Forward Orthogonal Deviations GMM and the Absence of Large Sample Bias

Robert F. Phillips*
Department of Economics
George Washington University

December 2022

Abstract

It is well-known that generalized method of moments (GMM) estimators of dynamic panel data models can have asymptotic bias if the number of time periods (T) and the number of cross-sectional units (n) are both large (Alvarez and Arellano, 2003). This conclusion, however, follows when all available instruments are used. This paper provides results supporting a more optimistic conclusion when less than all available instrumental variables are used. If the number of instruments used per period increases with T sufficiently slowly, the bias of GMM estimators based on the forward orthogonal deviations transformation (FOD-GMM) disappears as T and n increase, regardless of the relative rate of increase in T and n. Monte Carlo evidence is provided that corroborates this claim. Moreover, a large-n, large-T distribution result is provided for FOD-GMM.

Keywords: Forward orthogonal deviations; first difference; generalized method of moments; dynamic panel data; bias

*2115 G Street, NW, Suite 340, Washington DC, 20052; phone: 202-994-8619; fax: 202-994-6147; email: `rphil@gwu.edu`

1 Introduction

When estimating a regression with panel data, researchers are often concerned about unobserved individual-specific effects, typically referred to as "fixed effects". The concern is that they may be correlated with the explanatory variables. The fixed effects can be eliminated by subtracting from each variable its within-group average. If the explanatory variables are strictly exogenous, then applying ordinary least squares after removing the within-group averages produces an estimator—the fixed effects estimator—whose distribution has no asymptotic bias as the number of cross-sectional units (n) increases and the number of time periods (T) is fixed, or both n and T increase, or T increases with n fixed.[1] Consequently, under suitable conditions (see, e.g., Hansen, 2007), regardless of whether n is large, or T is large, or both are large, inference based on the fixed effects estimator is reliable.

But this conclusion is predicated on the regressors being strictly exogenous, which is not true when the model is dynamic. For a dynamic model, although the fixed effects estimator's distribution is correctly centered when $\lim n/T = 0$, it has a bias term if $\lim n/T > 0$ (Alvarez and Arellano, 2003). Consequently, if a dynamic panel data model is estimated with the fixed effects estimator, confidence intervals and test statistics based on the estimator can be misleading when n is large compared to T.

As a result, researchers proposed alternative estimators for dynamic panel data regressions specifically for the case where T is fixed and $n \to \infty$. For example, maximum-likelihood and quasi maximum-likehood estimators of dynamic panel data regressions have been proposed for the large-n, small-T case; see, for example, Alvarez and Arellano (2004), Anderson and Hsiao (1981), Binder et al. (2005), Hsiao et al. (2002), Kruiniger (2013), Moral-Benito (2013), and Phillips (2010, 2018), among others. Moreover, some well-known generalized method of moments (GMM) estimators of dynamic panel data models are suitable when n is large compared to T; see, for example, Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998), and Holtz-Eakin et al. (1988). But a drawback of these estimators is that their large sample distributions may not be correctly centered when T is not small compared to n (Alvarez and Arellano, 2003).

In this paper I provide conditions that, if satisfied, indicate a well-known class of GMM estimators have no asymptotic bias, even when T is not small compared to n. Like the fixed effects estimator, the data transformation I examine removes fixed effects by subtracting a within-group average from each observation. But unlike the fixed effects estimator, the GMM estimators studied here use only future values for the within-group averages. Specifically, I examine GMM based on the forward orthogonal deviations (FOD) transformation (FOD-GMM).

In the panel data literature the FOD transformation dates back to at least Arellano and Bover (1995). Arellano and Bover (1995) showed that, under suitable

[1] In cases where T increases, dependence across time among variables must be restricted (Hansen, 2007).

conditions, FOD-GMM and GMM based on the first difference (FD) transformation (FD-GMM) are identical. Alvarez and Arellano (2003) exploited this equivalence to establish the asymptotic properties of a FD/FOD-GMM estimator of the autoregressive parameter in a first-order autoregressive (AR(1)) panel data model. In particular, they showed that, if all available instrumental variables are used, then the FOD-GMM estimator is asymptotically biased if $\lim T/n = c$, with $0 < c < \infty$.

But this conclusion depends on the use of all available instrumental variables. Bun and Kiviet (2006) and Hsiao and Zhou (2017) provided results indicating that, if a fixed number of instruments are used each period, FOD-GMM estimators do not have large sample bias, regardless of the relative rate of increase in T and n. To obtain this conclusion Bun and Kiviet (2006) and Hsiao and Zhou (2017) studied specific models and instrument choices.

The first contribution of this paper is that it generalizes Bun and Kiviet's and Hsiao and Zhou's work. I provide results indicating FOD-GMM estimators do not have large sample bias—regardless of the relative rate of increase in T and n—for a much larger class of regression problems and instrument choices than that examined by Bun and Kiviet (2006) and Hsiao and Zhou (2017). Moreover, a new result provided in this paper indicates that the absence of large sample bias of FOD-GMM estimators depends, not on using a fixed number of instrumental variables per period, but rather on how fast the maximum number of instrumental variables used in a period increases with T. Another contribution provided in this paper is an asymptotic distribution result for FOD-GMM estimators.

Section 3 provides Monte Carlo evidence that corroborates the analytical results in Section 2. The Monte Carlo experiments illustrate that the bias of an FOD-GMM estimator, which uses current and recent lags of predetermined variables as instrumental variables, decreases as n, or T, or both increase. As predicted by the analytical results, the bias tends to be least when both n and T are large. As a result, confidence intervals are reliable when n and T are large, and given the close connection between confidence intervals and t-statistics, we can conclude that t-tests will also be reliable when n and T are large.

For the sake of comparison, I also provide Monte Carlo results for an FD-GMM estimator that exploits the same instrumental variables used by the FOD-GMM estimator. Although the FD- and FOD-GMM estimators are the same for the estimation problem Alvarez and Arellano (2003) studied, they are not the same when only current and recent lags of predetermined variables are used as instruments (see, e.g., Phillips, 2019). Consequently, we should not expect their sampling performance to be similar. Indeed, the bias of the FD-GMM estimator was usually greater than that of the FOD-GMM estimator, and consequently confidence intervals based on the FD-GMM estimator were, as expected, much less reliable. Moreover, the FD-GMM estimator was generally less efficient, though no theoretical explanation for this result is apparent.

2 Large sample results

2.1 The model and estimator

The regression model studied in this paper is

$$y_{i,t} = \boldsymbol{x}'_{i,t}\boldsymbol{\beta} + \eta_i + v_{i,t} \qquad (t = 1, \ldots, T,\ i = 1, \ldots, n),$$

where $\boldsymbol{x}'_{i,t} := (x_{i,t,1}, \ldots, x_{i,t,K})$ and $\boldsymbol{\beta}' := (\beta_1, \ldots, \beta_K)$ are vectors of regressors and parameters. Some or all of the $x_{i,t,k}$s may be lagged values of $y_{i,t}$. The term η_i is an unobserved individual-specific or fixed effect, and $v_{i,t}$ is an error term.

This paper examines estimation of $\boldsymbol{\beta}$ with FOD-GMM. The FOD transformation removes from each variable its within-group average over future periods. For example, the FOD transformed explanatory variables for the tth period are given by $\ddot{\boldsymbol{X}}_t := c_t\left(\boldsymbol{X}_t - \overline{\boldsymbol{X}}_t\right)$, where $\boldsymbol{X}'_t := (\boldsymbol{x}_{1,t}, \ldots, \boldsymbol{x}_{n,t})$, $\overline{\boldsymbol{X}}_t := (T-t)^{-1}\sum_{s=1}^{T-t}\boldsymbol{X}_{t+s}$, and $c_t^2 := (T-t)/(T-t+1)$. Moreover, the transformed values of the dependent variable in the tth period are $\ddot{\boldsymbol{y}}_t := c_t\left(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_t\right)$, where $\boldsymbol{y}'_t := (y_{1,t}, \ldots, y_{n,t})$ and $\overline{\boldsymbol{y}}_t := (T-t)^{-1}\sum_{s=1}^{T-t}\boldsymbol{y}_{t+s}$. The constant c_t ensures the transformed errors (the $\ddot{v}_{i,t}$s) are conditionally homoskedastic and uncorrelated if the original errors (the $v_{i,t}$s) are conditionally homoskedastic and uncorrelated (Arellano, 2003, p. 17).

