Sequential and efficient GMM estimation of dynamic short panel data models

Fei Jin[1], Lung-fei Lee[2], and Jihai Yu*[3]

[1]School of Economics, Fudan University, Shanghai 200433, China
[1]Shanghai Institute of International Finance and Economics, Shanghai 200433, China
[2]Department of Economics, The Ohio State University, Columbus, OH 43210, USA
[3]Guanghua School of Management, Peking University, Beijing 100871, China

March 18, 2020

Abstract

This paper considers generalized method of moments (GMM) and sequential GMM (SGMM) estimation of dynamic short panel data models. The efficient GMM motivated from the quasi maximum likelihood (QML) can avoid the use of many instrument variables (IV) for estimation. It can be asymptotically efficient as maximum likelihood estimators (MLE) when disturbances are normal, and can be more efficient than QML estimators when disturbances are not normal. The SGMM, which also incorporates many IVs, generalizes the minimum distance estimation originated in Hsiao et al. (2002). By focusing on the estimation of parameters of interest, the SGMM saves computational burden caused by nuisance parameters such as variances of disturbances. It is asymptotically as efficient as the corresponding GMM. In particular, the SGMM based on QML scores can generate a closed-form root estimator for the dynamic parameter, which is asymptotically as efficient as the QML estimator. Nuisance parameters can also be estimated efficiently by an additional SGMM step if they are of interest.

Keywords: Dynamic panel data, GMM, sequential GMM, efficiency, root estimator
JEL classification: C13, C18, C23

1 Introduction

Dynamic panel data (DPD) models are popular in empirical studies, as they control for unobserved individual effects and allow for state dependence. Due to individual effects and dynamic feature, maximum likelihood (ML) and quasi-maximum likelihood (QML) estimations of fixed effects DPD can cause an incidental parameter problem (Nickell, 1981; Hsiao, 1986), and the magnitude of the bias is of the order $O(1/T)$ where T is the number of time periods. To avoid the incidental parameter problem, the estimation method of instrumental variables (IV) is popular (see Anderson and Hsiao, 1981; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998; Bun and Kiviet, 2006, etc). When observations of panels

*Corresponding author. E-mail address: jihai.yu@gmail.com.

over time periods are short, instead of the IV method, another approach is to specify the initial condition and apply the QML estimation as in Hsiao et al. (2002), where individual effects are eliminated by first differences. In the current paper, we focus on dynamic panels with short time periods, and aim to estimate those panel models by the generalized method of moments (GMM) and sequential GMM (SGMM), where the information of initial conditions is critical and can be utilized to improve the efficiency of estimates. The GMM and SGMM are also applied to dynamic panel data models with time-varying exogenous variables.

With properly designed moment conditions, GMM estimates can be asymptotically as efficient as maximum likelihood (ML) estimates under normal disturbances but might be relatively more efficient than QML estimates when model disturbances are not normal.[1] For some situations, a GMM estimation approach can be relatively computationally simpler than the ML or QML estimates. Furthermore, in the existing literature, while the model is estimated by IV methods, there could be an issue on using many IVs as those many IV estimates might have large asymptotic biases. Such a problem can be overcome by a properly designed GMM estimation with a finite number of moments. Best GMM moments may also be constructed under some circumstances.

For DPD models that have not achieved stationarity due to a finite starting period, the variance of the initial starting period of the dependent variable will be a free parameter. With an initial consistent estimate of such a variance parameter, a minimum distance (MD) estimation method has been considered in Hsiao et al. (2002). However, the asymptotic distribution of the MD estimator would be influenced by the asymptotic distribution of the initial estimate of that free variance parameter. To overcome the issue of such an asymptotic variance estimate, the GMM and SGMM estimation can be used. An SGMM approach proposed in Jin and Lee (2018) can be asymptotically as efficient as the GMM and can be computationally simpler. If we take the variance parameter of the initial period dependent variable as a nuisance parameter and use a simple initial consistent estimate, the SGMM can focus on efficient estimation of remaining parameters of interest and avoid some computational burden. The SGMM uses a $C(\alpha)$-type transformation of moment vectors to eliminate the asymptotic impact of initial consistent estimators and to achieve asymptotic efficiency. In particular, we show that, for an SGMM that is based on the QML first order conditions but only estimates the dynamic parameter, a closed form root estimator exists and is asymptotically as efficient as the QML estimator.

The current paper is organized as follows. Section 2 studies the fixed effects pure DPD model with a short past, and Section 3 studies the fixed effects DPD model with exogenous variables. For each model, a general GMM estimation framework is motivated from QML scores. Efficient GMM under the normality

[1] In the following, the ML or QML estimates for fixed effects DPD models all refer to those based on first differenced equations of the dependent variable.

assumption of disturbances can be derived. We investigate some computationally simple and efficient SGMM estimates based on the efficient GMM and QML scores. For the fixed effects pure DPD model, the SGMM improves upon the computationally simple MD estimator in Hsiao et al. (2002) by eliminating the asymptotic impact of an initial variance estimator but yet can achieve asymptotic efficiency. Section 4 studies stationary fixed effects DPD models.[2] Monte Carlo results for various estimators are provided in Section 5. Section 6 concludes the paper and summarizes the contributions. Proofs, detailed algebra and additional Monte Carlo experiments are provided in a supplementary file available upon request. GMM and SGMM estimations of random effects DPD models can be similarly studied and we provide them in the supplementary file.[3]

2 Fixed effects pure DPD with a short past

In this section, we first introduce the fixed effects pure DPD model with a short past and its MD estimation. We show that a best IV for the MD estimation exists under some conditions, but the best IV is infeasible. We propose the efficient GMM and SGMM based on QML scores to overcome such a problem.

Consider the pure dynamic panel data model

$$Y_{nt} = \gamma_0 Y_{n,t-1} + \mathbf{c}_{n0} + V_{nt}, \quad t = 1, 2, \dots, T, \tag{2.1}$$

where $Y_{nt} = (y_{1t}, \dots, y_{nt})'$ at time t represents the vector of outcomes of all the n individuals with y_{it} representing the outcome of individual i at time t; $\mathbf{c}_{n0} = (c_{10}, c_{20}, \dots, c_{n0})'$ is the n-dimensional vector of individual effects, and $V_{nt} = (v_{1t}, \dots, v_{nt})'$ is the vector of disturbances of all n individuals. In this model, the disturbances v_{it}'s are i.i.d. $(0, \sigma_{v0}^2)$ across all individuals. In a fixed effects model, all the individual effects c_i's are treated as unknown fixed parameters, while in a random effect model, they are treated as random elements. In this section, we consider the fixed effects specification. As c_i's are n unknown parameters in a sample with n individuals but a finite number of T time periods, c_i creates an incidental parameter problem (Nickell, 1981). Therefore, it is desirable to eliminate the fixed effects c_i's before estimation. It is natural to perform the elimination by taking time difference (e.g., Hsiao et al. 2002). By taking first (time) difference, $\Delta Y_{nt} = \gamma_0 \Delta Y_{n,t-1} + \Delta V_{nt}$ where Δ denotes first difference. The estimating equation will consist of

$$\Delta Y_{nt} = \gamma_0 \Delta Y_{n,t-1} + \Delta V_{nt}, \; t = 2, \cdots, T, \tag{2.2}$$

together with the observation ΔY_{n1}. The ΔY_{n1} may be treated as the first period sample for the difference

[2] We note that, "stationarity" here refers to the situation that the process has started a long time ago.

[3] For stationary random effects DPD models, the quasi log likelihood function can be decomposed as a sum of the quasi log likelihood function of within equations and that of between equations (Lee and Yu, 2018). We use the decomposition to derive simple moment vectors, which can yield GMM estimators that are asymptotically as efficient as ML estimators under normal disturbances, but can be more efficient relative to QML estimators.

process in (2.2). For ΔY_{n1}, by continuous substitution, we have, up to the past m time periods,

$$\Delta Y_{n1} = \gamma_0^m \Delta Y_{n,-m+1} + \sum_{j=0}^{m-1} \gamma_0^j \Delta V_{n,1-j}.$$

In the case that the process has started from finite m periods ago, where m is unknown, $\mathrm{E}(\Delta Y_{n1}) = \gamma_0^m \mathrm{E}(\Delta Y_{n,-m+1})$ and $\mathrm{Var}(\Delta Y_{n1}) = \omega_0 \sigma_{v0}^2 I_n$ for some $\omega_0 > 1$, both of which are unknown values. Following Hsiao et al. (2002), we have the moment properties that $\mathrm{E}(\Delta Y_{n1}) = \kappa_0 l_n$, $\mathrm{Var}(\Delta Y_{n1}) = \omega_0 \sigma_{v0}^2 I_n$, $\mathrm{Cov}(\Delta Y_{n1}, \Delta V_{n2}) = -\sigma_{v0}^2 I_n$, and $\mathrm{Cov}(\Delta Y_{n1}, \Delta V_{nt}) = 0$ for $t \geq 3$. Thus, the $nT \times nT$ variance matrix of the disturbances is $\sigma_{v0}^2 [H_T(\omega_0) \otimes I_n]$, where

$$H_T(\omega) = \begin{pmatrix} \omega & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}. \tag{2.3}$$

The quasi log likelihood function for (2.2) with sample observations ΔY_{nt}, $t = 1, ..., T$, as if $[\Delta Y_{n1}' - \kappa_0 l_n', \Delta V_{n2}', \ldots, \Delta V_{nT}']'$ were normally distributed, is

$$\ln L_w(\theta) = -\frac{nT}{2} \ln(2\pi\sigma_v^2) - \frac{n}{2} \ln |H_T(\omega)| - \frac{1}{2\sigma_v^2} e_{nT}'(\theta_1)(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\theta_1), \tag{2.4}$$

where $\theta_1 = (\kappa, \gamma)'$, $\theta = (\kappa, \gamma, \omega, \sigma_v^2)'$, and $e_{nT}(\theta_1) = [\Delta Y_{n1}' - \kappa l_n', \Delta Y_{n2}' - \gamma \Delta Y_{n1}', \ldots, \Delta Y_{nT}' - \gamma \Delta Y_{n,T-1}']'$ with l_n being the n-dimensional vector consisting of all unit entries. This quasi log likelihood function has explored not only the main regression equations for ΔY_{nt} but also their variances. Hsiao et al. (2002) has considered the properties of a QML estimator that maximizes (2.4). The first order derivatives are

$$\frac{\partial \ln L_w(\theta)}{\partial \theta_1} = \frac{1}{\sigma_v^2} \Delta \mathbf{Z}_{n,T-1}' (H_T^{-1}(\omega) \otimes I_n) e_{nT}(\theta_1), \tag{2.5a}$$

$$\frac{\partial \ln L_w(\theta)}{\partial \omega} = -\frac{n}{2} \operatorname{tr}(H_T^{-1}(\omega) J_T) + \frac{1}{2\sigma_v^2} e_{nT}'(\theta_1) \left(H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \otimes I_n \right) e_{nT}(\theta_1), \tag{2.5b}$$

$$\frac{\partial \ln L_w(\theta)}{\partial \sigma_v^2} = -\frac{nT}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} e_{nT}'(\theta_1) \left(H_T^{-1}(\omega) \otimes I_n \right) e_{nT}(\theta_1), \tag{2.5c}$$

where $\Delta \mathbf{Z}_{n,T-1} = [\iota_{nT}, \Delta \mathbf{Y}_{n,T-1}]$ with $\iota_{nT} = [l_n', 0, \ldots, 0]'$, $\Delta \mathbf{Y}_{n,T-1} = [0, \Delta Y_{n1}', \ldots, \Delta Y_{n,T-1}']'$, and $J_T \equiv \frac{\partial H_T(\omega)}{\partial \omega}$ is a diagonal matrix with its $(1,1)$th element being 1 and all other elements being zero.

For further analysis, from Hsiao et al. (2002), by denoting $d = \frac{1}{1+T(\omega-1)}$, we have

$$H_T^{-1}(\omega) = d \cdot \begin{pmatrix} T & T-1 & T-2 & \ldots & 2 & 1 \\ T-1 & (T-1)\omega & (T-2)\omega & \ldots & 2\omega & \omega \\ T-2 & (T-2)\omega & (T-2)(2\omega-1) & \ldots & 2(2\omega-1) & 2\omega-1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 2\omega & 2(2\omega-1) & \ldots & 2((T-2)\omega-(T-3)) & (T-2)\omega-(T-3) \\ 1 & \omega & (2\omega-1) & \ldots & (T-2)\omega-(T-3) & (T-1)\omega-(T-2) \end{pmatrix}, \tag{2.6}$$

and $AH_TA' = D$, where

$$A = \begin{pmatrix} a_0 & & & & \\ a_0 & a_1 & & & \\ a_0 & a_1 & a_2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ a_0 & a_1 & a_2 & \dots & a_{T-1} \end{pmatrix}, \tag{2.7}$$

and $D = \text{diag}\{a_0a_1, a_1a_2, \dots, a_{T-1}a_T\}$ is a diagonal matrix formed by a_sa_{s+1} with $a_s = 1+s(\omega_0-1)$. Thus, $H_T^{-1} = A'D^{-1}A$ and it can be written as

$$H_T^{-1} = A'D^{-1/2} \cdot D^{-1/2}A. \tag{2.8}$$

This decomposition of H_T^{-1} will yield uncorrelated and homoskedastic transformed disturbances $(D^{-1/2}A \otimes I_n)e_{nT}$. Also, by denoting

$$F_T(\gamma) = \begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ \gamma & 1 & 0 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \gamma^{T-2} & \gamma^{T-3} & \dots & 1 & 0 \end{pmatrix}, \tag{2.9}$$

and $F_T^{(1)}(\gamma)$ as the first column of $F_T(\gamma)$, we have $\mathbf{\Delta Y}_{n,T-1} = (F_T(\gamma) \otimes I_n)e_{nT}(\theta_1) + \kappa F_T^{(1)}(\gamma) \otimes l_n$ for any value γ. In particular, $\mathbf{\Delta Y}_{n,T-1} = (F_T \otimes I_n)e_{nT} + \kappa_0 F_T^{(1)}(\gamma_0) \otimes l_n$, so that

$$\text{E}[\mathbf{\Delta Y}'_{n,T-1}(H_T \otimes I_n)^{-1}e_{nT}] = n\sigma_{v0}^2 \,\text{tr}(F_T) = 0. \tag{2.10}$$

With this orthogonality property for $\mathbf{\Delta Y}_{n,T-1}$, and strict exogeneity of l_{nT}, apparently $\text{E}[\mathbf{\Delta Z}'_{n,T-1}(H_T \otimes I_n)^{-1}e_{nT}] = 0$.

2.1 MD estimation and the search for the best IV estimate

The score vector (2.5) gives the moment $\mathbf{\Delta Z}'_{n,T-1} \cdot (H_T \otimes I_n)^{-1} \cdot e_{nT}$. However, $H_T(\omega)$ involves the unknown parameter ω if one would like to work with this moment equation. With a consistent initial estimate $\hat{\omega}$ of ω so that $H_T(\omega)$ can be consistently estimated, Hsiao et al. (2002) suggest an MD estimation of θ_1 by

$$\min_{\theta_1} e'_{nT}(\theta_1)(H_T^{-1}(\hat{\omega}) \otimes I_n)e_{nT}(\theta_1). \tag{2.11}$$

The first order condition of this MD estimation is the score (except the omission of a constant factor) in (2.5a). The resulting MD estimate $\hat{\theta}_{1,md}$ is a generalized instrumental variable (GIV) estimator as

$$\hat{\theta}_{1,md} = \left[\mathbf{\Delta Z}'_{n,T-1}(H_T^{-1}(\hat{\omega}) \otimes I_n)\mathbf{\Delta Z}_{n,T-1}\right]^{-1}\left[\mathbf{\Delta Z}'_{n,T-1}(H_T^{-1}(\hat{\omega}) \otimes I_n)\mathbf{\Delta Y}_{nT}\right],$$

where $\mathbf{\Delta Y}_{nT} = [\Delta Y'_{n1}, ..., \Delta Y'_{nT}]'$. However, the MD estimator is not really a usual GIV estimator in that the asymptotic distribution of the initial consistent estimate $\hat{\omega}$ has an impact on the asymptotic distribution of the MD estimator, even thought the MD estimator is consistent. The remaining issue of interest is whether a similar IV estimate could exist and be asymptotically efficient within a class of IV estimates, assuming that the disturbances in the model are normally distributed.

