Estimation of linear dynamic panel data models with time-invariant regressors

Sebastian Kripfganz[*] Claudia Schwarz[†]

Online Appendix

Appendix B One-stage GMM estimation

This appendix serves as a supplement to Section 3 in the main paper.

B.1 GMM estimation of linear dynamic panel data models

For the static model with strictly exogenous regressors $\mathbf{x}_{it}$, Hausman and Taylor (1981) propose an instrumental variable estimator that uses deviations from their within-group means, $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$, as instruments for the regressors $\mathbf{x}_{it}$, and the within-group means $\bar{\mathbf{x}}_{1i}$ as instruments for $\mathbf{f}_{2i}$.[1] The time-invariant regressors $\mathbf{f}_{1i}$ serve as their own instruments. We can extend this estimator to accommodate the dynamic setup by adding an appropriate instrument for the lagged dependent variable. For example, Anderson and Hsiao (1981) propose to use $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$. With $\mathbf{y}_{i,(-2)} = (y_{i0}, y_{i1}, \ldots, y_{i,T-2})'$, the resulting estimator satisfies the moment conditions $E[\mathbf{Z}_i'\mathbf{H}_i\mathbf{e}_i] = \mathbf{0}$ with

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{y}_{i,(-2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_{1i} & \mathbf{F}_{1i} \end{pmatrix}, \quad \text{and} \quad \mathbf{H}_i = \begin{pmatrix} \mathbf{D}_i \\ \mathbf{Q}_i \\ \mathbf{P}_i \end{pmatrix},$$

[*]Corresponding author: University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, UK. E-mail: S.Kripfganz@exeter.ac.uk

[†]European Central Bank, SSM Risk Analysis. E-mail: claudia.schwarz@ecb.int

[1]To improve on the efficiency of the estimator, Amemiya and MaCurdy (1986) propose to use all time periods of $\mathbf{x}_{1it}$ separately as instruments instead of the within-group means. Breusch et al. (1989) additionally suggest using the deviation of each individual time period from the within-group means as separate instruments.

for the $(T-1) \times T$ first-difference transformation matrix $\mathbf{D}_i = [(\mathbf{0}, \mathbf{I}_{T-1}) - (\mathbf{I}_{T-1}, \mathbf{0})]$, and the $T \times T$ idempotent and symmetric projection matrices $\mathbf{P}_i = \boldsymbol{\iota}_T(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T)^{-1}\boldsymbol{\iota}_T'$ and $\mathbf{Q}_i = \mathbf{I}_T - \mathbf{P}_i$, where $\mathbf{P}_i$ and $\mathbf{Q}_i$ transform the observations into within-group means and deviations from within-group means, respectively. Importantly, both $\mathbf{D}_i$ and $\mathbf{Q}_i$ are orthogonal to time-invariant variables. Due to the block-diagonal structure of $\mathbf{Z}_i$, only the instruments $(\mathbf{X}_{1i}, \mathbf{F}_{1i})$ in the lower-right block of $\mathbf{Z}_i$ are of use to identify $\boldsymbol{\gamma}$. Therefore, as in the static model of Hausman and Taylor (1981), a necessary condition for the identification of all coefficients $(\boldsymbol{\theta}', \boldsymbol{\gamma}')'$ with this extended estimator is $K_{x1} \geq K_{f2}$.

Since the above estimator does not exploit all model-implied moment conditions, it will be inefficient. Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) derive additional linear moment conditions for the model in first differences and in levels. Ahn and Schmidt (1995) add further nonlinear moment conditions valid under absence of serial correlation and under homoskedasticity of u_{it}. We present the moment conditions in Appendix A. For the model in first differences, $E[\mathbf{Z}_{di}'\mathbf{D}_i\mathbf{e}_i] = 0$, and in levels, $E[\mathbf{Z}_{li}'\mathbf{e}_i] = 0$, the moment conditions can be combined by defining

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{di} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li} \end{pmatrix}, \quad \text{and} \quad \mathbf{H}_i = \begin{pmatrix} \mathbf{D}_i \\ \mathbf{I}_T \end{pmatrix}.$$

With the full set of linear moment conditions, the instrument matrix for the first-differenced model is $\mathbf{Z}_{di} = (\mathbf{Z}_{dyi}, \mathbf{Z}_{dxi}, \mathbf{I}_{T-1} \otimes \mathbf{f}_i')$, where $\mathbf{Z}_{dyi} = \text{diag}(\mathbf{z}_{dyi2}', \mathbf{z}_{dyi3}', \ldots, \mathbf{z}_{dyiT}')$ and $\mathbf{Z}_{dxi} = \text{diag}(\mathbf{z}_{dxi2}', \mathbf{z}_{dxi3}', \ldots, \mathbf{z}_{dxiT}')$, and where $\mathbf{z}_{dyit} = (y_{i0}, y_{i1}, \ldots, y_{i,t-2})'$. Moreover, $\mathbf{z}_{dxit} = (\mathbf{x}_{i0}', \mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$ under strict exogeneity, while $\mathbf{z}_{dxit} = (\mathbf{x}_{i0}', \mathbf{x}_{i1}', \ldots, \mathbf{x}_{i,t-1}')'$ for predetermined regressors. For the untransformed model, $\mathbf{Z}_{li} = (\mathbf{Z}_{lyi}, \mathbf{Z}_{lx1i}, \mathbf{Z}_{lx2i}, \mathbf{F}_{1i})$, with $\mathbf{Z}_{lyi} = (\mathbf{0}, \text{diag}(\Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{i,T-1}))'$, $\mathbf{Z}_{lx1i} = ((\mathbf{x}_{1i0}, \mathbf{0})', \text{diag}(\mathbf{x}_{1i1}', \mathbf{x}_{1i2}', \ldots, \mathbf{x}_{1iT}'))$, and $\mathbf{Z}_{lx2i} = \text{diag}(\Delta\mathbf{x}_{2i1}', \Delta\mathbf{x}_{2i2}', \ldots, \Delta\mathbf{x}_{2iT}')$.

We can derive the GMM estimator as a minimum distance estimator based on the sample moments $N^{-1}\mathbf{Z}'\mathbf{H}\mathbf{e}$:

$$\left(\tilde{\boldsymbol{\theta}}', \tilde{\gamma}'\right)' = \arg \min_{(\mathbf{b}_1', \mathbf{b}_2')'} \left(\frac{1}{N}\mathbf{Z}'\mathbf{H}(\mathbf{y} - \mathbf{W}_{yx}\mathbf{b}_1 - \mathbf{F}\mathbf{b}_2)\right)' \mathbf{V}_N \left(\frac{1}{N}\mathbf{Z}'\mathbf{H}(\mathbf{y} - \mathbf{W}_{yx}\mathbf{b}_1 - \mathbf{F}\mathbf{b}_2)\right).$$

The closed-form solution for the whole parameter vector is given in equation (5), provided that $\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\mathbf{H}\mathbf{W}_{yxf}$ is nonsingular. Decompose the weighting matrix $\mathbf{V}_N = \mathbf{L}\mathbf{L}'$, and define $\mathbf{y}^* = \mathbf{L}'\mathbf{Z}'\mathbf{H}\mathbf{y}$, $\mathbf{W}_{yx}^* = \mathbf{L}'\mathbf{Z}'\mathbf{H}\mathbf{W}_{yx}$, and $\mathbf{F}^* = \mathbf{L}'\mathbf{Z}'\mathbf{H}\mathbf{F}$. The following partitioned regression result is useful:

$$\tilde{\boldsymbol{\theta}} = (\mathbf{W}_{yx}^*{}'\mathbf{Q}_F\mathbf{W}_{yx}^*)^{-1}\mathbf{W}_{yx}^*{}'\mathbf{Q}_F\mathbf{y}^*, \tag{B.1}$$

$$\tilde{\boldsymbol{\gamma}} = (\mathbf{F}^*{}'\mathbf{F}^*)^{-1}\mathbf{F}^*{}'\left(\mathbf{y}^* - \mathbf{W}_{yx}^*\tilde{\boldsymbol{\theta}}\right), \tag{B.2}$$

where $\mathbf{Q}_F = \mathbf{I}_{K_z} - \mathbf{F}^*(\mathbf{F}^*{}'\mathbf{F}^*)^{-1}\mathbf{F}^*{}'$ is an idempotent and symmetric projection matrix.

From Proposition 1, an unrestricted estimate of the variance matrix Ω of the joint asymptotic distribution can be obtained as $\tilde{\Omega} = \frac{1}{N}\sum_{i=1}^N \tilde{\boldsymbol{\varphi}}_i\tilde{\boldsymbol{\varphi}}_i'$, based on the estimated influence function of the one-stage GMM estimator:

$$\tilde{\boldsymbol{\varphi}}_i = N\left(\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N\mathbf{Z}'\mathbf{H}\mathbf{W}_{yxf}\right)^{-1}\mathbf{W}_{yxf}'\mathbf{H}'\mathbf{Z}\mathbf{V}_N(\mathbf{Z}_i'\mathbf{H}_i\tilde{\mathbf{e}}_i), \tag{B.3}$$

where $\tilde{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{W}_{yxi}\tilde{\boldsymbol{\theta}} - \mathbf{F}_i\tilde{\boldsymbol{\gamma}}$.

B.2 Asymptotic distribution of the GMM estimator

In the following, we present a sufficient set of assumptions for the identification of the coefficients and regularity conditions for consistency and asymptotic normality of the GMM estimator. We restrict the exposition again to linear moment conditions. For notational convenience, denote the moment vector by $\mathbf{m}_i(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{Z}_i'\mathbf{H}_i(\mathbf{y}_i - \mathbf{W}_{yxi}\mathbf{b}_1 - \mathbf{F}_i\mathbf{b}_2)$ such that $\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \mathbf{Z}_i'\mathbf{H}_i\mathbf{e}_i$. The corresponding negative gradient is $\mathbf{S}_i = \mathbf{Z}_i'\mathbf{H}_i\mathbf{W}_{yxfi}$. The latter is independent of the parameters due to the restriction to linear moment conditions.

Assumption B.1: The parameter space is the convex set $\Theta = (-1, 1) \times \mathbb{R}^{K_x} \times \mathbb{R}^{K_f}$ with $(\boldsymbol{\theta}', \boldsymbol{\gamma}') \in \Theta$.

Assumption B.2: A weak law of large numbers holds element-wise for the moment vector such that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{m}_i(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{m}(\mathbf{b}_1, \mathbf{b}_2)$ for all $(\mathbf{b}_1', \mathbf{b}_2')' \in \Theta$.

Assumption B.3: The moment conditions $E[\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})] = \mathbf{0}$ are satisfied.

Assumption B.4: A weak law of large numbers holds element-wise for the negative gradient of the moment vector such that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{S}_i = \mathbf{S}$.

Assumption B.5: A weak law of large numbers holds element-wise for the weighting matrix $\mathbf{V}_N$ such that $\text{plim}_{N\to\infty} \mathbf{V}_N = \mathbf{V}$ is positive semi-definite and $\mathbf{VS}$ has full column rank.

Lemma B.1: Under Assumptions B.1 to B.5, all parameters are identified.

Proof. By Assumptions B.2 and B.5 and by Slutsky's theorem, the GMM criterion function converges in probability to $\mathbf{m}(\mathbf{b}_1, \mathbf{b}_2)'\mathbf{V}\mathbf{m}(\mathbf{b}_1, \mathbf{b}_2) \geq 0$ for all $(\mathbf{b}_1', \mathbf{b}_2')' \in \Theta$. Denote the corresponding minimizer by $(\tilde{\boldsymbol{\theta}}_\infty', \tilde{\gamma}_\infty')'$. By Assumptions B.2 to B.4, the first-order condition is $\mathbf{S}'\mathbf{V}\mathbf{m}(\tilde{\boldsymbol{\theta}}_\infty, \tilde{\gamma}_\infty) = \mathbf{0}$ with $\mathbf{m}(\tilde{\boldsymbol{\theta}}_\infty, \tilde{\gamma}_\infty) = \mathbf{S}[(\boldsymbol{\theta}', \boldsymbol{\gamma}')' - (\tilde{\boldsymbol{\theta}}_\infty', \tilde{\gamma}_\infty')']$. By Assumption B.5, $\mathbf{S}'\mathbf{VS}$ is nonsingular. Hence, with Assumption B.1, $(\tilde{\boldsymbol{\theta}}_\infty', \tilde{\gamma}_\infty')' = (\boldsymbol{\theta}', \boldsymbol{\gamma}')'$ is the unique minimizer of the asymptotic GMM criterion function on Θ. □

Assumption B.6: The moment vector evaluated at the population values, $\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})$, is independently distributed across i with $\sup_{i,N} E[\|\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\|^{2+\delta}] < \infty$ for some $\delta > 0$, and where $\lim_{N\to\infty} N^{-1} \sum_{i=1}^N E[\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})'] = \Xi$ is a positive-definite matrix.[2]

Lemma B.2: Under Assumptions B.3 and B.6, $N^{-\frac{1}{2}} \sum_{i=1}^N \mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0, \Xi)$.

Proof. Let $\mathbf{c} \in \mathbb{R}^{K_z}$ with $\mathbf{c}'\mathbf{c} = 1$. By Assumption B.6 and the Cauchy-Schwarz inequality, $\sup_{i,N} E[|\mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})|^{2+\delta}] \leq \sup_{i,N} E[\|\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\|^{2+\delta}] < \infty$. With Ξ being positive definite, $0 < \lim_{N\to\infty} N^{-1} \sum_{i=1}^N E[\mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})'\mathbf{c}] = \mathbf{c}'\Xi\mathbf{c} < \infty$. Hence, for any $\delta > 0$,

$$
\begin{aligned}
0 &\leq \lim_{N\to\infty} \left(\sum_{i=1}^N E[\mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})'\mathbf{c}] \right)^{-(2+\delta)} \sum_{i=1}^N E[|\mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})|^{2+\delta}] \\
&\leq (\mathbf{c}'\Xi\mathbf{c})^{-(2+\delta)} \lim_{N\to\infty} N^{-\frac{\delta}{2}} \sup_{i,N} \sum_{i=1}^N E[|\mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})|^{2+\delta}] = 0.
\end{aligned}
$$

such that Lyapunov's condition (and thus Lindeberg's condition) is satisfied. With Assumption B.3, the Lindeberg-Feller central limit theorem implies $N^{-\frac{1}{2}} \sum_{i=1}^N \mathbf{c}'\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0, \mathbf{c}'\Xi\mathbf{c})$. The claim follows from the Cramér-Wold theorem. □

It remains to verify Proposition 1 from the main paper under Assumptions B.1 to B.6:

Proof of Proposition 1. The GMM estimator $(\tilde{\boldsymbol{\theta}}', \tilde{\boldsymbol{\gamma}}')'$ satisfies the first-order conditions $\left(N^{-1} \sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N \left(N^{-1} \sum_{i=1}^N \mathbf{m}_i(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\gamma}})\right) = \mathbf{0}$. Let $\tilde{I}$ be an indicator function for the

[2] $\|\cdot\|$ denotes the Frobenius norm.

event that $\left(N^{-1}\sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N \left(N^{-1}\sum_{i=1}^N \mathbf{S}_i\right)$ is nonsingular. With $\mathbf{m}_i(\tilde{\boldsymbol{\theta}}, \tilde{\gamma}) = \mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \mathbf{S}_i[(\tilde{\boldsymbol{\theta}}', \tilde{\gamma}')' - (\boldsymbol{\theta}', \boldsymbol{\gamma}')']$ and influence function $\boldsymbol{\varphi}_i = (\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}\mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})$ as defined in equation (7), it follows from the first-order conditions that

$$\begin{aligned}\sqrt{N}\begin{pmatrix}\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \tilde{\gamma} - \boldsymbol{\gamma}\end{pmatrix} &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\varphi}_i + (1 - \tilde{I})\sqrt{N}\begin{pmatrix}\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \tilde{\gamma} - \boldsymbol{\gamma}\end{pmatrix} \\ &+ \left[\tilde{I}\left(\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N \left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_i\right)\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N - (\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}\right] \\ &\qquad \times \left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})\right).\end{aligned}$$

By Lemma B.1, $\mathbf{S}'\mathbf{V}\mathbf{S}$ is nonsingular and thus $\tilde{I} \xrightarrow{p} 1$. By Lemma B.2 and by Slutsky's theorem, $N^{-\frac{1}{2}}\sum_{i=1}^N \boldsymbol{\varphi}_i \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, (\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}\Xi\mathbf{V}\mathbf{S}(\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\right)$. By Assumptions B.4 and B.5 and by Slutsky's theorem, $\tilde{I}\left(\left(N^{-1}\sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N \left(N^{-1}\sum_{i=1}^N \mathbf{S}_i\right)\right)^{-1}\left(N^{-1}\sum_{i=1}^N \mathbf{S}_i\right)' \mathbf{V}_N$ converges in probability to $(\mathbf{S}'\mathbf{V}\mathbf{S})^{-1}\mathbf{S}'\mathbf{V}$. Since $N^{-\frac{1}{2}}\sum_{i=1}^N \mathbf{m}_i(\boldsymbol{\theta}, \boldsymbol{\gamma})$ is bounded in probability by Lemma B.2, all right-hand side terms besides $N^{-\frac{1}{2}}\sum_{i=1}^N \boldsymbol{\varphi}_i$ converge to zero in probability. Hence, the GMM estimator is consistent. □

B.3 Feasible efficient GMM estimation

Blundell and Bond (1998) and Windmeijer (2000) emphasize that for dynamic panel data models, in general, efficient GMM estimation is infeasible without having a prior estimate of Ξ. An exception is the situation with homoskedastic error components, $E[u_{it}^2|\mathbf{Z}_i] = \sigma_u^2$ and $E[\alpha_i^2|\mathbf{Z}_i] = \sigma_\alpha^2$, and prior knowledge of $\tau = \sigma_\alpha^2/\sigma_u^2$. An optimal weighting matrix then is

$$\mathbf{V}_N = N\left[\mathbf{Z}'\mathbf{H}(\mathbf{I}_N \otimes \Phi)\mathbf{H}'\mathbf{Z}\right]^{-1}, \tag{B.4}$$

with $\Phi = \tau\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \mathbf{I}_T$ such that $\mathbf{V} = \sigma_u^2\Xi^{-1}$. When the estimator only involves moment conditions for the first-differenced model such that $\mathbf{H}'\mathbf{Z} = \mathbf{D}'\mathbf{Z}_d$, the optimal weighting matrix (B.4) boils down to $\mathbf{V}_N = N(\mathbf{Z}_d'\mathbf{D}\mathbf{D}'\mathbf{Z}_d)^{-1}$ independent of τ since $\mathbf{D}_i\Phi\mathbf{D}_i' = \mathbf{D}_i\mathbf{D}_i'$, as discussed by Arellano and Bond (1991).

When τ is unknown or homoskedasticity is too strong an assumption, it is common practice to proceed in two steps to obtain a feasible efficient GMM estimator. In the first step, choosing any positive-definite matrix $\mathbf{V}_N$ will yield consistent but generally inefficient estimates $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\gamma}}$. It is common practice to use a first-step weighting matrix of the form

$$\mathbf{V}_N = N\left[\mathbf{Z}'(\mathbf{I}_N \otimes \Phi^*)\mathbf{Z}\right]^{-1}, \tag{B.5}$$

with different choices for Φ^*. Among others, Arellano and Bover (1995) and Blundell and Bond (1998) use $\Phi^* = \mathbf{I}_{2T-1}$, while Blundell et al. (2001) take the first-order serial correlation in the first-differenced residuals into account by choosing

$$\Phi^* = \begin{pmatrix} \mathbf{D}_i\mathbf{D}_i' & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T \end{pmatrix}.$$

When σ_α^2 is small, Windmeijer (2000) suggests to use $\Phi^* = \mathbf{H}_i\mathbf{H}_i'$. In the latter case, the first-step weighting matrix (B.5) equals the optimal weighting matrix (B.4) under $\tau = 0$. A reasonable alternative is the weighting matrix (B.4) with an adequate choice (or prior estimate) of τ.

The second-step weighting matrix is formed as $\mathbf{V}_N = \tilde{\Xi}^{-1}$. With first-step residuals $\tilde{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{W}_{yxi}\tilde{\boldsymbol{\theta}} - \mathbf{F}_i\tilde{\boldsymbol{\gamma}}$, a consistent unrestricted estimate of Ξ can be obtained as $\tilde{\Xi} = N^{-1}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{H}_i\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i'\mathbf{H}_i'\mathbf{Z}_i$. Alternatively, $\tilde{\Xi} = N^{-1}\sum_{i=1}^N \mathbf{Z}_i'\mathbf{H}_i\tilde{\Phi}\mathbf{H}_i'\mathbf{Z}_i$ with an unrestricted estimate $\tilde{\Phi} = N^{-1}\sum_{i=1}^N \tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_i'$ or a restricted estimate $\tilde{\Phi} = \tilde{\sigma}_\alpha^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \tilde{\sigma}_u^2\mathbf{I}_T$. Consistent variance estimates $\tilde{\sigma}_\alpha^2$ and $\tilde{\sigma}_u^2$ can be obtained as follows:

$$\begin{aligned}
\tilde{\sigma}_e^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^2, \\
\tilde{\sigma}_\alpha^2 &= \frac{1}{NT(T-1)/2}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\tilde{e}_{it}\tilde{e}_{is}, \\
\tilde{\sigma}_u^2 &= \tilde{\sigma}_e^2 - \tilde{\sigma}_\alpha^2.
\end{aligned}$$

B.4 Finite-sample improvements

The importance of choosing an appropriate first-step weighting matrix should not be underestimated in applied work. Although the second-step GMM estimator is asymptotically unaffected, its finite-sample performance still depends on the choice of $\mathbf{V}_N$ in the first step. Windmeijer (2005) shows that asymptotic standard error estimates of the two-step GMM estimator can be severely downward biased in finite samples. He derives a finite-sample variance correction. Alternatives to the two-step GMM estimator that are targeted to improve the finite-sample performance include the iterated and the continuously updated GMM estimators, see for example Hansen et al. (1996). Recently, Seo and Shin (2016) suggested to average asymptotically efficient two-step estimators that are based on different initial weighting matrices to reduce the finite-sample variation.

