Supplement to "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

July 31, 2019

Contents

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

A Random effects DPD

A.1 The random effects pure DPD with a short past

In this subsection, we consider the random effects pure DPD model with a short past:

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

and y_{i0} is observable. The c_{i0} has mean zero and variance $\phi_{c0}\sigma_{v0}^2$, and is uncorrelated with v_{it}. For y_{i0}, by continuous substitution, we have

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

With an unknown and finite m, $y_{i,-m}$ might have some nonzero mean and unknown variance. We assume that $\mathrm{E}(y_{i0}) = \kappa_{m0}$ and $\mathrm{Var}(y_{i0}) = \varpi_{m0}\sigma_{v0}^2$, where ϖ_{m0} is a parameter free from γ_0. The variance matrix of $[y_{i0}, c_{i0} + v_{i1}, \dots, c_{i0} + v_{iT}]'$ is $\sigma_{v0}^2 \Omega_{T+1}$, where $\Omega_{T+1} = \begin{pmatrix} \varpi_{m0} & \eta_{m0} l_T' \\ \eta_{m0} l_T & \phi_{c0} l_T l_T' + I_T \end{pmatrix}$ and η_{m0} is a free parameter due to the unknown m. Then the quasi log likelihood function of the model is

$$\ln L_r(\theta) = -\frac{n(T+1)}{2} \ln(2\pi\sigma_v^2) - \frac{n}{2} \ln |\Omega_{T+1}(\theta_2)| - \frac{1}{2\sigma_v^2} \mathbf{U}_{nT}'(\theta_1)(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1), \tag{A.1}$$

where $\theta_1 = [\kappa_m, \gamma]'$, $\theta_2 = [\varpi_m, \eta_m, \phi_c]'$, $\theta = [\theta_1', \theta_2', \sigma_v^2]'$, $\Omega_{T+1}(\theta_2) = \begin{pmatrix} \varpi_m & \eta_m l_T' \\ \eta_m l_T & \phi_c l_T l_T' + I_T \end{pmatrix}$, $\mathbf{U}_{nT}(\theta_1) = \mathbf{Y}_{nT} - \mathbf{Z}_{nT}\theta_1$, $\mathbf{Y}_{nT} = [Y_{n0}', Y_{n1}', \dots, Y_{nT}']'$, and $\mathbf{Z}_{nT} = \begin{pmatrix} l_n & 0 \\ 0 & Y_{n0} \\ \vdots & \vdots \\ 0 & Y_{n,T-1} \end{pmatrix}$. Let θ_{2j} be the jth element of θ_2. The first order derivatives of $\ln L_r(\theta)$ are

$$\frac{\partial \ln L_r(\theta)}{\partial \theta_1} = \frac{1}{\sigma_v^2} \mathbf{Z}_{nT}'(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1),$$
$$\frac{\partial \ln L_r(\theta)}{\partial \theta_{2j}} = -\frac{n}{2} \operatorname{tr}\Big(\Omega_{T+1}^{-1}(\theta_2)\frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}}\Big) + \frac{1}{2\sigma_v^2} \mathbf{U}_{nT}'(\theta_1)\Big(\Omega_{T+1}^{-1}(\theta_2)\frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}}\Omega_{T+1}^{-1}(\theta_2) \otimes I_n\Big)\mathbf{U}_{nT}(\theta_1),$$
$$\frac{\partial \ln L_r(\theta)}{\partial \sigma_v^2} = -\frac{n(T+1)}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} \mathbf{U}_{nT}'(\theta_1)(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1).$$

A.1.1 Efficient GMM

For the pure random effects DPD model with a finite past, let $\mathcal{Y}_{n,T-1} = [0, Y_{n0}', \dots, Y_{n,T-1}']'$. Using $\mathcal{Y}_{n,T-1} = (F_{T+1} \otimes I_n)(\mathbf{U}_{nT} + \kappa_{m0} \iota_{n,T+1})$, the moment condition $\mathcal{Y}_{n,T-1}'(\Omega_{n,T+1}^{-1} \otimes I_n)\mathbf{U}_{nT} = \mathbf{U}_{nT}'(F_{T+1}'\Omega_{T+1}^{-1} \otimes I_n)\mathbf{U}_{nT} + \kappa_{m0}\iota_{n,T+1}'(F_{T+1}'\Omega_{T+1}^{-1} \otimes I_n)\mathbf{U}_{nT}$. Let the eigenvalue and eigenvector decomposition of Ω_{T+1}^{-1} be $\Omega_{T+1}^{-1} = A_1' A_2 A_1$, where A_2 is a diagonal matrix formed by the eigenvalues of Ω_{T+1}^{-1}. Then $\Omega_{T+1}^{-1} = \mathcal{B}_T' \mathcal{B}_T$, where $\mathcal{B}_T = A_2^{1/2} A_1$. Correspondingly, let $\Omega_{T+1}^{-1}(\theta_2) = \mathcal{B}_T'(\theta_2)\mathcal{B}_T(\theta_2)$. Thus, we may consider a GMM estimation with the following moment vector:

$$g_{nT}(\theta_1,\theta_2)=\frac{1}{n}\begin{pmatrix}\iota'_{n,T+1}(K_{1T}\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \vdots\\ \iota'_{n,T+1}(K_{m_1T}\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \vdots\\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1)\end{pmatrix}, \tag{A.2}$$

where C_{jT}'s have zero traces. Section E.3 shows how to compute the variance $\Sigma_T=\mathrm{Var}[\sqrt{n}g_{nT}(\theta_{10},\theta_{20})]$ by writing $\mathbf{U}_{nT}$ as a transformation of a vector of independent elements

$$\mathbb{V}_{nT}=[\mathbf{c}_{n0},\boldsymbol{\zeta}'_{n0},V'_{n1},\ldots,V'_{nT}]',$$

where $\mathbf{c}_{n0}=[c_{10},\ldots,c_{n0}]'$ and $\boldsymbol{\zeta}_{n0}\equiv[\zeta_{10},\ldots,\zeta_{n0}]'=\gamma_0^m[Y_{n,-m}-\mathrm{E}(Y_{n,-m})]+\sum_{j=0}^{m-1}\gamma_0^jV_{n,-j}$. Let $\hat{\Sigma}_{nT}$ be a consistent estimator of Σ_T. Then the optimal GMM estimator of $\theta_3=[\theta'_1,\theta'_2]'$ with $g_{nT}(\theta_1,\theta_2)$ is

$$\hat{\theta}_{3,gmm}=\arg\min_{\theta_3}g'_{nT}(\theta_1,\theta_2)\hat{\Sigma}_{nT}^{-1}g_{nT}(\theta_1,\theta_2). \tag{A.3}$$

Denote $\mathcal{B}_{\varpi T}=\frac{\partial\mathcal{B}_T(\theta_{20})}{\partial\varpi_m}$, $\mathcal{B}_{\eta T}=\frac{\partial\mathcal{B}_T(\theta_{20})}{\partial\eta_m}$, $\mathcal{B}_{\phi T}=\frac{\partial\mathcal{B}_T(\theta_{20})}{\partial\phi_c}$, $C^*_{1T}=\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}$, $C^*_{2T}=\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}-\frac{1}{T+1}\mathrm{tr}(\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1})I_{T+1}$, $C^*_{3T}=\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}-\frac{1}{T+1}\mathrm{tr}(\mathcal{B}_{\eta T}\mathcal{B}_T^{-1})I_{T+1}$ and $C^*_{4T}=\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}-\frac{1}{T+1}\mathrm{tr}(\mathcal{B}_{\phi T}\mathcal{B}_T^{-1})I_{T+1}$. The best moment vector under normality is

$$g^*_{nT}(\theta_1,\theta_2)=\begin{pmatrix}\iota'_{n,T+1}(\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \iota'_{n,T+1}(F'_{T+1}\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C^*_{1T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1)\\ \vdots\\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C^*_{4T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1)\end{pmatrix}, \tag{A.4}$$

which corresponds to the score vector, as expected and shown in Section D.1.

The analysis on the GMM estimation requires the following assumptions.

Assumption A.1. *The disturbances c_{i0}, $i=1,\ldots,n$, are i.i.d. $(0,\phi_{c0}\sigma_{v0}^2)$ and independent of v_{it}'s. For some $\eta>0$, $\mathrm{E}(|c_{i0}|^{4+\eta})<\infty$.*

Assumption A.2. *The process $\{y_{it}\}$ has started from m periods ago for a finite and unknown m, and $y_{i0}=\kappa_{m0}+\sum_{j=0}^{m-1}\gamma_0^jc_{i0}+\zeta_{i0}$, where ζ_{i0}'s are i.i.d with mean zero, $\mathrm{E}(|\zeta_{i0}|^{4+\eta})<\infty$ for some $\eta>0$, and ζ_{i0}'s are independent of v_{jt}'s and c_{j0}'s.*

Assumption A.3. *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 A.4. *When* $\kappa_{m0} \neq 0$, $\begin{pmatrix} (K_{1T})_{11} & (K_{1T}F_{T+1})_{11} \\ \vdots & \vdots \\ (K_{m_1T})_{11} & (K_{m_1T}F_{T+1})_{11} \end{pmatrix}$ *has full column rank, and*

$$\Lambda_1(\theta_2) = \begin{pmatrix} \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 1 & \\ & 0_{T\times T}\end{pmatrix}] & \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 0 & l'_T \\ l_T & 0_{T\times T}\end{pmatrix}] & \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 0 & \\ & l_Tl'_T\end{pmatrix}] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 1 & \\ & 0_{T\times T}\end{pmatrix}] & \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 0 & l'_T \\ l_T & 0_{T\times T}\end{pmatrix}] & \mathrm{tr}[\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2)\begin{pmatrix} 0 & \\ & l_Tl'_T\end{pmatrix}] \end{pmatrix}$$

has full column rank for any $\theta_2 \neq \theta_{20}$*; when* $\kappa_{m0} = 0$, $[(K_{1T})_{11}, \ldots, (K_{m_1T})_{11}] \neq 0$, *and* $[\Lambda_1(\theta_2), \Lambda_2(\theta_2)]$ *has full column rank for any* $\theta_2 \neq \theta_{20}$*, where*

$$\Lambda_2(\theta_2) = \begin{pmatrix} \mathrm{tr}[F'_{T+1}\mathcal{B}'_T(\theta_2)C^s_{1T}\mathcal{B}_T(\theta_2)\Omega_{T+1}] & \mathrm{tr}[F'_{T+1}\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2)F_{T+1}\Omega_{T+1}] \\ \vdots & \vdots \\ \mathrm{tr}[F'_{T+1}\mathcal{B}'_T(\theta_2)C^s_{m_2T}\mathcal{B}_T(\theta_2)\Omega_{T+1}] & \mathrm{tr}[F'_{T+1}\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2)F_{T+1}\Omega_{T+1}] \end{pmatrix}.$$

Assumption A.5. *When* $\kappa_{m0} \neq 0$,

$$\Lambda_3 = \begin{pmatrix} \mathrm{tr}(C^s_{1T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{1T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{1T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \\ \vdots & \vdots & \vdots \\ \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \end{pmatrix}$$

has full column rank; when $\kappa_{m0} = 0$,

$$\Lambda_4 = \begin{pmatrix} \mathrm{tr}(C^s_{1T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{1T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{1T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{1T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{tr}(C^s_{m_2T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \end{pmatrix}$$

has full column rank.

Assumption A.6. *The parameter space* Θ *of* θ *is compact,* $\varpi_m > 0$, $\phi_c > \frac{\eta_m^2}{\varpi_m} - \frac{1}{T}$, *and* θ_0 *is in the interior of* Θ.

Assumption A.1 summarizes conditions on individual effects c_{i0}'s in a random effects model. Assumption A.2 states the short past setting and the related conditions on y_{i0}. Assumption A.3 is a sufficient condition for the invertibility of the variance matrix of the moment vector. Assumption A.4 is a sufficient identification condition for the GMM estimation. Under Assumption A.5, the expected gradient matrix $G_T = \mathrm{E}(\frac{\partial g_{nT}(\theta_{10},\theta_{20})}{\partial \theta'_3})$ has full column rank, where

$$G_T = \begin{pmatrix} -(K_{1T})_{11} & -\kappa_{m0}(K_{1T}F_{T+1})_{11} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -(K_{m_1T})_{11} & -\kappa_{m0}(K_{m_1T}F_{T+1})_{11} & 0 & 0 & 0 \\ 0 & -\sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{1T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & -\sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\,\mathrm{tr}(C^s_{m_2T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \end{pmatrix}.$$

The conditions $\varpi_m > 0$ and $\phi_c > \frac{\eta_m^2}{\varpi_m} - \frac{1}{T}$ in Assumption A.6 ensure the positive-definiteness of $\Omega_{T+1}(\theta_2)$ so it has the decomposition $\mathcal{B}_T'(\theta_2)\mathcal{B}_T(\theta_2)$.

Let $\hat{\theta}_{qml}$ be the QML estimator that maximizes the log likelihood function (A.1), $\hat{\theta}_{3,qml}$ be a subvector of $\hat{\theta}_{qml}$ corresponding to θ_3, $\hat{\theta}_{3,gmm}^*$ be the optimal GMM estimator with the moment vector $g_{nT}^*(\theta_1, \theta_2)$ in (A.4), $G_T^* = \mathrm{E}(\frac{\partial g_{nT}^*(\theta_{10},\theta_{20})}{\partial \theta_3'})$ and $\Sigma_T^* = \mathrm{Var}[\sqrt{n} g_{nT}^*(\theta_{10}, \theta_{20})]$.

Theorem A.1. *Suppose that Assumptions 2.1–2.2 and A.1–A.6 are satisfied.*

(i) The optimal GMM estimator $\hat{\theta}_{3,gmm}$ in (A.3) *is consistent and follows the asymptotic distribution $\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30}) \xrightarrow{d} N(0, (G_T'\Sigma_T^{-1}G_T)^{-1})$.*

(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 $\Gamma_{T,\theta} = \big[\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}\big)\big]^{-1} \mathrm{E}\big(\frac{1}{n}\frac{\partial \ln L_w(\theta_0)}{\partial\theta}\frac{\partial \ln L_w(\theta_0)}{\partial\theta'}\big)\big[\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}\big)\big]^{-1}$ *with*

$$\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}\big) = \begin{pmatrix} \Pi_T + \frac{2\sigma_{v0}^4}{T+1}\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3\partial\sigma_v^2}\big)\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) & \mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3\partial\sigma_v^2}\big) \\ \mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) & \frac{T+1}{2\sigma_{v0}^4} \end{pmatrix},$$

$\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) = [0, 0, \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\kappa_m}), \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}), \frac{1}{2\sigma_{v0}^2}\mathrm{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\phi_c})]$ *and*

$$\Pi_T = \begin{pmatrix} \frac{1}{\sigma_{v0}^2}(\Omega_{T+1}^{-1})_{11} & \frac{\kappa_{m0}}{\sigma_{v0}^2}(\Omega_{T+1}^{-1}F_{T+1})_{11} & 0 & 0 & 0 \\ \frac{\kappa_{m0}}{\sigma_{v0}^2}(\Omega_{T+1}^{-1}F_{T+1})_{11} & \frac{\kappa_{m0}^2}{\sigma_{v0}^2}(F_{T+1}'\Omega_{T+1}^{-1}F_{T+1})_{11} + \mathrm{tr}(C_{1T}^{*s}C_{1T}^*) & -\mathrm{tr}(C_{1T}^{*s}C_{2T}^*) & -\mathrm{tr}(C_{1T}^{*s}C_{3T}^*) & -\mathrm{tr}(C_{1T}^{*s}C_{4T}^*) \\ 0 & -\mathrm{tr}(C_{2T}^{*s}C_{1T}^*) & \mathrm{tr}(C_{2T}^{*s}C_{2T}^*) & \mathrm{tr}(C_{2T}^{*s}C_{3T}^*) & \mathrm{tr}(C_{2T}^{*s}C_{4T}^*) \\ 0 & -\mathrm{tr}(C_{3T}^{*s}C_{1T}^*) & \mathrm{tr}(C_{3T}^{*s}C_{2T}^*) & \mathrm{tr}(C_{3T}^{*s}C_{3T}^*) & \mathrm{tr}(C_{3T}^{*s}C_{4T}^*) \\ 0 & -\mathrm{tr}(C_{4T}^{*s}C_{1T}^*) & \mathrm{tr}(C_{4T}^{*s}C_{2T}^*) & \mathrm{tr}(C_{4T}^{*s}C_{3T}^*) & \mathrm{tr}(C_{4T}^{*s}C_{4T}^*) \end{pmatrix}.$$

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

(iv) If $\mathbb{V}_{nT}$ is normally distributed, then:

(a) Among optimal GMM estimators with moments of the form (A.2), *$\hat{\theta}_{3,gmm}^*$ has the minimum asymptotic variance, i.e., $(G_T^{*\prime}\Sigma_T^{*-1}G_T^*)^{-1} \leq (G_T'\Sigma_T^{-1}G_T)^{-1}$, where $G_T^{*\prime}\Sigma_T^{*-1}G_T^* = \Pi_T$.*

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

A.1.2 SGMM

We may concentrate out both κ_m and σ_v^2 from the QML first order conditions and consider the SGMM estimation of γ_0, which yields a root estimator. By the first order derivative $\frac{\partial \ln L_r(\theta)}{\partial \kappa_m} = 0$, the estimate of κ_{m0} for given γ and θ_2 is $\kappa_m(\gamma,\theta_2) = [\iota'_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)\iota_{n,T+1}]^{-1}\iota'_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1})$. Substituting this estimate into the derivatives of $\ln L_r(\theta)$ yields the moments $g_{nT,1}(\gamma,\theta_2) = \frac{1}{n}\mathcal{Y}'_{n,T-1}(\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)M_{nT}(\theta_2)(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1})$ and $g_{nT,2}(\gamma,\theta_2)$, where

$$M_{nT}(\theta_2) = I_{n(T+1)} - \iota_{n,T+1}[\iota'_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)\iota_{n,T+1}]^{-1}\iota'_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)$$

and $g_{nT,2}(\gamma,\theta_2)$ is a 3×1 vector with the jth element being

$$\frac{1}{n}(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1})'M'_{nT}(\theta_2)\big[\Phi_{Tj}(\theta_2)\otimes I_n\big]M_{nT}(\theta_2)(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1}),$$

for $\Phi_{Tj}(\theta_2) = \Omega_{T+1}^{-1}(\theta_2)\frac{\partial\Omega_{T+1}(\theta_2)}{\partial\theta_{2j}}\Omega_{T+1}^{-1}(\theta_2) - \frac{\mathrm{tr}(\Omega_{T+1}^{-1}(\theta_2)\frac{\partial\Omega_{T+1}(\theta_2)}{\partial\theta_{2j}})}{T+1}\Omega_{T+1}^{-1}(\theta_2)$ for $j = 1,2,3$. Denote $\hat{C}_{nT,\gamma c} = \frac{\partial g_{nT,1}(\tilde{\gamma},\tilde{\theta}_2)}{\partial\theta'_2}(\frac{\partial g_{nT,2}(\tilde{\gamma},\tilde{\theta}_2)}{\partial\theta'_2})^{-1}$, $\tilde{M}_{nT} = M_{nT}(\tilde{\theta}_2)$, $\tilde{\Omega}_{T+1}^{-1} = \Omega_{T+1}^{-1}(\tilde{\theta}_2)$, and $\tilde{\Phi}_{Tj} = \Phi_{Tj}(\tilde{\theta}_2)$ for $j = 1,2,3$. The SGMM estimator $\hat{\gamma}$ of γ_0 is characterized by the quadratic equation

$$g_{nT,1}(\hat{\gamma},\tilde{\theta}_2) - \hat{C}_{nT,\gamma c}g_{nT,2}(\hat{\gamma},\tilde{\theta}_2) = -s_{nT,1}\hat{\gamma}^2 - s_{nT,2}\hat{\gamma} - s_{nT,3} = 0,$$

where

$$\begin{aligned}
s_{nT,1} &= \frac{1}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T1}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1},\ \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T2}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1},\\
&\qquad \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T3}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1}\big]',\\
s_{nT,2} &= \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1} - \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T1}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT},\\
&\qquad \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T2}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT},\ \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T3}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT}\big]',\\
s_{nT,3} &= \frac{1}{n}\hat{C}_{nT,\gamma c}\big[\mathbf{Y}'_{nT}\tilde{M}'_{nT}(\tilde{\Phi}_{T1}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT},\ \mathbf{Y}'_{nT}\tilde{M}'_{nT}(\tilde{\Phi}_{T2}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT},\\
&\qquad \mathbf{Y}'_{nT}\tilde{M}'_{nT}(\tilde{\Phi}_{T3}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT}\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1}\otimes I_n)\tilde{M}_{nT}\mathbf{Y}_{nT}.
\end{aligned}$$

Using $\mathbf{Y}_{nT} = \gamma_0\mathcal{Y}_{n,T-1} + \kappa_{m0}\iota_{n,T+1} + \mathbf{U}_{nT}$, we have $s_{nT,2} = -s_{nT,4} - 2\gamma_0 s_{nT,1}$, where

$$\begin{aligned}
s_{nT,4} &= \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T1}\otimes I_n)\tilde{M}_{nT}\mathbf{U}_{nT},\ \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T2}\otimes I_n)\tilde{M}_{nT}\mathbf{U}_{nT},\\
&\qquad \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T3}\otimes I_n)\tilde{M}_{nT}\mathbf{U}_{nT}\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1},
\end{aligned}$$

and $s_{nT,3} = \gamma_0 s_{nT,4} + \gamma_0^2 s_{nT,1} + o_p(1)$ as $\frac{1}{n}\mathbf{U}'_{nT}\tilde{M}'_{nT}(\tilde{\Phi}_{Tj}\otimes I_n)\tilde{M}_{nT}\mathbf{U}_{nT} = o_p(1)$ and $\frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1}\otimes I_n)\tilde{M}_{nT}\mathbf{U}_{nT} = o_p(1)$. Then the quadratic equation has the roots

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

It follows that 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 $s_{nT,4} < 0$. The $s_{nT,4}$ can be estimated by

$$\tilde{s}_{nT,4} = \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T1}\otimes I_n)\tilde{M}_{nT}(\mathbf{Y}_{nT}-\tilde{\gamma}\mathcal{Y}_{n,T-1}),\ \mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T2}\otimes I_n)\tilde{M}_{nT}(\mathbf{Y}_{nT}-\tilde{\gamma}\mathcal{Y}_{n,T-1}),$$
$$\mathcal{Y}'_{n,T-1}\tilde{M}'_{nT}(\tilde{\Phi}_{T3}\otimes I_n)\tilde{M}_{nT}(\mathbf{Y}_{nT}-\tilde{\gamma}\mathcal{Y}_{n,T-1})\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}^{-1}_{T+1}\otimes I_n)\tilde{M}_{nT}\mathcal{Y}_{n,T-1}$$

to determine the consistent root. Denote this root estimate by $\dot{\gamma}$. We may show that $\sqrt{n}[g_{nT,1}(\gamma_0,\theta_{20}), g'_{nT,2}(\gamma_0,\theta_{20})]' = \sqrt{n}g_{nT}(\gamma_0,\theta_{20}) + o_p(1)$, where

$$g_{nT}(\gamma_0,\theta_{20}) = \frac{1}{n}\begin{pmatrix}\mathbf{U}'_{nT}(F'_{T+1}\Omega^{-1}_{T+1}\otimes I_n)\mathbf{U}_{nT} + \kappa_{m0}\iota'_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1}\otimes I_n)M_{nT}\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T1}\otimes I_n)\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T2}\otimes I_n)\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T3}\otimes I_n)\mathbf{U}_{nT}\end{pmatrix}$$

with $\Phi_{Tj} = \Phi_{Tj}(\theta_{20})$ for $j = 1,2,3$, and $M_{nT} = M_{nT}(\theta_{20})$. We denote $\Sigma_{T,\gamma c} = \mathrm{Var}[\sqrt{n}g_{nT}(\gamma_0,\theta_{20})]$.

We may show that $\hat{C}_{nT,\gamma c} = C_{T,\gamma c} + o_p(1)$, where

$$C_{T,\gamma c} = -\big[\mathrm{tr}\big(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\varpi_m}\big), \mathrm{tr}\big(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}\big), \mathrm{tr}\big(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\phi_c}\big)\big]$$
$$\cdot\begin{pmatrix}\mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T1}}{\partial\varpi_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T1}}{\partial\eta_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T1}}{\partial\phi_c}\big) \\ \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T2}}{\partial\varpi_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T2}}{\partial\eta_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T2}}{\partial\phi_c}\big) \\ \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T3}}{\partial\varpi_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T3}}{\partial\eta_m}\big) & \mathrm{tr}\big(\Omega_{T+1}\frac{\partial\Phi_{T3}}{\partial\phi_c}\big)\end{pmatrix}^{-1}. \tag{A.5}$$

Under the following assumption, $s_{nT,1}$ has a nonzero probability limit.

Assumption A.7. $C_{T,\gamma c}\begin{pmatrix}\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T1}F_{T+1}\Omega_{T+1}) + \frac{1}{n}\kappa^2_{m0}\iota'_{nT}(F'_{T+1}\otimes I_n)M'_{nT}(\Phi_{T1}\otimes I_n)M_{nT}(F_{T+1}\otimes I_n)\iota_{nT} \\ \sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T2}F_{T+1}\Omega_{T+1}) + \frac{1}{n}\kappa^2_{m0}\iota'_{nT}(F'_{T+1}\otimes I_n)M'_{nT}(\Phi_{T2}\otimes I_n)M_{nT}(F_{T+1}\otimes I_n)\iota_{nT} \\ \sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T3}F_{T+1}\Omega_{T+1}) + \frac{1}{n}\kappa^2_{m0}\iota'_{nT}(F'_{T+1}\otimes I_n)M'_{nT}(\Phi_{T3}\otimes I_n)M_{nT}(F_{T+1}\otimes I_n)\iota_{nT}\end{pmatrix} \neq 0.$

Theorem A.2. *Under Assumptions 2.1–2.2 and A.1–A.7, the SGMM estimator $\dot{\gamma}$ is consistent and follows the asymptotic distribution*

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

where $R_{T,\gamma c} = [1, -C_{T,\gamma c}]$, and

$$G_{T,\gamma c} = -\begin{pmatrix}\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}F_{T+1}\Omega_{T+1}) + \frac{1}{n}\kappa^2_{m0}\iota'_{nT}(F'_{T+1}\Omega^{-1}_{T+1}\otimes I_n)M_{nT}(F_{T+1}\otimes I_n)\iota_{nT} \\ 2\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\varpi_m}) \\ 2\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}) \\ 2\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\phi_c})\end{pmatrix}.$$

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

A.2 The random effects DPD with exogenous variables and a short past

Consider the DPD model with exogenous variables and random effects:

$$y_{it} = \gamma_0 y_{i,t-1} + x_{it}\beta_0 + z_i b_0 + c_{i0} + v_{it}, \quad t = 1, \ldots, T, \tag{A.6}$$

and y_{i0} is observable. For y_{i0}, by continuous substitution, we have

$$y_{i0} = \gamma_0^m y_{i,-m} + x_{i0}\beta_0 + \sum_{j=1}^{m-1} \gamma_0^j x_{i,-j}\beta_0 + \sum_{j=0}^{m-1} \gamma_0^j z_i b_0 + \sum_{j=0}^{m-1} \gamma_0^j (c_{i0} + v_{i,-j}).$$

For the unobserved $x_{i,-1}, \ldots, x_{i,-m+1}$ and $y_{i,-m}$, we may use z_i and $\vec{x}_i = [x_{i0}, \ldots, x_{iT}]$ to predict them. Assume that $\mathrm{E}(y_{i,-m}|z_i, \vec{x}_i) = z_i\varrho + \vec{x}_i\pi$ 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 $\alpha^{(1)} = \gamma_0^m\varrho + \sum_{j=1}^{m-1}\gamma_0^j\varrho_j\beta_0 + \sum_{j=0}^{m-1}\gamma_0^j b_0$, $\alpha^{(2)} = \gamma_0^m\pi + \sum_{j=1}^{m-1}\gamma_0^j\pi_j\beta_0 + [\beta_0', 0, \ldots, 0]'$, $\xi_{i0} = \sum_{j=0}^{m-1}\gamma_0^j(c_{i0} + v_{i,-j}) + p_i$, and $p_i = \gamma_0^m[y_{i,-m} - \mathrm{E}(y_{i,-m}|z_i, \vec{x}_i)] + \sum_{j=1}^{m-1}\gamma_0^j[x_{i,-j} - \mathrm{E}(x_{i,-j}|z_i, \vec{x}_i)]\beta_0$ is the prediction error. Then the variance of y_{i0} is $\varpi_{m0}\sigma_{v0}^2$, where ϖ_{m0} is a free parameter. As $\mathrm{Cov}(y_{i0}, y_{it} - \gamma_0 y_{i,t-1}) = \phi_{c0}\sigma_{v0}^2\sum_{j=0}^{m-1}\gamma_0^j$ for $t \geq 1$, by denoting $\mathrm{Cov}(y_{i0}, y_{it} - \gamma_0 y_{i,t-1}) = \eta_{m0}\sigma_{v0}^2$, η_{m0} is a free parameter because m is finite but unknown.

The quasi log likelihood function of the model is

$$\ln L_r(\theta) = -\frac{n(T+1)}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|\Omega_{T+1}(\theta_2)| - \frac{1}{2\sigma_v^2}\mathbf{U}_{nT}'(\theta_1)(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1), \tag{A.7}$$

where $\theta_1 = [\alpha^{(1)'}, \alpha^{(2)'}, b', \beta', \gamma]'$, $\theta_2 = [\varpi_m, \eta_m, \phi_c]'$, $\theta = [\theta_1', \theta_2', \sigma_v^2]'$, $\Omega_{T+1}(\theta_2) = \begin{pmatrix} \varpi_m & \eta_m l_T' \\ \eta_m l_T & \phi_c l_T l_T' + I_T \end{pmatrix}$, $\mathbf{U}_{nT}(\theta_1) = \mathbf{Y}_{nT} - \mathbf{Z}_{nT}^*\theta_1$, $\mathbf{Y}_{nT} = [Y_{n0}', Y_{n1}', \ldots, Y_{nT}']'$, and

$$\mathbf{Z}_{nT}^* = \begin{pmatrix} Z_n & \vec{\mathbf{X}}_{nT} & 0 & 0 & 0 \\ 0 & 0 & Z_n & X_{n1} & Y_{n0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & Z_n & X_{nT} & Y_{n,T-1} \end{pmatrix}.$$

Let θ_{2j} be the jth element of θ_2. The first order derivatives of $\ln L_r(\theta)$ are

$$\begin{aligned}
\frac{\partial \ln L_r(\theta)}{\partial \theta_1} &= \frac{1}{\sigma_v^2}\mathbf{Z}_{nT}^{*'}(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1), \\
\frac{\partial \ln L_r(\theta)}{\partial \theta_{2j}} &= -\frac{n}{2}\operatorname{tr}\Big(\Omega_{T+1}^{-1}(\theta_2)\frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}}\Big) + \frac{1}{2\sigma_v^2}\mathbf{U}_{nT}'(\theta_1)\Big(\Omega_{T+1}^{-1}(\theta_2)\frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}}\Omega_{T+1}^{-1}(\theta_2) \otimes I_n\Big)\mathbf{U}_{nT}(\theta_1), \\
\frac{\partial \ln L_r(\theta)}{\partial \sigma_v^2} &= -\frac{n(T+1)}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\mathbf{U}_{nT}'(\theta_1)(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1).
\end{aligned}$$

A.2.1 Efficient GMM

For model (A.6), denote $\mathcal{Y}_{n,T-1} = [0, Y'_{n0}, Y'_{n1}, \ldots, Y'_{n,T-1}]'$. Using $\mathcal{Y}_{n,T-1} = (F_{T+1} \otimes I_n)(\mathbf{U}_{nT} + \Upsilon^*_{n,T+1}\theta_\Upsilon)$, where

$$\Upsilon^*_{n,T+1} = \begin{pmatrix} Z_n & \vec{\mathbf{X}}_{nT} & 0 & 0 \\ 0 & 0 & Z_n & X_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & Z_n & X_{nT} \end{pmatrix}$$

and $\theta_\Upsilon = [\alpha_0^{(1)\prime}, \alpha_0^{(2)\prime}, b'_0, \beta'_0]'$, we have $\mathcal{Y}'_{n,T-1}(\Omega^{-1}_{T+1} \otimes I_n)\mathbf{U}_{nT} = \mathbf{U}'_{nT}(F'_{T+1}\Omega^{-1}_{T+1} \otimes I_n)\mathbf{U}_{nT} + \theta'_\Upsilon \Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1} \otimes I_n)\mathbf{U}_{nT}$. As in the last subsection, let $\Omega^{-1}_{T+1} = \mathcal{B}'_T\mathcal{B}_T$ and $\Omega^{-1}_{T+1}(\theta_2) = \mathcal{B}'_T(\theta_2)\mathcal{B}_T(\theta_2)$. Then we may consider a GMM estimation with the following moment vector:

$$g_{nT}(\theta_1, \theta_2) = \frac{1}{n}\begin{pmatrix} \Upsilon^{*\prime}_{n,T+1}(K_{1T} \otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \vdots \\ \Upsilon^{*\prime}_{n,T+1}(K_{m_1T} \otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C_{1T}\mathcal{B}_T(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \vdots \\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C_{m_2T}\mathcal{B}_T(\theta_2) \otimes I_n)\mathbf{U}_{nT}(\theta_1) \end{pmatrix}, \tag{A.8}$$

where C_{jT}'s have zero traces. To compute $\Sigma_{nT} = \mathrm{Var}[\sqrt{n}g_{nT}(\theta_{10}, \theta_{20})]$, $\mathbf{U}_{nT}$ can be written as a linear transformation of a vector of independent elements

$$\mathbb{V}_{nT} = [\mathbf{c}_{n0}, \boldsymbol{\zeta}'_{n0}, V'_{n1}, \ldots, V'_{nT}]',$$

where $\boldsymbol{\zeta}_{n0} \equiv [\zeta_{10}, \ldots, \zeta_{n0}]' = \sum_{j=0}^{m-1} \gamma_0^j V_{n,-j} + [p_1, \ldots, p_n]'$. Let $\hat{\Sigma}_{nT}$ be a consistent estimator of $\lim_{n\to\infty} \Sigma_{nT}$. The optimal GMM estimator of $\theta_3 = [\theta'_1, \theta'_2]'$ with $g_{nT}(\theta_1, \theta_2)$ is

$$\hat{\theta}_{3,gmm} = \arg\min_{\theta_3} g'_{nT}(\theta_1, \theta_2)\hat{\Sigma}^{-1}_{nT}g_{nT}(\theta_1, \theta_2). \tag{A.9}$$

Under the normality assumption of $\mathbf{U}_{nT}$'s, $\Sigma_{nT} = \frac{\sigma^2_{v0}}{n}\Delta'_{nT}\Delta_{nT}$, where

$$\Delta_{nT} = \begin{pmatrix} (\mathcal{B}'^{-1}_T K'_{1T} \otimes I_n)\Upsilon^*_{n,T+1} & \ldots & (\mathcal{B}'^{-1}_T K'_{m_1T} \otimes I_n)\Upsilon^*_{n,T+1} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^s_{1T}) & \ldots & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^s_{m_2T}) \end{pmatrix}.$$

The asymptotic variance matrix of $\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30})$ is $\lim_{n\to\infty}(G'_{nT}\Sigma^{-1}_{nT}G_{nT})^{-1}$, where

$$G_{nT} = \mathrm{E}\Big(\frac{\partial g_{nT}(\theta_{10}, \theta_{20})}{\partial \theta'_3}\Big) = [G_{nT,1}, G_{nT,2}], \tag{A.10}$$

with

$$G_{nT,1}=\begin{pmatrix} -\frac{1}{n}\Upsilon^{*\prime}_{n,T+1}(K_{1T}\otimes I_n)\Upsilon^*_{n,T+1} & -\frac{1}{n}\Upsilon^{*\prime}_{n,T+1}(K_{1T}F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ \vdots & \vdots \\ -\frac{1}{n}\Upsilon^{*\prime}_{n,T+1}(K_{m_1T}\otimes I_n)\Upsilon^*_{n,T+1} & -\frac{1}{n}\Upsilon^{*\prime}_{n,T+1}(K_{m_1T}F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ 0 & -\sigma^2_{v0}\operatorname{tr}(C^s_{1T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) \\ \vdots & \vdots \\ 0 & -\sigma^2_{v0}\operatorname{tr}(C^s_{m_2T}\mathcal{B}_TF_{T+1}\mathcal{B}_T^{-1}) \end{pmatrix},$$

and

$$G_{nT,2}=\begin{pmatrix} 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \\ \sigma^2_{v0}\operatorname{tr}(C^s_{1T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\operatorname{tr}(C^s_{1T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\operatorname{tr}(C^s_{1T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \\ \vdots & \vdots & \vdots \\ \sigma^2_{v0}\operatorname{tr}(C^s_{m_2T}\mathcal{B}_{\varpi T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\operatorname{tr}(C^s_{m_2T}\mathcal{B}_{\eta T}\mathcal{B}_T^{-1}) & \sigma^2_{v0}\operatorname{tr}(C^s_{m_2T}\mathcal{B}_{\phi T}\mathcal{B}_T^{-1}) \end{pmatrix}.$$

Let $\Delta^*_{nT}=\begin{pmatrix}\Delta^*_{nT,1} & \\ & \Delta^*_{nT,2}\end{pmatrix}$, where $\Delta^*_{nT,1}=[(\mathcal{B}_T\otimes I_n)\Upsilon^*_{n,T+1},(\mathcal{B}_TF_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}]$ and

$$\Delta^*_{nT,2}=\left[\sqrt{\tfrac{n}{2}}\sigma_{v0}\operatorname{vec}((C^*_{1T})^s),\sqrt{\tfrac{n}{2}}\sigma_{v0}\operatorname{vec}((C^*_{2T})^s),\sqrt{\tfrac{n}{2}}\sigma_{v0}\operatorname{vec}((C^*_{3T})^s),\sqrt{\tfrac{n}{2}}\sigma_{v0}\operatorname{vec}((C^*_{4T})^s)\right]$$

with C^*_{jT} for $j=1,\ldots,4$ given below (A.3). Then $G_{nT}=\frac{1}{n}\Delta'_{nT}\Delta^*_{nT}\theta_\Delta$, where

$$\theta_\Delta=\begin{pmatrix} -I_{k_\Upsilon} & 0 & 0 & 0 & 0 \\ 0 & -\theta_\Upsilon & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

with k_Υ being the column dimension of $\Upsilon^*_{n,T+1}$. Thus, by the generalized Schwarz inequality, $G'_{nT}\Sigma^{-1}_{nT}G_{nT}\leq\frac{1}{n\sigma^2_{v0}}\theta'_\Delta\Delta^{*\prime}_{nT}\Delta^*_{nT}\theta_\Delta$, and the equality holds at $\Delta_{nT}=\Delta^*_{nT}$, i.e., the best moment vector under normal disturbances is[1]

$$g^*_{nT}(\theta_1,\theta_2)=\frac{1}{n}\begin{pmatrix} \Upsilon^{*\prime}_{n,T+1}(\Omega^{-1}_{T+1}\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1}\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C^*_{1T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \vdots \\ \mathbf{U}'_{nT}(\theta_1)(\mathcal{B}'_T(\theta_2)C^*_{4T}\mathcal{B}_T(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \end{pmatrix}, \tag{A.11}$$

which corresponds to the ML score vector.

We maintain the following assumptions for asymptotic analysis.

[1] Let $r_{nT}=[0_{1\times n},l'_{nT}]'$ and $s_{nT}=[l'_n,0_{1\times nT}]'$. As $r_{nT}=(F_{T+1}\otimes I_n)r_{nT}(1-\gamma)+(F_{T+1}\otimes I_n)s_{nT}$, the vectors r_{nT}, $(F_{T+1}\otimes I_n)r_{nT}$ and $(F_{T+1}\otimes I_n)s_{nT}$ are linearly dependent. Thus, if Z_n includes l_n, one of $(\Omega^{-1}_{T+1}\otimes I_n)r_{nT}$, $(\Omega^{-1}_{T+1}F_{T+1}\otimes I_n)r_{nT}$ and $(\Omega^{-1}_{T+1}F_{T+1}\otimes I_n)s_{nT}$ is redundant and should be removed from the set of IVs when implementing the best GMM.

Assumption A.8. *The process* $\{y_{it}\}$ *has started from a finite unknown* m *periods ago, and* $y_{i0} = z_i\alpha^{(1)} + \vec{x}_i\alpha^{(2)} + \sum_{j=0}^{m-1}\gamma_0^j c_{i0} + \zeta_{i0}$*, where* ζ_{i0}*'s are i.i.d with mean zero,* $\mathrm{E}(|\zeta_{i0}|^{4+\eta}) < \infty$ *for some* $\eta > 0$*, and* ζ_{i0}*'s are independent of* v_{jt}*'s for* $t \geq 1$ *and* c_{j0}*'s.*

Assumption A.9. C_{jT}*'s have zero traces and are linearly independent, and* $\lim_{n\to\infty}\frac{1}{n}[(K'_{1T}\otimes I_n)\Upsilon^*_{n,T+1},\ldots,(K'_{m_1T}\otimes I_n)\Upsilon^*_{n,T+1}]'[(K'_{1T}\otimes I_n)\Upsilon^*_{n,T+1},\ldots,(K'_{m_1T}\otimes I_n)\Upsilon^*_{n,T+1}]$ *has full rank.*

Assumption A.10. *When* $\theta_\Upsilon \neq 0$*,* $\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\Upsilon^{*\prime}_{n,T+1}(K_{1T}\otimes I_n)\Upsilon^*_{n,T+1} & \Upsilon^{*\prime}_{n,T+1}(K_{1T}F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ \vdots & \vdots \\ \Upsilon^{*\prime}_{n,T+1}(K_{m_1T}\otimes I_n)\Upsilon^*_{n,T+1} & \Upsilon^{*\prime}_{n,T+1}(K_{m_1T}F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon\end{pmatrix}$ *has full column rank, and* $\Lambda_1(\theta_2)$ *has full column rank for any* $\theta_2 \neq \theta_{20}$*; when* $\theta_\Upsilon = 0$*,* $\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\Upsilon^{*\prime}_{n,T+1}(K_{1T}\otimes I_n)\Upsilon^*_{n,T+1} \\ \vdots \\ \Upsilon^{*\prime}_{n,T+1}(K_{m_1T}\otimes I_n)\Upsilon^*_{n,T+1}\end{pmatrix}$ *has full column rank, and* $[\Lambda_1(\theta_2),\Lambda_2(\theta_2)]$ *has full column rank for any* $\theta_2 \neq \theta_{20}$*, where* $\Lambda_1(\theta_2)$ *and* $\Lambda_2(\theta_2)$ *are in Assumption A.4.*

Assumption A.11. *When* $\theta_\Upsilon \neq 0$*,* Λ_3 *has full column rank; when* $\theta_\Upsilon = 0$*,* Λ_4 *has full column rank, where* Λ_3 *and* Λ_4 *are in Assumption A.5.*

Assumption A.12. *The parameter space* Θ *of* θ *is compact,* $\varpi_m > 0$*,* $\phi_c > \frac{\eta_m^2}{\varpi_m} - \frac{1}{T}$*, and* θ_0 *is in the interior of* Θ*.*

Assumption A.13. $\lim_{n\to\infty}\frac{1}{n}\Upsilon^{*\prime}_{n,T+1}\Upsilon^*_{n,T+1}$ *has full rank.*

Assumptions A.8–A.12 correspond to Assumptions A.2–A.6 for the random effects pure DPD model with a short past. Assumption A.13 is sufficient identification condition for the QML estimator $\hat{\theta}_{qml}$ that maximizes (A.7). Let $\hat{\theta}_{3,qml}$ be a subvector of $\hat{\theta}_{qml}$ corresponding to θ_3, $\hat{\theta}^*_{3,gmm}$ be the best GMM estimator with the moment vector $g^*_{nT}(\theta_1,\theta_2)$ in (A.11) for the model with normal disturbances, $G^*_{nT} = \mathrm{E}(\frac{\partial g^*_{nT}(\theta_{10},\theta_{20})}{\partial\theta_3'})$, and $\Sigma^*_{nT} = \mathrm{Var}[\sqrt{n}g^*_{nT}(\theta_{10},\theta_{20})]$.