As the estimation equation involves the predetermined variables in difference, i.e., $\mathbf{\Delta Y}_{n,T-1}$, an IV can be a function of elements in $\mathbf{\Delta Y}_{n,T-1}$, in addition to other strictly exogenous variables, here the constant intercept term. Consider the IV matrix $Q_{nT} = [(K_T \otimes I_n)\iota_{nT}, (A_T \otimes I_n)\mathbf{\Delta Y}_{n,T-1}]$, where K_T and A_T are $T \times T$ constant matrices. The corresponding IV estimate is $\hat{\theta}_{1,iv}$ solved from the empirical moment $Q'_{nT} e_{nT}(\hat{\theta}_{1,iv}) = 0$. For Q_{nT} to be a valid IV, the orthogonality condition $\mathrm{E}(Q'_{nT} e_{nT}) = 0$ is required. With $\mathbf{\Delta Y}_{n,T-1} = (F_T \otimes I_n) e_{nT} + \kappa_0 F_T^{(1)} \otimes l_n$ and $\mathrm{E}(e_{nT} e'_{nT}) = \sigma^2_{v0} H_T \otimes I_n$, a selected A_T shall satisfy $\mathrm{tr}(F'_T A'_T H_T) = 0$. From the expressions of H_T in (2.3) and F_T in (2.9), we can see that $H_T F'_T$ is an upper triangular matrix with its $(1,1)$th element being 0. Then any matrix A_T that is lower triangular with zero diagonals (the first diagonal element can be arbitrary) will satisfy $\mathrm{tr}(F'_T A'_T H_T) = 0$, i.e.,

$$A_T = \begin{pmatrix} 0 & & & & \\ a_{21} & 0 & & & \\ a_{31} & a_{32} & 0 & & \\ \vdots & \vdots & \ddots & \ddots & \\ a_{T1} & a_{T2} & \dots & a_{T,T-1} & 0 \end{pmatrix}. \tag{2.12}$$

By doing so, for $(A_T \otimes I_n)\mathbf{\Delta Y}_{n,T-1}$ and corresponding disturbances $e_{nT} = (\Delta Y'_{n1} - \kappa_0 \iota'_{nT}, \Delta V'_{n2}, ..., \Delta V'_{nT})'$, arbitrary linear combinations of $\Delta y_{i1}, ..., \Delta y_{i,t-2}$ can be IV variables for Δv_{it} with $t \geq 3$. A simple choice of A_T can be $A_T = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}$, and that of K_T can be I_T so that $Q_{nT} = [\iota_{nT}, (0, 0, \Delta Y'_{n1}, ..., \Delta Y'_{n,T-2})']$. In the literature, most have used a sufficient number of time lagged variables for IVs. Thus, those correspond to the use of some A_T in (2.12). In addition to the class of IVs corresponding to (2.12), there might be other IVs, each of which in its general form might be a linear function of all ΔY_{nt}, $t = 1, ..., T-1$, even though e_{nT} is in certain recursive time order. Apparently, one example is $A_T = H_T^{-1}$, the choice implied by the score vector, which satisfies $\mathrm{tr}(F'_T A'_T H_T) = \mathrm{tr}(F_T) = 0$. However, there is a practical issue on using H_T^{-1} because it involves the unknown true parameter ω. But this example does motivate other possible A_T beyond those recursive time ones in (2.12). It is of interest to investigate whether a best A_T in the class of $\mathrm{tr}(F'_T A'_T H_T) = 0$ might exist, so that a corresponding IV matrix in the class $Q_{nT} = [(K_T \otimes I_n)\iota_{nT}, (A_T \otimes I_n)\mathbf{\Delta Y}_{n,T-1}]$ would yield an IV estimate with the smallest asymptotic variance. An IV estimate is

$$\hat{\theta}_{1,iv} = \left(Q'_{nT} \mathbf{\Delta Z}_{n,T-1}\right)^{-1} Q'_{nT} \mathbf{\Delta Y}_{nT} = \theta_{10} + \left(Q'_{nT} \mathbf{\Delta Z}_{n,T-1}\right)^{-1} Q'_{nT} e_{nT}. \tag{2.13}$$

As shown in the supplementary file, under normal disturbances, if A_T satisfies $\operatorname{tr}(F_T'A_T'H_TF_T'A_T'H_T)=0$, e.g., A_T belongs to (2.12) or $A_T=H_T^{-1}$, then $\operatorname{Var}(Q_{nT}'e_{nT})=\sigma_{v0}^2\operatorname{E}[Q_{nT}'(H_T\otimes I_n)Q_{nT}]$.[4] Thus, the asymptotic variance of the IV estimate in (2.13) under normality is

$$\sigma_{v0}^2\left(Q_{nT}'\Delta\mathbf{Z}_{n,T-1}\right)^{-1}Q_{nT}'(H_T\otimes I_n)Q_{nT}\left(\Delta\mathbf{Z}_{n,T-1}'Q_{nT}\right)^{-1}.$$

Its inverse is

$$\begin{aligned}&\frac{1}{\sigma_{v0}^2}\Delta\mathbf{Z}_{n,T-1}'Q_{nT}[Q_{nT}'(H_T\otimes I_T)Q_{nT}]^{-1}Q_{nT}'\Delta\mathbf{Z}_{n,T-1}\\&=\frac{1}{\sigma_{v0}^2}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1/2}\otimes I_T)\cdot(H_T^{1/2}\otimes I_T)Q_{nT}[Q_{nT}'(H_T\otimes I_T)Q_{nT}]^{-1}Q_{nT}'(H_T^{1/2}\otimes I_T)\cdot(H_T^{-1/2}\otimes I_T)\Delta\mathbf{Z}_{n,T-1}\\&\leq\frac{1}{\sigma_{v0}^2}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}\otimes I_T)\Delta\mathbf{Z}_{n,T-1},\end{aligned}$$

by the generalized Schwarz inequality, where the equality holds when $(H_T^{1/2}\otimes I_T)Q_{nT}=(H_T^{-1/2}\otimes I_T)\Delta\mathbf{Z}_{n,T-1}$, i.e., $K_T=A_T=H_T^{-1}$. Thus, under the normality assumption,[5] the best IV in the class of $Q_{nT}=[(K_T\otimes I_n)\iota_{nT},(A_T\otimes I_n)\Delta\mathbf{Y}_{n,T-1}]$, where A_T satisfies $\operatorname{tr}(F_T'A_T'H_T)=0$ and $\operatorname{tr}(F_T'A_T'H_TF_T'A_T'H_T)=0$, shall be $Q_{nT}^*=(H_T^{-1}\otimes I_n)\Delta\mathbf{Z}_{n,T-1}$.

Nevertheless, the best IV requires the use of H_T, which involves the unknown parameter ω_0. Thus, the "best" IV is infeasible. One may attempt to use an initial consistent estimate of ω_0 and hence a consistent estimate $\hat{H}_T$ to construct a feasible IV.[6] However, such a feasible IV would not achieve the same asymptotic variance of the infeasible best IV estimate. This issue has been pointed out by Maddala (1971) for a distributed lag model with serially correlated disturbances. The same issue has been recognized in Hsiao et al. (2002) for the dynamic panel model. With an initial consistent estimate $\tilde{\omega}$, the IV estimation with the empirical moment $\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}(\tilde{\omega})\otimes I_n)e_{nT}(\theta_1)$ is called an MD estimation in Hsiao et al. (2002). The asymptotic distribution of the MD estimator will depend on the asymptotic distribution of the initial estimate $\tilde{\omega}$. This can be seen from an asymptotic expansion of $\frac{1}{n}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}(\tilde{\omega})\otimes I_n)e_{nT}$ at ω_0, which has $\frac{1}{n}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}(\tilde{\omega})\otimes I_n)e_{nT}=\frac{1}{n}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}\otimes I_n)e_{nT}-\frac{1}{n}\Delta\mathbf{Z}_{n,T-1}'(H_T^{-1}\frac{\partial H_T}{\partial\omega}H_T^{-1}\otimes I_n)e_{nT}(\tilde{\omega}-\omega_0)+o_p(1)$, where $\operatorname{E}[\frac{1}{n}\Delta\mathbf{Y}_{n,T-1}'(H_T^{-1}\frac{\partial H_T}{\partial\omega}H_T^{-1}\otimes I_n)e_{nT}]=\sigma_{v0}^2\operatorname{tr}(F_T'H_T^{-1}\frac{\partial H_T}{\partial\omega})\neq 0$, since $\operatorname{tr}(F_T'H_T^{-1}\frac{\partial H_T}{\partial\omega})=\frac{1}{(1-\gamma_0)[1+T(\omega_0-1)]}(T-\frac{1-\gamma_0^T}{1-\gamma_0})$ when $\gamma_0\neq 1$ and $\operatorname{tr}(F_T'H_T^{-1}\frac{\partial H_T}{\partial\omega})=\frac{T(T-1)}{2[1+T(\omega_0-1)]}$ when $\gamma_0=1$.[7]

[4] On the other hand, if $\operatorname{tr}(F_T'A_T'H_TF_T'A_T'H_T)$ is not equal to zero, $\operatorname{Var}(Q_{nT}'e_{nT})$ would not be necessarily equal to $\sigma_{v0}^2\operatorname{E}[Q_{nT}'(H_T\otimes I_n)Q_{nT}]$.

[5] If the disturbances are not normal, from the proof of Theorem 3(iii), the asymptotic variance of the IV estimate $\hat{\theta}_{1,iv}$ is equal to that of an optimal GMM estimator. We show that there is no best GMM under non-normal disturbances in the supplementary file, so there is no best IV under non-normal disturbances.

[6] Initial consistent parameter estimates for various models considered in this paper are given in the supplementary file.

[7] As in Hsiao et al. (2002), when the process $\{y_{it}\}$ starts from a finite past, γ_0 can be 1. We thank an anonymous referee for pointing out this.

In the supplementary file, by restricting our attention to the case with true $\kappa_0 = 0$ being known, we elaborate on the impossibility that H_T could be consistently estimated without asymptotic impact on the feasible IV estimator. Thus, when T is small, there might not be a feasible "best" IV estimate. Without the availability of the best IV, in order to improve efficiency, one might be tempted to use more possible IVs for estimation. But the use of many IVs might give rise to a serious asymptotic bias problem as a by-product (see Bekker, 1994; Donald and Newey, 2001; Alvarez and Arellano, 2003; Chao and Swanson, 2005; Han and Phillips, 2006, etc). The "infeasible IV" issue occurs only for panels with a finite T, where ω in the first entry of $H_T(\omega)$ generates the efficiency issue for estimation. From the supplementary file, we see that under normality, the asymptotic precision of the MLE $\hat{\gamma}_{w,mle}$ is

$$\frac{1}{\sigma_{v0}^2} \mathrm{E}[\Delta \mathbf{Y}'_{n,T-1}(H_T^{-1} \otimes I_n)\Delta \mathbf{Y}_{n,T-1}] - \frac{2n}{T(T-1)(1-\gamma_0)^2}\Big(T - \frac{1-\gamma_0^T}{1-\gamma_0}\Big)^2 \tag{2.14}$$

when $\gamma_0 \neq 1$ and is $\frac{1}{\sigma_{v0}^2} \mathrm{E}[\Delta \mathbf{Y}'_{n,T-1}(H_T^{-1} \otimes I_n)\Delta \mathbf{Y}_{n,T-1}] - \frac{nT(T-1)}{2}$ when $\gamma_0 = 1$, which can be smaller than the asymptotic precision of the best IV estimate, if the latter would exist. For the case with T tending to infinity, asymptotically feasible best IV is possible for the case with $|\gamma_0| < 1$.[8]

2.2 Efficient GMM

We may consider a class of GMM estimators motivated from a direct application of the scores. Under the normality assumption of disturbances, the best moments exist, thus an efficient GMM can be constructed using these best moments. It is also possible to have a computationally simpler approach with a sequential GMM (SGMM), which treats some parameters as nuisance ones, and focus on estimation of remaining structural parameters of interest.

From the first order conditions in (2.5a)–(2.5c), we derive the moment conditions:

$$g_{nT,\kappa}(\theta_2) = \frac{1}{n}\iota'_{nT}(H_T^{-1}(\omega) \otimes I_n)e_{nT}(\theta_1), \tag{2.15a}$$

$$g_{nT,\gamma}(\theta_2) = \frac{1}{n}\Delta \mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega) \otimes I_n)e_{nT}(\theta_1), \tag{2.15b}$$

$$g_{nT,\omega}(\theta_2) = \frac{1}{n}e'_{nT}(\theta_1)(\Phi_T(\omega) \otimes I_n)e_{nT}(\theta_1), \tag{2.15c}$$

where $\theta_2 = (\kappa, \gamma, \omega)'$ and $\Phi_T(\omega) = H_T^{-1}(\omega)J_T H_T^{-1}(\omega) - \frac{1}{T}\operatorname{tr}(H_T^{-1}(\omega)J_T)H_T^{-1}(\omega)$. These moment conditions have mean zero at the true parameter values. We recognize that by using the identity $\Delta \mathbf{Y}_{n,T-1} = (F_T(\gamma) \otimes I_n)[e_{nT}(\theta_1) + \kappa\iota_{nT}]$, and denoting $B_T = D^{-1/2}A$ from (2.8) so that $H_T^{-1} = B'_T B_T$ and $H_T^{-1}(\omega) =$

[8] When T goes to infinity, the second component is dominated by the first one, so that the asymptotic precision of the MLE is asymptotically equal to that of the best IV estimate. The best IV estimation is possible by ignoring the first row of $H_T(\omega)$ or simply replacing it by $H_T(2)$ with 2 replacing ω. The approximation or replacement will be good when T becomes large.

$B_T'(\omega)B_T(\omega)$, the empirical moments (2.15a)–(2.15c) due to scores can be written as

$$g_{nT,\kappa}(\theta_2) = \frac{1}{n}\iota_{nT}'(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\theta_1), \tag{2.16}$$

$$\begin{aligned} g_{nT,\gamma}(\theta_2) = &\frac{1}{n}e_{nT}'(\theta_1)(B_T'(\omega)\otimes I_n)\left[(B_T'^{-1}(\omega)F_T'(\gamma)B_T'(\omega)\otimes I_n)\right](B_T(\omega)\otimes I_n)e_{nT}(\theta_1) \\ &+\frac{1}{n}\kappa\iota_{nT}'(F_T'(\gamma)H_T^{-1}(\omega)\otimes I_n)e_{nT}(\theta_1), \end{aligned} \tag{2.17}$$

and

$$g_{nT,\omega}(\theta_2) = \frac{1}{n}e_{nT}'(\theta_1)(B_T'(\omega)\otimes I_n)\Big[\Big(B_T(\omega)J_TB_T'(\omega) - \frac{1}{T}\operatorname{tr}(J_TH_T^{-1}(\omega))I_T\Big)\otimes I_n\Big](B_T(\omega)\otimes I_n)e_{nT}(\theta_1). \tag{2.18}$$

These empirical moments suggest a class of GMM estimation with linear moments

$$\iota_{nT}'(K_{jT}\otimes I_n)e_{nT}(\theta_1) \tag{2.19}$$

and quadratic moments

$$e_{nT}'(\theta_1)(B_T'(\omega)\otimes I_n)(C_{jT}\otimes I_n)(B_T(\omega)\otimes I_n)e_{nT}(\theta_1), \tag{2.20}$$

where C_{jT}'s can be constant matrices or matrices involving θ_2 with $\operatorname{tr}(C_{jT}) = 0$ at the true θ_{20}. Those matrices C_{jT}'s with their traces being zero will guarantee that the moment $\mathrm{E}[e_{nT}'(B_T'\otimes I_n)(C_{jT}\otimes I_n)(B_T\otimes I_n)e_{nT}] = 0$. To understand the moments in (2.16)–(2.18) from scores, we see that the implied equation (2.2) and the initial ΔY_{n1} are combined into a system $\Delta\mathbf{Y}_{nT} = \gamma_0\Delta\mathbf{Y}_{n,T-1} + \kappa_0\iota_{nT} + e_{nT}$ with the variance of e_{nT} being $\sigma_{v0}^2 H_T\otimes I_n$.[9] A relatively efficient estimate will explore both the main equation and the variance of disturbances for estimation. On the contrary, an IV approach has explored only the main regression equation but not the variance of disturbances, so that efficiency might be lost. For a possible efficient estimation, it is natural to consider the use of quadratic functions in terms of e_{nT}. For the DPD process, it is of interest to know that the moment involving $\Delta\mathbf{Y}_{n,T-1}$ and e_{nT} in a (weighted) product can be rewritten as a linear-quadratic function of e_{nT}. With $H_T^{-1} = B_T'B_T$, the pure DPD process (2.2) in difference can be transformed into

$$(B_T\otimes I_n)\Delta\mathbf{Y}_{nT} = \gamma_0(B_T\otimes I_n)\Delta\mathbf{Y}_{n,T-1} + \kappa_0(B_T\otimes I_n)\iota_{nT} + e_{nT}^*,$$

where $e_{nT}^* = (B_T\otimes I_n)e_{nT}$. The variance of e_{nT}^* is $\sigma_{v0}^2 I_{nT}$, hence the moment condition $\mathrm{E}[e_{nT}^{*\prime}(C_{jT}\otimes I_n)e_{nT}^*] = \sigma_{v0}^2\operatorname{tr}(C_{jT}) = 0$ for any $T\times T$ matrix C_{jT} with zero trace, i.e, $\operatorname{tr}(C_{jT}) = 0$. For the IV estimation with $Q_{nT} = (A_T\otimes I_n)\Delta\mathbf{Y}_{n,T-1}$, where A_T satisfies $\operatorname{tr}(F_T'A_T'H_T) = 0$, the empirical moment is

$$Q_{nT}'e_{nT}(\theta_1) = e_{nT}'(\theta_1)(B_T'(\omega)\otimes I_n)(A_{1T}(\theta_2)\otimes I_n)(B_T(\omega)\otimes I_n)e_{nT}(\theta_1) + \kappa\iota_{nT}'(F_T'(\gamma)A_T'\otimes I_n)e_{nT}(\theta_1),$$

[9] Recall that $\Delta\mathbf{Y}_{nT} = [\Delta Y_{n1}', \ldots, \Delta Y_{nT}']'$ and $\Delta\mathbf{Y}_{n,T-1} = [0, \Delta Y_{n1}', \ldots, \Delta Y_{n,T-1}']'$.

where $A_{1T}(\theta_2) = B_T'^{-1}(\omega)F_T'(\gamma)A_T'B_T^{-1}(\omega)$ has a zero trace at $\theta_2 = \theta_{20}$. This moment combines a linear moment and a quadratic moment with some specific weights. The quadratic moment in (2.17) has $C_{jT} = B_T'^{-1}(\omega)F_T'(\gamma)B_T'(\omega)$, and that in (2.18) has $C_{jT} = B_T(\omega)J_TB_T'(\omega) - \frac{1}{T}\operatorname{tr}(J_nH_T^{-1}(\omega))I_n$.

The GMM estimation can be implemented by treating θ_2 in $K_{jT}(\theta_2)$ and $C_{jT}(\theta_2)$ as unknown in addition to those in $B_T(\omega)$ and $e_{nT}(\theta_1)$. Such a GMM estimation will be a single step approach. On the other hand, the GMM approach may also be implemented with a two-step procedure, where in the first step, one derives a consistent estimate $\tilde{\theta}_2$ so that those K_{jT} and C_{jT} matrices can be consistently estimated by $K_{jT}(\tilde{\theta}_2)$ and $C_{jT}(\tilde{\theta}_2)$, then the suggested quadratic moments in terms of $(B_T(\omega)\otimes I_n)e_{nT}(\theta_1)$ for a GMM estimation is feasible.[10] One can show that a GMM estimate from a single step or a two-step estimation is asymptotically equivalent to the exact GMM estimator by using $K_{jT}(\theta_{20})$ and $C_{jT}(\theta_{20})$ as if they were known.[11]

Assume that we have m_1 such K_{jT} and m_2 such C_{jT}. Then, the vector of moment conditions is

$$g_{nT}(\theta_2) = \frac{1}{n}\begin{pmatrix} \iota_{nT}'(K_{1T}\otimes I_n)e_{nT}(\theta_1) \\ \vdots \\ \iota_{nT}'(K_{m_1T}\otimes I_n)e_{nT}(\theta_1) \\ e_{nT}'(\theta_1)(B_T'(\omega)C_{1T}B_T(\omega)\otimes I_n)e_{nT}(\theta_1) \\ \vdots \\ e_{nT}'(\theta_1)(B_T'(\omega)C_{m_2T}B_T(\omega)\otimes I_n)e_{nT}(\theta_1) \end{pmatrix}. \tag{2.21}$$

At the true θ_{20}, $\iota_{nT}'(K_{jT}\otimes I_n)e_{nT}(\theta_{10}) = \iota_{nT}'(K_{jT}B_T^{-1}\otimes I_n)e_{nT}^*$ and

$$e_{nT}'(\theta_{10})(B_T'(\omega_0)C_{m_2T}B_T(\omega_0)\otimes I_n)e_{nT}(\theta_{10}) = e_{nT}^{*\prime}(C_{m_2T}\otimes I_n)e_{nT}^*,$$

where $e_{nT}^* = (B_T\otimes I_n)e_{nT}$ is homoskedastic and uncorrelated. To derive the analytic form of the variance of $g_{nT}(\theta_{20})$, we can transform e_{nT} into homoskedastic errors as in $\mathbf{V}_{n,T+1}$ via $e_{nT} = (D_{T,T+1}\otimes I_n)\mathbf{V}_{n,T+1}$, where

$$D_{T,T+1} = \begin{pmatrix} -\sqrt{\omega_0 - 1} & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix},$$

$\mathbf{V}_{n,T+1} = [U_{n0}', V_{n1}', \ldots, V_{nT}']'$, and $U_{n0} = \frac{1}{\sqrt{\omega_0-1}}[-\gamma_0^m(\Delta Y_{n,-m+1} - \mathrm{E}\,\Delta Y_{n,-m+1}) + V_{n0} - \sum_{j=1}^{m-1}\gamma_0^j\Delta V_{n,1-j}]$ is independent of V_{nt} and has the same variance as that of V_{nt} for $t = 1, \ldots, T$. The transformation implies that $\Delta Y_{n1} = \kappa_0 l_n + V_{n1} - \sqrt{\omega_0 - 1}U_{n0}$. Notice that $D_{T,T+1}D_{T,T+1}' = H_T$. Then, at the true θ_{20}, elements of $g_{nT}(\theta_{20})$ are either linear or quadratic in $\mathbf{V}_{n,T+1}$, and $\mathrm{Var}(g_{nT}(\theta_{20}))$ generally involves the third and fourth moments of $\mathbf{V}_{n,T+1}$. Let μ_{3v} and μ_{3u} be the third moments of, respectively, v_{it} and

[10] We note that as contrary to later sequential GMM estimation, these moments are quadratic in e_{nT} but not quadratic in the parameter γ because $B_T(\gamma)$ is nonlinear in γ.