Moreover, GMM estimators might suffer from severe finite-sample distortions that arise from having too many instruments relative to the sample size, as stressed by Roodman (2009) among others. The instrument count can be reduced by forming linear combinations $\mathbf{Z}_i\mathbf{R}_i$ of the columns of $\mathbf{Z}_i$. For any deterministic transformation matrix $\mathbf{R}_i$, this also leads to a valid set of moment conditions, $E[\mathbf{R}_i'\mathbf{Z}_i'\mathbf{H}_i\mathbf{e}_i] = \mathbf{0}$. The GMM estimator (5) is then based on the transformed instruments $\mathbf{Z}_i\mathbf{R}_i$. In the following, we provide examples of the transformation matrix $\mathbf{R}_i$ that are relevant in practical applications.[3] We restrict our attention to block-diagonal versions of $\mathbf{R}_i$:

$$\mathbf{R}_i = \begin{pmatrix} \mathbf{R}_{di} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{li} \end{pmatrix},$$

such that $\mathbf{H}_i'\mathbf{Z}_i\mathbf{R}_i = (\mathbf{D}_i'\mathbf{Z}_{di}\mathbf{R}_{di}, \mathbf{Z}_{li}\mathbf{R}_{li})$. Similarly, we consider a block-diagonal partition of the transformation matrix for the first-differenced equation:

$$\mathbf{R}_{di} = \begin{pmatrix} \mathbf{R}_{dyi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{dxi} \otimes \mathbf{I}_{K_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{dfi} \otimes \mathbf{I}_{K_f} \end{pmatrix},$$

conformable for multiplication with the instruments matrix $\mathbf{Z}_{di}$ given in Appendix B.1.

Often, the instrument count is reduced by ignoring some of the available lags. This procedure is equivalent to the construction of a transformation matrix $\mathbf{R}_{di}$ that selects the appropriate columns of the full matrix $\mathbf{Z}_{di}$. As an example, the following matrices restrict the lag depth to $\kappa \geq 1$ for both the lagged dependent variable $y_{i,t-1}$ and strictly

[3] Mehrhoff (2009) provides similar transformation matrices for an AR(1) process.

exogenous regressors $\mathbf{x}_{it}$ while also discarding future values of the latter:

$$\mathbf{R}_{dyi} = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\kappa 2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_{\kappa 3} & & \vdots \\ \vdots & \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_{\kappa,T-1} \end{pmatrix}, \quad \mathbf{R}_{dxi} = \begin{pmatrix} \tilde{\mathbf{J}}_{\kappa 3} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{J}}_{\kappa 4} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & & \vdots \\ \vdots & \vdots & & \tilde{\mathbf{J}}_{\kappa T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_{\kappa,T+1} \end{pmatrix},$$

where $\mathbf{J}_{\kappa s} = \mathbf{I}_s$ if $s \leq \kappa$, and $\mathbf{J}_{\kappa s} = (\mathbf{0}, \mathbf{I}_\kappa)'$ with dimension $s \times \kappa$ if $s > \kappa$, and $\tilde{\mathbf{J}}_{\kappa s} = (\mathbf{J}'_{\kappa s}, \mathbf{0})'$ with dimension $(T+1) \times \min\{s, \kappa\}$. We could set $\mathbf{R}_{dfi} = \mathbf{I}_{T-1}$ in this case.

Alternatively, the dimension of the instruments matrix can be reduced by collapsing it into smaller blocks. The following transformation matrices linearly combine the columns of $\mathbf{Z}_{di}$, again for the case of strictly exogenous regressors $\mathbf{x}_{it}$:

$$\mathbf{R}_{dyi} = \begin{pmatrix} \mathbf{J}^*_{0,1,T-2} \\ \mathbf{J}^*_{0,2,T-3} \\ \vdots \\ \mathbf{J}^*_{0,T-2,1} \\ \mathbf{I}^*_{T-1} \end{pmatrix}, \quad \mathbf{R}_{dxi} = \begin{pmatrix} \mathbf{J}^*_{0,T+1,T-2} \\ \mathbf{J}^*_{1,T+1,T-3} \\ \vdots \\ \mathbf{J}^*_{T-3,T+1,1} \\ \mathbf{J}^*_{T-2,T+1,0} \end{pmatrix},$$

where $\mathbf{J}^*_{s_1,s_2,s_3} = (\mathbf{0}_{s_2 \times s_1}, \mathbf{I}^*_{s_2}, \mathbf{0}_{s_2 \times s_3})$ with dimension $s_2 \times (s_1 + s_2 + s_3)$, and $\mathbf{I}^*_{s_2}$ is the s_2-dimensional mirror identity matrix with ones on the antidiagonal and zeros elsewhere. $\mathbf{Z}_{dyi}\mathbf{R}_{dyi}$ now corresponds to the collapsed matrix described by Roodman (2009). As a consequence, the $T(T-1)/2$ moment conditions (A.1) are replaced by the $T-1$ conditions $E\left[\sum_{t=s}^{T} y_{i,t-s} \Delta u_{it}\right] = 0$, $s = 2, 3, \ldots, T$. Similarly, the information contained in the $K_x(T+1)(T-1)$ moment conditions (A.2) is condensed into $K_x(2T-1)$ conditions. The instruments block containing $\mathbf{f}_i$ can be collapsed by setting $\mathbf{R}_{dfi} = \boldsymbol{\iota}_{T-1}$. The implied K_f moment conditions are $E[\mathbf{f}_i(u_{iT} - u_{i1})] = \mathbf{0}$ instead of the $K_f(T-1)$ conditions (A.3). The transformation matrices can be further adjusted to combine the collapsing approach with the lag depth restriction.

The instruments for the level equation, for clarity ignoring the moment conditions $E[\mathbf{x}_{1i0} e_{i1}] = \mathbf{0}$, can be collapsed into a set of standard instruments by applying the following transformation:

$$\mathbf{R}_{li} = \begin{pmatrix} \boldsymbol{\iota}_{T-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota}_T \otimes \mathbf{I}_{K_{x1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\iota}_T \otimes \mathbf{I}_{K_{x2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{K_{f1}} \end{pmatrix},$$

such that $\mathbf{Z}_{li}\mathbf{R}_{li} = [(0, \Delta\mathbf{y}'_{i,(-1)})', \mathbf{X}_{1i}, \mathbf{D}_i\mathbf{X}_{2i}, \mathbf{F}_{1i}]$.

Appendix C Two-stage estimation

The additional information in this appendix complements Section 4 of the main paper.

C.1 Second-stage GMM estimation

The second-stage GMM estimator $\hat{\hat{\boldsymbol{\gamma}}}$ in equation (11) is obtained from the following minimization problem:

$$\hat{\hat{\boldsymbol{\gamma}}} = \arg\min_{\mathbf{b}_2} \left(\frac{1}{N}\mathbf{Z}'_\gamma(\mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}} - \mathbf{F}\mathbf{b}_2)\right)' \mathbf{V}_{\gamma N} \left(\frac{1}{N}\mathbf{Z}'_\gamma(\mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}} - \mathbf{F}\mathbf{b}_2)\right)',$$

where $\hat{\boldsymbol{\theta}}$ is a consistent first-stage estimator. With an appropriate estimate $\hat{\boldsymbol{\psi}}_{\theta i}$ of the first-stage influence function, an estimate of the second-stage influence function is readily obtained by replacing the probability limits in equation (7) with their sample analogues:

$$\hat{\hat{\boldsymbol{\psi}}}_{\gamma i} = N\left(\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{\gamma N}\mathbf{Z}'_\gamma\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{\gamma N}\left(\mathbf{Z}'_{\gamma i}\hat{\hat{\mathbf{e}}}_i - N^{-1}\mathbf{Z}'_\gamma\mathbf{W}_{yx}\hat{\boldsymbol{\psi}}_{\theta i}\right), \tag{C.1}$$

where $\hat{\hat{\mathbf{e}}}_i = \mathbf{y}_i - \mathbf{W}_{yxi}\hat{\boldsymbol{\theta}} - \mathbf{F}_i\hat{\hat{\boldsymbol{\gamma}}}$. An unrestricted estimate of the variance matrix Σ_γ of the asymptotic second-stage distribution can thus be obtained as $\hat{\hat{\Sigma}}_\gamma = N^{-1}\sum_{i=1}^N \hat{\hat{\boldsymbol{\psi}}}_{\gamma i}\hat{\hat{\boldsymbol{\psi}}}'_{\gamma i}$. Alternatively, an estimate could be obtained based on separate estimates of the components of Ξ_v in Proposition 2, $\hat{\hat{\Xi}}_v = \hat{\hat{\Xi}}_e + \hat{\hat{\mathbf{S}}}_\theta\hat{\Sigma}_\theta\hat{\hat{\mathbf{S}}}'_\theta - \hat{\hat{\Xi}}'_{\theta e}\hat{\hat{\mathbf{S}}}'_\theta - \hat{\hat{\mathbf{S}}}_\theta\hat{\hat{\Xi}}_{\theta e}$, with $\hat{\hat{\mathbf{S}}}_\theta = N^{-1}\mathbf{Z}'_\gamma\mathbf{W}_{yx}$ and $\hat{\hat{\Xi}}_{\theta e} = N^{-1}\sum_{i=1}^N \hat{\boldsymbol{\psi}}_{\theta i}\hat{\hat{\mathbf{e}}}'_i\mathbf{Z}_{\gamma i}$, and where $\hat{\Sigma}_\gamma$ is an estimate of the first-stage variance matrix. Finally, $\hat{\hat{\Xi}}_e$ can be obtained in an unrestricted or restricted way following along similar lines as in Appendix B.3.

C.2 Asymptotic distribution of the second-stage GMM estimator

In the following, we present a sufficient set of assumptions for the identification of the second-stage coefficients and regularity conditions for consistency and asymptotic normality of the second-stage GMM estimator. For notational convenience, denote the second-stage moment vector by $\mathbf{m}_{\gamma i}(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{Z}'_{\gamma i}(\mathbf{y}_i - \mathbf{W}_{yxi}\mathbf{b}_1 - \mathbf{F}_i\mathbf{b}_2)$ such that $\mathbf{m}_{\gamma i}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \mathbf{Z}'_{\gamma i}\mathbf{e}_i$. The corresponding negative gradients with respect to $\mathbf{b}_1$ and $\mathbf{b}_2$ are $\mathbf{S}_{\theta i} = \mathbf{Z}'_{\gamma i}\mathbf{W}_{yxi}$ and $\mathbf{S}_{\gamma i} = \mathbf{Z}'_{\gamma i}\mathbf{F}_i$, respectively. The GMM criterion function is minimized under the constraint $\mathbf{b}_1 = \hat{\boldsymbol{\theta}}$ for a consistent first-stage estimator $\hat{\boldsymbol{\theta}}$ that satisfies Assumption 4.

Assumption C.1: The first-stage parameter space is a convex set Θ_θ with $\boldsymbol{\theta} \in \Theta_\theta$, and the second-stage parameter space is the convex set $\Theta_\gamma = \mathbb{R}^{K_f}$ with $\boldsymbol{\gamma} \in \Theta_\gamma$.

Assumption C.2: A weak law of large numbers holds element-wise for the second-stage moment vector such that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{m}_{\gamma i}(\mathbf{b}_1, \mathbf{b}_2) = \mathbf{m}_\gamma(\mathbf{b}_1, \mathbf{b}_2)$ for all $\mathbf{b}_1 \in \Theta_\theta$ and $\mathbf{b}_2 \in \Theta_\gamma$.

Assumption C.3: The second-stage moment conditions $E[\mathbf{m}_{\gamma i}(\boldsymbol{\theta}, \boldsymbol{\gamma})] = \mathbf{0}$ are satisfied.

Assumption C.4: A weak law of large numbers holds element-wise for the negative gradients of the second-stage moment vector such that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{S}_{\theta i} = \mathbf{S}_\theta$ and $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \mathbf{S}_{\gamma i} = \mathbf{S}_\gamma$.

Assumption C.5: A weak law of large numbers holds element-wise for the second-stage weighting matrix $\mathbf{V}_N$ such that $\text{plim}_{N\to\infty} \mathbf{V}_{\gamma N} = \mathbf{V}_\gamma$ is positive semi-definite and $\mathbf{V}_\gamma\mathbf{S}_\gamma$ has full column rank.

Lemma C.1: Under Assumption 4 and Assumptions C.1 to C.5, all second-stage parameters are identified.

Proof. By Assumptions C.2 and C.5 and by Slutsky's theorem, the second-stage GMM criterion function converges in probability to $\mathbf{m}_\gamma(\mathbf{b}_1, \mathbf{b}_2)\mathbf{V}_\gamma\mathbf{m}_\gamma(\mathbf{b}_1, \mathbf{b}_2) \geq 0$ for all $\mathbf{b}_1 \in \Theta_\theta$ and $\mathbf{b}_2 \in \Theta_\gamma$. By Assumption 4, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$. Hence, by continuity of the moment vector and Assumption C.1, the constrained second-stage GMM criterion function convergences in probability to $\mathbf{m}_\gamma(\boldsymbol{\theta}, \mathbf{b}_2)\mathbf{V}_\gamma\mathbf{m}_\gamma(\boldsymbol{\theta}, \mathbf{b}_2) \geq 0$ for all $\mathbf{b}_2 \in \Theta_\gamma$. Denote the corresponding constrained minimizer by $\hat{\hat{\boldsymbol{\gamma}}}_\infty$. By Assumptions C.2 to C.4, the first-order condition is

$\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{m}_\gamma(\boldsymbol{\theta},\hat{\hat{\boldsymbol{\gamma}}}_\infty) = \mathbf{0}$ with $\mathbf{m}_\gamma(\boldsymbol{\theta},\hat{\hat{\boldsymbol{\gamma}}}_\infty) = \mathbf{S}_\gamma(\boldsymbol{\gamma} - \hat{\hat{\boldsymbol{\gamma}}}_\infty)$. By Assumption C.5, $\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma$ is nonsingular. Hence, with Assumption C.1, $\hat{\hat{\boldsymbol{\gamma}}}_\infty = \boldsymbol{\gamma}$ is the unique minimizer of the constrained asymptotic second-stage GMM criterion function on Θ_γ. □

Assumption C.6: The second-stage moment vector evaluated at the population values, $\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})$, is independently distributed across i with $\sup_{i,N} E[\|\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\|^{2+\delta}] < \infty$ for some $\delta > 0$, and where $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N E[\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})'] = \Xi_e$ is a positive-definite matrix.

Lemma C.2: Under Assumptions C.3 and C.6, $N^{-\frac{1}{2}}\sum_{i=1}^N \mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0,\Xi_e)$.

Proof. Let $\mathbf{c} \in \mathbb{R}^{K_{z\gamma}}$ with $\mathbf{c}'\mathbf{c} = 1$. By Assumption C.6 and the Cauchy-Schwarz inequality, $\sup_{i,N} E[|\mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})|^{2+\delta}] \leq \sup_{i,N} E[\|\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\|^{2+\delta}] < \infty$. With Ξ_e being positive definite, $0 < \lim_{N\to\infty} N^{-1}\sum_{i=1}^N E[\mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})'\mathbf{c}] = \mathbf{c}'\Xi_e\mathbf{c} < \infty$. Hence, for any $\delta > 0$,

$$
\begin{aligned}
0 &\leq \lim_{N\to\infty} \left(\sum_{i=1}^N E[\mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})'\mathbf{c}]\right)^{-(2+\delta)} \sum_{i=1}^N E[|\mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})|^{2+\delta}] \\
&\leq (\mathbf{c}'\Xi_e\mathbf{c})^{-(2+\delta)} \lim_{N\to\infty} N^{-\frac{\delta}{2}} \sup_{i,N} \sum_{i=1}^N E[|\mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})|^{2+\delta}] = 0.
\end{aligned}
$$

such that Lyapunov's condition (and thus Lindeberg's condition) is satisfied. With Assumption C.3, the Lindeberg-Feller central limit theorem implies $N^{-\frac{1}{2}}\sum_{i=1}^N \mathbf{c}'\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(0,\mathbf{c}'\Xi_e\mathbf{c})$. The claim follows from the Cramér-Wold theorem. □

Assumption C.7: The first-stage influence function $\boldsymbol{\psi}_{\theta i}$ is independently distributed across i with $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N E[\boldsymbol{\psi}_{\theta i}\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})'] = \Xi_{\theta e}$.

Under Assumption 4 and Assumptions C.1 to C.7, we can now verify Proposition 2 from the main paper.

Proof of Proposition 2. The constrained second-stage GMM estimator $\hat{\hat{\boldsymbol{\gamma}}}$ satisfies the first-order conditions $\left(N^{-1}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\left(N^{-1}\sum_{i=1}^N \mathbf{m}_{\gamma i}(\hat{\boldsymbol{\theta}},\hat{\hat{\boldsymbol{\gamma}}})\right) = \mathbf{0}$. Let $\hat{I}$ be an indicator function for the event that $\left(N^{-1}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\left(N^{-1}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)$ is nonsingular. With $\mathbf{m}_{\gamma i}(\hat{\boldsymbol{\theta}},\hat{\hat{\boldsymbol{\gamma}}}) = \mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma}) - \mathbf{S}_{\theta i}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \mathbf{S}_{\gamma i}(\hat{\hat{\boldsymbol{\gamma}}} - \boldsymbol{\gamma})$, first-stage influence function $\boldsymbol{\psi}_{\theta i}$ implicitly defined in Assumption 4, and second-stage influence function $\boldsymbol{\psi}_{\gamma i} =$

$(\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma(\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})-\mathbf{S}_\theta\boldsymbol{\psi}_{\theta i})$ as defined in equation (13), it follows from the first-order conditions that

$$\begin{aligned}\sqrt{N}\left(\hat{\hat{\boldsymbol{\gamma}}}-\boldsymbol{\gamma}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\psi}_{\gamma i} + (1-\hat{I})\sqrt{N}\left(\hat{\hat{\boldsymbol{\gamma}}}-\boldsymbol{\gamma}\right) \\ &+ \left[\hat{I}\left(\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\right. \\ &\qquad \left. - (\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma\right]\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})\right) \\ &- \left[\hat{I}\left(\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\gamma i}\right)'\mathbf{V}_{\gamma N}\left(\frac{1}{N}\sum_{i=1}^N \mathbf{S}_{\theta i}\right)\right. \\ &\qquad \left. - (\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\theta\right]\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) - (\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\theta o_p(1),\end{aligned}$$

where the last $o_p(1)$ term is the remainder term in Assumption 4. By Lemma C.1, $\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma$ is nonsingular and thus $\hat{I}\xrightarrow{p}1$. By Assumption 4, Lemma C.2, and Slutsky's theorem, $N^{-\frac{1}{2}}\sum_{i=1}^N(\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})-\mathbf{S}_\theta\boldsymbol{\psi}_{\theta i})\xrightarrow{d}\mathcal{N}(\mathbf{0},\Xi_v)$, where $\Xi_v=\Xi_e+\mathbf{S}_\theta\Sigma_\theta\mathbf{S}_\theta'-\Xi_{\theta e}'\mathbf{S}_\theta'-\mathbf{S}_\theta\Xi_{\theta e}$. Hence, $N^{-\frac{1}{2}}\sum_{i=1}^N\boldsymbol{\psi}_{\gamma i}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},(\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\mathbf{S}_\gamma'\mathbf{V}_\gamma\Xi_v\mathbf{V}_\gamma\mathbf{S}_\gamma(\mathbf{S}_\gamma'\mathbf{V}_\gamma\mathbf{S}_\gamma)^{-1}\right)$. By Assumptions C.4 and C.5 and by Slutsky's theorem, the expressions in square brackets converge in probability to zero. Since $N^{-\frac{1}{2}}\sum_{i=1}^N\mathbf{m}_{\gamma i}(\boldsymbol{\theta},\boldsymbol{\gamma})$ and $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ are bounded in probability by Lemma C.2 and Assumption 4, respectively, all right-hand side terms besides $N^{-\frac{1}{2}}\sum_{i=1}^N\boldsymbol{\psi}_{\gamma i}$ converge to zero in probability. Hence, the second-stage GMM estimator is consistent. □

C.3 First-stage GMM estimator

The above results hold for any first-stage estimator that satisfies Assumption 4. Consider a first-stage GMM estimator $\hat{\boldsymbol{\theta}}$ based on the moment conditions $E[\mathbf{Z}_{\theta i}'\mathbf{H}_i\boldsymbol{\epsilon}_i]=\mathbf{0}$ for the first-stage model (8), possibly making use of moment conditions for transformed and untransformed model representations. Moment conditions for the level model need to be adjusted accordingly compared to a one-stage procedure to account for the fact that the unobserved unit-specific component α_i is replaced by η_i, as defined in Section 4 of the main paper. It is then straightforward to obtain asymptotic results for the first-stage GMM estimator similar to those in Proposition 1 such that $\hat{\boldsymbol{\theta}}$ satisfies Assumption 4.