In addition to FOD transformed dependent and explanatory variables, an FOD-GMM estimator relies on instrumental variables. Let $\boldsymbol{z}_{i,t}$ denote a $q_t \times 1$ vector of instrumental variables for the ith individual in the tth period, and let $\boldsymbol{Z}'_t := (\boldsymbol{z}_{1,t}, \ldots, \boldsymbol{z}_{n,t})$. The tth period projection matrix based on these instruments is $\boldsymbol{P}_t := \boldsymbol{Z}_t\left(\boldsymbol{Z}'_t\boldsymbol{Z}_t\right)^{-1}\boldsymbol{Z}'_t$. Using these projections and the transformed regression variables, the FOD-GMM estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} := \left(\sum_{t=1}^{T-1}\ddot{\boldsymbol{X}}'_t\boldsymbol{P}_t\ddot{\boldsymbol{X}}_t\right)^{-1}\sum_{t=1}^{T-1}\ddot{\boldsymbol{X}}'_t\boldsymbol{P}_t\ddot{\boldsymbol{y}}_t$$

(Arellano, 2003, p. 154).

2.2 When there is no asymptotic bias

The reliability of large sample tests and confidence intervals based on an FOD-GMM estimator depends on whether or not $\sqrt{nT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically unbiased. To see when the distribution of $\sqrt{nT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is correctly centered at a vector of zeros, note that

$$\sqrt{nT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - \boldsymbol{A}_{n,T}^{-1}E\left(\boldsymbol{b}_{n,T}\right) = \boldsymbol{A}_{n,T}^{-1}\left[\boldsymbol{b}_{n,T} - E\left(\boldsymbol{b}_{n,T}\right)\right],$$

where $\boldsymbol{b}_{n,T} := (nT)^{-1/2}\sum_{t=1}^{T-1}\ddot{\boldsymbol{X}}'_t\boldsymbol{P}_t\ddot{\boldsymbol{v}}_t$, $\boldsymbol{A}_{n,T} := (nT)^{-1}\sum_{t=1}^{T-1}\ddot{\boldsymbol{X}}'_t\boldsymbol{P}_t\ddot{\boldsymbol{X}}_t$, $\ddot{\boldsymbol{v}}_t := c_t\left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t\right)$, $\boldsymbol{v}'_t := (v_{1,t}, \ldots, v_{n,t})$, and $\overline{\boldsymbol{v}}_t := (T-t)^{-1}\sum_{s=1}^{T-t}\boldsymbol{v}_{t+s}$. If, as the sample size increases, the mean of $\boldsymbol{b}_{n,T}$ does not converge to a vector of zeros, then we cannot expect $\sqrt{nT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ to be asymptotically unbiased. This point is illustrated with the following example.

The AR(1) panel data model provides a comparatively simple but instructive case. The AR(1) panel data model is

$$y_{i,t} = \beta y_{i,t-1} + \eta_i + v_{i,t},$$

with $|\beta| < 1$. For this case, $\sqrt{nT}(\widehat{\beta}-\beta) - E(b_{n,T})/a_{n,T} = [b_{n,T} - E(b_{n,T})]/a_{n,T}$, where $a_{n,T} = (nT)^{-1}\sum_{t=1}^{T-1} \ddot{\boldsymbol{y}}'_{t-1}\boldsymbol{P}_t\ddot{\boldsymbol{y}}_{t-1}$ and $b_{n,T} = (nT)^{-1/2}\sum_{t=1}^{T-1} \ddot{\boldsymbol{y}}'_{t-1}\boldsymbol{P}_t\ddot{\boldsymbol{v}}_t$, with $\ddot{\boldsymbol{y}}_{t-1} := c_t(\boldsymbol{y}_{t-1} - \overline{\boldsymbol{y}}_{t-1})$, $\boldsymbol{y}'_{t-1} := (y_{1,t-1},\ldots,y_{n,t-1})$, and $\overline{\boldsymbol{y}}_{t-1} := (T-t)^{-1}\sum_{s=1}^{T-t}\boldsymbol{y}_{t-1+s}$. Alvarez and Arellano (2003) provide conditions that imply

$$a_{n,T} \xrightarrow{p} \frac{\sigma^2}{1-\beta^2} \qquad \text{and} \qquad E(b_{n,T}) - \sqrt{\frac{T}{n}}\left(\frac{\sigma^2}{\beta-1}\right) \to 0 \qquad (n, T \to \infty),$$

where $\sigma^2 := \mathrm{var}(v_{i,t})$ (see Alvarez and Arellano, 2003, p. 1128, Lemma 2). Using these results, they showed that, if $\lim(T/n) = c$, with $0 \le c < \infty$, then

$$\sqrt{nT}\left(\widehat{\beta}-\beta\right) - \left[-\sqrt{T/n}\,(1+\beta)\right] \xrightarrow{d} N(0,\, 1-\beta^2).$$

(see Alvarez and Arellano, p. 1129, Theorem 2). Therefore, for the AR(1) panel data model, $\sqrt{nT}(\widehat{\beta}-\beta)$ has asymptotic bias of $-\sqrt{c}(1+\beta)$, which is zero only when $c = 0$. In other words, $\sqrt{nT}(\widehat{\beta}-\beta)$ is asymptotically unbiased only if $E(b_{n,T}) \to 0$, which only occurs when $T/n \to 0$.

This example is instructive for two reasons. First, it illustrates there will be asymptotic bias if we do not have $E(b_{n,T}) \to 0$. Hence, $E(b_{n,T}) \to 0$ is a critical requirement for the absence of asymptotic bias. The second reason the example is instructive is less obvious and is the point of this paper. The failure to get $E(b_{n,T}) \to 0$ when $c > 0$ is due the number of instrumental variables that are used. Alvarez and Arellano (2003) assumed all available moment restrictions are exploited. For the estimation problem they considered, using all available instrumental variables implies that the maximum number of instrumental variables used per period increases at the rate T increases. Theorem 1 shows that this matters, because, according to the theorem, if the maximum number of instrumental variables used each period increases at a rate no faster than T^δ, with $\delta < 1/2$, then $E(\boldsymbol{b}_{n,T}) \to \boldsymbol{0}$ regardless of how T and n increase.

In order to state sufficient conditions for this conclusion some additional definitions will be used. Specifically, let $\boldsymbol{w}_{i,t}$ be a column vector consisting of all of the distinct entries in $\left\{\left(\boldsymbol{x}'_{i,s}, \boldsymbol{z}'_{i,s}\right);\ s = 1,\ldots,t\right\}$. That is, $\boldsymbol{w}_{i,t}$ contains all the distinct values of the explanatory variables and instrumental variables for the ith individual from the first period up to the tth period. Next, set $\boldsymbol{u}'_i := (\boldsymbol{w}'_{i,T}, v_{i,1},\ldots,v_{i,T},\eta_i)$. Finally, let $\gamma_{k,t,s} := \text{cov}\left(v_{1,t}, x_{1,t+s,k} | \boldsymbol{w}_{1,t}\right)$, and let $b_{n,T,k}$ denote the kth entry in $\boldsymbol{b}_{n,T}$ ($k = 1, \ldots, K$).

Theorem 1 can now be stated.

Theorem 1. *Assume $T \geq 3$ and let $q_T^* := \max_{1 \leq t \leq T-1} q_t$. Further assume*

A1: the $\boldsymbol{u}_i$s are independent and identically distributed (i.i.d.) across i;

A2: rank$(\boldsymbol{Z}_t) = q_t$ with probability 1;

A3: $E(v_{1,t} | \boldsymbol{w}_{1,t}) = 0$; and

A4: $\left|\sum_{s=1}^{S} \gamma_{k,t,s}\right| \leq M$ with probability 1, for some finite M and all $t \geq 1$, $S \geq 1$, and $k = 1, \ldots, K$.

Then, if $q_T^* = O(T^{\delta})$, it follows that

$$E\left(b_{n,T,k}\right) = O\left(n^{-1/2} T^{\delta - 1/2} \ln T\right) \qquad (k = 1, \ldots, K).$$

Proofs are provided in Appendix A.

Theorem 1 shows how $E(\boldsymbol{b}_{n,T})$ depends on the maximum number of instrumental variables used per period. Specifically, if the maximum number of instrumental variables used per period increases with T at a rate no greater than T^{δ}, with $\delta < 1/2$, the conclusion of Theorem 1 implies $E(\boldsymbol{b}_{n,T}) \to \boldsymbol{0}$ for all sequences with n increasing, or T increasing, or both increasing.

Theorem 1 relies on several simplifying assumptions. The first three assumptions are straightforward. The fourth characterizes the linear association between the error in period t and the kth regressor s periods in the future relative to that period, conditional on the available information at time t. Because the conditional covariance between $v_{1,t}$ and $x_{1,t+s,k}$ is not restricted to be zero, Assumption A4 allows for predetermined regressors. Moreover, if the error term and future values of the explanatory variables are suitably weakly dependent, the bound on the sum of covariances in A4 will be satisfied. For example, Assumption A4 is satisfied by stationary Kth-order autoregressive panel data models. This fact is verified in Appendix B of the Supplemental Material (Phillips, 2022). That appendix also provides another dynamic panel data model that satisfies Assumption A4; a model that includes the model considered by Bun and Kiviet (2006) as a special case.