[11] See the supplementary file for a proof.

an element u_{i0} of U_{n0}. Define μ_{4v} and μ_{4u} similarly. Let $\mu_{3T} = \text{diag}(\mu_{3u}, \mu_{3v} l'_T)$ be a diagonal matrix formed by the vector $[\mu_{3u}, \mu_{3v} l'_T]$, $\text{vec}_{\text{D}}(A)$ for a square matrix A be a column vector formed by the diagonal elements of A, $A^s = A + A'$, and $\iota_T = [1, 0, \ldots, 0]'$ be a $T \times 1$ unit vector. By Lemma A.1 in Lin and Lee (2010), the variance of $\sqrt{n} g_{nT}(\theta_{20})$ is $\Sigma_T = \begin{pmatrix} \Sigma_{T,11} & \Sigma'_{T,21} \\ \Sigma_{T,21} & \Sigma_{T,22} \end{pmatrix}$, where $\Sigma_{T,11}$ is an $m_1 \times m_1$ matrix with its (r,s)th element being $\sigma^2_{v0}(K_{rT} H_T K'_{sT})_{11}$, $\Sigma_{T,21}$ is a $m_2 \times m_1$ matrix with its (r,s)th element being $\text{vec}_{\text{D}}{}'(D'_{T,T+1} B'_T C_{rT} B_T D_{T,T+1}) \mu_{3T} D'_{T,T+1} K'_{sT} \iota_T$, and $\Sigma_{T,22}$ is a $m_2 \times m_2$ matrix with its (r,s)th element being

$$\begin{aligned} &(\mu_{4u} - 3\sigma^4_{v0})(D'_{T,T+1} B'_T C_{rT} B_T D_{T,T+1})_{11} (D'_{T,T+1} B'_T C_{sT} B_T D_{T,T+1})_{11} \\ &+ (\mu_{4v} - 3\sigma^4_{v0}) \sum_{t=2}^{T+1} (D'_{T,T+1} B'_T C_{rT} B_T D_{T,T+1})_{tt} (D'_{T,T+1} B'_T C_{sT} B_T D_{T,T+1})_{tt} + \sigma^4_{v0} \, \text{tr}(C^s_{rT} C_{sT}). \end{aligned}$$

The optimal GMM estimator with the moment vector $g_{nT}(\theta_2)$ is

$$\hat{\theta}_{2,gmm} = \arg \min_{\theta_2 \in \Theta_2} g'_{nT}(\theta_2) \hat{\Sigma}^{-1}_{nT} g_{nT}(\theta_2), \tag{2.22}$$

where $\hat{\Sigma}_{nT}$ is a consistent estimator of Σ_T and Θ_2 is the parameter space of θ_2.

The above discussion assumes that v_{it}'s are i.i.d. It might be of interest to consider the case that v_{it}'s are independent but $\text{Var}(v^2_{it}) = \sigma^2_t$ depends on T and is unknown. In such a situation, the moments linear in $e_{nT}(\theta_1)$ of (2.21) are still valid, but those quadratic in $e_{nT}(\theta_1)$ might be not. We may investigate whether moments of the form $e'_{nT}(\theta_1)(C_{jT} \otimes I_n) e_{nT}(\theta_1)$ are valid or not. Note that $\text{Var}(e_{nT}) = D^*_{T,T+1} \Xi_{T+1} D^{*'}_{T,T+1} \otimes I_n$, where $\Xi_{T+1} = \text{diag}((\omega_0 - 1)\sigma^2_u, \sigma^2_1, \ldots, \sigma^2_T)$ with $\sigma^2_u = \text{E}(u^2_{i0})$, and $D^*_{T,T+1}$ is equal to $D_{T,T+1}$ except that the $(1,1)$th element of $D^*_{T,T+1}$ is -1. Then

$$\text{E}[e'_{nT}(\theta_{10})(C_{jT} \otimes I_n) e_{nT}(\theta_{10})] = n \cdot \text{tr}(C_{jT} D^*_{T,T+1} \Xi_{T+1} D^{*\prime}_{T,T+1}) = n \cdot \text{tr}(D^{*\prime}_{T,T+1} C_{jT} D^*_{T,T+1} \cdot \Xi_{T+1}).$$

As Ξ_{T+1} is a diagonal matrix, we may choose C_{jT}'s such that the diagonal elements of $D^{*\prime}_{T,T+1} C_{jT} D^*_{T,T+1}$ are zero, which implies that $\text{E}[e'_{nT}(\theta_{10})(C_{jT} \otimes I_n) e_{nT}(\theta_{10})] = 0$ and the quadratic moments are valid. However, due to the special form of $D^*_{T,T+1}$, $C_{jT} = [c_{jT,rs}]$ should satisfy $c_{jT,11} = c_{jT,TT} = 0$ and $c_{jT,r,r-1} = c_{jT,r-1,r-1} + c_{jT,rr} - c_{jT,r-1,r}$ for $r = 2, \ldots, T$. In that case, the best selection of C_{jT} becomes rather complex. Whether the best selection is possible or not remains an issue. For a possible panel model with infinite (large T) periods, this problem becomes even more challenging as it would be a model with infinite number of parameters as σ^2_t's are included.[12] So in this paper, we focus on the i.i.d. case of v_{it}'s and do not consider heterogeneity in v_{it}'s.

[12] We thank a referee for raising this issue.

If $\mathbf{V}_{n,T+1} \sim N(0, \sigma_{v0}^2 I_{n(T+1)})$,

$$\Sigma_T = \sigma_{v0}^2 \begin{pmatrix} (K_{1T}H_TK'_{1T})_{11} & \dots & (K_{1T}H_TK'_{m_1T})_{11} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (K_{m_1T}H_TK'_{1T})_{11} & \dots & (K_{m_1T}H_TK'_{m_1T})_{11} & 0 & \dots & 0 \\ 0 & \dots & 0 & \sigma_{v0}^2 \operatorname{tr}(C_{1T}C_{1T}^s) & \dots & \sigma_{v0}^2 \operatorname{tr}(C_{1T}C_{m_2T}^s) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{v0}^2 \operatorname{tr}(C_{m_2T}C_{1T}^s) & \dots & \sigma_{v0}^2 \operatorname{tr}(C_{m_2T}C_{m_2T}^s) \end{pmatrix}.$$

As $\operatorname{tr}(C_{jT}C_{kT}^s) = \frac{1}{2}\operatorname{tr}(C_{jT}^sC_{kT}^s)$ and $\operatorname{tr}(AB) = \operatorname{vec}'(A')\operatorname{vec}(B)$ for two conformable matrices A and B,

$$\Sigma_T = \sigma_{v0}^2 \Delta_T' \Delta_T, \tag{2.23}$$

where

$$\Delta_T = \begin{pmatrix} B_T'^{-1}K'_{1T}\iota_T & \dots & B_T'^{-1}K'_{m_1T}\iota_T & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{\sqrt{2}}{2}\sigma_{v0}\operatorname{vec}(C_{1T}^s) & \dots & \frac{\sqrt{2}}{2}\sigma_{v0}\operatorname{vec}(C_{m_2T}^s) \end{pmatrix}. \tag{2.24}$$

This variance form can be used to derive the best moment vector under normality, which is

$$g_{nT}^*(\theta_2) = \begin{pmatrix} \iota'_{nT}(H_T^{-1} \otimes I_n)e_{nT}(\theta_1) \\ \iota'_{nT}(F_T'H_T^{-1} \otimes I_n)e_{nT}(\theta_1) \\ e'_{nT}(\theta_1)(B_T'(\omega)C_{1T}^*B_T(\omega) \otimes I_n)e_{nT}(\theta_1) \\ e'_{nT}(\theta_1)(B_T'(\omega)C_{2T}^*B_T(\omega) \otimes I_n)e_{nT}(\theta_1) \end{pmatrix}, \tag{2.25}$$

where $C_{1T}^* = B_TF_TB_T^{-1}$ and $C_{2T}^* = B_{\omega T}B_T^{-1} - \frac{1}{T}\operatorname{tr}(B_{\omega T}B_T^{-1})I_T$ with $B_{\omega T} = \frac{\partial B_{\omega T}(\omega_0)}{\partial \omega}$. As expected from the asymptotic efficiency of the ML approach, the best moment vector under normality corresponds to the score vector.[13,14] While the score vector combines $\iota'_{nT}(F_T'H_T^{-1} \otimes I_n)e_{nT}(\theta_1)$ and $e'_{nT}(\theta_1)(B_T'(\omega)C_{1T}^*B_T(\omega) \otimes I_n)e_{nT}(\theta_1)$ linearly with specific weights, the GMM uses the two moments separately so the number of moments in (2.25) is over-identified. Hence, the corresponding optimal GMM could be efficient relative to the QML as the proper optimal weighting matrix is used for combining the set of linear and quadratic moments for estimation while the QML takes a specific combination of those moments into a score vector for estimation. The best GMM moments are over-identified but the moments of the score vector consist of exactly identified moments.[15] However, the score vector might not be the best combination if the disturbances were not normally distributed.

In the following, we present regularity conditions and asymptotic results on the GMM estimation.

[13]We may show that $(B_{\omega T}B_T^{-1})^s = -B_TJ_TB_T'$. See the proof of Theorem 3 in the supplementary file.

[14]If the disturbances are not normally distributed, we show in the supplementary file that the limiting variance of the GMM estimator $\hat{\theta}_{2,gmm}$ has a lower bound by the generalized Schwarz inequality, but the lower bound cannot be achieved. The reason is that $D_{T,T+1}B_T'C_{jT}^sB_TD_{T,T+1}$ needs to be a diagonal matrix for some C_{jT}^s, but this cannot be the case given the specific form of $D_{T,T+1}$.

[15]If the number of best moments is just identified, and the score vector and the best moment vector are linear transformations of each other, then their estimators would be the same. In this exactly identified moments case, the best GMM estimator would not have an asymptotic gain.

Assumption 2.1. *The disturbances* $\{v_{it}\}$*,* $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$*, are i.i.d. across* i *and* t *with zero mean, variance* σ_{v0}^2 *and* $\mathrm{E}(|v_{it}|^{4+\eta}) < \infty$ *for some* $\eta > 0$*.*

Assumption 2.2. $n \to \infty$ *and* T *is fixed.*

Assumption 2.3. *The process* $\{y_{it}\}$ *has started from a finite but unknown* m *periods ago,* $\Delta y_{i1} = \kappa_0 + v_{i1} - \sqrt{\omega_0 - 1} u_{i0}$*, where* u_{i0}*'s are i.i.d.* $(0, \sigma_{v0}^2)$*,* $\mathrm{E}(|u_{i0}|^{4+\eta}) < \infty$ *for some* $\eta > 0$*, and* u_{i0}*'s are independent of* v_{jt}*'s but with the same variance* σ_{v0}^2*.*

Assumption 2.4. C_{jT}*'s have zero traces and are linearly independent, and* $[K'_{1T,1\cdot}, \ldots, K'_{m_1T,1\cdot}]$ *has full column rank, where* $K_{jT,1\cdot}$ *is the first row of* K_{jT}*.*

Assumption 2.5. *When* $\kappa_0 \neq 0$*,* $\begin{pmatrix} (K_{1T})_{11} & (K_{1T}F_T)_{11} \\ \vdots & \vdots \\ (K_{m_1T})_{11} & (K_{m_1T}F_T)_{11} \end{pmatrix}$ *has full column rank, and*

$$[d_T(\omega)C_{1T}d'_T(\omega), \ldots, d_T(\omega)C_{m_2T}d'_T(\omega)] \neq 0$$

for any $\omega \neq \omega_0$*, where* $d_T(\omega) = [(a_0(\omega)a_1(\omega))^{-1/2}, (a_1(\omega)a_2(\omega))^{-1/2}, \ldots, (a_{T-1}(\omega)a_T(\omega))^{-1/2}]$ *with* $a_t(\omega) = 1 + t(\omega - 1)$*; when* $\kappa_0 = 0$*,* $(K_{jT})_{11} \neq 0$ *for some* $1 \leq j \leq m_1$*, and*

$$\begin{pmatrix} d_T(\omega)C_{1T}d'_T(\omega) & \mathrm{tr}[F'_TB'_T(\omega)C^s_{1T}B_T(\omega)H_T] & \mathrm{tr}[F'_TB'_T(\omega)C_{1T}B_T(\omega)F_TH_T] \\ \vdots & \vdots & \vdots \\ d_T(\omega)C_{m_2T}d'_T(\omega) & \mathrm{tr}[F'_TB'_T(\omega)C^s_{m_2T}B_T(\omega)H_T] & \mathrm{tr}[F'_TB'_T(\omega)C_{m_2T}B_T(\omega)F_TH_T] \end{pmatrix}$$

has full column rank for any ω *in its parameter space.*

Assumption 2.6. *When* $\kappa_0 \neq 0$*,* $\mathrm{tr}(C^s_{jT}B_{\omega T}B_T^{-1}) \neq 0$ *for some* $1 \leq j \leq m_2$*; when* $\kappa_0 = 0$*,*

$$\begin{pmatrix} \mathrm{tr}(C^s_{1T}B_TF_TB_T^{-1}) & \mathrm{tr}(C^s_{1T}B_{\omega T}B_T^{-1}) \\ \vdots & \vdots \\ \mathrm{tr}(C^s_{m_2T}B_TF_TB_T^{-1}) & \mathrm{tr}(C^s_{m_2T}B_{\omega T}B_T^{-1}) \end{pmatrix}$$

has full column rank.

Assumption 2.7. *The parameter space* Θ *of* θ *is compact,* $\omega > 1$*, and* θ_0 *is in the interior of* Θ*.*

Assumption 2.1 states the simple i.i.d. regularity condition on the disturbances v_{it}'s. The moment condition on v_{it} is needed for a proper central limit theorem. The large n and small T asymptotic in this paper is summarized in Assumption 2.2. While we focus on the small T case in this paper, the GMM and SGMM estimates in this paper can also be considered for a large T. Assumption 2.3 states the setting with a finite past and regularity conditions on Δy_{i1}. Assumption 2.4 is a sufficient condition for the nonsingularity of the variance of the moment vector so that the optimal GMM estimator in (2.22) can be formulated. Note that

$g_{nT}(\theta_2)$ only depends on the first rows of K_{jT}'s. Assumptions 2.5 is a sufficient identification condition on the corresponding GMM estimator. Under Assumption 2.6, the expected gradient matrix $G_T = \mathrm{E}(\frac{\partial g_{nT}(\theta_{20})}{\partial \theta_2'})$ has full column rank, so that $\hat{\theta}_{2,gmm}$ has the $\sqrt{n}$-rate of convergence. A compact parameter space in Assumption 2.7 is a usual condition on extremum estimators. The consistency of estimators does not require θ_0 to be in the interior of the parameter space, but the asymptotic distributions need it to avoid issues of a true parameter vector on the boundary of its space. For simplicity, we do not separately state conditions required for consistency and asymptotic distributions. Since $\omega_0 > 1$, we state explicitly that $\omega > 1$ so that $H_T(\omega)$ is positive definite and $H_T^{-1}(\omega) = B_T'(\omega) B_T(\omega)$ for any ω in its parameter space.

Let $\hat{\theta}_{qml}$ be the QML estimator that maximizes the log likelihood function (2.4), $\hat{\theta}_{2,qml}$ be a subvector of $\hat{\theta}_{qml}$ corresponding to θ_2, which as we recall is $(\kappa, \gamma, \omega)'$, $\hat{\theta}^*_{2,gmm}$ be the optimal GMM estimator with the over-identified moment vector $g^*_{nT}(\theta_2)$ in (2.25), $G_T^* = \mathrm{E}(\frac{\partial g^*_{nT}(\theta_{20})}{\partial \theta_2'})$ and $\Sigma_T^* = \mathrm{Var}[\sqrt{n} g^*_{nT}(\theta_{20})]$.

