**, 1417–1426.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models, *Econometrica*, **77**, 1447–1479.

Onatski, A. (2010). Determining the number of factors form empirical distribution of eigenvalues, *Rev. Econ. Stat.*, **92**, 1004–1016.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with cross section dependence, *Econometrica*, **74**, 967–1012.

Phillips, P. C. B. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence, *Econometrics Journal*, **6**, 217–259.

Phillips, P. C. B. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, *J. Econom.*, **127**, 162–188.

Sarafidis, V. and Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation, *Econometrics Journal*, **12**, 62–81.

Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis, *Econom. Rev.*, **31**, 483–531.

Sarafidis, V. and Yamagata, T. (2010). Instrumental variable estimation of dynamic linear panel data models with defactored regressors under cross-section dependence, mimeo.

Sarafidis, V., Yamagata, T. and Robertson, D. (2009). A test of cross section dependence for a linear dynamic panel model with regressors, *J. Econom.*, **148**, 149–161.

Semykina, A. and Wooldridge, J. M. (2012). Estimation of dynamic panel data models with sample selection, fothcoming in *J. Appl. Econ*.

Wansbeek, T. J. and Knaap, T. (1999). Estimating a dynamic panel data model with heterogeneous trends, *Ann. Econ. Stat.*, **55–56**, 331–349.