Bun and Kiviet (2006) and Hsiao and Zhou (2017) provide results that are related to Theorem 1. But neither Bun and Kiviet (2006) nor Hsiao and Zhou (2017) examined the possibility that asymptotic bias can vanish as the number of instrumental variables used per period increases, provided the maximum number of per period instrumental variables increases slowly enough. Moreover, the models and instrument choices these earlier papers studied satisfy the assumptions in Theorem 1 (see Appendix B of the Supplemental Material, Phillips, 2022), and, therefore, the estimation problems examined in Bun and Kiviet (2006) and Hsiao and Zhou (2017) are special cases covered by Theorem 1.

The conclusion of Theorem 1 implies $E(\boldsymbol{b}_{n,T}) \to \mathbf{0}$, for all sequences for which n, or T, or both increase, if the maximum number of instruments used per period does not increase with T too rapidly. For some estimation problems this restriction on the number of instrumental variables rules out using all available instrumental variables. However, in practice researchers often use no more than a fixed number of instrumental variables per period. In which case, if Assumptions A1 through A4 are satisfied, then $E(\boldsymbol{b}_{n,T}) \to \mathbf{0}$ for all sequences for which n, or T, or both increase.

The next theorem generalizes results in Hsiao and Zhou (2017). Hsiao and Zhou (2017) used sequential-limit analysis to obtain asymptotic distributions for FOD-GMM estimators of the AR(1) panel data model when one instrumental variable is used per period. Specifically, they considered limits as $n \to \infty$, and then $T \to \infty$.[2] Theorem 2 generalizes this approach to a much broader class of regressions and instrument choices.

Theorem 2 relies on additional assumptions. In order to state the first of these assumptions, let $\boldsymbol{w}^{*\prime}_{1,t} = (\boldsymbol{w}'_{1,t}, v_{1,0}, \ldots, v_{1,t-1})$ $(t \geq 1)$. Assumption A3* strengthens Assumption A3:

A3*: $E(v_{1,t}|\boldsymbol{w}^*_{1,t}) = 0$ $(t \geq 1)$.

Also assume

A5: $E\left(v^2_{1,t}|\boldsymbol{w}^*_{1,t}\right) = \sigma^2$ $(t \geq 1)$;

and

A6: the entries in $\boldsymbol{x}_{1,t}$ and $\boldsymbol{z}_{1,t}$ have finite second moments.

Assumption A6 implies the entries in the matrices $\boldsymbol{R}_t := E(\boldsymbol{z}_{1,t}\ddot{\boldsymbol{x}}'_{1,t})$ and $\boldsymbol{Q}_t := E(\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t})$ are finite. Moreover, Assumption A2 implies $\boldsymbol{Q}_t$ is positive definite. Hence, if Assumptions A2 and A6 are satisfied, we can define $\boldsymbol{\Omega}_t := \boldsymbol{R}'_t\boldsymbol{Q}^{-1}_t\boldsymbol{R}_t$. Also, let $\overline{\boldsymbol{\Omega}}_T := T^{-1}\sum_{t=1}^{T-1}\boldsymbol{\Omega}_t$. Finally, let $\omega_{t,j,k}$ denote the (j,k)th entry of $\boldsymbol{\Omega}_t$, and set $\nu_t := \max_{j,k}|\omega_{t,j,k}|$. Then the last set of assumptions for Theorem 2 are

A7: $\text{rank}(\boldsymbol{R}_t) = K$ $(t = 1, \ldots, T-1)$;

A8: $\overline{\boldsymbol{\Omega}} := \lim_{T\to\infty}\overline{\boldsymbol{\Omega}}_T$ exists, with $\overline{\boldsymbol{\Omega}}$ a positive definite matrix;
and

A9: $\lim_{T\to\infty} T^{-3/2}\sum_{t=1}^{T-1}\nu_t^{3/2} = 0$.

Theorem 2 can now be stated.

[2] Taking limits sequentially simplifies the analysis. But the simplification is only useful if it provides reliable approximations for large samples. For some sequences of n and T increasing, this will not be the case when the presence of asymptotic bias depends on what happens to T/n as n and T jointly increase. However, the conclusion of Theorem 1 indicates that, under the conditions described in this paper, bias disappears, as n and T increase, regardless of what happens to T/n.

Theorem 2. *Assume $q_T^* = O(T^\delta)$, with $0 \leq \delta < 1/2$, and assume A1, A2, A3*, and A4 through A9 are satisfied. Then*

$$\sqrt{nT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2 \overline{\boldsymbol{\Omega}}^{-1}\right) \qquad (n, T \to \infty)_{seq}.$$

The notation "$(n, T \to \infty)_{\text{seq}}$" means $n \to \infty$, then $T \to \infty$.

3 Monte Carlo simulations

This section provides results from Monte Carlo simulations. Using generated samples, FOD-GMM estimator biases, standard deviations, and confidence intervals were estimated for different sampling designs. For the sake of comparison, the bias, standard deviation, and confidence intervals for a FD-GMM estimator were also calculated.

3.1 Samples

Monte Carlo samples were generated using sampling schemes similar to those described in Bun and Kiviet (2006). The dependent variable in the regression model was always generated according to

$$y_{i,t} = \beta_1 y_{i,t-1} + \beta_2 x_{i,t} + \eta_i + v_{i,t} \qquad (t = -49, \ldots, -1, 0, 1, \ldots, T,\ i = 1, \ldots, n),$$

with $\beta_1 \in \{0.25, 0.75\}$ and $\beta_2 = 1 - \beta_1$. The start-up value $y_{i,-50}$ was set to zero. The error components $v_{i,t}$ and η_i were generated independently of one another as i.i.d. standard normal variates. Moreover, two schemes were used to generate the explanatory variable $x_{i,t}$. The two schemes—Scheme 1 and Scheme 2—were similar to those used by Bun and Kiviet (2006).

For Scheme 1, the time series for $x_{i,t}$ was constructed according to

$$\begin{aligned} x_{i,-50} &= \kappa_1 \eta_i + \varepsilon_{i,-50}, \\ x_{i,-49} &= \kappa_1 \eta_i + \frac{1}{1-\rho L}\left(\varepsilon_{i,-49} + \phi_1 v_{i,-50}\right) \\ x_{i,t} &= \kappa_1 \eta_i + \frac{1}{1-\rho L}\varepsilon_{i,t} + \phi_1 v_{i,t-1} \qquad (t = -48, \ldots, -1, 0, +1, \ldots, T), \end{aligned}$$

with $\rho \in \{0.50, 0.95\}$, $\kappa_1 \in \{-1, 0, +1\}$, and $\phi_1 \in \{-1, 0, +1\}$. Moreover, the $\varepsilon_{i,t}$s were generated as $\varepsilon_{i,t} \overset{i.i.d.}{\sim} U(-\sqrt{12}/2, \sqrt{12}/2)$ independently of the $v_{i,t}$s and η_is.

As for Scheme 2, $x_{i,t}$ was generated as

$$\begin{aligned} x_{i,-50} &= \kappa_2 \eta_i + \varepsilon_{i,-50}, \\ x_{i,t} &= \rho x_{i,t-1} + \phi_2 y_{i,t-1} + \kappa_2 \eta_i + \varepsilon_{i,t} \qquad (t = -49, \ldots, -1, 0, +1, \ldots, T). \end{aligned}$$

Table 1: Designs 1 through 18 ($\beta_1 = 0.25$).

	Designs								
	1	2	3	4	5	6	7	8	9
ρ	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
ϕ_1	−1.00	−1.00	−1.00	0.00	0.00	0.00	1.00	1.00	1.00
κ_1	−1.00	0.00	1.00	−1.00	0.00	1.00	−1.00	0.00	1.00
	Designs								
	10	11	12	13	14	15	16	17	18
ρ	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ϕ_1	−1.00	−1.00	−1.00	0.00	0.00	0.00	1.00	1.00	1.00
κ_1	−1.00	0.00	1.00	−1.00	0.00	1.00	−1.00	0.00	1.00

Following Bun and Kiviet (2006), I set $\phi_2 = \phi_1\beta_2(1-\rho)/(1+\beta_2\phi_1)$ and $\kappa_2 = \kappa_1(1-\rho-\phi_2) - \phi_2/\beta_2$.

For both Scheme 1 and Scheme 2, there were 36 experimental Designs, where a Design is a combination of parameter values. Table 1 lists Designs 1 through 18. For these designs I set $\beta_1 = 0.25$. Designs 19 through 36 were identical to Designs 1 through 18 except that, for Designs 19 through 36, β_1 was set to 0.75.