Theorem 1. *Suppose that Assumptions 2.1–2.7 are satisfied.*

(i) The optimal GMM estimator $\hat{\theta}_{2,gmm}$ in (2.22) is consistent and has the asymptotic distribution $\sqrt{n}(\hat{\theta}_{2,gmm} - \theta_{20}) \xrightarrow{d} N(0, (G_T' \Sigma_T^{-1} G_T)^{-1})$, where

$$G_T = \mathrm{E}\Big(\frac{\partial g_{nT}(\theta_{20})}{\partial \theta_2'}\Big) = -\begin{pmatrix} (K_{1T})_{11} & \kappa_0 (K_{1T} F_T)_{11} & 0 \\ \vdots & \vdots & \vdots \\ (K_{m_1 T})_{11} & \kappa_0 (K_{m_1 T} F_T)_{11} & 0 \\ 0 & \sigma_{v0}^2 \,\mathrm{tr}(C^s_{1T} B_T F_T B_T^{-1}) & -\sigma_{v0}^2 \,\mathrm{tr}(C^s_{1T} B_{\omega T} B_T^{-1}) \\ \vdots & \vdots & \vdots \\ 0 & \sigma_{v0}^2 \,\mathrm{tr}(C^s_{m_2 T} B_T F_T B_T^{-1}) & -\sigma_{v0}^2 \,\mathrm{tr}(C^s_{m_2 T} B_{\omega T} B_T^{-1}) \end{pmatrix}.$$

(ii) The QML estimator $\hat{\theta}_{qml}$ is consistent and follows the asymptotic distribution $\sqrt{n}(\hat{\theta}_{qml} - \theta_0) \xrightarrow{d} N(0, \Gamma_{T,\theta})$, where $\Gamma_{T,\theta} = [\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial \theta \partial \theta'})]^{-1} \mathrm{E}(\frac{1}{n}\frac{\partial \ln L_w(\theta_0)}{\partial \theta}\frac{\partial \ln L_w(\theta_0)}{\partial \theta'})[\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial \theta \partial \theta'})]^{-1}$ with[16]

$$\mathrm{E}\Big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial \theta \partial \theta'}\Big) = \begin{pmatrix} \frac{1}{\sigma_{v0}^2}(H_T^{-1})_{11} & * & * & * \\ \frac{\kappa_0}{\sigma_{v0}^2}(H_T^{-1} F_T)_{11} & \frac{\kappa_0^2}{\sigma_{v0}^2}(F_T' H_T^{-1} F_T)_{11} + \mathrm{tr}(F_T' H_T^{-1} F_T H_T) & * & * \\ 0 & \mathrm{tr}(F_T' H_T^{-1} J_T) & \frac{1}{2}\mathrm{tr}(H_T^{-1} J_T H_T^{-1} J_T) & * \\ 0 & 0 & \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(H_T^{-1} J_T) & \frac{T}{2\sigma_{v0}^4} \end{pmatrix}.$$

(iii) $\hat{\theta}^*_{2,gmm}$ is asymptotically efficient relative to $\hat{\theta}_{2,qml}$ in general, i.e., $(G_T^{*\prime} \Sigma_T^{*-1} G_T^*)^{-1} \leq \Gamma_{T,\theta_2}$, where Γ_{T,θ_2} is the asymptotic variance of $\hat{\theta}_{2,qml}$, which is a submatrix of $\Gamma_{T,\theta}$ corresponding to θ_2.

(iv) If $\mathbf{V}_{n,T+1} \sim N(0, \sigma_{v0}^2 I_{n(T+1)})$, then:

[16] The explicit expression of $\mathrm{E}(\frac{1}{n}\frac{\partial \ln L_w(\theta_0)}{\partial \theta}\frac{\partial \ln L_w(\theta_0)}{\partial \theta'})$ can be derived similarly as that of the variance matrix of $\sqrt{n} g_{nT}(\theta_{20})$, thus we omit it for simplicity. We can see that it does not depend on n.

(a) *Among optimal GMM estimators with moments of the form* (2.21), $\hat{\theta}^*_{2,gmm}$ *has the minimum asymptotic variance, i.e.,* $(G^{*\prime}_T \Sigma^{*-1}_T G^*_T)^{-1} \le (G'_T \Sigma^{-1}_T G_T)^{-1}$*, where*

$$G^{*\prime}_T \Sigma^{*-1}_T G^*_T = \begin{pmatrix} \frac{1}{\sigma^2_{v0}}(H^{-1}_T)_{11} & \frac{\kappa_0}{\sigma^2_{v0}}(H^{-1}_T F_T)_{11} & 0 \\ \frac{\kappa_0}{\sigma^2_{v0}}(H^{-1}_T F_T)_{11} & \frac{\kappa^2_0}{\sigma^2_{v0}}(F'_T H^{-1}_T F_T)_{11} + \operatorname{tr}(F'_T H^{-1}_T F_T H_T) & \operatorname{tr}(F'_T H^{-1}_T J_T) \\ 0 & \operatorname{tr}(F'_T H^{-1}_T J_T) & \frac{1}{2}\operatorname{tr}(H^{-1}_T J_T H^{-1}_T J_T) - \frac{1}{2T}\operatorname{tr}^2(H^{-1}_T J_T) \end{pmatrix}.$$

(b) $\hat{\theta}^*_{2,gmm}$ *has the same asymptotic variance as that of* $\hat{\theta}_{2,qml}$*, i.e.,* $(G^{*\prime}_T \Sigma^{*-1}_T G^*_T)^{-1} = \Gamma_{T,\theta_2}$.

Note that the relative efficiency in (*iii*) is possible because the best moments are over-identified for θ_2.

2.3 SGMM

For the above GMM with the efficient moment vector, ω appears in a highly nonlinear way in the moment vector. With an initial consistent estimator of $\tau_0 = (\kappa_0, \omega_0)'$, we may use the moment conditions (2.15a)–(2.15c) derived from the QML scores to define an SGMM estimator that focuses on the estimation of the parameter of interest γ_0. Such an SGMM estimator has a closed form expression and can be asymptotically efficient under normal disturbances.[17] As the asymptotic influence of an initial consistent estimate $\tilde{\omega}$ can be overcome, those SGMM estimators may also improve upon the feasible MD estimate. To derive an efficient estimator of γ_0, we may follow the approach in Jin and Lee (2018), which combines all moments with a $C(\alpha)$-type formulation so that initial consistent estimates of nuisance parameters can be plugged into combined $C(\alpha)$-moments to estimate only parameters of interest. With the proper $C(\alpha)$-type formulation, initial estimates will have no impact on the asymptotic distribution of the estimator of parameters of interest. In the special case that the difference between the number of total moments and the number of combined $C(\alpha)$-moments is equal to the number of nuisance parameters, the approach can generate an efficient estimator of parameters of interest. Alternatively, we may construct an SGMM estimator using concentrated moment conditions. Since $g_{nT,\kappa}(\theta_2) = 0$ yields a closed form solution of κ for given γ and ω,[18] we can substitute this solution into $g_{nT,\gamma}(\theta_2)$ and $g_{nT,\omega}(\theta_2)$ to derive concentrated moment conditions, and then base on these moments to consider an SGMM estimator of γ_0. This approach is in the spirit of Crepon et al. (1997) by estimation with concentrated moments. Since the number of moments reduced by concentration is equal to the number of parameters being concentrated, which is one here for κ, the concentrated moments do not lose information for the estimation of remaining parameters.[19] With concentrated moments, a further moment

[17]Instead of moments based on the score vector, one may used the best moments in (2.25) to obtain an SGMM estimate of γ. But as the number of moments involved is over-identified for γ, the corresponding SGMM would not have a tractable explicit expression. Such an SGMM estimation approach will be considered in a subsequent section on models with exogenous regressors. The moments in (2.25) could be regarded as a special case of the estimation with ι_{nT} as a regressor vector.

[18]The estimate of κ for given γ and ω is $[\iota'_{nT}(H^{-1}_T(\omega) \otimes I_n)\iota_{nT}]^{-1}\iota'_{nT}(H^{-1}_T(\omega) \otimes I_n)(\Delta \mathbf{Y}_{nT} - \gamma \Delta \mathbf{Y}_{n,T-1}) = \frac{1}{n} l'_n \Delta Y_{n1} + \frac{1}{n}\sum_{t=2}^T (1 - \frac{t-1}{T}) l'_n (\Delta Y_{nt} - \gamma \Delta Y_{n,t-1})$, which does not depend on ω.

[19]In our case, the concentration works on the solution of scores, so it is a method of elimination and substitution in solving a system of equations.

reduction is conducted for estimating the parameter of interest γ using the method in Jin and Lee (2018). Thus, the second approach is a combined one of those in Crepon et al. (1997) and Jin and Lee (2018).[20] The first approach is more general in the sense that it works whether or not some nuisance parameters can be concentrated out, while the second approach, when it can be applied, can partly simplify the first approach. In the above two approaches, the final moment condition is quadratic in the unknown parameter γ, so there is a closed-form consistent (and efficient) root estimator. We shall provide conditions that clarify which root of a quadratic equation is consistent and they can be easily employed in practice.[21]

Approach one: SGMM estimator of γ_0

Denote

$$g_{nT,\gamma}(\gamma,\tau) = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\theta_1),$$
$$g_{nT,\tau}(\gamma,\tau) = \frac{1}{n}[\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\theta_1),\ e'_{nT}(\theta_1)(\Phi_T(\omega)\otimes I_n)e_{nT}(\theta_1)]',$$

where $\tau = (\kappa,\omega)'$ and $\Phi_T(\omega)$ is defined below (2.15c). Here, the moments of κ and ω are collected in $g_{nT,\tau}(\gamma,\tau)$. Let $\tilde{\gamma}$ and $\tilde{\tau}$ be, respectively, $\sqrt{n}$-consistent estimators of γ_0 and τ_0. The SGMM estimator $\dot{\gamma}$ of γ_0 is characterized by the following equation

$$g_{nT,\gamma}(\dot{\gamma},\tilde{\tau}) - \hat{C}_{nT,\gamma}g_{nT,\tau}(\dot{\gamma},\tilde{\tau}) = 0, \tag{2.26}$$

where $\hat{C}_{nT,\gamma} = \frac{\partial g_{nT,\gamma}(\tilde{\gamma},\tilde{\tau})}{\partial\tau'}(\frac{\partial g_{nT,\tau}(\tilde{\gamma},\tilde{\tau})}{\partial\tau'})^{-1}$. By a Taylor expansion, one can see that, as an asymptotic orthogonality condition is satisfied, the initial estimate $\tilde{\tau}$ would not have an asymptotic impact on the moment $g_{nT,\gamma}(\gamma,\tilde{\tau}) - \hat{C}_{nT,\gamma}g_{nT,\tau}(\gamma,\tilde{\tau})$, so the asymptotic distribution of the initial estimate $\tilde{\tau}$ would not have an influence on the asymptotic distribution of the SGMM estimator $\dot{\gamma}$ of γ.[22] The $C(\alpha)$-moment $g_{nT,\gamma}(\gamma,\tilde{\tau}) - \hat{C}_{nT,\gamma}g_{nT,\tau}(\gamma,\tilde{\tau})$ has one less moment than the number of moments in $g_{nT}(\gamma,\tau) = [g_{nT,\gamma}(\gamma,\tau), g'_{nT,\tau}(\gamma,\tau)]'$, and the resulting SGMM estimator of γ_0 is asymptotically as efficient as that of γ_0 from the joint GMM estimation with the moment vector $g_{nT}(\gamma,\tau)$. As the GMM moments are constructed from scores, in the case of normality, the SGMM estimator $\dot{\gamma}$ is efficient relative to the MD estimator $\check{\gamma}$ of γ_0 in Hsiao et al. (2002), which is characterized by $\Delta\mathbf{Z}'_{n,T-1}(H_T^{-1}(\tilde{\omega})\otimes I_n)e_{nT}(\check{\theta}_1) = 0$. The combined $C(\alpha)$-moment is intended to improve robustness and retain efficiency of the $g_{nT,\gamma}(\gamma,\tau)$ moment for estimation. But while the moment

[20] The SGMM in Jin and Lee (2018) is asymptotically equivalent to the approach in Trognon and Gouriéroux (1990) applied to the GMM, which is derived by a first order Taylor expansion of the moment vector at the nuisance parameter estimator.

[21] Root estimators for spatial autoregressive models are considered in Jin and Lee (2012).

[22] The consistent estimation of $\text{plim}_{n\to\infty}\frac{\partial g_{nT,\gamma}(\gamma_0,\tau_0)}{\partial\tau'}(\frac{\partial g_{nT,\tau}(\gamma_0,\tau_0)}{\partial\tau'})^{-1}$ by $\hat{C}_{nT,\gamma}$ would not have an asymptotic influence on the moment equation due to its role as coefficients for linear combinations of valid moments.

$g_{nT,\gamma}(\gamma,\tau)$ is linear in γ, the combined $C(\alpha)$-moment is quadratic in γ. The importance of a score vector in terms of a quadratic function of e_{nT} is highly visible in the ML estimation of a short dynamic panel.

By denoting $\tilde{\Phi}_T = \Phi_T(\tilde{\omega})$ and $\tilde{H}_T = H_T(\tilde{\omega})$, the quadratic equation (2.26) of $\dot{\gamma}$ can be rewritten as

$$s_{nT,1}\dot{\gamma}^2 + s_{nT,2}\dot{\gamma} + s_{nT,3} = 0,$$

where $s_{nT,1} = \frac{1}{n}\hat{C}_{nT,\gamma}\big[0, \Delta\mathbf{Y}'_{n,T-1}(\tilde{\Phi}_T \otimes I_n)\Delta\mathbf{Y}_{n,T-1}\big]'$,

$$s_{nT,2} = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1} - \frac{1}{n}\hat{C}_{nT,\gamma}\big[\iota'_{nT}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1},\ 2\Delta\mathbf{Y}'_{n,T-1}(\tilde{\Phi}_T\otimes I_n)(\Delta\mathbf{Y}_{nT} - \tilde{\kappa}\iota_{nT})\big]',$$

and

$$s_{nT,3} = \frac{1}{n}\hat{C}_{nT,\gamma}\big[\iota'_{nT}(\tilde{H}_T^{-1}\otimes I_n)(\Delta\mathbf{Y}_{nT} - \tilde{\kappa}\iota_{nT}),\ (\Delta\mathbf{Y}_{nT} - \tilde{\kappa}\iota_{nT})'(\tilde{\Phi}_T\otimes I_n)(\Delta\mathbf{Y}_{nT} - \tilde{\kappa}\iota_{nT})\big]' - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)(\Delta\mathbf{Y}_{nT} - \tilde{\kappa}\iota_{nT}).$$

Using $\Delta\mathbf{Y}_{nT} = \gamma_0\Delta\mathbf{Y}_{n,T-1} + \kappa_0\iota_{nT} + e_{nT}$, we have $s_{nT,2} = -s_{nT,4} - 2\gamma_0 s_{nT,1} + o_p(1)$, where

$$s_{nT,4} = \frac{1}{n}\hat{C}_{nT,\gamma}\big[\iota'_{nT}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1},\ 2\Delta\mathbf{Y}'_{n,T-1}(\tilde{\Phi}_T\otimes I_n)e_{nT}\big]' - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1},$$

and $s_{nT,3} = \gamma_0 s_{nT,4} + \gamma_0^2 s_{nT,1} + o_p(1)$ as $\frac{1}{n}e'_{nT}(\tilde{\Phi}_T\otimes I_n)e_{nT} = o_p(1)$ and $\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)e_{nT} = o_p(1)$. The quadratic equation can have the solutions on γ as

$$\frac{-s_{nT,2} \pm \sqrt{s_{nT,2}^2 - 4s_{nT,1}s_{nT,3}}}{2s_{nT,1}} = \gamma_0 + \frac{s_{nT,4} \pm \sqrt{s_{nT,4}^2 + o_p(1)}}{2s_{nT,1}}.$$

Thus, the consistent root is $\frac{-s_{nT,2} - \sqrt{s_{nT,2}^2 - 4s_{nT,1}s_{nT,3}}}{2s_{nT,1}}$ if $s_{nT,4} \geq 0$, or $\frac{-s_{nT,2} + \sqrt{s_{nT,2}^2 - 4s_{nT,1}s_{nT,3}}}{2s_{nT,1}}$ if $s_{nT,4} < 0$. In practice, $s_{nT,4}$ can be estimated by

$$\tilde{s}_{nT,4} = \frac{1}{n}\hat{C}_{nT,\gamma}\big[\iota'_{nT}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}, 2\Delta\mathbf{Y}'_{n,T-1}(\tilde{\Phi}_T\otimes I_n)e_{nT}(\tilde{\theta}_1)\big]' - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}$$

to determine the consistent root.[23]

Approach two: SGMM estimator of γ_0 with concentrated moments

If we set $g_{nT,\kappa}(\theta_2)$ in (2.15a) to zero, then the estimate of κ for given γ and ω is $[\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)\iota_{nT}]^{-1}\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1})$.[24] Substituting this estimate into (2.15b) and (2.15c)

[23] The probability limit of $s_{nT,4}$ depends on ω_0, γ_0, T and σ_{v0}^2, and it can be positive or negative.

[24] This estimate can be further simplified to $\frac{1}{n}l'_n\Delta Y_{n1} + \frac{1}{n}\sum_{t=2}^T(1-\frac{t-1}{T})l'_n(\Delta Y_{nt} - \gamma\Delta Y_{n,t-1})$, which does not depend on ω. Using the form $[\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)\iota_{nT}]^{-1}\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1})$ simplifies the presentation of the concentrated moments.

yields the following two moment conditions:

$$g_{nT,\gamma c}(\gamma,\omega) = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega)\otimes I_n)M_{nT}(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1}),$$
$$g_{nT,\omega c}(\gamma,\omega) = \frac{1}{n}(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1})'M'_{nT}(\Phi_T(\omega)\otimes I_n)M_{nT}(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1}),$$

where $M_{nT} = I_{nT} - \iota_{nT}[\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n)\iota_{nT}]^{-1}\iota'_{nT}(H_T^{-1}(\omega)\otimes I_n) = I_{nT} - \varphi_T\otimes(\frac{1}{n}l_n l'_n)$ and $\varphi_T = \iota_T\cdot[1, 1-\frac{1}{T},\ldots,\frac{1}{T}]$ is a $T\times T$ matrix. Let $\tilde{\gamma}$ and $\tilde{\omega}$ be initial $\sqrt{n}$-consistent estimators of γ_0 and ω_0 respectively. Denote $\hat{C}_{nT,\gamma c} = \frac{\partial g_{nT,\gamma c}(\tilde{\gamma},\tilde{\omega})}{\partial\omega}(\frac{\partial g_{nT,\omega c}(\tilde{\gamma},\tilde{\omega})}{\partial\omega})^{-1}$. The SGMM estimator $\ddot{\gamma}$ is characterized by the equation

$$g_{nT,\gamma c}(\ddot{\gamma},\tilde{\omega}) - \hat{C}_{nT,\gamma c}g_{nT,\omega c}(\ddot{\gamma},\tilde{\omega}) = 0. \quad (2.27)$$

This equation can be rewritten as

$$s_{nT,1c}\ddot{\gamma}^2 + s_{nT,2c}\ddot{\gamma} + s_{nT,3c} = 0,$$

where $s_{nT,1c} = \frac{1}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}\Delta\mathbf{Y}_{n,T-1}$,

$$s_{nT,2c} = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)M_{nT}\Delta\mathbf{Y}_{n,T-1} - \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}\Delta\mathbf{Y}_{nT},$$

and $s_{nT,3c} = \frac{1}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{nT}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}\Delta\mathbf{Y}_{nT} - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)M_{nT}\Delta\mathbf{Y}_{nT}$. Using $\Delta\mathbf{Y}_{nT} = \gamma_0\Delta\mathbf{Y}_{n,T-1} + \kappa_0\iota_{nT} + e_{nT}$, we have $s_{nT,2c} = -s_{nT,4c} - 2\gamma_0 s_{nT,1c}$, where

$$s_{nT,4c} = \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}e_{nT} - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)M_{nT}\Delta\mathbf{Y}_{n,T-1},$$

and $s_{nT,3c} = \gamma_0 s_{nT,4c} + \gamma_0^2 s_{nT,1c} + o_p(1)$ as $\frac{1}{n}e'_{nT}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}e_{nT} = o_p(1)$ and $\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)M_{nT}e_{nT} = o_p(1)$. The quadratic equation can have the solutions on γ as

$$\frac{-s_{nT,2c}\pm\sqrt{s_{nT,2c}^2 - 4s_{nT,1c}s_{nT,3c}}}{2s_{nT,1c}} = \gamma_0 + \frac{s_{nT,4c}\pm\sqrt{s_{nT,4c}^2 + o_p(1)}}{2s_{nT,1c}}.$$

Thus, the consistent root is $\frac{-s_{nT,2c}-\sqrt{s_{nT,2c}^2-4s_{nT,1c}s_{nT,3c}}}{2s_{nT,1c}}$ if $s_{nT,4c}\geq 0$, or $\frac{-s_{nT,2c}+\sqrt{s_{nT,2c}^2-4s_{nT,1c}s_{nT,3c}}}{2s_{nT,1c}}$ if $s_{nT,4c} < 0$. The $S_{nT,4c}$ can be estimated by

$$\tilde{s}_{nT,4c} = \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}M'_{nT}(\tilde{\Phi}_T\otimes I_n)M_{nT}(\Delta\mathbf{Y}_{nT}-\tilde{\gamma}\Delta\mathbf{Y}_{n,T-1}) - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)M_{nT}\Delta\mathbf{Y}_{n,T-1}.$$

For the asymptotic variance of $\sqrt{n}[g_{nT,\gamma c}(\gamma_0,\omega_0), g_{nT,\omega c}(\gamma_0,\omega_0)]'$, using $Y_{n,T-1} = (F_T\otimes I_n)(e_{nT}+\kappa_0\iota_{nT})$, we may show that $\sqrt{n}[g_{nT,\gamma c}(\gamma_0,\omega_0), g_{nT,\omega c}(\gamma_0,\omega_0)]' = g_{nT,c}(\gamma_0,\omega_0)+o_p(1)$, where $g_{nT,c}(\gamma_0,\omega_0) = [\frac{1}{\sqrt{n}}e'_{nT}(F'_T H_T^{-1}\otimes I_n)e_{nT} + \frac{\kappa_0}{\sqrt{n}}\iota'_{nT}(F'_T H_T^{-1}\otimes I_n)M_{nT}e_{nT}, \frac{1}{\sqrt{n}}e'_{nT}(\Phi_T\otimes I_n)e_{nT}]'$. Then under regularity conditions,

$$\sqrt{n}[g_{nT,\gamma c}(\gamma_0,\omega_0), g_{nT,\omega c}(\gamma_0,\omega_0)]' \xrightarrow{d} N(0,\Sigma_{T,\gamma c}),$$

where $\Sigma_{T,\gamma c} = \mathrm{Var}[\sqrt{n} g_{nT,c}(\gamma_0, \omega_0)]$. This result can be used to derive the asymptotic distribution of $\ddot{\gamma}$.