We can adapt the partitioned regression result in equation (B.1) to the first-stage estimator, partialing out the intercept term $\bar{\eta}$:[4]

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{W}_{yx}^{*}{}'\mathbf{Q}_{\iota}\mathbf{W}_{yx}^{*}\right)^{-1}\mathbf{W}_{yx}^{*}{}'\mathbf{Q}_{\iota}\mathbf{y}^{*}, \tag{C.2}$$

where $\mathbf{Q}_{\iota} = \mathbf{I}_{K_z} - \boldsymbol{\iota}^{*}(\boldsymbol{\iota}^{*\prime}\boldsymbol{\iota}^{*})^{-1}\boldsymbol{\iota}^{*\prime}$ with $\boldsymbol{\iota}^{*} = \mathbf{L}'\mathbf{Z}'\mathbf{H}\boldsymbol{\iota}_{NT}$. An estimate of the corresponding first-stage influence function $\boldsymbol{\psi}_{\theta i}$ is then obtained as

$$\hat{\boldsymbol{\psi}}_{\theta i} = N\left(\mathbf{W}_{yx}^{*}{}'\mathbf{Q}_{\iota}\mathbf{W}_{yx}^{*}\right)^{-1}\mathbf{W}_{yx}^{*}{}'\mathbf{Q}_{\iota}\mathbf{L}'(\mathbf{Z}_{\theta i}'\mathbf{H}_i\hat{\boldsymbol{\epsilon}}_i), \tag{C.3}$$

where $\hat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{W}_{yxi}\hat{\boldsymbol{\theta}} - \hat{\bar{\eta}}\boldsymbol{\iota}_T$. Following Windmeijer (2005), a finite-sample correction term might be added for feasible efficient two-step GMM estimators to account for the first-step estimation error.[5]

C.4 Comparison of the one-stage and two-stage GMM estimators

The partitioned regression result in equations (B.1) and (B.2) is helpful to contrast the one-stage estimator $(\tilde{\boldsymbol{\theta}}', \tilde{\gamma}')'$ and the two-stage estimator $(\hat{\boldsymbol{\theta}}', \hat{\hat{\gamma}}')'$ in the special case considered below. As a preliminary step, partition the one-stage weighting matrix as

$$\mathbf{V}_N = \begin{pmatrix} \mathbf{V}_{dN} & \mathbf{V}_{dlN} \\ \mathbf{V}_{dlN}' & \mathbf{V}_{lN} \end{pmatrix},$$

conformable for multiplications $\mathbf{Z}_d\mathbf{V}_{dN}\mathbf{Z}_d'$ and $\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}_l'$, where $\mathbf{Z}_d$ and $\mathbf{Z}_l$ are stacked instruments matrices for the first-differenced and the level model, respectively.[6]

For the two-stage approach, let us assume in this appendix that the first-stage GMM estimator utilizes moment conditions $E[\mathbf{Z}_{di}'\mathbf{D}_i\mathbf{e}_i] = \mathbf{0}$ for the first-differenced model only. Thus,

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{W}_{yx}'\mathbf{H}'\mathbf{Z}_{\theta}\mathbf{V}_{\theta N}\mathbf{Z}_{\theta}'\mathbf{H}\mathbf{W}_{yx}\right)^{-1}\mathbf{W}_{yx}'\mathbf{H}'\mathbf{Z}_{\theta}\mathbf{V}_{\theta N}\mathbf{Z}_{\theta}'\mathbf{H}\mathbf{y}, \tag{C.4}$$

where $\mathbf{Z}_{\theta} = \mathbf{Z}_d$ and $\mathbf{H} = \mathbf{I}_N \otimes \mathbf{D}_i$. The second-stage estimator $\hat{\hat{\boldsymbol{\gamma}}}$ is given by equation (11). We can now make the following claim.

[4] An intercept is not needed if all instruments are orthogonal to time-invariant variables.

[5] The corresponding expression is easily deduced from equation (2.3) in Windmeijer (2005).

[6] See Appendix B.

Proposition C.1: It holds that $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$, with $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ given by equations (B.1) and (C.4), respectively, if $\mathbf{Z}_l'\mathbf{F}$ is non-singular and $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'$.

Proof. Observe that $\mathbf{F}'\mathbf{H}'\mathbf{Z} = (\mathbf{F}'\mathbf{D}'\mathbf{Z}_d, \mathbf{F}'\mathbf{Z}_l) = (\mathbf{0}, \mathbf{F}'\mathbf{Z}_l)$ since $\mathbf{D}\mathbf{F} = \mathbf{0}$. Consequently, $\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{F}'\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}_l'\mathbf{F}$. With $\mathbf{Z}_l'\mathbf{F}$ being non-singular, $(\mathbf{F}^{*\prime}\mathbf{F}^*)^{-1} = (\mathbf{Z}_l'\mathbf{F})^{-1}\mathbf{V}_{lN}^{-1}(\mathbf{F}'\mathbf{Z}_l)^{-1}$. Let $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'$. Then,

$$\mathbf{L}\mathbf{Q}_F\mathbf{L}' = \mathbf{V}_N - \mathbf{V}_N \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{lN}^{-1} \end{pmatrix} \mathbf{V}_N = \begin{pmatrix} \mathbf{V}_{\theta N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

such that after straightforward algebra equation (B.1) boils down to equation (C.4). Alternatively, if $\mathbf{Z}_d'\mathbf{D}'\mathbf{W}_{yx}$ is non-singular as well, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} = (\mathbf{Z}_d'\mathbf{D}'\mathbf{W}_{yx})^{-1}\mathbf{Z}_d'\mathbf{D}'\mathbf{y}$ independent of the choice of the weighting matrices. □

When $\mathbf{Z}_l'\mathbf{F}$ is non-singular, the coefficients $\boldsymbol{\gamma}$ are exactly identified because the time-invariant regressors are orthogonal to the instruments for the first-differenced model. But then the instruments for the level model cannot be used any more to identify the coefficients $\boldsymbol{\theta}$, and $\tilde{\boldsymbol{\theta}}$ consequently equals $\hat{\boldsymbol{\theta}}$ with an appropriate choice of the weighting matrix. A similar proposition holds for the coefficients $\boldsymbol{\gamma}$ under the additional restriction that the level instruments of the one-stage system GMM estimator equal the instruments of the second-stage GMM estimator, $\mathbf{Z}_l = \mathbf{Z}_\gamma$.

Proposition C.2: With $\mathbf{Z}_l = \mathbf{Z}_\gamma$, it holds that $\tilde{\gamma} = \hat{\hat{\gamma}}$, with $\tilde{\gamma}$ and $\hat{\hat{\gamma}}$ given by equations (B.2) and (11), respectively, if $\mathbf{Z}_\gamma'\mathbf{F}$ is non-singular and $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'$ with $\mathbf{V}_{dlN} = \mathbf{0}$.

Proof. With $\mathbf{F}^{*\prime}\mathbf{F}^* = \mathbf{F}'\mathbf{Z}_l\mathbf{V}_{lN}\mathbf{Z}_l'\mathbf{F}$ and $\mathbf{Z}_l = \mathbf{Z}_\gamma$, equation (B.2) can be written as

$$\tilde{\gamma} = (\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{lN}\mathbf{Z}_\gamma'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Z}_\gamma\mathbf{V}_{lN}(\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'\mathbf{Z}_d'\mathbf{D} + \mathbf{Z}_\gamma')(\mathbf{y} - \mathbf{W}_{yx}\tilde{\boldsymbol{\theta}}).$$

With $\mathbf{Z}_\gamma'\mathbf{F}$ being non-singular, this equation reduces further to

$$\tilde{\gamma} = (\mathbf{Z}_\gamma'\mathbf{F})^{-1}(\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'\mathbf{Z}_d'\mathbf{D} + \mathbf{Z}_\gamma')(\mathbf{y} - \mathbf{W}_{yx}\tilde{\boldsymbol{\theta}}).$$

Equation (11) becomes $\hat{\hat{\gamma}} = (\mathbf{Z}_\gamma'\mathbf{F})^{-1}\mathbf{Z}_\gamma'(\mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}})$ independent of $\mathbf{V}_{\gamma N}$. Consequently,

$\tilde{\gamma} = \hat{\tilde{\gamma}}$ if $\mathbf{V}_{dlN} = \mathbf{0}$ and $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$. The latter follows from Proposition C.1 by setting $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}' = \mathbf{V}_{dN}$. Alternatively, if $\mathbf{Z}_d'\mathbf{D}'\mathbf{W}$ is non-singular as well, $\mathbf{Z}_d'\mathbf{D}(\mathbf{y} - \mathbf{W}_{yx}\hat{\boldsymbol{\theta}}) = \mathbf{0}$ and again $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$ without any restriction on the weighting matrices. □

Taken together, Propositions C.1 and C.2 state that one-stage and two-stage GMM estimation are equivalent for a particular choice of the weighting matrices if both utilize the same linearly independent instruments for the model in levels and their number equals the count of time-invariant regressors.

Leaving aside the trivial case of just identified coefficients $\boldsymbol{\theta}$, we can now infer a statement on asymptotic efficiency. When $\mathbf{V}_N$ is the optimal weighting matrix for the estimator $\tilde{\boldsymbol{\theta}}$ according to Lemma 1, then an optimal weighting matrix for the estimator $\hat{\boldsymbol{\theta}}$ is given by $\mathbf{V}_{\theta N} = \mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'$ as can be easily seen by calculating the partitioned inverse of $\mathbf{V}_N$. This corresponds to the condition that is required by Proposition C.1. However, for equivalence of the one-stage and the two-stage estimators, Proposition C.2 requires a block-diagonal weighting matrix $\mathbf{V}_N$ of the one-stage estimator such that $\mathbf{V}_{dlN} = \mathbf{0}$. A relevant case where this would be indeed optimal is a restricted covariance structure of the error term, $E[\mathbf{e}_i\mathbf{e}_i'|\mathbf{Z}_i] = \sigma_\alpha^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma_u^2 \mathbf{I}_T$, together with time invariance of the level instruments $\mathbf{Z}_{li}$. In this case, the feasible efficient one-stage and two-stage GMM estimators will be (asymptotically) identical, and therefore also have the same variance. In general, it is clear that the restricted estimator with a block-diagonal weighting matrix is less efficient than the feasible efficient one-stage GMM estimator unless the optimal one-stage weighting matrix is indeed block diagonal asymptotically.

Remark C.1: If the optimal weighting matrices $\mathbf{V}_N$ or $\mathbf{V}_{\theta N}$ are based on separate initial consistent estimates, the equivalence of $\mathbf{V}_{\theta N}$ and $\mathbf{V}_{dN} - \mathbf{V}_{dlN}\mathbf{V}_{lN}^{-1}\mathbf{V}_{dlN}'$ only holds asymptotically, and the resulting feasible efficient estimators can be numerically different in finite samples, even if all other conditions of Propositions C.1 and C.2 are satisfied.

If the moment conditions for the model in levels outnumber the time-invariant regressors, the one-stage and the two-stage GMM estimators will generally be different.

C.5 First-stage QML estimator

As an alternative to a first-stage GMM estimator, we might want to use likelihood-based estimation techniques. If $K_{x2} = K_{f2} = 0$, we can immediately estimate model (1) with the random-effects QML estimator of Bhargava and Sargan (1983), without the need of a second stage. When this strong assumption does not hold, Hsiao et al. (2002) propose to estimate the coefficients of the time-varying regressors based on the first-differenced model:

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta u_{it}, \tag{C.5}$$

for the time periods $t = 2, 3, \ldots, T$. However, this procedure not only eliminates the incidental parameters α_i but also the time-invariant variables $\mathbf{f}_i$. The latter can be recovered with the two-stage approach described in Section 4 of the main paper.

Hsiao et al. (2002) derive the joint density of $\Delta \mathbf{y}_{i\cdot} = (\Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{iT})'$ conditional on the strictly exogenous variables $\Delta \mathbf{X}_{i\cdot} = (\Delta \mathbf{x}_{i1}, \Delta \mathbf{x}_{i2}, \ldots, \Delta \mathbf{x}_{iT})'$. Because Δy_{i0} is unobserved, the marginal density of the initial observations Δy_{i1} conditional on $\Delta \mathbf{X}_{i\cdot}$ cannot be obtained immediately from model (C.5). Instead, Hsiao et al. (2002) apply linear projection techniques to derive the following expression for the initial observations based on an additional stationarity assumption for the regressors $\mathbf{x}_{it}$:

$$\Delta y_{i1} = b + \sum_{s=1}^{T} \Delta \mathbf{x}_{is}' \boldsymbol{\pi}_s + \xi_{i1}, \tag{C.6}$$

with $E[\xi_{i1} | \Delta \mathbf{X}_{i\cdot}] = 0$, $E[\xi_{i1}^2] = \sigma_\xi^2$, $E[\xi_{i1} \Delta u_{i2}] = -\sigma_u^2$, and $E[\xi_{i1} \Delta u_{it}] = 0$ for $t = 3, 4, \ldots, T$. The $1 + K_x T$ coefficients $\boldsymbol{\pi} = (b, \boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \ldots, \boldsymbol{\pi}_T')'$ are additional nuisance parameters that need to be estimated jointly with the parameters of interest. Under homoskedasticity, the variance matrix of $\Delta\mathbf{u}_{i\cdot} = (\xi_{i1}, \Delta u_{i2}, \ldots, \Delta u_{iT})'$ is given by[7]

$$E[\Delta\mathbf{u}_{i\cdot}\Delta\mathbf{u}_{i\cdot}'] = \sigma_u^2\Omega^* = \sigma_u^2 \begin{pmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix},$$

where $\omega = \sigma_\xi^2/\sigma_u^2$. The log-likelihood function can now be set up for the transformed model $\Delta\mathbf{y}_{i\cdot} = \Delta\mathbf{W}_{yxi\cdot}\boldsymbol{\theta} + \Delta\ddot{\mathbf{X}}_i\boldsymbol{\pi} + \Delta\mathbf{u}_{i\cdot}$, where

$$\Delta\mathbf{W}_{yxi\cdot} = \begin{pmatrix} 0 & \mathbf{0} \\ \Delta\mathbf{y}_{i,(-1)} & \Delta\mathbf{X}_i \end{pmatrix}, \quad \Delta\ddot{\mathbf{X}}_i = \begin{pmatrix} 1 & \Delta\mathbf{x}_{i1}' & \Delta\mathbf{x}_{i2}' & \ldots & \Delta\mathbf{x}_{iT}' \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix},$$

and treating the errors as normally distributed.

Decompose $\Omega^{*-1} = \mathbf{A}'\mathbf{B}^{-1}\mathbf{A}$, where $\mathbf{A}$ is a $T \times T$ lower-triangular and $\mathbf{B}$ a diagonal matrix.[8] Moreover, let $\mathbf{P} = \mathbf{I}_N \otimes (\mathbf{B}^{-1/2}\mathbf{A})$. The QML estimator of $\boldsymbol{\theta}$ has a closed-form solution given by:

$$\hat{\boldsymbol{\theta}} = (\Delta\mathbf{W}_{yx\cdot}'\hat{\mathbf{P}}'\hat{\mathbf{Q}}_x\hat{\mathbf{P}}\Delta\mathbf{W}_{yx\cdot})^{-1}\Delta\mathbf{W}_{yx\cdot}'\hat{\mathbf{P}}'\hat{\mathbf{Q}}_x\hat{\mathbf{P}}\Delta\mathbf{y}_{i\cdot}, \tag{C.7}$$

where $\hat{\mathbf{Q}}_x = \mathbf{I}_{NT} - \hat{\mathbf{P}}\Delta\ddot{\mathbf{X}}(\Delta\ddot{\mathbf{X}}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\Delta\ddot{\mathbf{X}})^{-1}\Delta\ddot{\mathbf{X}}'\hat{\mathbf{P}}'$, and $\hat{\mathbf{P}}$ is a function of the variance estimate $\hat{\omega}$. The variance matrix of $\hat{\boldsymbol{\theta}}$ is the corresponding partition of the inverse negative Hessian matrix.

In our Monte Carlo simulations in Section 6 and the empirical application in Section 7 of the main paper, we obtain the estimate $\hat{\omega}$ by maximizing the concentrated log-likelihood function in terms of ω only, given the analytical first-order conditions for the remaining parameters. The initial values for the QML optimization are obtained in the following steps. First, we obtain consistent GMM estimates of λ and β, and a variance estimate of σ_u^2 from the corresponding first-differenced residuals. The nuisance parameters $\boldsymbol{\pi}$ are obtained as ordinary least squares estimates from the initial observations equation (C.6). Second,

[7]Hayakawa and Pesaran (2015) extend this likelihood approach to accommodate for heteroskedastic errors.

[8]See Hsiao et al. (2002) for details.

given those estimates we evaluate the first-order condition for the variance parameter ω. Third, we update the estimates of the other parameters based on their respective optimality conditions given this estimate of ω. Finally, we repeat the second and third step one more time to obtain a faster convergence of the subsequent Newton-Raphson algorithm.[9]

The second-stage estimator $\hat{\hat{\gamma}}$ for the coefficients of the time-invariant regressors is given by equation (11), and the joint asymptotic distribution of the first-stage and second-stage estimators follows from Assumption 4, Proposition 2, and Corollary 1. Finally, the influence function of the whole parameter vector including the ancillary parameters is given by the inverse negative expected Hessian matrix multiplied by the score function for unit i.[10] The influence function for the parameter vector $\hat{\boldsymbol{\theta}}$ is the corresponding partition.

Appendix D Testing the overidentifying restrictions

In this supplement to Section 5 of the main paper, we briefly discuss a generalization of the familiar Hausman (1978) test that is asymptotically equivalent to the difference-in-Hansen or second-stage Hansen (1982) test.[11] The idea of this test is to contrast the coefficient estimates for the time-varying regressors from the one-stage and the two-stage procedures:[12]

$$\hat{H}_\gamma = \left(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right)' \left(\hat{\Xi}_H\right)^+ \left(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\right) \xrightarrow{d} \chi^2_{\min(K_{z\gamma}-K_f, 1+K_x)}, \tag{D.1}$$

where $\hat{\Xi}_H$ is a consistent estimate of the asymptotic variance matrix of $\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}$, and $(\cdot)^+$ denotes a generalized inverse. Under the null hypothesis of no misspecification, both estimators are consistent, while the one-stage GMM estimator $\tilde{\boldsymbol{\theta}}$ with an optimal weighting matrix is efficient. When Assumption 3 is violated, $\tilde{\boldsymbol{\theta}}$ in general turns inconsistent in contrast to $\hat{\boldsymbol{\theta}}$. The two estimators differ only in the additional $K_{z\gamma} - K_f$ overidentifying restrictions. In most practical situations, this number is smaller than the number of

[9] See Kripfganz (2016) for details about the implementation and the necessary adjustments in the case of unbalanced panel data.

[10] See Newey and McFadden (1994), Section 3.

[11] In simulation results not shown, we find that the generalized Hausman test proves largely impractical in finite samples because it is substantially oversized. Its rejection rate is close to 30% in our baseline specification that we describe in Section 6 of the main paper.

[12] The coefficients of time dummies and similar variables that are asymptotically unaffected by the tested moment restrictions should be excluded from the comparison.

contrasted coefficients, $1 + K_x$, such that the asymptotic variance matrix Ξ_H cannot have full rank. The degrees of freedom of the generalized Hausman test then equal those of the corresponding difference-in-Hansen or second-stage Hansen test.[13]

With the influence function $\boldsymbol{\psi}_{\theta i}$ of the first-stage estimator and the respective partition $\boldsymbol{\varphi}_{\theta i}$ of the one-stage influence function, we can express the asymptotic covariance matrix of the test statistic as $\Xi_H = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N E[(\boldsymbol{\psi}_{\theta i} - \boldsymbol{\varphi}_{\theta i})(\boldsymbol{\psi}_{\theta i} - \boldsymbol{\varphi}_{\theta i})']$. From Proposition 1, the asymptotic variance matrix of $\tilde{\boldsymbol{\theta}}$ is $\Omega_\theta = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N E[\boldsymbol{\varphi}_{\theta i}\boldsymbol{\varphi}_{\theta i}']$, and from Assumption 4, the asymptotic variance matrix of $\hat{\boldsymbol{\theta}}$ is $\Sigma_\theta = \lim_{N\to\infty} \sum_{i=1}^N E[\boldsymbol{\psi}_{\theta i}\boldsymbol{\psi}_{\theta i}']$. Following the original insights by Hausman (1978), $\Xi_H = \Sigma_\theta - \Omega_\theta$, given that $\tilde{\boldsymbol{\theta}}$ is an efficient estimator under the null hypothesis. However, the estimate $\hat{\Xi}_H = \hat{\Sigma}_\theta - \tilde{\Omega}_\theta$ is not guaranteed to be positive semidefinite in finite samples which might result in a negative, and therefore unusable, estimate $\hat{H}_\gamma$. A robust estimate that does not rely on the efficiency of the one-stage estimator can be constructed similar to White (1982) by using consistent estimates of the variance matrices and influence functions:

$$\hat{\Xi}_H = \hat{\Sigma}_\theta + \tilde{\Omega}_\theta - \frac{1}{N}\sum_{i=1}^N \hat{\boldsymbol{\psi}}_{\theta i}\tilde{\boldsymbol{\varphi}}_{\theta i}' - \frac{1}{N}\sum_{i=1}^N \tilde{\boldsymbol{\varphi}}_{\theta i}\hat{\boldsymbol{\psi}}_{\theta i}'. \tag{D.2}$$

Appendix E Monte Carlo simulation

This appendix contains additional information about the data-generating process and further results from our Monte Carlo simulation in Section 6 of the main paper.