Theorem A.3. *Suppose that Assumptions 2.1–2.2, 3.2, A.1 and A.8–A.13 are satisfied.*

(i) The optimal GMM estimator $\hat{\theta}_{3,gmm}$ *in* (A.9) *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} *is in* (A.10).

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

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

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*

$$\mathrm{E}(-\frac{1}{n}\frac{\partial^2\ln L_w(\theta_0)}{\partial\theta\partial\theta'}) = \begin{pmatrix}\Pi_{nT} + \frac{2\sigma_{v0}^4}{T+1}\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2\ln L_w(\theta_0)}{\partial\theta_3\partial\sigma_v^2}\big)\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2\ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) & \mathrm{E}\big(-\frac{1}{n}\frac{\partial^2\ln L_w(\theta_0)}{\partial\theta_3\partial\sigma_v^2}\big) \\ \mathrm{E}\big(-\frac{1}{n}\frac{\partial^2\ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) & \frac{T+1}{2\sigma_{v0}^4}\end{pmatrix},$$

$\mathrm{E}\big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta_3'\partial\sigma_v^2}\big) = [0, \frac{1}{2\sigma_{v0}^2}\operatorname{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\kappa_m}), \frac{1}{2\sigma_{v0}^2}\operatorname{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}), \frac{1}{2\sigma_{v0}^2}\operatorname{tr}(\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\phi_c})]$, $\Pi_{nT} = \begin{pmatrix}\Pi_{nT,1} & \\ & 0_{3\times 3}\end{pmatrix} + \Pi_{nT,2}$,

$$\Pi_{nT,1} = \frac{1}{n\sigma_{v0}^2}\begin{pmatrix} \Upsilon_{n,T+1}^{*\prime}(\Omega_{T+1}^{-1}\otimes I_n)\Upsilon_{n,T+1}^{*} & \Upsilon_{n,T+1}^{*\prime}(\Omega_{T+1}^{-1}F_{T+1}\otimes I_n)\Upsilon_{n,T+1}^{*}\theta_\Upsilon \\ \theta_\Upsilon'\Upsilon_{n,T+1}^{*\prime}(F_{T+1}'\Omega_{T+1}^{-1}\otimes I_n)\Upsilon_{n,T+1}^{*} & \theta_\Upsilon'\Upsilon_{n,T+1}^{*\prime}(F_{T+1}'\Omega_{T+1}^{-1}F_{T+1}\otimes I_n)\Upsilon_{n,T+1}^{*}\theta_\Upsilon \end{pmatrix}$$

and

$$\Pi_{nT,2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \operatorname{tr}(C_{1T}^{*s}C_{1T}^{*}) & -\operatorname{tr}(C_{1T}^{*s}C_{2T}^{*}) & -\operatorname{tr}(C_{1T}^{*s}C_{3T}^{*}) & -\operatorname{tr}(C_{1T}^{*s}C_{4T}^{*}) \\ 0 & -\operatorname{tr}(C_{2T}^{*s}C_{1T}^{*}) & \operatorname{tr}(C_{2T}^{*s}C_{2T}^{*}) & \operatorname{tr}(C_{2T}^{*s}C_{3T}^{*}) & \operatorname{tr}(C_{2T}^{*s}C_{4T}^{*}) \\ 0 & -\operatorname{tr}(C_{3T}^{*s}C_{1T}^{*}) & \operatorname{tr}(C_{3T}^{*s}C_{2T}^{*}) & \operatorname{tr}(C_{3T}^{*s}C_{3T}^{*}) & \operatorname{tr}(C_{3T}^{*s}C_{4T}^{*}) \\ 0 & -\operatorname{tr}(C_{4T}^{*s}C_{1T}^{*}) & \operatorname{tr}(C_{4T}^{*s}C_{2T}^{*}) & \operatorname{tr}(C_{4T}^{*s}C_{3T}^{*}) & \operatorname{tr}(C_{4T}^{*s}C_{4T}^{*}) \end{pmatrix}.$$

(iii) $\hat\theta_{3,gmm}^*$ *is asymptotically efficient relative to* $\hat\theta_{3,qml}$ *in general, i.e.,* $(G_{nT}^{*\prime}\Sigma_{nT}^{*-1}G_{nT}^*)^{-1} \le \Gamma_{nT,\theta_3}$*, where* $G_{nT}^{*\prime}\Sigma_{nT}^{*-1}G_{nT}^* = \Pi_{nT}$*,* 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 $\mathbb{V}_{nT}$ *is normally distributed, then:*

(a) Among optimal GMM estimators with moments of the form (A.8)*,* $\hat\theta_{3,gmm}^*$ *has the minimum asymptotic variance, i.e.,* $(G_{nT}^{*\prime}\Sigma_{nT}^{*-1}G_{nT}^*)^{-1} \le (G_{nT}'\Sigma_{nT}^{-1}G_{nT})^{-1}$*.*

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

A.2.2 SGMM

Denote $\delta_1 = [b', \beta', \gamma]'$ and $\theta_4 = [\alpha^{(1)'}, \alpha^{(2)'}, \theta_2']'$. Let $\tilde\delta_1$ and $\tilde\theta_4$ be, respectively, $\sqrt{n}$-consistent estimators of δ_{10} and θ_{40}, $\tilde F_{T+1} = F_{T+1}(\tilde\gamma)$ and $\tilde\Omega_{T+1} = \Omega_{T+1}(\tilde\theta_2)$. An SGMM estimator of $\delta_{10} = [b_0', \beta_0', \gamma_0]'$ can be based on the efficient moment vector under normality $g_{nT}(\delta_1, \theta_4) = [g_{nT,1}'(\delta_1,\theta_4), g_{nT,2}'(\delta_1,\theta_4)]'$, where

$$g_{nT,1}(\delta_1,\theta_4) = \frac{1}{n}\begin{pmatrix} \mathbb{X}_{nT}'(\tilde\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \Upsilon_{n,T+1}^{*\prime}(\tilde F_{T+1}'\tilde\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}_{nT}'(\theta_1)(\tilde F_{T+1}'\Omega_{T+1}^{-1}(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \end{pmatrix},$$

$$g_{nT,2}(\delta_1,\theta_4) = \frac{1}{n}\begin{pmatrix} \Upsilon_{n,T+1}'(\tilde\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}_{nT}'(\theta_1)(\Phi_{T1}(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}_{nT}'(\theta_1)(\Phi_{T2}(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \\ \mathbf{U}_{nT}'(\theta_1)(\Phi_{T3}(\theta_2)\otimes I_n)\mathbf{U}_{nT}(\theta_1) \end{pmatrix},$$

with $\mathbb{X}_{nT} = \begin{pmatrix} 0 & 0 \\ Z_n & X_{n1} \\ \vdots & \vdots \\ Z_n & X_{nT}\end{pmatrix}$ and $\Upsilon_{n,T+1} = \begin{pmatrix} Z_n & \vec{\mathbf{X}}_{nT} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0\end{pmatrix}$. Let $\Sigma_{nT,\delta_1} = \operatorname{Var}[\sqrt{n}g_{nT,b}(\delta_{10},\theta_{40})]$, where $g_{nT,b}(\delta_1,\theta_4)$ is the vector derived by replacing, respectively, $\tilde\Omega_{T+1}$ and $\tilde F_{T+1}$ in $g_{nT}(\delta_1,\theta_4)$ by $\Omega_{T+1}(\theta_2)$ and $F_{T+1}(\gamma)$.

Then we have the following SGMM estimator:

$$\dot{\delta}_1 = \arg\min_{\delta_1}[\hat{R}_{nT,\delta_1} g_{nT}(\delta_1, \tilde{\theta}_4)]'(\hat{R}_{nT,\delta_1}\hat{\Sigma}_{nT,\delta_1}\hat{R}'_{nT,\delta_1})^{-1}[\hat{R}_{nT,\delta_1} g_{nT}(\delta_1, \tilde{\theta}_4)], \tag{A.12}$$

where $\hat{R}_{nT,\delta_1} = [I, -\frac{\partial g_{nT,1}(\tilde{\delta}_1,\tilde{\theta}_4)}{\partial \theta_4'}(\frac{\partial g_{nT,2}(\tilde{\delta}_1,\tilde{\theta}_4)}{\partial \theta_4'})^{-1}]$ with I being an identity matrix conformable with $g_{nT,1}(\delta_1, \theta_4)$, and $\hat{\Sigma}_{nT,\delta_1}$ is a consistent estimator of $\lim_{n\to\infty} \Sigma_{nT,\delta_1}$.

We can also focus on an SGMM estimator of γ based on the QML first order conditions. With $\alpha^{(1)}$, $\alpha^{(2)}$, b, β and σ_v^2 being concentrated out from the QML first order conditions, we derive the moments $g_{nT,1}(\gamma, \theta_2) = \frac{1}{n}\mathcal{Y}'_{n,T-1}(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)M^*_{nT}(\theta_2)(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1})$ and $g_{nT,2}(\gamma, \theta_2)$, where

$$M^*_{nT}(\theta_2) = I_{n(T+1)} - \Upsilon^*_{n,T+1}[\Upsilon^{*'}_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n)\Upsilon^*_{n,T+1}]^{-1}\Upsilon^{*'}_{n,T+1}(\Omega_{T+1}^{-1}(\theta_2) \otimes I_n),$$

and $g_{nT,2}(\gamma, \theta_2)$ is a 3×1 vector with its jth element being

$$\frac{1}{n}(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1})'M^{*'}_{nT}(\theta_2)\big[\Phi_{Tj}(\theta_2) \otimes I_n\big]M^*_{nT}(\theta_2)(\mathbf{Y}_{nT} - \gamma\mathcal{Y}_{n,T-1}).$$

Denote $\hat{C}_{nT,\gamma c} = \frac{\partial g_{nT,1}(\tilde{\gamma},\tilde{\theta}_2)}{\partial \theta_2'}(\frac{\partial g_{nT,2}(\tilde{\gamma},\tilde{\theta}_2)}{\partial \theta_2'})^{-1}$ and $\tilde{M}^*_{nT} = M^*_{nT}(\tilde{\theta}_2)$. The SGMM estimator $\ddot{\gamma}$ of γ_0 is characterized by the quadratic equation

$$g_{nT,1}(\ddot{\gamma}, \tilde{\theta}_2) - \hat{C}_{nT,\gamma c} g_{nT,2}(\ddot{\gamma}, \tilde{\theta}_2) = -s_{nT,1}\ddot{\gamma}^2 - s_{nT,2}\ddot{\gamma} - s_{nT,3} = 0, \tag{A.13}$$

where

$$\begin{aligned}
s_{nT,1} &= \frac{1}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T1} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1},\ \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T2} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1},\\
&\quad \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T3} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1}\big]',\\
s_{nT,2} &= \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1} - \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T1} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT},\\
&\quad \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T2} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT},\ \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T3} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT}\big]',\\
s_{nT,3} &= \frac{1}{n}\hat{C}_{nT,\gamma c}\big[\mathbf{Y}'_{nT}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T1} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT},\ \mathbf{Y}'_{nT}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T2} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT},\\
&\quad \mathbf{Y}'_{nT}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T3} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT}\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1} \otimes I_n)\tilde{M}^*_{nT}\mathbf{Y}_{nT}.
\end{aligned}$$

Using $\mathbf{Y}_{nT} = \gamma_0\mathcal{Y}_{n,T-1} + \Upsilon^*_{n,T+1}\theta_\Upsilon + \mathbf{U}_{nT}$, we have $s_{nT,2} = -s_{nT,4} - 2\gamma_0 s_{nT,1}$, where

$$\begin{aligned}
s_{nT,4} &= \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T1} \otimes I_n)\tilde{M}^*_{nT}\mathbf{U}_{nT},\ \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T2} \otimes I_n)\tilde{M}^*_{nT}\mathbf{U}_{nT},\\
&\quad \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T3} \otimes I_n)\tilde{M}^*_{nT}\mathbf{U}_{nT}\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1},
\end{aligned}$$

and $s_{nT,3} = \gamma_0 s_{nT,4} + \gamma_0^2 s_{nT,1} + o_p(1)$ as $\frac{1}{n}\mathbf{U}'_{nT}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{Tj} \otimes I_n)\tilde{M}^*_{nT}\mathbf{U}_{nT} = o_p(1)$ and $\frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}_{T+1}^{-1} \otimes I_n)\tilde{M}^*_{nT}\mathbf{U}_{nT} = o_p(1)$. Then the quadratic equation has the roots

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

It follows that 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 $s_{nT,4} < 0$. The $s_{nT,4}$ can be estimated by

$$\tilde{s}_{nT,4} = \frac{2}{n}\hat{C}_{nT,\gamma c}\big[\mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T1} \otimes I_n)\tilde{M}^*_{nT}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathcal{Y}_{n,T-1}),\ \mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T2} \otimes I_n)\tilde{M}^*_{nT}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathcal{Y}_{n,T-1}),$$
$$\mathcal{Y}'_{n,T-1}\tilde{M}^{*'}_{nT}(\tilde{\Phi}_{T3} \otimes I_n)\tilde{M}^*_{nT}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathcal{Y}_{n,T-1})\big]' - \frac{1}{n}\mathcal{Y}'_{n,T-1}(\tilde{\Omega}^{-1}_{T+1} \otimes I_n)\tilde{M}^*_{nT}\mathcal{Y}_{n,T-1}.$$

We can show that $\sqrt{n}[g'_{nT,1}(\gamma_0,\theta_{20}), g'_{nT,2}(\gamma_0,\theta_{20})]' = \sqrt{n}g_{nT,\gamma c}(\gamma_0,\theta_{20}) + o_p(1)$, where

$$g_{nT,\gamma c}(\gamma_0,\theta_{20}) = \frac{1}{n}\begin{pmatrix} \theta'_\Upsilon \Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1} \otimes I_n)M^*_{nT}\mathbf{U}_{nT} + \mathbf{U}'_{nT}(F'_{T+1}\Omega^{-1}_{T+1} \otimes I_n)\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T1} \otimes I_n)\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T2} \otimes I_n)\mathbf{U}_{nT} \\ \mathbf{U}'_{nT}(\Phi_{T2} \otimes I_n)\mathbf{U}_{nT} \end{pmatrix}$$

with $M^*_{nT} = M^*_{nT}(\theta_{20})$. Denote $\Sigma_{nT,\gamma c} = \mathrm{Var}[\sqrt{n}g_{nT,\gamma c}(\gamma_0,\theta_{20})]$.

The following assumptions are maintained for asymptotic analysis.

Assumption A.14. $\lim_{n\to\infty} \frac{1}{n}\begin{pmatrix} \mathbb{X}'_{nT}(\Omega^{-1}_{T+1}\otimes I_n)M_{nT}[\mathbb{X}_{nT},(F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon] \\ \Upsilon^{*'}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1}\otimes I_n)M_{nT}[\mathbb{X}_{nT},(F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon] \end{pmatrix}$ *has full column rank, where* $M_{nT} = I_{n(T+1)} - \Upsilon_{n,T+1}[\Upsilon'_{n,T+1}(\Omega^{-1}_{T+1} \otimes I_n)\Upsilon_{n,T+1}]^{-1}\Upsilon'_{n,T+1}(\Omega^{-1}_{T+1} \otimes I_n)$.

Assumption A.15. $\lim_{n\to\infty} C_{T,\gamma c}\begin{pmatrix} \sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T1}F_{T+1}\Omega_{T+1})+\frac{1}{n}\theta'_\Upsilon\Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\otimes I_n)M^{*\prime}_{nT}(\Phi_{T1}\otimes I_n)M^*_{nT}(F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ \sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T2}F_{T+1}\Omega_{T+1})+\frac{1}{n}\theta'_\Upsilon\Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\otimes I_n)M^{*\prime}_{nT}(\Phi_{T2}\otimes I_n)M^*_{nT}(F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ \sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Phi_{T3}F_{T+1}\Omega_{T+1})+\frac{1}{n}\theta'_\Upsilon\Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\otimes I_n)M^{*\prime}_{nT}(\Phi_{T3}\otimes I_n)M^*_{nT}(F_{T+1}\otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \end{pmatrix}$ $\neq 0$, *where* $C_{T,\gamma c}$ *is in* (A.5).

Assumption A.14 is a sufficient identification condition for the SGMM estimator $\dot{\delta}_1$. Under Assumption A.15, (A.13) is quadratic in γ in the limit.

Theorem A.4. *Suppose that Assumptions 2.1–2.2, 3.2, A.1 and A.8–A.13 are satisfied.*

(i) If Assumption A.14 also holds, the SGMM estimator $\dot{\delta}_1$ *is consistent and has the asymptotic distribution*

$$\sqrt{n}(\dot{\delta}_1 - \delta_{10}) \xrightarrow{d} N(0, \lim_{n\to\infty}[G'_{nT,\delta_1}R'_{nT,\delta_1}(R_{nT,\delta_1}\Sigma_{nT,\delta_1}R'_{nT,\delta_1})^{-1}R_{nT,\delta_1}G_{nT,\delta_1}]^{-1}),$$

where

$$G_{nT,\delta_1} = -\frac{1}{n}\begin{pmatrix} \mathbb{X}'_{nT}(\Omega^{-1}_{T+1} \otimes I_n)\mathbb{X}_{nT} & \mathbb{X}'_{nT}(\Omega^{-1}_{T+1}F_{T+1} \otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ \Upsilon^{*'}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1} \otimes I_n)\mathbb{X}_{nT} & \Upsilon^{*'}_{n,T+1}(F'_{T+1}\Omega^{-1}_{T+1}F_{T+1} \otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ 0 & n\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}F_{T+1}\Omega_{T+1}) \\ \Upsilon'_{n,T+1}(\Omega^{-1}_{T+1} \otimes I_n)\mathbb{X}_{nT} & \Upsilon'_{n,T+1}(\Omega^{-1}_{T+1}F_{T+1} \otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ 0 & 2n\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\varpi_m}) \\ 0 & 2n\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}) \\ 0 & 2n\sigma^2_{v0}\,\mathrm{tr}(F'_{T+1}\Omega^{-1}_{T+1}\frac{\partial\Omega_{T+1}}{\partial\phi_c}) \end{pmatrix}$$

and $R_{nT,\delta_1} = [I, -C_{nT,\delta_1}]$ *with*

$$C_{nT,\delta_1} = \begin{pmatrix} \mathbb{X}'_{nT}(\Omega_{T+1}^{-1} \otimes I_n)\Upsilon_{n,T+1}[\Upsilon'_{n,T+1}(\Omega_{T+1}^{-1} \otimes I_n)\Upsilon_{n,T+1}]^{-1} & 0 \\ \Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\Omega_{T+1}^{-1} \otimes I_n)\Upsilon_{n,T+1}[\Upsilon'_{n,T+1}(\Omega_{T+1}^{-1} \otimes I_n)\Upsilon_{n,T+1}]^{-1} & 0 \\ 0 & C_{T,\gamma c} \end{pmatrix}.$$

The asymptotic distribution of $\dot{\delta}_1$ *is the same as that of the GMM estimator of* δ_1 *in Theorem A.3(i).*

(ii) If Assumption A.15 also holds, the SGMM estimator $\ddot{\gamma}$ *is consistent and follows the asymptotic distribution*

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

where $R_{nT,\gamma c} = [1, -C_{T,\gamma c}]$ *and*

$$G_{nT,\gamma c} = -\begin{pmatrix} \sigma_{v0}^2 \operatorname{tr}(F'_{T+1}\Omega_{T+1}^{-1}F_{T+1}\Omega_{T+1}) + \frac{1}{n}\theta'_\Upsilon \Upsilon^{*\prime}_{n,T+1}(F'_{T+1}\Omega_{T+1}^{-1} \otimes I_n)M^*_{nT}(F_{T+1} \otimes I_n)\Upsilon^*_{n,T+1}\theta_\Upsilon \\ 2\sigma_{v0}^2 \operatorname{tr}(F'_{T+1}\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\varpi_m}) \\ 2\sigma_{v0}^2 \operatorname{tr}(F'_{T+1}\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\eta_m}) \\ 2\sigma_{v0}^2 \operatorname{tr}(F'_{T+1}\Omega_{T+1}^{-1}\frac{\partial\Omega_{T+1}}{\partial\phi_c}) \end{pmatrix}.$$

The asymptotic distribution of $\ddot{\gamma}$ *is the same as that of the QML estimator of* γ *in Theorem A.3(ii).*

A.3 Stationary random effects DPD

In this section, we consider stationary random effects DPD models. As the covariance matrix of the overall disturbance vector of the random effects model would be restricted functions of the structural parameters, it is desirable to relate the quasi log likelihood function of this random component model $\ln L_r(\theta)$ to the quasi log likelihood function of the within equations model. For that purpose, we realize that the joint density function $f_r(\mathbf{Y}_{nT})$ of $\mathbf{Y}_{nT} = [Y'_{n0}, \cdots, Y'_{nT}]'$ and the joint density function $f_w(\Delta\mathbf{Y}_{nT})$, where $\Delta\mathbf{Y}_{nT} = (\Delta Y'_{n1}, \cdots, \Delta Y'_{nT})'$, are related by

$$f_r(\mathbf{Y}_{nT}) = f_w(\Delta\mathbf{Y}_{nT})f_b(Y_{n0}|\Delta\mathbf{Y}_{nT}),$$

where $f_b(Y_{n0}|\Delta\mathbf{Y}_{nT})$ is simply the conditional density of Y_{n0} conditional on $\Delta\mathbf{Y}_{nT}$.[2] This follows because there is a one-to-one correspondence between $\mathbf{Y}_{nT}$ and $(Y'_{n0}, \Delta Y'_{n1}, \cdots, \Delta Y'_{nT})'$ with the determinant of its Jacobian matrix being one, and the density of the latter can be decomposed into the product of a conditional density and a marginal density. If the disturbances of the model are normally distributed, then the expression of the conditional density will be derived via its conditional mean and variance-covariance. Without normality, those densities represent only quasi-likelihood functions, but the decomposition can be derived via a quadratic partition matrix decomposition, which is from Lee and Yu (2018) and shown in

[2] In the presence of exogenous variables, they are conditional arguments in the three densities but are omitted for simplicity.

Section D. Anyhow, the random effects model can be decomposed into a "within equations" model and a "between equation" model. By the decomposition, we hope to see whether one can derive computationally more tractable moments for GMM estimation, which take into account that γ imposes constraints on the variance matrix Ω_{T+1}. Another interest of the model decomposition is that, as moments have already been considered for the within equation with stationarity, additional moments from the between equation highlight additional information from random effects compared with fixed effects models. For the DPD model with exogenous variables, as unobserved past values of exogenous variables are unknown but can be predicted, it has an additional prediction error, so SGMM estimation can also be used by concentrating on structural parameters.

A.3.1 Random effects stationary pure DPD

Consider the pure DPD model with random effects

$$Y_{nt} = \gamma Y_{n,t-1} + \mathbf{c}_{n0} + V_{nt}, \tag{A.14}$$

where $t = 1, \cdots, T$; c_i, $i = 1, \cdots, n$ are randomly distributed with $(0, \sigma_{c0}^2)$, and Y_{n0} is observable. Assume that the process has started long time ago and is stationary, then

$$Y_{n0} = \frac{1}{1-\gamma_0}\mathbf{c}_{n0} + \sum_{s=0}^{\infty} \gamma_0^s V_{n,-s}. \tag{A.15}$$

The variance matrix of $[Y'_{n0}, Y'_{n1} - \gamma_0 Y'_{n0}, \cdots, Y'_{nT} - \gamma_0 Y'_{n,T-1}]'$ is $\sigma_{v0}^2(\Omega_{T+1} \otimes I_n)$, where

$$\Omega_{T+1} = \begin{pmatrix} \frac{\phi_{c0}}{(1-\gamma_0)^2} + \frac{1}{1-\gamma_0^2} & \frac{\phi_{c0}}{1-\gamma_0} l'_T \\ \frac{\phi_{c0}}{1-\gamma_0} l_T & \phi_{c0} l_T l'_T + I_T \end{pmatrix}$$

with $\phi_{c0} = \frac{\sigma_{c0}^2}{\sigma_{v0}^2}$. The quasi log likelihood function is[3]

$$\ln L_r(\theta) = -\frac{n(T+1)}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|\Omega_{T+1}(\gamma, \phi_c)| - \frac{1}{2\sigma_v^2}\mathbf{U}'_{nT}(\gamma)(\Omega_{T+1}^{-1}(\gamma, \phi_c) \otimes I_n)\mathbf{U}_{nT}(\gamma), \tag{A.16}$$

where $\theta = (\gamma, \sigma_v^2, \phi_c)'$ and $U_{nT}(\gamma) = [Y'_{n0}, (Y_{n1} - \gamma Y_{n0})', \cdots, (Y_{nT} - \gamma Y_{n,T-1})']'$.

By the decomposition of the quasi log likelihood of the random effects model, the quasi log likelihood of the within equations is

$$\ln L_w(\gamma, \sigma_v^2) = -\frac{nT}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|H_T(\gamma)| - \frac{1}{2\sigma_v^2}(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1})'(H_T^{-1}(\gamma)\otimes I_n)(\Delta\mathbf{Y}_{nT} - \gamma\Delta\mathbf{Y}_{n,T-1}),$$

[3]Under the normality assumption of disturbances, Alvarez and Arellano (2003) consider a log likelihood function of $[Y'_{n1}, \dots, Y'_{nT}]'$ given Y_{n0} and show the consistency and asymptotic normality of their ML estimator. Since the log likelihood function of $[Y'_{n0}, \dots, Y'_{nT}]'$ is equal to sum of that of Y_{n0} and that of $[Y'_{n1}, \dots, Y'_{nT}]'$ given Y_{n0}, the MLE derived with our log likelihood function will be asymptotically more efficient than that in Alvarez and Arellano (2003) under the same regularity conditions.

where $\mathbf{\Delta Y}_{nT} = [\Delta Y'_{n1}, \ldots, \Delta Y'_{nT}]'$ and $\mathbf{\Delta Y}_{n,T-1} = [0, \Delta Y'_{n1}, \ldots, \Delta Y'_{n,T-1}]'$, and the quasi log likelihood of the between equation is

$$\ln L_b(\gamma, \sigma_\xi^2) = -\frac{n}{2}\ln(2\pi\sigma_\xi^2) - \frac{1}{2\sigma_\xi^2}\xi'_n(\gamma)\xi_n(\gamma),$$

where $\sigma_\xi^2 = \frac{\sigma_c^2}{(1-\gamma)^2} + \frac{\sigma_v^2}{(1-\gamma)(T+1-(T-1)\gamma)}$ and

$$\begin{aligned}\xi_n(\gamma) &= Y_{n0} + \frac{1}{T+1-(T-1)\gamma}\Big[T\Delta Y_{n1} + \sum_{t=2}^{T}(T+1-t)(\Delta Y_{nt} - \gamma\Delta Y_{n,t-1})\Big]\\ &= \frac{1}{T+1-(T-1)\gamma}\Big(\sum_{t=0}^{T} Y_{nt} - \gamma\sum_{t=1}^{T-1} Y_{nt}\Big).\end{aligned}$$

From the between model, it is apparent that only γ and σ_ξ^2 can be identified but not for σ_c^2 and σ_v^2. Thus, for convenience, it is desirable to reparameterize the parameter vector $(\gamma, \sigma_v^2, \phi_c)$ in the random effects model to the vector $(\gamma, \sigma_v^2, \sigma_\xi^2)$. Because $\sigma_\xi^2 = \sigma_v^2[\frac{\phi_c}{(1-\gamma)^2} + \frac{1}{(1-\gamma)(T+1-(T-1)\gamma)}]$, one has $\phi_c = \frac{\sigma_\xi^2}{\sigma_v^2}(1-\gamma)^2 - \frac{(1-\gamma)}{T+1-(T-1)\gamma}$. Thus the quasi log likelihood of the random effects model in terms of $(\gamma, \sigma_v^2, \sigma_\xi^2)$ is simply $\ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2) = \ln L_r(\gamma, \sigma_v^2, \phi_c(\gamma, \sigma_v^2, \sigma_\xi^2))$, and

$$\ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2) = \ln L_w(\gamma, \sigma_v^2) + \ln L_b(\gamma, \sigma_\xi^2).$$

It follows that the scores of $\ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2)$ of the random effects model can be written as sums of scores from the within and between equations, i.e.,

$$\begin{aligned}\frac{\partial \ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2)}{\partial\gamma} &= \frac{\partial \ln L_w(\gamma, \sigma_v^2)}{\partial\gamma} + \frac{\partial \ln L_b(\gamma, \sigma_\xi^2)}{\partial\gamma},\\ \frac{\partial \ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2)}{\partial\sigma_v^2} &= \frac{\partial \ln L_w(\gamma, \sigma_v^2)}{\partial\sigma_v^2},\\ \frac{\partial \ln L_r(\gamma, \sigma_v^2, \sigma_\xi^2)}{\partial\sigma_\xi^2} &= \frac{\partial \ln L_b(\gamma, \sigma_\xi^2)}{\partial\sigma_\xi^2}.\end{aligned}$$

The QMLEs of σ_v^2 and σ_ξ^2 given γ are solved respectively from $\frac{\partial \ln L_w(\gamma, \hat\sigma_v^2)}{\partial\sigma_v^2} = 0$ and $\frac{\partial \ln L_b(\gamma, \hat\sigma_\xi^2)}{\partial\sigma_\xi^2} = 0$. They are simply

$$\hat\sigma_v^2(\gamma) = \frac{1}{nT}(\mathbf{\Delta Y}_{nT} - \gamma\mathbf{\Delta Y}_{n,T-1})'(H_T^{-1}(\gamma)\otimes I_n)(\mathbf{\Delta Y}_{nT} - \gamma\mathbf{\Delta Y}_{n,T-1}) \quad \text{and} \quad \hat\sigma_\xi^2(\gamma) = \frac{1}{n}\xi'_n(\gamma)\xi_n(\gamma).$$

For the scores with γ, we have

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

and

$$\frac{\partial \ln L_b(\gamma,\sigma_\xi^2)}{\partial\gamma} = -\frac{1}{\sigma_\xi^2}\frac{\partial \xi_n'(\gamma)}{\partial\gamma}\xi_n(\gamma) = -\frac{1}{\sigma_\xi^2[T+1-(T-1)\gamma]^3}Y_{nb}'\Big(\sum_{t=0}^{T}Y_{nt}-\gamma\sum_{t=1}^{T-1}Y_{nt}\Big),$$

where

$$Y_{nb} = \sum_{t=1}^{T-1}(Y_{n0}+Y_{nT}-2Y_{nt}) = -\sum_{t=1}^{T}(T+1-2t)\Delta Y_{nt},$$

because $\frac{\partial\xi_n(\gamma)}{\partial\gamma} = \frac{1}{[T+1-(T-1)\gamma]^2}Y_{nb}$. While the score $\frac{\partial \ln L_r(\gamma,\sigma_v^2,\sigma_\xi^2)}{\partial\gamma}$ as a sum of the two corresponding scores of the within and between equations is rather nonlinear, one may directly consider implied moments from the two separate score vectors of the within and between equations. These scores motivate three moments:

$$g_{n1}(\gamma) = \frac{1}{n}\Delta\mathbf{Y}_{n,T-1}'(H_T^{-1}(\gamma)\otimes I_n)(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1}),$$
$$g_{n2}(\gamma) = \frac{1}{n}(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1})'\big(\Phi_T(\gamma)\otimes I_n\big)(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1}),$$

where $\Phi_T(\gamma) = H_T^{-1}(\gamma)\frac{\partial H_T(\gamma)}{\partial\gamma}H_T^{-1}(\gamma) - \frac{1}{T}\operatorname{tr}\big(\frac{\partial H_T(\gamma)}{\partial\gamma}H_T^{-1}(\gamma)\big)H_T^{-1}(\gamma)$, and

$$g_{n3}(\gamma) = \frac{1}{n}Y_{nb}'\Big(\sum_{t=0}^{T}Y_{nt}-\gamma\sum_{t=1}^{T-1}Y_{nt}\Big).$$

The moments g_{n1} and g_{n2} are from the stationary within model. Correspondingly, a family of moments for GMM estimation can be based on the quadratic moments $\frac{1}{n}e_{nT}'(\gamma)(B_T'(\gamma)A_TB_T(\gamma)\otimes I_n)e_{nT}(\gamma)$, for some $T\times T$ constant (or consistently estimated) matrix A_T as in (2.12) such that $\operatorname{tr}(H_TA_TF_T)=0$. Corresponding to g_{n3}, define $\xi_n^*(\gamma) = \frac{1}{T}(\sum_{t=0}^{T}Y_{nt}-\gamma\sum_{t=1}^{T-1}Y_{nt})$. At the true γ_0, $\xi_n^*(=\xi_n^*(\gamma_0))$ is proportional to the residual ξ_n of the between equation model, so ξ_n^* and ΔY_{nt} for $t=1,\cdots,T$ are perpendicular and hence uncorrelated by construction. Also, as Y_{nb} is a linear combination of elements of $\Delta\mathbf{Y}_{nT} = (\Delta Y_{n1}',\cdots,\Delta Y_{nT}')'$, this suggests a class of IV moments with IV matrices of the form $Q_n^* = (c_T\otimes I_n)\Delta\mathbf{Y}_{nT}$ for some constant (or consistently estimated) $1\times T$ row vector c_T. The class of moments consists of

$$g_n(\gamma) = \frac{1}{n}\begin{pmatrix} e_{nT}'(\gamma)(B_T'(\gamma)A_{1T}B_T(\gamma)\otimes I_n)e_{nT}(\gamma)\\ \vdots \\ e_{nT}'(\gamma)(B_T'(\gamma)A_{m_1T}B_T(\gamma)\otimes I_n)e_{nT}(\gamma)\\ \Delta\mathbf{Y}_{nT}'(c_{1T}'\otimes I_n)\xi_n^*(\gamma)\\ \vdots \\ \Delta\mathbf{Y}_{nT}'(c_{m_2T}'\otimes I_n)\xi_n^*(\gamma)\end{pmatrix},$$

which has a total of arbitrarily finite $m = m_1+m_2$ moments.[4] From Section 4, the best choice of

[4] It is of interest to note that, for the pure DPD model, $\Delta\mathbf{Y}_{nT}$ is directly related to e_{nT}. From the definitions of $\Delta\mathbf{Y}_{nT}$, $\Delta\mathbf{Y}_{n,T-1}$ and $e_{nT}(\gamma)$, we have the relation $\Delta\mathbf{Y}_{nT} = \gamma\Delta\mathbf{Y}_{n,T-1}+e_{nT}(\gamma) = [\gamma F_T(\gamma)\otimes I_n + I_{nT}]e_{nT}(\gamma) = (F_T^*(\gamma)\otimes I_n)e_{nT}(\gamma)$, where $F_T^*(\gamma) = \begin{pmatrix} 1 & & & \\ \gamma & 1 & & \\ \vdots & \vdots & \ddots & \\ \gamma^{T-1} & \gamma^{T-2} & \cdots & 1\end{pmatrix}$. Hence, $Q_n^* = (c_T\otimes I_n)\Delta\mathbf{Y}_{nT} = (c_TF_T^*\otimes I_n)e_{nT}$ with $F_T^* = F_T^*(\gamma_0)$.

$A_{1T}, \ldots, A_{m_1 T}$ under normality has $m_1 = 1$ with $A_{1T} = K_T - \frac{1}{T}\operatorname{tr}(K_T)I_T = B_T F_T B_T^{-1} - (B_{\gamma T}B_T^{-1} - \frac{1}{T}\operatorname{tr}(B_{\gamma T}B_T^{-1})I_T)$, implied from the score of the within equation. From the score of the between equation, we see that the best choice of $c_{1T}, \ldots, c_{m_2 T}$ has $m_2 = 1$ with $(c_T \otimes I_n)\Delta\mathbf{Y}_{nT} = -\sum_{t=1}^{T}(T+1-2t)\Delta Y_{nt} = Y_{nb}$.

A.3.2 Random effects stationary DPD with exogenous variables

For the random effects DPD model with exogenous variables in (A.6), we consider the case that the process started from the infinite past. According to the analysis in Section A.2,

$$y_{i0} = x_{i0}\beta_0 + \sum_{j=1}^{\infty}\gamma_0^j x_{i,-j}\beta_0 + \frac{1}{1-\gamma_0}z_i b_0 + \frac{1}{1-\gamma_0}c_{i0} + \sum_{j=0}^{\infty}\gamma_0^j v_{i,-j}.$$

With the assumptions $\mathrm{E}(x_{i,-j}|z_i, \vec{x}_i) = z_i\varrho_j + \vec{x}_i\pi_j$ for $j \geq 1$,

$$y_{i0} = z_i\alpha^{(1)} + \vec{x}_i\alpha^{(2)} + \xi_{i0},$$

where $\alpha^{(1)} = \sum_{j=1}^{\infty}\gamma_0^j\varrho_j\beta_0 + \frac{1}{1-\gamma_0}b_0$, $\alpha^{(2)} = \sum_{j=1}^{\infty}\gamma_0^j\pi_j\beta_0 + [\beta_0', 0, \ldots, 0]'$, $\xi_{i0} = \frac{1}{1-\gamma_0}c_{i0} + \sum_{j=0}^{\infty}\gamma_0^j v_{i,-j} + p_i$, and $p_i = \sum_{j=1}^{\infty}\gamma_0^j[x_{i,-j} - \mathrm{E}(x_{i,-j}|z_i, \vec{x}_i)]\beta_0$ is the prediction error. Note that in general, there are too many ϱ's and π's to be identifiable, but $\alpha^{(1)}$ and $\alpha^{(2)}$, which summarize the influence of exogenous variables at time 0 and their predictions for y_{i0}, are identifiable. Denote the $(T+1)$-dimensional vector of the overall disturbances for the individual i as

$$\mathbf{U}_{iT} = \Big[p_i + \frac{c_i}{1-\gamma_0} + \sum_{j=0}^{\infty}\gamma_0^j v_{i,-j},\ c_i + v_{i1},\ \cdots,\ c_i + v_{iT}\Big]',$$

which has the variance $\sigma_{v0}^2\Omega_{T+1}$, where

$$\Omega_{T+1} = \begin{pmatrix} \phi_{p0} + \frac{\phi_{c0}}{(1-\gamma_0)^2} + \frac{1}{1-\gamma_0^2} & \frac{\phi_{c0}}{1-\gamma_0}l_T' \\ \frac{\phi_{c0}}{1-\gamma_0}l_T & \phi_{c0}l_T l_T' + I_T \end{pmatrix},$$

with $\phi_{c0} = \frac{\sigma_{c0}^2}{\sigma_{v0}^2}$, $\phi_{p0} = \frac{\sigma_{p0}^2}{\sigma_{v0}^2}$ and σ_{p0}^2 being the variance of p_i. Thus, the parameters for estimation for the random effects model consist of γ, b, β, σ_v^2, ϕ_c and ϕ_p. The γ is not only involved as a coefficient of the time lag $y_{i,t-1}$ from the score of the quasi log likelihood of the random effects model but also in the variance matrix Ω_{T+1}.

With the model decomposition, the corresponding within equations have

$$\Delta Y_{nt} = \gamma_0\Delta Y_{n,t-1} + \Delta X_{nt}\beta_0 + \Delta V_{nt}$$

for $t = 2, \cdots, T$, and the first difference at $t = 1$ gives

$$\Delta Y_{n1} = (\gamma_0 - 1)(Z_n\alpha^{(1)} + \vec{\mathbf{X}}_{nT}\alpha^{(2)}) + X_{n1}\beta_0 + Z_n b_0 + \xi_{n1} = Z_n\kappa_{w1} + \vec{\mathbf{X}}_{nT}\kappa_{w2} + \Xi_n,$$

where $\kappa_{w1} = (\gamma_0 - 1)\alpha^{(1)} + b_0 = (\gamma_0 - 1)\sum_{j=1}^{\infty} \gamma_0^j \varrho_j \beta_0$, $\kappa_{w2} = (\gamma_0 - 1)\alpha^{(2)} + [0, \beta_0', 0, \ldots, 0]' = (\gamma_0 - 1)\sum_{j=1}^{\infty} \gamma_0^j \pi_j \beta_0 + [(\gamma_0 - 1)\beta_0', \beta_0', 0, \ldots, 0]'$, $\Xi_n = [\xi_{11}, \ldots, \xi_{n1}]'$ with $\xi_{i1} = (\gamma_0 - 1)p_i + \sum_{j=0}^{\infty} \gamma_0^j \Delta v_{i,1-j}$, and $\vec{\mathbf{X}}_{nT} = [\vec{x}_1', \ldots, \vec{x}_n']'$. There is a one-to-one correspondence between $[\alpha^{(1)}, \alpha^{(2)}]$ and $[\kappa_{w1}, \kappa_{w2}]$. The variance of the disturbance vector $(\xi_{i1}, \Delta v_{i2}, \cdots, \Delta v_{iT})'$ is $\sigma_{v0}^2 H_T(\omega_0)$, where $\omega = \frac{2}{1+\gamma} + (\gamma - 1)^2 \phi_p$. Because ϕ_p is an unknown parameter, ω in this variance matrix may be regarded as a free parameter because it is not constrained by γ.