Equation (2.26) is quadratic in the limit under the following Assumption 2.8, and so is (2.27) under Assumption 2.9.

Assumption 2.8. $\sigma_{v0}^2 \operatorname{tr}(F_T' \Phi_T F_T H_T) + \kappa_0^2 (F_T' \Phi_T F_T)_{11} \neq 0$.

Assumption 2.9. $\sigma_{v0}^2 \operatorname{tr}(F_T' \Phi_T F_T H_T) + \kappa_0^2 \big[\big(F_T - \frac{(F_T' H_T^{-1})_{11}}{(H_T^{-1})_{11}} I_T\big)' \Phi_T \big(F_T - \frac{(F_T' H_T^{-1})_{11}}{(H_T^{-1})_{11}} I_T\big)\big]_{11} \neq 0$.

Theorem 2. *Suppose that Assumptions 2.1–2.7 are satisfied.*

(i) If Assumption 2.8 also holds, the SGMM estimator $\dot{\gamma}$ is consistent and follows the asymptotic distribution

$$\sqrt{n}(\dot{\gamma} - \gamma_0) \xrightarrow{d} N(0, (R_{T\gamma} G_{T\gamma})^{-2} R_{T\gamma} \Sigma_{T\gamma} R_{T\gamma}'),$$

where $\Sigma_{T\gamma} = \mathrm{Var}(\sqrt{n} g_{nT}(\gamma_0, \tau_0))$, $R_{T\gamma} = [1, -\kappa_0 \frac{(F_T' H_T^{-1})_{11}}{(H_T^{-1})_{11}}, -\frac{\operatorname{tr}(F_T' H_T^{-1} J_T)}{\operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) - \frac{1}{T} \operatorname{tr}^2(H_T^{-1} J_T)}]$ *and*

$$G_{T\gamma} = -[\sigma_{v0}^2 \operatorname{tr}(F_T' H_T^{-1} F_T H_T) + \kappa_0^2 (F_T' H_T^{-1} F_T)_{11}, \kappa_0 (F_T' H_T^{-1})_{11}, 2\sigma_{v0}^2 \operatorname{tr}(F_T' H_T^{-1} J_T)]'.$$

The asymptotic distribution of $\dot{\gamma}$ is the same as that of the QML estimator $\hat{\gamma}_{qml}$ of γ in Theorem 1(ii).

(ii) If Assumption 2.9 also holds, the SGMM estimator $\ddot{\gamma}$ is consistent and follows the asymptotic distribution

$$\sqrt{n}(\ddot{\gamma} - \gamma_0) \xrightarrow{d} N(0, (R_{T,\gamma c} G_{T,\gamma c})^{-2} R_{T,\gamma c} \Sigma_{T\gamma,c} R_{T,\gamma c}'),$$

where $\Sigma_{T,\gamma c} = \mathrm{Var}(\sqrt{n} g_{nT,c}(\gamma_0, \tau_0))$, $R_{T,\gamma c} = [1, -\frac{\operatorname{tr}(F_T' H_T^{-1} J_T)}{\operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) - \frac{1}{T} \operatorname{tr}^2(H_T^{-1} J_T)}]$ *and*

$$G_{T,\gamma c} = -\big[\sigma_{v0}^2 \operatorname{tr}(F_T' H_T^{-1} F_T H_T) + \kappa_0^2 (F_T' H_T^{-1} F_T)_{11} - \kappa_0^2 \frac{[(F_T' H_T^{-1})_{11}]^2}{(H_T^{-1})_{11}}, 2\sigma_{v0}^2 \operatorname{tr}(F_T' H_T^{-1} J_T)\big]'.$$

The asymptotic distribution of $\ddot{\gamma}$ is the same as that of the QML estimator $\hat{\gamma}_{qml}$ of γ in Theorem 1(ii).

Note that the asymptotic variances in Theorem 2(i)–(ii) are equal, though they are written in different forms.

3 Fixed effects DPD with exogenous variables

Consider the fixed effects DPD model with exogenous variables:

$$y_{it} = \gamma_0 y_{i,t-1} + x_{it} \beta_0 + c_{i0} + v_{it}, \quad t = 1, \ldots, T,$$

and y_{i0} is observable. In this model, c_{i0} absorbs all time-invariant regressors. For y_{i0}, by continuous substitution, we have

$$y_{i0} = \gamma_0^m y_{i,-m} + x_{i0}\beta_0 + \sum_{j=1}^{m-1} \gamma_0^j x_{i,-j}\beta_0 + c_{i0} \sum_{j=0}^{m-1} \gamma_0^j + \sum_{j=0}^{m-1} \gamma_0^j v_{i,-j}.$$

For c_{i0} and the unobserved $x_{i,-1}, \ldots, x_{i,-m+1}$ and $y_{i,-m}$, we may use $\vec{x}_i = [x_{i0}, \ldots, x_{iT}]$ and observable time-invariant regressors z_i to predict them. Assume that $\mathrm{E}(c_{i0}|z_i, \vec{x}_i) = z_i\varrho_c + \vec{x}_i\pi_c$, $\mathrm{E}(y_{i,-m}|z_i, \vec{x}_i) = z_i\varrho + \vec{x}_i\pi$ and $\mathrm{E}(x_{i,-j}|z_i, \vec{x}_i) = z_i\varrho_j + \vec{x}_i\pi_j$ for $j = 1, \ldots, m-1$. Then, $y_{i0} = z_i\alpha^{(1)} + \vec{x}_i\alpha^{(2)} + \xi_{i0}$, where $\alpha^{(1)} = \gamma_0^m\varrho + \sum_{j=1}^{m-1}\gamma_0^j\varrho_j\beta_0 + \varrho_c\sum_{j=0}^{m-1}\gamma_0^j$, $\alpha^{(2)} = \gamma_0^m\pi + [\beta_0', 0, \ldots, 0]' + \sum_{j=1}^{m-1}\gamma_0^j\pi_j\beta_0 + \pi_c\sum_{j=0}^{m-1}\gamma_0^j$, $\xi_{i0} = \sum_{j=0}^{m-1}\gamma_0^j v_{i,-j} + p_i$, and $p_i = \gamma_0^m[y_{i,-m} - \mathrm{E}(y_{i,-m}|z_i, \vec{x}_i)] + \sum_{j=1}^{m-1}\gamma_0^j[x_{i,-j} - \mathrm{E}(x_{i,-j}|z_i, \vec{x}_i)]\beta_0 + [c_{i0} - \mathrm{E}(c_{i0}|z_i, \vec{x}_i)]\sum_{j=0}^{m-1}\gamma_0^j$ is the prediction error. It follows that

$$\Delta y_{i1} = [z_i, \vec{x}_i]\alpha_0 + \xi_{i1},$$

where $\alpha_0 = \left[(\gamma_0 - 1)\alpha^{(1)'} + \varrho_c', (\gamma_0 - 1)\alpha^{(2)'} + \pi_c' + [0, \beta_0', 0, \ldots, 0]\right]'$ is a free parameter vector, and $\xi_{i1} = (\gamma_0 - 1)\xi_{i0} + v_{i1} + [c_{i0} - \mathrm{E}(c_{i0}|z_i, \vec{x}_i)] = v_{i1} + (\gamma_0 - 1)\sum_{j=0}^{m-1}\gamma_0^j v_{i,-j} + p_i^*$ is the error term with $p_i^* = (\gamma_0 - 1)\gamma_0^m[y_{i,-m} - \mathrm{E}(y_{i,-m}|z_i, \vec{x}_i)] + (\gamma_0 - 1)\sum_{j=1}^{m-1}\gamma_0^j[x_{i,-j} - \mathrm{E}(x_{i,-j}|z_i, \vec{x}_i)]\beta_0 + \gamma_0^m[c_{i0} - \mathrm{E}(c_{i0}|z_i, \vec{x}_i)]$ being the overall prediction error. As in Hsiao et al. (2002), p_i^*'s are assumed to be i.i.d. with zero mean and a finite variance. With this assumption, ξ_{i1} has overall zero mean and finite variance. Let the variance of ξ_{i1} be $\sigma_{v0}^2\omega_0$, where ω_0 is a free parameter due to the prediction errors in ξ_{i1}. Thus, regardless whether the process has started from a finite m or infinite past as $m \to \infty$, due to the prediction error on exogenous variables in the past, the variance parameter ω of ξ_{i1} is free from a restriction with γ in this model.

The quasi log likelihood function for the within model of $\Delta\mathbf{Y}_{nT}$ is

$$\ln L_w(\theta) = -\frac{nT}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|H_T(\omega)| - \frac{1}{2\sigma_v^2}e_{nT}'(\alpha, \delta)(H_T^{-1}(\omega) \otimes I_n)e_{nT}(\alpha, \delta), \quad (3.1)$$

where $\delta = (\beta', \gamma)'$, $\theta = (\alpha', \delta', \omega, \sigma_v^2)'$, $e_{nT}(\alpha, \delta) = \Delta\mathbf{Y}_{nT} - \Delta\mathbf{Z}_{nT}\delta - \Upsilon_{nT}\alpha$,

$$\Delta\mathbf{Z}_{nT} = \begin{pmatrix} 0 & 0 \\ \Delta X_{n2} & \Delta Y_{n1} \\ \vdots & \vdots \\ \Delta X_{nT} & \Delta Y_{n,T-1} \end{pmatrix} \text{ and } \Upsilon_{nT} = \begin{pmatrix} Z_n & \vec{\mathbf{X}}_{nT} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix},$$

for $Z_n = [z_1', \ldots, z_n']'$ and $\overrightarrow{\mathbf{X}}_{nT} = [\vec{x}_1', ..., \vec{x}_n']'$. The first order derivatives of (3.1) are

$$\begin{aligned}
\frac{\partial \ln L_w(\theta)}{\partial \alpha} &= \frac{1}{\sigma_v^2} \Upsilon_{nT}'(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\alpha, \delta), \\
\frac{\partial \ln L_w(\theta)}{\partial \delta} &= \frac{1}{\sigma_v^2} \Delta \mathbf{Z}_{nT}'(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\alpha, \delta), \\
\frac{\partial \ln L_w(\theta)}{\partial \omega} &= -\frac{n}{2} \operatorname{tr}[H_T^{-1}(\omega) J_T] + \frac{1}{2\sigma_v^2} e_{nT}'(\alpha, \delta)(H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \otimes I_n) e_{nT}(\alpha, \delta), \\
\frac{\partial \ln L_w(\theta)}{\partial \sigma_v^2} &= -\frac{nT}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} e_{nT}'(\alpha, \delta)(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\alpha, \delta).
\end{aligned}$$

3.1 Efficient GMM

For the fixed effects DPD model with exogenous variables, $\Delta \mathbf{Y}_{n,T-1} = (F_T \otimes I_n)(e_{nT} + \Delta \mathbf{X}_{nT} \beta_0 + \Upsilon_{nT} \alpha_0)$, where $\Delta \mathbf{X}_{nT} = [0, \Delta X_{n2}', \ldots, \Delta X_{nT}']'$. Then the moment condition $\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} \otimes I_n) e_{nT} = e_{nT}'(F_T' H_T^{-1} \otimes I_n) e_{nT} + (\Delta \mathbf{X}_{nT} \beta_0 + \Upsilon_{nT} \alpha_0)'(F_T' H_T^{-1} \otimes I_n) e_{nT}$ is linear-quadratic in e_{nT}. Let $\mathbf{X}_{nT}^* = [\Upsilon_{nT}, \Delta \mathbf{X}_{nT}]$. From the score vector of the quasi log likelihood function, we may consider the following moment vector for a GMM estimation:

$$g_{nT}(\theta_3) = \frac{1}{n} \begin{pmatrix} \mathbf{X}_{nT}^{*\prime}(K_{1T} \otimes I_n) e_{nT}(\alpha, \delta) \\ \vdots \\ \mathbf{X}_{nT}^{*\prime}(K_{m_1 T} \otimes I_n) e_{nT}(\alpha, \delta) \\ e_{nT}'(\alpha, \delta)(B_T'(\omega) C_{1T} B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta) \\ \vdots \\ e_{nT}'(\alpha, \delta)(B_T'(\omega) C_{m_2 T} B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta) \end{pmatrix}, \tag{3.2}$$

where $\theta_3 = [\alpha', \delta', \omega]'$, $e_{nT}(\alpha, \delta) = \Delta \mathbf{Y}_{nT} - \Delta \mathbf{Z}_{nT} \delta - \Upsilon_{nT} \alpha$, and C_{jT}'s have zero traces. Let $\Sigma_{nT} = \operatorname{Var}[\sqrt{n} g_{nT}(\theta_{30})]$ and $\hat{\Sigma}_{nT}$ be a consistent estimator of $\lim_{n\to\infty} \Sigma_{nT}$. The optimal GMM estimator with $g_{nT}(\theta_3)$ is

$$\hat{\theta}_{3,gmm} = \arg \min_{\theta_3 \in \Theta_3} g_{nT}'(\theta_3) \hat{\Sigma}_{nT}^{-1} g_{nT}(\theta_3), \tag{3.3}$$

where Θ_3 is the parameter space of θ_3. When the disturbances are normally distributed, we may show by the generalized Schwarz inequality that the best moment vector among moment vectors of the form (3.2) is

$$g_{nT}^*(\theta_3) = \frac{1}{n} \begin{pmatrix} \mathbf{X}_{nT}^{*\prime}(H_T^{-1} \otimes I_n) e_{nT}(\alpha, \delta) \\ \mathbf{X}_{nT}^{*\prime}(F_T' H_T^{-1} \otimes I_n) e_{nT}(\alpha, \delta) \\ e_{nT}'(\alpha, \delta)(B_T'(\omega) C_{1T}^* B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta) \\ e_{nT}'(\alpha, \delta)(B_T'(\omega) C_{2T}^* B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta) \end{pmatrix}, \tag{3.4}$$

where $C_{1T}^* = B_T F_T B_T^{-1}$ and $C_{2T}^* = B_{\omega T} B_T^{-1} - \frac{1}{T} \operatorname{tr}(B_{\omega T} B_T^{-1}) I_T$ with $B_{\omega T} = \frac{\partial B_T(\omega_0)}{\partial \omega}$. This moment vector corresponds to the QML score vector.

We give required regularity conditions below and present asymptotic results on the GMM estimation.