The effect of how each series was initialized was eliminated by dropping the first 50 values of each generated time series. Therefore, estimation was based on the values $(x_{i,0}, y_{i,0}), (x_{i,1}, y_{i,1}), \ldots, (x_{i,T}, y_{i,T})$ $(i = 1, \ldots, n)$.

Finally, four combinations of (T, n) were considered: small-T, small-n $(25, 25)$; small-T, large-n $(25, 100)$; large-T, small-n $(100, 25)$; and large-T, large-n $(100, 100)$. These sample sizes were selected to see how the bias of the FOD-GMM estimator changed with sample size. The large-T, small-n sample size is not usually considered in the panel data literature, but the theory in Section 2 predicts that the bias of the FOD-GMM estimator should be less for $(T, n) = (100, 25)$ than for $(T, n) = (25, 25)$. Therefore, experiments were conducted to investigate this prediction.

3.2 Estimators

For each Design and sample size, 5,000 samples were generated.[3] After a sample was generated, an FOD-GMM estimate was calculated using the instrumental vari-

[3] The data were generated using GAUSS. Moreover, all calculations were performed with GAUSS.

ables $\boldsymbol{z}_{i,1}' = (y_{i,0}, x_{i,0}, x_{i,1})$ and $\boldsymbol{z}_{i,t}' = (y_{i,t-2}, y_{i,t-1}, x_{i,t-2}, x_{i,t-1}, x_{i,t})$ $(t = 2, \ldots, T-1$, $i = 1, \ldots, n)$. Then bias and standard deviation estimates were calculated. I also calculated standard errors by taking the square roots of the diagonal entries of $\widehat{\sigma}^2 \left(\sum_{t=1}^{T-1} \ddot{\boldsymbol{X}}_t' \boldsymbol{P}_t \ddot{\boldsymbol{X}}_t\right)^{-1}$, where $\widehat{\sigma}^2 := \sum_{t=1}^{T-1}(\ddot{\boldsymbol{y}}_t - \ddot{\boldsymbol{X}}_t\widehat{\boldsymbol{\beta}})'(\ddot{\boldsymbol{y}}_t - \ddot{\boldsymbol{X}}_t\widehat{\boldsymbol{\beta}})/[n(T-1)]$. Using the standard errors and parameter estimates, 90- and 95-percent confidence intervals were calculated for β_1 and β_2.

For the sake of comparison, estimates of a FD-GMM estimator's bias, standard deviation, and confidence intervals were also calculated for β_1 and β_2. For the FD-GMM estimator, the transformed data are $\tilde{\boldsymbol{y}}' := (\tilde{\boldsymbol{y}}_1', \ldots, \tilde{\boldsymbol{y}}_n')$, where $\tilde{\boldsymbol{y}}_i' := (y_{i,2} - y_{i,1}, \ldots, y_{i,T} - y_{i,T-1})$; and $\tilde{\boldsymbol{X}}' := \left(\tilde{\boldsymbol{X}}_1' \ldots \tilde{\boldsymbol{X}}_n'\right)$, where the tth row of matrix $\tilde{\boldsymbol{X}}_i$ is $\tilde{\boldsymbol{x}}_{i,t+1}' := \boldsymbol{x}_{i,t+1}' - \boldsymbol{x}_{i,t}'$ $(t = 1, \ldots, T-1)$. Moreover, I used the same instrumental variables for the FD-GMM estimator that I used for the FOD-GMM estimator. Specifically, for the instrument matrix, I set $\tilde{\boldsymbol{Z}}' := \left(\tilde{\boldsymbol{Z}}_1' \ldots, \tilde{\boldsymbol{Z}}_n'\right)$, where $\tilde{\boldsymbol{Z}}_i$ is a block-diagonal matrix, with $\boldsymbol{z}_{i,t}'$ in the tth-diagonal block $(t = 1, \ldots, T-1)$. Then, the one-step FD-GMM estimator is

$$\widetilde{\boldsymbol{\beta}} := \left[\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Z}}\left(\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i'\boldsymbol{G}\tilde{\boldsymbol{Z}}_i\right)^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{X}}\right]^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Z}}\left(\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i'\boldsymbol{G}\tilde{\boldsymbol{Z}}_i\right)^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{y}},$$

where $\boldsymbol{G}$ is a $(T-1) \times (T-1)$ matrix with twos running down the main diagonal, minus ones just above and below the main diagonal, and zeros everywhere else (Arellano and Bond, 1991). Standard errors for the entries in $\widetilde{\boldsymbol{\beta}}$ were calculated using the square roots of the diagonal entries in $\widetilde{\sigma}^2 \left[\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Z}}\left(\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i'\boldsymbol{G}\tilde{\boldsymbol{Z}}_i\right)^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{X}}\right]^{-1}$, where $\widetilde{\sigma}^2 := \sum_{i=1}^{n}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{\beta}})'(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\widetilde{\boldsymbol{\beta}})/[2n(T-1)]$.

3.3 Simulation results

Two schemes, each with 36 designs and four cases of (T, n), led to a total of 288 Monte Carlo experiments. This produced a large number of bias, standard deviation, and confidence interval estimates. But the main conclusions drawn from the experiments can be illustrated with only some of the simulated data.

For example, the bias estimate results were qualitatively similar for Schemes 1 and 2. Therefore, only Scheme 1 bias estimates are provided here.[4] Tables 2 and 3 provide Scheme 1 bias estimates for $\sqrt{nT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{nT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$, where $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}$ denote the FOD- and FD-GMM estimators. Table 2 provides bias estimates for Designs 1 through 18 $(\beta_1 = 0.25)$, and Table 3 provides estimates for Designs 19 through 36 $(\beta_1 = 0.75)$.

[4]Scheme 2 bias estimates are provided in Appendix C of the Supplemental Material (Phillips,

Table 2: Scheme 1: Bias estimates for Designs 1 through 18 ($\beta_1 = 0.25$).

Designs:	1	2	3	4	5	6	7	8	9
$(T,\, n)$				Bias of $\sqrt{nT}$(FOD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−0.01	0.00	0.02	−0.23	−0.26	−0.26	−0.25	−0.25	−0.28
(25, 100):	0.02	0.00	0.02	−0.11	−0.12	−0.16	−0.12	−0.14	−0.18
(100, 25):	0.01	−0.01	0.01	−0.10	−0.09	−0.11	−0.12	−0.11	−0.13
(100, 100):	−0.02	0.01	0.01	−0.04	−0.05	−0.06	−0.04	−0.05	−0.08
				Bias of $\sqrt{nT}$(FOD estimator of $\beta_2 - \beta_2$)					
(25, 25):	0.13	0.11	0.14	0.10	0.10	0.01	0.03	0.01	−0.02
(25, 100):	0.08	0.08	0.08	0.06	0.03	−0.02	0.01	0.01	−0.02
(100, 25):	0.05	0.05	0.05	0.07	0.05	0.02	0.02	0.01	−0.01
(100, 100):	0.03	0.05	0.04	0.03	−0.01	0.04	0.01	0.00	0.02
				Bias of $\sqrt{nT}$(FD estimator of $\beta_1 - \beta_1$)					
(25, 25):	0.11	0.10	0.12	−0.91	−0.93	−0.94	−0.95	−0.99	−1.00
(25, 100):	0.10	0.07	0.09	−0.46	−0.48	−0.51	−0.50	−0.54	−0.55
(100, 25):	0.27	0.25	0.25	−1.54	−1.53	−1.51	−1.66	−1.68	−1.67
(100, 100):	0.13	0.17	0.16	−0.80	−0.83	−0.81	−0.86	−0.88	−0.90
				Bias of $\sqrt{nT}$(FD estimator of $\beta_2 - \beta_2$)					
(25, 25):	0.64	0.55	0.57	0.15	0.06	−0.13	−0.02	−0.08	−0.18
(25, 100):	0.36	0.33	0.32	0.09	0.00	−0.10	−0.02	−0.05	−0.11
(100, 25):	1.07	1.05	1.02	0.16	0.13	0.05	−0.03	−0.05	−0.10
(100, 100):	0.57	0.59	0.56	0.06	0.00	0.04	−0.02	−0.05	−0.04

Table 2 is continued on the next page.

Table 2 continued.