For the between equation, it is a cross sectional equation with n observations aggregated. Let $\mu_{n0} = Z_n \alpha^{(1)} + \vec{\mathbf{X}}_{nT} \alpha^{(2)}$ and $\mu_{n1} = Z_n \kappa_{w1} + \vec{\mathbf{X}}_{nT} \kappa_{w2}$. For any possible values of parameters, denote $\kappa_{b1}(\gamma, b, \omega, \kappa_{w1}) = -\frac{1}{[1+T(\omega-1)](1-\gamma)} \kappa_{w1} + \frac{1}{1-\gamma} b$,

$$\kappa_{b2}(\gamma, \beta, \omega, \kappa_{w2}) = -\frac{1}{[1 + T(\omega - 1)](1 - \gamma)} \kappa_{w2} + \frac{1}{[1 + T(\omega - 1)](1 - \gamma)} \big[T(\omega - 1)\beta', \beta', 0, \ldots, 0\big]',$$

and $\phi_b(\gamma, \omega) = \frac{\omega - 1}{[1+T(\omega-1)](1-\gamma)}$. At the true parameters, they are reduced respectively to the values κ_{b1}, κ_{b2} and ϕ_b. As derived in Lee and Yu (2018) and shown in Section D.3, the between equation is

$$\begin{aligned} Y_{n0} &= \mu_{n0} - \frac{1}{1 + T(\omega_0 - 1)} \Big[(1 - \gamma_0)\phi_{p0} + \frac{1}{1 + \gamma_0}\Big] \Big[T(\Delta Y_{n1} - \mu_{n1}) + \sum_{t=2}^{T} (T + 1 - t)(\Delta Y_{nt} - \gamma_0 \Delta Y_{n,t-1} - \Delta X_{nt}\beta_0)\Big] \\ &\quad + \tilde{\xi}_{n0} \\ &= Z_n \kappa_{b1} + \vec{\mathbf{X}}_{nT} \kappa_{b2} - \phi_b \Big[T(\Delta Y_{n1} - \Delta X_{n1}\beta_0) + \sum_{t=2}^{T} (T + 1 - t)(\Delta Y_{nt} - \gamma_0 \Delta Y_{n,t-1} - \Delta X_{nt}\beta_0)\Big] + \tilde{\xi}_{n0}, \end{aligned}$$

where $\tilde{\xi}_{n0}$ has zero mean and variance matrix $\sigma_{\tilde{\xi}}^2 \otimes I_n$ with

$$\begin{aligned} \sigma_{\tilde{\xi}}^2 &= \sigma_{v0}^2 \Big[\frac{\phi_{c0}}{(1 - \gamma_0)^2} + \frac{1}{1 - \gamma_0^2} + \phi_{p0} - \big(\frac{1}{1 + \gamma_0} + (1 - \gamma_0)\phi_{p0}\big)^2 \frac{T}{1 + T(\omega_0 - 1)}\Big] \\ &= \sigma_{v0}^2 \Big[\frac{\phi_{c0}}{(1 - \gamma_0)^2} + \frac{1}{1 - \gamma_0^2} + \frac{1}{(1 - \gamma_0)^2}(\omega_0 - \frac{2}{1 + \gamma_0}) - \frac{T(\omega_0 - 1)^2}{(1 - \gamma_0)^2 [1 + T(\omega_0 - 1)]}\Big] \end{aligned}$$

as $\phi_{p0} = \frac{1}{(1-\gamma_0)^2}(\omega_0 - \frac{2}{1+\gamma_0})$.

The random effects model has the parameter vector $(\gamma, b, \beta, \kappa_{w1}, \kappa_{w2}, \sigma_v^2, \phi_p, \phi_c)$. As there is a one-to-one correspondence between the vectors $(\gamma, b, \beta, \kappa_{w1}, \kappa_{w2}, \sigma_v^2, \phi_p, \phi_c)$ and $(\gamma, b, \beta, \kappa_{w1}, \kappa_{w2}, \sigma_v^2, \omega, \sigma_{\tilde{\xi}}^2)$, it is more convenient to adopt the later as the parameters of the model. Let $\theta_1 = (\gamma, \beta', \kappa_{w1}', \kappa_{w2}', \omega)'$. Then,

$$\ln L_r(\theta_1, b, \sigma_v^2, \sigma_{\tilde{\xi}}^2) = \ln L_w(\theta_1, \sigma_v^2) + \ln L_b(\theta_1, b, \sigma_{\tilde{\xi}}^2).$$

For the between equation alone, not all the parameters listed in $\ln L_b$ can be identifiable, but only some functions of them are. However, for setting up the SGMM estimation, this is not a concern as we are only interested in the scores of the sum of $\ln L_w$ and $\ln L_b$; equivalently the scores of the random effects model.[5]

[5]However, for a Hausman test, one has to care about parameters which can be identified in each of the models in order to get a correct degree of freedom for a Hausman test.

The advantage of using the model decomposition is that the variance matrix $H_T(\omega)$ of the within model is a function of the free parameter ω, and the variance of the between equation is simply a scalar, which is free after reparameterization. As the main structural parameters of interest in the random effects model are γ, β and b, one might design asymptotically efficient GMM estimation for them while regarding the parameters ω, $\sigma^2_{\tilde{\xi}}$ and some others such as κ_{w1} and κ_{w2} as nuisance parameters.

The within equations can be rewritten as

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

and

$$\Delta Y_{n1} = Z_n \kappa_{w1} + \vec{\mathbf{X}}_{nT}\kappa_{w2} + \Xi_n.$$

The quasi log likelihood function of the within equations is then

$$\ln L_w(\theta_1, \sigma_v^2) = -\frac{nT}{2}\ln(2\pi\sigma_v^2) - \frac{n}{2}\ln|H_T(\omega)| \\ - \frac{1}{2\sigma_v^2}\begin{pmatrix} \Delta Y_{n1} - Z_n\kappa_{w1} - \vec{\mathbf{X}}_{nT}\kappa_{w2} \\ \Delta Y_{n2} - \gamma\Delta Y_{n1} - \Delta X_{n2}\beta \\ \vdots \\ \Delta Y_{nT} - \gamma\Delta Y_{n,T-1} - \Delta X_{nT}\beta \end{pmatrix}' (H_T^{-1}(\omega)\otimes I_n) \begin{pmatrix} \Delta Y_{n1} - Z_n\kappa_{w1} - \vec{\mathbf{X}}_{nT}\kappa_{w2} \\ \Delta Y_{n2} - \gamma\Delta Y_{n1} - \Delta X_{n2}\beta \\ \vdots \\ \Delta Y_{nT} - \gamma\Delta Y_{n,T-1} - \Delta X_{nT}\beta \end{pmatrix}.$$

For the between equations, the quasi log likelihood function is

$$\ln L_b(\theta_1, b, \sigma^2_{\tilde{\xi}}) = -\frac{n}{2}\ln(2\pi\sigma^2_{\tilde{\xi}}) - \frac{1}{2\sigma^2_{\tilde{\xi}}}\tilde{\xi}_n'(\theta_1, b)\tilde{\xi}_n(\theta_1, b),$$

where

$$\tilde{\xi}_n(\theta_1, b) = Y_{n0} - Z_n\kappa_{b1}(\gamma, b, \omega, \kappa_{w1}) - \vec{\mathbf{X}}_{nT}\kappa_{b2}(\gamma, \beta, \omega, \kappa_{w2}) \\ + \phi_b(\gamma, \omega)\Big[T(\Delta Y_{n1} - \Delta X_{n1}\beta) + \sum_{t=2}^{T}(T+1-t)(\Delta Y_{nt} - \Delta Y_{n,t-1}\gamma - \Delta X_{nt}\beta)\Big].$$

Even though the score vector of $\ln L_r$ will be a summation of the corresponding score vectors of $\ln L_w$ and $\ln L_b$, the scores of those quasi log likelihood functions with their composite coefficients provide valid but simpler moments. Those moments can be used as basic moment relations for GMM estimation, and their optimal weighting matrix would provide proper combinations, which are asymptotically equivalent to the scores of the random effects likelihood function under the normal distribution.

One may derive moments from the within model via its scores. To simplify notations, denote $\Xi_n(\kappa_{w1}, \kappa_{w2}) = \Delta Y_{n1} - Z_n\kappa_{w1} - \vec{\mathbf{X}}_{nT}\kappa_{w2}$ and $\Delta V_{nt}(\gamma, \beta) = \Delta Y_{nt} - \Delta Y_{n,t-1}\gamma - X_{nt}\beta$ for $t = 2, \cdots, T$. The following moments motivated from the scores of the within equations correspond to, respectively, $[\gamma, \beta', \kappa_{w1}, \kappa_{w2}]'$ and ω:

$$[\Delta\mathbf{Z}_{nT}, \Upsilon_{nT}]'\big(H_T^{-1}(\omega)\otimes I_n\big)e_{nT}(\kappa_w, \gamma, \beta),$$

and

$$e'_{nT}(\kappa_w,\gamma,\beta)\big(\Phi_T(\omega)\otimes I_n\big)e_{nT}(\kappa_w,\gamma,\beta),$$

where $e_{nT}(\kappa_w,\gamma,\beta) = [\Xi'_n(\kappa_{w1},\kappa_{w2}),\Delta V'_{n2}(\gamma,\beta),\ldots,\Delta V'_{nT}(\gamma,\beta)]'$, $\Delta\mathbf{Z}_{nT} = \begin{pmatrix} 0 & 0 \\ \Delta Y_{n1} & \Delta X_{n2} \\ \vdots & \vdots \\ \Delta Y_{n,T-1} & \Delta X_{nT}\end{pmatrix}$, $\Upsilon_{nT} = \begin{pmatrix} Z_n & \vec{\mathbf{X}}_{nT} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0\end{pmatrix}$, and $\Phi_T(\omega) = H_T^{-1}(\omega)\frac{\partial H_T(\omega)}{\partial\omega}H_T^{-1}(\omega) - \frac{1}{T}\operatorname{tr}\big(\frac{\partial H_T(\omega)}{\partial\omega}H_T^{-1}(\omega)\big)H_T^{-1}(\omega)$.

For the between model, its scores take the form $\frac{\partial\tilde{\xi}'_n(\theta_1,b)}{\partial\theta}\tilde{\xi}_n(\theta_1,b)$. The moments corresponding to those with respect to κ_{w1}, κ_{w2}, $\phi_b(\gamma,\omega)$, $(\phi_b(\gamma,\omega)\beta)$, and $(\phi_b(\gamma,\omega)\gamma)$ are, respectively,

$$Z'_n\tilde{\xi}_n(\theta_1,b),\quad \vec{\mathbf{X}}'_{nT}\tilde{\xi}_n(\theta_1,b),\quad \Big(\sum_{t=1}^{T}(T+1-t)\Delta Y_{nt}\Big)'\tilde{\xi}_n(\theta_1,b),$$
$$\Big(\sum_{t=1}^{T}(T+1-t)\Delta X_{nt}\Big)'\tilde{\xi}_n(\theta_1,b),\ \text{and}\ \Big(\sum_{t=2}^{T}(T+1-t)\Delta Y_{n,t-1}\Big)'\tilde{\xi}_n(\theta_1,b).$$

As $\frac{1}{1-\gamma}$ and its products with b and β appear in $\tilde{\xi}_n(\theta_1,b)$ in addition to other nuisance parameters, the above moments are nonlinear in the structural parameters γ, b and β. Given the moments $\vec{\mathbf{X}}'_{nT}\tilde{\xi}_n(\theta_1,b)$, the moments $[\sum_{t=1}^{T}(T+1-t)\Delta X_{nt}]'\tilde{\xi}_n(\theta_1,b)$ are redundant, because $\sum_{t=1}^{T}(T+1-t)\Delta X_{nt}$ is linear of $X_{n0},\cdots,X_{nT}$.

With the moments derived from the within and between models, we may implement an SGMM estimation with an initial consistent estimate $\tilde{\theta} = [\tilde{\gamma},\tilde{\beta}',\tilde{\kappa}'_{w1},\tilde{\kappa}'_{w2},\tilde{\omega},\tilde{b}']'$ of θ_0. We consider the case that the SGMM efficiently estimates $\delta_1 = [\gamma,\beta',b']'$ by treating $\delta_2 = [\kappa'_{w1},\kappa'_{w2},\omega]'$ as a nuisance parameter vector. Let

$$g_{nT,1}(\delta_1,\delta_2) = \begin{pmatrix} \Delta\mathbf{Z}'_{nT}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\kappa_w,\gamma,\beta) \\ Z'_n\tilde{\xi}_n(\theta_1,b) \\ \vec{\mathbf{X}}'_{nT}\tilde{\xi}_n(\theta_1,b) \\ \big(\sum_{t=1}^{T}(T+1-t)\Delta Y_{nt}\big)'\tilde{\xi}_n(\theta_1,b) \\ \big(\sum_{t=2}^{T}(T+1-t)\Delta Y_{n,t-1}\big)'\tilde{\xi}_n(\theta_1,b)\end{pmatrix}, \tag{A.17}$$

and

$$g_{nT,2}(\delta_1,\delta_2) = \begin{pmatrix} \Upsilon'_{n,T+1}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\kappa_w,\gamma,\beta) \\ e'_{nT}(\kappa_w,\gamma,\beta)(\Phi_T(\omega)\otimes I_n)e_{nT}(\kappa_w,\gamma,\beta)\end{pmatrix} \tag{A.18}$$

where $g_{nT,2}(\delta_1,\delta_2)$ contains moment conditions derived from the scores of the within model with respect to δ_2. Denote $\hat{C}_{nT} = \frac{\partial g_{nT,1}(\tilde{\delta}_1,\tilde{\delta}_2)}{\partial\delta'_2}\big(\frac{\partial g_{nT,2}(\tilde{\delta}_1,\tilde{\delta}_2)}{\partial\delta'_2}\big)^{-1}$ and let $\hat{\Sigma}_{nT}$ be a consistent estimator of the limiting variance matrix of $\sqrt{n}[g'_{nT,1}(\delta_{10},\delta_{20}),g'_{nT,2}(\delta_{10},\delta_{20})]'$. The SGMM estimator $\hat{\delta}_1$ of δ_{10} is derived from

$$\min_{\delta_1}[g_{nT,1}(\delta_1,\tilde{\delta}_2) - \hat{C}_{nT}g_{nT,2}(\delta_1,\tilde{\delta}_2)]'\big([I_{k_1},-\hat{C}_{nT}]\hat{\Sigma}_{nT}[I_{k_1},-\hat{C}_{nT}]'\big)^{-1}[g_{nT,1}(\delta_1,\tilde{\delta}_2) - \hat{C}_{nT}g_{nT,2}(\delta_1,\tilde{\delta}_2)],$$

where k_1 is the number of moments in $g_{nT,1}(\delta_1, \delta_2)$. Since the difference between the total number of moments from the within and between models and the number of the moments in $g_{nT,1}(\delta_1, \tilde{\delta}_2) - \hat{C}_{nT} g_{nT,2}(\delta_1, \tilde{\delta}_2)$ is equal to the number of nuisance parameters in δ_2, the SGMM estimator $\hat{\delta}_1$ is asymptotically as efficient as the joint GMM estimator of δ_{10} that uses all the moments from the within and between models, i.e., the QML estimator of δ_{10} in the random effects model.

B Proofs of theorems

We only prove Theorems 3–5, since the proofs of the rest of theorems are similar to them.

B.1 Proof of Theorem 3

(i) We first prove the uniform convergence that $\sup_{\theta_3 \in \Theta_3} \|g_{nT}(\theta_3) - \mathrm{E}[g_{nT}(\theta_3)]\| = o_p(1)$. As $\Delta \mathbf{Y}_{nT} = e_{nT} + \gamma_0 \Delta \mathbf{Y}_{n,T-1} + \mathbf{X}^*_{nT} \binom{\alpha_0}{\beta_0}$ and $\Delta \mathbf{Y}_{n,T-1} = (F_T \otimes I_n)(e_{nT} + \mathbf{X}^*_{nT} \binom{\alpha_0}{\beta_0})$,

$$e_{nT}(\alpha, \delta) = ((I_T + (\gamma_0 - \gamma) F_T) \otimes I_n) e_{nT} + f_{nT}(\alpha, \delta), \tag{B.1}$$

where $f_{nT}(\alpha, \delta) = \mathbf{X}^*_{nT} \binom{\alpha_0 - \alpha}{\beta_0 - \beta} + (\gamma_0 - \gamma)(F_T \otimes I_n) \mathbf{X}^*_{nT} \binom{\alpha_0}{\beta_0}$. Then

$$\begin{aligned} &\frac{1}{n} \mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta) - \frac{1}{n} \mathrm{E}[\mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta)] \\ &= \frac{1}{n} \mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT} + \frac{1}{n} (\gamma_0 - \gamma) \mathbf{X}^{*\prime}_{nT} (K_{jT} F_T \otimes I_n) e_{nT}. \end{aligned}$$

For a $T \times T$ matrix A_T with bounded elements and a column vector a conformable with $\mathbf{X}^*_{nT}$, $\mathrm{E}[\frac{1}{n} a' \mathbf{X}^{*\prime}_{nT} (A_T \otimes I_n) e_{nT} \cdot \frac{1}{n} e'_{nT} (A'_T \otimes I_n) \mathbf{X}^*_{nT} a] = \frac{\sigma^2_{v0}}{n^2} a' \mathbf{X}^{*\prime}_{nT} (A_T H_T A'_T \otimes I_n) \mathbf{X}^*_{nT} a \leq \frac{c \sigma^2_{v0}}{n^2} a' \mathbf{X}^{*\prime}_{nT} \mathbf{X}^*_{nT} a = o(1)$ for some constant c, since H_T is positive definite. Then $\frac{1}{n} \mathbf{X}^{*\prime}_{nT} (A_T \otimes I_n) e_{nT} = o_p(1)$. Thus, $\frac{1}{n} \mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta) - \frac{1}{n} \mathrm{E}[X^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta)] = o_p(1)$. Since $\frac{1}{n} \mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta) - \frac{1}{n} \mathrm{E}[\mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta)]$ is linear in γ, $\sup_{\theta_3 \in \Theta_3} \| \frac{1}{n} \mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta) - \frac{1}{n} \mathrm{E}[\mathbf{X}^{*\prime}_{nT} (K_{jT} \otimes I_n) e_{nT}(\alpha, \delta)] \| = o_p(1)$. For moments quadratic in $e_{nT}(\alpha, \delta)$,

$$\begin{aligned} \Xi_{nT}(\theta_3) &\equiv \frac{1}{n} e'_{nT}(\alpha, \delta)(B'_T(\omega) C_{jT} B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta) - \frac{1}{n} \mathrm{E}[e'_{nT}(\alpha, \delta)(B'_T(\omega) C_{jT} B_T(\omega) \otimes I_n) e_{nT}(\alpha, \delta)] \\ &= \frac{1}{n} f'_{nT}(\alpha, \delta)(B'_T(\omega) C^s_{jT} B_T(\omega) F^*_T \otimes I_n) e_{nT} \\ &\quad + \frac{1}{n} e'_{nT} (F^{*\prime}_T B'_T(\omega) C_{jT} B_T(\omega) F^*_T \otimes I_n) e_{nT} - \sigma^2_{v0} \operatorname{tr}[F^{*\prime}_T B'_T(\omega) C_{jT} B_T(\omega) F^*_T H_T], \end{aligned}$$

where $F^*_T = I_T + (\gamma_0 - \gamma) F_T$. By Lemma A.1 in Lin and Lee (2010),

$$\operatorname{Var}\big(\tfrac{1}{n} e'_{nT} (A_T \times I_n) e_{nT}\big) = \operatorname{Var}\big(\tfrac{1}{n} \mathbf{V}'_{nT} (D'_{T,T+1} A_T D_{T,T+1} \times I_n) \mathbf{V}_{nT}\big) = o(1)$$

for a $T\times T$ matrix A_T with bounded elements. Hence, $\Xi_{nT}(\theta_3)=o_p(1)$. Since the parameter space Θ_3 of θ_3 is bounded, it is straightforward to show that $\sup_{\theta_3\in\Theta_3}\|\frac{\partial \Xi_{nT}(\theta_3)}{\partial\theta_3}\|=O_p(1)$. Therefore, $\sup_{\theta_3\in\Theta_3}|\Xi_{nT}(\theta_3)|=o_p(1)$. It follows that $\sup_{\theta_3\in\Theta_3}\|g_{nT}(\theta_3)-\mathrm{E}[g_{nT}(\theta_3)]\|=o_p(1)$.

We next prove that $\lim_{n\to\infty}\mathrm{E}[g_{nT}(\theta_3)]$ is uniquely zero at $\theta_3=\theta_{30}$. With $e_{nT}(\alpha,\delta)$ in (B.1),

$$\frac{1}{n}\,\mathrm{E}[\mathbf{X}_{nT}^{*\prime}(K_{jT}\otimes I_n)e_{nT}(\alpha,\delta)]=\frac{1}{n}\mathbf{X}_{nT}^{*\prime}(K_{jT}\otimes I_n)\mathbf{X}_{nT}^{*}\begin{pmatrix}\alpha_0-\alpha\\ \beta_0-\beta\end{pmatrix}+\frac{1}{n}(\gamma_0-\gamma)\mathbf{X}_{nT}^{*\prime}(K_{jT}F_T\otimes I_n)\mathbf{X}_{nT}^{*}\begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix},$$

and

$$\begin{aligned}&\frac{1}{n}\,\mathrm{E}[e_{nT}^{\prime}(\alpha,\delta)(B_T^{\prime}(\omega)C_{jT}B_T(\omega)\otimes I_n)e_{nT}(\alpha,\delta)]\\&=\frac{1}{n}f_{nT}^{\prime}(\alpha,\delta)[B_T^{\prime}(\omega)C_{jT}B_T(\omega)\otimes I_n]f_{nT}(\alpha,\delta)+\sigma_{v0}^2\,\mathrm{tr}[(I_T+(\gamma_0-\gamma)F_T)^{\prime}B_T^{\prime}(\omega)C_{jT}B_T(\omega)(I_T+(\gamma_0-\gamma)F_T)H_T].\end{aligned}$$

Denote $g_{nT}(\theta_3)=[g_{nT,1}^{\prime}(\theta_3),g_{nT,2}^{\prime}(\theta_3)]^{\prime}$, where $g_{nT,2}(\theta_3)$ consists of the last m_2 elements of $g_{nT}(\theta_3)$. When $(\alpha_0^{\prime},\beta_0^{\prime})^{\prime}\neq 0$, since $\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\mathbf{X}_{nT}^{*\prime}(K_{1T}\otimes I_n)\mathbf{X}_{nT}^{*} & X_{nT}^{*\prime}(K_{1T}F_T\otimes I_n)\mathbf{X}_{nT}^{*}\binom{\alpha_0}{\beta_0}\\ \vdots & \vdots\\ \mathbf{X}_{nT}^{*\prime}(K_{m_1T}\otimes I_n)\mathbf{X}_{nT}^{*} & X_{nT}^{*\prime}(K_{m_1T}F_T\otimes I_n)\mathbf{X}_{nT}^{*}\binom{\alpha_0}{\beta_0}\end{pmatrix}$ has full column rank, $\lim_{n\to\infty}\mathrm{E}[g_{nT,1}(\theta_3)]=0$ implies that $(\alpha^{\prime},\beta^{\prime},\gamma)^{\prime}=(\alpha_0^{\prime},\beta_0^{\prime},\gamma_0)^{\prime}$; when $(\alpha_0^{\prime},\beta_0^{\prime})^{\prime}=0$, since $\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix}\mathbf{X}_{nT}^{*\prime}(K_{1T}\otimes I_n)\mathbf{X}_{nT}^{*}\\ \vdots\\ \mathbf{X}_{nT}^{*\prime}(K_{m_1T}\otimes I_n)\mathbf{X}_{nT}^{*}\end{pmatrix}$ has full column rank, $\lim_{n\to\infty}\mathrm{E}[g_{nT,1}(\theta_3)]=0$ implies that $(\alpha^{\prime},\beta^{\prime})^{\prime}=(\alpha_0^{\prime},\beta_0^{\prime})^{\prime}=0$. In both cases, $\lim_{n\to\infty}\mathrm{E}[g_{nT,1}(\theta_3)]=0$ implies that $f_{nT}(\alpha,\delta)=0$. Note that

$$\begin{aligned}&\mathrm{tr}[(I_T+(\gamma_0-\gamma)F_T)^{\prime}B_T^{\prime}(\omega)C_{jT}B_T(\omega)(I_T+(\gamma_0-\gamma)F_T)H_T]\\&=\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)H_T]+(\gamma_0-\gamma)\,\mathrm{tr}[F_T^{\prime}B_T^{\prime}(\omega)C_{jT}^{s}B_T(\omega)H_T]+(\gamma_0-\gamma)^2\,\mathrm{tr}[F_T^{\prime}B_T^{\prime}(\omega)C_{jT}B_T(\omega)F_TH_T]\end{aligned}$$

As C_{jT} has a zero trace, $\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)H_T(\omega)]=\mathrm{tr}(C_{jT})=0$. Thus,

$$\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)H_T]=\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)(H_T-H_T(\omega))]=(\omega_0-\omega)\,\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)J_T].$$

With $B_T=D^{-1/2}A$, where $D^{-1/2}A$ is in (2.8), $\mathrm{tr}[B_T^{\prime}(\omega)C_{jT}B_T(\omega)J_T]=\mathrm{tr}[C_{jT}\,\mathrm{diag}(d_T(\omega))l_Tl_T^{\prime}\,\mathrm{diag}(d_T(\omega))]=d_T(\omega)C_{jT}d_T^{\prime}(\omega)$, where $d_T(\omega)$ is in Assumption 3.4. It follows that $\lim_{n\to\infty}\mathrm{E}[g_{nT}(\theta_3)]$ is uniquely zero at $\theta_3=\theta_{30}$ under Assumption 3.4.

We further prove that $\lim_{n\to\infty}\Sigma_{nT}$ is invertible under Assumption 3.3. Let $\mu_{3T}=\mathrm{diag}(\mu_{3u},\mu_{3v}l_T^{\prime})$, $c_T=\mathrm{diag}(\frac{1}{2}(\mu_{4u}-\sigma_{v0}^4-\mu_{3u}^2/\sigma_{v0}^2)^{1/2},\frac{1}{2}(\mu_{4v}-\sigma_{v0}^4-\mu_{3v}^2/\sigma_{v0}^2)^{1/2}l_T^{\prime})$,

$$\bar{C}_{rT}^{s}=\frac{\sqrt{2}\sigma_{v0}^2}{2}[D_{T,T+1}^{\prime}B_T^{\prime}C_{rT}^{s}B_TD_{T,T+1}-\mathrm{diag}(D_{T,T+1}^{\prime}B_T^{\prime}C_{rT}^{s}B_TD_{T,T+1})]+c_T\cdot\mathrm{diag}(D_{T,T+1}^{\prime}B_T^{\prime}C_{rT}^{s}B_TD_{T,T+1}),$$

$\Delta_{1T}=[\mathrm{vec}(\bar{C}_{1T}^{s}),\ldots,\mathrm{vec}(\bar{C}_{m_2T}^{s})]$, $\Delta_{2nT}=[\frac{\sigma_{v0}}{\sqrt{n}}(D_{T,T+1}^{\prime}K_{1T}^{\prime}\otimes I_n)\mathbf{X}_{nT}^{*},\ldots,\frac{\sigma_{v0}}{\sqrt{n}}(D_{T,T+1}^{\prime}K_{m_1T}^{\prime}\otimes I_n)\mathbf{X}_{nT}^{*}]$, and

$$\Delta_{3nT}=\Big[\frac{1}{2\sqrt{n}\sigma_{v0}}\mu_{3T}\,\mathrm{vec}_{\mathrm{D}}(D_{T,T+1}^{\prime}B_T^{\prime}C_{1T}^{s}B_TD_{T,T+1})\otimes l_n,\ldots,\frac{1}{2\sqrt{n}\sigma_{v0}}\mu_{3T}\,\mathrm{vec}_{\mathrm{D}}(D_{T,T+1}^{\prime}B_T^{\prime}C_{m_2T}^{s}B_TD_{T,T+1})\otimes l_n\Big].$$

Then $\Sigma_{nT} = \Delta'_{nT}\Delta_{nT}$, where

$$\Delta_{nT} = \begin{pmatrix} 0 & \Delta_{1T} \\ \Delta_{2nT} & \Delta_{3nT} \end{pmatrix}.$$

For a vector of constants $a = [a'_1, a_2, \ldots, a_{1+m_2}]'$, where a_1 is conformable with Δ_{2nT},

$$\Delta_{nT} a = \begin{pmatrix} \operatorname{vec}(\bar{C}^s_T) \\ \Delta_{2nT} a_1 + \frac{1}{2\sqrt{n}\sigma_{v0}} \mu_{3T} \operatorname{vec}_{\mathrm{D}}(D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}) \otimes l_n \end{pmatrix}$$

and $\lim_{n\to\infty} a'\Sigma_{nT} a = \operatorname{vec}'(\bar{C}^s_T)\operatorname{vec}(\bar{C}^s_T) + \lim_{n\to\infty}[\Delta_{2nT} a_1 + \frac{1}{2\sqrt{n}\sigma_{v0}} \mu_{3T} \operatorname{vec}_{\mathrm{D}}(D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}) \otimes l_n]'[\Delta_{2nT} a_1 + \frac{1}{2\sqrt{n}\sigma_{v0}} \mu_{3T} \operatorname{vec}_{\mathrm{D}}(D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}) \otimes l_n]$, where $C^s_T = \sum_{j=1}^{m_2} a_{1+j} C^s_{jT}$ and

$$\bar{C}^s_T = \frac{\sqrt{2}\sigma^2_{v0}}{2}[D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1} - \operatorname{diag}(D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1})] + c_T \cdot \operatorname{diag}(D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}).$$

We shall show that $D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}$ cannot a nonzero diagonal matrix. Suppose that

$$D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1} = A_T \tag{B.2}$$

is a nonzero diagonal matrix. Then $C^s_T = B_T D_{T,T+1} A_T D'_{T,T+1} B'_T$. Substituting this equation into (B.2) yields

$$D'_{T,T+1}(D_{T,T+1}D'_{T,T+1})^{-1} D_{T,T+1} A_T D'_{T,T+1}(D_{T,T+1}D'_{T,T+1})^{-1} D_{T,T+1} = A_T, \tag{B.3}$$

which holds if and only if columns of A_T are in the column space of $D'_{T,T+1}$. Since A_T is diagonal, we investigate whether the $(T+1)\times 1$ unit vector $a_T = [0, \ldots, 0, 1, 0, \ldots, 0]'$ is in the column space of $D'_{T,T+1}$, where the rth element of a_T is 1 with $1 < r < T+1$ and the rest of elements are zero. Since $D'_{T,T+1}[\alpha_1, \ldots, \alpha_T]' = [-\alpha_1\sqrt{\omega_0 - 1}, \alpha_1 - \alpha_2, \alpha_2 - \alpha_3, \ldots, \alpha_{T-1} - \alpha_T, \alpha_T]'$, $D'_{T,T+1}[\alpha_1, \ldots, \alpha_T]' = a_T$ requires all α_i's to be zero and also $\alpha_{r-1} - \alpha_r = 1$, which cannot hold simultaneously. In addition, there exist no α_j's such that $D'_{T,T+1}[\alpha_1, \ldots, \alpha_T]' = [1, 0, \ldots, 0]'$ or $[0, \ldots, 0, 1]'$. Thus, (B.3) does not hold and $D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}$ cannot a nonzero diagonal matrix. Hence, $\lim_{n\to\infty} a'\Sigma_{nT} a = 0$ implies that $D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1} = 0$. Pre-multiplying this equation by $B_T D_{T,T+1}$ and post-multiplying by $D'_{T,T+1} B'_T$ yield $C^s_T = 0$. Since C_{jT}'s are linearly independent, $C^s_T = 0$ implies $a_2 = \cdots = a_{1+m_2} = 0$. Then $\lim_{n\to\infty} a'\Sigma_{nT} a = \lim_{n\to\infty} a'_1 \Delta'_{2nT}\Delta_{2nT} a_1 = \lim_{n\to\infty} \frac{\sigma^2_{v0}}{n} a'_1[(K'_{1T} \otimes I_n)\mathbf{X}^*_{nT}, \ldots, (K'_{m_1 T} \otimes I_n)\mathbf{X}^*_{nT}]'(H_T \otimes I_n)[(K'_{1T} \otimes I_n)\mathbf{X}^*_{nT}, \ldots, (K'_{m_1 T} \otimes I_n)\mathbf{X}^*_{nT}] a_1 \geq \lim_{n\to\infty} \frac{c\sigma^2_{v0}}{n} a'_1[(K'_{1T} \otimes I_n)\mathbf{X}^*_{nT}, \ldots, (K'_{m_1 T} \otimes I_n)\mathbf{X}^*_{nT}]'[(K'_{1T} \otimes I_n)\mathbf{X}^*_{nT}, \ldots, (K'_{m_1 T} \otimes I_n)\mathbf{X}^*_{nT}] a_1$ for some constant c. Hence, $\lim_{n\to\infty} a'\Sigma_{nT} a = 0$ implies $a_1 = 0$ under Assumption 3.3. Therefore, $\lim_{n\to\infty}\Sigma_{nT}$ is invertible.

With the invertibility of $\lim_{n\to\infty}\Sigma_{nT}$, $\hat{\Sigma}_{nT} = \Sigma_{nT} + o_p(1)$ and $\sup_{\theta_3\in\Theta_3} |g_{nT}(\theta_3) - \mathrm{E}[g_{nT}(\theta_3)]| = o_p(1)$,

$$\sup_{\theta_3\in\Theta_3} |g'_{nT}(\theta_3)\hat{\Sigma}^{-1}_{nT} g_{nT}(\theta_3) - \mathrm{E}[g'_{nT}(\theta_3)]\Sigma^{-1}_{nT}\,\mathrm{E}[g_{nT}(\theta_3)]| = o_p(1). \tag{B.4}$$

It is straightforward to show that $\mathrm{E}[g'_{nT}(\theta_3)]\Sigma_{nT}^{-1}\mathrm{E}[g_{nT}(\theta_3)]$ is equicontinuous by the mean value theorem. As Θ_3 is compact, $\mathrm{E}[g'_{nT}(\theta_3)]\Sigma_{nT}^{-1}\mathrm{E}[g_{nT}(\theta_3)]$ is uniformly equicontinuous. With the uniform convergence in (B.4), identification uniqueness, and the uniform equicontinuity of $\mathrm{E}[g'_{nT}(\theta_3)]\Sigma_{nT}^{-1}\mathrm{E}[g_{nT}(\theta_3)]$, the consistency of $\hat{\theta}_{3,gmm}$ follows.

Applying the mean value theorem to the first order condition of $\hat{\theta}_{3,gmm}$ yields $0 = \frac{\partial g'_{nT}(\hat{\theta}_{3,gmm})}{\partial\theta_3}\hat{\Sigma}_{nT}^{-1}g_{nT}(\hat{\theta}_{3,gmm}) = \frac{\partial g'_{nT}(\hat{\theta}_{3,gmm})}{\partial\theta_3}\hat{\Sigma}_{nT}^{-1}[g_{nT}(\theta_{30}) + \frac{\partial g_{nT}(\bar{\theta}_3)}{\partial\theta'_3}(\hat{\theta}_{3,gmm} - \theta_{30})]$, where $\bar{\theta}_3$ lies between $\hat{\theta}_{3,gmm}$ and θ_{30}. Then

$$\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30}) = -\Big(\frac{\partial g'_{nT}(\hat{\theta}_{3,gmm})}{\partial\theta_3}\hat{\Sigma}_{nT}^{-1}\frac{\partial g_{nT}(\bar{\theta}_3)}{\partial\theta'_3}\Big)^{-1}\frac{\partial g'_{nT}(\hat{\theta}_{3,gmm})}{\partial\theta_3}\hat{\Sigma}_{nT}^{-1}\sqrt{n}g_{nT}(\theta_{30}).$$

We may prove by the mean value theorem that $\frac{\partial g_{nT}(\bar{\theta}_3)}{\partial\theta'_3} = \frac{\partial g_{nT}(\theta_{30})}{\partial\theta'_3} + o_p(1)$. Since terms in $\frac{\partial g_{nT}(\theta_{30})}{\partial\theta'_3}$ have the forms $\frac{1}{n}\mathbf{X}^{*\prime}_{nT}(A_T\otimes I_n)\iota_{nT}$, $\frac{1}{n}\mathbf{X}^{*\prime}_{nT}(A_T\otimes I_n)e_{nT}$ or $\frac{1}{n}e'_{nT}(A_T\otimes I_n)e_{nT}$, where A_T is a $T\times T$ matrix with bounded elements, $\frac{\partial g_{nT}(\theta_{30})}{\partial\theta'_3} - G_{nT} = o_p(1)$, where $G_{nT} = \mathrm{E}(\frac{\partial g_{nT}(\theta_{30})}{\partial\theta'_3})$. It is straightforward to see that $\lim_{n\to\infty}G_{nT}$ has full column rank under Assumptions 3.4 and 3.5. Hence,

$$\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30}) = -(G'_{nT}\Sigma_{nT}^{-1}G_{nT})^{-1}G'_{nT}\Sigma_{nT}^{-1}\sqrt{n}g_{nT}(\theta_{30}) + o_p(1).$$

Note that $\mathbf{X}^{*\prime}_{nT}(K_{jT}\otimes I_n)e_{nT} = \mathbf{X}^{*\prime}_{nT}(K_{jT}D_{T,T+1}\otimes I_n)\mathbf{V}_{nT}$ and

$$e'_{nT}(B'_TC_{jT}B_T\otimes I_n)e_{nT} = \mathbf{V}'_{nT}(D'_{T,T+1}B'_TC_{jT}B_TD_{T,T+1}\otimes I_n)\mathbf{V}_{nT},$$

where $D'_{T,T+1}B'_TC_{jT}B_TD_{T,T+1}\otimes I_n$ is bounded in both row and column sum norms, and elements of $\mathbf{V}_{nT}$ are independent with mean zero. Thus, by Theorem 1 in Kelejian and Prucha (2001), $\sqrt{n}g_{nT}(\theta_{30}) \xrightarrow{d} N(0, \lim_{n\to\infty}\Sigma_{nT})$. It follows that $\sqrt{n}(\hat{\theta}_{3,gmm} - \theta_{30}) \xrightarrow{d} N(0, \lim_{n\to\infty}(G'_{nT}\Sigma_{nT}^{-1}G_{nT})^{-1})$.

(ii) We first prove the consistency of $\hat{\theta}_{qml}$. Note that $\frac{1}{n}\ln L_w(\theta) - \frac{1}{n}\mathrm{E}[\ln L_w(\theta)] = -\frac{1}{2n\sigma_v^2}e'_{nT}(\alpha,\delta)(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) + \frac{1}{2n\sigma_v^2}\mathrm{E}[e'_{nT}(\alpha,\delta)(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta)]$, where

$$\begin{aligned}\frac{1}{n}\mathrm{E}[\ln L_w(\theta)] = &-\frac{T}{2}\ln(2\pi\sigma_v^2) - \frac{1}{2}\ln|H_T(\omega)| - \frac{1}{2n\sigma_v^2}f'_{nT}(\alpha,\delta)(H_T^{-1}(\omega)\otimes I_n)f_{nT}(\alpha,\delta)\\ &-\frac{\sigma_{v0}^2}{2\sigma_v^2}\mathrm{tr}[(I_T + (\gamma_0-\gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0-\gamma)F_T)H_T].\end{aligned}$$

Then we may show that $\sup_{\theta\in\Theta}|\frac{1}{n}\ln L_w(\theta) - \frac{1}{n}\mathrm{E}[\ln L_w(\theta)]| = o_p(1)$ and $\frac{1}{n}\mathrm{E}[\ln L_w(\theta)]$ is uniformly equicontinuous on Θ, by arguments similar to those for the GMM estimator $\hat{\theta}_{3,gmm}$. For the consistency of $\hat{\theta}_{qml}$, it remains to show the identification uniqueness. Let

$$b_{nT}(\theta_1) \equiv \min_{\sigma^2}\frac{1}{n}\mathrm{E}[\ln L_w(\theta)] = -\frac{T}{2}\ln(2\pi) - \frac{T}{2} - \frac{1}{2}\ln|H_T(\omega)| - \frac{T}{2}\ln[\bar{\sigma}_v^2(\theta_1)],$$

where

$$\bar{\sigma}_v^2(\theta_1) = \frac{1}{nT}f'_{nT}(\alpha,\delta)(H_T^{-1}(\omega)\otimes I_n)f_{nT}(\alpha,\delta) + \frac{\sigma_{v0}^2}{T}\mathrm{tr}[(I_T + (\gamma_0-\gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0-\gamma)F_T)H_T],$$

with $\theta_1 = [\alpha', \gamma, \beta', \omega]'$. Let the eigenvalues of $A_T = H_T^{1/2}(I_T + (\gamma_0 - \gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0 - \gamma)F_T)H_T^{1/2}$ be $\lambda_1, \ldots, \lambda_T$, which are all positive numbers since A_T is positive definite. By the inequality of arithmetic and geometric means,

$$
\begin{aligned}
&\frac{1}{T}\operatorname{tr}[(I_T + (\gamma_0 - \gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0 - \gamma)F_T)H_T] = \frac{1}{T}\sum_{t=1}^{T}\lambda_t \\
&\geq \Big(\prod_{t=1}^{T}\lambda_t\Big)^{1/T} = |H_T^{1/2}(I_T + (\gamma_0 - \gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0 - \gamma)F_T)H_T^{1/2}|^{1/T} \\
&= [|H_T|^{1/2} \cdot |I_T + (\gamma_0 - \gamma)F_T| \cdot |H_T(\omega)|^{-1} \cdot |I_T + (\gamma_0 - \gamma)F_T| \cdot |H_T|^{1/2}]^{1/T} = |H_T|^{1/T}|H_T(\omega)|^{-1/T}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
b_{nT}(\theta_1) &\leq -\frac{T}{2}\ln(2\pi) - \frac{T}{2} - \frac{1}{2}\ln|H_T(\omega)| - \frac{T}{2}\ln\Big[\frac{\sigma_{v0}^2}{T}\operatorname{tr}((I_T + (\gamma_0 - \gamma)F_T)'H_T^{-1}(\omega)(I_T + (\gamma_0 - \gamma)F_T)H_T)\Big] \\
&\leq -\frac{T}{2}\ln(2\pi) - \frac{T}{2} - \frac{1}{2}\ln|H_T(\omega)| - \frac{T}{2}\ln[\sigma_{v0}^2|H_T|^{1/T}|H_T(\omega)|^{-1/T}] \\
&= -\frac{T}{2}[\ln(2\pi\sigma_{v0}^2) + 1] - \frac{1}{2}\ln|H_T| = b_{nT}(\theta_{10}).
\end{aligned}
$$

The equalities are attained if and only if $f_{nT}'(\alpha, \delta)(H_T^{-1}(\omega) \otimes I_n)f_{nT}(\alpha, \delta) = 0$ and $\lambda_1 = \cdots = \lambda_T$. The latter implies that there is some $\lambda > 0$ such that $\lambda I_T = H_T^{1/2}[I_T + (\gamma_0 - \gamma)F_T]'H_T^{-1}(\omega)[I_T + (\gamma_0 - \gamma)F_T]H_T^{1/2}$, i.e.,

$$
\lambda H_T^{-1} = [I_T + (\gamma_0 - \gamma)F_T]'H_T^{-1}(\omega)[I_T + (\gamma_0 - \gamma)F_T]. \tag{B.5}
$$

Taking the determinants of both sides of (B.5) yields $\lambda^T|H_T^{-1}| = |H_T^{-1}(\omega)|$, i.e., $\lambda^T[1 + T(\omega - 1)] = 1 + T(\omega_0 - 1)$. As $[I_T + (\gamma_0 - \gamma)F_T]$ is lower-triangular, the (T, T)th elements of both sides of (B.5) implies that $\lambda\frac{(T-1)\omega_0-(T-2)}{1+T(\omega_0-1)} = \frac{(T-1)\omega-(T-2)}{1+T(\omega-1)}$. This equation and $\lambda^T[1 + T(\omega - 1)] = 1 + T(\omega_0 - 1)$ imply that $\frac{[(T-1)\omega_0-(T-2)]^T}{[1+T(\omega_0-1)]^{T-1}} = \frac{[(T-1)\omega-(T-2)]^T}{[1+T(\omega-1)]^{T-1}}$. The $\frac{[(T-1)\omega-(T-2)]^T}{[1+T(\omega-1)]^{T-1}}$ has a positive first order derivative with respect to ω, so it is increasing in ω. Thus, $\omega = \omega_0$ and $\lambda = 1$. Then (B.5) becomes $(\gamma_0 - \gamma)[F_T'H_T^{-1} + H_T^{-1}F_T + (\gamma_0 - \gamma)F_T'H_T^{-1}F_T] = 0$. The (T, T)th element of $F_T'H_T^{-1} + H_T^{-1}F_T + (\gamma_0 - \gamma)F_T'H_T^{-1}F_T$ is $\frac{(T-1)\omega_0-(T-2)}{1+T(\omega_0-1)}(\gamma_0 - \gamma)$, which is not zero if $\gamma \neq \gamma_0$. Hence (B.5) implies that $\gamma = \gamma_0$. Since the numerical values of the eigenvalues of $(H_T^{-1}(\omega) \times I_n)$ are equal to those of $H_T^{-1}(\omega)$, $f_{nT}'(\alpha, \delta)(H_T^{-1}(\omega) \otimes I_n)f_{nT}(\alpha, \delta) \geq c \cdot f_{nT}(\alpha, \delta)f_{nT}'(\alpha, \delta) \geq 0$ for some constant c. Under Assumption 3.6 and $\gamma = \gamma_0$, $\lim_{n\to\infty} f_{nT}(\alpha, \delta)f_{nT}'(\alpha, \delta)$ is uniquely zero at $(\alpha', \beta')' = (\alpha_0', \beta_0')'$. Thus, $\lim_{n\to\infty} b_{nT}(\theta_1)$ is uniquely maximized at $\theta_1 = \theta_{10}$. As $\bar{\sigma}_v^2(\theta_{10}) = \sigma_{v0}^2$, $\lim_{n\to\infty}\frac{1}{n}\operatorname{E}[\ln L_w(\theta)]$ is uniquely maximized at $\theta = \theta_0$. It follows that $\hat{\theta}_{qml}$ is consistent.