Assumption 3.1. *The process* $\{y_{it}\}$ *has started from either the infinite past or a finite but unknown* m *periods ago,* $\Delta y_{i1} = [z_i, \vec{x}_i]\alpha_0 + v_{i1} - \sqrt{\omega_0 - 1}u_{i0}$*, where* u_{i0}*'s are i.i.d.* $(0, \sigma_{v0}^2)$*,* $\mathrm{E}(|u_{i0}|^{4+\eta}) < \infty$ *for some* $\eta > 0$*, and* u_{i0}*'s are independent of* v_{jt}*'s, even though they have the same variance* σ_{v0}^2*.*

Assumption 3.2. X_{nt} *and* Z_n *are nonstochastic such that* $\sup_{l,t,n} \frac{1}{n}\sum_{i=1}^n |x_{it,l}|^{2+\eta} < \infty$ *and* $\sup_{l,n} \frac{1}{n}\sum_{i=1}^n |z_{i,l}|^{2+\eta} < \infty$ *for some* $\eta > 0$*, where* $x_{it,l}$ *is the* (i,l)*th element of* X_{nt} *and* $z_{i,l}$ *is the* (i,l)*th element of* Z_n*.*

Assumption 3.3. C_{jT}*'s have zero traces and are linearly independent, and* $\lim_{n\to\infty} \frac{1}{n}[(K_{1T}' \otimes I_n)\mathbf{X}_{nT}^*, \ldots, (K_{m_1T}' \otimes I_n)\mathbf{X}_{nT}^*]'[(K_{1T}' \otimes I_n)\mathbf{X}_{nT}^*, \ldots, (K_{m_1T}' \otimes I_n)\mathbf{X}_{nT}^*]$ *has full rank.*

Assumption 3.4. *When* $(\alpha_0', \beta_0')' \neq 0$*,* $\lim_{n\to\infty} \frac{1}{n}\begin{pmatrix} \mathbf{X}_{nT}^{*\prime}(K_{1T}\otimes I_n)\mathbf{X}_{nT}^* & \mathbf{X}_{nT}^{*\prime}(K_{1T}F_T\otimes I_n)\mathbf{X}_{nT}^*\binom{\alpha_0}{\beta_0} \\ \vdots & \vdots \\ \mathbf{X}_{nT}^{*\prime}(K_{m_1T}\otimes I_n)\mathbf{X}_{nT}^* & \mathbf{X}_{nT}^{*\prime}(K_{m_1T}F_T\otimes I_n)\mathbf{X}_{nT}^*\binom{\alpha_0}{\beta_0} \end{pmatrix}$ *has full column rank, and* $[d_T(\omega)C_{1T}d_T'(\omega), \ldots, d_T(\omega)C_{m_2T}d_T'(\omega)] \neq 0$ *for any* $\omega \neq \omega_0$*, where*

$$d_T(\omega) = [(a_0(\omega)a_1(\omega))^{-1/2}, (a_1(\omega)a_2(\omega))^{-1/2}, \ldots, (a_{T-1}(\omega)a_T(\omega))^{-1/2}]$$

with $a_t(\omega) = 1 + t(\omega - 1)$*; when* $(\alpha_0', \beta_0')' = 0$*,* $\lim_{n\to\infty} \frac{1}{n}\begin{pmatrix} \mathbf{X}_{nT}^{*\prime}(K_{1T}\otimes I_n)\mathbf{X}_{nT}^* \\ \vdots \\ \mathbf{X}_{nT}^{*\prime}(K_{m_1T}\otimes I_n)\mathbf{X}_{nT}^* \end{pmatrix}$ *has full column rank, and*

$$\begin{pmatrix} d_T(\omega)C_{1T}d_T'(\omega) & \mathrm{tr}[F_T'B_T'(\omega)C_{1T}^sB_T(\omega)H_T] & \mathrm{tr}[F_T'B_T'(\omega)C_{1T}B_T(\omega)F_TH_T] \\ \vdots & \vdots & \vdots \\ d_T(\omega)C_{m_2T}d_T'(\omega) & \mathrm{tr}[F_T'B_T'(\omega)C_{m_2T}^sB_T(\omega)H_T] & \mathrm{tr}[F_T'B_T'(\omega)C_{m_2T}B_T(\omega)F_TH_T] \end{pmatrix}$$

has full column rank for any ω *in its parameter space.*

Assumption 3.5. *When* $(\alpha_0', \beta_0')' \neq 0$*,* $\mathrm{tr}(C_{jT}^sB_{\omega T}B_T^{-1}) \neq 0$ *for some* $1 \leq j \leq m_2$*; when* $(\alpha_0', \beta_0')' = 0$*,*

$$\begin{pmatrix} \mathrm{tr}(C_{1T}^sB_TF_TB_T^{-1}) & \mathrm{tr}(C_{1T}^sB_{\omega T}B_T^{-1}) \\ \vdots & \vdots \\ \mathrm{tr}(C_{m_2T}^sB_TF_TB_T^{-1}) & \mathrm{tr}(C_{m_2T}^sB_{\omega T}B_T^{-1}) \end{pmatrix}$$

has full column rank.

Assumption 3.6. $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}_{nT}^{*\prime}\mathbf{X}_{nT}^*$ *has full rank.*

Assumption 3.7. *The parameter space* Θ *of* θ *is compact,* $\omega > 1$*, and* θ_0 *in the interior of* Θ*.*

Assumption 3.1 states the prediction equation of Δy_{i1} using exogenous variables. In Assumption 3.2, X_{nt} and Z_n are assumed to be nonstochastic for simplicity, and the existence of empirical moments of order $2 + \eta$ is the requirement of a proper central limit theorem. Assumption 3.3 is a sufficient condition for the nonsingularity of the limiting variance of the moment vector. It requires $\mathbf{X}_{nT}^*$ to have full column rank for large enough n and K_{jT}'s to be linearly independent, in addition to the linear independence of C_{jT}'s. Assumptions 3.4 and 3.6 are sufficient identification conditions for, respectively, the GMM estimator and the QML estimator $\hat{\theta}_{qml}$ that maximizes the log likelihood function (3.1). Under Assumption 3.5, the expected gradient matrix $G_{nT} = \mathrm{E}(\frac{\partial g_{nT}(\theta_{30})}{\partial \theta_3'})$ has full column rank for large enough n. Assumption 3.7 is a usual assumption on the parameter space.

Let $\hat{\theta}_{3,qml}$ be the QML estimator of θ_3, which is a subvector of $\hat{\theta}_{qml}$ corresponding to θ_3, $\hat{\theta}^*_{3,gmm}$ be the optimal GMM estimator with the moment vector $g^*_{nT}(\theta_3)$ in (3.4), $G^*_{nT} = \mathrm{E}(\frac{\partial g^*_{nT}(\theta_{30})}{\partial \theta_3'})$ and $\Sigma^*_{nT} = \mathrm{Var}[\sqrt{n} g^*_{nT}(\theta_{30})]$.

Theorem 3. *Suppose that Assumptions 2.1–2.2 and 3.1–3.7 are satisfied.*

(i) The optimal GMM estimator $\hat{\theta}_{3,gmm}$ *in* (3.3) *is consistent and has the asymptotic distribution* $\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30}) \xrightarrow{d} N(0, \lim_{n\to\infty}(G'_{nT}\Sigma^{-1}_{nT}G_{nT})^{-1})$*, where*

$$G_{nT} = -\frac{1}{n}\begin{pmatrix} \mathbf{X}^{*\prime}_{nT}(K_{1T}\otimes I_n)\mathbf{X}^*_{nT} & \mathbf{X}^{*\prime}_{nT}(K_{1T}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} & 0 \\ \vdots & \vdots & \vdots \\ \mathbf{X}^{*\prime}_{nT}(K_{m_1T}\otimes I_n)\mathbf{X}^*_{nT} & \mathbf{X}^{*\prime}_{nT}(K_{m_1T}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} & 0 \\ 0 & n\sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}B_TF_TB_T^{-1}) & -n\sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}B_{\omega T}B_T^{-1}) \\ \vdots & \vdots & \vdots \\ 0 & n\sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}B_TF_TB_T^{-1}) & -n\sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}B_{\omega T}B_T^{-1}) \end{pmatrix}.$$

(ii) The QML estimator $\hat{\theta}_{qml}$ *is consistent and follows the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{qml} - \theta_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}\Gamma_{nT,\theta}\big),$$

where $\Gamma_{nT,\theta} = [\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})]^{-1}\,\mathrm{E}(\frac{1}{n}\frac{\partial \ln L_w(\theta_0)}{\partial\theta}\frac{\partial \ln L_w(\theta_0)}{\partial\theta'})[\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})]^{-1}$ *with*

$$\mathrm{E}\Big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}\Big) = \begin{pmatrix} \frac{1}{n\sigma^2_{v0}}\mathbf{X}^{*\prime}_{nT}(H_T^{-1}\otimes I_n)\mathbf{X}^*_{nT} & * & * & * \\ \frac{1}{n\sigma^2_{v0}}\binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F_T'H_T^{-1}\otimes I_n)\mathbf{X}^*_{nT} & \Omega_{nT,22} & * & * \\ 0 & \mathrm{tr}(F_T'H_T^{-1}J_T) & \frac{1}{2}\,\mathrm{tr}(H_T^{-1}J_TH_T^{-1}J_T) & * \\ 0 & 0 & \frac{1}{2\sigma^2_{v0}}\,\mathrm{tr}(H_T^{-1}J_T) & \frac{T}{2\sigma^4_{v0}} \end{pmatrix},$$

and $\Omega_{nT,22} = \frac{1}{n\sigma^2_{v0}}\binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F_T'H_T^{-1}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} + \mathrm{tr}(F_T'H_T^{-1}F_TH_T)$.

(iii) $\hat{\theta}^*_{3,gmm}$ *is asymptotically efficient relative to* $\hat{\theta}_{3,qml}$ *in general, i.e.,* $(G^{*\prime}_{nT}\Sigma^{*-1}_{nT}G^*_{nT})^{-1} \leq \Gamma_{nT,\theta_3}$*, where* Γ_{nT,θ_3} *is the asymptotic variance of* $\hat{\theta}_{3,qml}$*, which is a submatrix of* $\Gamma_{nT,\theta}$ *corresponding to* θ_3.

(iv) If $\mathbf{V}_{n,T+1} \sim N(0, \sigma^2_{v0}I_{n(T+1)})$*, then:*

(a) *Among optimal GMM estimators with moments of the form (3.2), $\hat{\theta}^*_{3,gmm}$ has the minimum asymptotic variance, i.e., $(G^{*\prime}_{nT}\Sigma^{*-1}_{nT}G^*_{nT})^{-1} \leq (G'_{nT}\Sigma^{-1}_{nT}G_{nT})^{-1}$, where*

$$G^{*\prime}_T\Sigma^{*-1}_T G^*_T = \begin{pmatrix} \frac{1}{n\sigma^2_{v0}}\mathbf{X}^{*\prime}_{nT}(H^{-1}_T \otimes I_n)\mathbf{X}^*_{nT} & * & * \\ \frac{1}{n\sigma^2_{v0}}\binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F'_T H^{-1}_T \otimes I_n)\mathbf{X}^*_{nT} & \Omega_{nT,22} & * \\ 0 & \mathrm{tr}(F'_T H^{-1}_T J_T) & \frac{1}{2}\mathrm{tr}(H^{-1}_T J_T H^{-1}_T J_T) - \frac{1}{2T}\mathrm{tr}^2(H^{-1}_T J_T) \end{pmatrix}.$$

(b) *$\hat{\theta}^*_{3,gmm}$ has the same asymptotic variance as that of $\hat{\theta}_{3,qml}$, i.e., $(G^{*\prime}_{nT}\Sigma^{*-1}_{nT}G^*_{nT})^{-1} = \Gamma_{nT,\theta_3}$.*

3.2 SGMM

We may consider an efficient SGMM estimator of $\delta_0 = [\beta'_0, \gamma_0]'$ based on the efficient GMM. Let $\tilde{\delta} = [\tilde{\beta}', \tilde{\gamma}]'$ and $\tilde{\alpha}_1$ be, respectively, $\sqrt{n}$-consistent estimators of δ_0 and $\alpha_{10} = [\alpha'_0, \omega_0]'$. Denote the moment vector by $g_{nT}(\delta, \alpha_1) = [g'_{nT,1}(\delta, \alpha_1), g'_{nT,2}(\delta, \alpha_1)]'$, where

$$g_{nT,1}(\delta, \alpha_1) = \frac{1}{n}\begin{pmatrix} \Delta\mathbf{X}'_{nT}(\tilde{H}^{-1}_T \otimes I_n)e_{nT}(\alpha, \delta) \\ \mathbf{X}^{*\prime}_{nT}(\tilde{F}'_T\tilde{H}^{-1}_T \otimes I_n)e_{nT}(\alpha, \delta) \\ e'_{nT}(\alpha, \delta)(\tilde{F}'_T H^{-1}_T(\omega) \otimes I_n)e_{nT}(\alpha, \delta) \end{pmatrix},$$

$$g_{nT,2}(\delta, \alpha_1) = \frac{1}{n}\begin{pmatrix} \Upsilon'_{nT}(\tilde{H}^{-1}_T \otimes I_n)e_{nT}(\alpha, \delta) \\ e'_{nT}(\alpha, \delta)(\Phi_T(\omega) \otimes I_n)e_{nT}(\alpha, \delta) \end{pmatrix},$$

with $\alpha_1 = [\alpha', \omega]'$, $\tilde{F}_T = F_T(\tilde{\gamma})$, $\tilde{H}_T = H_T(\tilde{\omega})$ and $\Phi_T(\omega)$ is defined below (2.15c), which will be asymptotically efficient under normality. In $g_{nT}(\delta, \alpha_1)$, the estimates $\tilde{\gamma}$ and $\tilde{\omega}$ in $\tilde{F}_T$ and $\tilde{H}_T$ do not affect the asymptotic distribution of the SGMM estimator, so we use $\tilde{F}_T$ and $\tilde{H}_T$ directly. Let $\Sigma_{nT,\delta} = \mathrm{Var}[\sqrt{n}g_{nT,b}(\delta_0, \alpha_{10})]$ and $\hat{\Sigma}_{nT,\delta}$ be a consistent estimator of $\lim_{n\to\infty}\Sigma_{nT,\delta}$, where $g_{nT,b}(\delta, \alpha_1)$ is the moment vector obtained by replacing $\tilde{F}_T$ and $\tilde{H}_T$ in $g_{nT}(\delta, \alpha_1)$ with, respectively, $F_T(\gamma)$ and $H_T(\omega)$. Denote $\hat{C}_{nT,\delta} = \frac{\partial g_{nT,1}(\tilde{\delta}, \tilde{\alpha}_1)}{\partial \alpha'_1}(\frac{\partial g_{nT,2}(\tilde{\delta}, \tilde{\alpha}_1)}{\partial \alpha'_1})^{-1}$ and $\hat{R}_{nT,\delta} = [I, -\hat{C}_{nT,\delta}]$, where I is an identity matrix conformable with $g_{nT,1}(\delta, \alpha_1)$. Then we have the SGMM estimator of δ_0:

$$\hat{\delta} = \arg\min_{\delta}[\hat{R}_{nT,\delta}g_{nT}(\delta, \tilde{\alpha}_1)]'(\hat{R}_{nT,\delta}\hat{\Sigma}_{nT,\delta}\hat{R}'_{nT,\delta})^{-1}\hat{R}_{nT,\delta}g_{nT}(\delta, \tilde{\alpha}_1). \tag{3.5}$$

If we would like to focus on the estimation of only γ, which has a closed form solution of the estimate and can also be asymptotically efficient under normal disturbances, we may concentrate out α, β and σ^2_v from the QML first order conditions to derive the following two moment conditions:[25]

$$g_{nT,1}(\gamma, \omega) = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H^{-1}_T(\omega) \otimes I_n)M^*_{nT}(\omega)(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1}),$$
$$g_{nT,2}(\gamma, \omega) = \frac{1}{n}(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1})'M^{*\prime}_{nT}(\omega)[\Phi_T(\omega) \otimes I_n]M^*_{nT}(\omega)(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1}),$$

[25] As in Section 2, we can also directly follow the approach in Jin and Lee (2018) to construct an SGMM estimator of γ using moment conditions derived from the QML first order conditions. On the other hand, we do not use $g_{nT}(\delta, \alpha_1)$ to construct an SGMM estimator of only γ due to an identification issue. As shown below, by using the concentrated moments derived from the QML first order conditions, we can have closed-form roots of γ and investigate which root is consistent.

where $M^*_{nT}(\omega) = I_{nT} - \mathbf{X}^*_{nT}[\mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)\mathbf{X}^*_{nT}]^{-1}\mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)$. Denote $\hat{C}_{nT,\gamma c} = \frac{\partial g_{nT,1}(\tilde{\gamma},\tilde{\omega})}{\partial \omega}(\frac{\partial g_{nT,2}(\tilde{\gamma},\tilde{\omega})}{\partial \omega})^{-1}$ and $\tilde{M}^*_{nT} = M^*_{nT}(\tilde{\omega})$. The SGMM estimator $\dot{\gamma}$ is characterized by the quadratic equation of $\dot{\gamma}$:

$$g_{nT,1}(\dot{\gamma},\tilde{\omega}) - \hat{C}_{nT,\gamma c} g_{nT,2}(\dot{\gamma},\tilde{\omega}) = -s_{nT,1}\dot{\gamma}^2 - s_{nT,2}\dot{\gamma} - s_{nT,3} = 0, \tag{3.6}$$

where $s_{nT,1} = \frac{1}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T \otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{n,T-1}$,

$$s_{nT,2} = \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{n,T-1} - \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{nT},$$

and $s_{nT,3} = \frac{1}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{nT}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{nT} - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{nT}$. Using $\Delta\mathbf{Y}_{nT} = \gamma_0\Delta\mathbf{Y}_{n,T-1} + \Delta\mathbf{X}_{nT}\beta_0 + \Upsilon_{nT}\alpha_0 + e_{nT}$, we have $s_{nT,2} = -s_{nT,4} - 2\gamma_0 s_{nT,1}$, where

$$s_{nT,4} = \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T\otimes I_n)\tilde{M}^*_{nT}e_{nT} - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{n,T-1},$$

and $s_{nT,3} = \gamma_0 s_{nT,4} + \gamma_0^2 s_{nT,1} + o_p(1)$ as $\frac{1}{n}e'_{nT}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T\otimes I_n)\tilde{M}^*_{nT}e_{nT} = o_p(1)$ and $\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\tilde{M}^*_{nT}e_{nT} = o_p(1)$. The quadratic equation has the solutions

$$\frac{-s_{nT,2} \pm \sqrt{s^2_{nT,2} - 4s_{nT,1}s_{nT,3}}}{2s_{nT,1}} = \gamma_0 + \frac{s_{nT,4} \pm \sqrt{s^2_{nT,4} + o_p(1)}}{2s_{nT,1}}.$$

Thus, the consistent root is $\frac{-s_{nT,2}-\sqrt{s^2_{nT,2}-4s_{nT,1}s_{nT,3}}}{2s_{nT,1}}$ if $s_{nT,4} \geq 0$, or $\frac{-s_{nT,2}+\sqrt{s^2_{nT,2}-4s_{nT,1}s_{nT,3}}}{2s_{nT,1}}$ if $s_{nT,4} < 0$. The $s_{nT,4}$ can be estimated by

$$\tilde{s}_{nT,4} = \frac{2}{n}\hat{C}_{nT,\gamma c}\Delta\mathbf{Y}'_{n,T-1}\tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T\otimes I_n)\tilde{M}^*_{nT}(\Delta\mathbf{Y}_{nT} - \tilde{\gamma}\Delta\mathbf{Y}_{n,T-1}) - \frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(\tilde{H}_T^{-1}\otimes I_n)\tilde{M}^*_{nT}\Delta\mathbf{Y}_{n,T-1}.$$

Using $\mathbf{Y}_{n,T-1} = (F_T\otimes I_n)(e_{nT} + \mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0})$, we may show that $[\sqrt{n}g_{nT,1}(\gamma_0,\omega_0), \sqrt{n}g_{nT,2}(\gamma_0,\omega_0)]' = \sqrt{n}g_{nT}(\gamma_0,\omega_0) + o_p(1)$, where

$$g_{nT}(\gamma_0,\omega_0) = \frac{1}{n}\begin{pmatrix} e'_{nT}(F'_T H_T^{-1}\otimes I_n)e_{nT} + \binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F'_T H_T^{-1}\otimes I_n)M^*_{nT}e_{nT} \\ e'_{nT}(\Phi_T\otimes I_n)e_{nT}\end{pmatrix}$$

with $M^*_{nT} = M^*_{nT}(\omega_0)$. Denote $\Sigma_{nT,\gamma c} = \mathrm{Var}[\sqrt{n}g_{nT}(\gamma_0,\omega_0)]$.