Designs:	10	11	12	13	14	15	16	17	18
$(T,\ n)$				Bias of $\sqrt{nT}$(FOD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−0.01	−0.03	−0.02	−0.18	−0.22	−0.22	−0.22	−0.25	−0.24
(25, 100):	0.00	−0.00	0.01	−0.11	−0.10	−0.12	−0.13	−0.16	−0.16
(100, 25):	−0.02	−0.02	−0.03	−0.10	−0.11	−0.09	−0.10	−0.11	−0.14
(100, 100):	−0.01	−0.01	−0.03	−0.05	−0.04	−0.05	−0.04	−0.06	−0.06
				Bias of $\sqrt{nT}$(FOD estimator of $\beta_2 - \beta_2$)					
(25, 25):	0.11	0.12	0.12	0.14	0.17	0.13	0.03	0.06	0.04
(25, 100):	0.08	0.06	0.08	0.05	0.06	0.09	0.03	0.03	0.01
(100, 25):	0.04	0.03	0.04	0.10	0.09	0.08	0.06	0.07	0.11
(100, 100):	0.02	0.02	0.03	0.04	0.03	0.05	0.02	0.04	0.03
				Bias of $\sqrt{nT}$(FD estimator of $\beta_1 - \beta_1$)					
(25, 25):	0.09	0.08	0.06	−0.72	−0.76	−0.76	−0.79	−0.82	−0.82
(25, 100):	0.08	0.09	0.07	−0.39	−0.37	−0.39	−0.43	−0.46	−0.45
(100, 25):	0.08	0.04	0.03	−1.18	−1.20	−1.17	−1.29	−1.30	−1.33
(100, 100):	0.08	0.06	0.05	−0.63	−0.63	−0.63	−0.67	−0.71	−0.69
				Bias of $\sqrt{nT}$(FD estimator of $\beta_2 - \beta_2$)					
(25, 25):	0.62	0.61	0.56	0.40	0.31	0.16	0.01	−0.01	−0.09
(25, 100):	0.38	0.36	0.33	0.15	0.12	0.09	0.01	−0.01	−0.06
(100, 25):	0.84	0.79	0.80	0.61	0.57	0.53	0.22	0.20	0.23
(100, 100):	0.48	0.46	0.47	0.25	0.23	0.24	0.08	0.08	0.08

Note: Each bias estimate is based on 5,000 samples.

Table 3: Scheme 1: Bias estimates for Designs 19 through 36 ($\beta_1 = 0.75$).

Designs:	19	20	21	22	23	24	25	26	27
(T, n)				Bias of $\sqrt{nT}$(FOD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−1.11	−1.41	−1.50	−0.90	−1.11	−1.23	−0.75	−0.94	−1.06
(25, 100):	−0.78	−1.41	−1.83	−0.63	−0.87	−1.13	−0.45	−0.63	−0.86
(100, 25):	−0.43	−0.54	−0.60	−0.38	−0.48	−0.52	−0.30	−0.37	−0.45
(100, 100):	−0.29	−0.46	−0.62	−0.21	−0.34	−0.39	−0.16	−0.24	−0.34
				Bias of $\sqrt{nT}$(FOD estimator of $\beta_2 - \beta_2$)					
(25, 25):	−0.19	−0.42	−0.56	0.19	0.01	−0.20	0.16	0.17	0.09
(25, 100):	−0.16	−0.49	−0.81	0.16	0.01	−0.29	0.11	0.13	0.06
(100, 25):	−0.05	−0.10	−0.19	0.15	0.05	−0.04	0.12	0.09	0.07
(100, 100):	−0.02	−0.12	−0.26	0.09	0.02	−0.05	0.07	0.09	0.06
				Bias of $\sqrt{nT}$(FD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−3.80	−4.68	−5.30	−3.01	−3.34	−3.64	−2.73	−2.98	−3.16
(25, 100):	−2.30	−3.16	−4.17	−1.76	−2.00	−2.37	−1.59	−1.75	−1.91
(100, 25):	−4.31	−4.39	−4.55	−3.65	−3.73	−3.75	−3.08	−3.13	−3.13
(100, 100):	−2.50	−2.61	−2.78	−2.05	−2.11	−2.16	−1.72	−1.74	−1.78
				Bias of $\sqrt{nT}$(FD estimator of $\beta_2 - \beta_2$)					
(25, 25):	−1.03	−1.63	−2.12	0.06	−0.45	−1.01	0.28	0.09	−0.16
(25, 100):	−0.70	−1.19	−1.82	0.01	−0.30	−0.76	0.16	0.05	−0.11
(100, 25):	−1.05	−1.14	−1.30	0.07	−0.14	−0.27	0.49	0.39	0.32
(100, 100):	−0.66	−0.76	−0.90	−0.03	−0.13	−0.20	0.25	0.23	0.17

Table 3 is continued on the next page.

Table 3 continued.

Designs:	28	29	30	31	32	33	34	35	36
(T, n)				Bias of $\sqrt{nT}$(FOD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−0.47	−0.48	−0.50	−0.69	−0.71	−0.66	−0.66	−0.71	−0.70
(25, 100):	−0.32	−0.33	−0.33	−0.45	−0.48	−0.47	−0.45	−0.47	−0.48
(100, 25):	−0.16	−0.18	−0.17	−0.25	−0.27	−0.25	−0.25	−0.29	−0.26
(100, 100):	−0.10	−0.10	−0.11	−0.15	−0.15	−0.15	−0.16	−0.15	−0.17
				Bias of $\sqrt{nT}$(FOD estimator of $\beta_2 - \beta_2$)					
(25, 25):	0.13	0.10	0.09	0.29	0.26	0.20	0.16	0.15	0.12
(25, 100):	0.03	0.03	0.01	0.16	0.15	0.06	0.09	0.06	0.02
(100, 25):	0.11	0.09	0.08	0.18	0.19	0.16	0.16	0.18	0.15
(100, 100):	0.06	0.04	0.05	0.12	0.10	0.09	0.09	0.09	0.08
				Bias of $\sqrt{nT}$(FD estimator of $\beta_1 - \beta_1$)					
(25, 25):	−1.65	−1.73	−1.71	−2.02	−2.01	−1.92	−1.92	−1.94	−1.95
(25, 100):	−1.00	−1.12	−1.14	−1.09	−1.14	−1.19	−1.08	−1.08	−1.08
(100, 25):	−1.62	−1.63	−1.62	−1.98	−2.00	−1.97	−1.96	−1.99	−1.95
(100, 100):	−0.96	−0.94	−0.99	−1.10	−1.11	−1.11	−1.08	−1.08	−1.07
				Bias of $\sqrt{nT}$(FD estimator of $\beta_2 - \beta_2$)					
(25, 25):	−0.07	−0.21	−0.26	0.22	−0.05	−0.26	0.17	0.04	−0.14
(25, 100):	−0.17	−0.28	−0.33	0.08	−0.07	−0.34	0.08	−0.02	−0.13
(100, 25):	0.34	0.30	0.26	0.76	0.71	0.60	0.64	0.61	0.54
(100, 100):	0.10	0.08	0.05	0.35	0.28	0.24	0.32	0.27	0.27

Note: Each bias estimate is based on 5,000 samples.

According to the results in Section 2, the FOD-GMM estimator's bias is predicted to disappear as both T and n increase. This claim is confirmed by the data in Tables 2 and 3. For some designs the absolute bias estimates for $\sqrt{nT}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ are small for all sample sizes. On the other hand, for other designs the bias estimates are substantial when $(T, n) = (25, 25)$ but much smaller when $(T, n) = (100, 100)$. Moreover, T and n need not both be large in order to get a reduction in absolute bias. The analytical results in Section 2 predicts that the bias becomes less as either T or n increase, and the simulation data confirm this prediction. The absolute bias of the FOD-GMM estimator is usually smaller for both the $(T, n) = (25, 100)$ and the $(T, n) = (100, 25)$ sample sizes than for the $(T, n) = (25, 25)$ sample size. Bias is usually smallest, however, when both T and n are large.

Table 4 shows the effect of bias on the reliability of confidence intervals. The table provides the coverage of estimated 95-percent confidence intervals for Scheme 1.[5] Both 95-percent and 90-percent confidence intervals were calculated, but, for the sake of brevity, only the coverage of the estimated 95-percent confidence intervals is provided. I report results for only 95-percent confidence intervals because they are representative. In particular, when the coverage of a 95-percent confidence interval approximated 95 percent, the coverage of the 90-percent confidence interval likewise approximated 90 percent. On the other hand, if the coverage of a 95-percent confidence interval did not approximate 95 percent well, the 90-percent confidence interval did not approximate 90 percent. Additionally, again for the sake of brevity, I report the coverage of 95-percent confidence intervals only for the $(T, n) = (25, 25)$ and $(T, n) = (100, 100)$ sample sizes, for these are arguably the extreme cases in terms of the FOD-GMM estimator's bias. Therefore, these sample sizes should provide the extremes in terms of poor and good approximate coverage of the FOD confidence intervals. Specifically, the coverage should be best for the large-T, large-n cases and poorest for the small-T, small-n cases.

This prediction is confirmed by the data in Table 4. For Designs 1 through 18 the estimated coverage of the FOD 95-percent confidence intervals approximates 95 percent in most cases. This is true for both the $(T, n) = (25, 25)$ and $(T, n) = (100, 100)$ sample sizes. However, when T and n are both small, the approximation deteriorates for Designs 19 through 36, especially for β_1, which is 0.75 for those designs. On the other hand, even when $\beta_1 = 0.75$, the coverage of the FOD 95-percent confidence intervals is usually close to 95 percent when T and n are both large.