Applying the mean value theorem to the first order condition for $\hat{\theta}_{qml}$ yields $0 = \frac{\partial \ln L_w(\hat{\theta}_{qml})}{\partial\theta} = \frac{\partial \ln L_w(\theta_0)}{\partial\theta} + \frac{\partial^2 \ln L_w(\check{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}_{qml} - \theta_0)$, where $\check{\theta}$ lies between $\hat{\theta}_{qml}$ and θ_0. Then $\sqrt{n}(\hat{\theta}_{qml} - \theta_0) = (-\frac{1}{n}\frac{\partial^2 \ln L_w(\check{\theta})}{\partial\theta\partial\theta'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_w(\theta_0)}{\partial\theta}$. We may show that $\frac{1}{n}\frac{\partial^2 \ln L_w(\check{\theta})}{\partial\theta\partial\theta'} = \frac{1}{n}\operatorname{E}(\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}) + o_p(1)$, as for the proof of $\frac{\partial g_{nT}(\hat{\theta}_{3,gmm})}{\partial\theta_3'} = \operatorname{E}(\frac{\partial g_{nT}(\theta_{30})}{\partial\theta_3'}) +$

$o_p(1)$. The explicit expression of $\frac{1}{n}\,\mathrm{E}(-\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})$ given in the theorem can be derived by using $\Delta\mathbf{Y}_{n,T-1} = (F_T \otimes I_n)(e_{nT} + \mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0})$. For a block matrix $E = \left(\begin{smallmatrix} A_1 & A_2 \\ A_3 & A_4 \end{smallmatrix}\right)$, where A_1 and A_4 are square matrices and A_4 is invertible,

$$\begin{pmatrix} I & -A_2A_4^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}\begin{pmatrix} I & 0 \\ -A_4^{-1}A_3 & I \end{pmatrix} = \begin{pmatrix} A_1 - A_2A_4^{-1}A_3 & 0 \\ 0 & A_4 \end{pmatrix}. \tag{B.6}$$

If E is symmetric and A_4 is invertible, then E is invertible when $A_1 - A_2A_4^{-1}A_3$ is invertible. Partition $\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})$ into a 2×2 block matrix $\left(\begin{smallmatrix} A_1 & A_2 \\ A_3 & A_4 \end{smallmatrix}\right)$ such that A_4 is a scalar. Applying (B.6) to $\frac{1}{n}\,\mathrm{E}(-\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})$ then shows that $A_1 - A_2A_4^{-1}A_3 = \frac{1}{n\sigma_{v0}^2}A_5'A_5$, where

$$A_5 = \begin{pmatrix} (B_T \otimes I_n)\mathbf{X}^*_{nT} & (B_TF_T \otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} & 0 \\ 0 & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C_1^s) & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C_2^s) \end{pmatrix} \tag{B.7}$$

with $C_1 = B_TF_TB_T^{-1}$ and $C_2 = \frac{1}{2}B_TJ_TB_T' - \frac{1}{2T}\,\mathrm{tr}(B_TJ_TB_T')I_T$. It follows that $\lim_{n\to\infty}(A_1 - A_2A_4^{-1}A_3)$ is invertible under Assumption 3.6. Thus, $\lim_{n\to\infty}\frac{1}{n}\,\mathrm{E}(-\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})$ is invertible. The asymptotic distribution of $\hat{\theta}_{qml}$ then follows.

(iii) Note that $\hat{\theta}_{3,qml}$ solves the equation $g_{nT,a}(\theta_3) = 0$, where

$$g_{nT,a}(\theta_3) = \begin{pmatrix} \mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ \Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)\big\{[H_T^{-1}(\omega)J_TH_T^{-1}(\omega) - \frac{\mathrm{tr}[H_T^{-1}(\omega)J_T]}{T}H_T^{-1}(\omega)]\otimes I_n\big\}e_{nT}(\alpha,\delta) \end{pmatrix}.$$

With $\Delta\mathbf{Y}_{n,T-1} = (F_T(\gamma)\otimes I_n)[e_{nT}(\alpha,\delta) + \mathbf{X}^*_{nT}\binom{\alpha}{\beta}]$, $g_{nT,a}(\theta_3) = A(\theta_3)g_{nT,b}(\theta_3)$, where $A(\theta_3) = \left(\begin{smallmatrix} I_{k_{x^*}} & 0 & 0 & 0 \\ 0 & [\alpha',\beta'] & 1 & 0 \\ 0 & 0 & 0 & 1 \end{smallmatrix}\right)$ and

$$g_{nT,b}(\theta_3) = \begin{pmatrix} \mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ \mathbf{X}^{*\prime}_{nT}(F_T'(\gamma)H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)(F_T'(\gamma)H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)\big\{[H_T^{-1}(\omega)J_TH_T^{-1}(\omega) - \frac{\mathrm{tr}[H_T^{-1}(\omega)J_T]}{T}H_T^{-1}(\omega)]\otimes I_n\big\}e_{nT}(\alpha,\delta) \end{pmatrix} \tag{B.8}$$

$$= \begin{pmatrix} \mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ \mathbf{X}^{*\prime}_{nT}(F_T'(\gamma)H_T^{-1}(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)(B_T'(\omega)\cdot B_T'^{-1}(\omega)F_T'(\gamma)B_T'(\omega)\cdot B_T(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)\big\{B_T'(\omega)[B_T(\omega)J_TB_T'(\omega) - \frac{\mathrm{tr}[B_T(\omega)J_TB_T'(\omega)]}{T}I_n]B_T(\omega)\otimes I_n\big\}e_{nT}(\alpha,\delta) \end{pmatrix}, \tag{B.9}$$

with k_{x^*} being the number of columns in $\mathbf{X}^*_{nT}$. By a mean value theorem expansion of $A(\hat{\theta}_{3,qml})g_{nT,b}(\hat{\theta}_{3,qml}) = 0$, the asymptotic variance of $\hat{\theta}_{3,qml}$ is $\lim_{n\to\infty}[G'_{nT,b}A'(A\Sigma_{nT,b}A')^{-1}AG_{nT,b}]^{-1}$, where $A = A(\theta_{30})$ and $G_{nT,b} = \mathrm{E}(\frac{\partial g_{nT,b}(\theta_{30})}{\partial\theta_3'})$. Note that $G'_{nT,b}A'(A\Sigma_{nT,b}A')^{-1}AG_{nT,b} = G'_{nT,b}\Sigma_{nT,b}^{-1/2}\cdot\Sigma_{nT,b}^{1/2}A'(A\Sigma_{nT,b}A')^{-1}A\Sigma_{nT,b}^{1/2}\cdot\Sigma_{nT,b}^{-1/2}G_{nT,b} \le G'_{nT,b}\Sigma_{nT,b}^{-1}G_{nT,b}$, by the generalized Schwarz inequality. We can easily show that the optimal GMM estimator with the moment vector

$$g_{nT,c}(\theta_3) = \begin{pmatrix} \mathbf{X}^{*\prime}_{nT}(H_T^{-1}\otimes I_n)e_{nT}(\alpha,\delta) \\ \mathbf{X}^{*\prime}_{nT}(F_T'H_T^{-1}\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)(B_T'(\omega)\cdot B_T'^{-1}F_T'B_T'\cdot B_T(\omega)\otimes I_n)e_{nT}(\alpha,\delta) \\ e'_{nT}(\alpha,\delta)\big\{B_T'(\omega)[B_TJ_TB_T' - \frac{\mathrm{tr}[B_TJ_TB_T']}{T}I_n]B_T(\omega)\otimes I_n\big\}e_{nT}(\alpha,\delta) \end{pmatrix}$$

has the asymptotic variance $\lim_{n\to\infty}(G'_{nT,b}\Sigma_{nT,b}^{-1}G_{nT,b})^{-1}$. Furthermore, $B_T J_T B_T'$ in $g_{nT,c}(\theta_3)$ relates to $B_{\omega T}B_T^{-1}$ as follows. Using $H_T^{-1}(\omega) = B_T'(\omega)B_T(\omega)$ and $B_T(\omega) = B_T'^{-1}(\omega)H_T^{-1}(\omega) = (H_T(\omega)B_T'(\omega))^{-1}$, we have $B_{\omega T} = \frac{\partial B_T}{\partial \omega} = -(H_T B_T')^{-1}(\frac{\partial H_T}{\partial \omega}B_T' + H_T B'_{\omega T})(H_T B_T')^{-1} = -B_T(\frac{\partial H_T}{\partial \omega}B_T' + H_T B'_{\omega T})B_T$. Thus, $B_{\omega T}B_T^{-1} = -B_T(\frac{\partial H_T}{\partial \omega}B_T' + H_T B'_{\omega T}) = -B_T\frac{\partial H_T}{\partial \omega}B_T' - B_T(B_T'B_T)^{-1}B'_{\omega T} = -B_T\frac{\partial H_T}{\partial \omega}B_T' - B_T'^{-1}B'_{\omega T}$. It follows that

$$-B_T\frac{\partial H_T}{\partial \omega}B_T' = B_{\omega T}B_T^{-1} + (B_{\omega T}B_T^{-1})'. \tag{B.10}$$

Thus, $g_{nT,c}(\theta_3)$ is equal to $g^*_{nT}(\theta_3)$ in (3.4) except that the last element of $g_{nT,c}(\theta_3)$ is equal to -2 times the last element of $g^*_{nT}(\theta_3)$. Hence, $G'_{nT,b}\Sigma_{nT,b}^{-1}G_{nT,b} = G^{*\prime}_{nT}\Sigma_{nT}^{*-1}G^*_{nT}$. It follows that $(G^{*\prime}_{nT}\Sigma_{nT}^{*-1}G^*_{nT})^{-1} \leq [G'_{nT,b}A'(A\Sigma_{nT,b}A')^{-1}AG_{nT,b}]^{-1} = \Gamma_{nT,\theta_3}$, where Γ_{nT,θ_3} is the asymptotic variance of $\hat\theta_{3,qml}$.

(iv) Under the normality assumption of $\mathbf{V}_{nT}$, $\Sigma_{nT} = \frac{\sigma_{v0}^2}{n}\Delta'_{nT}\Delta_{nT}$, where

$$\Delta_{nT} = \begin{pmatrix} (B_T'^{-1}K'_{1T}\otimes I_n)\mathbf{X}^*_{nT} & \dots & (B_T'^{-1}K'_{m_1T}\otimes I_n)\mathbf{X}^*_{nT} & 0 & \dots & 0 \\ 0 & \dots & 0 & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^s_{1T}) & \dots & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^s_{m_2T}) \end{pmatrix}.$$

As $\frac{\partial e_{nT}(\alpha,\delta)}{\partial \theta_3'} = [-\Upsilon_{n,T+1}, -\Delta\mathbf{Z}_{nT}, 0]$, $G_{nT} = \frac{1}{n}\Delta'_{nT}\Delta^*_{nT}\theta_\Delta$, where

$$\Delta^*_{nT} = \begin{pmatrix} (B_T\otimes I_n)\mathbf{X}^*_{nT} & (B_TF_T\otimes I_n)\mathbf{X}^*_{nT} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}((C^*_{1T})^s) & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}((C^*_{2T})^s) \end{pmatrix},$$

$$\theta_\Delta = \begin{pmatrix} -I_{k_{x^*}} & 0 & 0 \\ 0 & -\binom{\alpha_0}{\beta_0} & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$C^*_{1T} = B_TF_TB_T^{-1}$, and $C^*_{2T} = B_{\omega T}B_T^{-1} - \frac{1}{T}\mathrm{tr}(B_{\omega T}B_T^{-1})I_T$ with $B_{\omega T} = \frac{\partial B_T(\omega_0)}{\partial \omega}$. Thus, by the generalized Schwarz inequality,

$$G'_{nT}\Sigma_{nT}^{-1}G_{nT} \leq \frac{1}{n\sigma_{v0}^2}\theta'_\Delta\Delta^{*\prime}_{nT}\Delta^*_{nT}\theta_\Delta, \tag{B.11}$$

and the equality holds at $\Delta_{nT} = \Delta^*_{nT}$, i.e., $m_1 = 2$, $m_2 = 2$, $K_{1T} = H_T^{-1}$, $K_{2T} = F_T'H_T^{-1}$, $C_{1T} = C^*_{1T}$ and $C_{2T} = C^*_{2T}$. Note that $B_T'C^*_{1T}B_T = H_T^{-1}F_T$, and $(B_{\omega T}B_T^{-1})^s = -B_TJ_TB_T'$ by (B.10), so the best moment vector under normality corresponds to the score vector as expected. Since $\Delta^*_{nT}\theta_\Delta = \begin{pmatrix} -(B_T\otimes I_n)\mathbf{X}^*_{nT} & -(B_TF_T\otimes I_n)\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} & 0 \\ 0 & -\sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^{*s}_{1T}) & \sqrt{\frac{n}{2}}\sigma_{v0}\,\mathrm{vec}(C^{*s}_{2T}) \end{pmatrix}$ and $G^{*\prime}_{nT}\Sigma_{nT}^{*-1}G^*_{nT} = \frac{1}{n\sigma_{v0}^2}\theta'_\Delta\Delta^{*\prime}_{nT}\Delta^*_{nT}\theta_\Delta$, $\lim_{n\to\infty}G^{*\prime}_{nT}\Sigma_{nT}^{*-1}G^*_{nT}$ is invertible under Assumption 3.6.

For $\hat\theta_{qml}$, when $\mathbf{V}_{nT}$ is normal, as $\ln L_w(\theta)$ is the true log likelihood function, $\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}) = \mathrm{E}(\frac{1}{n}\frac{\partial \ln L_w(\theta_0)}{\partial\theta}\frac{\partial \ln L_w(\theta_0)}{\partial\theta'})$. Then $\sqrt{n}(\hat\theta_{qml} - \theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}[\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})]^{-1})$. With the partition $\mathrm{E}(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}) = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$ such that A_4 is a scalar, the asymptotic variance of $\sqrt{n}(\hat\theta_{3,qml} - \theta_{30})$ is $\lim_{n\to\infty}(A_1 - A_2A_4^{-1}A_3)^{-1} = \lim_{n\to\infty}(\frac{1}{n\sigma_{v0}^2}A_5'A_5)^{-1}$, where $\frac{1}{n\sigma_{v0}^2}A_5'A_5 = G^{*\prime}_{nT}\Sigma_{nT}^{*-1}G^*_{nT}$ for A_5 in (B.7). Hence, $\hat\theta^*_{3,gmm}$ and $\hat\theta_{3,qml}$ have the same asymptotic variance when $\mathbf{V}_{nT}$ is normal. □

B.2 Proof of Theorem 4

(*i*) Denote $g_{nT,b}(\delta,\alpha_1) = [g'_{nT,1b}(\delta,\alpha_1), g'_{nT,2b}(\delta,\alpha_1)]'$. We may first show that $\hat{C}_{nT,\delta} = C_{nT,\delta} + o_p(1)$, where $C_{nT,\delta} = \mathrm{E}(\frac{\partial g_{nT,1b}(\delta_0,\alpha_{10})}{\partial \alpha_1'})[\mathrm{E}(\frac{\partial g_{nT,2b}(\delta_0,\alpha_{10})}{\partial \alpha_1'})]^{-1}$ and its explicit expression is given in the theorem. Then we can show by the mean value theorem that $\hat{R}_{nT,\delta} g_{nT}(\delta,\tilde{\alpha}_1) = g_{nT}(\delta) + o_p(1)$, where

$$g_{nT}(\delta) = \frac{1}{n}\begin{pmatrix} \Delta \mathbf{X}'_{nT}(H_T^{-1}\otimes I_n)M_{nT}e_{nT}(\delta,\alpha_{10}) \\ \mathbf{X}^{*\prime}_{nT}(F_T'H_T^{-1}\otimes I_n)M_{nT}e_{nT}(\delta,\alpha_{10}) \\ e'_{nT}(\delta,\alpha_{10})[(F_T'H_T^{-1} + \frac{\mathrm{tr}(F_T'H_T^{-1}J_T)}{\mathrm{tr}(\frac{\partial \Phi_T}{\partial \omega}H_T)}\Phi_T)\otimes I_n]e_{nT}(\delta,\alpha_{10}) \end{pmatrix}.$$

Thus, as argued in the proof of Theorem 3, $\lim_{n\to\infty}\mathrm{E}[g_{nT}(\delta)]$ is uniquely zero at $\delta=\delta_0$ under Assumption 3.8. With the identification uniqueness, the consistency of $\hat{\delta}$ can be proved by an argument similar to that for $\hat{\theta}_{4,gmm}$.

For the asymptotic distribution, by the mean value theorem,

$$\begin{aligned} 0 &= \frac{\partial g'_{nT}(\hat{\delta},\tilde{\alpha}_1)}{\partial \delta}\hat{R}'_{nT,\delta}(\hat{R}_{nT,\delta}\hat{\Sigma}_{nT,\delta}\hat{R}'_{nT,\delta})^{-1}\hat{R}_{nT,\delta}g_{nT}(\hat{\delta},\tilde{\alpha}_1) \\ &= \frac{\partial g'_{nT}(\hat{\delta},\tilde{\alpha}_1)}{\partial \delta}\hat{R}'_{nT,\delta}(\hat{R}_{nT,\delta}\hat{\Sigma}_{nT,\delta}\hat{R}'_{nT,\delta})^{-1}\big[\hat{R}_{nT,\delta}g_{nT}(\delta_0,\alpha_{10}) + \hat{R}_{nT,\delta}\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \delta'}(\hat{\delta}-\delta_0) \\ &\quad + \hat{R}_{nT,\delta}\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \alpha_1'}(\tilde{\alpha}_1-\alpha_{10})\big], \end{aligned}$$

where $\check{\delta}$ lies between δ_0 and $\hat{\delta}$, and $\check{\alpha}_1$ lies between α_{10} and $\tilde{\alpha}_1$. Thus,

$$\begin{aligned} \sqrt{n}(\hat{\delta}-\delta_0) = &-\big[\frac{\partial g'_{nT}(\hat{\delta},\tilde{\alpha}_1)}{\partial \delta}\hat{R}'_{nT,\delta}(\hat{R}_{nT,\delta}\hat{\Sigma}_{nT,\delta}\hat{R}'_{nT,\delta})^{-1}\hat{R}_{nT,\delta}\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \delta'}\big]^{-1}\frac{\partial g'_{nT}(\hat{\delta},\tilde{\alpha}_1)}{\partial \delta}\hat{R}'_{nT,\delta}(\hat{R}_{nT,\delta}\hat{\Sigma}_{nT,\delta}\hat{R}'_{nT,\delta})^{-1} \\ &\cdot\big[\hat{R}_{nT,\delta}\sqrt{n}g_{nT}(\delta_0,\alpha_{10}) + \hat{R}_{nT,\delta}\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \alpha_1'}\sqrt{n}(\tilde{\alpha}_1-\alpha_{10})\big]. \end{aligned}$$

We can show that $\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \alpha_1'} = G_{nT,\alpha_1} + o_p(1)$ and $\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \delta'} = G_{nT,\delta} + o_p(1)$, where $G_{nT,\alpha_1} = \mathrm{E}(\frac{\partial g_{nT,b}(\delta_0,\alpha_{10})}{\partial \alpha_1'})$ and $G_{nT,\delta} = \mathrm{E}(\frac{\partial g_{nT,b}(\delta_0,\alpha_{10})}{\partial \delta'})$. Since $R_{nT,\delta}G_{nT,\alpha_1} = [I, -C_{nT,\delta}]G_{nT,\alpha_1} = 0$, $\hat{R}_{nT,\delta}\frac{\partial g_{nT}(\check{\delta},\check{\alpha}_1)}{\partial \alpha_1'} = o_p(1)$. As in Theorem 3, $\lim_{n\to\infty}G_{nT,\delta}$ and $\lim_{n\to\infty}\Sigma_{nT,\delta}$ have full column rank under the maintained assumptions. Hence,

$$\begin{aligned} &\sqrt{n}(\hat{\delta}-\delta_0) \\ &= -[G'_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}G_{nT,\delta}]^{-1}G'_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}\sqrt{n}g_{nT}(\delta_0,\alpha_{10}) + o_p(1) \\ &\xrightarrow{d} N(0, \lim_{n\to\infty}[G'_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}G_{nT,\delta}]^{-1}). \end{aligned}$$

For the asymptotic variance of the GMM estimator $\hat{\delta}_{gmm}$ with the moment vector (3.4), we may reorder

(3.4) to be the new

$$g_{nT}(\theta_3)=\frac{1}{n}\begin{pmatrix}\Delta\mathbf{X}'_{nT}(H_T^{-1}\otimes I_n)e_{nT}(\alpha,\delta)\\ \mathbf{X}^{*\prime}_{nT}(F'_TH_T^{-1}\otimes I_n)e_{nT}(\alpha,\delta)\\ e'_{nT}(\alpha,\delta)(B'_T(\omega)C^*_{1T}B_T(\omega)\otimes I_n)e_{nT}(\alpha,\delta)\\ \Upsilon'_{n,T+1}(H_T^{-1}\otimes I_n)e_{nT}(\alpha,\delta)\\ e'_{nT}(\alpha,\delta)(B'_T(\omega)C^*_{2T}B_T(\omega)\otimes I_n)e_{nT}(\alpha,\delta)\end{pmatrix},$$

and denote $G_{nT}=\frac{\partial g_{nT}(\theta_{30})}{\partial\theta'_3}$ and $\Sigma_{nT,\delta}=\mathrm{Var}(\sqrt{n}g_{nT}(\theta_{30}))$ corresponding to this new $g_{nT}(\theta_3)$. Then, as $G_{nT}=[G_{nT,\delta},G_{nT,\alpha_1}]$, by Theorem 3($i$) and the block matrix inverse formula, the asymptotic variance of $\sqrt{n}(\hat{\delta}_{gmm}-\delta_0)$ is

$$\begin{aligned}&[G'_{nT,\delta}\Sigma^{-1}_{nT,\delta}G_{nT,\delta}-G'_{nT,\delta}\Sigma^{-1}_{nT,\delta}G_{nT,\alpha_1}(G'_{nT,\alpha_1}\Sigma^{-1}_{nT,\delta}G_{nT,\alpha_1})^{-1}G'_{nT,\alpha_1}\Sigma^{-1}_{nT,\delta}G_{nT,\delta}]^{-1}\\&=[G'_{nT,\delta}\Sigma^{-1/2}_{nT,\delta}(I_{k_g}-\Sigma^{-1/2}_{nT,\delta}G_{nT,\alpha_1}(G'_{nT,\alpha_1}\Sigma^{-1}_{nT,\delta}G_{nT,\alpha_1})^{-1}G'_{nT,\alpha_1}\Sigma^{-1/2}_{nT,\delta})\Sigma^{-1/2}_{nT,\delta}G_{nT,\delta}]^{-1},\end{aligned}\tag{B.12}$$

where k_g is the number of moments in $g_{nT}(\theta_3)$. Note that $(\Sigma^{1/2}_{nT,\delta}R'_{nT,\delta})'\Sigma^{-1/2}_{nT,\delta}G_{nT,\alpha_1}=R_{nT,\delta}G_{nT,\alpha_1}=0$ and

$$[\Sigma^{1/2}_{nT,\delta}R'_{nT,\delta},\Sigma^{-1/2}_{nT,\delta}G_{nT,\alpha_1}]'[\Sigma^{1/2}_{nT,\delta}R'_{nT,\delta},\Sigma^{-1/2}_{nT,\delta}G_{nT,\alpha_1}]=\begin{pmatrix}R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta} & \\ & G'_{nT,\alpha_1}\Sigma^{-1}_{nT,\delta}G_{nT,\alpha_1}\end{pmatrix}$$

has full rank for large enough n. By Exercise (3.17) on pp. 71–72 of Ruud (2000),

$$I_{k_g}=\Sigma^{-1/2}_{nT,\delta}G_{nT,\alpha_1}(G'_{nT,\alpha_1}\Sigma^{-1}_{nT,\delta}G_{nT,\alpha_1})^{-1}G'_{nT,\alpha_1}\Sigma^{-1/2}_{nT,\delta}+\Sigma^{1/2}_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}\Sigma^{1/2}_{nT,\delta}.$$

Hence, by (B.12), the asymptotic variance of $\sqrt{n}(\hat{\delta}_{gmm}-\delta_0)$ is $[G'_{nT,\delta}\Sigma^{-1/2}_{nT,\delta}\cdot\Sigma^{1/2}_{nT,\delta}R'_{nT,\delta}(R_{nT,\delta}\Sigma_{nT,\delta}R'_{nT,\delta})^{-1}R_{nT,\delta}\Sigma^{1/2}_{nT,\delta}\cdot\Sigma^{-1/2}_{nT,\delta}G_{nT,\delta}]^{-1}$, the same as that of $\sqrt{n}(\hat{\delta}-\delta_0)$.

(ii) Note that

$$\begin{aligned}\frac{\partial g_{nT,1}(\gamma,\omega)}{\partial\omega}&=-\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega)J_TH_T^{-1}(\omega)\otimes I_n)M^*_{nT}(\omega)(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1})\\&\quad+\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega)\otimes I_n)\frac{\partial M^*_{nT}(\omega)}{\partial\omega}(\Delta\mathbf{Y}_{nT}-\gamma\Delta\mathbf{Y}_{n,T-1}),\end{aligned}$$

where $\frac{\partial M^*_{nT}(\omega)}{\partial\omega}=\mathbf{X}^{*\prime}_{nT}[\mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)\mathbf{X}^*_{nT}]^{-1}\mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)J_TH_T^{-1}(\omega)\otimes I_n)M^*_{nT}(\omega)=(H_T^{1/2}(\omega)\otimes I_n)P_{nT}(\omega)(H_T^{-1/2}(\omega)J_TH_T^{-1}(\omega)\otimes I_n)M^*_{nT}(\omega)$ with

$$P_{nT}(\omega)=(H_T^{-1/2}(\omega)\otimes I_n)\mathbf{X}^{*\prime}_{nT}[\mathbf{X}^{*\prime}_{nT}(H_T^{-1}(\omega)\otimes I_n)\mathbf{X}^*_{nT}]^{-1}\mathbf{X}^{*\prime}_{nT}(H_T^{-1/2}(\omega)\otimes I_n)$$

being a projection matrix. In addition, $M^*_{nT}(\omega)=I_{nT}-(H_T^{1/2}(\omega)\otimes I_n)P_{nT}(\omega)(H_T^{-1/2}(\omega)\otimes I_n)$. By Lemma A.9 in Lee (2004), $\frac{1}{n}\operatorname{tr}(P_{nT}(\omega_0)A_{nT})=o(1)$ for an $nT\times nT$ matrix that is bounded in both row and column sum norms. Hence, using $\Delta\mathbf{Y}_{n,T-1}=(F_T\otimes I_n)(e_{nT}+\mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0})$,

$$\frac{\partial g_{nT,1}(\tilde{\gamma},\tilde{\omega})}{\partial\omega}=-\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}J_TH_T^{-1}\otimes I_n)M^*_{nT}e_{nT}+\frac{1}{n}\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\frac{\partial M^*_{nT}}{\partial\omega}e_{nT}+o_p(1)$$

$$= -\frac{1}{n}e'_{nT}(F'_T H_T^{-1} J_T H_T^{-1} \otimes I_n) M^*_{nT} e_{nT} + \frac{1}{n} e'_{nT}(F'_T H_T^{-1} \otimes I_n) \frac{\partial M^*_{nT}}{\partial \omega} e_{nT} + o_p(1)$$
$$= -\sigma^2_{v0} \operatorname{tr}(F'_T H_T^{-1} J_T) + \frac{\sigma^2_{v0}}{n} \operatorname{tr}[(H_T F'_T H_T^{-1} \otimes I_n) \frac{\partial M^*_{nT}}{\partial \omega}] + o_p(1)$$
$$= -\sigma^2_{v0} \operatorname{tr}(F'_T H_T^{-1} J_T) + o_p(1).$$

Similarly, $\frac{\partial g_{nT,2}(\tilde{\gamma},\tilde{\omega})}{\partial \omega} = \sigma^2_{v0} \operatorname{tr}(\frac{\partial \Phi_T(\omega_0)}{\partial \omega} H_T) + o_p(1) = -\sigma^2_{v0}[\operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) - \frac{1}{T} \operatorname{tr}^2(H_T^{-1} J_T)]$. Thus, $\hat{C}_{nT,\gamma c} = \operatorname{tr}(F'_T H_T^{-1} J_T)/[\operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) - \frac{1}{T} \operatorname{tr}^2(H_T^{-1} J_T)] + o_p(1)$, where $\operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) - \frac{1}{T} \operatorname{tr}^2(H_T^{-1} J_T) = \frac{T(T-1)}{[1+T(\omega_0-1)]^2} > 0$, $\operatorname{tr}(F'_T H_T^{-1} J_T) = \frac{1}{[1+T(\omega_0-1)](1-\gamma_0)}(T - \frac{1-\gamma_0^T}{1-\gamma_0}) \neq 0$ if $\gamma_0 \neq 1$, and $\operatorname{tr}(F'_T H_T^{-1} J_T) = \frac{T(T-1)}{2}$ if $\gamma_0 = 1$. We may also show that $\frac{1}{n}\Delta \mathbf{Y}'_{n,T-1} \tilde{M}^{*\prime}_{nT}(\tilde{\Phi}_T \otimes I_n) \tilde{M}^*_{nT} \Delta \mathbf{Y}_{n,T-1} = \sigma^2_{v0} \operatorname{tr}(F'_T \Phi_T F_T H_T) + \frac{1}{n}\binom{\alpha_0}{\beta_0}' \mathbf{X}^{*\prime}_{nT}(F'_T \otimes I_n) M^{*\prime}_{nT}(\Phi_T \otimes I_n) M^*_{nT}(F_T \times I_n) \mathbf{X}^*_{nT}\binom{\alpha_0}{\beta_0} + o_p(1)$. Thus, $s_{nT,1}$ is nonzero in the limit under Assumption 3.9, and (3.6) is quadratic in γ in the limit. The consistency of $\dot{\gamma}$ follows as described in the main text. The asymptotic distribution of $\dot{\gamma}$ follows by an argument similar to that for $\hat{\delta}$ in (i). The QML estimator $[\hat{\gamma}_{qml}, \hat{\omega}_{qml}]'$ solves the equation $[g_{nT,1}(\gamma,\omega), g_{nT,2}(\gamma,\omega)] = 0$, then the proof for the asymptotic equivalence of $\dot{\gamma}$ and $\hat{\gamma}_{qml}$ is similar to that for the asymptotic equivalence of $\hat{\delta}$ and $\hat{\delta}_{gmm}$ in (i). □

B.3 Proof of Theorem 5

We only show the identification condition of the GMM estimation and the derivation of best moments, since the rest of proof is similar to that of Theorem 3.

Since $e_{nT}(\gamma) = [(I_T + (\gamma_0 - \gamma)F_T) \otimes I_n] e_{nT}$, $\frac{1}{n} \operatorname{E}[e'_{nT}(\gamma)(B'_T(\gamma) A_{jT} B_T(\gamma) \otimes I_n) e_{nT}(\gamma)] = \sigma^2_{v0} \operatorname{tr}[(I_T + (\gamma_0 - \gamma)F_T)' B'_T(\gamma) A_{jT} B_T(\gamma)(I_T + (\gamma_0 - \gamma)F_T) H_T] = \sigma^2_{v0} \operatorname{tr}[B'_T(\gamma) A_{jT} B_T(\gamma) H_T] + \sigma^2_{v0}(\gamma_0 - \gamma) \operatorname{tr}[F'_T B'_T(\gamma) A^s_{jT} B_T(\gamma) H_T] + \sigma^2_{v0}(\gamma_0 - \gamma)^2 \operatorname{tr}[F'_T B'_T(\gamma) A_{jT} B_T(\gamma) F_T H_T]$, where $\operatorname{tr}[B'_T(\gamma) A_{jT} B_T(\gamma) H_T] = \operatorname{tr}[B'_T(\gamma) A_{jT} B_T(\gamma)(H_T - H_T(\gamma))]$ can be written as $\frac{2(\gamma-\gamma_0)}{(1+\gamma_0)(1+\gamma)} d_T(\gamma) A_{jT} d'_T(\gamma)$ with $d_T(\gamma)$ given in Assumption 4.3. Thus, under Assumption 4.3, $\lim_{n\to\infty} \operatorname{E}[g_{nT}(\gamma)]$ is uniquely zero at $\gamma = \gamma_0$.

To derive the best moments, denote $B_{\gamma T}(\gamma) = \frac{\partial B_T(\gamma)}{\partial \gamma}$. As $\Delta \mathbf{Y}_{n,T-1} = (F_T \otimes I_n) e_{nT}$,

$$\frac{\partial[(B_T(\gamma_0) \otimes I_n) e_{nT}(\gamma_0)]}{\partial \gamma} = (B_{\gamma T}(\gamma_0) \otimes I_n) e_{nT}(\gamma_0) - (B_T(\gamma_0) \otimes I_n) \Delta \mathbf{Y}_{n,T-1}$$
$$= ((B_{\gamma T} - B_T F_T) \otimes I_n) e_{nT}.$$

Then the jth element of $G_{nT} = \operatorname{E}(\frac{\partial g_{nT}(\gamma_0)}{\partial \gamma})$ is $\frac{1}{n} \operatorname{E}\{e'_{nT}[B'_T A^s_{jT}(B_{\gamma T} - B_T F_T) \otimes I_n] e_{nT}\} = -\sigma^2_{v0} \operatorname{tr}(A^s_{jT} K_T) = -\frac{\sigma^2_{v0}}{2} \operatorname{tr}(A^s_{jT} K^s_T) = -\frac{\sigma^2_{v0}}{2} \operatorname{tr}(A^s_{jT}(K^s_T - \frac{\operatorname{tr}(K^s_T)}{T} I_T))$, where $K_T = F^*_T - B_{\gamma T} B_T^{-1}$ with $F^*_T = B_T F_T B_T^{-1}$. When the disturbances are normal, the (j,k)th element of Σ_{nT} is $\frac{\sigma^4_{v0}}{2} \operatorname{tr}(A^s_{jT} A^s_{kT})$. Denote $\Delta_{nT} = [\operatorname{vec}(A^s_{1T}), \ldots, \operatorname{vec}(A^s_{m_1 T})]$. Then

$$G'_{nT} \Sigma^{-1}_{nT} G_{nT} = \frac{1}{2} \operatorname{vec}'\big(K^s_T - \frac{\operatorname{tr}(K^s_T)}{T} I_T\big) \Delta_{nT}(\Delta'_{nT} \Delta_{nT})^{-1} \Delta'_{nT} \operatorname{vec}\big(K^s_T - \frac{\operatorname{tr}(K^s_T)}{T} I_T\big)$$

$$\leq \frac{1}{2}\,\mathrm{vec}'\big(K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T\big)\,\mathrm{vec}\big(K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T\big) = \mathrm{tr}\big[\big(K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T\big)K_T\big],$$

by the generalized Schwarz inequality. The second equality can be attained if $\Delta_{nT} = \mathrm{vec}(K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T)$ or $\Delta_{nT} = [\mathrm{vec}((F_T^*)^s), \mathrm{vec}((B_{\gamma T}B_T^{-1})^s - \frac{\mathrm{tr}((B_{\gamma T}B_T^{-1})^s)}{T}I_n)]$, i.e., $m_1 = 1$ with $A_{1T} = K_T - \frac{\mathrm{tr}(K_T)}{T}I_n$, or $m_1 = 2$ with $A_{1T} = F_T^*$ and $A_{2T} = B_{\gamma T}B_T^{-1} - \frac{\mathrm{tr}(B_{\gamma T}B_T^{-1})}{T}I_T$.[6] We see that $F_T^* = B_T F_T B_T^{-1}$ corresponds to the linear component (4.5) in the score. As in the proof of Theorem 3, $B_{\gamma T}B_T^{-1}$ is related to the quadratic component $B_T\frac{\partial H_T}{\partial\gamma}B_T'$ in the score (4.6) via the relation

$$-B_T\frac{\partial H_T}{\partial\gamma}B_T' = B_{\gamma T}B_T^{-1} + (B_{\gamma T}B_T^{-1})'. \tag{B.13}$$

Using this relation, we can show that $\mathrm{tr}^{-1}\big[\big(K_T^s - \frac{\mathrm{tr}(K_T^s)}{T}I_T\big)K_T\big] = [2\,\mathrm{tr}(F_T'H_T^{-1}\frac{\partial H_T}{\partial\gamma}) + \mathrm{tr}(F_T'H_T^{-1}F_TH_T) + \frac{1}{2}\,\mathrm{tr}(H_T^{-1}\frac{\partial H_T}{\partial\gamma}H_T^{-1}\frac{\partial H_T}{\partial\gamma}) - \frac{1}{2T}\,\mathrm{tr}^2(H_T^{-1}\frac{\partial H_T}{\partial\gamma})]^{-1}$, which is equal to the asymptotic variance of $\sqrt{n}(\hat\gamma_{qml} - \gamma_0)$ under normal disturbances. □

C Algebra for fixed effects pure DPD

C.1 Variance of an IV estimate

In this section, we consider the variance of the IV estimator

$$\hat\theta_{1,iv} = (Q_{nT}'\Delta\mathbf{Z}_{n,T-1})^{-1}Q_{nT}'\Delta\mathbf{Y}_{nT} = \theta_{10} + (Q_{nT}'\Delta\mathbf{Z}_{n,T-1})^{-1}Q_{nT}'e_{nT}$$

in Section 2.1 for the fixed effects pure DPD model with a short past, where $Q_{nT} = [(K_T\otimes I_n)\iota_{nT}, (A_T\otimes I_n)\Delta\mathbf{Y}_{n,T-1}]$. We use the transformation $e_{nT} = (D_{T,T+1}\otimes I_n)\mathbf{V}_{nT}$ in terms of $\mathbf{V}_{nT} = (U_{n0}', V_{n1}', ..., V_{nT}')'$, where U_{n0} is independent of V_{nt}'s but has the same variance as that of V_{nt} for $t\geq 1$. Thus, we have

$$\begin{aligned}
&\mathrm{Var}(\Delta\mathbf{Y}_{n,T-1}'(A_T'\otimes I_n)e_{nT})\\
&= \mathrm{Var}[(e_{nT} + \kappa_0\iota_{nT})'(F_T'A_T'\otimes I_n)e_{nT}]\\
&= \mathrm{Var}[\mathbf{V}_{nT}'(D_{T,T+1}'F_T'A_T'D_{T,T+1}\otimes I_n)\mathbf{V}_{nT} + \kappa_0\iota_{nT}'(F_T'A_T'D_{T,T+1}\otimes I_n)\mathbf{V}_{nT}]\\
&= n\sigma_{v0}^4\,\mathrm{tr}(F_T'A_T'H_TF_T'A_T'H_T + F_T'A_T'H_TA_TF_TH_T) + \kappa_0^2\sigma_{v0}^2\iota_{nT}'(F_T'A_T'H_TA_TF_T\otimes I_n)\iota_{nT},
\end{aligned}$$

under the normality assumption of $\mathbf{V}_{nT}$. If A_T belongs to (2.12) or $A_T = H_T^{-1}$, we have $\mathrm{tr}(F_T'A_T'H_TF_T'A_T'H_T) = 0$ so that

$$\mathrm{Var}(\Delta\mathbf{Y}_{n,T-1}'(A_T'\otimes I_n)e_{nT}) = n\sigma_{v0}^4\,\mathrm{tr}(F_T'A_T'H_TA_TF_TH_T) + \kappa_0^2\sigma_{v0}^2\iota_{nT}'(F_T'A_T'H_TA_TF_T\otimes I_n)\iota_{nT}$$

[6] Here, as $B_{\gamma T}$ is highly nonlinear which might cause numerical imprecision in A_{1T} and K_T, we can use $-\frac{1}{2}B_T\frac{\partial H_T}{\partial\gamma}B_T'$ to replace $B_{\gamma T}B_T^{-1}$ in a quadratic moment due to the relation (B.13).

$$= \sigma_{v0}^2 \, \mathrm{E}[\Delta \mathbf{Y}'_{n,T-1}(A'_T H_T A_T \otimes I_n)\Delta \mathbf{Y}_{n,T-1}].$$

Also, under the normality assumption of $\mathbf{V}_{nT}$,

$$\begin{aligned}
&\mathrm{E}[\Delta \mathbf{Y}'_{n,T-1}(A'_T \otimes I_n)e_{nT}e'_{nT}(K_T \otimes I_n)\iota_{nT}] \\
&= \mathrm{E}[(e_{nT} + \kappa_0 \iota_{nT})'(F'_T A'_T \otimes I_n)e_{nT}e'_{nT}(K_T \otimes I_n)\iota_{nT}] \\
&= \mathrm{E}[\mathbf{V}'_{nT}(D'_{T,T+1}F'_T A'_T D_{T,T+1} \otimes I_n)\mathbf{V}_{nT}\mathbf{V}'_{nT}(D'_{T,T+1}K_T \otimes I_n)\iota_{nT}] + \kappa_0 \sigma_{v0}^2 \iota'_{nT}(F'_T A'_T H_T K_T \otimes I_n)\iota_{nT} \\
&= \kappa_0 \sigma_{v0}^2 \iota'_{nT}(F'_T A'_T H_T K_T \otimes I_n)\iota_{nT} \\
&= \sigma_{v0}^2 \, \mathrm{E}[\Delta \mathbf{Y}'_{n,T-1}(A'_T \otimes I_n) \cdot (H_T \otimes I_n) \cdot (K_T \otimes I_n)\iota_{nT}].
\end{aligned}$$

Hence, under normal disturbances, if $\mathrm{tr}(F'_T A'_T H_T F'_T A'_T H_T) = 0$, then $\mathrm{Var}(Q'_{nT}e_{nT}) = \sigma_{v0}^2 \, \mathrm{E}[Q'_{nT}(H_T \otimes I_n)Q_{nT}]$. It follows that the asymptotic variance of the IV estimate under normality is

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

C.2 MLE of γ vs the infeasible best IV for the case with known true $\kappa_0 = 0$ and ω a free parameter

For the fixed effects pure DPD model with a finite past, where ω is a free parameter for estimation, we can investigate the variance of the MLE of γ and compare it with that of the infeasible best IV estimate. For simplicity, we assume the beginning $\Delta Y_{n,-m+1}$ has mean zero so that $\kappa_0 = 0$.