E.1 Data-generating process

The data-generating process is presented in Section 6.1 and the estimators are described in Table 1 of the main paper. In the following, we provide some additional clarification about the derivation of the parameter restrictions. To control the correlation of the time-varying regressors with the time-invariant variables, we want to express the coefficients in

[13] The Hausman test is essentially a test of linear combinations of moment restrictions. See Ruud (2000, Chapter 22) and Baum et al. (2003) for a general discussion of GMM specification tests.

equation (19) as a function of the correlation coefficients in matrix

$$\mathbf{\Psi} = \begin{pmatrix} 1 & \rho_{f1,f2} & \rho_{f1,z} & \rho_{f1,\alpha} \\ \rho_{f1,f2} & 1 & \rho_{f2,z} & \rho_{f2,\alpha} \\ \rho_{f1,z} & \rho_{f2,z} & 1 & \rho_{z,\alpha} \\ \rho_{f1,\alpha} & \rho_{f2,\alpha} & \rho_{z,\alpha} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \rho_{f2,z} & \rho_{f2,\alpha} \\ 0 & \rho_{f2,z} & 1 & 0 \\ 0 & \rho_{f2,\alpha} & 0 & 1 \end{pmatrix},$$

as well as $\rho_{x1,f2}$ and $\rho_{x2,\alpha}$.

Given that x_{1it} and x_{2it} are covariance stationary, their data-generating process can be written as

$$x_{kit} = \frac{\pi_{k1}}{1-\phi_k} f_{1i} + \frac{\pi_{k2}}{1-\phi_k} f_{2i} + \frac{\kappa_k}{1-\phi_k}\alpha_i + \sum_{s=0}^{\infty} \phi_k^s \epsilon_{ki,t-s}, \quad k = 1,2. \tag{E.1}$$

By imposing that f_{1i} is uncorrelated with f_{2i} and α_i, respectively, we can derive

$$Cov(x_{kit}, f_{1i}) = \frac{\pi_{k1}\sigma_{f1}^2}{1-\phi_k}, \tag{E.2}$$

$$Cov(x_{kit}, f_{2i}) = \frac{\pi_{k2}\sigma_{f2}^2 + \kappa_k \rho_{f2,\alpha}\sigma_{f2}\sigma_\alpha}{1-\phi_k}, \tag{E.3}$$

$$Cov(x_{kit}, \alpha_i) = \frac{\pi_{k2}\rho_{f2,\alpha}\sigma_{f2}\sigma_\alpha + \kappa_k\sigma_\alpha^2}{1-\phi_k}, \tag{E.4}$$

$k = 1, 2$. Hence, f_{1i} is uncorrelated with both x_{1it} and x_{2it} if $\pi_{11} = \pi_{21} = 0$. To ensure the exogeneity of x_{1it} with respect to α_i, equation (E.4) requires $\kappa_1 = -\pi_{12}\rho_{f2,\alpha}\sigma_{f2}/\sigma_\alpha$. Then, using the fact that ϵ_{1it} is distributed independently and identically over time with $Var(\epsilon_{1it}) = (1-\phi^2)\sigma_{\epsilon 1}^2$,

$$Var(x_{1it}) = \frac{\pi_{12}^2\sigma_{f2}^2 + \kappa_1^2\sigma_\alpha^2 + 2\pi_{12}\kappa_1\rho_{f2,\alpha}\sigma_{f2}\sigma_\alpha}{(1-\phi_1)^2} + \sigma_{\epsilon 1}^2 = \frac{\pi_{12}^2(1-\rho_{f2,\alpha}^2)\sigma_{f2}^2}{(1-\phi_1)^2} + \sigma_{\epsilon 1}^2, \tag{E.5}$$

which is no longer a function of σ_α^2. From the definition of $\rho_{x1,f2}$, we can then solve for the parameter π_{12}:

$$\pi_{12} = \frac{(1-\phi_1)\rho_{x1,f2}\sigma_{\epsilon 1}}{\sqrt{(1-\rho_{f2,\alpha}^2)(1-\rho_{f2,\alpha}^2-\rho_{x1,f2}^2)}\sigma_{f2}}.$$

Similarly, to achieve a zero correlation between x_{2it} and f_{2i}, equation (E.3) tells us to set

$\pi_{22} = -\kappa_2 \rho_{f2,\alpha} \sigma_\alpha / \sigma_{f2}$. With

$$Var(x_{2it}) = \frac{\pi_{22}^2 \sigma_{f2}^2 + \kappa_2^2 \sigma_\alpha^2 + 2\pi_{22}\kappa_2 \rho_{f2,\alpha}\sigma_{f2}\sigma_\alpha}{(1-\phi_2)^2} + \sigma_{\epsilon 2}^2 = \frac{\kappa_2^2 (1-\rho_{f2,\alpha}^2)\sigma_\alpha^2}{(1-\phi_2)^2} + \sigma_{\epsilon 2}^2, \quad \text{(E.6)}$$

we can solve for the remaining parameter κ_2 from the definition of $\rho_{x2,\alpha}$:

$$\kappa_2 = \frac{(1-\phi_2)\rho_{x2,\alpha}\sigma_{\epsilon 2}}{\sqrt{(1-\rho_{f2,\alpha}^2)(1-\rho_{f2,\alpha}^2 - \rho_{x2,\alpha}^2)}\sigma_\alpha}.$$

In the following, we provide details on the derivation of the population value of the coefficient of determination for the first-differenced regression model:

$$R_{\Delta y}^2 = 1 - \frac{Var(\Delta y_{it} | \Delta x_{1it}, \Delta x_{1i,t-1}, \ldots, \Delta x_{2it}, \Delta x_{2i,t-1}, \ldots)}{Var(\Delta y_{it})}.$$

Given stationarity, the first-differenced data-generating processes are

$$\begin{aligned} \Delta y_{it} &= \lambda \Delta y_{i,t-1} + \beta_1 \Delta x_{1it} + \beta_2 \Delta x_{2it} + \Delta u_{it} \\ &= \beta_1 \sum_{s=0}^{\infty} \lambda^s \Delta x_{1i,t-s} + \beta_2 \sum_{s=0}^{\infty} \lambda^s \Delta x_{2i,t-s} + \sum_{s=0}^{\infty} \lambda^s \Delta u_{i,t-s}, \end{aligned} \quad \text{(E.7)}$$

and

$$\Delta x_{kit} = \phi_k \Delta x_{ki,t-1} + \Delta \epsilon_{kit} = \sum_{s=0}^{\infty} \phi_k^s \Delta \epsilon_{ki,t-s}, \quad k = 1, 2. \quad \text{(E.8)}$$

Since u_{it} is independent and identically distributed over time with $Var(u_{it}) = (1 - \lambda^2)\sigma_u^2$, such that $Var(\Delta u_{it}) = 2(1-\lambda^2)\sigma_u^2$, $Cov(\Delta u_{it}, \Delta u_{i,t-1}) = -(1-\lambda^2)\sigma_u^2$, and $Cov(\Delta u_{it}, \Delta u_{i,t-s}) = 0$ for all $s > 1$, it directly follows from equation (E.7) that

$$Var(\Delta y_{it} | \Delta x_{1it}, \Delta x_{1i,t-1}, \ldots, \Delta x_{2it}, \Delta x_{2i,t-1}, \ldots) = 2(1-\lambda)\sigma_u^2. \quad \text{(E.9)}$$

Because x_{1it} and x_{2it} are uncorrelated with each other and with u_{it}, the unconditional variance of Δy_{it} can be decomposed as

$$\begin{aligned} Var(\Delta y_{it}) = \beta_1^2 Var\left(\sum_{s=0}^{\infty} \lambda^s \Delta x_{1i,t-s}\right) + \beta_2^2 Var\left(\sum_{s=0}^{\infty} \lambda^s \Delta x_{2i,t-s}\right) \\ + Var(\Delta y_{it} | \Delta x_{1it}, \Delta x_{1i,t-1}, \ldots, \Delta x_{2it}, \Delta x_{2i,t-1}, \ldots), \end{aligned}$$

with

$$Var\left(\sum_{s=0}^{\infty}\lambda^s\Delta x_{ki,t-s}\right)=\left(\sum_{s=0}^{\infty}\lambda^{2s}Var(\Delta x_{kit})+2\sum_{s=0}^{\infty}\sum_{l=s+1}^{\infty}\lambda^{s+l}Cov(\Delta x_{ki,t-s},\Delta x_{ki,t-l})\right),$$

$k=1,2$. We thus need the variance

$$Var(\Delta x_{kit})=2(1-\phi_k)\sigma_{\epsilon k}^2,\quad k=1,2, \tag{E.10}$$

obtained analogously to equation (E.9), and the autocovariances

$$Cov(\Delta x_{ki,t-s},\Delta x_{ki,t-l})=-\phi_k^{l-s-1}(1-\phi_k)^2\sigma_{\epsilon k}^2,\quad k=1,2, \tag{E.11}$$

such that

$$Var(\Delta y_{it})=2(1-\lambda)\left(\frac{\beta_1^2(1-\phi_1)}{(1-\lambda^2)(1-\lambda\phi_1)}\sigma_{\epsilon 1}^2+\frac{\beta_2^2(1-\phi_2)}{(1-\lambda^2)(1-\lambda\phi_2)}\sigma_{\epsilon 2}^2+\sigma_u^2\right). \tag{E.12}$$

With the conditional variance (E.9) and the unconditional variance (E.12) of Δy_{it}, the population value of the coefficient of determination results as

$$\begin{aligned}R_{\Delta y}^2&=\frac{\beta_1^2(1-\phi_1)(1-\lambda\phi_2)\sigma_{\epsilon 1}^2+\beta_2^2(1-\phi_2)(1-\lambda\phi_1)\sigma_{\epsilon 2}^2}{\beta_1^2(1-\phi_1)(1-\lambda\phi_2)\sigma_{\epsilon 1}^2+\beta_2^2(1-\phi_2)(1-\lambda\phi_1)\sigma_{\epsilon 2}^2+(1-\lambda^2)(1-\lambda\phi_1)(1-\lambda\phi_2)\sigma_u^2}\\&=\frac{(1-\phi_1)\sigma_{\epsilon 1}^2+(1-\phi_2)\sigma_{\epsilon 2}^2}{(1-\phi_1)\sigma_{\epsilon 1}^2+(1-\phi_2)\sigma_{\epsilon 2}^2+(1-\lambda^2)\sigma_u^2},\end{aligned} \tag{E.13}$$

where the last equality follows by setting $\beta_k=\sqrt{1-\lambda\phi_k}$ for each $k=1,2$.

Finally, the equations used to initialize the data-generating processes are:

$$y_{i,-50}=\frac{\beta_1}{1-\lambda}x_{1i,-50}+\frac{\beta_2}{1-\lambda}x_{2i,-50}+\frac{\gamma_1}{1-\lambda}f_{1i}+\frac{\gamma_2}{1-\lambda}f_{2i}+\frac{\kappa_y}{1-\lambda}\alpha_i+\nu_i, \tag{E.14}$$

$$x_{ki,-50}=\frac{\pi_{k1}}{1-\phi_k}f_{1i}+\frac{\pi_{k2}}{1-\phi_k}f_{2i}+\frac{\kappa_k}{1-\phi_k}\alpha_i+\xi_{ki},\quad k=1,2, \tag{E.15}$$

where $(\nu_i,\xi_{1i},\xi_{2i})\overset{i.i.d.}{\sim}\mathcal{N}\left(\mathbf{0},\operatorname{diag}(\sigma_u^2,\sigma_{\epsilon 1}^2,\sigma_{\epsilon 2}^2)\right)$.

E.2 Simulation results for misspecified estimators

All of the estimators considered in Section 6 of the main paper are consistent given the data-generating process in our baseline scenario. Here, we highlight the adverse consequences of incorrect assumptions. In Table 6, we leave the data-generating process unchanged but classify the endogenous time-varying regressor x_{2it} as uncorrelated with the unobserved unit-specific effects α_i. Instead, x_{1it} is treated as correlated with α_i. Clearly, the first-stage coefficients λ, β_1, and β_2 of the two-stage estimators are completely unaffected because the modified regressor classification comes into effect only in the second stage. The one-stage estimators instead suffer from an endogeneity bias due to x_{2it} being used as an invalid instrument. While the bias and RMSE for the short-run coefficients still appear to be relatively small, they become substantial when we look at the long-run coefficients.

It might seem odd that even the estimates for the second-stage coefficients γ_1 and γ_2 of the two-stage estimators are hardly affected. This is due to the fact that x_{2it} is uncorrelated with both time-invariant regressors and the exogenous time-invariant instrument z_i is sufficient to identify the coefficient of the endogenous time-invariant regressor f_{2i}. Nevertheless, the second-stage Hansen test has a high power to detect the misspecification, as shown in Table 3 of the main paper. In particular, it is more powerful than the difference-in-Hansen test.

In Table 7, we aggravate the misspecification problem by generating z_i as positively correlated with α_i. It thus becomes an invalid instrumental variable and the second-stage estimates for γ_2 are now severely biased as well. Interestingly, however, the corresponding one-stage estimates become much less biased. The effects from the misspecification of x_{2it} and z_i apparently almost offset each other in this example.

In Table 8, we consider further types of estimator misspecifications given our baseline data-generating process. First, we look at one-stage and two-stage GMM estimators that replace Assumption 3 by a Mundlak (1978) assumption as in Remarks 3 and 5 (M-sGMM1 and M-sGMM2). The Mundlak (1978) projection has no effect on the first stage of M-sGMM1 which is identical to the first stage of the sGMM2 estimator in Table 2. For the one-stage estimator M-sGMM1, the results are slightly worse in comparison to sGMM1 but the incorrect assumption about the correlation between the observed and unobserved time-invariant variables apparently does not bite much. This picture changes for the coefficient γ_2 of the endogenous time-invariant regressor. Both M-sGMM1 and M-sGMM2 are significantly biased because the Mundlak (1978) assumption is not in line with the data-generating process. For the short-run coefficient, the bias of the two-stage estimator is slightly less pronounced, while the picture is reversed for the long-run coefficient.

The Mundlak (1978) assumption can also be used to construct the one-stage Bhargava and Sargan (1983) QML estimator (M-QML1). This estimator would be more efficient than the two-stage Hsiao et al. (2002) QML estimator if this assumption was correct. Here, their performance hardly differs with the exception of the coefficient γ_2 that is only correctly identified by the two-stage estimator, as evident in Table 8.

To demonstrate the implications of neglected dynamics, we furthermore consider a static Hausman and Taylor (1981) GMM estimator (HT-GMM1) and its two-stage variant (FE-IV2) in Table 8. Both use the exogenous variables x_{1it}, f_{1i}, and z_i as standard instruments, HT-GMM1 in the first stage and FE-IV2 in the second stage. The coefficients of the time-varying regressors x_{1it} and x_{2it} are instrumented with the deviations from their own within-group means, $x_{1it} - \bar{x}_{1i}$ and $x_{2it} - \bar{x}_{2i}$, respectively.[14] Because of the omitted lagged dependent variable, these estimators cannot distinguish between short-run and long-run coefficients. On average, the regression coefficient β_1 is estimated to be 0.82 by HT-GMM1 and 0.74 by FE-IV2. For β_2, the average estimates are 0.75 and 0.74, respectively. Given a true short-run coefficient of 0.6 and a true long-run coefficient of 3, it is apparent that the static-model estimators yield biased short-run effects and cannot be interpreted as long-run coefficients. For the time-invariant regressors, the opposite is the case. The average estimates for γ_1 are 3.0 with both estimators. They are thus very precise estimates of the long-run coefficient. For the coefficient γ_2 of the endogenous time-invariant regressor, the respective average estimates are 3.97 and 3.79, which are still approximations of the long-run coefficient but with substantial upward bias. The results are in line with the findings of Egger and Pfaffermayr (2004b).

Despite the use of the nonlinear moment conditions that are implied by the absence of serial correlation in the idiosyncratic error term, the Ahn and Schmidt (1995) estimators nlGMM1 and nlGMM2 remained inferior to the sGMM1 and sGMM2 estimators, respectively, in our baseline data-generating process. The additional efficiency gains of the latter rest on Assumption A.1 that is implied by mean stationarity. In Table 9, we

[14] The FE-IV2 corresponds to the Pesaran and Zhou (2018) estimator with the just-identified fixed-effects estimator in the first stage.

consider a violation of this assumption. Instead of drawing the initial observations from the joint stationary distribution 50 time points before the estimation sample starts, we now set $y_{i0} = x_{1i0} = x_{2i0} = 0$ without burn-in period. This creates a correlation of the initial-period change, and thus of all subsequent periods' change, with the unobserved unit-specific effects such that the moment conditions (A.7) and (A.8) become invalid. As expected, the nonlinear GMM estimators and the two-stage QML estimator are robust to this modification of the data-generating process, while the system GMM estimators become substantially biased.

E.3 Simulation results under different variance and covariance set-ups

Binder et al. (2005) and Bun and Windmeijer (2010) emphasize that a high variance ratio of the unit-specific effects to the idiosyncratic error term can result in a weak-instruments problem for the GMM estimators. In Tables 10 and 11, we modify our baseline scenario by reducing the variance ratio first to $\omega = 1$ and then further to $\omega = 1/3$. For the coefficients of the time-varying regressors, this alteration reduces both the bias and the RMSE of the GMM estimators, as expected. The nlGMM2 estimator benefits most because it is now less prone to produce outlying estimates of λ close to unity. The first-stage QML estimator is unaffected because it performs the estimation on the first-differenced model. For the coefficients of the time-invariant regressors, the picture is different. They tend to be less precisely estimated, even though the RMSE of the short-run coefficient γ_1 initially shrinks slightly. Given that the first-stage estimates of the QML2 estimator remain unchanged, its worsened second-stage performance provides the most genuine evidence that a lower variance ratio is unfavorable for the time-invariant regressors.

Changing the variance ratio ω does not alter the relative variances among the time-invariant model components. In Table 12, we double the variance of α_i to $\sigma_\alpha^2 = 2$, while keeping the variance ratio fixed at its initial value $\omega = 3$. There is not much of an effect on the coefficients of the time-varying regressors. In particular, the first-stage QML estimator is again unaffected by construction. However, the precision of the coefficients of the observed regressors f_{1i} and f_{2i} deteriorates because they now explain less of the time-invariant variation across units.

As another experiment, we increase the signal-to-noise ratio to $\tau = 2$ such that the variance parameters $\sigma^2_{\epsilon 1}$ and $\sigma^2_{\epsilon 2}$ quadruple, everything else as in the baseline scenario. The results improve unequivocally for all coefficients and estimators, as can be seen in Table 13.

In the baseline scenario, the endogenous time-invariant regressor f_{2i} was twice as strongly correlated with the time-invariant external instrument z_i than with the time-varying internal instrument x_{1it}. In Table 14, we swap these correlations to analyze their importance for the identification of the coefficient γ_2. It turns out that both the bias and the RMSE increase noticeably for this coefficient when the internal instrument is more relevant than the external. It thus appears to be preferable to use strong external instruments, if available, compared to an identification strategy that primarily builds on internal instruments. This observation is confirmed by a scenario in which the correlation between z_i and f_{2i} is lowered such that $\rho_{x1,f2} = \rho_{z,f2} = 0.2$. The results in Table 15 demonstrate that the RMSE of γ_2 substantially increases in this situation compared to the baseline scenario.

E.4 Simulation results under different degrees of persistence

The persistence of the time-varying variables plays an important role in determining the estimator properties in dynamic panel data models. In the baseline scenario, both the dependent variable y_{it} and the independent variables x_{1it} and x_{2it} are already quite persistent. In Table 16, we get even closer to the nonstationary case by increasing the autoregressive parameters to $\lambda = \phi_1 = \phi_2 = 0.9$. Perhaps surprisingly at first glance, the estimators for the short-run coefficients of the time-varying regressors tend to have a lower RMSE. This result might appear counterintuitive in the light of the potential weak-instruments problem that can arise as $\lambda \to 1$.[15] It can be explained by the observation that unity effectively serves as an upper barrier for the distribution of $\hat{\lambda}$ which reduces the variance of the estimators. In particular for the system GMM estimators, the resulting distribution is left-skewed. The nlGMM2 estimator is the only estimator with estimates exceeding this threshold, resulting in severe distortions when estimating the long-run coefficients. Also for the other estimators, it becomes more challenging to obtain precise estimates of the long-run coefficients. The coefficients of the time-invariant regressors are less precisely

[15] See for example Blundell and Bond (1998) and Bun and Sarafidis (2015).

estimated, not least because they now explain a smaller share of the total variance of the dependent variable.

Tables 17 to 24 similarly show the simulation results for less persistent data-generating processes. In comparison to the baseline scenario, already for $\lambda = \phi_1 = \phi_2 = 0.7$ the problem of too many estimates of λ close to unity disappears which is reflected in a reasonably small RMSE for the long-run coefficients with all estimators. For the short-run coefficients of the time-varying regressors, the RMSE initially increases, until around a true value of $\lambda = 0.5$. For the sGMM1 and the QML estimators, the turnaround point occurs at a slightly lower persistence. The argument made above for the high-persistence case is just reversed. The upper barrier for $\hat{\lambda}$ is now farther away such that the distribution has more room to spread out. For very low persistence levels, the RMSE starts to improve again for the autoregressive coefficient λ but not for the coefficients β_1 and β_2. On the other side, less persistence in the time-varying regressors is always preferable to obtain more precise estimates of the coefficients γ_1 and γ_2 of the time-invariant regressors, both regarding the short-run and long-run estimates.

An interesting observation can be made in the case with no history dependence at all. The true values of the long-run coefficients equal those of the short-run coefficients and the estimation of a static model would be appropriate. However, despite the estimation uncertainty that is additionally introduced by unnecessarily estimating the coefficient λ in a dynamic model, the long-run estimates $\hat{\gamma}_1/(1-\hat{\lambda})$ are actually more precise than the short-run estimates $\hat{\gamma}_1$ for the coefficient γ_1, and similarly for γ_2, as indicated by the RMSE in Table 24.