Finally, the standard deviations of the FD-GMM and FOD-GMM estimators were also estimated. Rather than report all of the standard deviation estimates, the main conclusion drawn from calculating them can be summarized briefly: not only did the FOD transformation generally produce less biased estimates, it led to more ef-

2022).

[5]For Scheme 2, qualitatively similar results are provided in Appendix C of the Supplemental Material (Phillips, 2022).

Table 4: Scheme 1: Coverage of FD and FOD 95-percent confidence intervals.

(T, n)	Designs: 1	2	3	4	5	6	7	8	9
	Percentage of FOD intervals covering β_1								
$(25, 25)$:	95.1	94.9	95.3	93.8	93.8	94.6	94.4	94.5	93.9
$(100, 100)$:	95.1	95.4	95.2	95.1	95.3	94.6	94.7	94.7	95.2
	Percentage of FOD intervals covering β_2								
$(25, 25)$:	94.5	94.7	94.5	94.2	94.8	95.2	94.9	95.1	94.6
$(100, 100)$:	94.5	94.8	95.3	95.0	95.3	95.4	94.6	94.9	95.0
	Percentage of FD intervals covering β_1								
$(25, 25)$:	94.6	94.3	94.1	84.0	83.7	84.2	81.9	80.5	79.6
$(100, 100)$:	94.6	95.3	94.9	87.4	87.6	86.0	84.4	83.5	83.6
	Percentage of FD intervals covering β_2								
$(25, 25)$:	90.4	91.6	91.2	93.9	94.4	93.9	94.3	94.7	93.4
$(100, 100)$:	92.0	91.9	92.4	94.3	94.7	95.4	94.7	94.6	94.7

(T, n)	Designs: 10	11	12	13	14	15	16	17	18
	Percentage of FOD intervals covering β_1								
$(25, 25)$:	95.0	95.1	95.0	93.8	94.2	94.6	94.2	94.4	94.3
$(100, 100)$:	94.8	95.5	95.1	95.1	95.3	95.5	94.8	94.7	95.0
	Percentage of FOD intervals covering β_2								
$(25, 25)$:	94.1	95.0	94.8	94.4	95.1	94.7	94.7	95.2	95.4
$(100, 100)$:	95.0	95.1	95.3	94.7	94.6	95.1	94.9	94.9	94.7
	Percentage of FD intervals covering β_1								
$(25, 25)$:	94.4	94.4	94.8	87.3	86.7	87.2	85.5	85.1	84.3
$(100, 100)$:	94.6	95.5	95.0	88.8	89.7	89.7	88.1	87.1	88.5
	Percentage of FD intervals covering β_2								
$(25, 25)$:	91.0	90.2	91.2	92.9	94.1	94.2	94.1	94.3	94.2
$(100, 100)$:	92.2	92.5	92.1	94.3	94.3	94.9	94.8	94.9	94.8

Table 4 is continued on the next page.

Table 4 continued.

(T,n)	Designs: 19	20	21	22	23	24	25	26	27
	Percentage of FOD intervals covering β_1								
$(25,25)$:	88.1	88.1	87.9	87.5	85.5	85.2	88.6	85.7	85.2
$(100,100)$:	94.7	94.3	93.2	94.6	94.1	94.2	95.1	93.7	94.3
	Percentage of FOD intervals covering β_2								
$(25,25)$:	93.7	92.9	92.3	94.5	94.5	93.8	95.0	94.8	95.2
$(100,100)$:	95.0	94.4	94.3	94.8	95.4	94.3	95.1	94.2	95.2
	Percentage of FD intervals covering β_1								
$(25,25)$:	50.8	43.7	38.7	41.8	38.1	33.9	44.5	38.4	35.3
$(100,100)$:	72.3	71.1	69.6	64.8	63.7	63.2	67.4	66.5	65.6
	Percentage of FD intervals covering β_2								
$(25,25)$:	81.3	70.3	63.0	93.6	90.7	82.1	93.2	94.4	92.3
$(100,100)$:	90.9	89.2	88.0	94.6	94.5	94.2	94.2	93.2	94.4

(T,n)	Designs: 28	29	30	31	32	33	34	35	36
	Percentage of FOD intervals covering β_1								
$(25,25)$:	91.5	92.0	91.8	88.6	88.2	89.3	89.1	87.9	88.6
$(100,100)$:	94.3	95.1	94.2	94.9	94.5	94.6	94.9	94.8	94.7
	Percentage of FOD intervals covering β_2								
$(25,25)$:	95.1	94.4	94.5	93.4	93.6	93.6	94.4	94.2	94.4
$(100,100)$:	94.8	94.6	94.9	94.4	94.6	94.4	94.3	94.6	95.0
	Percentage of FD intervals covering β_1								
$(25,25)$:	75.2	74.7	75.8	58.5	59.1	60.9	59.3	56.2	56.1
$(100,100)$:	88.4	88.6	87.4	81.8	81.5	80.4	78.6	78.6	78.6
	Percentage of FD intervals covering β_2								
$(25,25)$:	90.7	91.8	91.2	90.0	89.8	89.3	92.0	92.0	91.6
$(100,100)$:	94.9	94.7	94.8	93.7	94.1	94.4	92.6	93.1	93.2

Note: Each coverage estimate is based on 5,000 samples.

ficient estimates. As already noted, there were a total of 288 experiments, and for each experiment the standard deviations of the FOD-GMM estimators of β_1 and β_2 were estimated, for a total of 576 FOD standard deviation estimates. Similarly, 576 standard deviation estimates were obtained for the FD-GMM estimators of β_1 and β_2. When the standard deviation estimates of the FD- and FOD-GMM estimators were compared, only nine FD standard deviation estimates, out of a total of 576, were smaller than the corresponding FOD standard deviation estimates. In other words, FOD-GMM was rarely less efficient than FD-GMM. Moreover, for those cases for which the FOD-GMM estimator was less efficient, the maximum reduction in standard deviation from using the FD-GMM estimator rather than the FOD-GMM estimator was only about seven percent. On the other hand, for over 98 percent of the experiments the FOD-GMM estimator was more efficient, and for those experiments the maximum reduction in standard deviation from using the FOD-GMM estimator rather than the FD-GMM estimator was about 47 percent. The median reduction was approximately 19 percent.

However, this dominance of the FOD-GMM estimator in terms of efficiency may be specific to the experiments considered here. Unlike the reductions in bias from using the FOD transformation, I know of no theoretical explanation for why FOD-GMM should be superior to FD-GMM in terms of efficiency. For the experiments considered in this section, neither the FD-GMM estimator nor the FOD-GMM estimator exploited all available moment restrictions. But the two estimators exploited the same number of moment restrictions, and both are efficient for the moment restrictions they exploited. However, the restrictions they exploited differ.

4 Conclusion

When using panel data to estimate a dynamic regression model, researchers are often concerned about unobserved time-invariant effects. These effects can be removed with a suitable transformation. Many transformations are available, but due to historical priority, differencing the data has become a widely adopted approach. An alternative is the FOD transformation. The FOD transformation is computationally fast, even when both n and T are large (Phillips, 2020), which is an advantage for simulation studies and bootstrapping. Moreover, the analytical and Monte Carlo results provided in this paper indicate that FOD-GMM estimators have little large sample bias, regardless of whether n is large, or T is large, or both are large. This conclusion is useful because the reliability of large sample confidence intervals and test statistics based on an estimator depend on the estimator being asymptotically unbiased. Key restrictions for FOD-GMM estimators to have little bias are that the regressors be predetermined, the instrumental variables should not be weak, and the number of instruments used in any given period cannot increase with T too fast.

Appendix A: Proofs

The proofs of Theorems 1 and 2 rely on a few preliminary results. Lemma A.1 is used in the proof of Theorem 1. Theorem 2, on the other hand, draws on Lemma A.2.