The quasi log likelihood function is

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

where $\theta = [\gamma, \sigma_v^2, \omega]'$ and $e_{nT}(\gamma) = (\Delta Y'_{n1}, \Delta Y'_{n2} - \gamma \Delta Y'_{n1}, ..., \Delta Y'_{nT} - \gamma \Delta Y'_{n,T-1})'$. The first order derivatives are

$$\begin{aligned}
\frac{\partial \ln L_w(\theta)}{\partial \gamma} &= \frac{1}{\sigma_v^2}\Delta \mathbf{Y}'_{n,T-1}(H_T^{-1}(\omega) \otimes I_n)e_{nT}(\gamma), \\
\frac{\partial \ln L_w(\theta)}{\partial \sigma_v^2} &= -\frac{nT}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}e'_{nT}(\gamma)(H_T^{-1}(\omega) \otimes I_n)e_{nT}(\gamma), \\
\frac{\partial \ln L_w(\theta)}{\partial \omega} &= \frac{1}{2\sigma_v^2}e'_{nT}(\gamma)(H_T^{-1}(\omega)J_T H_T^{-1}(\omega) \otimes I_n)e_{nT}(\gamma) - \frac{n}{2}\mathrm{tr}\left(J_T H_T^{-1}(\omega)\right),
\end{aligned}$$

where $J_T = \begin{pmatrix} 1 & 0 & \\ & \ddots & \\ & & 0 \end{pmatrix}$ and $\Delta \mathbf{Y}_{n,T-1} = [0, \Delta Y'_{n1}, ..., \Delta Y'_{n,T-1}]'$. At the true parameter vector,

$$\frac{\partial \ln L_w(\theta_0)}{\partial \gamma} = \frac{1}{\sigma_{v0}^2}\Delta \mathbf{Y}'_{n,T-1}(H_T^{-1} \otimes I_n) \cdot e_{nT},$$

$$
\begin{aligned}
\frac{\partial \ln L_w(\theta_0)}{\partial \sigma_v^2} &= -\frac{nT}{2\sigma_{v0}^2} + \frac{1}{2\sigma_{v0}^4} e_{nT}'(H_T^{-1} \otimes I_n) e_{nT}, \\
\frac{\partial \ln L_w(\theta_0)}{\partial \omega} &= \frac{1}{2\sigma_{v0}^2} e_{nT}' \left(H_T^{-1} J_T H_T^{-1} \otimes I_n \right) e_{nT} - \frac{n}{2} \operatorname{tr}\left(J_T H_T^{-1} \right).
\end{aligned}
$$

From the scores, we see that $\mathrm{E}[\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} \otimes I_n) e_{nT}] = 0$. The second order derivatives are

$$
\begin{aligned}
\frac{\partial^2 \ln L_w(\theta)}{\partial \gamma^2} &= -\frac{1}{\sigma_v^2} \Delta \mathbf{Y}_{n,T-1}'(H_T^{-1}(\omega) \otimes I_n) \Delta \mathbf{Y}_{n,T-1}, \\
\frac{\partial^2 \ln L_w(\theta)}{\partial (\sigma_v^2)^2} &= \frac{nT}{2\sigma_v^4} - \frac{1}{\sigma_v^6} e_{nT}'(\gamma)(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\gamma), \\
\frac{\partial^2 \ln L_w(\theta)}{\partial \omega^2} &= -\frac{1}{\sigma_v^2} e_{nT}'(\gamma)(H_T^{-1}(\omega) J_T H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \otimes I_n) e_{nT}(\gamma) + \frac{n}{2} \operatorname{tr}\left(J_T H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \right), \\
\frac{\partial^2 \ln L_w(\theta)}{\partial \gamma \partial \sigma_v^2} &= -\frac{1}{\sigma_v^4} \Delta \mathbf{Y}_{n,T-1}'(H_T^{-1}(\omega) \otimes I_n) e_{nT}(\gamma), \\
\frac{\partial^2 \ln L_w(\theta)}{\partial \gamma \partial \omega} &= -\frac{1}{\sigma_v^2} \Delta \mathbf{Y}_{n,T-1}'(H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \otimes I_n) e_{nT}(\gamma), \\
\frac{\partial^2 \ln L_w(\theta)}{\partial \sigma_v^2 \partial \omega} &= -\frac{1}{2\sigma_v^4} e_{nT}'(\gamma)(H_T^{-1}(\omega) J_T H_T^{-1}(\omega) \otimes I_n) e_{nT}(\gamma).
\end{aligned}
$$

Because $\mathrm{E}(e_{nT} e_{nT}') = \sigma_{v0}^2 H_T \otimes I_n$ and $\mathrm{E}[\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} \otimes I_n) e_{nT}] = 0$, we see that at true values,

$$
\begin{aligned}
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial \gamma^2}\Big) &= -\frac{1}{\sigma_{v0}^2} \mathrm{E}[\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} \otimes I_n) \Delta \mathbf{Y}_{n,T-1}], \\
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial (\sigma_v^2)^2}\Big) &= -\frac{nT}{2\sigma_{v0}^4}, \\
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial \omega^2}\Big) &= -\frac{n}{2} \operatorname{tr}(J_T H_T^{-1} J_T H_T^{-1}), \\
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial \gamma \partial \sigma_v^2}\Big) &= 0, \\
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial \gamma \partial \omega}\Big) &= -\frac{1}{\sigma_{v0}^2} \mathrm{E}[\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} J_T H_T^{-1} \otimes I_n) e_{nT}], \\
\mathrm{E}\Big(\frac{\partial^2 \ln L_w(\theta_0)}{\partial \sigma_v^2 \partial \omega}\Big) &= -\frac{n}{2\sigma_{v0}^2} \operatorname{tr}(J_T H_T^{-1}).
\end{aligned}
$$

Using $\Delta \mathbf{Y}_{n,T-1} = (F_T \otimes I_n) e_{nT}$, we have $\mathrm{E}[\Delta \mathbf{Y}_{n,T-1}' \left(H_T^{-1} J_T H_T^{-1} \otimes I_n \right) e_{nT}] = n\sigma_{v0}^2 \operatorname{tr}(F_T' H_T^{-1} J_T)$. Therefore, the corresponding information matrix is

$$
\mathrm{E}\Big(-\frac{1}{n} \frac{\partial^2 \ln L_w(\theta_0)}{\partial \theta \partial \theta'}\Big) = \frac{1}{n} \begin{pmatrix} \frac{1}{\sigma_{v0}^2} \mathrm{E}[\Delta \mathbf{Y}_{n,T-1}'(H_T^{-1} \otimes I_n) \Delta \mathbf{Y}_{n,T-1}] & 0 & * \\ 0 & \frac{nT}{2\sigma_{v0}^4} & * \\ n \cdot \operatorname{tr}(F_T' H_T^{-1} J_T) & \frac{n}{2\sigma_{v0}^2} tr(H_T^{-1} J_T) & \frac{n}{2} \operatorname{tr}(H_T^{-1} J_T H_T^{-1} J_T) \end{pmatrix}.
$$

By using the form of F_T in (8) and the form of H_T^{-1} in (5), we have $\operatorname{tr}(F_T' H_T^{-1} J_T) = d[(T-1) + (T-2)\gamma_0 + ... + \gamma_0^{T-2}]$, which is equal to $\frac{d}{1-\gamma_0}(T - \frac{1-\gamma_0^T}{1-\gamma_0})$ when $\gamma_0 \neq 1$ and is equal to $\frac{T(T-1)d}{2}$ when $\gamma_0 = 1$, where

$d = \frac{1}{1+T(\omega_0-1)}$. Also, $\mathrm{tr}(H_T^{-1}J_TH_T^{-1}J_T) = d^2T^2$ and $\mathrm{tr}(H_T^{-1}J_T) = dT$. Thus,

$$\mathrm{E}\Big(-\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'}\Big) = \frac{1}{n}\begin{pmatrix} \frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}] & 0 & * \\ 0 & \frac{nT}{2\sigma_{v0}^4} & * \\ n\,\mathrm{tr}(F_T'H_T^{-1}J_T) & \frac{nTd}{2\sigma_{v0}^2} & \frac{n}{2}(Td)^2 \end{pmatrix}.$$

Therefore, by denoting the $(1,1)$th element of $(-\,\mathrm{E}\,\frac{1}{n}\frac{\partial^2 \ln L_w(\theta_0)}{\partial\theta\partial\theta'})^{-1}$ as $\hat{\sigma}^2_{\gamma,mle}$, we have

$$\begin{aligned}
\big(\frac{1}{n}\hat{\sigma}^2_{\gamma,mle}\big)^{-1} &= \frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}] - n^2\,\mathrm{tr}^2(F_T'H_T^{-1}J_T)\left[\begin{pmatrix} \frac{nT}{2\sigma_{v0}^4} & \frac{nTd}{2\sigma_{v0}^2} \\ \frac{nTd}{2\sigma_{v0}^2} & \frac{n}{2}(Td)^2 \end{pmatrix}^{-1}\right]_{22} \\
&= \frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}] - n^2\,\mathrm{tr}^2(F_T'H_T^{-1}J_T)\left(\frac{n}{2}(Td)^2 - \frac{\big(\frac{nTd}{2\sigma_{v0}^2}\big)^2}{\frac{nT}{2\sigma_{v0}^4}}\right)^{-1} \\
&= \frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}] - n\cdot\mathrm{tr}^2(F_T'H_T^{-1}J_T)\Big(\frac{T(T-1)d^2}{2}\Big)^{-1} \\
&< \frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}].
\end{aligned}$$

Thus, the asymptotic precision of the MLE of γ with a finite T under normality would be even smaller than $\frac{1}{\sigma_{v0}^2}\,\mathrm{E}[\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1}\otimes I_n)\Delta\mathbf{Y}_{n,T-1}]$. This implies that any consistent IV estimate cannot treat H_T as a given or known variance matrix in an IV estimation. The infeasible best IV estimation cannot be achievable by any feasible version. Otherwise, the infeasible best IV as described in the text would be asymptotically more efficient relative to the MLE of γ, which contradicts the asymptotic efficiency theory of the ML approach.

C.3 Non-existence of a best GMM for the DPD model with a short past under non-normal disturbances

In this section, we show that the GMM estimator $\hat{\theta}_{2,gmm}$ in (2.22) has a lower bound by the generalized Schwarz inequality, but this bound cannot be attained.

Let $\mu_{3T} = \mathrm{diag}(\mu_{3u}, \mu_{3v}l_T')$ and $c_T = \mathrm{diag}(\frac{1}{2}(\mu_{4u} - \sigma_{v0}^4 - \mu_{3u}^2/\sigma_{v0}^2)^{1/2}, \frac{1}{2}(\mu_{4v} - \sigma_{v0}^4 - \mu_{3v}^2/\sigma_{v0}^2)^{1/2}l_T')$. We may write G_T given in Theorem 1(i) as $G_T = \Delta_T'\theta_T$, where Δ_T is in (2.24) and

$$\theta_T = \begin{pmatrix} \mathrm{vec}(\frac{1}{2\sigma_{v0}^2}c_T^{-1}\,\mathrm{diag}(\mu_{3T}D'_{T,T+1}H_T^{-1}\iota_T)) & \theta_{T,12} & \theta_{T,13} \\ -\frac{1}{\sigma_{v0}}D'_{T,T+1}H_T^{-1}\iota_T & -\frac{1}{\sigma_{v0}}D'_{T,T+1}H_T^{-1}F_T\iota_T & 0 \end{pmatrix}$$

with

$$\begin{aligned}
\theta_{T,12} = \mathrm{vec}\big(&\frac{1}{2\sigma_{v0}^2}c_T^{-1}\,\mathrm{diag}(\mu_{3T}D'_{T,T+1}H_T^{-1}F_T\iota_T)) - \frac{\sigma_{v0}^2}{2}c_T^{-1}\,\mathrm{diag}(D'_{T,T+1}B_T'(B_TF_TB_T^{-1})^sB_TD_{T,T+1}) \\
&- \frac{\sqrt{2}}{2}[D'_{T,T+1}B_T'(B_TF_TB_T^{-1})^sB_TD_{T,T+1} - \mathrm{diag}(D'_{T,T+1}B_T'(B_TF_TB_T^{-1})^sB_TD_{T,T+1})]\big),
\end{aligned}$$

$$\theta_{T,13} = \mathrm{vec}\Big(\frac{\sigma_{v0}^2}{2} c_T^{-1} \,\mathrm{diag}(D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1})$$
$$+ \frac{\sqrt{2}}{2} [D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1} - \mathrm{diag}(D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1})]\Big),$$

and $C^*_{2T} = B_{\omega T} B_T^{-1} - \frac{tr(B_{\omega T} B_T^{-1})}{T} I_T$. Since each $D'_{T,T+1} B'_T C^s_{jT} B_T D_{T,T+1}$ has a zero trace,

$$G_T = \Delta'_T \tilde{\theta}_T,$$

where

$$\tilde{\theta}_T = \begin{pmatrix} \mathrm{vec}(\frac{1}{2\sigma_{v0}^2} c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_1}{T+1} l_{T+1} l'_{T+1}) \mu_{3T} D'_{T,T+1} H_T^{-1} \iota_T)) & \tilde{\theta}_{T,12} & \tilde{\theta}_{T,13} \\ -\frac{1}{\sigma_{v0}} D'_{T,T+1} H_T^{-1} \iota_T & -\frac{1}{\sigma_{v0}} D'_{T,T+1} H_T^{-1} F_T \iota_T & 0 \end{pmatrix}$$

with

$$\tilde{\theta}_{T,12} = \mathrm{vec}\Big(\frac{1}{2\sigma_{v0}^2} c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_2}{T+1} l_{T+1} l'_{T+1}) \mu_{3T} D'_{T,T+1} H_T^{-1} F_T \iota_T))$$
$$- \frac{\sigma_{v0}^2}{2} c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_3}{T+1} l_{T+1} l'_{T+1}) \,\mathrm{vec_D}(D'_{T,T+1} B'_T (B_T F_T B_T^{-1})^s B_T D_{T,T+1}))$$
$$- \frac{\sqrt{2}}{2} [D'_{T,T+1} B'_T (B_T F_T B_T^{-1})^s B_T D_{T,T+1} - \mathrm{diag}(D'_{T,T+1} B'_T (B_T F_T B_T^{-1})^s B_T D_{T,T+1})]\Big),$$

and

$$\tilde{\theta}_{T,13} = \mathrm{vec}\Big(\frac{\sigma_{v0}^2}{2} c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_4}{T+1} l_{T+1} l'_{T+1}) \,\mathrm{vec_D}(D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1}))$$
$$+ \frac{\sqrt{2}}{2} [D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1} - \mathrm{diag}(D'_{T,T+1} B'_T (C^*_{2T})^s B_T D_{T,T+1})]\Big)$$

for some specific constants $d_1, \ldots, d_4$. Thus,

$$G'_T \Sigma_T^{-1} G_T = \tilde{\theta}'_T \Delta_T (\Delta'_T \Delta_T)^{-1} \Delta'_T \tilde{\theta}_T \leq \tilde{\theta}'_T \tilde{\theta}_T$$

by the generalized Schwarz inequality, and the second equality is attained if there exists $A_T = \begin{pmatrix} A_{1T} & A_{2T} \\ A_{3T} & A_{4T} \end{pmatrix}$ conformable with Δ_T such that $\Delta_T A_T = \tilde{\theta}_T$, i.e.,

$$\begin{pmatrix} \Delta_{1T} A_{3T} & \Delta_{1T} A_{4T} \\ \Delta_{2T} A_{1T} + \Delta_{3T} A_{3T} & \Delta_{2T} A_{2T} + \Delta_{3T} A_{4T} \end{pmatrix}$$
$$= \begin{pmatrix} \mathrm{vec}(\frac{1}{2\sigma_{v0}^2} c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_1}{T+1} l_{T+1} l'_{T+1}) \mu_{3T} D'_{T,T+1} H_T^{-1} \iota_T)) & \tilde{\theta}_{T,12} & \tilde{\theta}_{T,13} \\ -\frac{1}{\sigma_{v0}} D'_{T,T+1} H_T^{-1} \iota_T & -\frac{1}{\sigma_{v0}} D'_{T,T+1} H_T^{-1} F_T \iota_T & 0 \end{pmatrix}. \tag{C.1}$$

From the above equation, $\mathrm{vec}(c_T^{-1} \,\mathrm{diag}((I_{T+1} - \frac{d_1}{T+1} l_{T+1} l'_{T+1}) \mu_{3T} D'_{T,T+1} H_T^{-1} \iota_T)))$ is in the column space of Δ_{1T}. Since $\bar{C}^s_{jT}$ is a sum of a matrix with a zero diagonal and a diagonal matrix, there must exist some symmetric matrix C^s_T, which is a linear combination of C^s_{jT}'s, such that $D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}$ is a diagonal matrix. However, we show in the proof of Theorem 1 that $D'_{T,T+1} B'_T C^s_T B_T D_{T,T+1}$ cannot be a diagonal matrix.

C.4 Asymptotic equivalence of three efficient GMM estimates

For the within equation of the stationary fixed effects pure DPD model, we consider the GMM estimation with the empirical moment (for simplicity, we present the case with a single moment)

$$e'_{nT}(\gamma)(B'_T(\omega)\otimes I_n)(A_T(\gamma)\otimes I_n)(B_T(\omega)\otimes I_n)e_{nT}(\gamma),$$

where $A_T(\gamma)$ has its trace being zero and ω is a function of γ. We shall show the asymptotic equivalence of three GMMEs of γ_0 below, where the first one is infeasible but provides a standard for comparison:

1) $A_T(\gamma)$ is known as if it is $A_T(\gamma_0)$. The GMM estimate $\hat{\gamma}_e$ is derived by solving the moment

$$e'_{nT}(\hat{\gamma}_e)(B'_T(\hat{\omega}_e)\otimes I_n)(A_T(\gamma_0)\otimes I_n)(B_T(\hat{\omega}_e)\otimes I_n)e_{nT}(\hat{\gamma}_e)=0,$$

where $\hat{\omega}_e$ is the function ω evaluated at $\hat{\gamma}_e$.

2) We have an initial $\sqrt{n}$ consistent estimate $\tilde{\gamma}$, and use the feasible $A_T(\tilde{\gamma})$. The feasible GMM estimate is $\hat{\gamma}_f$ derived from

$$e'_{nT}(\hat{\gamma}_f)(B'_T(\hat{\omega}_f)\otimes I_n)(A_T(\tilde{\gamma})\otimes I_n)(B_T(\hat{\omega}_f)\otimes I_n)e_{nT}(\hat{\gamma}_f)=0,$$

where $\hat{\omega}_f$ is the function ω evaluated at $\hat{\gamma}_f$.

3) Another approach is to treat γ in $A_T(\gamma)$ in addition to other terms as unknown for estimation. Such a GMM estimation is a continuous updating (CU) approach with its estimate $\hat{\gamma}_c$ derived from

$$e'_{nT}(\hat{\gamma}_c)(B'_T(\hat{\omega}_c)\otimes I_n)(A_T(\hat{\gamma}_c)\otimes I_n)(B_T(\hat{\omega}_c)\otimes I_n)e_{nT}(\hat{\gamma}_c)=0,$$

where $\hat{\omega}_c$ is the function ω evaluated at $\hat{\gamma}_c$.

Their asymptotic equivalence can be shown by a Taylor expansion for each case in the derivation of the asymptotic distribution of a GMM estimator. The key is that $\mathrm{tr}[A_T(\gamma)]=0$, so is $\mathrm{tr}[\frac{\partial A_T(\gamma)}{\partial\gamma}]=0$. For simplicity, denote $s_{nT}(\gamma)=(B_T(\omega)\otimes I_n)e_{nT}(\gamma)$.

1) For $\hat{\gamma}_e$, by a Taylor expansion,

$$0=s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0)+[\frac{\partial s'_{nT}(\bar{\gamma})}{\partial\gamma}(A^s_T(\gamma_0)\otimes I_n)s_{nT}(\bar{\gamma})](\hat{\gamma}_e-\gamma_0),$$

where $A^s_T(\gamma)=A_T(\gamma)+A'_T(\gamma)$. From this expansion, we can derive the asymptotically normal distribution of $\sqrt{n}(\hat{\gamma}_e-\gamma_0)$ via

$$\sqrt{n}(\hat{\gamma}_e-\gamma_0)=-\big[\frac{1}{n}\frac{\partial s'_{nT}(\bar{\gamma})}{\partial\gamma}(A^s_T(\gamma_0)\otimes I_n)s_{nT}(\bar{\gamma})\big]^{-1}\frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0),$$

by a uniform law of large numbers (ULLN) and a central limit theorem (CLT) for quadratic forms.

2) For $\hat{\gamma}_f$, by a similar Taylor expansion, we have

$$\sqrt{n}(\hat{\gamma}_f - \gamma_0) = -\big[\frac{1}{n}\frac{\partial s'_{nT}(\bar{\gamma})}{\partial \gamma}(A^s_T(\tilde{\gamma}) \otimes I_n)s_{nT}(\bar{\gamma})\big]^{-1}\frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\tilde{\gamma})s_{nT}(\gamma_0).$$

Here, the inverse matrix will converge to the same limit as that in 1). For the second term, by expanding the term involving $\tilde{\gamma}$ at γ_0,

$$\begin{aligned}\frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\tilde{\gamma})s_{nT}(\gamma_0) &= \frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0) + \frac{1}{n}s'_{nT}(\gamma_0)\frac{\partial A_T(\bar{\gamma})}{\partial \gamma}s_{nT}(\gamma_0)\cdot\sqrt{n}(\tilde{\gamma} - \gamma_0)\\ &= \frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0) + o_P(1),\end{aligned}$$

where the second equality follows by a ULLN for the convergence of a sample average of a quadratic form to its mean, which is zero in this case because $\frac{\partial A_T(\gamma_0)}{\partial \gamma}$ has its trace being zero. By comparing 1) with 2), we see the asymptotic equivalence of the two estimators.

3) For the third estimator $\hat{\gamma}_c$, a Taylor expansion gives

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_c - \gamma_0) &= -\big[\frac{1}{n}\frac{\partial s'_{nT}(\bar{\gamma})}{\partial \gamma}(A^s_T(\bar{\gamma}) \otimes I_n)s_{nT}(\bar{\gamma}) + \frac{1}{n}s'_{nT}(\bar{\gamma})\frac{\partial A_T(\bar{\gamma})}{\partial \gamma}s_{nT}(\bar{\gamma})\big]^{-1}\frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0)\\ &= -\big[\frac{1}{n}\frac{\partial s'_{nT}(\bar{\gamma})}{\partial \gamma}(A^s_T(\bar{\gamma}) \otimes I_n)s_{nT}(\bar{\gamma}) + o_P(1)\big]^{-1}\frac{1}{\sqrt{n}}s'_{nT}(\gamma_0)A_T(\gamma_0)s_{nT}(\gamma_0),\end{aligned}$$

because $\frac{1}{n}s'_{nT}(\bar{\gamma})\frac{\partial A_T(\bar{\gamma})}{\partial \gamma}s_{nT}(\bar{\gamma}) \xrightarrow{p} 0$ as $\frac{\partial A_T(\bar{\gamma})}{\partial \gamma}$ has its trace being zero. We see that the three estimators are asymptotically equivalent.

C.5 Scoring method for the fixed effects pure stationary DPD model

For the fixed effects pure stationary DPD model, while the SGMM does not apply due to the restriction $\omega = \frac{2}{1+\gamma}$, we may consider an estimator $\check{\gamma}$ generated by the scoring method (Rao, 1965). With the score in (4.4) and an initial consistent estimator $\tilde{\gamma}$ of γ_0,

$$\check{\gamma} = \tilde{\gamma} - \Big(\frac{\partial S_{nT}(\tilde{\gamma})}{\partial \gamma}\Big)^{-1}S_{nT}(\tilde{\gamma}),$$

where

$$\begin{aligned}\frac{\partial S_{nT}(\gamma)}{\partial \gamma} &= -\Delta\mathbf{Y}'_{n,T-1}\big(H_T^{-1}(\gamma)\otimes I_n\big)\Delta\mathbf{Y}_{n,T-1} - 2\Delta\mathbf{Y}'_{n,T-1}\big(R_T(\gamma)\otimes I_n\big)e_{nT}(\gamma)\\ &\quad - e'_{nT}(\gamma)\Big(H_T^{-1}(\gamma)\frac{\partial H_T(\gamma)}{\partial \gamma}R_T(\gamma)\otimes I_n\Big)e_{nT}(\gamma) + \frac{1}{2}e'_{nT}(\gamma)\Big(H_T^{-1}(\gamma)\frac{\partial^2 H_T(\gamma)}{\partial \gamma^2}H_T^{-1}(\gamma)\otimes I_n\Big)e_{nT}(\gamma)\\ &\quad - \frac{1}{2T}\operatorname{tr}\Big(\frac{\partial^2 H_T(\gamma)}{\partial \gamma^2}H_T^{-1}(\gamma) - \frac{\partial H_T(\gamma)}{\partial \gamma}R_T(\gamma)\Big)e'_{nT}(\gamma)\big(H_T^{-1}(\gamma)\otimes I_n\big)e_{nT}(\gamma)\\ &\quad + \frac{1}{2T}\operatorname{tr}\Big(\frac{\partial H_T(\gamma)}{\partial \gamma}H_T^{-1}(\gamma)\Big)\Big[2\Delta\mathbf{Y}'_{n,T-1}\big(H_T^{-1}(\gamma)\otimes I_n\big)e_{nT}(\gamma) + e'_{nT}(\gamma)\big(R_T(\gamma)\otimes I_n\big)e_{nT}(\gamma)\Big],\end{aligned}$$

with $R_T(\gamma) = H_T^{-1}(\gamma)\frac{\partial H_T(\gamma)}{\partial \gamma}H_T^{-1}(\gamma)$. This estimator is asymptotically as efficient as the QML estimator.

D Algebra for random effects DPD

D.1 Efficient GMM for the random effects pure DPD model with a short past

For the random effects pure DPD model, under the normality assumption of $\mathbb{V}_{nT}$, the variance of the moment vector (A.2) at the true parameters is

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

where

$$\Delta_T = \begin{pmatrix} \mathcal{B}_T'^{-1} K_{1T}' \iota_{T+1} & \dots & \mathcal{B}_T'^{-1} K_{m_1T}' \iota_{T+1} & 0 & \dots & 0 \\ 0 & \dots & 0 & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}(C_{1T}^s) & \dots & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}(C_{m_2T}^s) \end{pmatrix}.$$

Then, $G_T = \Delta_T' \Delta_T^* \theta_\Delta$, where

$$\Delta_T^* = \begin{pmatrix} \mathcal{B}_T \iota_{T+1} & \mathcal{B}_T F_{T+1} \iota_{T+1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}((C_{1T}^*)^s) & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}((C_{2T}^*)^s) & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}((C_{3T}^*)^s) & \sqrt{\tfrac{1}{2}}\sigma_{v0} \operatorname{vec}((C_{4T}^*)^s) \end{pmatrix}$$

and

$$\theta_\Delta = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -\kappa_{m0} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus, by the generalized Schwarz inequality, $G_T' \Sigma_T^{-1} G_T = \frac{1}{\sigma_{v0}^2} \theta_\Delta' \Delta_T^{*\prime} \Delta_T (\Delta_T' \Delta_T)^{-1} \Delta_T' \Delta_T^* \theta_\Delta \leq \frac{1}{\sigma_{v0}^2} \theta_\Delta' \Delta_T^{*\prime} \Delta_T^* \theta_\Delta$, and the equality holds at $\Delta_T = \Delta_T^*$, i.e., $m_1 = 2$, $m_2 = 4$, $K_{1T} = \Omega_{T+1}^{-1}$, $K_{2T} = F_{T+1}' \Omega_{T+1}^{-1}$, and $C_{jT} = C_{jT}^*$ for $j = 1, \dots, 4$. Estimating the unknown θ_{20} in C_{jT}^*'s by using an initial consistent estimator of θ_{20} will not affect the asymptotic distribution of the GMM estimator, since C_{jT}^*'s have zero traces and an orthogonality condition can hold. As in Section B.1, $-\mathcal{B}_T \frac{\partial \Omega_{T+1}}{\partial \varpi} \mathcal{B}_T' = \mathcal{B}_{\varpi T} \mathcal{B}_T^{-1} + (\mathcal{B}_{\varpi T} B_T^{-1})'$. Then $\mathbf{U}_{nT}' (\mathcal{B}_T' (C_{2T}^* + C_{2T}^{*\prime}) \mathcal{B}_T \otimes I_n) \mathbf{U}_{nT} = -\mathbf{U}_{nT}' (\Phi_{T1}(\theta_{20}) \otimes I_n) \mathbf{U}_{nT}$ and, similarly, $\mathbf{U}_{nT}' (\mathcal{B}_T' (C_{jT}^* + C_{jT}^{*\prime}) \mathcal{B}_T \otimes I_n) \mathbf{U}_{nT} = -\mathbf{U}_{nT}' (\Phi_{T,j-1}(\theta_{20}) \otimes I_n) \mathbf{U}_{nT}$ for $j = 3$ and 4, where $\Phi_{Tj}(\theta_2) = \Omega_{T+1}^{-1}(\theta_2) \frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}} \Omega_{T+1}^{-1}(\theta_2) - \frac{\operatorname{tr}\left(\Omega_{T+1}^{-1}(\theta_2) \frac{\partial \Omega_{T+1}(\theta_2)}{\partial \theta_{2j}}\right)}{T+1} \Omega_{T+1}^{-1}(\theta_2)$ for $j = 1, 2, 3$. Thus, the best moment vector under normality corresponds to the score vector.

D.2 Likelihood decomposition of the random effects pure stationary DPD model

For the random effects pure stationary model, it has the specification

$$Y_{nt} = \gamma Y_{n,t-1} + U_{nt}, \quad U_{nt} = c_n + V_{nt},$$

for all t. The observations of Y_{nt} for $t = 0, 1, \cdots, T$ are available. The density $f(Y_{n0}, Y_{n1}, \cdots, Y_{nT})$ of $(Y_{n0}, Y_{n1}, \cdots, Y_{nT})$ is constructed as if U_{nt}'s were normally distributed.

By taking time differences, the above equations imply that

$$\Delta Y_{nt} = \gamma \Delta Y_{n,t-1} + \Delta V_{nt},$$

for all t. Let $f(Y_{n0}, \Delta Y_{n1}, \cdots, \Delta Y_{nT})$ be the density of $(Y_{n0}, \Delta Y_{n1}, \cdots, \Delta Y_{nT})$, which is constructed as if Y_{n0} and ΔY_{nt} were normally distributed. Consider the transformation from $(Y_{n0}, U_{n1}, \cdots, U_{nT})$ to $(Y_{n0}, \Delta Y_{n1}, \Delta V_{n2}, \cdots, \Delta V_{nT})$. Because $\Delta Y_{n1} = (\gamma_0 - 1)Y_{n0} + U_{n1}$, we have the transformation

$$(Y_{n0}', \Delta Y_{n1}', \Delta V_{n2}', \cdots, \Delta V_{nT}')' = (L_{T+1} \otimes I_n)(Y_{n0}', U_{n1}', \cdots, U_{nT}')',$$

where

$$L_{T+1} = \begin{pmatrix} 1 & & & & \\ (\gamma_0 - 1) & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}$$

with its determinant being one. By the recursive structure of the autoregressive process, the determinant of the relevant Jacobian transformation of disturbances to dependent variables is one, and hence $f(Y_{n0}, Y_{n1}, \cdots, Y_{nT}) = f(Y_{n0}, \Delta Y_{n1}, \cdots, \Delta Y_{nT})$. The variance matrix of $(Y_{n0}', U_{n1}', \cdots, U_{nT}')'$ is $\sigma_{v0}^2 \Omega_{T+1} \otimes I_n$, where $\Omega_{T+1} = \begin{pmatrix} \frac{\phi_{c0}}{(1-\gamma_0)^2} + \frac{1}{1-\gamma_0^2} & \frac{\phi_{c0}}{1-\gamma_0} l_T' \\ \frac{\phi_{c0}}{1-\gamma_0} l_T & \phi_{c0} l_T l_T' + I_n \end{pmatrix}$, and hence the variance matrix of $(Y_{n0}', \Delta Y_{n1}', \Delta V_{n2}', \cdots, \Delta V_{nT}')'$ is $\sigma_{v0}^2 \Sigma_{T+1} \otimes I_n$, where $\Sigma_{T+1} = L_{T+1} \Omega_{T+1} L_{T+1}'$. In a block matrix form, $\Sigma_{T+1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, where $\Sigma_{11} = \frac{\phi_{c0}}{(1-\gamma_0)^2} + \frac{1}{1-\gamma_0^2}$, $\Sigma_{12} = (-\frac{1}{1+\gamma_0}, 0 \cdots 0)$ and $\Sigma_{22} = H_T(\omega_0)$ with $\omega_0 = \frac{2}{1+\gamma_0}$. By the partitioned quadratic formula

$$\begin{pmatrix} Y_{n0} \\ \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix}' \Sigma_{T+1}^{-1} \begin{pmatrix} Y_{n0} \\ \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix} = \begin{pmatrix} \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix}' \Sigma_{22}^{-1} \begin{pmatrix} \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix}$$

$$+ \left(Y_{n0} - \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix} \right)' (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \left(Y_{n0} - \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} \Delta Y_{n1} \\ \Delta V_{n2} \\ \vdots \\ \Delta V_{nT} \end{pmatrix} \right),$$

and the determinant decomposition $|\Sigma_{T+1}| = |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}| \cdot |\Sigma_{22}|$, $f(Y_{n0}, \Delta Y_{n1}, \Delta V_{n2}, \cdots, \Delta V_{nT})$ can then be decomposed into the product of $f_w(Y_{n0}|\Delta Y_{n1}, \Delta V_{n2}, \cdots, \Delta V_{nT})$ and $f_b(\Delta Y_{n1}, \Delta V_{n2}, \cdots, \Delta V_{nT})$.

D.3 Likelihood decomposition of the random effects stationary DPD model with exogenous variables

For the random effects model (A.6) with exogenous variables, by taking time differences,

$$\Delta Y_{nt} = \gamma_0 \Delta Y_{n,t-1} + \Delta X_{nt}\beta_0 + \Delta V_{nt},$$

for $t \geq 2$. With the approximation of unobserved exogenous variables in Section 5.2, because $\Delta Y_{n1} = (\gamma_0 - 1)Y_{n0} + X_{n1}\beta_0 + Z_n b_0 + U_{n1}$, we have the transformation

$$((Y_{n0}-\mu_{n0})', (\Delta Y_{n1}-(\gamma_0-1)\mu_{n0}-X_{n1}\beta_0-Z_n b_0)', \Delta V_{n2}', \cdots, \Delta V_{nT}')' = (L_{T+1}\otimes I_n)((Y_{n0}-\mu_{n0})', U_{n1}', \cdots, U_{nT}')',$$

where $\mu_{n0} = Z_n\alpha^{(1)} + \vec{\mathbf{X}}_{nT}\alpha^{(2)}$. The variance matrix of $(Y_{n0}', \Delta Y_{n1}', \Delta V_{n2}', \cdots, \Delta V_{nT}')'$ is $\sigma_{v0}^2\Sigma_{T+1}\otimes I_n$, where $\Sigma_{T+1} = L_{T+1}\Omega_{T+1}L_{T+1}'$ and $\Omega_{T+1} = \begin{pmatrix} \phi_{p0}+\frac{\phi_{c0}}{(1-\gamma_0)^2}+\frac{1}{1-\gamma_0^2} & \frac{\phi_{c0}}{1-\gamma_0}l_T' \\ \frac{\phi_{c0}}{1-\gamma_0}l_T & \phi_{c0}l_T l_T' + I_n \end{pmatrix}$. In a block matrix form, $\Sigma_{T+1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ where $\Sigma_{11} = \phi_{p0} + \frac{\phi_{c0}}{(1-\gamma_0)^2} + \frac{1}{1-\gamma_0^2}$, $\Sigma_{12} = (-(1-\gamma_0)\phi_{p0} - \frac{1}{1+\gamma_0}, 0, \ldots, 0)$ and $\Sigma_{22} = H_T(\omega_0)$ with $\omega_0 = (1-\gamma_0)^2\phi_{p0} + \frac{2}{1+\gamma_0}$. With the above transformation and Σ_{T+1}, the likelihood decomposition can be done similarly to that for the pure random effects DPD model.

D.4 Scoring method for the random effects pure stationary DPD model

For the random effects pure stationary DPD model, we may also consider an estimator generated by the scoring method. From (A.16), the first order derivatives of the quasi log likelihood function are:

$$\begin{aligned}
\frac{\partial \ln L_r(\theta)}{\partial \gamma} &= -\frac{n}{2}\operatorname{tr}\Big(\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big) + \frac{1}{\sigma_v^2}\mathcal{Y}_{n,T-1}'\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma) \\
&\quad + \frac{1}{2\sigma_v^2}\mathbf{U}_{nT}'(\gamma)\Big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\Big)\mathbf{U}_{nT}(\gamma), \\
\frac{\partial \ln L_r(\theta)}{\partial \sigma_v^2} &= -\frac{n(T+1)}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\mathbf{U}_{nT}'(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma), \\
\frac{\partial \ln L_r(\theta)}{\partial \phi_c} &= -\frac{n}{2}\operatorname{tr}\Big(\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big) + \frac{1}{2\sigma_v^2}\mathbf{U}_{nT}'(\gamma)\Big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\Big)\mathbf{U}_{nT}(\gamma),
\end{aligned}$$

where $\mathcal{Y}_{n,T-1} = [0, Y_{n0}', \ldots, Y_{n,T-1}']'$. From $\frac{\partial \ln L_r(\theta)}{\partial \sigma_v^2} = 0$, for given $\theta_1 = [\gamma, \phi_c]'$, the estimate of σ_{v0}^2 is $\sigma_v^2(\theta_1) = \frac{1}{n(T+1)}\mathbf{U}_{nT}'(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$. Substituting this expression into $\frac{\partial \ln L_r(\theta)}{\partial \gamma} = 0$ and $\frac{\partial \ln L_r(\theta)}{\partial \phi_c} = 0$, we see that the QML estimate $\hat{\theta}_1$ of θ_{10} is characterized by $S_{nT}(\hat{\theta}_1) = [S_{nT,1}(\hat{\theta}_1), S_{nT,2}(\hat{\theta}_1)]' =$

0, where

$$S_{nT,1}(\theta_1) = \mathcal{Y}'_{n,T-1}\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma) + \frac{1}{2}\mathbf{U}'_{nT}(\gamma)\Big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\Big)\mathbf{U}_{nT}(\gamma)$$
$$- \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma),$$
$$S_{nT,2}(\theta_1) = \frac{1}{2}\mathbf{U}'_{nT}(\gamma)\Big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\Big)\mathbf{U}_{nT}(\gamma)$$
$$- \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma).$$

With an initial consistent estimator $\tilde\theta_1$ of θ_{10}, the estimator $\check\theta_1$ from the scoring method is $\check\theta_1 = \tilde\theta_1 - \big(\frac{\partial S_{nT}(\tilde\theta_1)}{\partial\theta_1'}\big)^{-1}S_{nT}(\tilde\theta_1)$. The explicit expressions for the components of $\frac{\partial S_{nT}(\theta_1)}{\partial\theta_1'}$ are

$$\frac{\partial S_{nT,1}(\theta_1)}{\partial\gamma} = -2\mathcal{Y}'_{n,T-1}(R_{T+1,\gamma}(\theta_1)\otimes I_n)\mathbf{U}_{nT}(\gamma) - \mathcal{Y}'_{n,T-1}(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n)\mathcal{Y}_{n,T-1}$$
$$- \mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}R_{T+1,\gamma}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2}\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma^2}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$- \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma^2}\Omega_{T+1}^{-1}(\gamma,\phi_c) - \frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}R_{T+1,\gamma}(\theta_1)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{T+1}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathcal{Y}'_{n,T-1}(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(R_{T+1,\gamma}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma),$$
$$\frac{\partial S_{nT,1}(\theta_1)}{\partial\phi_c} = -\mathcal{Y}'_{n,T-1}(R_{T+1,\phi_c}(\theta_1)\otimes I_n)\mathbf{U}_{nT}(\gamma) - \mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}R_{T+1,\gamma}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2}\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$- \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c) - \frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}R_{T+1,\phi_c}(\theta_1)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(R_{T+1,\phi_c}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma),$$
$$\frac{\partial S_{nT,2}(\theta_1)}{\partial\gamma} = -\mathcal{Y}'_{n,T-1}(R_{T+1,\phi_c}(\theta_1)\otimes I_n)\mathbf{U}_{nT}(\gamma) - \mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}R_{T+1,\gamma}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2}\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$- \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial^2\Omega_{T+1}(\gamma,\phi_c)}{\partial\gamma\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c) - \frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}R_{T+1,\gamma}(\theta_1)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{T+1}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathcal{Y}'_{n,T-1}\big(\Omega_{T+1}^{-1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma)$$
$$+ \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial\Omega_{T+1}(\gamma,\phi_c)}{\partial\phi_c}\Omega_{T+1}^{-1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(R_{T+1,\gamma}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma),$$

$$
\begin{aligned}
\frac{\partial S_{nT,2}(\theta_1)}{\partial \phi_c} &= -\mathbf{U}'_{nT}(\gamma)\Big(\Omega^{-1}_{T+1}(\gamma,\phi_c)\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}R_{T+1,\phi_c}(\theta_1)\otimes I_n\Big)\mathbf{U}_{nT}(\gamma) \\
&\quad + \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}R_{T+1,\phi_c}(\theta_1)\Big)\mathbf{U}'_{nT}(\gamma)\big(\Omega^{-1}_{T+1}(\gamma,\phi_c)\otimes I_n\big)\mathbf{U}_{nT}(\gamma) \\
&\quad + \frac{1}{2(T+1)}\operatorname{tr}\Big(\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}\Omega^{-1}_{T+1}(\gamma,\phi_c)\Big)\mathbf{U}'_{nT}(\gamma)\big(R_{T+1,\phi_c}(\theta_1)\otimes I_n\big)\mathbf{U}_{nT}(\gamma),
\end{aligned}
$$

where $R_{T+1,\gamma}(\theta_1) = \Omega^{-1}_{T+1}(\gamma,\phi_c)\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \gamma}\Omega^{-1}_{T+1}(\gamma,\phi_c)$ and $R_{T+1,\phi_c}(\theta_1) = \Omega^{-1}_{T+1}(\gamma,\phi_c)\frac{\partial \Omega_{T+1}(\gamma,\phi_c)}{\partial \phi_c}\Omega^{-1}_{T+1}(\gamma,\phi_c)$.