E.5 Simulation results with different sample sizes

A reduction of the sample size has the expected effect that all of the statistics worsen. In Table 25, we reduced the cross-sectional sample size to $N = 100$ units. The most dramatic deterioration of the estimators' performance is observed for the long-run coefficients that are estimated with substantial imprecision. In particular the two-stage estimators suffer from a larger number of estimates $\hat{\lambda}$ that are close to or beyond unity. A large sample size is definitely critical if the interest is on precise long-run coefficients.

Table 26 highlights for $N = 650$ that all estimators show the expected convergence behavior when the cross-sectional sample size grows. However, it is worth emphasizing that the underestimation of the second-stage standard errors without the correction formula in Proposition 2, and therefore also the size distortion of the Wald tests, remains at about the same magnitude. The first-stage estimation error affects the asymptotic distribution and is relevant for any sample size.

Tables 27 and 28 display the simulation results for alternative time horizons, $T = 3$ and $T = 10$, respectively. Given that we have taken effective measures to limit the proliferation of instruments, increasing the number of time periods helps to improve the accuracy of the GMM estimators. For the QML estimator of Hsiao et al. (2002), it needs to be highlighted that it has a bimodal distribution when $T = 3$.[16] As a consequence, its RMSE is substantially larger. The corresponding second-stage estimator clearly suffers as well from the poor first-stage performance. When T increases, this unfavorable effect disappears and the two-stage QML estimator can unfold its full potential. Unlike an increase in the cross-sectional dimension N, adding more time periods appears to reduce the distortions from the ignored standard error correction.

Appendix F Empirical application

F.1 Background and data description

In this appendix, we provide additional details about the data used by Egger and Pfaffermayr (2004a) and in Section 7 of our main paper, augmented with the dummy variables from the GeoDist data base (Mayer and Zignago, 2011).[17] The data on U.S. outward FDI form an unbalanced panel data set with annual observations for 341 bilateral industry level relationships and a time span from 1989 to 1999. It suffers from truncation because data points with FDI reported as zero are excluded from the sample.[18] In addition, some data points are missing because they could be associated with individual companies, and some are genuinely unavailable. Table 29 provides summary statistics for the final estimation sample that includes 2,767 observations. Given the set of 69 countries, 7 manufacturing industries, and 11 time periods, this implies a coverage of only 52 percent. With the data

[16] Juodis (2018) provides a detailed discussion of this peculiarity.

[17] Summary statistics are unavailable in the original paper by Egger and Pfaffermayr (2004a). We have computed the statistics in this appendix based on the data set in the Journal of Applied Econometrics Data Archive.

[18] Egger and Pfaffermayr (2004a) log transform their data and thus exclude zeros for obvious reasons.

at hand, it is not possible to carry out a systematic analysis about the magnitude or direction of a potential sample selection bias. If truncation from below was the major cause, we would overestimate the conditional mean of the dependent variable but the effect on the coefficients would remain unclear.

The data set of Egger and Pfaffermayr (2004a) has another peculiarity. All independent variables are constant across industry within any partner country. Table 29 counts these observations multiple times, depending on the number of nonmissing industries for a given country-year pair. In contrast, Table 30 provides the corresponding summary statistics for distinct observations only. In a strongly balanced panel, the variable means in both tables would coincide. Large differences indicate that sample selection is correlated with the respective variable. Here, we might be concerned that missing values are more likely to occur the stronger the similarity is between the U.S. and the partner country in terms of GDP, and the smaller the relative advantage of the U.S. is in terms of the factor endowments. Given that these variables have relatively little variation over time, we might be confident that the country-industry specific effects can account for most of this correlation. Although not uniformly, we also observe that the number of missing observations tends to decline over time. Any potential concern in this regard might be alleviated by the inclusion of time effects in all regressions. Finally, we do not find evidence that sample selection is related to the key time-invariant variable, geographical distance.

Based on a New Trade Theory gravity model, Egger and Pfaffermayr (2004a) expect a positive impact of bilateral GDP on real outward FDI. The similarity of the country size (in terms of GDP) is presumed to have an effect on (horizontal) FDI, too, but its sign is a priori unclear. An important role in their model play the relative factor endowments in physical capital, human capital, and labor. The first two of them are expected to be supportive of FDI while the last one should have the opposite sign. For the variable of main interest, geographical distance, the model predicts a positive effect in a scenario of predominantly vertical FDI and an ambiguous effect when horizontal MNEs dominate.

F.2 Additional empirical results

In our Monte Carlo simulation, we considered an alternative GMM estimator with the nonlinear moment conditions (A.4) proposed by Ahn and Schmidt (1995). We present the respective empirical results for the dynamic model with geographical distance as the only time-invariant regressor in the first two columns of Table 31. The estimates reveal again a substantial degree of history dependence, while all other explanatory variables turn statistically insignificant. The Hausman-Taylor instruments are the same as in the last three columns of Table 4 in the main paper. The resulting overidentifying restriction is not rejected by the difference-in-Hansen and the second-stage Hansen test.

For completeness, column 3 of Table 31 displays the estimates from the two-stage analogue of the system GMM estimator with the Mundlak (1978) projection and the additional time-invariant control variables in the last column of Table 5.[19] The qualitative conclusions do not differ. With the Mundlak CRE approach, a one-stage QML alternative is available. This is the Bhargava and Sargan (1983) QML estimator that is based on the model in levels. Similarly to the QML estimator of Hsiao et al. (2002), it accounts for the endogeneity of the lagged dependent variable by modeling the marginal distribution of the initial observations. The results are shown in column 4 of Table 31. Despite the loss of observations due to time series gaps, the results are remarkably similar to the M-sGMM estimates.

Finally, we make once again use of the robustness property of the two-stage approach by applying the Mundlak (1978) assumption only in the second stage. The Bhargava and Sargan (1983) QML estimator is not applicable in the first stage if some of the time-varying regressors (besides the lagged dependent variable) are allowed to be correlated with the unobserved effect. We thus have to resort again to the Hsiao et al. (2002) first-difference QML estimator. The first-stage results in column 5 of Table 31 are identical to those in column 5 of Table 5. In the second stage, only the common border and the colonial relationship dummies are found to statistically significantly affect the formation of outward FDI, both with the expected positive sign.

In another model specification, we have added the absolute value of the relative physical capital endowment as a regressor because it is standard practice to include the main effects of interaction terms. However, this main effect turned out to be clearly statistically insignificant. For brevity, we do not report these results.

[19] Besides the 5 time-invariant explanatory variables and the regression constant, the second-stage regression includes the 7 within-group averages of the strictly exogenous time-varying regressors and 9 within-group averages of the time dummies, adding up to a total of 22 coefficients.

References

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

Amemiya, T. and T. E. MaCurdy (1986). Instrumental-variable estimation of an error-components model. *Econometrica 54*, 869–880.

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

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

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

Baum, C. F., M. E. Schaffer, and S. Stillman (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal 3*, 1–31.

Bhargava, A. and J. D. Sargan (1983). Estimating dynamic random effects models from panel data covering short time periods. *Econometrica 51*, 1635–1659.

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*, 795–837.

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

Blundell, R., S. R. Bond, and F. Windmeijer (2001). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In B. H. Baltagi, T. F. Fomby, and R. C. Hill (Eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Volume 15 of *Advances in Econometrics*, pp. 53–91. Emerald Group Publishing.

Breusch, T. S., G. E. Mizon, and P. Schmidt (1989). Efficient estimation using panel data. *Econometrica 57*, 695–700.

Bun, M. J. G. and V. Sarafidis (2015). Dynamic panel data models. In B. H. Baltagi

(Ed.), *The Oxford Handbook of Panel Data*, Chapter 3, pp. 76–110. Oxford: Oxford University Press.

Bun, M. J. G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal 13*, 95–126.

Egger, P. and M. Pfaffermayr (2004a). Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics 19*, 227–246.

Egger, P. and M. Pfaffermayr (2004b). Estimating long and short run effects in static panel models. *Econometric Reviews 23*, 199–214.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics 14*, 262–280.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*, 1251–1271.

Hausman, J. A. and W. E. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*, 1377–1398.

Hayakawa, K. and M. H. Pesaran (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models. *Journal of Econometrics 188*, 111–134.

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

Juodis, A. (2018). First difference transformation in panel VAR models: Robustness, estimation, and inference. *Econometric Reviews*, 650–693.

Kripfganz, S. (2016). Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal 16*, 1013–1038.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distance measures: The GeoDist database. CEPII Working Paper 2011-25, Centre d'études prospectives et d'informations internationales.

Mehrhoff, J. (2009). A solution to the problem of too many instruments in dynamic panel data GMM. Discussion Paper, Series 1: Economic Studies 31/2009, Deutsche Bundesbank.

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

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: North-Holland.

Pesaran, M. H. and Q. Zhou (2018). Estimation of time-invariant effects in static panel data models. *Econometric Reviews 37*, 1137–1171.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*, 135–158.

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

Seo, M. H. and Y. Shin (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics 195*, 169–186.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

Windmeijer, F. (2000). Efficiency comparisons for a system GMM estimator in dynamic panel data models. In R. D. H. Heijmans, D. S. G. Pollock, and A. Sattora (Eds.), *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker*, Chapter 11, pp. 175–184. Dordrecht: Kluwer Academic Publishers.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*, 25–51.

Tables

Table 6: Simulation results: coefficient estimates, invalid instruments (i)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0523	0.0541	0.9479		0.9768					
	sGMM2	0.0062	0.0273	0.0802		0.9848					
	nlGMM1	0.0649	0.0671	0.9498		0.9754					
	nlGMM2	0.0091	0.0341	0.0619		0.9539					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	-0.0103	0.0444	0.0526		1.0192		1.0308	1.1456	0.5359	1.0000
	sGMM2	0.0016	0.0450	0.0520		1.0001		0.1655	0.5415	0.0464	0.9763
	nlGMM1	0.0043	0.0480	0.0495		1.0148		1.5499	1.7070	0.6309	0.9861
	nlGMM2	0.0014	0.0458	0.0494		1.0160		0.2515	2.5280	0.0392	0.4411
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0755	0.0836	0.5347		1.0210		1.6015	1.6394	0.9986	0.9556
	sGMM2	0.0017	0.0448	0.0521		1.0060		0.1661	0.5385	0.0452	0.9826
	nlGMM1	0.0518	0.0671	0.2339		1.0070		1.8741	1.9274	0.9985	0.9558
	nlGMM2	0.0017	0.0460	0.0486		1.0137		0.2520	2.4380	0.0374	0.4513
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0517	0.9945
γ_1	sGMM1	-0.1591	0.1665	0.8733		0.9815		-0.0138	0.1848	0.0521	0.9982
	sGMM2	-0.0187	0.0884	0.0801	0.4611	0.9795	0.3977	0.0001	0.1778	0.0492	1.0045
	nlGMM1	-0.1962	0.2044	0.9065		0.9840		-0.0112	0.1908	0.0534	0.9950
	nlGMM2	-0.0271	0.1076	0.0612	0.5110	0.9630	0.3265	-0.0001	0.2413	0.0477	0.8731
	QML2	0.0017	0.0714	0.0483	0.3333	1.0023	0.4970	0.0018	0.1771	0.0484	1.0033
γ_2	sGMM1	-0.1607	0.1820	0.5092		0.9730		-0.0296	0.4825	0.0525	0.9673
	sGMM2	-0.0068	0.1343	0.0685	0.2137	0.9412	0.6424	0.0564	0.5137	0.0787	0.8923
	nlGMM1	-0.2108	0.2281	0.6831		0.9885		-0.1269	0.5107	0.0492	0.9821
	nlGMM2	-0.0153	0.1523	0.0753	0.2822	0.9357	0.5630	0.0524	0.6940	0.0763	0.8028
	QML2	0.0187	0.1274	0.0684	0.1683	0.9170	0.7057	0.0821	0.5199	0.0890	0.8765

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}, \rho_{z,\alpha}) = (0.2, 0.4, 0.3, 0.3, 0)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is the same as in the baseline scenario. In the construction of the estimators, the collapsed moment condition (3) for x_{1it} is replaced by an identical collapsed moment condition for x_{2it}.

Table 7: Simulation results: coefficient estimates, invalid instruments (ii)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0524	0.0542	0.9443		0.9757					
	sGMM2	0.0062	0.0273	0.0802		0.9837					
	nlGMM1	0.0657	0.0679	0.9521		0.9784					
	nlGMM2	0.0091	0.0341	0.0619		0.9543					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	-0.0107	0.0444	0.0541		1.0225		1.0308	1.1466	0.5274	1.0013
	sGMM2	0.0016	0.0449	0.0504		1.0011		0.1647	0.5412	0.0445	0.9760
	nlGMM1	0.0044	0.0479	0.0484		1.0145		1.5769	1.7329	0.6432	0.9889
	nlGMM2	0.0014	0.0458	0.0494		1.0157		0.2558	2.3980	0.0387	0.4414
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0751	0.0833	0.5282		1.0228		1.6022	1.6413	0.9977	0.9572
	sGMM2	0.0017	0.0448	0.0515		1.0056		0.1655	0.5392	0.0469	0.9806
	nlGMM1	0.0501	0.0657	0.2191		1.0094		1.8906	1.9445	0.9984	0.9623
	nlGMM2	0.0017	0.0460	0.0494		1.0132		0.2560	2.2899	0.0372	0.4554
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.1577	0.1654	0.8597		0.9812		-0.0028	0.1880	0.0488	1.0017
	sGMM2	-0.0188	0.0888	0.0806	0.4485	0.9798	0.4039	-0.0012	0.1813	0.0508	1.0057
	nlGMM1	-0.1972	0.2054	0.9053		0.9847		-0.0010	0.1944	0.0492	0.9993
	nlGMM2	-0.0273	0.1081	0.0625	0.5030	0.9629	0.3320	-0.0012	0.2369	0.0482	0.8849
	QML2	0.0014	0.0719	0.0489	0.3270	0.9999	0.5040	0.0005	0.1807	0.0520	1.0045
γ_2	sGMM1	0.0972	0.1444	0.1263		0.9535		1.7164	1.7872	0.9555	0.9596
	sGMM2	0.3543	0.3956	0.5486	0.8989	0.9549	0.4988	1.9141	1.9843	0.9811	0.8928
	nlGMM1	0.0245	0.1188	0.0498		0.9717		1.6348	1.7133	0.9044	0.9808
	nlGMM2	0.3397	0.3983	0.4417	0.8534	0.9401	0.4184	1.9051	2.0493	0.9617	0.7655
	QML2	0.3930	0.4227	0.7884	0.9531	0.9375	0.5833	1.9473	2.0171	0.9847	0.8753

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}, \rho_{z,\alpha}) = (0.2, 0.4, 0.3, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by increasing $\rho_{z,\alpha}$ from 0 to 0.3. In the construction of the estimators, the collapsed moment condition (3) for x_{1it} is replaced by an identical collapsed moment condition for x_{2it}, as in Table 6.

Table 8: Simulation results: coefficient estimates, misspecified estimators

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	M-sGMM1	0.0029	0.0258	0.0671		0.9714					
	M-sGMM2	0.0062	0.0273	0.0802		0.9848					
	M-QML1	0.0022	0.0194	0.0584		0.9817					
β_1	M-sGMM1	-0.0018	0.0446	0.0495		1.0080		0.0939	0.5219	0.0481	0.9728
	M-sGMM2	0.0016	0.0450	0.0520		1.0001		0.1655	0.5415	0.0464	0.9763
	M-QML1	0.0005	0.0339	0.0485		1.0110		0.0601	0.3932	0.0470	0.8825
	HT-GMM1	0.2220	0.2333	0.7902		1.1936		-2.1780	2.1792	1.0000	1.1936
	FE-IV2	0.1414	0.1554	0.5921		0.9967		-2.2586	2.2595	1.0000	0.9967
β_2	M-sGMM1	-0.0021	0.0445	0.0484		1.0117		0.0920	0.5202	0.0492	0.9756
	M-sGMM2	0.0017	0.0448	0.0521		1.0060		0.1661	0.5385	0.0452	0.9826
	M-QML1	-0.0026	0.0347	0.0549		0.9937		0.0442	0.3809	0.0480	0.8941
	HT-GMM1	0.1484	0.1626	0.5610		1.0580		-2.2516	2.2526	1.0000	1.0580
	FE-IV2	0.1432	0.1572	0.6036		0.9928		-2.2568	2.2577	1.0000	0.9928
γ_1	M-sGMM1	-0.0111	0.0837	0.0689		0.9738		-0.0118	0.1637	0.0554	0.9780
	M-sGMM2	-0.0185	0.0873	0.0797	0.5014	0.9800	0.3594	0.0007	0.1597	0.0513	0.9965
	M-QML1	-0.0064	0.0660	0.0586		0.9867		0.0008	0.1588	0.0524	0.9951
	HT-GMM1	2.4017	2.4108	1.0000		1.0184		0.0017	0.2098	0.0478	1.0184
	FE-IV2	2.4015	2.4102	1.0000	1.0000	0.9939	0.9938	0.0015	0.2053	0.0537	0.9939
γ_2	M-sGMM1	0.1741	0.2055	0.3466		0.9747		0.9263	0.9420	0.9995	0.9701
	M-sGMM2	0.1643	0.1993	0.2802	0.8543	0.9853	0.2807	0.9419	0.9567	0.9996	0.9852
	M-QML1	0.1808	0.2000	0.5825		0.9827		0.9464	0.9611	0.9997	0.9799
	HT-GMM1	3.3707	3.4083	0.9996		0.9477		0.9707	1.0943	0.5422	0.9477
	FE-IV2	3.1865	3.2237	0.9996	0.9996	0.9460	0.9426	0.7865	0.9257	0.4268	0.9460

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is the same as in the baseline scenario. The estimator M-sGMM1 uses the within-group averages $\bar{x}_{1i}$ and $\bar{x}_{2i}$ as additional regresssors. The collapsed moment condition (3) for x_{1it} is replaced by the moment conditions (4) for $\bar{x}_{1i}$, $\bar{x}_{2i}$, and $\bar{x}_{1i}$, $\bar{f}_{2i}$. The estimator M-sGMM2 applies the same modifications in the second stage. The estimators HT-GMM1 and FE-IV2 constrain $\lambda = 0$. In contrast to sGMM1 and sGMM2, the collapsed moment conditions (A.1) for $y_{i,t-s}$ and (A.2) for $x_{1i,t-s}$ and $x_{2i,t-s}$ are replaced by the collapsed moment conditions (3) for $x_{1it} - \bar{x}_{1i}$ and $x_{2it} - \bar{x}_{2i}$. For these static-model estimators, the estimates of the long-run coefficients equal those of the short-run coefficients.

Table 9: Simulation results: coefficient estimates, no mean stationarity

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0843	0.0873	0.9579		1.0110					
	sGMM2	0.2933	0.2934	1.0000		1.2324					
	nlGMM1	0.0003	0.0144	0.0494		1.0051					
	nlGMM2	0.0009	0.0109	0.0548		0.9888					
	QML2	0.0000	0.0097	0.0486		1.0036					
β_1	sGMM1	-0.0851	0.0934	0.5413		1.0694		1.6194	1.8881	0.2128	0.9938
	sGMM2	-0.0051	0.0484	0.0259		1.1561		-9.4727	9.5292	1.0000	1.1630
	nlGMM1	-0.0005	0.0393	0.0507		0.9970		0.0155	0.2781	0.0505	0.9936
	nlGMM2	-0.0002	0.0489	0.0529		1.0000		0.0208	0.2950	0.0515	0.9920
	QML2	-0.0002	0.0366	0.0492		0.9984		0.0055	0.2280	0.0500	0.9988
β_2	sGMM1	0.0935	0.1020	0.5803		1.0679		3.2327	3.4992	0.9438	0.9831
	sGMM2	0.0023	0.0481	0.0264		1.1553		-9.5545	9.6120	1.0000	1.1673
	nlGMM1	0.0022	0.0485	0.0492		1.0032		0.0295	0.3222	0.0462	1.0008
	nlGMM2	0.0011	0.0486	0.0502		1.0031		0.0267	0.2909	0.0461	0.9957
	QML2	0.0005	0.0364	0.0500		1.0020		0.0089	0.2248	0.0479	1.0015
γ_1	sGMM1	-0.0987	0.1071	0.6404		1.0190		1.4637	1.6476	0.4940	0.9885
	sGMM2	-0.3395	0.3400	1.0000	1.0000	1.0148	1.0594	-5.8294	5.8436	1.0000	1.1921
	nlGMM1	-0.0025	0.0405	0.0543		0.9921		0.0015	0.2322	0.0536	0.9817
	nlGMM2	-0.0008	0.0379	0.0514	0.0658	0.9945	0.9367	0.0145	0.2100	0.0549	0.9827
	QML2	0.0001	0.0374	0.0506	0.0620	0.9956	0.9481	0.0053	0.2037	0.0546	0.9809
γ_2	sGMM1	0.1659	0.1928	0.4382		0.9189		3.7368	3.8115	0.9993	1.0814
	sGMM2	-0.4285	0.4327	1.0000	1.0000	0.8246	0.8150	-4.8550	4.9026	0.9992	0.9198
	nlGMM1	0.0018	0.0901	0.0470		1.0057		0.0210	0.4551	0.0426	1.0107
	nlGMM2	-0.0014	0.0880	0.0433	0.0539	1.0080	0.9678	0.0107	0.4461	0.0439	1.0072
	QML2	-0.0002	0.0862	0.0460	0.0493	1.0087	0.9883	0.0033	0.4349	0.0434	1.0078

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by setting $y_{i0} = x_{1i0} = x_{2i0} = 0$ with no burn-in period.