The following assumptions are needed for the SGMM estimators.

Assumption 3.8. $\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\Delta\mathbf{X}'_{nT}(H_T^{-1}\otimes I_n)M_{nT}[\Delta\mathbf{X}_{nT},(F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0}] \\ \mathbf{X}^{*\prime}_{nT}(F'_T H_T^{-1}\otimes I_n)M_{nT}[\Delta\mathbf{X}_{nT},(F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0}]\end{pmatrix}$ *has full column rank.*

Assumption 3.9. $\lim_{n\to\infty}\frac{1}{n}\binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F'_T\otimes I_n)M^{*\prime}_{nT}(\Phi_T\otimes I_n)M^*_{nT}(F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} + \sigma^2_{v0}\,\mathrm{tr}(F'_T\Phi_T F_T H_T) \neq 0.$

Assumptions 3.8 is a sufficient identification condition for $\hat{\delta}$. It requires $[\alpha'_0, \beta'_0]' \neq 0$. Under Assumption 3.9, (3.6) is quadratic in γ in the limit.

Theorem 4. *Suppose that Assumptions 2.1–2.2 and 3.1–3.7 are satisfied.*

(i) If Assumption 3.8 also holds, the SGMM estimation $\hat{\delta}$ in (3.5) is consistent and has the asymptotic distribution

$$\sqrt{n}(\hat{\delta}-\delta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}[G'_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}G_{nT,\delta}]^{-1}),$$

where

$$G_{nT,\delta} = -\frac{1}{n}\begin{pmatrix} \Delta\mathbf{X}'_{nT}(H_T^{-1}\otimes I_n)\Delta\mathbf{X}_{nT} & \Delta\mathbf{X}'_{nT}(H_T^{-1}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} \\ \mathbf{X}^{*\prime}_{nT}(F'_TH_T^{-1}\otimes I_n)\Delta\mathbf{X}_{nT} & \mathbf{X}^{*\prime}_{nT}(F'_TH_T^{-1}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} \\ 0 & n\sigma^2_{v0}\operatorname{tr}(F'_TH_T^{-1}F_TH_T) \\ \Upsilon'_{nT}(H_T^{-1}\otimes I_n)\Delta\mathbf{X}_{nT} & \Upsilon'_{nT}(H_T^{-1}F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} \\ 0 & 2n\sigma^2_{v0}\operatorname{tr}(F'_TH_T^{-1}J_T) \end{pmatrix},$$

and $R_{nT,\delta} = [I, -C_{nT,\delta}]$ with

$$C_{nT,\delta} = \begin{pmatrix} \Delta\mathbf{X}'_{nT}(H_T^{-1}\otimes I_n)\Upsilon_{nT}[\Upsilon'_{nT}(H_T^{-1}\otimes I_n)\Upsilon_{nT}]^{-1} & 0 \\ \mathbf{X}^{*\prime}_{nT}(F'_TH_T^{-1}\otimes I_n)\Upsilon_{nT}[\Upsilon'_{nT}(H_T^{-1}\otimes I_n)\Upsilon_{nT}]^{-1} & 0 \\ 0 & \frac{\operatorname{tr}(F'_TH_T^{-1}J_T)}{\operatorname{tr}(H_T^{-1}J_TH_T^{-1}J_T)-\frac{1}{T}\operatorname{tr}^2(H_T^{-1}J_T)} \end{pmatrix}.$$

The asymptotic distribution of $\hat{\delta}$ is the same as that of the GMM estimator $\hat{\delta}_{gmm}$ in Theorem 3(i) with the moment vector (3.4).

(ii) If Assumption 3.9 also holds, the SGMM estimator $\dot{\gamma}$ is consistent and follows the asymptotic distribution

$$\sqrt{n}(\dot{\gamma}-\gamma_0) \xrightarrow{d} N(0, \lim_{n\to\infty}(R_{nT,\gamma c}G_{nT,\gamma c})^{-2}R_{nT,\gamma c}\Sigma_{nT,\gamma c}R'_{nT,\gamma c}),$$

where $R_{nT,\gamma c} = [1, -\frac{\operatorname{tr}(F'_TH_T^{-1}J_T)}{\operatorname{tr}(H_T^{-1}J_TH_T^{-1}J_T)-\frac{1}{T}\operatorname{tr}^2(H_T^{-1}J_T)}]$ and

$$G_{nT,\gamma c} = -[\sigma^2_{v0}\operatorname{tr}(F'_TH_T^{-1}F_TH_T)+\tfrac{1}{n}\binom{\alpha_0}{\beta_0}'\mathbf{X}^{*\prime}_{nT}(F'_TH_T^{-1}\otimes I_n)M^*_{nT}(F_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0}, 2\sigma^2_{v0}\operatorname{tr}(F'_TH_T^{-1}J_T)]'.$$

The asymptotic distribution of $\dot{\gamma}$ is the same as that of the QML estimator $\hat{\gamma}_{qml}$ of γ in Theorem 3(ii).

4 Stationary fixed effects DPD

In this section, we consider the pure stationary fixed effects DPD model where the process has started a long time ago. In this situation, $\omega = \frac{2}{1+\gamma}$ will no longer be a free parameter.[26] Another model of interest is the stationary fixed effects DPD model with exogenous variables. However, to approximate the unobservable past exogenous variables, we have the need to introduce a prediction error, and hence ω would become a free parameter. Under such a situation, the efficient GMM estimation in Section 3.1 would apply.

[26] κ no longer exists, and the mean of ΔY_{n1} is zero.

For the pure stationary fixed effects DPD model, the estimation equation (2.2) has now the $T \times T$ variance matrix

$$H_T(\gamma) = \begin{pmatrix} \frac{2}{1+\gamma} & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix} \tag{4.1}$$

and $H_T = H_T(\gamma_0)$. For the QML estimation, the quasi log likelihood function for (2.2) with sample observations ΔY_{nt}, $t = 1, \ldots, T$, is

$$\ln L_w(\gamma, \sigma_v^2) = -\frac{nT}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|H_T(\gamma)| - \frac{1}{2\sigma_v^2} e'_{nT}(\gamma)(H_T^{-1}(\gamma) \otimes I_n) e_{nT}(\gamma), \tag{4.2}$$

where $e_{nT}(\gamma) = (\Delta Y'_{n1}, \Delta Y'_{n2} - \gamma \Delta Y'_{n1}, ..., \Delta Y'_{nT} - \gamma \Delta Y'_{n,T-1})'$. The first order derivatives of (4.2) are

$$\begin{aligned} \frac{\partial \ln L_w(\gamma, \sigma_v^2)}{\partial \gamma} &= \frac{1}{\sigma_v^2} \mathbf{\Delta Y}'_{n,T-1}(H_T^{-1}(\gamma) \otimes I_n) e_{nT}(\gamma) \\ &\quad + \frac{1}{2\sigma_v^2} e'_{nT}(\gamma)\big(H_T^{-1}(\gamma)\frac{\partial H_T(\gamma)}{\partial \gamma} H_T^{-1}(\gamma) \otimes I_n\big) e_{nT}(\gamma) - \frac{n}{2}\operatorname{tr}\big(\frac{\partial H_T(\gamma)}{\partial \gamma} H_T^{-1}(\gamma)\big), \\ \frac{\partial \ln L_w(\gamma, \sigma_v^2)}{\partial \sigma_v^2} &= -\frac{nT}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} e'_{nT}(\gamma)(H_T^{-1}(\gamma) \otimes I_n) e_{nT}(\gamma), \end{aligned} \tag{4.3}$$

where $\mathbf{\Delta Y}_{n,T-1} \equiv (0, \Delta Y'_{n1}, \Delta Y'_{n2}, \ldots, \Delta Y'_{n,T-1})'$. The QMLE $\hat{\theta} = (\hat{\gamma}, \hat{\sigma}_v^2)'$ will satisfy the gradient vector of the quasi log likelihood in (4.3) being set to zero. Given $\hat{\gamma}$, we have $\hat{\sigma}_v^2(\hat{\gamma}) = \frac{1}{nT} e'_{nT}(\hat{\gamma})(H_T^{-1}(\hat{\gamma}) \otimes I_n) e_{nT}(\hat{\gamma})$. For $\hat{\gamma}$, it is characterized by the score equation:

$$\begin{aligned} &\mathbf{\Delta Y}'_{n,T-1} \cdot (H_T^{-1}(\hat{\gamma}) \otimes I_n) \cdot e_{nT}(\hat{\gamma}) \\ &+ \frac{1}{2} e'_{nT}(\hat{\gamma})\big(H_T^{-1}(\hat{\gamma})\frac{\partial H_T(\hat{\gamma})}{\partial \gamma} H_T^{-1}(\hat{\gamma}) \otimes I_n\big) e_{nT}(\hat{\gamma}) - \frac{tr\big(\frac{\partial H_T(\hat{\gamma})}{\partial \gamma} H_T^{-1}(\hat{\gamma})\big)}{2T} e'_{nT}(\hat{\gamma})(H_T^{-1}(\hat{\gamma}) \otimes I_n) e_{nT}(\hat{\gamma}) = 0. \end{aligned} \tag{4.4}$$

From the score vector, the model implies two moment conditions:

$$\mathrm{E}[\mathbf{\Delta Y}'_{n,T-1}(H_T^{-1} \otimes I_n) e_{nT}] = \mathrm{E}[e'_{nT}(F'_T \otimes I_n)(H_T^{-1} \otimes I_n) e_{nT}] = 0$$

and

$$\mathrm{E}\Big\{e'_{nT}\big[H_T^{-1}\frac{\partial H_T}{\partial \gamma} H_T^{-1} \otimes I_n - \frac{\operatorname{tr}(\frac{\partial H_T}{\partial \gamma} H_T^{-1})}{T} \cdot H_T^{-1} \otimes I_n\big] e_{nT}\Big\} = 0.$$

Therefore, these suggest two empirical moments $\mathbf{\Delta Y}'_{n,T-1}(H_T^{-1}(\gamma) \otimes I_n) e_{nT}(\gamma)$ and

$$e'_{nT}(\gamma)\Big(H_T^{-1}(\gamma)\big(\frac{\partial H_T(\gamma)}{\partial \gamma} H_T^{-1}(\gamma) - \frac{\operatorname{tr}(\frac{\partial H_T(\gamma)}{\partial \gamma} H_T^{-1}(\gamma))}{T} I_T\big) \otimes I_n\Big) e_{nT}(\gamma)$$

for a GMM estimation. A direct GMM approach is to use these two moment conditions to implement a GMM estimation.

Alternatively, we would like a possible GMM framework which can incorporate both IVs and MD estimation approaches in the literature as members in it, and also an efficient GMM estimation might be possible. Because ω is now a function of γ but not a free parameter, the previous sequential GMM approach by regarding ω as if it was a free parameter is possible but it would not retain the constraint that ω is a function of γ in the second step estimation. Therefore, such a sequential GMM approach could not achieve asymptotic efficiency. In order to achieve possible asymptotic efficiency, we suggest an alternative GMM estimation approach below. However, as ω is a nonlinear function of γ, nonlinear moments in γ seem not to be avoidable.

We recognize that by using the identity $\Delta\mathbf{Y}_{n,T-1} = (F_T(\gamma) \otimes I_n)e_{nT}(\gamma)$, and denoting $B_T = D^{-1/2}A$ from (2.8) so that $H_T^{-1} = B_T'B_T$, the two corresponding empirical moments due to scores can be written as

$$e_{nT}'(\gamma)(B_T'(\gamma) \otimes I_n)\left[(B_T'^{-1}(\gamma)F_T'(\gamma)B_T'(\gamma) \otimes I_n)\right](B_T(\gamma) \otimes I_n)e_{nT}(\gamma) \tag{4.5}$$

and

$$e_{nT}'(\gamma)(B_T'(\gamma) \otimes I_n)\Big[B_T(\gamma)\frac{\partial H_T(\gamma)}{\partial \gamma}B_T'(\gamma) \otimes I_n - \frac{\mathrm{tr}(\frac{\partial H_T(\gamma)}{\partial \gamma}H_T^{-1}(\gamma))}{T}I_T \otimes I_n\Big](B_T(\gamma) \otimes I_n)e_{nT}(\gamma). \tag{4.6}$$

As γ_0 is in B_T, these empirical moments suggest a class of GMM estimation with moments of the form

$$e_{nT}'(\gamma)(B_T'(\gamma) \otimes I_n)(A_{jT} \otimes I_n)(B_T(\gamma) \otimes I_n)e_{nT}(\gamma), \tag{4.7}$$

where A_{jT}'s can be constant matrices or matrices involving γ, but at the true γ_0, they have the property $\mathrm{tr}(A_{jT}) = 0$. Those matrices A_{jT}'s with their traces being zero will guarantee that the moment $\mathrm{E}[e_{nT}'(B_T' \otimes I_n)(A_{jT} \otimes I_n)(B_T \otimes I_n)e_{nT}] = 0$.

Assume that we have m_1 such A_{jT} for $j = 1, ..., m_1$. Then, the vector of moment conditions is

$$g_{nT}(\gamma) = \frac{1}{n}\begin{pmatrix} e_{nT}'(\gamma)(B_T'(\gamma)A_{1T}B_T(\gamma) \otimes I_n)e_{nT}(\gamma) \\ \vdots \\ e_{nT}'(\gamma)(B_T'(\gamma)A_{m_1T}B_T(\gamma) \otimes I_n)e_{nT}(\gamma)\end{pmatrix}, \tag{4.8}$$

and the optimal GMM estimator with $g_{nT}(\gamma)$ is

$$\hat{\gamma} = \arg\min_{\gamma} g_{nT}'(\gamma)\hat{\Sigma}_{nT}^{-1}g_{nT}(\gamma), \tag{4.9}$$

where $\hat{\Sigma}_{nT}$ is a consistent estimator of the limiting variance matrix of $\sqrt{n}g_{nT}(\gamma_0)$. Due to the nature of the score vector which is correctly specified under normal disturbances and the GMM moments are motivated from those scores, the best moment vector under normal disturbances is

$$g_{nT}^*(\gamma) = \frac{1}{n}\begin{pmatrix} e_{nT}'(\gamma)(B_T'(\gamma)A_{1T}^*B_T(\gamma) \otimes I_n)e_{nT}(\gamma) \\ e_{nT}'(\gamma)(B_T'(\gamma)A_{2T}^*B_T(\gamma) \otimes I_n)e_{nT}(\gamma)\end{pmatrix}, \tag{4.10}$$

where $A_{1T}^* = B_T F_T B_T^{-1}$ and $A_{2T}^* = B_{\gamma T} B_T^{-1} - \frac{\operatorname{tr}(B_{\gamma T} B_T^{-1})}{T} I_T$ with $B_{\gamma T} = \frac{\partial B_T(\gamma_0)}{\partial \gamma}$. Alternatively, it can be simply $\frac{1}{n} e_{nT}'(\gamma)(B_T'(\gamma) K_{1T} B_T(\gamma) \otimes I_n) e_{nT}(\gamma)$, where $K_{1T} = K_T^s - \frac{\operatorname{tr}(K_T^s)}{T} I_T$ with $K_T = B_T F_T B_T^{-1} - B_{\gamma T} B_T^{-1}$. While using A_{1T}^* and B_{2T}^* separately yields a GMM estimate that is asymptotically as efficient as the ML estimate in the case of normal disturbances, it may generate a relatively efficient GMM estimate due to optimal weighting than the QML estimate in the case of non-normal disturbances.

We maintain the following assumptions for the GMM estimation.

Assumption 4.1. *The process $\{y_{it}\}$ has started from the infinite past and $|\gamma_0| < 1$.*

Assumption 4.2. *A_{jT}'s have zero traces and are linearly independent.*

Assumption 4.3. *For any $\gamma \neq \gamma_0$,*

$$\frac{2}{(1+\gamma_0)(1+\gamma)} \begin{pmatrix} d_T(\gamma) A_{1T} d_T'(\gamma) \\ \vdots \\ d_T(\gamma) A_{m_1 T} d_T'(\gamma) \end{pmatrix} + (\gamma_0 - \gamma) \begin{pmatrix} \operatorname{tr}[F_T' B_T'(\gamma) A_{1T}^s B_T(\gamma) H_T] \\ \vdots \\ \operatorname{tr}[F_T' B_T'(\gamma) A_{m_1 T}^s B_T(\gamma) H_T] \end{pmatrix} + (\gamma_0 - \gamma)^2 \begin{pmatrix} \operatorname{tr}[F_T' B_T'(\gamma) A_{1T} B_T(\gamma) F_T H_T] \\ \vdots \\ \operatorname{tr}[F_T' B_T'(\gamma) A_{m_1 T} B_T(\gamma) F_T H_T] \end{pmatrix} \neq 0,$$

where $d_T(\gamma) = [(a_0(\gamma) a_1(\gamma))^{-1/2}, (a_1(\gamma) a_2(\gamma))^{-1/2}, \ldots, (a_{T-1}(\gamma) a_T(\gamma))^{-1/2}]$ with $a_t(\gamma) = 1 + t(\frac{2}{1+\gamma} - 1)$.

Assumption 4.4. $\operatorname{tr}(A_j^s K_T^s) \neq 0$ *for at least one $1 \leq j \leq m_1$.*

Assumption 4.5. *γ_0 is in the interior of the compact parameter space of γ.*

Since we study in this section the stationary case that the process $\{y_{it}\}$ has started from the infinite past, the condition $|\gamma_0| < 1$ in Assumption 4.1 is needed. Assumption 4.2 ensures the nonsingularity of the variance matrix of the moment vector. Assumption 4.3 is a sufficient identification condition. Under Assumption 4.4, the expected gradient $G_T = \operatorname{E}(\frac{\partial g_{nT}(\gamma_0)}{\partial \gamma})$ is a nonzero vector. Assumption 4.5 is a familiar condition on the parameter space.

Let $\hat{\theta}_{qml}$ be the QML estimator that maximizes (4.2), $\hat{\gamma}_{qml}$ be the first element of $\hat{\theta}_{qml}$, $\hat{\gamma}_{gmm}^*$ be the optimal GMM estimator with the moment vector $g_{nT}^*(\gamma)$, $G_T^* = \operatorname{E}(\frac{\partial g_{nT}^*(\gamma_0)}{\partial \gamma})$ and $\Sigma_T = \operatorname{Var}[\sqrt{n} g_{nT}^*(\gamma_0)]$.

Theorem 5. *Suppose that Assumptions 2.1–2.2 and 4.1–4.5 are satisfied.*

(i) The optimal GMM estimator $\hat{\gamma}$ in (4.9) is consistent and has the asymptotic distribution

$$\sqrt{n}(\hat{\gamma}_{gmm} - \gamma_0) \xrightarrow{d} N(0, (G_T' \Sigma_T^{-1} G_T)^{-1}),$$

where $\Sigma_T = \operatorname{Var}(\sqrt{n} g_{nT}(\gamma_0))$ and $G_T = \operatorname{E}(\frac{\partial g_{nT}(\gamma_0)}{\partial \gamma}) = -\sigma_{v0}^2 [\operatorname{tr}(A_{1T}^s K_T), \ldots, \operatorname{tr}(A_{m_1 T}^s K_T)]'$.