E Initial estimates and estimation of variance matrices of moment vectors

E.1 Fixed effects pure DPD model with a short past

Consider the pure DPD model with fixed effects:

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

and Y_{n0} is observable. The process started m periods ago. For the first-differenced equations $\Delta Y_{nt} = \gamma_0 \Delta Y_{n,t-1} + \Delta V_{nt}$, we use the moment condition $Z'_i \Delta v_i$, where the IV matrix for each individual is from

$$
Z_{i2} = \begin{pmatrix} y_{i0} & 0 & 0 & \cdots & 0 & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,T-3} & y_{i,T-2} \end{pmatrix},
$$

which uses the two most recent lags as IVs. From Blundell and Bond (1998), their GMM estimates use those IVs and also the optimal weighting matrix $A_n = (\frac{1}{n}\sum_{i=1}^n Z'_i \Phi_1 Z_i)^{-1}$ where $\Phi_1 = \sigma^2_{v0}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}$ is the $(T-1)\times(T-1)$ variance matrix of $(\Delta v_{i2}, ..., \Delta v_{iT})'$. Thus, our initial estimate could be $\tilde{\gamma} = (\Delta y'_{-1} Z_2 A_n Z'_2 \Delta y_{-1})^{-1}(\Delta y'_{-1} Z_2 A_n Z'_2 \Delta y)$ where $\Delta y_{-1} = (\Delta y'_{-1,1}, ..., \Delta y'_{-1,n})'$ with $\Delta y_{-1,i} = (\Delta y_{i1}, \cdots, \Delta y_{i,T-1})'$, $Z'_2 = (Z'_{12}, \cdots, Z'_{n2})$, and $\Delta y = (\Delta y'_1, \ldots, \Delta y'_n)'$ with $\Delta y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$. An initial consistent estimate of κ_0 is $\tilde{\kappa} = \frac{1}{n}\sum_{i=1}^n \Delta y_{i1} + \frac{1}{n}\sum_{i=1}^n \sum_{t=2}^T (1 - \frac{t-1}{T})(\Delta y_{it} - \tilde{\gamma}\Delta y_{i,t-1})$. Let $\Delta \tilde{V}_{nt} = \Delta Y_{nt} - \tilde{\gamma}\Delta Y_{n,t-1}$ for $t = 2, \ldots, T$. Then σ^2_{v0} and ω_0 can be, respectively, estimated by $\tilde{\sigma}^2_v = \frac{1}{2n(T-1)}\sum_{t=2}^T \Delta \tilde{V}'_{nt}\Delta \tilde{V}_{nt}$ and $\tilde{\omega} = \frac{1}{n\tilde{\sigma}^2_v}\sum_{i=1}^n (\Delta y_{i1} - \tilde{\kappa})^2$.

With $e_{nT}(\theta_1) = [\Delta Y'_{n1} - \kappa l'_n, \Delta Y'_{n2} - \gamma \Delta Y'_{n1}, \ldots, \Delta Y'_{nT} - \gamma \Delta Y'_{n,T-1}]'$, $e_{nT} \equiv e_{nT}(\theta_{10}) = [\Delta Y'_{n1} - \kappa_0 l'_n, \Delta V'_{n2}, \ldots, \Delta V'_{nT}]'$. For the transformation $e_{nT} = (D_{T,T+1}\otimes I_n)\mathbf{V}_{n,T+1}$ in Section 2.2, the elements of $\mathbf{V}_{n,T+1} = [U'_{n0}, V'_{n1}, \ldots, V'_{nT}]'$ are independent with the same variance σ^2_{v0}. Then $\Delta Y_{n1} - \kappa_0 l_n = V_{n1} - \sqrt{\omega_0 - 1}U_{n0}$. Let $\mu_{ju} = \mathrm{E}(u^j_{i0})$ and $\mu_{jv} = \mathrm{E}(v^j_{it})$ for $j = 3$ and 4. We note that $\mathrm{E}[(\Delta v_{it})^3] = 0$, so the third sample moments of Δv_{it} cannot be used to estimate μ_{3v}, but we can use products of Δv_{it}'s with different t and $\Delta y_{i1} - \kappa_0$ to estimate μ_{3v}. For a product of three terms with $\Delta y_{i1} - \kappa_0$ and Δv_{it}, where $t = 2, \ldots, T$, only the following cases have nonzero means:

(i) $\mathrm{E}[(\Delta y_{i1} - \kappa_0)^3] = \mu_{3v} - (\omega_0 - 1)^{3/2}\mu_{3u}$;

(ii) $\mathrm{E}[(\Delta y_{i1} - \kappa_0)\Delta v_{i2}\Delta v_{i2}] = \mu_{3v}$;

(iii) $\mathrm{E}[(\Delta y_{i1} - \kappa_0)^2\Delta v_{i2}] = -\mu_{3v}$;

(iv) $\mathrm{E}(\Delta v_{it}\Delta v_{it}\Delta v_{i,t-1}) = \mu_{3v}$ for $3 \leq t \leq T$;

(v) $\mathrm{E}(\Delta v_{it}\Delta v_{it}\Delta v_{i,t+1}) = -\mu_{3v}$ for $2 \leq t \leq T-1$.

Then we might have the estimate

$$\tilde{\mu}_{3v} = \frac{1}{2n(T-1)}\sum_{i=1}^{n}\Bigg((\Delta y_{i1} - \tilde{\kappa})(\Delta\tilde{v}_{i2})^2 - (\Delta y_{i1} - \tilde{\kappa})^2\Delta\tilde{v}_{i2} + (\Delta\tilde{v}_{iT})^2\Delta\tilde{v}_{i,T-1} - (\Delta\tilde{v}_{i2})^2\Delta\tilde{v}_{i3} + \sum_{t=3}^{T-1}(\Delta\tilde{v}_{it})^2(\Delta\tilde{v}_{i,t-1} - \Delta\tilde{v}_{i,t+1})\Bigg)$$

of μ_{3v}, and the estimate $\tilde{\mu}_{3u} = (\tilde{\omega}-1)^{-3/2}[\tilde{\mu}_{3v} - \frac{1}{n}\sum_{i=1}^{n}(\Delta y_{i1} - \tilde{\kappa})^3]$ of μ_{3u}. The fourth moments of v_{it} and u_{i0} can be, respectively, estimated by $\tilde{\mu}_{4v} = \frac{1}{2n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta\tilde{v}_{it})^4 - 3\tilde{\sigma}_v^4$, and $\tilde{\mu}_{4u} = \frac{1}{(\tilde{\omega}-1)^2}[\frac{1}{n}\sum_{i=1}^{n}(\Delta y_{i1} - \tilde{\kappa})^4 - \tilde{\mu}_{4v} - 6(\tilde{\omega}-1)\tilde{\sigma}_v^4]$.

E.2 Fixed effects DPD model with exogenous variables

For the fixed effects DPD model with exogenous variables, we note that $\Delta\mathbf{Y}_{n,T-1} = (F_T \otimes I_n)(e_{nT} + \Delta\mathbf{X}_{nT}\beta_0 + \Upsilon_{nT}\alpha_0)$, where $\Delta\mathbf{X}_{nT} = [0, \Delta X'_{n2}, \ldots, \Delta X'_{nT}]'$. Then $\Delta\mathbf{Y}'_{n,T-1}(H_T^{-1} \otimes I_n)e_{nT} = (e_{nT} + \Delta\mathbf{X}_{nT}\beta_0 + \Upsilon_{nT}\alpha_0)'(F'_T H_T^{-1} \otimes I_n)e_{nT}$ is linear-quadratic in e_{nT}, and we may use this form to compute the covariance matrix of the moment vector. The OLS estimate of α_0 is $\tilde{\alpha} = [(Z_n, \vec{\mathbf{X}}_{nT})'(Z_n, \vec{\mathbf{X}}_{nT})]^{-1}(Z_n, \vec{\mathbf{X}}_{nT})'\Delta Y_{n1}$, where $\vec{\mathbf{X}}_{nT} = [\vec{x}'_1, ..., \vec{x}'_n]'$. To estimate $\delta_0 = [\beta'_0, \gamma_0]'$, we can use the moment condition $Z'_i\Delta v_i$ where the IV matrix for each individual is from

$$Z^*_{i2} = \begin{pmatrix} \Delta x_{i2} & y_{i0} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \Delta x_{i3} & y_{i0} & y_{i1} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & y_{i,T-4} & y_{i,T-3} & y_{i,T-2} \end{pmatrix}.$$

The optimal GMM estimates for our initial estimate could be $\tilde{\delta} = (\Delta y^{*\prime}_{-1}Z^*_2 A_n Z^{*\prime}_2\Delta y^*_{-1})^{-1}(\Delta y^{*\prime}_{-1}Z^*_2 A_n Z^{*\prime}_2\Delta y)$, where $\Delta y^*_{-1} = [\Delta y^{*\prime}_{-1,1}, \ldots, \Delta y^{*\prime}_{-1,n}]'$ with $\Delta y^*_{-1,i} = [[\Delta x_{i2}, \Delta y_{i1}]', \ldots, [\Delta x_{iT}, \Delta y_{i,T-1}]']'$, and Z^*_2 consists of

Z_{i2}^* for all $i = 1, \ldots, n$. Let $\Delta\tilde{V}_{nt} = \Delta Y_{nt} - [\Delta X_{nt}, \Delta Y_{n,t-1}]\tilde{\delta}$ for $t = 2, \ldots, T$. Replacing $\Delta y_{i1} - \tilde{\kappa}$ in the last subsection by $\Delta y_{i1} - (Z_n, \vec{\mathbf{X}}_{nT})\tilde{\alpha}$ and using the new $\Delta\tilde{V}_{nt}$'s, we can similarly derive other parameter estimates as those for the fixed effects pure DPD model.

E.3 Random effects pure DPD model with a short past

For the random effects pure DPD model with a short past, note that $y_{i0} = \gamma_0^m \mathrm{E}(y_{i,-m}) + \xi_{i0}$, where $\xi_{i0} = c_{i0}\sum_{j=0}^{m-1}\gamma_0^j + \zeta_{i0}$ with $\zeta_{i0} = \gamma_0^m[y_{i,-m} - \mathrm{E}(y_{i,-m})] + \sum_{j=0}^{m-1}\gamma_0^j v_{i,-j}$. Denote $k_m = \sum_{j=0}^{m-1}\gamma_0^j$. Then $k_m = \eta_{m0}/\phi_{c0}$ with the parameterization in Section A.1. Thus, by exploring additivity of random effects, prediction errors and noises, we have $\mathbf{U}_{nT} = (\chi_T \otimes I_n)\mathbb{V}_{nT}$, where $\chi_T = \begin{pmatrix} k_m & 1 & & & \\ 1 & & 1 & & \\ \vdots & & & \ddots & \\ 1 & & & & 1 \end{pmatrix}$ is a $(T+1)\times(T+2)$ matrix, and $\mathbb{V}_{nT} = [\mathbf{c}_{n0}', \boldsymbol{\zeta}_{n0}', V_{n1}', \ldots, V_{nT}']'$ with $\boldsymbol{\zeta}_{n0} = [\zeta_{10}, \ldots, \zeta_{n0}]'$. For $n(T+1)\times n(T+1)$ nonstochastic matrices $A = [a_{ij}]$ and $B = [b_{ij}]$, $\mathbf{U}_{nT}'A\mathbf{U}_{nT} = \mathbb{V}_{nT}'(\chi_T'\otimes I_n)A(\chi_T\otimes I_n)\mathbb{V}_{nT}$ and $\mathbf{U}_{nT}'B\mathbf{U}_{nT} = \mathbb{V}_{nT}'(\chi_T'\otimes I_n)B(\chi_T\otimes I_n)\mathbb{V}_{nT}$ are quadratic in $\mathbb{V}_{nT}$ of which all its entries are independent even though they are not identically distributed. Then $\mathrm{E}[\mathbf{U}_{nT}'A\mathbf{U}_{nT}\mathbf{U}_{nT}']$ and the covariance of $\mathbf{U}_{nT}'A\mathbf{U}_{nT}$ and $\mathbf{U}_{nT}'B\mathbf{U}_{nT}$ can be computed using Lemma A.1 in Lin and Lee (2010). We note that the relation $[0, Y_{n0}', \ldots, Y_{n,T-1}']' = (F_{T+1}\otimes I_n)(\mathbf{U}_{nT} + \kappa_0[l_n', 0_{1\times nT}]')$ can be used to rewrite the moment condition $[0, Y_{n0}', \ldots, Y_{n,T-1}']'(\Omega_{T+1}^{-1}\otimes I_n)\mathbf{U}_{nT}$ into a linear-quadratic form of $\mathbf{U}_{nT}$.

An initial consistent estimate of γ_0 can be the same as that for the pure fixed effects model with a finite m in Section E.1. As $\Delta y_{i1} = (\gamma_0 - 1)y_{i0} + c_{i0} + v_{i1}$, an initial consistent estimator $\tilde{\kappa}_m$ of $\kappa_{m0} = \mathrm{E}(y_{i0})$ is $\frac{1}{\tilde{\gamma}-1}\tilde{\kappa}^*$, where $\tilde{\kappa}^*$ is the estimate of $\mathrm{E}(\Delta y_{i1})$ in Section E.1. With $\Delta\tilde{V}_{nt} = \Delta Y_{nt} - \tilde{\gamma}\Delta Y_{n,t-1}$ for $t = 2, \ldots, T$, similar to the fixed effects model, we have the following estimates of σ_{v0}^2, μ_{3v} and μ_{4v}: $\tilde{\sigma}_v^2 = \frac{1}{2n(T-1)}\sum_{t=2}^T \Delta\tilde{V}_{nt}'\Delta\tilde{V}_{nt}$,

$$\tilde{\mu}_{3v} = \frac{1}{2n(T-1)}\sum_{i=1}^n\Big([\Delta y_{i1} - (\tilde{\gamma}-1)\tilde{\kappa}_m](\Delta\tilde{v}_{i2})^2 - [\Delta y_{i1} - (\tilde{\gamma}-1)\tilde{\kappa}_m]^2\Delta\tilde{v}_{i2} + (\Delta\tilde{v}_{iT})^2\Delta\tilde{v}_{i,T-1} - (\Delta\tilde{v}_{i2})^2\Delta\tilde{v}_{i3}$$
$$+ \sum_{t=3}^{T-1}(\Delta\tilde{v}_{it})^2(\Delta\tilde{v}_{i,t-1} - \Delta\tilde{v}_{i,t+1})\Big),$$

and $\tilde{\mu}_{4v} = \frac{1}{2n(T-1)}\sum_{i=1}^n\sum_{t=2}^T(\Delta\tilde{v}_{it})^4 - 3\tilde{\sigma}_v^4$. Consistent estimates of ϖ_{m0}, η_{m0} and ϕ_{c0} can be, respectively, $\tilde{\varpi}_m = \frac{1}{n\tilde{\sigma}_v^2}\sum_{i=1}^n(y_{i0} - \tilde{\kappa}_m)^2$, $\tilde{\eta}_m = \frac{1}{nT\tilde{\sigma}_v^2}\sum_{i=1}^n\sum_{t=1}^T(y_{i0} - \tilde{\kappa}_m)(y_{it} - \tilde{\gamma}y_{i,t-1})$, and $\tilde{\phi}_c = \frac{1}{nT\tilde{\sigma}_v^2}\sum_{i=1}^n\sum_{t=1}^T(y_{it} - \tilde{\gamma}y_{i,t-1})^2 - 1$. Thus, an estimate of k_m is $\tilde{k}_m = \tilde{\eta}_m/\tilde{\phi}_c$. The consistent estimates of μ_{3c}, μ_{4c}, σ_ζ^2, $\mu_{3\zeta}$ and $\mu_{4\zeta}$ are, respectively, $\tilde{\mu}_{3c} = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T(y_{it} - \tilde{\gamma}y_{i,t-1})^3 - \mu_{3v}$, $\tilde{\mu}_{4c} = \frac{1}{n}\sum_{i=1}^n\sum_{t=1}^T(y_{it} - \tilde{\gamma}y_{i,t-1})^4 - \tilde{\mu}_{4v} - 6\tilde{\phi}_c\tilde{\sigma}_v^4$, $\tilde{\sigma}_\zeta^2 = (\tilde{\varpi}_m - \tilde{k}_m^2\tilde{\phi}_c)\tilde{\sigma}_v^2$, $\tilde{\mu}_{3\zeta} = \frac{1}{n}\sum_{i=1}^n(y_{i0} - \tilde{\kappa}_m)^3 - \tilde{k}_m^3\tilde{\mu}_{3c}$, and $\tilde{\mu}_{4\zeta} = \frac{1}{n}\sum_{i=1}^n(y_{i0} - \tilde{\kappa}_m)^4 - \tilde{k}_m^4\tilde{\mu}_{4c} - 6\tilde{k}_m^2\tilde{\phi}_c\tilde{\sigma}_v^2\tilde{\sigma}_\zeta^2$.

E.4 Random effects DPD model with a finite past and exogenous variables

For the random effects DPD model with a short past and exogenous variables, the OLS estimator $\tilde{\alpha} = [(Z_n, \vec{\mathbf{X}}_{nT})'(Z_n, \vec{\mathbf{X}}_{nT})]^{-1}(Z_n, \vec{\mathbf{X}}_{nT})'Y_{n0}$ is a consistent estimator of α_0, and we may use the generalized least squares estimator $[\tilde{b}', \tilde{\beta}', \tilde{\gamma}]'$ of $[b_0', \beta_0', \gamma_0]'$ as its initial consistent estimator. Let $\Delta\tilde{V}_{nt} = \Delta Y_{nt} - \Delta Y_{i,t-1}\tilde{\gamma} - \Delta X_{it}\tilde{\beta}$ for $t = 2, \ldots, T$. Using these new $\Delta\tilde{V}_{nt}$'s and taking into account exogenous variables by replacing $y_{i0} - \tilde{\kappa}_m$, $\Delta y_{i1} - (\tilde{\gamma} - 1)\tilde{\kappa}_m$ and $y_{it} - \tilde{\gamma}y_{i,t-1}$ in the last subsection with, respectively, $y_{i0} - [z_i, \vec{x}_i]\tilde{\alpha}$, $\Delta y_{i1} - (\tilde{\gamma} - 1)[z_i, \vec{x}_i]\tilde{\alpha} - x_{i1}\tilde{\beta} - z_i\tilde{b}$ and $y_{it} - \tilde{\gamma}y_{i,t-1} - x_{it}\tilde{\beta} - z_i\tilde{b}$, other parameters can be similarly estimated.

E.5 Stationary random effects DPD model with exogenous variables

For the computation of the variance matrix of the moment vector $[g_{nT,1}'(\delta_{10}, \delta_{20}), g_{nT,2}'(\delta_{10}, \delta_{20})]'$, where $g_{nT,1}(\delta_1, \delta_2)$ and $g_{nT,2}(\delta_1, \delta_2)$ are in (A.17) and (A.18), we note that $[0, \Delta Y_{n1}', \ldots, \Delta Y_{n,T-1}']' = (F_T \otimes I_n)(e_{nT} + \Upsilon_{n,T+1}\kappa_w + \Delta\mathbf{X}_{nT}\beta_0)$, and $[\Delta Y_{n1}', \ldots, \Delta Y_{nT}']' = (F_T^* \otimes I_n)(e_{nT} + \Upsilon_{n,T+1}\kappa_w + \Delta\mathbf{X}_{nT}\beta_0)$, where $F_T^* = \begin{pmatrix} 1 & & & \\ \gamma_0 & 1 & & \\ \vdots & \vdots & \ddots & \\ \gamma_0^{T-1} & \gamma_0^{T-2} & \cdots & 1 \end{pmatrix}$ and $\Delta\mathbf{X}_{nT} = [0, \Delta X_{n2}', \ldots, \Delta X_{nT}']'$. Then $\sum_{t=1}^T (T+1-t)\Delta Y_{nt} = ([T, T-1, \ldots, 1]F_T^* \otimes I_n)(e_{nT} + \Upsilon_{n,T+1}\kappa_w + \Delta\mathbf{X}_{nT}\beta_0)$ and $\sum_{t=2}^T (T+1-t)\Delta Y_{n,t-1} = ([0, T-1, T-2, \ldots, 1]F_T \otimes I_n)(e_{nT} + \Upsilon_{n,T+1}\kappa_w + \Delta\mathbf{X}_{nT}\beta_0)$. It follows that elements of $g_{nT,1}(\delta_{10}, \delta_{20})$ and $g_{nT,2}(\delta_{10}, \delta_{20})$ are sums of the forms $a_{1,nT}'(B_T \otimes I_n)e_{nT}$, $e_{nT}'(B_T \otimes I_n)e_{nT}$, $a_{2n}'\tilde{\xi}_n$, or $e_{nT}'(a_{3T} \otimes I_n)\tilde{\xi}_n$, where $a_{1,nT}$, B_T, a_{2n}, a_{3T} are, respectively, $nT \times 1$, $T \times T$, $n \times 1$ and $T \times 1$ exogenous matrices or vectors. Denote $q_i = p_i + \sum_{j=0}^\infty \gamma_0^j v_{i,-j}$. Then, $\xi_{i0} = q_i + \frac{1}{1-\gamma_0}c_{i0}$, $\xi_{i1} = (\gamma_0 - 1)p_i + \sum_{j=0}^\infty \gamma_0^j \Delta v_{i,1-j} = (\gamma_0 - 1)q_i + v_{i1}$, and $\tilde{\xi}_{i0} = \xi_{i0} + \phi_b[T\xi_{i1} + \sum_{t=2}^T (T+1-t)\Delta v_{it}] = \frac{1}{1+T(\omega_0-1)}q_i + \frac{1}{1-\gamma_0}c_{i0} + \phi_b \sum_{t=1}^T v_{it}$. Denote $\mathbf{q}_n = [q_1, \ldots, q_n]'$ and $\mathcal{V}_{nT} = [\mathbf{c}_{n0}', \mathbf{q}_n', V_{n1}', \ldots, V_{nT}']'$. Then,

$$\tilde{\xi}_n = \big[\frac{1}{1-\gamma_0}I_n, \frac{1}{1+T(\omega_0-1)}I_n, \phi_b(l_T' \otimes I_n)\big]\mathcal{V}_{nT},$$

and

$$e_{nT} = [0_{nT\times n}, (\gamma_0 - 1)(1, 0, \ldots, 0)' \otimes I_n, \Psi_T \otimes I_n]\mathcal{V}_{nT},$$

where $\Psi_T = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}$ is a $T \times T$ matrix. Hence, $g_{nT,1}(\delta_{10}, \delta_{20})$ and $g_{nT,2}(\delta_{10}, \delta_{20})$ are linear-quadratic forms of $\mathcal{V}_{nT}$. To compute the variance matrix of $\sqrt{n}[g_{nT,1}'(\delta_{10}, \delta_{20}), g_{nT,2}'(\delta_{10}, \delta_{20})]'$, we need estimates for the third and fourth moments of elements of $\mathcal{V}_{nT}$.

Let $\sigma_q^2 = \mathrm{E}(q_i^2)$, $\mu_{jv} = \mathrm{E}(v_{it}^j)$, $\mu_{jc} = \mathrm{E}(c_{i0}^j)$ and $\mu_{jq} = \mathrm{E}(q_i^j)$ for $j = 3$ and 4. The initial consistent estimates $\tilde{\gamma}$, $\tilde{\beta}$, $\tilde{b}$, $\tilde{\sigma}_v^2$, $\tilde{\mu}_{3v}$ and $\tilde{\mu}_{4v}$ can be the same as those in the last subsection. The equation for Δy_{i1} can be used to estimate moments of q_i. The OLS estimate of κ_w is $\tilde{\kappa}_w = [(Z_n, \vec{\mathbf{X}}_{nT})'(Z_n, \vec{\mathbf{X}}_{nT})]^{-1}(Z_n, \vec{\mathbf{X}}_{nT})\Delta Y_{n1}$. Then ω_0, σ_q^2, μ_{3q} and μ_{4q} can be, respectively, estimated by $\tilde{\omega} = \frac{1}{n\tilde{\sigma}_v^2}\sum_{i=1}^n[\Delta y_{i1} - (z_i, \vec{x}_i)\tilde{\kappa}_w]^2$, $\tilde{\sigma}_q^2 = \frac{1}{(\tilde{\gamma}-1)^2}\{\frac{1}{n}\sum_{i=1}^n[\Delta y_{i1} - (z_i, \vec{x}_i)\tilde{\kappa}_w]^2 - \tilde{\sigma}_v^2\}$, $\tilde{\mu}_{3q} = \frac{1}{(\tilde{\gamma}-1)^3}\{\frac{1}{n}\sum_{i=1}^n[\Delta y_{i1} - (z_i, \vec{x}_i)\tilde{\kappa}_w]^3 - \tilde{\mu}_{3v}\}$, and $\tilde{\mu}_{4q} = \frac{1}{(\tilde{\gamma}-1)^4}\{\frac{1}{n}\sum_{i=1}^n[\Delta y_{i1} - (z_i, \vec{x}_i)\tilde{\kappa}_w]^4 - \tilde{\mu}_{4v} - 6(\tilde{\gamma}-1)^2\tilde{\sigma}_v^2\tilde{\sigma}_q^2\}$. The equation $y_{i0} = z_i\alpha^{(1)} + \vec{x}_i\alpha^{(2)} + \xi_{i0}$ can be used to estimate moments of c_i. The OLS estimate of $\alpha = [\alpha^{(1)\prime}, \alpha^{(2)\prime}]'$ is $\tilde{\alpha} = [(Z_n, \vec{\mathbf{X}}_{nT})'(Z_n, \vec{\mathbf{X}}_{nT})]^{-1}(Z_n, \vec{\mathbf{X}}_{nT})'Y_{n0}$. Then ϕ_{c0}, μ_{3c} and μ_{4c} can be estimated by, respectively, $\tilde{\phi}_c = \frac{1}{\tilde{\sigma}_v^2}(1-\tilde{\gamma})^2\{\frac{1}{n}\sum_{i=1}^n[y_{i0} - (z_i, \vec{x}_i)\tilde{\alpha}]^2 - \tilde{\sigma}_q^2\}$, $\tilde{\mu}_{3c} = (1-\tilde{\gamma})^3\{\frac{1}{n}\sum_{i=1}^n[y_{i0} - (z_i, \vec{x}_i)\tilde{\alpha}]^3 - \tilde{\mu}_{3q}\}$, and $\tilde{\mu}_{4c} = (1-\tilde{\gamma})^4\{\frac{1}{n}\sum_{i=1}^n[y_{i0} - (z_i, \vec{x}_i)\tilde{\alpha}]^4 - \tilde{\mu}_{4q} - \frac{6}{(1-\tilde{\gamma})^2}\tilde{\phi}_c\tilde{\sigma}_v^2\tilde{\sigma}_q^2\}$.

F Additional Monte Carlo results

F.1 More Monte Carlo results with normal disturbances

In this section, we present complete Monte Carlo results with normal disturbances including those in the main text.

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 values from 0.2 to 0.9. 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. With different values of γ_0 for each n and T, finite sample properties of these estimators are summarized in Tables S.1–S.16.

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 equations together. The ML-W and ML-R are ML estimates assuming that the process has achieved stationarity, where ML-W corresponds to the fixed effects specification and ML-R corresponds to the random effects specification of the individual effects. The ML-C is the ML estimate that allows the initial period to have unrestricted mean and variance, and MD-W and MD-R are minimum distance estimates similar to ML-W and ML-R. The G-W and G-R are efficient GMM estimates under fixed and random effects model respectively with a short history, and G-WS and G-RS are corresponding efficient GMM estimates with a long history that has achieved stationarity. The SG-W1 and SG-W2 are the sequential GMM estimate for the fixed effects model under different approaches. Correspondingly, SG-R1 and SG-R2 are sequential GMM estimates under the random effects model. The Rao-W and Rao-R are the score method estimates based on Rao (1965) corresponding to fixed effects and random effects specifications.[7]

Tables S.1–S.2 are the results for the case with $\gamma_0 = 0.2$ under the pure DPD, where Table S.1 is for the DGP with $m = 20$ and Table S.2 is for the DGP with $m = 1$. For the $m = 20$ case, we see that when T is short ($T = 3$), the bias of LSDV is large. The FD-W and FD-S have some bias, while the ML-W, ML-R, ML-C, MD-W and MD-R have small biases. The efficient GMM estimate which uses quadratic moments has small bias and small median absolute deviation overall, and GS-W and GS-R that assume a stationary process have a smaller variation. The SG-W1 and SG-W2 also have small biases, but have a larger deviation than the efficient GMM estimates. The SG-R1 and SG-R2 have a similar bias but a smaller deviation than the corresponding SG-W1 and SG-W2 estimates, and they are similar to the efficient GMM estimates G-W and G-R. The score method estimates Rao-W and Rao-R also have small biases, but Rao-R has a relatively larger bias and idr. For the $m = 1$ case, the biases of ML-R, MD-R, G-R and GS-R become larger, because they require a stationary process which is clearly violated when $m = 1$. Especially, the FD-S has a larger bias when $m = 1$. 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, although the ML-W and G-W make the assumption of a steady state in constructing the likelihood function or moment conditions, their biases are still very small. The Rao-W and Rao-R have a larger bias when $m = 1$, and Rao-R additionally has a larger idr when T is short. Other estimates such as LSDV and FD-W have similar performances compared to the case of a long history ($m = 20$). To sum up, for the case when the dynamic coefficient is small, 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

[7] The explicit expressions are in Sections C.5 and D.4.

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}. We first use the random effects DGP to obtain various estimates for the within and random effects equations. For an alternative setting, 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. Tables S.3 and S.4 report the results with $\gamma_0 = 0.2$ where the DGPs are, respectively, a random effects model and a fixed effects model.

From Tables S.3 and S.4, we see that sequential GMM estimates under within equations are unbiased under both DGPs of random and fixed effects, where $\mathbf{c}_{n0}$ and $\bar{X}_{nT}$ may or may not be correlated. However, when the true DGP is a random effects model where $\mathbf{c}_{n0}$ and $\bar{X}_{nT}$ are not correlated, the random effects estimates SG-R1 and SG-R2 do have a smaller (empirical) deviation than SG-W1 and SG-W2. When the true DGP is a fixed effects model where $\mathbf{c}_{n0}$ and $\bar{X}_{nT}$ are correlated, the random effects estimate has a larger bias, especially when T is large. The SG-RS is consistent and efficient under the random effects DGP, and has a bias when the DGP is a fixed effects model. For other estimates, the bias corrected LSDV, FD-W and MD-W have similar performances under fixed and random effects DGPs, while FD-S and MD-R have poor performances when the DGP is a fixed effects model. Thus, for the case with exogenous variables, all the estimates have satisfactory performance when the DGP is a random effect model. When the DGP is a fixed effect model, the estimates based on random effects specification would have poor performance. Under the random effects DGP, compared with FD-W in Arellano and Bond (1991) and FD-S in Blundell and Bond (1998), the efficient and sequential GMM estimates have smaller deviations when T is small, but have similar bias on average. Compared with MD estimates under fixed and random effects specifications, the efficient and sequential GMM estimates have similar biases as MD estimates, but they have a smaller deviation than MD estimates (especially for MD-W). Compared with ML estimates, the efficient and sequential GMM estimates have similar biases and deviations.

Tables S.5–S.8, S.9–S.12 and S.13–S.16 are the results with $\gamma_0 = 0.5$, $\gamma_0 = 0.8$ and $\gamma_0 = 0.9$. From Table S.5 with $\gamma_0 = 0.5$, we see that the results are similar to Table S.1 as the magnitude of bias increases proportionally with γ_0 under $m = 20$. However, when $m = 1$, most estimates have large bias except for sequential estimates. When γ_0 increases to 0.8 or 0.9, the performances of various estimates will change a lot. From Table S.9 with $\gamma_0 = 0.8$ and Table S.13 with $\gamma_0 = 0.9$, we see that when $m = 20$, only ML estimates and efficient GMM estimates GS-W/GS-R have small biases, while other estimates including sequential GMM estimates have large biases. When $m = 1$, these ML and efficient GMM estimates have large bias, while those sequential GMM estimates have small biases along with ML-C. Thus, although SGMM estimates have satisfactory performance when $\gamma_0 = 0.2$, it does not when $\gamma_0 = 0.8$. When the history of DGP is short so that $m = 1$, the sequential GMM estimates perform much better as seen from Table S.2 and Table S.10. Also, for the case with exogenous variables, we see that when the dynamic coefficient is 0.5, the performance of estimates are similar to the case of $\gamma_0 = 0.2$. When γ_0 is equal to 0.8 or 0.9, these estimates are still satisfactory.

F.2 Monte Carlo results with non-normal disturbances

In this subsection, we present simulation results where 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, we see that performances of various estimates are similar.

References

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

Kelejian, H.H. and Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.

Lee, L.F., 2004. A supplement to "Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models". Department of Economics, The Ohio State University, Columbus, OH.

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

Ruud, P.A., 2000. *An Introduction to Classical Econometric Theory.* Oxford University Press.

Table S.1: Estimates under $\gamma = 0.2, m = 20$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R
n=100,T=3								
mb	-0.152	-0.015	0.009	0.006	0.006	-0.001	0.004	-0.006
mad	0.051	0.111	0.080	0.063	0.059	0.077	0.070	0.103
idr	0.223	0.417	0.310	0.266	0.256	0.308	0.275	0.379
n=100,T=10								
mb	-0.016	-0.010	0.011	0.000	-0.001	-0.004	-0.001	-0.004
mad	0.023	0.033	0.030	0.022	0.022	0.023	0.025	0.026
idr	0.088	0.121	0.115	0.091	0.091	0.101	0.094	0.094
n=300,T=3								
mb	-0.149	-0.003	0.001	0.002	-0.001	0.003	-0.005	-0.003
mad	0.033	0.057	0.046	0.042	0.041	0.046	0.044	0.053
idr	0.128	0.229	0.170	0.154	0.145	0.172	0.162	0.208
n=300,T=10								
mb	-0.013	-0.003	0.007	0.002	0.002	0.000	0.000	0.000
mad	0.014	0.019	0.017	0.015	0.014	0.014	0.016	0.015
idr	0.055	0.076	0.070	0.057	0.057	0.058	0.058	0.061

	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.011	-0.014	0.001	0.014	0.006	0.006	0.009	-0.001	0.054
mad	0.085	0.089	0.063	0.065	0.092	0.092	0.087	0.063	0.097
idr	0.354	0.387	0.270	0.262	0.396	0.396	0.341	0.268	0.446
n=100,T=10									
mb	-0.002	-0.002	-0.001	0.000	0.009	0.009	0.000	-0.001	0.000
mad	0.024	0.024	0.022	0.023	0.026	0.026	0.024	0.022	0.023
idr	0.096	0.093	0.092	0.091	0.112	0.112	0.093	0.091	0.090
n=300,T=3									
mb	0.002	-0.002	0.001	0.001	0.009	0.009	0.005	0.001	0.017
mad	0.045	0.046	0.040	0.041	0.050	0.050	0.046	0.041	0.045
idr	0.174	0.176	0.155	0.143	0.214	0.214	0.179	0.154	0.164
n=300,T=10									
mb	0.001	0.001	0.001	0.002	0.005	0.005	0.002	0.002	0.002
mad	0.015	0.014	0.014	0.015	0.016	0.016	0.014	0.015	0.014
idr	0.057	0.056	0.058	0.057	0.061	0.061	0.057	0.057	0.057

1. mb is median bias, md is median absolute deviation, and idr is 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 S.2: Estimates under $\gamma = 0.2, m = 1$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.153	-0.030	0.100	0.002	0.025	-0.007	-0.007	0.016	
mad	0.055	0.124	0.091	0.070	0.070	0.070	0.079	0.098	
idr	0.223	0.442	0.324	0.258	0.254	0.301	0.289	0.395	
n=100,T=10									
mb	-0.013	-0.014	0.032	0.004	0.007	-0.003	0.005	0.019	
mad	0.023	0.036	0.032	0.024	0.024	0.025	0.026	0.026	
idr	0.088	0.139	0.123	0.093	0.094	0.095	0.095	0.100	
n=300,T=3									
mb	-0.145	0.000	0.111	0.012	0.032	0.000	0.014	0.034	
mad	0.032	0.069	0.054	0.042	0.042	0.044	0.045	0.060	
idr	0.121	0.276	0.201	0.149	0.151	0.169	0.163	0.229	
n=300,T=10									
mb	-0.016	-0.005	0.027	0.001	0.004	-0.001	0.004	0.019	
mad	0.013	0.021	0.020	0.014	0.014	0.014	0.016	0.016	
idr	0.053	0.079	0.073	0.055	0.054	0.057	0.058	0.060	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.024	-0.025	-0.007	0.033	0.004	0.004	0.003	-0.009	0.082
mad	0.078	0.077	0.067	0.073	0.087	0.087	0.083	0.069	0.141
idr	0.354	0.390	0.262	0.277	0.444	0.444	0.390	0.260	0.746
n=100,T=10									
mb	0.001	0.000	0.003	0.009	0.014	0.014	0.002	0.004	0.008
mad	0.024	0.024	0.024	0.024	0.029	0.029	0.023	0.024	0.024
idr	0.094	0.092	0.093	0.094	0.112	0.112	0.091	0.093	0.096
n=300,T=3									
mb	-0.002	-0.008	0.008	0.034	0.000	0.000	-0.005	0.009	0.079
mad	0.047	0.046	0.041	0.042	0.051	0.051	0.045	0.042	0.075
idr	0.170	0.172	0.149	0.153	0.227	0.227	0.177	0.150	0.431
n=300,T=10									
mb	-0.001	-0.001	0.001	0.006	0.003	0.003	-0.001	0.001	0.005
mad	0.014	0.014	0.014	0.015	0.014	0.014	0.014	0.014	0.014
idr	0.055	0.055	0.055	0.055	0.057	0.057	0.054	0.055	0.055

1. mb is median bias, md is median absolute deviation, and idr is 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 S.3: Estimates under $\gamma = 0.2, \beta = 1$, RE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	-0.043	-0.002	-0.018	0.000	-0.004	-0.002	-0.009	0.000	-0.007	0.002	-0.005	0.002	-0.011	0.000
mad	0.040	0.052	0.055	0.054	0.049	0.045	0.040	0.052	0.036	0.045	0.050	0.053	0.040	0.044
idr	0.155	0.198	0.232	0.201	0.197	0.172	0.151	0.201	0.130	0.170	0.195	0.198	0.172	0.170
n=100,T=10														
mb	-0.003	0.002	-0.009	0.004	0.002	0.000	-0.002	0.002	-0.002	0.001	0.009	-0.004	-0.002	0.001
mad	0.016	0.022	0.017	0.022	0.017	0.023	0.016	0.022	0.016	0.022	0.022	0.023	0.017	0.022
idr	0.060	0.088	0.063	0.089	0.063	0.086	0.059	0.088	0.059	0.086	0.092	0.096	0.063	0.085
n=300,T=3														
mb	-0.041	-0.006	-0.004	-0.004	0.001	-0.005	-0.003	-0.004	-0.003	-0.003	-0.001	-0.003	-0.004	-0.006
mad	0.024	0.033	0.035	0.032	0.029	0.027	0.024	0.032	0.021	0.025	0.031	0.033	0.027	0.025
idr	0.095	0.118	0.130	0.119	0.110	0.106	0.094	0.118	0.079	0.100	0.122	0.120	0.103	0.101
n=300,T=10														
mb	-0.003	0.001	-0.004	0.001	-0.001	0.000	-0.001	0.001	-0.002	0.001	0.002	0.000	-0.002	0.001
mad	0.010	0.014	0.010	0.013	0.010	0.013	0.010	0.014	0.010	0.014	0.013	0.014	0.010	0.014
idr	0.036	0.050	0.040	0.050	0.040	0.048	0.036	0.050	0.035	0.049	0.045	0.051	0.035	0.048
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.012	0.000	-0.008	-0.002	0.001	0.001	0.009		-0.008	-0.003	0.001		0.005	-0.004
mad	0.041	0.053	0.038	0.045	0.043	0.051	0.045		0.036	0.045	0.043		0.036	0.045
idr	0.170	0.202	0.150	0.172	0.170	0.203	0.184		0.140	0.172	0.170		0.141	0.174
n=100,T=10														
mb	-0.001	0.001	0.001	0.000	0.005	-0.002	0.005		0.001	-0.001	0.004		0.012	-0.006
mad	0.017	0.022	0.017	0.022	0.017	0.022	0.018		0.017	0.022	0.017		0.018	0.022
idr	0.063	0.088	0.059	0.085	0.064	0.087	0.069		0.060	0.086	0.064		0.065	0.089
n=300,T=3														
mb	-0.004	-0.004	-0.005	-0.003	0.001	-0.004	0.002		-0.004	-0.004	0.001		0.001	-0.004
mad	0.025	0.032	0.021	0.025	0.027	0.033	0.028		0.022	0.025	0.027		0.022	0.026
idr	0.098	0.115	0.084	0.101	0.100	0.117	0.104		0.084	0.101	0.100		0.084	0.101
n=300,T=10														
mb	0.000	0.000	-0.001	0.001	0.001	0.000	0.001		-0.001	0.001	0.001		0.004	0.000
mad	0.010	0.014	0.010	0.014	0.010	0.014	0.010		0.010	0.014	0.010		0.010	0.014
idr	0.036	0.050	0.036	0.049	0.037	0.049	0.037		0.036	0.049	0.037		0.037	0.049

1. mb is median bias, md is median absolute deviation, and idr is 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 S.4: Estimates under $\gamma = 0.2, \beta = 1$, FE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	-0.045	-0.007	-0.027	-0.003	0.306	0.081	-0.014	-0.003	0.442	0.232	-0.002	0.000	0.170	0.357
mad	0.041	0.056	0.070	0.055	0.080	0.062	0.042	0.055	0.055	0.071	0.049	0.054	0.105	0.105
idr	0.161	0.199	0.258	0.205	0.295	0.229	0.166	0.201	0.245	0.252	0.193	0.203	0.439	0.461
n=100,T=10														
mb	-0.003	0.003	-0.015	0.006	0.069	-0.009	-0.001	0.002	0.006	0.015	0.007	-0.003	0.182	-0.010
mad	0.015	0.022	0.018	0.020	0.019	0.021	0.015	0.021	0.015	0.022	0.021	0.024	0.035	0.025
idr	0.060	0.082	0.068	0.084	0.076	0.085	0.059	0.082	0.061	0.082	0.087	0.090	0.146	0.092
n=300,T=3														
mb	-0.039	-0.003	-0.002	-0.001	0.334	0.081	-0.002	-0.001	0.444	0.229	0.001	-0.001	0.186	0.373
mad	0.023	0.028	0.040	0.029	0.055	0.033	0.024	0.028	0.049	0.051	0.028	0.029	0.072	0.082
idr	0.094	0.113	0.161	0.118	0.21	0.134	0.089	0.114	0.225	0.208	0.111	0.114	0.297	0.394
n=300,T=10														
mb	-0.002	-0.001	-0.004	0.000	0.084	-0.016	0.000	-0.002	0.008	0.012	0.004	-0.002	0.196	-0.014
mad	0.011	0.013	0.013	0.013	0.014	0.014	0.011	0.013	0.011	0.013	0.012	0.013	0.033	0.015
idr	0.037	0.048	0.046	0.049	0.053	0.050	0.037	0.048	0.037	0.048	0.045	0.049	0.123	0.057
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.020	-0.005	0.179	0.121	-0.003	-0.003	0.001		0.173	0.119	-0.003		0.391	0.316
mad	0.048	0.053	0.082	0.062	0.045	0.054	0.047		0.077	0.063	0.045		0.067	0.093
idr	0.181	0.202	0.320	0.217	0.179	0.203	0.196		0.29	0.221	0.179		0.275	0.348
n=100,T=10														
mb	-0.001	0.002	0.012	0.012	0.004	-0.001	0.007		0.011	0.012	0.004		0.137	0.114
mad	0.016	0.022	0.018	0.022	0.016	0.023	0.018		0.018	0.022	0.016		0.058	0.036
idr	0.065	0.083	0.07	0.084	0.066	0.080	0.074		0.070	0.084	0.066		0.590	0.132
n=300,T=3														
mb	-0.002	-0.001	0.207	0.127	0.002	-0.001	0.004		0.202	0.129	0.002		0.404	0.305
mad	0.025	0.028	0.051	0.035	0.026	0.028	0.025		0.049	0.035	0.026		0.055	0.082
idr	0.101	0.114	0.195	0.139	0.103	0.115	0.102		0.184	0.138	0.103		0.226	0.291
n=300,T=10														
mb	0.001	-0.002	0.012	0.009	0.002	-0.003	0.002		0.012	0.010	0.002		0.133	0.115
mad	0.011	0.014	0.012	0.014	0.011	0.014	0.010		0.012	0.014	0.011		0.041	0.027
idr	0.037	0.048	0.045	0.049	0.039	0.049	0.039		0.045	0.049	0.039		0.582	0.110