Table 10: Simulation results: coefficient estimates, lower variance ratio ($\omega = 1$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	-0.0001	0.0164	0.0506		0.9969					
	sGMM2	0.0021	0.0219	0.0581		0.9968					
	nlGMM1	-0.0008	0.0220	0.0486		1.0018					
	nlGMM2	0.0038	0.0293	0.0515		0.9850					
	QML2	-0.0004	0.0206	0.0482		1.0004					
β_1	sGMM1	-0.0008	0.0322	0.0489		1.0109		0.0096	0.2331	0.0509	0.9994
	sGMM2	0.0017	0.0433	0.0513		1.0006		0.0781	0.4236	0.0466	0.9908
	nlGMM1	-0.0001	0.0359	0.0506		1.0086		0.0126	0.2781	0.0500	0.9944
	nlGMM2	0.0011	0.0456	0.0481		1.0140		0.1435	0.6457	0.0453	0.9003
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0023	0.0413	0.0511		1.0039		0.0318	0.3371	0.0480	1.0041
	sGMM2	0.0019	0.0431	0.0497		1.0062		0.0790	0.4214	0.0467	0.9968
	nlGMM1	0.0020	0.0456	0.0512		1.0050		0.0351	0.4204	0.0493	0.9984
	nlGMM2	0.0014	0.0458	0.0488		1.0123		0.1446	0.6350	0.0446	0.9169
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.0032	0.0648	0.0543		0.9937		-0.0173	0.2120	0.0518	0.9960
	sGMM2	-0.0067	0.0761	0.0564	0.2968	0.9952	0.5398	-0.0011	0.2083	0.0479	1.0081
	nlGMM1	-0.0005	0.0781	0.0524		0.9961		-0.0147	0.2117	0.0537	0.9913
	nlGMM2	-0.0114	0.0966	0.0493	0.3998	0.9939	0.4275	0.0011	0.2110	0.0431	1.0247
	QML2	0.0016	0.0743	0.0475	0.2806	1.0047	0.5554	0.0016	0.2063	0.0467	1.0104
γ_2	sGMM1	-0.0024	0.1248	0.0469		0.9942		-0.0204	0.5325	0.0515	0.9881
	sGMM2	-0.0145	0.1513	0.0499	0.2272	0.9948	0.6213	-0.0648	0.5898	0.0430	0.9906
	nlGMM1	-0.0010	0.1376	0.0486		0.9927		-0.0284	0.5451	0.0500	0.9795
	nlGMM2	-0.0210	0.1845	0.0524	0.3193	0.9927	0.5112	-0.0995	0.6909	0.0382	0.9832
	QML2	0.0019	0.1424	0.0486	0.1899	0.9961	0.6623	-0.0169	0.5456	0.0477	0.9969

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 1$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing ω from 3 to 1.

Table 11: Simulation results: coefficient estimates, lower variance ratio ($\omega = 1/3$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	-0.0006	0.0147	0.0483		1.0008					
	sGMM2	0.0003	0.0187	0.0500		1.0020					
	nlGMM1	-0.0017	0.0204	0.0484		1.0064					
	nlGMM2	0.0006	0.0264	0.0504		0.9992					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	0.0002	0.0294	0.0470		1.0137		0.0029	0.1932	0.0503	1.0004
	sGMM2	0.0018	0.0415	0.0517		1.0026		0.0409	0.3658	0.0470	0.9956
	nlGMM1	0.0012	0.0339	0.0478		1.0144		0.0006	0.2341	0.0493	1.0071
	nlGMM2	0.0009	0.0455	0.0490		1.0125		0.0733	0.5218	0.0518	0.9664
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0506	0.9974
β_2	sGMM1	0.0019	0.0400	0.0516		1.0039		0.0182	0.3134	0.0478	1.0054
	sGMM2	0.0020	0.0414	0.0512		1.0067		0.0420	0.3642	0.0466	1.0011
	nlGMM1	0.0018	0.0456	0.0507		1.0045		0.0155	0.3932	0.0490	1.0013
	nlGMM2	0.0013	0.0457	0.0493		1.0108		0.0749	0.5187	0.0493	0.9740
	QML2	0.0003	0.0348	0.0532		0.9961		0.0263	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.0007	0.0712	0.0499		0.9996		-0.0123	0.2811	0.0488	0.9993
	sGMM2	-0.0017	0.0767	0.0472	0.1622	1.0066	0.7148	-0.0027	0.2775	0.0455	1.0096
	nlGMM1	0.0030	0.0829	0.0480		1.0084		-0.0100	0.2818	0.0479	1.0157
	nlGMM2	-0.0020	0.0963	0.0441	0.2592	1.0109	0.5761	0.0001	0.2821	0.0418	1.0285
	QML2	0.0015	0.0826	0.0466	0.1945	1.0079	0.6650	0.0010	0.2759	0.0454	1.0148
γ_2	sGMM1	-0.0001	0.1566	0.0452		0.9988		-0.0173	0.7174	0.0462	0.9951
	sGMM2	-0.0081	0.1920	0.0461	0.1919	0.9984	0.6569	-0.0652	0.8324	0.0435	0.9936
	nlGMM1	0.0051	0.1684	0.0439		1.0117		-0.0113	0.7337	0.0462	1.0059
	nlGMM2	-0.0066	0.2366	0.0462	0.2874	0.9995	0.5391	-0.0904	0.9844	0.0397	0.9927
	QML2	0.0029	0.1899	0.0450	0.1882	0.9976	0.6652	-0.0260	0.7855	0.0445	1.0002

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 1/3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1, 1, 1, 1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing ω from 3 to 1/3.

Table 12: Simulation results: coefficient estimates, higher variance of the unit-specific error component

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0027	0.0212	0.0646		0.9929					
	sGMM2	0.0062	0.0273	0.0787		0.9856					
	nlGMM1	0.0033	0.0258	0.0586		0.9960					
	nlGMM2	0.0113	0.0355	0.0671		0.9425					
	QML2	-0.0004	0.0206	0.0482		1.0004					
β_1	sGMM1	-0.0027	0.0385	0.0552		1.0029		0.0572	0.3455	0.0535	0.9972
	sGMM2	0.0014	0.0450	0.0507		0.9995		0.1652	0.5396	0.0474	0.9782
	nlGMM1	-0.0025	0.0403	0.0554		0.9960		0.0814	0.4019	0.0525	0.9836
	nlGMM2	0.0016	0.0459	0.0493		1.0185		0.2893	2.1651	0.0352	0.4287
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0027	0.0437	0.0525		1.0028		0.0906	0.4187	0.0463	1.0005
	sGMM2	0.0019	0.0448	0.0516		1.0054		0.1673	0.5381	0.0467	0.9843
	nlGMM1	0.0026	0.0458	0.0509		1.0077		0.1184	0.5090	0.0437	0.9886
	nlGMM2	0.0023	0.0462	0.0474		1.0165		0.2962	1.9377	0.0337	0.4677
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.0137	0.0814	0.0705		0.9849		-0.0284	0.2519	0.0555	0.9880
	sGMM2	-0.0186	0.0943	0.0811	0.3392	0.9801	0.5122	0.0011	0.2451	0.0494	1.0032
	nlGMM1	-0.0149	0.0919	0.0668		0.9880		-0.0256	0.2515	0.0646	0.9663
	nlGMM2	-0.0334	0.1158	0.0707	0.4141	0.9590	0.4181	0.0016	0.2562	0.0480	1.0012
	QML2	0.0019	0.0790	0.0498	0.2275	1.0029	0.6187	0.0028	0.2439	0.0505	1.0028
γ_2	sGMM1	-0.0129	0.1456	0.0533		0.9866		-0.0364	0.6040	0.0499	0.9800
	sGMM2	-0.0305	0.1634	0.0647	0.2068	0.9867	0.6703	-0.0845	0.6266	0.0472	0.9865
	nlGMM1	-0.0228	0.1582	0.0656		0.9756		-0.0828	0.6267	0.0576	0.9555
	nlGMM2	-0.0507	0.1927	0.0728	0.2900	0.9717	0.5671	-0.1394	0.8532	0.0385	0.8751
	QML2	0.0011	0.1449	0.0519	0.1308	0.9930	0.7660	-0.0147	0.5895	0.0497	0.9900

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1, 1, 1, \sqrt{2})$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by increasing σ^2_α from 1 to 2.

Table 13: Simulation results: coefficient estimates, higher signal-to-noise ratio

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0004	0.0101	0.0511		0.9946					
	sGMM2	0.0019	0.0136	0.0576		0.9967					
	nlGMM1	0.0003	0.0124	0.0489		1.0012					
	nlGMM2	0.0018	0.0153	0.0495		0.9907					
	QML2	-0.0003	0.0106	0.0493		1.0011					
β_1	sGMM1	-0.0009	0.0184	0.0522		1.0069		0.0078	0.1552	0.0528	0.9975
	sGMM2	0.0008	0.0223	0.0495		1.0033		0.0466	0.2466	0.0455	0.9943
	nlGMM1	-0.0005	0.0193	0.0510		1.0070		0.0118	0.1760	0.0541	0.9973
	nlGMM2	0.0004	0.0227	0.0491		1.0086		0.0481	0.2761	0.0418	0.9833
	QML2	0.0000	0.0171	0.0475		1.0149		0.0042	0.1767	0.0499	1.0028
β_2	sGMM1	0.0014	0.0211	0.0523		1.0031		0.0210	0.1932	0.0490	1.0040
	sGMM2	0.0009	0.0223	0.0500		1.0059		0.0471	0.2462	0.0455	0.9967
	nlGMM1	0.0011	0.0228	0.0516		1.0048		0.0224	0.2253	0.0495	1.0037
	nlGMM2	0.0005	0.0229	0.0500		1.0067		0.0486	0.2747	0.0424	0.9895
	QML2	0.0001	0.0174	0.0529		0.9961		0.0048	0.1769	0.0485	1.0019
γ_1	sGMM1	-0.0046	0.0475	0.0566		0.9897		-0.0170	0.1826	0.0537	0.9922
	sGMM2	-0.0057	0.0539	0.0580	0.2051	0.9926	0.6566	-0.0007	0.1781	0.0490	1.0042
	nlGMM1	-0.0040	0.0520	0.0543		0.9964		-0.0149	0.1824	0.0534	0.9958
	nlGMM2	-0.0052	0.0577	0.0524	0.2396	0.9978	0.6139	0.0016	0.1781	0.0482	1.0094
	QML2	0.0012	0.0477	0.0487	0.1475	1.0045	0.7442	0.0019	0.1774	0.0492	1.0040
γ_2	sGMM1	-0.0030	0.0995	0.0489		0.9861		-0.0137	0.4436	0.0516	0.9811
	sGMM2	-0.0128	0.1122	0.0490	0.1587	0.9907	0.7189	-0.0468	0.4656	0.0471	0.9873
	nlGMM1	-0.0067	0.1059	0.0471		0.9893		-0.0344	0.4581	0.0509	0.9803
	nlGMM2	-0.0108	0.1203	0.0502	0.1948	0.9913	0.6704	-0.0412	0.4848	0.0472	0.9898
	QML2	0.0008	0.1006	0.0525	0.1115	0.9913	0.8031	-0.0058	0.4366	0.0496	0.9897

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 2$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by increasing τ from 0.5 to 2.

Table 14: Simulation results: coefficient estimates, swapped correlations

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0023	0.0213	0.0593		0.9921					
	sGMM2	0.0065	0.0273	0.0829		0.9831					
	nlGMM1	0.0043	0.0290	0.0585		0.9937					
	nlGMM2	0.0095	0.0342	0.0631		0.9504					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	-0.0027	0.0417	0.0513		0.9980		0.0553	0.3884	0.0481	0.9929
	sGMM2	0.0020	0.0452	0.0502		0.9993		0.1728	0.5438	0.0457	0.9753
	nlGMM1	-0.0013	0.0439	0.0517		0.9982		0.1246	0.5297	0.0479	0.9799
	nlGMM2	0.0018	0.0461	0.0492		1.0151		0.3072	2.4508	0.0387	0.4135
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0029	0.0432	0.0527		1.0029		0.0860	0.4154	0.0468	0.9997
	sGMM2	0.0017	0.0448	0.0517		1.0059		0.1711	0.5400	0.0461	0.9814
	nlGMM1	0.0030	0.0459	0.0519		1.0075		0.1526	0.5805	0.0429	0.9800
	nlGMM2	0.0019	0.0460	0.0493		1.0141		0.3044	2.2015	0.0367	0.4454
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.0109	0.0741	0.0642		0.9856		-0.0202	0.1842	0.0540	0.9894
	sGMM2	-0.0196	0.0884	0.0816	0.4593	0.9778	0.4025	0.0003	0.1806	0.0477	1.0086
	nlGMM1	-0.0162	0.0943	0.0613		0.9937		-0.0178	0.1856	0.0557	0.9915
	nlGMM2	-0.0284	0.1079	0.0626	0.5121	0.9603	0.3313	0.0034	0.2118	0.0459	0.9542
	QML2	0.0017	0.0714	0.0479	0.3305	1.0025	0.4978	0.0021	0.1775	0.0502	1.0067
γ_2	sGMM1	-0.0117	0.1486	0.0548		0.9875		-0.0520	0.5524	0.0491	0.9805
	sGMM2	-0.0406	0.1817	0.0701	0.4449	0.9844	0.4129	-0.1606	0.6593	0.0429	0.9822
	nlGMM1	-0.0276	0.1894	0.0633		0.9875		-0.1335	0.6801	0.0436	0.9759
	nlGMM2	-0.0567	0.2228	0.0627	0.5011	0.9655	0.3377	-0.2818	2.3854	0.0320	0.4688
	QML2	0.0029	0.1424	0.0507	0.2999	0.9990	0.5251	-0.0191	0.5087	0.0450	0.9988

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.4, 0.2, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by swapping the correlations $(\rho_{x1,f2}, \rho_{z,f2})$ from $(0.2, 0.4)$ to $(0.4, 0.2)$.

Table 15: Simulation results: coefficient estimates, equalized correlations

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0016	0.0205	0.0570		0.9982					
	sGMM2	0.0062	0.0273	0.0804		0.9845					
	nlGMM1	0.0024	0.0264	0.0525		1.0031					
	nlGMM2	0.0091	0.0341	0.0620		0.9540					
	QML2	-0.0004	0.0206	0.0482		1.0003					
β_1	sGMM1	-0.0031	0.0390	0.0534		1.0043		0.0378	0.3441	0.0485	1.0012
	sGMM2	0.0016	0.0449	0.0512		1.0003		0.1653	0.5414	0.0464	0.9763
	nlGMM1	-0.0013	0.0408	0.0521		1.0038		0.0782	0.4249	0.0477	0.9972
	nlGMM2	0.0014	0.0458	0.0498		1.0159		0.2532	2.4868	0.0390	0.4405
	QML2	0.0000	0.0341	0.0473		1.0145		0.0251	0.3563	0.0507	0.9974
β_2	sGMM1	0.0027	0.0431	0.0522		1.0044		0.0717	0.4003	0.0450	1.0056
	sGMM2	0.0017	0.0448	0.0516		1.0059		0.1659	0.5387	0.0457	0.9822
	nlGMM1	0.0025	0.0458	0.0512		1.0065		0.1070	0.5120	0.0443	0.9962
	nlGMM2	0.0017	0.0460	0.0486		1.0136		0.2536	2.3898	0.0375	0.4520
	QML2	0.0003	0.0348	0.0532		0.9961		0.0262	0.3574	0.0518	0.9945
γ_1	sGMM1	-0.0091	0.0721	0.0617		0.9945		-0.0215	0.1861	0.0497	1.0038
	sGMM2	-0.0187	0.0887	0.0774	0.4492	0.9809	0.4103	0.0002	0.1837	0.0440	1.0194
	nlGMM1	-0.0111	0.0874	0.0597		1.0037		-0.0194	0.1879	0.0517	1.0024
	nlGMM2	-0.0271	0.1078	0.0595	0.4994	0.9654	0.3400	0.0003	0.2547	0.0418	0.8802
	QML2	0.0017	0.0716	0.0472	0.3220	1.0053	0.5072	0.0018	0.1798	0.0467	1.0173
γ_2	sGMM1	-0.0036	0.1730	0.0422		0.9883		-0.0168	0.7273	0.0447	0.9829
	sGMM2	-0.0391	0.2101	0.0567	0.2648	0.9829	0.5873	-0.1586	0.8545	0.0369	0.9774
	nlGMM1	-0.0216	0.2000	0.0506		0.9886		-0.1119	0.8116	0.0415	0.9762
	nlGMM2	-0.0541	0.2471	0.0557	0.3350	0.9753	0.5014	-0.2398	1.9361	0.0304	0.6558
	QML2	0.0027	0.1753	0.0454	0.1592	0.9956	0.7044	-0.0201	0.7199	0.0407	0.9911

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.4, 0.2, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by equalizing the correlations $(\rho_{x1,f2}, \rho_{z,f2})$ from $(0.2, 0.4)$ to $(0.2, 0.2)$.

Table 16: Simulation results: coefficient estimates, higher persistence ($\lambda = 0.9$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0020	0.0144	0.0729		0.9876					
	sGMM2	0.0088	0.0241	0.1394		0.9716					
	nlGMM1	0.0019	0.0193	0.0587		1.0036					
	nlGMM2	0.0129	0.0346	0.0909		0.9520					
	QML2	-0.0003	0.0209	0.0522		0.9892					
β_1	sGMM1	-0.0046	0.0337	0.0637		1.0033		0.1105	0.5713	0.0511	0.9971
	sGMM2	0.0011	0.0431	0.0529		0.9986		0.7567	1.7903	0.0442	0.8764
	nlGMM1	-0.0037	0.0390	0.0539		0.9997		0.1705	0.7474	0.0401	0.9976
	nlGMM2	0.0011	0.0451	0.0516		1.0091		2.8631	84.2138	0.0454	6.3901
	QML2	-0.0001	0.0326	0.0482		1.0117		0.2290	1.7042	0.0714	0.7258
β_2	sGMM1	0.0028	0.0401	0.0512		1.0059		0.2182	0.8502	0.0468	0.9982
	sGMM2	0.0012	0.0429	0.0531		1.0048		0.7565	1.7833	0.0431	0.8799
	nlGMM1	0.0016	0.0447	0.0524		1.0059		0.2862	1.1592	0.0440	0.9767
	nlGMM2	0.0013	0.0451	0.0509		1.0098		2.7056	79.0527	0.0422	6.3184
	QML2	0.0001	0.0332	0.0528		0.9949		0.2309	1.7729	0.0712	0.7013
γ_1	sGMM1	-0.0118	0.0683	0.0796		0.9848		-0.0320	0.2685	0.0545	0.9947
	sGMM2	-0.0387	0.1075	0.1398	0.6823	0.9693	0.2418	-0.0004	0.2745	0.0421	1.0280
	nlGMM1	-0.0111	0.0879	0.0615		1.0018		-0.0300	0.2709	0.0518	1.0063
	nlGMM2	-0.0562	0.1528	0.0882	0.7389	0.9543	0.1713	0.0047	3.1828	0.0318	6.2864
	QML2	0.0015	0.0951	0.0475	0.5867	0.9910	0.2759	0.0035	0.2682	0.0448	1.0205
γ_2	sGMM1	-0.0109	0.1003	0.0629		0.9886		-0.0415	0.6767	0.0500	0.9871
	sGMM2	-0.0565	0.1607	0.1227	0.5290	0.9752	0.3700	-0.3217	1.0212	0.0361	0.9431
	nlGMM1	-0.0146	0.1183	0.0664		1.0066		-0.0934	0.7050	0.0485	0.9944
	nlGMM2	-0.0801	0.2241	0.0886	0.6080	0.9595	0.2671	-1.0603	27.3269	0.0226	6.2834
	QML2	0.0016	0.1428	0.0498	0.4047	0.9925	0.4215	-0.0923	1.0761	0.0437	0.8040

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.9$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by increasing λ, ϕ_1, and ϕ_2 from 0.8 to 0.9, respectively.

Table 17: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.7$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0014	0.0225	0.0529		1.0030					
	sGMM2	0.0045	0.0285	0.0593		0.9905					
	nlGMM1	0.0015	0.0273	0.0521		1.0016					
	nlGMM2	0.0066	0.0316	0.0503		0.9779					
	QML2	-0.0005	0.0206	0.0478		1.0033					
β_1	sGMM1	-0.0012	0.0394	0.0507		1.0048		0.0190	0.2090	0.0523	0.9994
	sGMM2	0.0017	0.0462	0.0503		1.0017		0.0649	0.2996	0.0469	0.9838
	nlGMM1	-0.0006	0.0406	0.0507		1.0035		0.0273	0.2347	0.0546	0.9933
	nlGMM2	0.0014	0.0468	0.0481		1.0155		0.0888	0.3475	0.0370	0.9571
	QML2	0.0001	0.0357	0.0458		1.0155		0.0071	0.1985	0.0498	1.0062
β_2	sGMM1	0.0026	0.0448	0.0510		1.0022		0.0340	0.2456	0.0499	1.0022
	sGMM2	0.0019	0.0460	0.0507		1.0076		0.0657	0.2984	0.0440	0.9891
	nlGMM1	0.0028	0.0469	0.0511		1.0060		0.0425	0.2858	0.0491	0.9964
	nlGMM2	0.0018	0.0470	0.0491		1.0144		0.0903	0.3472	0.0366	0.9600
	QML2	0.0004	0.0363	0.0523		0.9985		0.0082	0.2003	0.0500	0.9977
γ_1	sGMM1	-0.0080	0.0692	0.0593		0.9948		-0.0161	0.1455	0.0555	0.9894
	sGMM2	-0.0106	0.0786	0.0607	0.3062	0.9887	0.5339	0.0006	0.1410	0.0504	1.0015
	nlGMM1	-0.0078	0.0779	0.0586		0.9993		-0.0141	0.1452	0.0566	0.9855
	nlGMM2	-0.0155	0.0853	0.0545	0.3355	0.9906	0.4923	0.0012	0.1408	0.0496	1.0059
	QML2	0.0016	0.0648	0.0492	0.2031	1.0051	0.6517	0.0014	0.1405	0.0500	1.0017
γ_2	sGMM1	-0.0069	0.1239	0.0498		0.9900		-0.0168	0.3429	0.0510	0.9814
	sGMM2	-0.0173	0.1343	0.0559	0.1774	0.9881	0.7050	-0.0299	0.3457	0.0473	0.9866
	nlGMM1	-0.0117	0.1345	0.0530		0.9890		-0.0337	0.3542	0.0521	0.9736
	nlGMM2	-0.0232	0.1432	0.0540	0.2124	0.9911	0.6605	-0.0359	0.3571	0.0460	0.9907
	QML2	0.0007	0.1176	0.0503	0.1125	0.9935	0.8106	-0.0054	0.3302	0.0496	0.9891

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.7$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.7, respectively.