(ii) The QML estimator $\hat{\theta}_{qml}$ is consistent and has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{qml} - \theta_0) \xrightarrow{d} N(0, \Gamma_{T,\theta}),$$

where $\Gamma_{T,\theta} = [\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\gamma_0,\sigma_{v0}^2)}{\partial\theta\partial\theta'})]^{-1}\frac{1}{n}\mathrm{E}(\frac{\partial \ln L_w(\gamma_0,\sigma_{v0}^2)}{\partial\theta}\frac{\partial \ln L_w(\gamma_0,\sigma_{v0}^2)}{\partial\theta'})[\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\gamma_0,\sigma_{v0}^2)}{\partial\theta\partial\theta'})]^{-1}$ *with*

$$\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\gamma_0,\sigma_{v0}^2)}{\partial\theta\partial\theta'}) = \begin{pmatrix} 2\,\mathrm{tr}(F_T' H_T^{-1}\frac{\partial H_T}{\partial\gamma}) + \mathrm{tr}(F_T' H_T^{-1} F_T H_T) + \frac{1}{2}\mathrm{tr}(H_T^{-1}\frac{\partial H_T}{\partial\gamma}H_T^{-1}\frac{\partial H_T}{\partial\gamma}) & \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(H_T^{-1}\frac{\partial H_T}{\partial\gamma}) \\ \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(H_T^{-1}\frac{\partial H_T}{\partial\gamma}) & \frac{T}{2\sigma_{v0}^4} \end{pmatrix}.$$

(iii) $\hat\gamma_{gmm}^*$ *is asymptotically efficient relative to* $\hat\gamma_{qml}$ *in general, i.e.,* $(G_T^{*\prime}\Sigma_T^{*-1}G_T^*)^{-1} \le \Gamma_{T,\gamma}$*, where* $\Gamma_{T,\gamma}$ *is the* $(1,1)$*th element of* $\Gamma_{T,\theta}$ *in (ii).*

(iv) If $v_{it} \sim N(0,\sigma_{v0}^2)$*, then:*

(a) Among optimal GMM estimators with moments of the form (4.8), $\hat\gamma_{gmm}^*$ *has the minimum asymptotic variance, i.e.,* $(G_T^{*\prime}\Sigma_T^{*-1}G_T^*)^{-1} \le (G_T'\Sigma_T^{-1}G_T)^{-1}$*, where* $G_T^{*\prime}\Sigma_T^{*-1}G_T^* = \mathrm{tr}((K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T)K_T) = 2\,\mathrm{tr}(F_T' H_T^{-1}\frac{\partial H_T}{\partial\gamma}) + \mathrm{tr}(F_T' H_T^{-1} F_T H_T) + \frac{1}{2}\mathrm{tr}(H_T^{-1}\frac{\partial H_T}{\partial\gamma}H_T^{-1}\frac{\partial H_T}{\partial\gamma}) - \frac{1}{2T}\mathrm{tr}^2(H_T^{-1}\frac{\partial H_T}{\partial\gamma})$*.*

(b) $\hat\gamma_{gmm}^*$ *has the same asymptotic variance as that of* $\hat\gamma_{qml}$*, i.e.,* $(G_T^{*\prime}\Sigma_T^{*-1}G_T^*)^{-1} = \Gamma_{T,\gamma}$*.*

5 Monte Carlo

We investigate the performances of various GMM estimators for the dynamic panel, and compare them with least square dummy variables (LSDV) estimates, MLEs and MD estimates under different values of n and T. Samples are generated from

$$y_{it} = \gamma_0 y_{i,t-1} + c_{i0} + v_{it}, \quad t = 1, 2, \cdots, T,$$

where γ_0 takes the value of 0.5. The c_{i0} and v_{it} are generated from independent standard normal distributions. We generated the dynamic panel data with $m+T$ periods ($m = 20$) where the starting value is from $N(0, I_n)$, and then take the last T periods as our sample. By doing so, the initial value in the estimation is close to the steady state. We also use the first period of the simulated data as the initial observation in the estimation sample (so that $m = 1$ and the process is away from its steady state). We use $T = 3$, 10, and $n = 100, 300$. For each set of generated sample observations, we calculate various estimators and evaluate their biases. We do this 500 times to obtain the median bias (mb), median absolute deviation (mad), and interdecile range (idr) which is the difference between the 0.9 and 0.1 quantiles in the empirical distribution. Finite sample properties of these estimators are summarized in Tables 1–2 for each n and T.[27]

For these estimates, LSDV is the bias corrected LSDV estimate in Hahn and Kuersteiner (2002), FD-W is the IV estimate from first differenced equations using two lagged values as IVs, FD-S is the system GMM estimate in Blundell and Bond (1988), which combines the moments of first differenced equations and level

[27] Detailed Monte Carlo results for other values of γ_0 (0.2, 0.8, 0.9) are presented in the supplement file.

equations together. The ML-W is the ML estimate assuming that the process has achieved stationarity. The ML-C is the ML estimate that allows the initial period to have unrestricted mean and variance, and MD-W is the minimum distance estimate. The G-W is the efficient GMM estimate with a short history, and G-WS is the corresponding efficient GMM estimate with a long history that has achieved stationarity. The SG-W1 and SG-W2 are the two sequential GMM estimates in Section 2.3. The Rao-W is the score method estimate based on Rao (1965).

Table 1 reports the results for the case with $\gamma_0 = 0.5$ under the pure DPD, including both the DGP with $m = 20$ and $m = 1$. For the $m = 20$ case, we see that when T is short ($T = 3$), the bias of LSDV is large. The FD-W, FD-S, ML-C and MD-W have some bias, while the ML-W has small bias. The efficient GMM estimate which uses quadratic moments has small bias and small median absolute deviation overall, and GS-W that assumes a stationary process has a smaller variation. The SG-W1 and SG-W2 also have small biases, but have a larger deviation than the efficient GMM estimates under stationary process. The score method estimate Rao-W also has small biases. For the $m = 1$ case, the FD-S has a larger bias. This is consistent with the theoretical prediction because the system GMM estimate requires that initial observations are uncorrelated with the individual effects, which is satisfied if the process has started a long time ago. However, various sequential GMM estimates still perform well. Also, the biases of the ML-W, G-W and Rao-W have a larger bias when $m = 1$. Other estimates such as LSDV and FD-W have similar performances compared to the case of a long history ($m = 20$) when n or T is large. To sum up, for the case when the dynamic coefficient is moderate, the efficient and sequential GMM estimates have satisfactory performance. Compared with the FD-W in Arellano and Bond (1991) and the FD-S in Blundell and Bond (1998), the efficient and sequential GMM estimates have smaller biases, and efficient GMM estimates have a smaller deviation. Particularly, the system GMM estimates in Blundell and Bond (1998) have large biases when $m = 1$ under a small T. Also, the efficient and sequential GMM estimates have similar performance as MD estimates in Hsiao et al. (2002), except that MD estimates have a larger deviation when the DGP has a short history ($m = 1$). Compared with ML estimates, the efficient and sequential GMM estimates have similar performance on average in terms of biases. However, the deviations of sequential GMM estimates are larger than those of ML estimates, while deviations of efficient GMM estimates are similar to those of ML estimates.

We also investigate the case with exogenous variables. The DGP is

$$y_{it} = \gamma_0 y_{i,t-1} + z_i b_0 + x_{it}\beta_0 + c_{i0} + v_{it}, \quad t = 1, 2, \cdots, T,$$

where x_{it} is generated from an AR(1) process $x_{it} = 0.5x_{i,t-1} + \epsilon_{it}$. We assign $b_0 = 1$ for the intercept term in the DGP and $\beta_0 = 1$ for x_{it}. The dynamic coefficient is $\gamma_0 = 0.5$. We use the fixed effects DGP where individual effects are generated by

$$\mathbf{c}_{n0} = \bar{X}_{nT}\pi_0 + \zeta_n,$$

with standard normally distributed ζ_n and $\pi_0 = 1$. For the initial period approximation of unobserved past regressors, we use the linear approximation. Table 2 presents the result.

From Table 2, we see that the bias corrected LSDV, FD-W and MD-W have similar performance when $T = 10$, while FD-S has large bias under short T. The efficient and sequential GMM estimates have similar biases as MD estimates, but they have a smaller deviation than MD estimates. Comparing with ML estimates, the efficient and sequential GMM estimates have similar biases and deviations.

In the above DGPs, the disturbances are normally distributed. We see that the deviations of GMM estimates are larger than those of MLEs. We conduct additional simulations to investigate performance of various estimates when the disturbances are not normal. Non-normal errors are generated from the exponential distribution with mean 1, where its kurtosis is $4! = 24$. Comparing the cases with normal disturbances and non-normal disturbances,[28] we see that the performances of various estimates are similar, and the efficient and sequential GMM estimates do not have a smaller deviation than those of MLEs under non-normal disturbances.

6 Conclusion

This paper investigates various GMM estimation of short dynamic panel data models including efficient GMM and sequential GMM estimation. For the efficient GMM estimation, we make use of the score vector of the quasi maximum likelihood (QML) estimation. These GMM estimators can be as efficient as maximum likelihood estimators when disturbances are normal, and more efficient than QML estimators when disturbances are not normal. Alternative sequential GMMs based on the score moments and the efficient GMM estimation are also discussed. For the sequential GMM estimation, we focus on the estimation of parameters of interest, thus it reduces some computational burden caused by nuisance parameters. For those sequential GMM estimation of γ based on score moments, estimates with analytical expressions as root estimates are available. Monte Carlo experiments are conducted to compare various estimates for dynamic panel data in the literature, and the performances of efficient and sequential GMM estimates are satisfactory. Compared with the FD-W in Arellano and Bond (1991) and FD-S in Blundell and Bond (1998), the efficient and sequential GMM estimates have a smaller bias under the pure DPD model, and have a smaller deviation when exogenous variables are present. Compared with MD estimates, these GMM estimates have a smaller

[28] Due to space limit, we present the details of simulation results under non-normal disturbances for $\gamma_0 = 0.2, \gamma_0 = 0.5, \gamma_0 = 0.8, \gamma_0 = 0.9$ in the supplementary file.

deviation for the short history case under the pure DPD model. Compared with the ML estimates, the efficient and sequential GMM estimates have similar performance on average under different settings.

Acknowledgements

We are grateful to the Editor Esfandiar Maasoumi, Co-Editor Tong Li, and two anonymous referees for their valuable and helpful comments. Fei Jin gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71833004) and Program for Innovative Research Team of Shanghai University of Finance and Economics.

References

Alvarez, J. and M. Arellano, 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

Anderson, T.W. and C. Hsiao, 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598–606.

Arellano, M. and O. Bond, 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 O. Bover, 1995. Another look at the instrumental-variable estimation of error-components models. *Journal of Econometrics* 68, 29–51.

Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.

Blundell, R. and S. Bond, 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.

Bun, M.J.G. and J.F. Kiviet, 2006. The effects of dynamic feedbacks and LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132, 409–444.

Chao, J.C. and N.R. Swanson, 2005. Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Crepon, B., F. Kramarz, and A. Trognon, 1997. Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models. *Journal of Econometrics* 82, 135–156.

Donald, C.G. and W.K. Newey, 2001. Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Hahn, J. and G. Kuersteiner, 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Han, C. and P.C.B. Phillips, 2006. GMM with many moment conditions. *Econometrica* 74, 147–192.

Hsiao, C., 1986. *Analysis of Panel Data*. Cambridge University Press.

Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu, 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107–150.

Jin, F. and Lee, L.F., 2012. Approximated likelihood and root estimators for spatial interaction in spatial autoregressive models. *Regional Science and Urban Economics* 42, 446–458.

Jin, F. and Lee, L.F., 2018. Robust and efficient sequential GMM and GEL estimation and tests. Working paper.

Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.

Lee, L.F. and J. Yu, 2014. Efficient GMM estimation of spatial dynamic panel data models. *Journal of Econometrics* 180, 174–197.

Lee, L.F. and J. Yu, 2018. Decomposition of dynamic panel data models: within and between equations. Working paper.

Lin, Xu and L.F. Lee, 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34–52.

Maddala, G.S., 1971. Generalized least squares with an estimated variance covariance matrix. *Econometrica* 39, 24–33.

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

Rao, C.R., 1965. *Linear Statistical Inference and its Applications*. Hoboken, NJ, US: John Wiley & Sons Inc.

Trognon, A., Gouriéroux, C., 1990. A note on the efficiency of two-step estimation methods, in: Champsaur, P. (Ed.), Essays in Honor of Edmond Malinvaud 3. MIT Press, 232–248.

Yang, Z., 2018. Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics* 205, 423–447.

Table 1: Estimates under $\gamma = 0.5$

$m = 20$	LSDV	FD-W	FD-S	ML-W	ML-C	MD-W	G-W	GS-W	SG-W1	SG-W2	Rao-W
n=100,T=3											
mb	-0.217	-0.039	0.002	0.001	-0.057	-0.010	-0.026	-0.009	0.006	0.006	-0.005
mad	0.061	0.191	0.112	0.074	0.113	0.096	0.176	0.075	0.136	0.136	0.073
idr	0.241	0.683	0.407	0.278	0.495	0.370	0.935	0.294	0.561	0.563	0.282
n=100,T=10											
mb	-0.031	-0.021	0.015	-0.001	-0.316	-0.006	-0.008	-0.003	0.015	0.015	-0.003
mad	0.023	0.043	0.034	0.024	0.048	0.025	0.028	0.023	0.037	0.037	0.023
idr	0.088	0.157	0.132	0.085	0.383	0.093	0.138	0.085	0.206	0.206	0.084
n=300,T=3											
mb	-0.212	-0.017	-0.004	0.004	-0.007	-0.005	0.000	0.000	0.028	0.028	0.002
mad	0.038	0.096	0.057	0.046	0.072	0.054	0.081	0.046	0.088	0.088	0.046
idr	0.153	0.368	0.225	0.168	0.301	0.198	0.395	0.168	0.358	0.358	0.173
n=300,T=10											
mb	-0.028	-0.007	0.008	0.001	-0.290	-0.001	0.000	0.000	0.011	0.011	0.000
mad	0.015	0.024	0.019	0.015	0.042	0.016	0.015	0.015	0.020	0.020	0.015
idr	0.054	0.094	0.079	0.053	0.366	0.058	0.060	0.053	0.080	0.080	0.052
$m = 1$	LSDV	FD-W	FD-S	ML-W	ML-C	MD-W	G-W	GS-W	SG-W1	SG-W2	Rao-W
n=100,T=3											
mb	-0.155	-0.273	0.277	0.154	-0.139	-0.034	-0.149	0.100	-0.046	-0.046	0.054
mad	0.061	0.302	0.060	0.068	0.095	0.137	0.233	0.090	0.100	0.101	0.114
idr	0.248	1.224	0.227	0.304	0.408	0.533	0.858	0.340	0.552	0.551	0.485
n=100,T=10											
mb	-0.021	-0.034	0.154	0.044	-0.360	0.047	-0.006	0.040	0.011	0.011	0.042
mad	0.023	0.055	0.034	0.025	0.086	0.026	0.027	0.025	0.031	0.031	0.024
idr	0.086	0.211	0.133	0.097	0.496	0.100	0.121	0.093	0.133	0.133	0.094
n=300,T=3											
mb	-0.150	-0.093	0.304	0.153	-0.120	0.051	-0.018	0.124	-0.008	-0.008	0.119
mad	0.037	0.215	0.031	0.048	0.119	0.099	0.088	0.050	0.088	0.088	0.055
idr	0.143	0.886	0.129	0.181	0.403	0.414	0.808	0.205	0.478	0.478	0.235
n=300,T=10											
mb	-0.024	-0.009	0.162	0.044	-0.388	0.054	-0.001	0.041	0.004	0.004	0.043
mad	0.011	0.031	0.021	0.015	0.033	0.017	0.012	0.015	0.015	0.015	0.015
idr	0.049	0.120	0.080	0.058	0.457	0.062	0.054	0.057	0.060	0.060	0.058

1. mb is the median bias, md is the median absolute deviation, and idr is the interdecile range which is the difference between the 0.9 and 0.1 quantiles in the empirical distribution.
2. LSDV is the bias corrected LSDV, FD-W is an IV estimate for first differenced equations, and FD-S is the system GMM in Blundell and Bond (1988).

Table 2: Estimates under $\gamma = 0.5, \beta = 1$, FE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		MD-W:γ, β		G-W:γ, β		SG-W1:γ, β		SG-W2:γ
n=100,T=3.															
mb	-0.020	-0.006	-0.049	-0.009	0.275	0.074	-0.009	-0.004	-0.008	0.001	-0.014	-0.006	-0.001	-0.003	0.002
mad	0.048	0.054	0.082	0.055	0.050	0.060	0.046	0.053	0.057	0.054	0.052	0.053	0.049	0.054	0.057
idr	0.180	0.211	0.328	0.211	0.196	0.216	0.183	0.207	0.237	0.214	0.194	0.208	0.192	0.207	0.240
n=100,T=10															
mb	-0.001	0.002	-0.015	0.006	0.083	-0.010	-0.001	0.002	0.009	-0.004	0.001	0.002	0.007	-0.002	0.012
mad	0.014	0.022	0.016	0.020	0.018	0.022	0.014	0.021	0.024	0.025	0.014	0.021	0.015	0.022	0.020
idr	0.055	0.082	0.063	0.083	0.073	0.088	0.055	0.082	0.113	0.089	0.057	0.083	0.061	0.084	0.088
n=300,T=3															
mb	-0.015	-0.003	-0.006	0.000	0.299	0.073	0.000	-0.001	-0.001	0.000	-0.001	-0.002	0.004	-0.001	0.005
mad	0.028	0.029	0.048	0.031	0.029	0.032	0.027	0.029	0.035	0.029	0.030	0.029	0.030	0.028	0.030
idr	0.113	0.117	0.189	0.126	0.126	0.132	0.105	0.116	0.136	0.114	0.115	0.119	0.115	0.117	0.122
n=300,T=10															
mb	0.000	-0.001	-0.004	0.000	0.097	-0.019	0.001	-0.001	0.005	-0.002	0.002	-0.002	0.003	-0.003	0.003
mad	0.008	0.013	0.011	0.013	0.012	0.013	0.008	0.013	0.011	0.014	0.008	0.013	0.009	0.013	0.009
idr	0.033	0.046	0.039	0.047	0.045	0.049	0.032	0.047	0.045	0.049	0.033	0.047	0.035	0.047	0.035

1. mb is the median bias, md is the median absolute deviation, and idr is the interdecile range which is the difference between the 0.9 and 0.1 quantiles in the empirical distribution.
2. LSDV is the bias corrected LSDV, FD-W is an IV estimate for first differenced equations, and FD-S is the system GMM in Blundell and Bond (1988).