Lemma A.1. *Assume rank$(\boldsymbol{Z}_t) = q_t$ with probability 1 and define $p_{n,t} := n\, \boldsymbol{z}'_{1,t} \left(\boldsymbol{Z}'_t \boldsymbol{Z}_t\right)^{-1} \boldsymbol{z}_{1,t}$. If $\boldsymbol{z}_{i,t}$ is identically distributed across i, then $E(p_{n,t}) = q_t$.*

Proof: By a well-known result for the trace of a projection matrix, we have $\operatorname{tr}(\boldsymbol{P}_t) = \operatorname{rank}(\boldsymbol{Z}_t)$. And $\operatorname{rank}(\boldsymbol{Z}_t) = q_t$ with probability 1 by assumption. But $\operatorname{tr}(\boldsymbol{P}_t) = \sum_{i=1}^{n} \boldsymbol{z}'_{i,t} \left(\boldsymbol{Z}'_t \boldsymbol{Z}_t\right)^{-1} \boldsymbol{z}_{i,t}$. Hence, $\sum_{i=1}^{n} E\left[\boldsymbol{z}'_{i,t} \left(\boldsymbol{Z}'_t \boldsymbol{Z}_t\right)^{-1} \boldsymbol{z}_{i,t}\right] = q_t$. Therefore, if $\boldsymbol{z}_{i,t}$ is identically distributed across i, then $nE\left[\boldsymbol{z}'_{1,t} \left(\boldsymbol{Z}'_t \boldsymbol{Z}_t\right)^{-1} \boldsymbol{z}_{1,t}\right] = q_t$.

Lemma A.2. *Assume $E(v_{1,t}|\boldsymbol{w}^*_{1,t}) = 0$ and $E(v^2_{1,t}|\boldsymbol{w}^*_{1,t}) = \sigma^2$ $(t \geq 1)$. Also, assume $E(\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t}) = \boldsymbol{Q}_t$ $(t \geq 1)$. Then, for $t \geq 1$, $r \geq 0$, and $1 \leq t + r \leq T - 1$, we have $E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t+r}) = \sigma^2 \boldsymbol{Q}_t$ or $\boldsymbol{0}$ according as $r = 0$ or $r > 0$.*

Proof: Note that $E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t+r}) = E[E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}|\boldsymbol{w}^*_{1,t+r})\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t+r}]$. Moreover,

$$\begin{aligned}
E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}|\boldsymbol{w}^*_{1,t+r}) &= c_t c_{t+r} E\left(v_{1,t}v_{1,t+r}|\boldsymbol{w}^*_{1,t+r}\right) \\
&\quad - \frac{c_t c_{t+r}}{T-t-r} \sum_{s=1}^{T-t-r} E\left(v_{1,t}v_{1,t+r+s}|\boldsymbol{w}^*_{1,t+r}\right) \\
&\quad - c_t c_{t+r} \frac{1}{T-t} \sum_{s=1}^{T-t} E\left(v_{1,t+s}v_{1,t+r}|\boldsymbol{w}^*_{1,t+r}\right) \\
&\quad + \frac{c_t c_{t+r}}{(T-t)(T-t-r)} \sum_{s=1}^{T-t} \sum_{j=1}^{T-t-r} E\left(v_{1,t+s}v_{1,t+r+j}|\boldsymbol{w}^*_{1,t+r}\right).
\end{aligned} \tag{A.1}$$

In order to evaluate the right-hand side of Eq. (A.1), note that the assumptions of the lemma imply that the $\ddot{v}_{1,t}$s are conditionally homoskedastic and uncorrelated across t. Specifically, for $\ell = k = j$, we have $E(v_{1,\ell}v_{1,k}|\boldsymbol{w}^*_{1,j}) = \sigma^2$. For $\ell = k > j$, we get $E[E(v_{1,\ell}v_{1,k}|\boldsymbol{w}^*_{1,k})|\boldsymbol{w}^*_{1,j}] = E(\sigma^2|\boldsymbol{w}^*_{1,j}) = \sigma^2$. If $\ell < k = j$, then $E(v_{1,\ell}v_{1,k}|\boldsymbol{w}^*_{1,j}) = v_{1,\ell}E(v_{1,k}|\boldsymbol{w}^*_{1,k}) = 0$. Finally, if $\ell < k$ and $k > j$, then $E(v_{1,\ell}v_{1,k}|\boldsymbol{w}^*_{1,j}) = E[v_{1,\ell}E(v_{1,k}|\boldsymbol{w}^*_{1,k})|\boldsymbol{w}^*_{1,j}] = E(v_{1,\ell}0|\boldsymbol{w}^*_{1,j}) = 0$. Hence,

$$E(v_{1,\ell}v_{1,k}|\boldsymbol{w}^*_{1,j}) = \sigma^2 \text{ or } 0 \text{ according as } \ell = k \text{ or } \ell < k \quad (k \geq j). \tag{A.2}$$

Suppose $r = 0$. Then from (A.2), we see that the first term on the right-hand side of Eq. (A.1) is $c_t^2\sigma^2$, the second and third terms are zero, and the last term is $c_t^2\sigma^2/(T-t)$. Hence, $E(\ddot{v}^2_{1,t}|\boldsymbol{w}^*_{1,t}) = \sigma^2$, and thus $E(\ddot{v}^2_{1,t}\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t}|\boldsymbol{w}^*_{1,t}) = \sigma^2\boldsymbol{Q}_t$.

Suppose $r > 0$. Then the first two terms on the right-hand side of (A.1) are zero. The third term equals $-c_t c_{t+r}\sigma^2/(T-t)$, and the last term is $c_t c_{t+r}\sigma^2/(T-t)$. Hence, $E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}|\boldsymbol{w}^*_{1,t+r}) = 0$, and thus $E(\ddot{v}_{1,t}\ddot{v}_{1,t+r}\boldsymbol{z}_{1,t}\boldsymbol{z}'_{1,t+r}|\boldsymbol{w}^*_{1,t+r}) = \boldsymbol{0}$.

A.1 Theorem 1 proof

First note that

$$\sqrt{nT}\boldsymbol{b}_{n,T} = \sum_{t=1}^{T-1} c_t^2 \boldsymbol{X}_t' \boldsymbol{P}_t \left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t\right) - \sum_{t=1}^{T-1} c_t^2 \overline{\boldsymbol{X}}_t' \boldsymbol{P}_t \left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t\right). \tag{A.3}$$

Moreover, Assumption A1 implies $(\boldsymbol{w}_{i,t}', v_{i,t+s})$ is independent of $\boldsymbol{w}_{j,t}$, for $i \neq j$. Thus, if we set $\boldsymbol{w}_t' := (\boldsymbol{w}_{1,t}', \ldots, \boldsymbol{w}_{n,t}')$, then $E\left(v_{i,t+s}|\boldsymbol{w}_t\right) = E\left(v_{i,t+s}|\boldsymbol{w}_{i,t}\right)$. And, A1, A3, and the law of iterated expectations imply $E\left(v_{i,t+s}|\boldsymbol{w}_{i,t}\right) = E\left[E\left(v_{i,t+s}|\boldsymbol{w}_{i,t+s}\right)|\boldsymbol{w}_{i,t}\right] = E\left(0|\boldsymbol{w}_{i,t}\right) = 0$ $(s \geq 0)$. Hence, $E\left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t|\boldsymbol{w}_t\right) = \boldsymbol{0}$, and thus $E\left[\boldsymbol{X}_t'\boldsymbol{P}_t\left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t\right)\right] = E\left[\boldsymbol{X}_t'\boldsymbol{P}_t E\left(\boldsymbol{v}_t - \overline{\boldsymbol{v}}_t|\boldsymbol{w}_t\right)\right] = \boldsymbol{0}$. From this observation and Eq. (A.3), it follows that

$$\sqrt{nT}E(\boldsymbol{b}_{n,T}) = -\sum_{t=1}^{T-1} \frac{1}{T-t+1}\left(\boldsymbol{s}_{1,T-t} - \boldsymbol{s}_{2,T-t}\right), \tag{A.4}$$

where $\boldsymbol{s}_{1,T-t} := \sum_{s=1}^{T-t} E\left(\boldsymbol{X}_{t+s}'\boldsymbol{P}_t\boldsymbol{v}_t\right)$ and $\boldsymbol{s}_{2,T-t} := \sum_{s=1}^{T-t} E\left(\boldsymbol{X}_{t+s}'\boldsymbol{P}_t\overline{\boldsymbol{v}}_t\right)$.