Table 18: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.6$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0013	0.0243	0.0548		1.0057					
	sGMM2	0.0033	0.0289	0.0544		0.9937					
	nlGMM1	0.0015	0.0282	0.0520		1.0027					
	nlGMM2	0.0053	0.0297	0.0511		0.9947					
	QML2	-0.0005	0.0209	0.0511		1.0016					
β_1	sGMM1	-0.0007	0.0417	0.0485		1.0043		0.0115	0.1565	0.0524	0.9978
	sGMM2	0.0016	0.0473	0.0495		1.0029		0.0321	0.2002	0.0470	0.9874
	nlGMM1	-0.0003	0.0426	0.0488		1.0030		0.0161	0.1690	0.0516	0.9950
	nlGMM2	0.0013	0.0480	0.0476		1.0130		0.0423	0.2094	0.0426	0.9909
	QML2	0.0001	0.0374	0.0452		1.0152		0.0027	0.1370	0.0499	1.0050
β_2	sGMM1	0.0028	0.0462	0.0518		1.0023		0.0211	0.1766	0.0475	0.9995
	sGMM2	0.0020	0.0472	0.0503		1.0084		0.0329	0.1992	0.0482	0.9938
	nlGMM1	0.0031	0.0481	0.0522		1.0054		0.0260	0.1965	0.0492	0.9967
	nlGMM2	0.0019	0.0482	0.0500		1.0121		0.0437	0.2102	0.0430	0.9899
	QML2	0.0004	0.0379	0.0505		1.0008		0.0036	0.1384	0.0503	0.9949
γ_1	sGMM1	-0.0075	0.0688	0.0584		0.9972		-0.0126	0.1220	0.0558	0.9893
	sGMM2	-0.0064	0.0731	0.0567	0.2117	0.9958	0.6427	0.0008	0.1181	0.0512	1.0003
	nlGMM1	-0.0074	0.0742	0.0583		1.0004		-0.0108	0.1217	0.0553	0.9863
	nlGMM2	-0.0102	0.0750	0.0547	0.2254	1.0053	0.6269	0.0011	0.1180	0.0502	1.0022
	QML2	0.0015	0.0630	0.0507	0.1410	1.0039	0.7501	0.0012	0.1179	0.0500	1.0006
γ_2	sGMM1	-0.0062	0.1288	0.0491		0.9904		-0.0121	0.2792	0.0504	0.9837
	sGMM2	-0.0117	0.1327	0.0490	0.1184	0.9897	0.7946	-0.0161	0.2791	0.0485	0.9875
	nlGMM1	-0.0104	0.1365	0.0512		0.9877		-0.0222	0.2874	0.0519	0.9757
	nlGMM2	-0.0161	0.1355	0.0506	0.1327	0.9969	0.7770	-0.0181	0.2829	0.0473	0.9910
	QML2	0.0001	0.1205	0.0485	0.0853	0.9927	0.8794	-0.0039	0.2711	0.0499	0.9889

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.6$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.6, respectively.

Table 19: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.5$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0012	0.0254	0.0530		1.0053					
	sGMM2	0.0026	0.0289	0.0525		0.9953					
	nlGMM1	0.0015	0.0283	0.0523		1.0037					
	nlGMM2	0.0044	0.0285	0.0516		0.9995					
	QML2	-0.0006	0.0211	0.0521		0.9988					
β_1	sGMM1	-0.0004	0.0438	0.0472		1.0042		0.0076	0.1242	0.0518	0.9969
	sGMM2	0.0015	0.0485	0.0473		1.0041		0.0183	0.1482	0.0511	0.9894
	nlGMM1	-0.0003	0.0446	0.0476		1.0032		0.0101	0.1304	0.0507	0.9978
	nlGMM2	0.0014	0.0493	0.0453		1.0104		0.0243	0.1495	0.0480	0.9968
	QML2	0.0000	0.0390	0.0443		1.0144		0.0011	0.1053	0.0493	1.0031
β_2	sGMM1	0.0028	0.0475	0.0529		1.0024		0.0146	0.1365	0.0488	0.9974
	sGMM2	0.0020	0.0484	0.0512		1.0081		0.0191	0.1474	0.0492	0.9971
	nlGMM1	0.0032	0.0494	0.0534		1.0043		0.0178	0.1470	0.0482	0.9972
	nlGMM2	0.0020	0.0495	0.0515		1.0088		0.0256	0.1504	0.0472	0.9939
	QML2	0.0004	0.0395	0.0509		1.0025		0.0020	0.1064	0.0519	0.9936
γ_1	sGMM1	-0.0070	0.0686	0.0563		0.9974		-0.0100	0.1055	0.0556	0.9898
	sGMM2	-0.0041	0.0700	0.0535	0.1584	0.9997	0.7264	0.0008	0.1022	0.0512	0.9998
	nlGMM1	-0.0068	0.0716	0.0567		1.0010		-0.0083	0.1052	0.0546	0.9888
	nlGMM2	-0.0072	0.0701	0.0542	0.1610	1.0078	0.7251	0.0010	0.1021	0.0507	1.0008
	QML2	0.0015	0.0627	0.0490	0.1093	1.0027	0.8143	0.0010	0.1021	0.0506	1.0000
γ_2	sGMM1	-0.0055	0.1312	0.0488		0.9914		-0.0089	0.2343	0.0509	0.9865
	sGMM2	-0.0088	0.1332	0.0495	0.0933	0.9907	0.8537	-0.0103	0.2363	0.0507	0.9885
	nlGMM1	-0.0086	0.1370	0.0502		0.9882		-0.0144	0.2409	0.0507	0.9791
	nlGMM2	-0.0123	0.1338	0.0505	0.0973	0.9964	0.8480	-0.0113	0.2382	0.0485	0.9912
	QML2	-0.0004	0.1243	0.0495	0.0711	0.9920	0.9180	-0.0034	0.2316	0.0502	0.9890

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.5$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.5, respectively.

Table 20: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.4$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0011	0.0260	0.0518		1.0038					
	sGMM2	0.0021	0.0288	0.0510		0.9959					
	nlGMM1	0.0015	0.0280	0.0491		1.0040					
	nlGMM2	0.0037	0.0275	0.0510		0.9994					
	QML2	-0.0005	0.0213	0.0515		0.9972					
β_1	sGMM1	-0.0003	0.0457	0.0474		1.0041		0.0053	0.1027	0.0524	0.9968
	sGMM2	0.0015	0.0497	0.0470		1.0045		0.0116	0.1172	0.0500	0.9909
	nlGMM1	-0.0003	0.0465	0.0472		1.0034		0.0065	0.1059	0.0517	0.9998
	nlGMM2	0.0014	0.0505	0.0474		1.0083		0.0154	0.1165	0.0496	0.9975
	QML2	-0.0001	0.0407	0.0441		1.0132		0.0005	0.0864	0.0483	1.0020
β_2	sGMM1	0.0028	0.0488	0.0516		1.0022		0.0108	0.1107	0.0488	0.9965
	sGMM2	0.0020	0.0497	0.0492		1.0072		0.0124	0.1166	0.0503	0.9991
	nlGMM1	0.0034	0.0508	0.0531		1.0029		0.0131	0.1168	0.0491	0.9974
	nlGMM2	0.0021	0.0509	0.0526		1.0056		0.0166	0.1173	0.0500	0.9934
	QML2	0.0004	0.0411	0.0496		1.0036		0.0013	0.0871	0.0522	0.9941
γ_1	sGMM1	-0.0065	0.0683	0.0546		0.9971		-0.0080	0.0928	0.0557	0.9906
	sGMM2	-0.0028	0.0682	0.0528	0.1268	1.0013	0.7890	0.0008	0.0901	0.0512	0.9996
	nlGMM1	-0.0061	0.0698	0.0553		1.0014		-0.0065	0.0926	0.0540	0.9916
	nlGMM2	-0.0052	0.0677	0.0545	0.1240	1.0065	0.7947	0.0009	0.0900	0.0507	1.0001
	QML2	0.0014	0.0630	0.0486	0.0930	1.0022	0.8581	0.0009	0.0900	0.0511	0.9996
γ_2	sGMM1	-0.0048	0.1318	0.0480		0.9922		-0.0065	0.2003	0.0509	0.9887
	sGMM2	-0.0071	0.1341	0.0478	0.0790	0.9914	0.8935	-0.0075	0.2054	0.0509	0.9892
	nlGMM1	-0.0066	0.1364	0.0498		0.9889		-0.0088	0.2059	0.0511	0.9821
	nlGMM2	-0.0099	0.1341	0.0477	0.0812	0.9955	0.8925	-0.0080	0.2066	0.0488	0.9915
	QML2	-0.0008	0.1277	0.0498	0.0642	0.9915	0.9415	-0.0031	0.2024	0.0503	0.9890

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.4$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.4, respectively.

Table 21: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.3$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0010	0.0263	0.0491		1.0020					
	sGMM2	0.0017	0.0284	0.0520		0.9961					
	nlGMM1	0.0014	0.0275	0.0506		1.0038					
	nlGMM2	0.0031	0.0267	0.0509		0.9984					
	QML2	-0.0005	0.0214	0.0505		0.9972					
β_1	sGMM1	-0.0003	0.0475	0.0477		1.0036		0.0037	0.0877	0.0532	0.9966
	sGMM2	0.0013	0.0509	0.0475		1.0041		0.0079	0.0970	0.0508	0.9920
	nlGMM1	-0.0004	0.0484	0.0468		1.0033		0.0043	0.0894	0.0513	1.0006
	nlGMM2	0.0014	0.0519	0.0471		1.0062		0.0104	0.0958	0.0497	0.9973
	QML2	-0.0001	0.0424	0.0435		1.0118		0.0002	0.0739	0.0481	1.0018
β_2	sGMM1	0.0029	0.0501	0.0531		1.0019		0.0083	0.0930	0.0496	0.9963
	sGMM2	0.0019	0.0510	0.0487		1.0061		0.0087	0.0965	0.0513	1.0003
	nlGMM1	0.0035	0.0522	0.0520		1.0014		0.0101	0.0969	0.0525	0.9976
	nlGMM2	0.0021	0.0522	0.0525		1.0030		0.0115	0.0966	0.0502	0.9931
	QML2	0.0004	0.0427	0.0498		1.0043		0.0010	0.0744	0.0510	0.9959
γ_1	sGMM1	-0.0060	0.0679	0.0564		0.9971		-0.0065	0.0826	0.0554	0.9915
	sGMM2	-0.0019	0.0670	0.0503	0.1070	1.0020	0.8360	0.0008	0.0803	0.0514	0.9994
	nlGMM1	-0.0054	0.0686	0.0557		1.0019		-0.0051	0.0825	0.0534	0.9941
	nlGMM2	-0.0037	0.0663	0.0537	0.1032	1.0050	0.8443	0.0008	0.0803	0.0511	0.9998
	QML2	0.0012	0.0632	0.0501	0.0802	1.0022	0.8894	0.0008	0.0803	0.0508	0.9994
γ_2	sGMM1	-0.0041	0.1309	0.0477		0.9925		-0.0047	0.1731	0.0514	0.9900
	sGMM2	-0.0061	0.1351	0.0474	0.0696	0.9917	0.9209	-0.0059	0.1814	0.0501	0.9896
	nlGMM1	-0.0047	0.1348	0.0509		0.9894		-0.0049	0.1783	0.0509	0.9843
	nlGMM2	-0.0082	0.1348	0.0477	0.0701	0.9949	0.9216	-0.0061	0.1822	0.0486	0.9917
	QML2	-0.0013	0.1303	0.0492	0.0591	0.9910	0.9567	-0.0030	0.1794	0.0506	0.9889

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.3$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.3, respectively.

Table 22: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.2$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0010	0.0263	0.0495		1.0006					
	sGMM2	0.0015	0.0279	0.0521		0.9966					
	nlGMM1	0.0013	0.0269	0.0485		1.0038					
	nlGMM2	0.0026	0.0259	0.0514		0.9980					
	QML2	-0.0004	0.0214	0.0503		0.9980					
β_1	sGMM1	-0.0003	0.0493	0.0479		1.0027		0.0027	0.0767	0.0526	0.9960
	sGMM2	0.0012	0.0523	0.0487		1.0028		0.0056	0.0831	0.0513	0.9925
	nlGMM1	-0.0005	0.0503	0.0471		1.0026		0.0029	0.0777	0.0513	1.0005
	nlGMM2	0.0014	0.0532	0.0480		1.0044		0.0073	0.0819	0.0489	0.9968
	QML2	-0.0002	0.0441	0.0456		1.0102		0.0001	0.0651	0.0478	1.0018
β_2	sGMM1	0.0029	0.0514	0.0516		1.0018		0.0067	0.0803	0.0503	0.9970
	sGMM2	0.0019	0.0524	0.0497		1.0052		0.0064	0.0826	0.0508	1.0014
	nlGMM1	0.0035	0.0535	0.0514		1.0003		0.0080	0.0830	0.0509	0.9979
	nlGMM2	0.0022	0.0536	0.0511		1.0013		0.0083	0.0825	0.0517	0.9937
	QML2	0.0004	0.0444	0.0486		1.0049		0.0009	0.0654	0.0505	0.9981
γ_1	sGMM1	-0.0054	0.0673	0.0566		0.9976		-0.0053	0.0741	0.0549	0.9925
	sGMM2	-0.0013	0.0661	0.0496	0.0898	1.0026	0.8717	0.0007	0.0722	0.0516	0.9993
	nlGMM1	-0.0048	0.0675	0.0550		1.0026		-0.0040	0.0740	0.0521	0.9961
	nlGMM2	-0.0026	0.0653	0.0527	0.0891	1.0041	0.8804	0.0007	0.0721	0.0513	0.9995
	QML2	0.0011	0.0632	0.0492	0.0750	1.0024	0.9126	0.0008	0.0722	0.0510	0.9993
γ_2	sGMM1	-0.0034	0.1286	0.0498		0.9922		-0.0034	0.1507	0.0514	0.9904
	sGMM2	-0.0054	0.1355	0.0481	0.0632	0.9917	0.9401	-0.0048	0.1618	0.0500	0.9897
	nlGMM1	-0.0028	0.1324	0.0514		0.9895		-0.0022	0.1556	0.0521	0.9856
	nlGMM2	-0.0069	0.1352	0.0480	0.0631	0.9945	0.9413	-0.0049	0.1625	0.0477	0.9917
	QML2	-0.0017	0.1320	0.0489	0.0570	0.9904	0.9670	-0.0028	0.1604	0.0508	0.9888

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.2$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.2, respectively.

Table 23: Simulation results: coefficient estimates, lower persistence ($\lambda = 0.1$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0009	0.0260	0.0504		1.0000					
	sGMM2	0.0012	0.0273	0.0536		0.9978					
	nlGMM1	0.0012	0.0261	0.0519		1.0040					
	nlGMM2	0.0022	0.0251	0.0532		0.9982					
	QML2	-0.0003	0.0213	0.0504		0.9990					
β_1	sGMM1	-0.0003	0.0511	0.0498		1.0015		0.0019	0.0684	0.0518	0.9954
	sGMM2	0.0011	0.0537	0.0499		1.0011		0.0040	0.0729	0.0532	0.9926
	nlGMM1	-0.0006	0.0521	0.0486		1.0017		0.0019	0.0691	0.0506	0.9998
	nlGMM2	0.0013	0.0546	0.0491		1.0027		0.0053	0.0720	0.0497	0.9964
	QML2	-0.0003	0.0459	0.0481		1.0087		0.0001	0.0586	0.0474	1.0018
β_2	sGMM1	0.0029	0.0528	0.0505		1.0021		0.0055	0.0708	0.0517	0.9983
	sGMM2	0.0019	0.0537	0.0493		1.0046		0.0048	0.0725	0.0497	1.0027
	nlGMM1	0.0036	0.0549	0.0505		0.9995		0.0066	0.0729	0.0527	0.9982
	nlGMM2	0.0022	0.0549	0.0496		1.0002		0.0062	0.0723	0.0516	0.9949
	QML2	0.0005	0.0460	0.0497		1.0055		0.0009	0.0588	0.0516	1.0004
γ_1	sGMM1	-0.0049	0.0664	0.0539		0.9987		-0.0043	0.0668	0.0547	0.9935
	sGMM2	-0.0008	0.0650	0.0505	0.0795	1.0034	0.8996	0.0007	0.0651	0.0519	0.9992
	nlGMM1	-0.0042	0.0663	0.0550		1.0034		-0.0032	0.0667	0.0516	0.9976
	nlGMM2	-0.0018	0.0644	0.0528	0.0794	1.0037	0.9073	0.0007	0.0651	0.0514	0.9994
	QML2	0.0009	0.0630	0.0509	0.0695	1.0026	0.9303	0.0007	0.0651	0.0512	0.9992
γ_2	sGMM1	-0.0027	0.1251	0.0497		0.9915		-0.0024	0.1316	0.0513	0.9902
	sGMM2	-0.0048	0.1354	0.0489	0.0598	0.9915	0.9538	-0.0041	0.1453	0.0503	0.9897
	nlGMM1	-0.0013	0.1291	0.0530		0.9892		-0.0004	0.1365	0.0526	0.9861
	nlGMM2	-0.0060	0.1351	0.0475	0.0597	0.9940	0.9549	-0.0041	0.1458	0.0481	0.9917
	QML2	-0.0020	0.1327	0.0498	0.0551	0.9899	0.9743	-0.0026	0.1442	0.0511	0.9887

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.1$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0.1, respectively.

Table 24: Simulation results: coefficient estimates, no persistence ($\lambda = 0$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0008	0.0255	0.0513		1.0000					
	sGMM2	0.0010	0.0265	0.0539		0.9995					
	nlGMM1	0.0010	0.0253	0.0525		1.0043					
	nlGMM2	0.0018	0.0244	0.0542		0.9986					
	QML2	-0.0002	0.0210	0.0485		0.9997					
β_1	sGMM1	-0.0003	0.0528	0.0493		1.0001		0.0014	0.0619	0.0487	0.9947
	sGMM2	0.0010	0.0552	0.0504		0.9991		0.0030	0.0653	0.0529	0.9923
	nlGMM1	-0.0006	0.0539	0.0493		1.0007		0.0012	0.0625	0.0505	0.9989
	nlGMM2	0.0013	0.0560	0.0493		1.0014		0.0039	0.0645	0.0503	0.9961
	QML2	-0.0003	0.0476	0.0487		1.0073		0.0001	0.0537	0.0478	1.0018
β_2	sGMM1	0.0029	0.0541	0.0497		1.0026		0.0046	0.0635	0.0479	1.0000
	sGMM2	0.0018	0.0551	0.0501		1.0043		0.0037	0.0648	0.0479	1.0040
	nlGMM1	0.0037	0.0563	0.0515		0.9991		0.0055	0.0653	0.0529	0.9985
	nlGMM2	0.0022	0.0563	0.0501		0.9997		0.0048	0.0647	0.0521	0.9962
	QML2	0.0005	0.0477	0.0487		1.0062		0.0009	0.0536	0.0504	1.0025
γ_1	sGMM1	-0.0044	0.0652	0.0525		0.9999		-0.0036	0.0603	0.0538	0.9944
	sGMM2	-0.0005	0.0638	0.0509	0.0718	1.0042	0.9215	0.0006	0.0589	0.0516	0.9990
	nlGMM1	-0.0037	0.0650	0.0543		1.0040		-0.0026	0.0602	0.0524	0.9986
	nlGMM2	-0.0012	0.0633	0.0527	0.0722	1.0037	0.9277	0.0006	0.0589	0.0514	0.9992
	QML2	0.0008	0.0624	0.0506	0.0662	1.0027	0.9441	0.0006	0.0589	0.0512	0.9991
γ_2	sGMM1	-0.0021	0.1204	0.0510		0.9905		-0.0016	0.1149	0.0511	0.9896
	sGMM2	-0.0044	0.1344	0.0488	0.0564	0.9910	0.9639	-0.0035	0.1309	0.0498	0.9894
	nlGMM1	-0.0001	0.1249	0.0527		0.9886		0.0007	0.1199	0.0538	0.9862
	nlGMM2	-0.0053	0.1342	0.0475	0.0570	0.9934	0.9646	-0.0035	0.1312	0.0487	0.9915
	QML2	-0.0022	0.1323	0.0501	0.0539	0.9894	0.9797	-0.0025	0.1300	0.0509	0.9886

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The data-generating process is modified by reducing λ, ϕ_1, and ϕ_2 from 0.8 to 0, respectively.