1. mb is median bias, md is median absolute deviation, and idr is 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 S.5: Estimates under $\gamma = 0.5, m = 20$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.217	-0.039	0.002	0.001	0.001	-0.057	-0.010	-0.036	
mad	0.061	0.191	0.112	0.074	0.072	0.113	0.096	0.189	
idr	0.241	0.683	0.407	0.278	0.269	0.495	0.370	0.673	
n=100,T=10									
mb	-0.031	-0.021	0.015	-0.001	-0.001	-0.316	-0.006	-0.017	
mad	0.023	0.043	0.034	0.024	0.023	0.048	0.025	0.034	
idr	0.088	0.157	0.132	0.085	0.086	0.383	0.093	0.124	
n=300,T=3									
mb	-0.212	-0.017	-0.004	0.004	0.001	-0.007	-0.005	-0.018	
mad	0.038	0.096	0.057	0.046	0.047	0.072	0.054	0.093	
idr	0.153	0.368	0.225	0.168	0.162	0.301	0.198	0.370	
n=300,T=10									
mb	-0.028	-0.007	0.008	0.001	0.000	-0.290	-0.001	-0.005	
mad	0.015	0.024	0.019	0.015	0.015	0.042	0.016	0.020	
idr	0.054	0.094	0.079	0.053	0.052	0.366	0.058	0.078	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.026	-0.010	-0.009	0.013	0.006	0.006	-0.007	-0.005	0.117
mad	0.176	0.193	0.075	0.074	0.136	0.136	0.126	0.073	0.175
idr	0.935	0.907	0.294	0.309	0.561	0.563	0.518	0.282	0.833
n=100,T=10									
mb	-0.008	-0.008	-0.003	0.000	0.015	0.015	-0.001	-0.003	0.001
mad	0.028	0.025	0.023	0.024	0.037	0.037	0.025	0.023	0.024
idr	0.138	0.103	0.085	0.088	0.206	0.206	0.105	0.084	0.092
n=300,T=3									
mb	0.000	-0.002	0.000	0.000	0.028	0.028	-0.001	0.002	0.064
mad	0.081	0.083	0.046	0.046	0.088	0.088	0.074	0.046	0.083
idr	0.395	0.469	0.168	0.168	0.358	0.358	0.283	0.173	0.406
n=300,T=10									
mb	0.000	0.000	0.000	0.000	0.011	0.011	0.001	0.000	0.001
mad	0.015	0.015	0.015	0.015	0.020	0.020	0.016	0.015	0.015
idr	0.060	0.058	0.053	0.053	0.080	0.080	0.059	0.052	0.053

1. mb is median bias, md is median absolute deviation, and idr is 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 S.6: Estimates under $\gamma = 0.5, m = 1$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.155	-0.273	0.277	0.154	0.206	-0.139	-0.034	-0.079	
mad	0.061	0.302	0.060	0.068	0.070	0.095	0.137	0.171	
idr	0.248	1.224	0.227	0.304	0.263	0.408	0.533	0.679	
n=100,T=10									
mb	-0.021	-0.034	0.154	0.044	0.057	-0.360	0.047	0.068	
mad	0.023	0.055	0.034	0.025	0.025	0.086	0.026	0.032	
idr	0.086	0.211	0.133	0.097	0.100	0.496	0.100	0.118	
n=300,T=3									
mb	-0.150	-0.093	0.304	0.153	0.211	-0.120	0.051	0.034	
mad	0.037	0.215	0.031	0.048	0.050	0.119	0.099	0.146	
idr	0.143	0.886	0.129	0.181	0.186	0.403	0.414	0.580	
n=300,T=10									
mb	-0.024	-0.009	0.162	0.044	0.058	-0.388	0.054	0.080	
mad	0.011	0.031	0.021	0.015	0.016	0.033	0.017	0.019	
idr	0.049	0.120	0.080	0.058	0.062	0.457	0.062	0.077	

	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.149	-0.142	0.100	0.162	-0.046	-0.046	-0.038	0.054	-0.466
mad	0.233	0.233	0.090	0.097	0.100	0.101	0.101	0.114	0.558
idr	0.858	1.003	0.340	0.382	0.552	0.551	0.589	0.485	2.109
n=100,T=10									
mb	-0.006	-0.004	0.040	0.075	0.011	0.011	0.001	0.042	0.145
mad	0.027	0.026	0.025	0.030	0.031	0.031	0.025	0.024	0.069
idr	0.121	0.102	0.093	0.111	0.133	0.133	0.094	0.094	0.297
n=300,T=3									
mb	-0.018	-0.016	0.124	0.174	-0.008	-0.008	-0.010	0.119	-0.342
mad	0.088	0.103	0.050	0.061	0.088	0.088	0.080	0.055	0.346
idr	0.808	0.938	0.205	0.274	0.478	0.478	0.480	0.235	1.392
n=300,T=10									
mb	-0.001	-0.002	0.041	0.075	0.004	0.004	-0.002	0.043	0.128
mad	0.012	0.013	0.015	0.018	0.015	0.015	0.012	0.015	0.035
idr	0.054	0.053	0.057	0.066	0.060	0.060	0.052	0.058	0.137

1. mb is median bias, md is median absolute deviation, and idr is 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 S.7: Estimates under $\gamma = 0.5, \beta = 1$, RE DGP

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	-0.020	-0.003	-0.037	-0.002	0.000	-0.007	-0.007	-0.001	-0.002	0.000	-0.010	0.003	-0.027	-0.004
mad	0.047	0.053	0.064	0.054	0.050	0.046	0.043	0.053	0.037	0.044	0.054	0.053	0.048	0.045
idr	0.188	0.206	0.282	0.208	0.202	0.172	0.179	0.208	0.125	0.174	0.237	0.206	0.220	0.175
n=100,T=10														
mb	-0.002	0.001	-0.011	0.005	0.003	-0.001	-0.002	0.002	-0.002	0.001	0.012	-0.005	-0.006	0.002
mad	0.015	0.022	0.015	0.021	0.016	0.022	0.015	0.022	0.014	0.021	0.025	0.023	0.015	0.022
idr	0.055	0.087	0.060	0.087	0.057	0.086	0.055	0.087	0.055	0.084	0.131	0.098	0.056	0.084
n=300,T=3														
mb	-0.017	-0.005	-0.007	-0.003	0.003	-0.007	-0.002	-0.002	-0.002	-0.003	-0.002	-0.003	-0.007	-0.007
mad	0.026	0.034	0.043	0.035	0.030	0.027	0.027	0.033	0.021	0.025	0.036	0.034	0.035	0.024
idr	0.117	0.120	0.161	0.125	0.121	0.102	0.109	0.119	0.075	0.100	0.148	0.125	0.127	0.102
n=300,T=10														
mb	-0.001	0.001	-0.005	0.001	0.000	0.000	-0.001	0.001	0.000	0.001	0.003	-0.001	-0.002	0.002
mad	0.009	0.013	0.010	0.013	0.011	0.013	0.009	0.013	0.009	0.014	0.013	0.013	0.009	0.013
idr	0.034	0.049	0.039	0.049	0.035	0.048	0.033	0.049	0.032	0.047	0.049	0.051	0.033	0.047

	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.015	-0.002	-0.003	-0.004	-0.001	0.001	0.012		-0.005	-0.005	-0.001		0.001	-0.003
mad	0.050	0.054	0.041	0.045	0.046	0.052	0.057		0.039	0.046	0.046		0.037	0.045
idr	0.199	0.206	0.153	0.176	0.192	0.207	0.224		0.148	0.175	0.192		0.138	0.172
n=100,T=10														
mb	0.000	0.001	0.001	-0.001	0.005	-0.002	0.009		0.001	-0.001	0.005		0.019	-0.005
mad	0.015	0.022	0.013	0.022	0.014	0.022	0.018		0.014	0.021	0.014		0.018	0.022
idr	0.063	0.087	0.056	0.084	0.060	0.085	0.091		0.056	0.083	0.060		0.070	0.089
n=300,T=3														
mb	-0.005	-0.004	-0.004	-0.005	0.002	-0.003	0.002		-0.005	-0.006	0.002		0.001	-0.004
mad	0.028	0.032	0.022	0.026	0.029	0.034	0.031		0.024	0.026	0.029		0.021	0.026
idr	0.111	0.119	0.079	0.100	0.117	0.120	0.117		0.082	0.100	0.117		0.079	0.101
n=300,T=10														
mb	0.000	0.000	0.001	0.001	0.002	-0.001	0.002		0.001	0.001	0.002		0.007	-0.002
mad	0.009	0.013	0.009	0.014	0.009	0.013	0.010		0.009	0.014	0.009		0.010	0.014
idr	0.033	0.049	0.032	0.048	0.036	0.049	0.036		0.032	0.047	0.036		0.037	0.048

1. mb is median bias, md is median absolute deviation, and idr is 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 S.8: Estimates under $\gamma = 0.5, \beta = 1$, FE DGP

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	-0.020	-0.006	-0.049	-0.009	0.275	0.074	-0.009	-0.004	0.314	0.185	-0.008	0.001	0.147	0.349
mad	0.048	0.054	0.082	0.055	0.050	0.060	0.046	0.053	0.028	0.066	0.057	0.054	0.106	0.125
idr	0.180	0.211	0.328	0.211	0.196	0.216	0.183	0.207	0.118	0.222	0.237	0.214	0.455	0.542
n=100,T=10														
mb	-0.001	0.002	-0.015	0.006	0.083	-0.010	-0.001	0.002	0.010	0.015	0.009	-0.004	0.192	-0.010
mad	0.014	0.022	0.016	0.020	0.018	0.022	0.014	0.021	0.014	0.022	0.024	0.025	0.028	0.025
idr	0.055	0.082	0.063	0.083	0.073	0.088	0.055	0.082	0.058	0.082	0.113	0.089	0.118	0.092
n=300,T=3														
mb	-0.015	-0.003	-0.006	0.000	0.299	0.073	0.000	-0.001	0.314	0.186	-0.001	0.000	0.173	0.340
mad	0.028	0.029	0.048	0.031	0.029	0.032	0.027	0.029	0.024	0.042	0.035	0.029	0.069	0.075
idr	0.113	0.117	0.189	0.126	0.126	0.132	0.105	0.116	0.110	0.179	0.136	0.114	0.252	0.354
n=300,T=10														
mb	0.000	-0.001	-0.004	0.000	0.097	-0.019	0.001	-0.001	0.012	0.012	0.005	-0.002	0.203	-0.016
mad	0.008	0.013	0.011	0.013	0.012	0.013	0.008	0.013	0.009	0.013	0.011	0.014	0.025	0.014
idr	0.033	0.046	0.039	0.047	0.045	0.049	0.032	0.047	0.033	0.047	0.045	0.049	0.096	0.054
	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.014	-0.006	0.198	0.096	-0.001	-0.003	0.002		0.190	0.099	-0.001		0.266	0.335
mad	0.052	0.053	0.068	0.062	0.049	0.054	0.057		0.067	0.062	0.049		0.032	0.091
idr	0.194	0.208	0.255	0.220	0.192	0.207	0.240		0.248	0.221	0.192		0.149	0.361
n=100,T=10														
mb	0.001	0.002	0.016	0.012	0.007	-0.002	0.012		0.016	0.012	0.007		0.378	0.106
mad	0.014	0.021	0.016	0.022	0.015	0.022	0.020		0.016	0.022	0.015		0.034	0.040
idr	0.057	0.083	0.063	0.084	0.061	0.084	0.088		0.064	0.084	0.061		0.313	0.150
n=300,T=3														
mb	-0.001	-0.002	0.231	0.102	0.004	-0.001	0.005		0.221	0.105	0.004		0.276	0.311
mad	0.030	0.029	0.041	0.032	0.030	0.028	0.030		0.042	0.033	0.030		0.031	0.080
idr	0.115	0.119	0.151	0.133	0.115	0.117	0.122		0.152	0.131	0.115		0.120	0.309
n=300,T=10														
mb	0.002	-0.002	0.016	0.009	0.003	-0.003	0.003		0.016	0.009	0.003		0.391	0.093
mad	0.008	0.013	0.01	0.013	0.009	0.013	0.009		0.010	0.013	0.009		0.018	0.025
idr	0.033	0.047	0.038	0.047	0.035	0.047	0.035		0.038	0.047	0.035		0.284	0.108

1. mb is median bias, md is median absolute deviation, and idr is 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 S.9: Estimates under $\gamma = 0.8, m = 20$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.287	-0.348	-0.064	0.001	-0.005	-0.348	-0.153	-0.354	
mad	0.069	0.362	0.163	0.084	0.079	0.283	0.162	0.348	
idr	0.286	1.588	0.667	0.295	0.295	0.826	0.563	1.159	
n=100,T=10									
mb	-0.060	-0.099	0.034	0.001	-0.001	-0.009	-0.035	-0.092	
mad	0.024	0.084	0.042	0.023	0.023	0.033	0.033	0.074	
idr	0.091	0.338	0.154	0.086	0.087	0.145	0.123	0.305	
n=300,T=3									
mb	-0.284	-0.141	-0.021	0.000	0.000	-0.064	-0.069	-0.141	
mad	0.041	0.245	0.090	0.045	0.044	0.215	0.108	0.238	
idr	0.165	0.979	0.373	0.172	0.178	0.814	0.359	0.72	
n=300,T=10									
mb	-0.059	-0.038	0.023	-0.001	-0.001	0.001	-0.015	-0.033	
mad	0.015	0.044	0.024	0.013	0.013	0.021	0.021	0.040	
idr	0.056	0.198	0.098	0.048	0.050	0.082	0.080	0.184	

	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.277	-0.188	-0.023	0.014	-0.15	-0.149	-0.325	-0.063	0.005
mad	0.339	0.323	0.102	0.104	0.148	0.148	0.411	0.100	0.546
idr	0.951	0.968	0.408	0.361	0.453	0.453	1.461	0.553	3.419
n=100,T=10									
mb	-0.054	-0.053	-0.002	0.002	-0.023	-0.023	-0.104	-0.010	0.027
mad	0.116	0.146	0.023	0.023	0.055	0.055	0.106	0.023	0.039
idr	0.375	0.471	0.087	0.092	0.224	0.225	0.480	0.085	0.182
n=300,T=3									
mb	-0.08	-0.024	-0.014	0.001	-0.068	-0.068	-0.138	-0.017	-0.018
mad	0.166	0.178	0.051	0.053	0.142	0.142	0.238	0.048	0.348
idr	0.909	0.702	0.205	0.214	0.356	0.356	0.798	0.201	4.246
n=300,T=10									
mb	-0.011	-0.010	-0.003	0.000	-0.001	-0.001	-0.036	-0.004	0.011
mad	0.031	0.048	0.014	0.014	0.040	0.040	0.050	0.013	0.018
idr	0.297	0.443	0.050	0.052	0.161	0.161	0.252	0.049	0.072

1. mb is median bias, md is median absolute deviation, and idr is 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 S.10: Estimates under $\gamma = 0.8, m = 1$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R
n=100,T=3								
mb	-0.017	-0.065	0.291	0.190	0.089	-0.007	0.237	0.161
mad	0.069	0.155	0.038	0.000	0.008	0.074	0.078	0.100
idr	0.277	0.701	0.156	0.000	0.031	0.312	0.314	0.397
n=100,T=10								
mb	0.000	-0.042	0.166	0.190	0.153	0.001	0.129	0.116
mad	0.018	0.051	0.015	0.000	0.003	0.017	0.023	0.027
idr	0.067	0.209	0.059	0.000	0.014	0.068	0.085	0.102
n=300,T=3								
mb	-0.012	-0.029	0.301	0.190	0.089	-0.005	0.243	0.178
mad	0.043	0.099	0.021	0.000	0.005	0.043	0.051	0.066
idr	0.163	0.378	0.075	0.000	0.019	0.166	0.189	0.247
n=300,T=10								
mb	-0.001	-0.012	0.175	0.190	0.154	0.000	0.141	0.130
mad	0.010	0.029	0.008	0.000	0.002	0.010	0.012	0.015
idr	0.036	0.123	0.031	0.000	0.008	0.034	0.050	0.059

	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.021	-0.010	0.190	0.190	-0.004	-0.004	-0.010	0.593	-0.242
mad	0.102	0.124	0.000	0.000	0.094	0.094	0.088	0.114	0.194
idr	0.569	0.575	0.000	0.000	0.265	0.267	0.255	0.433	0.670
n=100,T=10									
mb	-0.001	-0.003	0.190	0.190	0.000	0.000	-0.003	0.252	-0.073
mad	0.025	0.025	0.000	0.000	0.023	0.023	0.018	0.023	0.034
idr	0.151	0.136	0.017	0.000	0.090	0.090	0.073	0.090	0.130
n=300,T=3									
mb	-0.006	0.011	0.190	0.190	-0.007	-0.007	-0.012	0.613	-0.190
mad	0.048	0.056	0.000	0.000	0.058	0.058	0.051	0.060	0.140
idr	0.198	0.276	0.000	0.000	0.207	0.207	0.191	0.234	0.468
n=300,T=10									
mb	0.001	0.001	0.190	0.190	-0.001	-0.001	-0.003	0.255	-0.059
mad	0.010	0.010	0.000	0.000	0.011	0.011	0.010	0.013	0.023
idr	0.038	0.038	0.002	0.000	0.047	0.047	0.040	0.051	0.089

1. mb is median bias, md is median absolute deviation, and idr is 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 S.11: Estimates under $\gamma = 0.8, \beta = 1$, RE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	0.069	0.024	-0.052	-0.013	0.003	-0.012	0.001	0.003	0.002	-0.002	-0.016	0.001	-0.052	-0.013
mad	0.066	0.060	0.083	0.063	0.050	0.049	0.052	0.055	0.030	0.046	0.063	0.059	0.069	0.046
idr	0.256	0.224	0.327	0.234	0.199	0.172	0.208	0.216	0.114	0.172	0.292	0.225	0.253	0.180
n=100,T=10														
mb	0.012	0.000	-0.014	0.002	0.009	0.000	-0.001	0.001	-0.001	0.001	0.015	0.000	-0.015	0.003
mad	0.013	0.022	0.013	0.021	0.014	0.022	0.012	0.022	0.012	0.021	0.028	0.023	0.013	0.022
idr	0.052	0.086	0.055	0.086	0.055	0.084	0.047	0.085	0.045	0.083	0.168	0.094	0.051	0.083
n=300,T=3														
mb	0.069	0.025	-0.012	-0.005	0.004	-0.01	-0.003	-0.003	-0.003	-0.004	-0.002	0.001	-0.016	-0.011
mad	0.037	0.035	0.054	0.037	0.028	0.027	0.029	0.034	0.016	0.026	0.044	0.036	0.044	0.026
idr	0.155	0.131	0.203	0.141	0.121	0.103	0.117	0.126	0.066	0.100	0.174	0.143	0.157	0.105
n=300,T=10														
mb	0.012	0.000	-0.005	0.000	0.004	-0.001	0.000	0.000	0.000	0.000	0.004	0.000	-0.005	0.001
mad	0.008	0.013	0.009	0.013	0.009	0.013	0.007	0.013	0.007	0.013	0.013	0.013	0.008	0.013
idr	0.028	0.047	0.036	0.047	0.033	0.046	0.026	0.048	0.025	0.046	0.056	0.047	0.031	0.047

	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.004	0.001	0.008	-0.008	-0.007	0.000	0.007		-0.007	-0.013	-0.006		-0.004	0.000
mad	0.055	0.057	0.034	0.045	0.050	0.057	0.066		0.046	0.047	0.050		0.032	0.047
idr	0.214	0.215	0.135	0.175	0.205	0.215	0.262		0.183	0.181	0.205		0.123	0.178
n=100,T=10														
mb	0.000	0.001	0.005	0.000	0.003	0.000	0.012		0.001	0.000	0.003		0.015	-0.002
mad	0.013	0.022	0.012	0.022	0.013	0.022	0.021		0.012	0.023	0.013		0.017	0.022
idr	0.052	0.087	0.048	0.083	0.050	0.085	0.133		0.047	0.083	0.050		0.071	0.089
n=300,T=3														
mb	-0.007	-0.003	0.002	-0.007	0.003	0.001	0.000		-0.003	-0.012	0.003		-0.001	-0.003
mad	0.032	0.035	0.018	0.026	0.032	0.034	0.033		0.027	0.026	0.033		0.020	0.028
idr	0.129	0.127	0.069	0.099	0.135	0.132	0.149		0.112	0.107	0.135		0.075	0.103
n=300,T=10														
mb	0.000	0.000	0.001	0.000	0.004	0.000	0.005		0.000	0.000	0.004		0.011	-0.001
mad	0.008	0.013	0.007	0.013	0.009	0.013	0.009		0.007	0.013	0.009		0.009	0.013
idr	0.026	0.047	0.026	0.047	0.031	0.048	0.038		0.025	0.047	0.031		0.033	0.049

1. mb is median bias, md is median absolute deviation, and idr is 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 S.12: Estimates under $\gamma = 0.8, \beta = 1$, FE DGP

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	0.068	0.025	-0.05	-0.014	0.148	0.057	-0.002	0.001	0.154	0.124	-0.006	0.001	0.092	0.332
mad	0.067	0.060	0.082	0.062	0.018	0.057	0.049	0.054	0.008	0.056	0.062	0.059	0.059	0.154
idr	0.239	0.230	0.305	0.232	0.076	0.208	0.203	0.219	0.035	0.194	0.246	0.227	0.285	0.740
n=100,T=10														
mb	0.012	0.003	-0.014	0.003	0.083	0.008	0.000	0.003	0.021	0.021	0.011	0.003	0.154	0.013
mad	0.014	0.022	0.013	0.021	0.013	0.023	0.012	0.021	0.018	0.022	0.026	0.024	0.009	0.024
idr	0.052	0.084	0.054	0.085	0.058	0.087	0.048	0.084	0.183	0.086	0.158	0.089	0.037	0.089
n=300,T=3														
mb	0.075	0.027	-0.012	-0.004	0.155	0.055	-0.001	0.000	0.155	0.120	-0.003	0.001	0.115	0.269
mad	0.038	0.036	0.049	0.032	0.010	0.030	0.029	0.034	0.007	0.033	0.039	0.031	0.039	0.088
idr	0.147	0.131	0.186	0.133	0.041	0.123	0.117	0.123	0.031	0.135	0.146	0.125	0.162	0.396
n=300,T=10														
mb	0.013	-0.001	-0.004	0.000	0.095	0.003	0.000	-0.001	0.022	0.016	0.005	-0.001	0.157	0.007
mad	0.008	0.013	0.008	0.013	0.008	0.013	0.007	0.013	0.010	0.014	0.012	0.013	0.007	0.014
idr	0.028	0.047	0.032	0.046	0.031	0.048	0.026	0.046	0.041	0.047	0.048	0.046	0.026	0.051
	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.004	-0.002	0.139	0.064	0.003	-0.001	0.013		0.135	0.065	0.003		0.119	0.349
mad	0.057	0.053	0.022	0.058	0.054	0.055	0.062		0.022	0.058	0.053		0.013	0.103
idr	0.216	0.215	0.113	0.206	0.201	0.220	0.244		0.114	0.211	0.202		0.059	0.400
n=100,T=10														
mb	0.002	0.003	0.020	0.015	0.003	0.003	0.013		0.020	0.015	0.003		0.157	0.214
mad	0.013	0.022	0.015	0.022	0.013	0.022	0.023		0.015	0.022	0.013		0.009	0.056
idr	0.054	0.084	0.063	0.085	0.051	0.085	0.128		0.062	0.085	0.051		0.113	0.224
n=300,T=3														
mb	-0.004	0.000	0.145	0.067	0.006	0.002	0.004		0.143	0.068	0.006		0.124	0.342
mad	0.033	0.034	0.012	0.029	0.031	0.032	0.034		0.012	0.028	0.032		0.011	0.086
idr	0.127	0.125	0.051	0.12	0.123	0.126	0.133		0.056	0.120	0.122		0.046	0.334
n=300,T=10														
mb	0.000	-0.001	0.023	0.011	0.003	-0.001	0.005		0.023	0.011	0.003		0.161	0.206
mad	0.007	0.013	0.010	0.013	0.007	0.013	0.008		0.010	0.013	0.007		0.004	0.039
idr	0.027	0.046	0.039	0.047	0.028	0.047	0.037		0.039	0.047	0.028		0.020	0.139

1. mb is median bias, md is median absolute deviation, and idr is 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 S.13: Estimates under $\gamma = 0.9, m = 20$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.292	-0.647	-0.074	0.028	0.024	-0.175	-0.28	-0.629	
mad	0.075	0.418	0.155	0.062	0.069	0.303	0.172	0.406	
idr	0.295	1.694	0.762	0.212	0.222	0.889	0.649	1.483	
n=100,T=10									
mb	-0.073	-0.384	0.042	0.015	0.014	-0.083	-0.105	-0.342	
mad	0.025	0.202	0.028	0.022	0.023	0.059	0.043	0.175	
idr	0.094	0.713	0.120	0.087	0.088	0.194	0.172	0.611	
n=300,T=3									
mb	-0.291	-0.342	0.042	0.022	0.023	-0.635	-0.153	-0.319	
mad	0.042	0.333	0.046	0.048	0.047	0.206	0.150	0.323	
idr	0.170	1.572	0.303	0.154	0.161	0.861	0.527	1.204	
n=300,T=10									
mb	-0.070	-0.218	0.062	0.018	0.019	-0.126	-0.060	-0.189	
mad	0.016	0.147	0.011	0.013	0.013	0.023	0.040	0.134	
idr	0.056	0.536	0.047	0.047	0.047	0.162	0.152	0.488	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.467	-0.445	0.011	0.025	-0.231	-0.231	-0.716	-0.100	-0.678
mad	0.274	0.276	0.079	0.065	0.106	0.106	0.473	0.163	0.694
idr	0.861	1.089	0.312	0.299	0.415	0.415	1.695	1.190	3.576
n=100,T=10									
mb	-0.194	-0.169	0.011	0.016	-0.110	-0.110	-0.451	-0.023	0.075
mad	0.043	0.106	0.024	0.025	0.041	0.041	0.157	0.031	0.324
idr	0.270	0.412	0.098	0.100	0.169	0.169	0.590	0.128	1.365
n=300,T=3									
mb	-0.179	-0.215	0.007	0.011	-0.183	-0.183	-0.363	-0.022	-0.425
mad	0.278	0.261	0.063	0.062	0.118	0.118	0.360	0.069	0.455
idr	0.843	0.896	0.233	0.224	0.339	0.339	1.296	0.563	2.416
n=300,T=10									
mb	-0.199	-0.166	0.015	0.018	-0.077	-0.077	-0.265	-0.002	0.182
mad	0.049	0.115	0.014	0.014	0.042	0.042	0.177	0.019	0.143
idr	0.309	0.424	0.051	0.052	0.139	0.138	0.581	0.073	1.038

1. mb is median bias, md is median absolute deviation, and idr is 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 S.14: Estimates under $\gamma = 0.9, m = 1$

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	0.090	-0.028	0.272	0.090	0.004	-0.001	0.308	0.216	
mad	0.083	0.095	0.034	0.000	0.006	0.065	0.052	0.061	
idr	0.312	0.447	0.139	0.000	0.023	0.262	0.219	0.247	
n=100,T=10									
mb	0.031	-0.008	0.110	0.090	0.074	0.000	0.145	0.129	
mad	0.016	0.021	0.014	0.000	0.002	0.013	0.011	0.011	
idr	0.059	0.082	0.057	0.000	0.007	0.049	0.038	0.042	
n=300,T=3									
mb	0.089	-0.014	0.276	0.090	0.004	-0.004	0.311	0.219	
mad	0.051	0.065	0.019	0.000	0.004	0.038	0.035	0.043	
idr	0.181	0.237	0.069	0.000	0.014	0.141	0.129	0.149	
n=300,T=10									
mb	0.032	-0.001	0.113	0.090	0.074	0.000	0.147	0.132	
mad	0.008	0.011	0.008	0.000	0.001	0.006	0.005	0.006	
idr	0.034	0.049	0.031	0.000	0.004	0.026	0.023	0.025	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.020	-0.014	0.090	0.090	-0.002	-0.001	0.000	0.858	-0.142
mad	0.075	0.094	0.000	0.000	0.043	0.043	0.034	0.125	0.096
idr	0.351	0.336	0.000	0.000	0.202	0.202	0.175	0.448	0.446
n=100,T=10									
mb	0.000	-0.006	0.090	0.090	0.003	0.003	-0.002	0.311	-0.029
mad	0.013	0.016	0.000	0.000	0.014	0.014	0.014	0.015	0.012
idr	0.054	0.097	0.000	0.000	0.057	0.057	0.059	0.061	0.048
n=300,T=3									
mb	-0.004	-0.003	0.090	0.090	-0.007	-0.007	-0.012	0.873	-0.113
mad	0.038	0.043	0.000	0.000	0.035	0.035	0.026	0.069	0.082
idr	0.141	0.191	0.000	0.000	0.146	0.146	0.114	0.258	0.309
n=300,T=10									
mb	0.000	-0.001	0.090	0.090	0.002	0.002	-0.001	0.312	-0.028
mad	0.006	0.007	0.000	0.000	0.007	0.007	0.007	0.009	0.007
idr	0.027	0.036	0.000	0.000	0.029	0.029	0.030	0.033	0.028

1. mb is median bias, md is median absolute deviation, and idr is 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 S.15: Estimates under $\gamma = 0.9, \beta = 1$, RE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	0.139	0.052	-0.058	-0.021	0.018	-0.014	0.021	0.007	-0.003	-0.002	-0.014	0.005	-0.094	-0.027
mad	0.079	0.064	0.096	0.071	0.042	0.049	0.069	0.059	0.032	0.046	0.071	0.061	0.075	0.061
idr	0.295	0.260	0.376	0.259	0.186	0.180	0.184	0.229	0.121	0.178	0.311	0.242	0.311	0.247
n=100,T=10														
mb	0.032	0.006	-0.017	-0.002	0.017	0.001	-0.001	0.001	-0.004	0.000	0.015	0.005	-0.030	-0.003
mad	0.015	0.022	0.014	0.023	0.014	0.022	0.012	0.023	0.011	0.021	0.033	0.028	0.016	0.022
idr	0.058	0.087	0.055	0.085	0.049	0.085	0.042	0.086	0.040	0.083	0.194	0.108	0.067	0.084
n=300,T=3														
mb	0.130	0.057	-0.019	-0.008	0.028	-0.007	0.004	0.005	-0.007	-0.005	-0.005	0.001	-0.043	-0.011
mad	0.046	0.037	0.059	0.040	0.024	0.028	0.046	0.037	0.018	0.026	0.047	0.037	0.042	0.030
idr	0.179	0.147	0.221	0.150	0.111	0.106	0.148	0.139	0.074	0.104	0.202	0.151	0.158	0.125
n=300,T=10														
mb	0.033	0.006	-0.006	-0.001	0.012	0.001	-0.001	0.000	-0.004	0.000	0.006	0.003	-0.012	0.000
mad	0.008	0.013	0.009	0.013	0.009	0.012	0.007	0.013	0.006	0.013	0.016	0.014	0.009	0.013
idr	0.032	0.048	0.036	0.048	0.032	0.047	0.024	0.047	0.022	0.048	0.064	0.048	0.033	0.048
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.005	-0.001	0.014	-0.011	-0.008	-0.002	0.016		-0.022	-0.021	-0.007		-0.021	0.000
mad	0.050	0.058	0.030	0.047	0.048	0.057	0.036		0.064	0.050	0.050		0.029	0.051
idr	0.196	0.216	0.142	0.179	0.196	0.216	0.168		0.252	0.193	0.194		0.106	0.190
n=100,T=10														
mb	-0.001	0.001	0.006	0.000	0.000	0.000	0.012		0.000	-0.001	0.000		0.006	-0.002
mad	0.013	0.023	0.012	0.022	0.013	0.022	0.024		0.012	0.022	0.013		0.015	0.022
idr	0.048	0.083	0.046	0.085	0.047	0.085	0.189		0.047	0.085	0.047		0.085	0.090
n=300,T=3														
mb	-0.004	-0.002	0.006	-0.006	0.002	0.001	0.006		-0.007	-0.017	0.002		-0.013	-0.002
mad	0.034	0.036	0.016	0.028	0.034	0.034	0.037		0.042	0.029	0.034		0.024	0.027
idr	0.129	0.131	0.065	0.101	0.138	0.138	0.125		0.178	0.113	0.137		0.087	0.106
n=300,T=10														
mb	0.001	0.000	0.003	0.000	0.003	0.000	0.008		0.000	0.000	0.003		0.007	0.000
mad	0.008	0.013	0.007	0.013	0.008	0.013	0.010		0.007	0.013	0.008		0.009	0.013
idr	0.026	0.047	0.025	0.047	0.028	0.047	0.051		0.026	0.047	0.028		0.033	0.048

1. mb is median bias, md is median absolute deviation, and idr is 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 S.16: Estimates under $\gamma = 0.9, \beta = 1$, FE DGP

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	0.177	0.075	-0.023	-0.008	0.093	0.042	0.002	0.003	0.092	0.116	0.005	0.005	0.067	0.309
mad	0.088	0.066	0.061	0.059	0.009	0.055	0.051	0.056	0.002	0.049	0.055	0.066	0.035	0.230
idr	0.295	0.278	0.213	0.220	0.037	0.202	0.173	0.214	0.013	0.194	0.236	0.231	0.193	0.995
n=100,T=10														
mb	0.035	0.008	-0.009	0.000	0.058	0.018	0.001	0.002	0.002	0.020	0.011	0.007	0.100	0.019
mad	0.013	0.023	0.010	0.021	0.009	0.023	0.009	0.021	0.011	0.023	0.025	0.026	0.004	0.025
idr	0.051	0.086	0.041	0.086	0.036	0.087	0.037	0.085	0.042	0.089	0.176	0.104	0.014	0.091
n=300,T=3														
mb	0.175	0.074	-0.009	-0.004	0.095	0.042	0.000	0.002	0.093	0.104	-0.002	0.001	0.072	0.260
mad	0.047	0.039	0.037	0.030	0.004	0.029	0.029	0.033	0.001	0.034	0.034	0.032	0.026	0.170
idr	0.178	0.158	0.134	0.125	0.019	0.120	0.136	0.128	0.008	0.132	0.128	0.126	0.160	1.036
n=300,T=10														
mb	0.035	0.005	-0.002	-0.002	0.064	0.015	0.001	-0.001	0.002	0.016	0.006	0.001	0.102	0.011
mad	0.007	0.013	0.006	0.013	0.005	0.013	0.005	0.013	0.006	0.014	0.011	0.013	0.002	0.015
idr	0.028	0.048	0.022	0.046	0.020	0.049	0.021	0.046	0.024	0.049	0.048	0.049	0.010	0.055
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.001	-0.001	0.090	0.048	0.002	-0.002	0.024		0.088	0.049	0.002		0.057	0.492
mad	0.046	0.057	0.008	0.053	0.044	0.056	0.032		0.009	0.054	0.044		0.010	0.130
idr	0.170	0.210	0.042	0.195	0.161	0.209	0.145		0.044	0.198	0.161		0.047	0.538
n=100,T=10														
mb	0.000	0.002	0.021	0.016	0.001	0.002	0.009		0.020	0.016	0.001		0.024	0.266
mad	0.010	0.021	0.016	0.022	0.010	0.022	0.022		0.016	0.022	0.010		0.033	0.118
idr	0.039	0.085	0.064	0.087	0.039	0.084	0.168		0.064	0.087	0.039		0.128	0.490
n=300,T=3														
mb	0.001	0.001	0.092	0.048	0.003	0.001	0.009		0.091	0.050	0.003		0.062	0.475
mad	0.029	0.032	0.004	0.028	0.025	0.032	0.030		0.005	0.030	0.025		0.008	0.117
idr	0.109	0.125	0.020	0.118	0.101	0.123	0.100		0.022	0.118	0.101		0.033	0.445
n=300,T=10														
mb	0.001	-0.001	0.026	0.015	0.003	0.000	0.005		0.025	0.015	0.003		0.039	0.295
mad	0.005	0.013	0.012	0.013	0.005	0.013	0.007		0.011	0.013	0.005		0.028	0.097
idr	0.021	0.046	0.043	0.049	0.022	0.046	0.038		0.042	0.049	0.022		0.073	0.356

1. mb is median bias, md is median absolute deviation, and idr is 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 S.17: Estimates under $\gamma = 0.2, m = 20$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.152	-0.029	0.000	-0.005	-0.001	0.002	-0.005	-0.022	
mad	0.067	0.110	0.076	0.083	0.076	0.089	0.086	0.101	
idr	0.238	0.403	0.274	0.299	0.283	0.309	0.298	0.380	
n=100,T=10									
mb	-0.017	-0.007	0.010	-0.003	-0.003	-0.008	-0.002	-0.004	
mad	0.026	0.036	0.031	0.027	0.027	0.029	0.031	0.029	
idr	0.090	0.135	0.114	0.094	0.095	0.108	0.110	0.104	
n=300,T=3									
mb	-0.149	-0.011	-0.006	-0.001	-0.001	0.003	-0.009	-0.009	
mad	0.040	0.060	0.047	0.048	0.045	0.051	0.049	0.055	
idr	0.156	0.234	0.179	0.191	0.179	0.216	0.181	0.218	
n=300,T=10									
mb	-0.017	-0.003	0.003	-0.001	-0.001	-0.003	-0.003	-0.002	
mad	0.014	0.019	0.016	0.015	0.014	0.014	0.016	0.015	
idr	0.053	0.074	0.064	0.056	0.056	0.055	0.063	0.060	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.015	-0.003	-0.007	0.004	0.017	0.016	0.025	-0.011	0.038
mad	0.099	0.097	0.086	0.078	0.110	0.108	0.097	0.083	0.113
idr	0.541	0.442	0.302	0.283	0.457	0.459	0.346	0.297	0.454
n=100,T=10									
mb	-0.008	-0.004	-0.003	-0.002	0.014	0.014	-0.001	-0.003	-0.001
mad	0.032	0.027	0.027	0.026	0.037	0.037	0.027	0.027	0.027
idr	0.133	0.095	0.094	0.094	0.169	0.169	0.095	0.094	0.095
n=300,T=3									
mb	-0.003	-0.002	-0.003	0.000	0.020	0.020	0.006	-0.003	0.019
mad	0.055	0.046	0.050	0.046	0.058	0.058	0.052	0.048	0.055
idr	0.218	0.193	0.188	0.176	0.271	0.270	0.221	0.188	0.208
n=300,T=10									
mb	-0.002	-0.002	-0.001	0.000	0.007	0.007	-0.001	-0.001	-0.001
mad	0.015	0.015	0.015	0.014	0.017	0.017	0.014	0.015	0.014
idr	0.056	0.054	0.055	0.056	0.064	0.064	0.054	0.055	0.056

1. mb is median bias, md is median absolute deviation, and idr is 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 S.18: Estimates under $\gamma = 0.2, m = 1$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.150	-0.021	0.092	0.009	0.031	-0.011	0.002	0.014	
mad	0.060	0.111	0.087	0.076	0.075	0.080	0.086	0.095	
idr	0.234	0.440	0.326	0.314	0.285	0.302	0.317	0.355	
n=100,T=10									
mb	-0.016	-0.008	0.031	0.002	0.005	-0.005	0.004	0.018	
mad	0.024	0.037	0.033	0.025	0.026	0.025	0.031	0.027	
idr	0.095	0.131	0.120	0.099	0.098	0.108	0.109	0.104	
n=300,T=3									
mb	-0.144	-0.002	0.105	0.011	0.033	-0.001	0.011	0.026	
mad	0.038	0.062	0.051	0.049	0.048	0.050	0.050	0.053	
idr	0.144	0.247	0.187	0.179	0.171	0.183	0.183	0.208	
n=300,T=10									
mb	-0.015	-0.003	0.027	0.001	0.004	0.000	0.004	0.020	
mad	0.014	0.021	0.020	0.015	0.015	0.014	0.018	0.017	
idr	0.053	0.079	0.072	0.056	0.056	0.055	0.070	0.062	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.041	-0.023	0.001	0.037	0.018	0.018	0.007	0.000	0.053
mad	0.097	0.091	0.081	0.079	0.111	0.110	0.089	0.081	0.158
idr	0.521	0.478	0.306	0.291	0.505	0.506	0.383	0.312	1.262
n=100,T=10									
mb	-0.006	-0.001	0.001	0.007	0.018	0.018	0.000	0.002	0.006
mad	0.029	0.024	0.025	0.026	0.036	0.036	0.024	0.025	0.026
idr	0.140	0.099	0.098	0.097	0.170	0.170	0.100	0.098	0.101
n=300,T=3									
mb	-0.006	-0.005	0.011	0.035	0.005	0.005	0.002	0.009	0.089
mad	0.053	0.050	0.049	0.048	0.061	0.061	0.053	0.050	0.094
idr	0.207	0.185	0.178	0.169	0.246	0.246	0.202	0.180	0.619
n=300,T=10									
mb	0.000	0.000	0.001	0.007	0.008	0.008	0.001	0.001	0.005
mad	0.015	0.015	0.015	0.015	0.018	0.018	0.014	0.015	0.015
idr	0.055	0.055	0.056	0.057	0.066	0.066	0.055	0.056	0.058

1. mb is median bias, md is median absolute deviation, and idr is 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 S.19: Estimates under $\gamma = 0.2, \beta = 1$, RE DGP with non-normal disturbances