Table 25: Simulation results: coefficient estimates, smaller sample size ($N = 100$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0064	0.0385	0.0814		0.9869					
	sGMM2	0.0229	0.0557	0.1486		0.9576					
	nlGMM1	0.0081	0.0492	0.0772		0.9740					
	nlGMM2	0.0364	0.0739	0.1264		0.9295					
	QML2	0.0005	0.0405	0.0498		0.9674					
β_1	sGMM1	-0.0103	0.0740	0.0649		0.9867		0.1452	0.6612	0.0549	0.9834
	sGMM2	0.0038	0.0893	0.0548		0.9943		0.8018	4.1892	0.0419	0.5544
	nlGMM1	-0.0120	0.0792	0.0683		0.9826		0.2486	0.9428	0.0524	0.9133
	nlGMM2	0.0030	0.0914	0.0553		1.0055		-17.3201	1843.3787	0.0349	278.2778
	QML2	0.0002	0.0652	0.0489		0.9985		0.1416	1.1905	0.0649	0.7527
β_2	sGMM1	0.0083	0.0850	0.0590		0.9905		0.2727	0.8865	0.0469	0.9706
	sGMM2	0.0051	0.0903	0.0607		0.9815		0.8131	4.5151	0.0425	0.5315
	nlGMM1	0.0063	0.0903	0.0599		0.9871		0.4132	1.3552	0.0504	0.8818
	nlGMM2	0.0034	0.0921	0.0565		0.9952		-20.5159	2171.7355	0.0376	278.2973
	QML2	0.0003	0.0653	0.0513		0.9950		0.1429	1.1083	0.0655	0.7934
γ_1	sGMM1	-0.0317	0.1367	0.0931		0.9865		-0.0640	0.3808	0.0657	0.9822
	sGMM2	-0.0718	0.1783	0.1479	0.5078	0.9630	0.3872	-0.0092	0.4721	0.0473	0.9321
	nlGMM1	-0.0356	0.1657	0.0901		0.9693		-0.0587	0.3906	0.0670	0.9946
	nlGMM2	-0.1113	0.2309	0.1323	0.5819	0.9402	0.3042	0.1688	22.0373	0.0382	273.6874
	QML2	-0.0022	0.1415	0.0526	0.3327	0.9743	0.4882	-0.0045	0.3648	0.0543	0.9985
γ_2	sGMM1	-0.0240	0.2307	0.0637		0.9618		-0.0623	0.9391	0.0529	0.9394
	sGMM2	-0.1045	0.2894	0.1208	0.3615	0.9571	0.5383	-0.3355	1.6760	0.0320	0.8011
	nlGMM1	-0.0372	0.2573	0.0824		0.9632		-0.1259	1.0028	0.0496	0.9523
	nlGMM2	-0.1563	0.3531	0.1177	0.4516	0.9583	0.4481	2.0152	237.5422	0.0222	275.7977
	QML2	-0.0049	0.2412	0.0465	0.2019	0.9725	0.6569	-0.0687	1.0264	0.0415	0.9301

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 100$.
Note: See the notes for Table 2 in the main paper. The cross-sectional sample size N is reduced from 350 to 100.

Table 26: Simulation results: coefficient estimates, larger sample size ($N = 650$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0009	0.0143	0.0547		0.9902					
	sGMM2	0.0036	0.0198	0.0695		0.9809					
	nlGMM1	0.0008	0.0182	0.0543		0.9929					
	nlGMM2	0.0037	0.0228	0.0538		0.9785					
	QML2	0.0001	0.0151	0.0478		1.0023					
β_1	sGMM1	-0.0016	0.0269	0.0526		0.9987		0.0189	0.2237	0.0547	0.9900
	sGMM2	0.0004	0.0324	0.0475		1.0064		0.0880	0.3676	0.0490	0.9834
	nlGMM1	-0.0011	0.0283	0.0511		1.0017		0.0275	0.2619	0.0513	0.9841
	nlGMM2	0.0002	0.0333	0.0463		1.0111		0.1012	0.4399	0.0423	0.9648
	QML2	0.0002	0.0252	0.0458		1.0066		0.0184	0.2573	0.0464	0.9996
β_2	sGMM1	0.0016	0.0310	0.0493		1.0132		0.0386	0.2797	0.0505	0.9955
	sGMM2	0.0011	0.0322	0.0462		1.0137		0.0918	0.3702	0.0481	0.9797
	nlGMM1	0.0011	0.0329	0.0466		1.0141		0.0443	0.3383	0.0466	0.9925
	nlGMM2	0.0007	0.0330	0.0467		1.0201		0.1037	0.4420	0.0434	0.9624
	QML2	0.0002	0.0252	0.0481		1.0080		0.0184	0.2562	0.0479	1.0039
γ_1	sGMM1	-0.0053	0.0505	0.0610		0.9881		-0.0122	0.1324	0.0537	0.9944
	sGMM2	-0.0110	0.0641	0.0719	0.4394	0.9773	0.4047	-0.0003	0.1301	0.0475	1.0041
	nlGMM1	-0.0047	0.0605	0.0560		0.9920		-0.0111	0.1327	0.0540	0.9904
	nlGMM2	-0.0109	0.0731	0.0565	0.4866	0.9804	0.3538	0.0005	0.1305	0.0469	1.0063
	QML2	-0.0001	0.0524	0.0516	0.3324	0.9980	0.4955	0.0004	0.1300	0.0485	1.0017
γ_2	sGMM1	-0.0046	0.0844	0.0528		0.9848		-0.0137	0.3229	0.0516	0.9833
	sGMM2	-0.0164	0.1016	0.0650	0.2695	0.9865	0.5779	-0.0389	0.3376	0.0503	0.9923
	nlGMM1	-0.0069	0.0950	0.0526		0.9868		-0.0294	0.3331	0.0474	0.9795
	nlGMM2	-0.0159	0.1140	0.0543	0.3201	0.9892	0.5139	-0.0416	0.3574	0.0452	0.9962
	QML2	-0.0010	0.0867	0.0507	0.1893	0.9947	0.6791	-0.0100	0.3180	0.0486	0.9902

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 6$, and $N = 650$.
Note: See the notes for Table 2 in the main paper. The cross-sectional sample size N is increased from 350 to 650.

Table 27: Simulation results: coefficient estimates, smaller sample size ($T = 3$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0036	0.0305	0.0675		0.9944					
	sGMM2	0.0168	0.0504	0.1393		0.9662					
	nlGMM1	-0.0029	0.0456	0.0511		0.9959					
	nlGMM2	0.0040	0.0722	0.0652		0.9753					
	QML2	0.0046	0.0630	0.0504		0.9433					
β_1	sGMM1	-0.0057	0.0547	0.0556		0.9974		0.0782	0.4289	0.0446	1.0078
	sGMM2	0.0030	0.0724	0.0456		1.0144		0.5691	2.2452	0.0442	0.6129
	nlGMM1	0.0002	0.0632	0.0532		0.9989		0.0809	0.7258	0.0495	0.8511
	nlGMM2	0.0006	0.0761	0.0437		1.0193		0.2432	35.7086	0.0915	18.2589
	QML2	0.0004	0.0678	0.0471		0.9978		0.5423	8.4714	0.0886	1.4245
β_2	sGMM1	0.0037	0.0695	0.0540		0.9976		0.1540	0.6636	0.0475	0.9863
	sGMM2	0.0016	0.0736	0.0511		0.9984		0.5602	2.1268	0.0430	0.6349
	nlGMM1	-0.0009	0.0765	0.0518		0.9993		0.1381	1.2284	0.0702	0.7776
	nlGMM2	-0.0009	0.0772	0.0496		1.0061		0.1788	41.0247	0.0961	18.3025
	QML2	-0.0001	0.0676	0.0494		1.0008		0.5449	9.1259	0.0867	1.4018
γ_1	sGMM1	-0.0129	0.0999	0.0685		0.9917		-0.0104	0.1982	0.0514	1.0076
	sGMM2	-0.0515	0.1541	0.1383	0.6659	0.9678	0.2529	-0.0035	0.2226	0.0426	1.0042
	nlGMM1	0.0069	0.1436	0.0538		0.9872		-0.0107	0.2021	0.0517	1.0098
	nlGMM2	-0.0122	0.2218	0.0625	0.7178	0.9737	0.1811	-0.0045	1.0798	0.0349	13.3417
	QML2	-0.0138	0.1947	0.0463	0.6657	0.9427	0.2032	-0.0112	0.5464	0.0380	1.3735
γ_2	sGMM1	-0.0111	0.1470	0.0542		0.9941		-0.0166	0.5023	0.0477	0.9952
	sGMM2	-0.0737	0.2315	0.1210	0.4993	0.9725	0.3838	-0.2317	1.0334	0.0343	0.7798
	nlGMM1	0.0073	0.1906	0.0581		0.9936		-0.0372	0.5544	0.0453	0.9658
	nlGMM2	-0.0169	0.3224	0.0556	0.5642	0.9827	0.2853	-0.1007	13.3519	0.0380	17.3573
	QML2	-0.0196	0.2839	0.0432	0.4997	0.9546	0.3189	-0.2178	3.9745	0.0382	1.3414

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 3$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The time horizon T is reduced from 6 to 3.

Table 28: Simulation results: coefficient estimates, larger sample size ($T = 10$)

		short-run coefficients						long-run coefficients			
		Bias	RMSE	Size	(uncorr.)	SE/SD	(uncorr.)	Bias	RMSE	Size	SE/SD
λ	sGMM1	0.0015	0.0141	0.0582		0.9865					
	sGMM2	0.0034	0.0180	0.0654		0.9791					
	nlGMM1	0.0044	0.0176	0.0680		0.9870					
	nlGMM2	0.0129	0.0257	0.0967		0.9412					
	QML2	-0.0001	0.0109	0.0509		1.0054					
β_1	sGMM1	-0.0007	0.0280	0.0526		0.9897		0.0322	0.2393	0.0530	0.9916
	sGMM2	0.0013	0.0322	0.0526		0.9922		0.0845	0.3418	0.0491	0.9856
	nlGMM1	-0.0023	0.0292	0.0575		0.9869		0.0753	0.2837	0.0511	0.9921
	nlGMM2	0.0027	0.0330	0.0526		1.0071		0.2719	0.5456	0.0313	0.9268
	QML2	0.0003	0.0230	0.0561		0.9853		0.0079	0.1849	0.0482	1.0003
β_2	sGMM1	0.0015	0.0314	0.0561		0.9850		0.0457	0.2803	0.0573	0.9824
	sGMM2	0.0010	0.0324	0.0552		0.9878		0.0836	0.3447	0.0505	0.9775
	nlGMM1	0.0030	0.0331	0.0534		0.9923		0.1072	0.3516	0.0480	0.9795
	nlGMM2	0.0024	0.0333	0.0528		1.0005		0.2712	0.5480	0.0338	0.9222
	QML2	-0.0001	0.0230	0.0548		0.9875		0.0059	0.1856	0.0485	0.9959
γ_1	sGMM1	-0.0085	0.0553	0.0653		0.9856		-0.0206	0.1776	0.0589	0.9861
	sGMM2	-0.0104	0.0631	0.0698	0.3014	0.9821	0.5401	-0.0001	0.1713	0.0525	1.0014
	nlGMM1	-0.0167	0.0638	0.0719		0.9907		-0.0183	0.1788	0.0571	0.9860
	nlGMM2	-0.0386	0.0832	0.1024	0.4407	0.9549	0.4348	0.0008	0.1717	0.0512	1.0042
	QML2	0.0005	0.0475	0.0508	0.1610	0.9997	0.7205	0.0010	0.1712	0.0529	0.9992
γ_2	sGMM1	-0.0077	0.0983	0.0469		0.9974		-0.0226	0.4073	0.0470	0.9977
	sGMM2	-0.0160	0.1072	0.0568	0.1759	0.9948	0.7120	-0.0377	0.4115	0.0442	1.0008
	nlGMM1	-0.0181	0.1080	0.0588		0.9950		-0.0343	0.4196	0.0470	0.9975
	nlGMM2	-0.0554	0.1322	0.0895	0.3007	0.9743	0.5997	-0.1085	0.4515	0.0436	0.9973
	QML2	0.0001	0.0887	0.0467	0.0895	0.9977	0.8644	-0.0040	0.3911	0.0486	0.9968

Simulation design according to the data-generating process in Section 6.1: $\lambda = 0.8$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \sqrt{1-\lambda^2}$, $\omega = 3$, $\tau = 0.5$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \text{diag}(1,1,1,1)$, $(\rho_{x1,f2}, \rho_{z,f2}, \rho_{x2,\alpha}, \rho_{f2,\alpha}) = (0.2, 0.4, 0.3, 0.3)$, $T = 10$, and $N = 350$.
Note: See the notes for Table 2 in the main paper. The time horizon T is increased from 6 to 10.

Table 29: Summary statistics: estimation sample

	Obs.	Mean	Std. Dev.	Min.	Max.
ln(outward FDI)$_{ijt}$	2,767	4.677	2.372	0	9.697
ln(distance)$_i$	2,767	8.594	0.351	7.560	9.234
ln(distance)$_i$ × ln(relative capital-labor ratio)$_{it}$	2,767	-0.971	12.234	-18.386	29.066
ln(bilateral GDP)$_{it}$	2,767	27.085	0.650	26.285	29.394
ln(bilateral GDP)$_{it}$ × |ln(relative physical capital endowment)$_{it}$|	2,767	39.421	31.008	0.095	128.683
ln(similarity in country size)$_{it}$	2,767	-1.358	0.806	-4.483	-0.693
ln(relative physical capital endowment)$_{it}$	2,767	0.805	1.681	-3.074	4.841
ln(relative human capital endowment)$_{it}$	2,767	0.045	0.813	-1.526	2.854
ln(relative labor endowment)$_{it}$	2,767	0.927	1.458	-3.324	5.148
common border$_i$	2,767	0.048	0.215	0	1
common language, official$_i$	2,767	0.316	0.465	0	1
common language, other$_i$	2,767	0.570	0.495	0	1
colonial relationship$_i$	2,767	0.104	0.305	0	1

Note: See Egger and Pfaffermayr (2004a) and Mayer and Zignago (2011) for a variable description. Subscripts i, j, and t indicate variation across country, sector, and year, respectively. The columns display the number of observations, the mean and standard deviation, and the minimum and maximum value.

Table 30: Summary statistics: distinct observations

	Obs.	Mean	Std. Dev.	Min.	Max.
ln(outward FDI)$_{ijt}$	2,767	4.677	2.372	0	9.697
ln(distance)$_i$	69	8.611	0.350	7.560	9.234
ln(distance)$_i$ × ln(relative capital-labor ratio)$_{it}$	677	1.664	12.624	-18.386	29.066
ln(bilateral GDP)$_{it}$	677	26.911	0.562	26.285	29.394
ln(bilateral GDP)$_{it}$ × |ln(relative physical capital endowment)$_{it}$|	677	48.129	34.696	0.095	128.683
ln(similarity in country size)$_{it}$	677	-1.629	0.993	-4.483	-0.693
ln(relative physical capital endowment)$_{it}$	677	1.390	1.737	-3.074	4.841
ln(relative human capital endowment)$_{it}$	677	0.241	0.883	-1.526	2.854
ln(relative labor endowment)$_{it}$	677	1.206	1.486	-3.324	5.148
common border$_i$	69	0.029	0.169	0	1
common language, official$_i$	69	0.275	0.450	0	1
common language, other$_i$	69	0.536	0.502	0	1
colonial relationship$_i$	69	0.058	0.235	0	1

Note: See Egger and Pfaffermayr (2004a) and Mayer and Zignago (2011) for a variable description. Subscripts i, j, and t indicate variation across country, sector, and year, respectively. The columns show the number of distinct observations, the corresponding mean and standard deviation, and the minimum and maximum value.

Table 31: Estimation results: additional estimates

$\ln(\text{outward FDI})_{it}$	nlGMM1[b]	nlGMM2[b]	M-sGMM2	M-QML1	M-QML2
$\ln(\text{outward FDI})_{i,t-1}$	0.913	0.945	0.916	0.927	0.802
	(0.082)***	(0.069)***	(0.066)***	(0.056)***	(0.054)***
$\ln(\text{distance})_i$ ×	-0.132	0.060	-0.099	0.101	-0.842
$\ln(\text{rel. capital-labor ratio})_{it}$	(0.713)	(0.584)	(0.141)	(0.519)	(0.520)
$\ln(\text{bilateral GDP})_{it}$	0.680	0.596	1.237	1.639	2.305
	(0.614)	(0.695)	(0.719)*	(0.588)***	(0.605)***
$\ln(\text{bilateral GDP})_{it}$ ×	-0.006	-0.016	-0.007	-0.008	-0.009
$\lvert\ln(\text{rel. phys. capital endowment})_{it}\rvert$	(0.012)	(0.013)	(0.007)	(0.006)	(0.006)
$\ln(\text{similarity in country size})_{it}$	-0.032	-0.054	0.436	0.839	0.840
	(0.389)	(0.502)	(0.233)*	(0.466)*	(0.420)**
$\ln(\text{rel. phys. capital endowment})_{it}$	1.268	-0.307	1.385	-0.324	7.926
	(6.064)	(5.054)	(1.384)	(4.630)	(4.594)*
$\ln(\text{rel. human capital endowment})_{it}$	0.013	0.084	-0.040	0.060	0.081
	(0.146)	(0.152)	(0.134)	(0.108)	(0.090)
$\ln(\text{rel. labor endowment})_{it}$	-1.114	0.581	-0.741	1.854	-6.275
	(6.219)	(5.150)	(1.228)	(4.276)	(4.282)
$\ln(\text{distance})_i$	0.366	-0.722	-0.055	-0.038	-0.082
	(0.887)	(1.309)	(0.054)	(0.051)	(0.100)
common border_i			0.134	0.162	0.443
			(0.169)	(0.175)	(0.192)**
common language, official_i			0.053	0.006	0.110
			(0.061)	(0.044)	(0.067)
common language, other_i			-0.036	-0.027	-0.043
			(0.039)	(0.036)	(0.068)
colonial relationship_i			0.065	0.084	0.174
			(0.057)	(0.055)	(0.066)***
observations	2,198	2,198	2,198	1,614	1,664
units	337	337	337	173	227
1st stage					
instruments	60	58	58		
constant	yes	yes	yes	yes	no
year dummies	1991–1999	1991–1999	1991–1999	1991–1999	1991–1999
	χ^2_9=11.16	χ^2_9=9.11	χ^2_9=13.50	χ^2_9=29.61	χ^2_9=25.42
	[0.265]	[0.427]	[0.141]	[0.001]***	[0.003]***
Mundlak				χ^2_7=10.56	
				[0.159]	
Arellano-Bond	z=-0.014	z=-0.031	z=-0.016		
	[0.989]	[0.975]	[0.987]		
Hansen	χ^2_{41}=44.43	χ^2_{40}=42.62	χ^2_{40}=44.27		
	[0.239]	[0.359]	[0.296]		
difference-in-Hansen	χ^2_1=1.81				
	[0.178]				
2nd stage					
instruments		3	22		22
constant		yes	yes		yes
Mundlak			χ^2_7=6.50		χ^2_7=11.21
			[0.482]		[0.130]
Hansen		χ^2_1=0.48			
		[0.487]			
short-run marg. eff. of $\ln(\text{distance})_i$					
evaluated at the 5th percentile	0.636	-0.841	0.140	-0.237	1.574
	(1.836)	(1.717)	(0.274)	(1.031)	(1.051)
evaluated at the mean	0.383	-0.729	-0.043	-0.050	0.021
	(0.913)	(1.309)	(0.054)	(0.088)	(0.131)
evaluated at the 95th percentile	0.010	-0.559	-0.324	0.235	-2.356
	(1.907)	(2.078)	(0.392)	(1.393)	(1.382)*
long-run marg. eff. of $\ln(\text{distance})_i$					
evaluated at the 5th percentile	7.210	-15.412	1.676	-3.232	7.933
	(17.669)	(43.339)	(2.868)	(14.718)	(5.386)
evaluated at the mean	4.406	-13.369	-0.519	-0.684	0.105
	(9.499)	(33.346)	(0.708)	(1.397)	(0.662)
evaluated at the 95th percentile	0.116	-10.242	-3.876	3.215	-11.874
	(22.013)	(39.864)	(3.811)	(19.592)	(7.094)*

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Note: See Egger and Pfaffermayr (2004a) and Mayer and Zignago (2011) for a data description. We abbreviate the estimators as follows: "sGMM" and "nlGMM" refer to GMM estimators that are described in Table 1. "QML2" refers to the Hsiao et al. (2002) QML estimator in the first stage and "QML1" to the Bhargava and Sargan (1983) QML estimator. The prefix "M-" denotes a Mundlak (1978) projection (including time dummies) as discussed in Remark 3 or 5. The trailing numbers 1 or 2 denote one-stage and two-stage estimators, respectively. The exogenous variables according to Assumption 3 are the similarity in country size[(b)] and the relative human capital endowment[(b)]. Standard errors robust to serial correlation and heteroskedasticity are in parentheses. The test statistics are a Wald test for the joint insignificance of the time dummies, a Wald test for joint insignificance of the within-group averages of the time-varying regressors (excluding time dummies), the Arellano and Bond (1991) test for no second-order serial correlation in the first-differenced residuals, as well as the Hansen and difference-in-Hansen tests discussed in Section 5. The respective p-values are in brackets. The marginal effects of $\ln(\text{distance})_i$ are evaluated at the 5th percentile (-1.966), the mean (-0.122), and the 95th percentile (2.701) of $\ln(\text{rel. capital-labor ratio})_{it}$ for the full sample of 2,767 observations. The long-run marginal effects in dynamic models are obtained as the short-run marginal effects divided by one minus the coefficient of the lagged dependent variable.