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	-0.037	-0.001	-0.013	0.000	0.005	-0.004	-0.007	0.000	-0.001	0.001	0.006	0.003	-0.002	-0.001
mad	0.050	0.050	0.054	0.050	0.052	0.047	0.045	0.050	0.039	0.046	0.062	0.050	0.047	0.047
idr	0.186	0.202	0.249	0.200	0.196	0.183	0.169	0.201	0.144	0.177	0.262	0.201	0.178	0.187
n=100,T=10														
mb	-0.002	-0.001	-0.010	0.002	0.002	-0.003	-0.002	-0.001	-0.002	-0.001	0.011	-0.008	-0.001	-0.002
mad	0.015	0.023	0.019	0.023	0.018	0.023	0.015	0.023	0.015	0.023	0.026	0.026	0.016	0.024
idr	0.065	0.090	0.069	0.090	0.067	0.089	0.064	0.090	0.064	0.087	0.111	0.104	0.066	0.088
n=300,T=3														
mb	-0.037	-0.002	-0.003	0.002	0.005	0.000	-0.003	0.002	-0.003	0.000	0.005	0.002	0.001	-0.001
mad	0.028	0.033	0.037	0.033	0.03	0.029	0.027	0.033	0.023	0.028	0.038	0.032	0.027	0.028
idr	0.105	0.115	0.132	0.118	0.110	0.104	0.103	0.117	0.087	0.105	0.135	0.121	0.105	0.101
n=300,T=10														
mb	-0.002	0.000	-0.003	0.001	0.001	-0.001	-0.001	-0.001	-0.001	0.000	0.005	-0.001	0.000	0.000
mad	0.008	0.013	0.010	0.013	0.010	0.013	0.008	0.013	0.008	0.012	0.012	0.013	0.009	0.013
idr	0.032	0.049	0.038	0.049	0.036	0.050	0.032	0.049	0.032	0.050	0.049	0.052	0.032	0.050
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.007	0.000	-0.005	-0.002	0.006	0.000	0.027		-0.003	-0.002	0.006		0.009	-0.004
mad	0.053	0.050	0.041	0.048	0.051	0.050	0.059		0.042	0.048	0.051		0.043	0.045
idr	0.191	0.201	0.163	0.180	0.185	0.201	0.228		0.155	0.177	0.185		0.158	0.182
n=100,T=10														
mb	-0.001	-0.002	0.002	-0.003	0.005	-0.004	0.009		0.002	-0.003	0.005		0.012	-0.008
mad	0.017	0.023	0.016	0.024	0.017	0.023	0.022		0.015	0.024	0.017		0.021	0.024
idr	0.071	0.09	0.065	0.088	0.070	0.089	0.097		0.065	0.088	0.070		0.082	0.090
n=300,T=3														
mb	0.000	0.002	-0.001	-0.001	0.007	0.002	0.013		0.000	-0.001	0.007		0.006	-0.001
mad	0.027	0.033	0.023	0.027	0.027	0.033	0.030		0.023	0.027	0.027		0.023	0.029
idr	0.107	0.116	0.090	0.103	0.103	0.116	0.115		0.087	0.105	0.103		0.084	0.106
n=300,T=10														
mb	0.000	0.000	0.001	-0.001	0.002	-0.002	0.003		0.001	-0.001	0.002		0.005	-0.002
mad	0.009	0.013	0.008	0.012	0.009	0.013	0.008		0.009	0.012	0.009		0.009	0.013
idr	0.035	0.050	0.032	0.050	0.034	0.050	0.036		0.032	0.050	0.034		0.033	0.051

1. mb is median bias, md is median absolute deviation, and idr is 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 S.20: Estimates under $\gamma = 0.2, \beta = 1$, FE DGP with non-normal disturbances

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	-0.042	-0.008	-0.017	-0.003	0.323	0.081	-0.010	-0.006	0.437	0.233	0.007	-0.001	0.188	0.374
mad	0.046	0.057	0.063	0.057	0.066	0.062	0.042	0.057	0.055	0.072	0.061	0.058	0.095	0.107
idr	0.169	0.203	0.243	0.207	0.272	0.241	0.166	0.202	0.251	0.259	0.257	0.206	0.447	0.481
n=100,T=10														
mb	-0.002	0.001	-0.012	0.004	0.072	-0.010	-0.002	0.001	0.006	0.015	0.012	-0.007	0.185	-0.010
mad	0.016	0.021	0.019	0.022	0.024	0.024	0.016	0.021	0.017	0.021	0.029	0.024	0.041	0.025
idr	0.061	0.084	0.073	0.087	0.082	0.086	0.062	0.083	0.064	0.082	0.126	0.107	0.160	0.096
n=300,T=3														
mb	-0.039	0.000	-0.007	0.003	0.327	0.084	-0.005	0.003	0.435	0.227	0.005	0.004	0.184	0.367
mad	0.027	0.033	0.040	0.033	0.053	0.038	0.026	0.033	0.055	0.055	0.036	0.033	0.069	0.082
idr	0.104	0.121	0.162	0.125	0.207	0.145	0.100	0.124	0.239	0.199	0.141	0.124	0.288	0.366
n=300,T=10														
mb	-0.002	0.002	-0.004	0.002	0.085	-0.014	-0.001	0.001	0.007	0.014	0.005	-0.002	0.192	-0.013
mad	0.008	0.013	0.011	0.013	0.013	0.014	0.009	0.013	0.009	0.014	0.014	0.014	0.035	0.016
idr	0.034	0.051	0.045	0.050	0.052	0.054	0.034	0.051	0.035	0.051	0.053	0.053	0.126	0.057
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.008	-0.004	0.195	0.124	0.004	-0.007	0.025		0.193	0.123	0.004		0.395	0.325
mad	0.045	0.056	0.077	0.060	0.042	0.058	0.056		0.071	0.061	0.042		0.061	0.090
idr	0.186	0.205	0.292	0.221	0.179	0.204	0.228		0.275	0.227	0.179		0.269	0.335
n=100,T=10														
mb	-0.001	0.001	0.011	0.013	0.005	-0.002	0.010		0.011	0.013	0.005		0.167	0.100
mad	0.018	0.022	0.019	0.022	0.018	0.021	0.023		0.019	0.022	0.018		0.109	0.044
idr	0.067	0.087	0.069	0.082	0.067	0.084	0.101		0.069	0.082	0.067		0.616	0.179
n=300,T=3														
mb	-0.005	0.002	0.201	0.133	0.003	0.002	0.008		0.197	0.132	0.003		0.398	0.305
mad	0.030	0.033	0.054	0.037	0.028	0.034	0.029		0.048	0.037	0.028		0.060	0.075
idr	0.106	0.123	0.205	0.140	0.108	0.127	0.122		0.186	0.141	0.108		0.258	0.280
n=300,T=10														
mb	-0.001	0.000	0.010	0.012	0.003	-0.001	0.004		0.010	0.012	0.003		0.142	0.107
mad	0.009	0.013	0.011	0.014	0.009	0.013	0.010		0.010	0.013	0.009		0.054	0.032
idr	0.036	0.051	0.041	0.05	0.033	0.051	0.038		0.041	0.050	0.033		0.594	0.127

1. mb is median bias, md is median absolute deviation, and idr is 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 S.21: Estimates under $\gamma = 0.5, m = 20$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.219	-0.052	-0.015	-0.004	-0.003	-0.080	-0.013	-0.059	
mad	0.082	0.171	0.104	0.096	0.089	0.137	0.102	0.173	
idr	0.303	0.652	0.365	0.355	0.320	0.578	0.380	0.644	
n=100,T=10									
mb	-0.032	-0.022	0.017	-0.001	-0.001	-0.315	-0.005	-0.017	
mad	0.026	0.044	0.033	0.026	0.026	0.058	0.030	0.035	
idr	0.089	0.172	0.127	0.094	0.092	0.398	0.110	0.140	
n=300,T=3									
mb	-0.215	-0.030	-0.009	-0.003	-0.006	-0.015	-0.010	-0.024	
mad	0.050	0.094	0.059	0.056	0.052	0.082	0.060	0.095	
idr	0.195	0.385	0.232	0.207	0.197	0.384	0.214	0.397	
n=300,T=10									
mb	-0.030	-0.008	0.005	-0.001	-0.001	-0.291	-0.003	-0.006	
mad	0.014	0.023	0.020	0.013	0.014	0.063	0.016	0.020	
idr	0.053	0.093	0.074	0.056	0.056	0.381	0.061	0.077	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.047	-0.001	-0.016	0.003	0.022	0.022	-0.003	-0.009	0.074
mad	0.206	0.199	0.097	0.093	0.144	0.146	0.138	0.097	0.181
idr	0.985	0.869	0.365	0.344	0.628	0.628	0.516	0.342	0.786
n=100,T=10									
mb	-0.016	-0.007	-0.002	-0.001	0.007	0.007	0.001	-0.002	0.002
mad	0.049	0.029	0.026	0.026	0.064	0.064	0.031	0.026	0.027
idr	0.269	0.110	0.095	0.094	0.441	0.442	0.109	0.094	0.099
n=300,T=3									
mb	-0.012	0.007	-0.006	0.001	0.021	0.021	-0.002	-0.006	0.061
mad	0.105	0.109	0.056	0.052	0.099	0.099	0.074	0.056	0.095
idr	0.766	0.607	0.220	0.197	0.418	0.417	0.322	0.208	0.510
n=300,T=10									
mb	-0.004	-0.003	-0.002	-0.001	0.016	0.016	-0.001	-0.001	0.000
mad	0.020	0.016	0.013	0.014	0.027	0.027	0.016	0.013	0.014
idr	0.188	0.057	0.057	0.057	0.148	0.148	0.059	0.056	0.057

1. mb is median bias, md is median absolute deviation, and idr is 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 S.22: Estimates under $\gamma = 0.5, m = 1$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.153	-0.245	0.268	0.154	0.205	-0.145	-0.026	-0.071	
mad	0.073	0.291	0.061	0.096	0.086	0.094	0.133	0.163	
idr	0.292	1.246	0.232	0.378	0.280	0.417	0.538	0.646	
n=100,T=10									
mb	-0.023	-0.030	0.151	0.043	0.056	-0.348	0.048	0.066	
mad	0.022	0.056	0.035	0.026	0.026	0.107	0.028	0.036	
idr	0.094	0.198	0.135	0.108	0.112	0.498	0.108	0.120	
n=300,T=3									
mb	-0.151	-0.062	0.302	0.149	0.208	-0.124	0.063	0.057	
mad	0.045	0.209	0.037	0.059	0.056	0.110	0.099	0.148	
idr	0.168	0.875	0.132	0.214	0.195	0.407	0.398	0.58	
n=300,T=10									
mb	-0.024	-0.010	0.161	0.044	0.057	-0.379	0.054	0.080	
mad	0.013	0.032	0.023	0.015	0.016	0.041	0.017	0.022	
idr	0.048	0.126	0.082	0.059	0.060	0.458	0.066	0.076	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.115	-0.121	0.099	0.165	-0.026	-0.026	-0.026	0.060	-0.456
mad	0.198	0.261	0.101	0.108	0.125	0.125	0.117	0.122	0.467
idr	0.759	0.987	0.429	0.429	0.599	0.600	0.589	0.539	1.785
n=100,T=10									
mb	-0.009	-0.005	0.039	0.074	0.012	0.012	-0.001	0.042	0.143
mad	0.032	0.024	0.025	0.030	0.043	0.044	0.024	0.025	0.070
idr	0.200	0.107	0.105	0.115	0.292	0.294	0.103	0.107	0.386
n=300,T=3									
mb	-0.024	-0.030	0.122	0.178	-0.002	-0.002	-0.001	0.123	-0.267
mad	0.090	0.115	0.058	0.067	0.097	0.097	0.094	0.066	0.315
idr	0.771	0.911	0.227	0.257	0.505	0.505	0.495	0.248	1.349
n=300,T=10									
mb	-0.001	-0.002	0.040	0.076	0.009	0.009	-0.001	0.043	0.127
mad	0.014	0.014	0.015	0.017	0.019	0.019	0.014	0.015	0.037
idr	0.055	0.053	0.057	0.069	0.083	0.083	0.053	0.058	0.157

1. mb is median bias, md is median absolute deviation, and idr is 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 S.23: Estimates under $\gamma = 0.5, \beta = 1$, RE DGP with non-normal disturbances

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	-0.013	-0.003	-0.025	-0.003	0.001	-0.006	-0.002	-0.002	0.001	-0.001	0.005	0.003	-0.020	-0.004
mad	0.059	0.051	0.072	0.053	0.055	0.046	0.051	0.051	0.035	0.047	0.081	0.054	0.057	0.046
idr	0.233	0.202	0.295	0.201	0.201	0.183	0.192	0.201	0.137	0.179	0.328	0.209	0.228	0.186
n=100,T=10														
mb	-0.001	-0.002	-0.011	0.002	0.003	-0.002	0.000	-0.002	-0.001	-0.001	0.014	-0.010	-0.005	-0.001
mad	0.014	0.024	0.016	0.023	0.016	0.023	0.013	0.024	0.013	0.024	0.031	0.028	0.015	0.024
idr	0.057	0.090	0.064	0.090	0.065	0.087	0.056	0.090	0.055	0.087	0.18	0.110	0.059	0.088
n=300,T=3														
mb	-0.014	-0.001	-0.007	0.002	0.002	-0.001	0.004	0.002	-0.001	0.000	0.006	0.001	-0.005	-0.001
mad	0.037	0.032	0.044	0.032	0.029	0.029	0.030	0.033	0.022	0.027	0.047	0.032	0.034	0.028
idr	0.129	0.117	0.155	0.116	0.116	0.104	0.116	0.115	0.083	0.106	0.172	0.121	0.127	0.106
n=300,T=10														
mb	0.000	0.000	-0.003	0.001	0.002	-0.001	0.000	-0.001	0.000	0.000	0.005	-0.002	-0.002	0.000
mad	0.007	0.012	0.009	0.013	0.009	0.012	0.007	0.012	0.008	0.013	0.015	0.013	0.008	0.012
idr	0.030	0.049	0.035	0.048	0.034	0.050	0.029	0.050	0.028	0.049	0.060	0.052	0.030	0.049
	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.002	-0.001	-0.001	-0.003	0.003	-0.001	0.023		-0.004	-0.004	0.003		0.002	-0.003
mad	0.059	0.052	0.040	0.047	0.051	0.052	0.077		0.038	0.046	0.051		0.040	0.046
idr	0.218	0.201	0.148	0.183	0.206	0.198	0.308		0.148	0.179	0.207		0.157	0.183
n=100,T=10														
mb	0.002	-0.003	0.003	-0.003	0.004	-0.004	0.012		0.003	-0.003	0.004		0.011	-0.006
mad	0.016	0.024	0.014	0.024	0.015	0.023	0.024		0.014	0.024	0.015		0.020	0.024
idr	0.064	0.092	0.059	0.088	0.063	0.09	0.149		0.056	0.087	0.063		0.078	0.091
n=300,T=3														
mb	0.002	0.002	0.000	-0.001	0.011	0.002	0.018		-0.001	-0.001	0.011		0.003	-0.001
mad	0.031	0.033	0.024	0.027	0.028	0.034	0.036		0.023	0.027	0.028		0.022	0.029
idr	0.115	0.117	0.086	0.104	0.112	0.116	0.145		0.082	0.105	0.112		0.084	0.106
n=300,T=10														
mb	0.001	-0.001	0.001	-0.001	0.005	-0.002	0.006		0.001	-0.001	0.005		0.009	-0.003
mad	0.008	0.013	0.008	0.013	0.008	0.013	0.010		0.008	0.013	0.008		0.009	0.013
idr	0.030	0.050	0.029	0.049	0.030	0.049	0.039		0.029	0.049	0.030		0.032	0.050

1. mb is median bias, md is median absolute deviation, and idr is 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 S.24: Estimates under $\gamma = 0.5, \beta = 1$, FE DGP with non-normal disturbances

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	-0.021	-0.005	-0.036	-0.005	0.283	0.073	-0.006	-0.003	0.315	0.186	0.002	0.004	0.163	0.365
mad	0.056	0.057	0.082	0.059	0.048	0.060	0.049	0.058	0.028	0.066	0.085	0.059	0.102	0.125
idr	0.218	0.206	0.307	0.220	0.184	0.229	0.185	0.207	0.124	0.228	0.323	0.215	0.436	0.534
n=100,T=10														
mb	-0.002	0.001	-0.013	0.003	0.084	-0.011	-0.002	0.001	0.010	0.015	0.010	-0.005	0.195	-0.010
mad	0.015	0.021	0.016	0.021	0.020	0.023	0.014	0.021	0.015	0.021	0.034	0.028	0.032	0.025
idr	0.054	0.085	0.065	0.086	0.077	0.086	0.053	0.085	0.057	0.085	0.188	0.117	0.124	0.097
n=300,T=3														
mb	-0.016	0.002	-0.014	0.001	0.294	0.076	0.000	0.004	0.31	0.183	0.000	0.004	0.171	0.347
mad	0.036	0.032	0.048	0.034	0.030	0.037	0.031	0.032	0.028	0.047	0.046	0.034	0.067	0.075
idr	0.132	0.124	0.203	0.128	0.120	0.138	0.117	0.122	0.121	0.171	0.172	0.123	0.270	0.350
n=300,T=10														
mb	0.000	0.001	-0.003	0.002	0.097	-0.015	0.000	0.001	0.013	0.014	0.006	-0.002	0.202	-0.014
mad	0.008	0.013	0.010	0.014	0.012	0.014	0.008	0.013	0.009	0.014	0.016	0.014	0.027	0.015
idr	0.030	0.049	0.040	0.050	0.046	0.052	0.030	0.049	0.032	0.049	0.066	0.052	0.098	0.055
	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.011	-0.003	0.209	0.099	0.000	-0.001	0.018		0.203	0.098	0.001		0.267	0.334
mad	0.052	0.058	0.066	0.060	0.056	0.058	0.077		0.064	0.061	0.056		0.035	0.093
idr	0.207	0.204	0.245	0.220	0.198	0.201	0.337		0.238	0.224	0.198		0.143	0.326
n=100,T=10														
mb	-0.001	0.001	0.015	0.011	0.004	-0.001	0.013		0.015	0.011	0.004		0.381	0.101
mad	0.016	0.021	0.017	0.022	0.017	0.021	0.028		0.016	0.022	0.017		0.034	0.047
idr	0.057	0.086	0.062	0.084	0.057	0.084	0.198		0.062	0.084	0.057		0.346	0.190
n=300,T=3														
mb	-0.002	0.002	0.222	0.106	0.005	0.004	0.012		0.216	0.108	0.005		0.273	0.315
mad	0.032	0.033	0.042	0.036	0.034	0.033	0.038		0.041	0.035	0.034		0.030	0.078
idr	0.124	0.124	0.164	0.134	0.124	0.122	0.156		0.162	0.137	0.124		0.132	0.289
n=300,T=10														
mb	0.001	0.000	0.015	0.011	0.005	-0.001	0.007		0.015	0.011	0.005		0.397	0.085
mad	0.008	0.013	0.01	0.014	0.008	0.013	0.011		0.010	0.014	0.008		0.017	0.029
idr	0.031	0.049	0.038	0.050	0.033	0.049	0.046		0.037	0.050	0.033		0.273	0.110

1. mb is median bias, md is median absolute deviation, and idr is 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 S.25: Estimates under $\gamma = 0.8, m = 20$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.283	-0.309	-0.092	-0.004	-0.006	-0.414	-0.153	-0.316	
mad	0.102	0.324	0.161	0.088	0.088	0.265	0.144	0.308	
idr	0.367	1.369	0.600	0.306	0.300	0.830	0.512	1.054	
n=100,T=10									
mb	-0.066	-0.096	0.035	-0.004	-0.004	-0.012	-0.034	-0.091	
mad	0.024	0.088	0.039	0.022	0.022	0.035	0.034	0.079	
idr	0.091	0.377	0.152	0.088	0.092	0.139	0.128	0.332	
n=300,T=3									
mb	-0.286	-0.149	-0.025	-0.003	-0.003	-0.080	-0.070	-0.148	
mad	0.058	0.248	0.091	0.048	0.049	0.241	0.109	0.245	
idr	0.231	0.977	0.370	0.188	0.179	0.791	0.363	0.737	
n=300,T=10									
mb	-0.061	-0.038	0.020	-0.001	-0.001	-0.003	-0.015	-0.036	
mad	0.014	0.049	0.025	0.014	0.014	0.020	0.020	0.046	
idr	0.053	0.197	0.092	0.053	0.054	0.082	0.083	0.184	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.228	-0.194	-0.019	0.013	-0.125	-0.125	-0.309	-0.069	-0.026
mad	0.327	0.324	0.107	0.115	0.176	0.176	0.353	0.099	0.473
idr	0.984	0.972	0.411	0.379	0.486	0.486	1.323	0.469	3.176
n=100,T=10									
mb	-0.133	-0.061	-0.008	-0.003	-0.048	-0.048	-0.08	-0.012	0.026
mad	0.110	0.133	0.023	0.024	0.097	0.097	0.092	0.023	0.042
idr	0.463	0.484	0.093	0.092	0.437	0.437	0.508	0.096	0.185
n=300,T=3									
mb	-0.130	-0.013	-0.015	0.005	-0.079	-0.079	-0.143	-0.017	-0.002
mad	0.222	0.207	0.055	0.056	0.135	0.135	0.243	0.052	0.290
idr	0.934	0.719	0.227	0.242	0.382	0.382	0.783	0.218	3.238
n=300,T=10									
mb	-0.021	-0.005	-0.003	-0.001	0.001	0.001	-0.034	-0.004	0.012
mad	0.100	0.062	0.014	0.014	0.070	0.070	0.047	0.014	0.020
idr	0.425	0.407	0.053	0.054	0.301	0.301	0.236	0.054	0.079

1. mb is median bias, md is median absolute deviation, and idr is 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 S.26: Estimates under $\gamma = 0.8, m = 1$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.018	-0.066	0.289	0.190	0.088	-0.010	0.236	0.158	
mad	0.087	0.166	0.041	0.000	0.010	0.095	0.083	0.107	
idr	0.353	0.689	0.166	0.000	0.040	0.418	0.320	0.378	
n=100,T=10									
mb	-0.003	-0.044	0.165	0.190	0.153	-0.003	0.13	0.116	
mad	0.016	0.051	0.016	0.000	0.004	0.016	0.021	0.025	
idr	0.066	0.195	0.059	0.000	0.016	0.066	0.081	0.094	
n=300,T=3									
mb	-0.010	-0.013	0.304	0.190	0.090	-0.005	0.252	0.188	
mad	0.053	0.101	0.021	0.000	0.006	0.061	0.054	0.069	
idr	0.199	0.380	0.084	0.000	0.021	0.224	0.196	0.255	
n=300,T=10									
mb	-0.001	-0.015	0.173	0.190	0.154	0.000	0.140	0.130	
mad	0.010	0.029	0.008	0.000	0.002	0.010	0.012	0.015	
idr	0.040	0.111	0.031	0.000	0.009	0.038	0.048	0.056	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.021	-0.034	0.190	0.190	0.010	0.011	-0.013	0.587	-0.241
mad	0.118	0.172	0.000	0.000	0.090	0.089	0.103	0.121	0.201
idr	0.552	0.579	0.000	0.000	0.307	0.308	0.258	0.480	0.640
n=100,T=10									
mb	-0.005	-0.007	0.190	0.190	0.003	0.003	-0.006	0.251	-0.073
mad	0.027	0.024	0.000	0.000	0.028	0.028	0.018	0.025	0.034
idr	0.178	0.137	0.021	0.000	0.126	0.126	0.076	0.099	0.123
n=300,T=3									
mb	-0.009	0.017	0.190	0.190	0.005	0.005	-0.002	0.609	-0.168
mad	0.058	0.077	0.000	0.000	0.074	0.074	0.067	0.065	0.144
idr	0.272	0.302	0.000	0.000	0.229	0.229	0.205	0.267	0.468
n=300,T=10									
mb	0.002	0.001	0.190	0.190	0.003	0.003	-0.002	0.257	-0.059
mad	0.010	0.011	0.000	0.000	0.014	0.014	0.011	0.014	0.022
idr	0.044	0.043	0.003	0.000	0.061	0.061	0.041	0.054	0.086

1. mb is median bias, md is median absolute deviation, and idr is 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 S.27: Estimates under $\gamma = 0.8, \beta = 1$, RE DGP with non-normal disturbances

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	0.074	0.023	-0.045	-0.017	0.004	-0.012	-0.002	0.000	0.000	-0.002	-0.002	0.003	-0.052	-0.012
mad	0.091	0.057	0.093	0.061	0.047	0.046	0.062	0.055	0.031	0.044	0.090	0.064	0.072	0.048
idr	0.320	0.226	0.360	0.228	0.189	0.183	0.242	0.206	0.116	0.182	0.417	0.241	0.266	0.191
n=100,T=10														
mb	0.013	-0.002	-0.014	-0.002	0.009	-0.003	0.000	-0.003	-0.001	-0.002	0.014	-0.002	-0.015	0.000
mad	0.012	0.023	0.015	0.023	0.014	0.023	0.012	0.023	0.011	0.023	0.042	0.024	0.014	0.023
idr	0.052	0.089	0.057	0.090	0.055	0.087	0.046	0.089	0.043	0.089	0.295	0.101	0.053	0.088
n=300,T=3														
mb	0.073	0.026	-0.012	0.003	0.004	-0.004	0.005	0.004	0.000	-0.001	0.004	0.002	-0.014	-0.006
mad	0.049	0.038	0.054	0.035	0.028	0.028	0.031	0.035	0.018	0.027	0.055	0.035	0.043	0.029
idr	0.187	0.135	0.198	0.133	0.111	0.105	0.129	0.122	0.067	0.105	0.212	0.140	0.156	0.109
n=300,T=10														
mb	0.013	-0.001	-0.005	0.000	0.004	-0.001	0.000	0.000	0.000	0.000	0.005	-0.001	-0.005	0.000
mad	0.008	0.012	0.008	0.013	0.008	0.012	0.007	0.013	0.007	0.012	0.020	0.013	0.008	0.012
idr	0.029	0.048	0.032	0.047	0.029	0.048	0.025	0.048	0.024	0.048	0.091	0.050	0.027	0.048
	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.008	-0.004	0.007	-0.007	-0.009	-0.001	0.019		-0.011	-0.015	-0.009		-0.006	-0.004
mad	0.059	0.055	0.033	0.046	0.057	0.055	0.092		0.047	0.046	0.057		0.036	0.049
idr	0.224	0.206	0.128	0.171	0.214	0.203	0.283		0.193	0.179	0.215		0.152	0.185
n=100,T=10														
mb	0.002	-0.002	0.004	-0.003	0.003	-0.002	0.009		0.001	-0.003	0.003		0.003	-0.004
mad	0.013	0.023	0.013	0.023	0.012	0.023	0.037		0.013	0.023	0.012		0.021	0.022
idr	0.051	0.09	0.048	0.089	0.049	0.089	0.325		0.045	0.089	0.049		0.092	0.089
n=300,T=3														
mb	0.006	0.005	0.004	-0.003	0.005	0.003	0.017		0.001	-0.008	0.005		0.003	0.000
mad	0.034	0.035	0.018	0.028	0.030	0.034	0.045		0.027	0.028	0.030		0.019	0.028
idr	0.127	0.121	0.068	0.104	0.119	0.120	0.184		0.107	0.111	0.119		0.073	0.105
n=300,T=10														
mb	0.001	0.000	0.002	-0.001	0.002	0.000	0.010		0.001	0.000	0.002		0.009	-0.002
mad	0.007	0.013	0.007	0.012	0.007	0.013	0.012		0.007	0.012	0.007		0.008	0.012
idr	0.027	0.048	0.025	0.048	0.026	0.047	0.069		0.025	0.048	0.026		0.032	0.047

1. mb is median bias, md is median absolute deviation, and idr is 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 S.28: Estimates under $\gamma = 0.8, \beta = 1$, FE DGP with non-normal disturbances

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	0.070	0.022	-0.036	-0.015	0.150	0.055	-0.004	-0.004	0.154	0.127	0.002	0.010	0.109	0.312
mad	0.079	0.063	0.089	0.064	0.018	0.056	0.055	0.061	0.009	0.057	0.097	0.062	0.051	0.162
idr	0.316	0.245	0.316	0.239	0.080	0.211	0.215	0.225	0.039	0.187	0.387	0.266	0.294	0.709
n=100,T=10														
mb	0.012	0.001	-0.014	0.002	0.083	0.006	-0.001	0.002	0.021	0.018	0.004	0.001	0.154	0.012
mad	0.013	0.022	0.013	0.022	0.014	0.023	0.012	0.022	0.019	0.023	0.040	0.025	0.009	0.023
idr	0.048	0.086	0.050	0.085	0.052	0.090	0.045	0.086	0.179	0.087	0.392	0.097	0.038	0.093
n=300,T=3														
mb	0.075	0.030	-0.008	-0.001	0.155	0.060	0.002	0.004	0.153	0.126	0.003	0.004	0.112	0.271
mad	0.047	0.037	0.051	0.035	0.010	0.034	0.035	0.035	0.007	0.036	0.057	0.037	0.037	0.089
idr	0.186	0.145	0.197	0.138	0.042	0.130	0.132	0.129	0.032	0.138	0.205	0.139	0.156	0.393
n=300,T=10														
mb	0.013	0.001	-0.004	0.001	0.094	0.005	0.000	0.001	0.021	0.019	0.010	0.000	0.157	0.009
mad	0.008	0.013	0.007	0.013	0.008	0.013	0.007	0.013	0.011	0.013	0.024	0.013	0.007	0.015
idr	0.030	0.048	0.032	0.048	0.032	0.050	0.026	0.048	0.054	0.050	0.102	0.048	0.028	0.053

	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.013	-0.004	0.138	0.061	-0.013	-0.003	0.018		0.136	0.062	-0.013		0.121	0.368
mad	0.057	0.060	0.021	0.057	0.054	0.060	0.086		0.022	0.057	0.054		0.016	0.102
idr	0.219	0.220	0.107	0.211	0.197	0.217	0.308		0.110	0.211	0.197		0.068	0.384
n=100,T=10														
mb	0.000	0.001	0.019	0.012	0.002	0.002	0.009		0.019	0.012	0.002		0.155	0.204
mad	0.012	0.021	0.016	0.021	0.012	0.022	0.039		0.015	0.022	0.012		0.011	0.059
idr	0.046	0.086	0.057	0.088	0.047	0.085	0.294		0.057	0.088	0.047		0.147	0.243
n=300,T=3														
mb	0.003	0.002	0.144	0.072	0.002	0.003	0.011		0.142	0.072	0.002		0.124	0.342
mad	0.036	0.035	0.012	0.032	0.034	0.036	0.044		0.013	0.033	0.034		0.011	0.082
idr	0.134	0.131	0.056	0.132	0.129	0.128	0.200		0.057	0.133	0.129		0.050	0.325
n=300,T=10														
mb	0.001	0.001	0.022	0.014	0.002	0.001	0.011		0.022	0.014	0.002		0.162	0.203
mad	0.007	0.013	0.010	0.013	0.007	0.013	0.014		0.009	0.013	0.007		0.005	0.046
idr	0.027	0.048	0.038	0.049	0.028	0.048	0.085		0.038	0.048	0.028		0.030	0.176

1. mb is median bias, md is median absolute deviation, and idr is 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 S.29: Estimates under $\gamma = 0.9, m = 20$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	-0.296	-0.581	-0.055	0.017	0.008	-0.215	-0.258	-0.549	
mad	0.105	0.380	0.128	0.073	0.076	0.326	0.164	0.373	
idr	0.388	1.566	0.685	0.223	0.234	0.884	0.588	1.401	
n=100,T=10									
mb	-0.074	-0.368	0.045	0.014	0.015	-0.083	-0.099	-0.327	
mad	0.026	0.165	0.025	0.023	0.023	0.055	0.040	0.149	
idr	0.099	0.639	0.114	0.089	0.090	0.204	0.155	0.549	
n=300,T=3									
mb	-0.294	-0.363	0.036	0.020	0.018	-0.277	-0.163	-0.343	
mad	0.062	0.349	0.046	0.045	0.045	0.371	0.158	0.342	
idr	0.244	1.427	0.267	0.162	0.171	0.860	0.496	1.125	
n=300,T=10									
mb	-0.070	-0.235	0.060	0.017	0.019	-0.128	-0.068	-0.205	
mad	0.016	0.148	0.011	0.014	0.014	0.023	0.041	0.136	
idr	0.061	0.553	0.043	0.049	0.050	0.153	0.145	0.503	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.331	-0.392	0.002	0.026	-0.225	-0.225	-0.622	-0.097	-0.554
mad	0.298	0.319	0.088	0.064	0.135	0.135	0.432	0.150	0.630
idr	0.919	0.999	0.323	0.290	0.451	0.450	1.620	1.034	3.040
n=100,T=10									
mb	-0.195	-0.183	0.008	0.013	-0.106	-0.106	-0.433	-0.020	0.131
mad	0.040	0.096	0.026	0.025	0.047	0.048	0.157	0.031	0.394
idr	0.291	0.422	0.094	0.092	0.208	0.208	0.548	0.121	1.802
n=300,T=3									
mb	-0.192	-0.248	0.005	0.009	-0.179	-0.179	-0.373	-0.022	-0.427
mad	0.261	0.294	0.065	0.065	0.137	0.137	0.373	0.074	0.460
idr	0.870	0.878	0.226	0.22	0.359	0.359	1.207	0.465	2.743
n=300,T=10									
mb	-0.198	-0.187	0.012	0.015	-0.083	-0.083	-0.284	-0.004	0.183
mad	0.050	0.095	0.014	0.014	0.045	0.045	0.185	0.020	0.147
idr	0.313	0.406	0.051	0.051	0.162	0.163	0.586	0.072	1.134

1. mb is median bias, md is median absolute deviation, and idr is 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 S.30: Estimates under $\gamma = 0.9, m = 1$ with non-normal disturbances

	LSDV	FD-W	FD-S	ML-W	ML-R	ML-C	MD-W	MD-R	
n=100,T=3									
mb	0.082	-0.027	0.268	0.090	0.004	-0.010	0.305	0.210	
mad	0.102	0.112	0.035	0.000	0.008	0.082	0.059	0.064	
idr	0.403	0.444	0.152	0.000	0.031	0.343	0.230	0.232	
n=100,T=10									
mb	0.030	-0.009	0.109	0.090	0.074	-0.002	0.145	0.128	
mad	0.015	0.021	0.016	0.000	0.002	0.012	0.009	0.011	
idr	0.060	0.076	0.056	0.000	0.008	0.054	0.036	0.039	
n=300,T=3									
mb	0.093	-0.005	0.279	0.090	0.005	-0.005	0.315	0.224	
mad	0.062	0.064	0.019	0.000	0.004	0.050	0.039	0.046	
idr	0.230	0.239	0.079	0.000	0.016	0.193	0.139	0.157	
n=300,T=10									
mb	0.031	-0.002	0.113	0.090	0.074	0.000	0.147	0.132	
mad	0.010	0.011	0.008	0.000	0.001	0.008	0.006	0.006	
idr	0.035	0.045	0.031	0.000	0.005	0.027	0.023	0.024	
	G-W	G-R	GS-W	GS-R	SG-W1	SG-W2	SG-R	Rao-W	Rao-R
n=100,T=3									
mb	-0.022	-0.019	0.090	0.090	0.000	0.000	0.000	0.858	-0.139
mad	0.094	0.119	0.000	0.000	0.040	0.040	0.028	0.133	0.105
idr	0.403	0.371	0.000	0.000	0.215	0.215	0.190	0.499	0.432
n=100,T=10									
mb	-0.003	-0.005	0.090	0.090	0.003	0.003	0.000	0.310	-0.028
mad	0.014	0.016	0.000	0.000	0.017	0.017	0.015	0.016	0.012
idr	0.073	0.085	0.000	0.000	0.075	0.075	0.060	0.063	0.046
n=300,T=3									
mb	-0.007	-0.008	0.090	0.090	0.000	0.000	-0.008	0.868	-0.098
mad	0.044	0.060	0.000	0.000	0.036	0.036	0.032	0.070	0.082
idr	0.176	0.236	0.000	0.000	0.160	0.160	0.136	0.291	0.304
n=300,T=10									
mb	0.000	0.000	0.090	0.090	0.004	0.004	-0.002	0.312	-0.027
mad	0.007	0.008	0.000	0.000	0.009	0.009	0.008	0.009	0.007
idr	0.028	0.032	0.000	0.000	0.037	0.037	0.031	0.038	0.029

1. mb is median bias, md is median absolute deviation, and idr is 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 S.31: Estimates under $\gamma = 0.9, \beta = 1$, RE DGP with non-normal disturbances

	LSDV:γ, β		FD-W:γ, β		FD-S:γ, β		ML-W:γ, β		ML-R:γ, β		MD-W:γ, β		MD-R:γ, β	
n=100,T=3.														
mb	0.132	0.055	-0.058	-0.022	0.020	-0.007	0.009	-0.001	-0.005	-0.008	-0.003	0.004	-0.091	-0.027
mad	0.109	0.070	0.103	0.066	0.042	0.048	0.081	0.052	0.031	0.048	0.098	0.072	0.076	0.056
idr	0.393	0.269	0.399	0.255	0.172	0.180	0.200	0.210	0.130	0.192	0.485	0.275	0.291	0.237
n=100,T=10														
mb	0.033	0.006	-0.017	-0.005	0.017	-0.001	0.000	-0.002	-0.004	-0.002	0.006	0.002	-0.031	-0.005
mad	0.014	0.024	0.015	0.022	0.014	0.023	0.011	0.023	0.011	0.023	0.050	0.032	0.017	0.022
idr	0.060	0.090	0.058	0.093	0.054	0.086	0.043	0.090	0.039	0.088	0.533	0.129	0.068	0.091
n=300,T=3														
mb	0.141	0.060	-0.018	-0.002	0.025	-0.004	0.014	0.011	-0.004	0.000	0.000	0.003	-0.049	-0.003
mad	0.058	0.044	0.052	0.038	0.023	0.029	0.057	0.038	0.021	0.029	0.057	0.037	0.043	0.035
idr	0.227	0.155	0.220	0.147	0.103	0.106	0.163	0.132	0.076	0.104	0.243	0.163	0.161	0.139
n=300,T=10														
mb	0.032	0.005	-0.006	-0.001	0.012	0.001	0.000	0.000	-0.003	0.000	0.005	0.002	-0.011	-0.001
mad	0.009	0.013	0.009	0.013	0.008	0.013	0.007	0.013	0.006	0.013	0.025	0.014	0.008	0.013
idr	0.034	0.048	0.033	0.048	0.030	0.048	0.025	0.047	0.022	0.047	0.129	0.055	0.032	0.047

	G-W:γ, β		G-R:γ, β		SG-W1:γ, β		SG-W2:γ		SG-R1:γ, β		SG-R2:γ		SG-RS:γ, β	
n=100,T=3.														
mb	-0.010	-0.008	0.011	-0.008	-0.008	-0.007	0.025		-0.025	-0.020	-0.009		-0.024	-0.006
mad	0.056	0.057	0.031	0.044	0.057	0.055	0.033		0.065	0.046	0.058		0.033	0.051
idr	0.215	0.209	0.135	0.176	0.215	0.207	0.208		0.242	0.192	0.215		0.134	0.196
n=100,T=10														
mb	0.001	-0.001	0.007	-0.002	0.001	-0.002	0.001		0.000	-0.003	0.001		-0.004	-0.005
mad	0.013	0.023	0.013	0.023	0.012	0.023	0.049		0.012	0.022	0.012		0.019	0.023
idr	0.047	0.091	0.050	0.088	0.045	0.089	0.258		0.048	0.087	0.045		0.113	0.091
n=300,T=3														
mb	0.005	0.005	0.006	-0.003	0.002	0.004	0.021		-0.010	-0.012	0.002		-0.010	0.005
mad	0.031	0.035	0.016	0.028	0.030	0.035	0.029		0.040	0.030	0.030		0.021	0.030
idr	0.120	0.122	0.062	0.107	0.112	0.123	0.137		0.161	0.119	0.112		0.081	0.111
n=300,T=10														
mb	0.000	-0.001	0.003	0.000	0.000	-0.001	0.012		0.000	0.000	0.000		0.007	-0.001
mad	0.006	0.013	0.006	0.013	0.007	0.013	0.016		0.007	0.013	0.007		0.009	0.012
idr	0.026	0.047	0.026	0.047	0.026	0.047	0.067		0.026	0.048	0.026		0.034	0.049

1. mb is median bias, md is median absolute deviation, and idr is 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 S.32: Estimates under $\gamma = 0.9, \beta = 1$, FE DGP with non-normal disturbances

	LSDV:γ,β		FD-W:γ,β		FD-S:γ,β		ML-W:γ,β		ML-R:γ,β		MD-W:γ,β		MD-R:γ,β	
n=100,T=3.														
mb	0.174	0.066	-0.015	-0.015	0.093	0.040	-0.003	-0.004	0.092	0.116	0.007	0.010	0.076	0.260
mad	0.093	0.073	0.060	0.065	0.009	0.054	0.053	0.060	0.002	0.050	0.085	0.068	0.029	0.222
idr	0.361	0.294	0.231	0.224	0.040	0.212	0.176	0.225	0.015	0.182	0.427	0.297	0.204	1.049
n=100,T=10														
mb	0.033	0.008	-0.008	-0.001	0.057	0.018	-0.001	0.000	0.001	0.019	-0.007	-0.001	0.100	0.017
mad	0.012	0.022	0.008	0.022	0.007	0.023	0.008	0.022	0.009	0.023	0.044	0.028	0.003	0.024
idr	0.044	0.087	0.036	0.087	0.031	0.091	0.031	0.086	0.039	0.088	0.406	0.122	0.015	0.096
n=300,T=3														
mb	0.180	0.081	-0.007	-0.002	0.095	0.045	0.002	0.004	0.093	0.106	0.004	0.004	0.076	0.253
mad	0.057	0.042	0.037	0.036	0.005	0.033	0.033	0.035	0.001	0.036	0.050	0.038	0.025	0.170
idr	0.224	0.175	0.140	0.136	0.020	0.128	0.139	0.135	0.009	0.143	0.199	0.143	0.164	0.935
n=300,T=10														
mb	0.033	0.008	-0.003	0.001	0.063	0.018	0.000	0.001	0.001	0.020	0.008	0.003	0.102	0.014
mad	0.007	0.014	0.006	0.013	0.005	0.014	0.005	0.013	0.006	0.013	0.022	0.015	0.003	0.015
idr	0.028	0.050	0.022	0.048	0.020	0.051	0.020	0.049	0.026	0.054	0.129	0.060	0.011	0.057

	G-W:γ,β		G-R:γ,β		SG-W1:γ,β		SG-W2:γ		SG-R1:γ,β		SG-R2:γ		SG-RS:γ,β	
n=100,T=3.														
mb	-0.011	-0.008	0.090	0.044	-0.010	-0.008	0.028		0.089	0.046	-0.01		0.056	0.518
mad	0.048	0.061	0.009	0.054	0.044	0.061	0.035		0.009	0.055	0.045		0.011	0.141
idr	0.170	0.217	0.044	0.206	0.168	0.215	0.234		0.049	0.208	0.168		0.065	0.533
n=100,T=10														
mb	-0.001	0.001	0.019	0.016	-0.001	0.001	-0.003		0.018	0.016	-0.001		0.041	0.267
mad	0.008	0.022	0.014	0.023	0.008	0.022	0.053		0.014	0.023	0.008		0.037	0.161
idr	0.032	0.088	0.056	0.087	0.032	0.086	0.193		0.056	0.088	0.032		0.315	0.666
n=300,T=3														
mb	0.001	0.001	0.092	0.052	-0.001	0.001	0.015		0.091	0.054	-0.001		0.061	0.473
mad	0.029	0.036	0.005	0.032	0.028	0.036	0.035		0.005	0.033	0.028		0.008	0.116
idr	0.111	0.135	0.022	0.123	0.108	0.132	0.123		0.024	0.130	0.108		0.037	0.475
n=300,T=10														
mb	0.000	0.001	0.025	0.018	0.000	0.001	0.008		0.024	0.018	0.000		0.041	0.293
mad	0.005	0.013	0.011	0.013	0.005	0.013	0.015		0.011	0.013	0.005		0.031	0.116
idr	0.022	0.049	0.043	0.050	0.022	0.049	0.159		0.043	0.050	0.022		0.086	0.478

1. mb is median bias, md is median absolute deviation, and idr is 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).