In order to evaluate $\boldsymbol{s}_{1,T-t}$, note that the kth entry of $E\left(\boldsymbol{X}_{t+s}'\boldsymbol{P}_t\boldsymbol{v}_t\right)$ is

$$\sum_{i=1}^{n} E\left(x_{i,t+s,k}\boldsymbol{z}_{i,t}'(\boldsymbol{Z}_t'\boldsymbol{Z}_t)^{-1}\boldsymbol{Z}_t'\boldsymbol{v}_t\right) = \sum_{i=1}^{n} E\left[\boldsymbol{z}_{i,t}'(\boldsymbol{Z}_t'\boldsymbol{Z}_t)^{-1}\sum_{j=1}^{n}\boldsymbol{z}_{j,t}E\left(v_{j,t}x_{i,t+s,k}|\boldsymbol{w}_t\right)\right] \tag{A.5}$$

And, by the law of iterated expectations, $E\left(v_{j,t}x_{i,t+s,k}|\boldsymbol{w}_t\right) = E\left[x_{i,t+s,k}E\left(v_{j,t}|\boldsymbol{w}_t, \boldsymbol{x}_{i,t+s}\right)|\boldsymbol{w}_t\right]$. Given $(\boldsymbol{w}_{j,t}', v_{j,t})$ is independent of $\boldsymbol{w}_{i,T}$ for $j \neq i$, it follows that $E\left(v_{j,t}|\boldsymbol{w}_t, \boldsymbol{x}_{i,t+s}\right) = E\left(v_{j,t}|\boldsymbol{w}_{j,t}\right)$ for $j \neq i$. Moreover, $E\left(v_{j,t}|\boldsymbol{w}_{j,t}\right) = 0$ by A1 and A3. Hence, $E\left(v_{j,t}x_{i,t+s,k}|\boldsymbol{w}_t\right) = 0$ for $j \neq i$. On the other hand, $E\left(v_{j,t}x_{i,t+s,k}|\boldsymbol{w}_t\right) = E\left(v_{i,t}x_{i,t+s,k}|\boldsymbol{w}_t\right)$ for $j = i$. Hence, $\sum_{j=1}^{n}\boldsymbol{z}_{j,t}E\left(v_{j,t}x_{i,t+s,k}|\boldsymbol{w}_t\right) = \boldsymbol{z}_{i,t}E\left(v_{i,t}x_{i,t+s,k}|\boldsymbol{w}_t\right)$, which implies the right-hand side of (A.5) is

$$\sum_{i=1}^{n} E\left[\boldsymbol{z}_{i,t}'(\boldsymbol{Z}_t'\boldsymbol{Z}_t)^{-1}\boldsymbol{z}_{i,t}E\left(v_{i,t}x_{i,t+s,k}|\boldsymbol{w}_t\right)\right] = E\left[p_{n,t}E\left(v_{1,t}x_{1,t+s,k}|\boldsymbol{w}_t\right)\right], \tag{A.6}$$

where the equality in Eq. (A.6) follows from the fact that the $\boldsymbol{u}_i$s are identically distributed and $p_{n,t} = n\,\boldsymbol{z}_{1,t}'\left(\boldsymbol{Z}_t'\boldsymbol{Z}_t\right)^{-1}\boldsymbol{z}_{1,t}$. Moreover, A1 implies $E\left(v_{1,t}x_{1,t+s,k}|\boldsymbol{w}_t\right) = E\left(v_{1,t}x_{1,t+s,k}|\boldsymbol{w}_{1,t}\right)$. Furthermore, from A3 and the definition of conditional covariance, we have $\gamma_{k,t,s} = E\left(v_{1,t}x_{1,t+s,k}|\boldsymbol{w}_{1,t}\right)$. The preceding observations imply the kth entry of $E\left(\boldsymbol{X}_{t+s}'\boldsymbol{P}_t\boldsymbol{v}_t\right)$ is $E\left(p_{n,t}\gamma_{k,t,s}\right)$. This conclusion and the definition of $\boldsymbol{s}_{1,T-t}$

implies the kth entry in $\boldsymbol{s}_{1,T-t}$ is

$$s_{1,T-t,k} = E\left(p_{n,t}\sum_{s=1}^{T-t}\gamma_{k,t,s}\right) \qquad (1 \leq t \leq T-1).$$

Moreover,

$$|s_{1,T-t,k}| \leq E\left(p_{n,t}\left|\sum_{s=1}^{T-t}\gamma_{k,t,s}\right|\right) \leq E\left(p_{n,t}\right)M = q_t M, \tag{A.7}$$

where the second inequality in (A.7) follows from A4 and the equality on the far right-hand side follows from Lemma A.1.

Next, the entries in $\boldsymbol{s}_{2,T-t}$ are evaluated. To that end, note that by arguments similar to those used to establish Eq. (A.6), the kth entry in $E\left(\boldsymbol{X}'_{t+s}\boldsymbol{P}_t\overline{\boldsymbol{v}}_t\right)$ is $E\left(p_{n,t}\overline{v}_{1,t}x_{1,t+s,k}\right)$. Moreover, $E\left(p_{n,t}v_{1,t+r}x_{1,t+s,k}\right) = E\left[p_{n,t}x_{1,t+s,k}E\left(v_{1,t+r}|\boldsymbol{w}_{t+s}\right)\right] = 0$ if $r \geq s$. Hence, for $s = 1$, we get $(T-t)\,E\left(p_{n,t}\overline{v}_{1,t}x_{1,t+s,k}\right) = E(p_{n,t}x_{1,t+1,k}\sum_{r=1}^{T-t}v_{1,t+r}) = 0$. And, when $t = T-1$, it must be that $s = 1$. Thus, for $t = T-1$ and $s = 1$, we get $(T-t)\,E\left(p_{n,t}\overline{v}_{1,t}x_{1,t+s,k}\right) = E\left(p_{n,T-1}v_{1,T}x_{1,T,k}\right) = 0$. On the other hand, for $1 \leq t \leq T-2$ and $s \geq 2$, we have

$$\begin{aligned}(T-t)E\left(p_{n,t}\overline{v}_{1,t}x_{1,t+s,k}\right) &= E\left(p_{n,t}x_{1,t+s,k}\sum_{r=1}^{T-t}v_{1,t+r}\right)\\ &= E\left(p_{n,t}x_{1,t+s,k}\sum_{r=1}^{s-1}v_{1,t+r}\right)\\ &= E\left[p_{n,t}\sum_{r=1}^{s-1}E\left(v_{1,t+r}x_{1,t+r+(s-r),k}|\boldsymbol{w}_{t+r}\right)\right]\\ &= E\left[p_{n,t}\sum_{r=1}^{s-1}\gamma_{k,t+r,s-r}\right].\end{aligned} \tag{A.8}$$

Moreover, for $T \geq 3$,

$$
\begin{aligned}
\sum_{s=2}^{T-t} \sum_{r=1}^{s-1} \gamma_{k,t+r,s-r} &= \underset{(r=1)}{\gamma_{k,t+1,1}} && (s=2) \\
&+ \underset{(r=1)}{\gamma_{k,t+1,2}} + \underset{(r=2)}{\gamma_{k,t+2,1}} && (s=3) \\
&+ \cdots && \text{(A.9)} \\
&+ \underset{(r=1)}{\gamma_{k,t+1,T-t-1}} + \underset{(r=2)}{\gamma_{k,t+2,T-t-2}} + \cdots + \underset{(r=T-t-1)}{\gamma_{k,T-1,1}} && (s=T-t) \\
&= \sum_{r=1}^{T-t-1} \sum_{j=1}^{T-t-r} \gamma_{k,t+r,j}.
\end{aligned}
$$

It follows from Eq.s (A.8) and (A.9), $E\left(p_{n,T-1} v_{1,T} x_{1,T,k}\right) = 0$, and the definition of $\boldsymbol{s}_{2,T-t}$ that, for $T \geq 3$, the kth entry in $\boldsymbol{s}_{2,T-t}$ is

$$
\begin{aligned}
s_{2,T-t,k} &= \frac{1}{T-t} E\left(p_{n,t} \sum_{r=1}^{T-t-1} \sum_{j=1}^{T-t-r} \gamma_{k,t+r,j}\right) && (1 \leq t \leq T-2) \\
&= 0 && (t = T-1).
\end{aligned}
$$

Hence, $|s_{2,T-t,k}| = 0$ for $t = T-1$, and, for $1 \leq t \leq T-2$, we have

$$
\begin{aligned}
|s_{2,T-t,k}| &\leq \frac{1}{T-t} E\left(p_{n,t} \sum_{r=1}^{T-t-1} \left|\sum_{j=1}^{T-t-r} \gamma_{k,t+r,j}\right|\right) \\
&\leq E\left(p_{n,t}\right) \frac{1}{T-t} \sum_{r=1}^{T-t-1} M \qquad\qquad \text{(A.10)} \\
&\leq q_t M.
\end{aligned}
$$

Expression (A.4), an obvious inequality, and the inequalities in (A.7) and (A.10) imply

$$
|E\left(b_{n,T,k}\right)| \leq \frac{1}{\sqrt{nT}} \sum_{t=1}^{T-1} \frac{1}{T-t+1} \left(|s_{1,T-t,k}| + |s_{2,T-t,k}|\right) \leq \frac{2}{\sqrt{nT}} M \sum_{t=1}^{T-1} \frac{q_t}{T-t+1}.
$$

Moreover, given $q_t \leq q_T^*$, for all t, we have $\sum_{t=1}^{T-1} q_t/(T-t+1) \leq q_T^* \sum_{t=1}^{T-1} (T-t+1)^{-1}$. And, $\sum_{t=1}^{T-1} (T-t+1)^{-1} = \sum_{t=1}^{T} t^{-1} - 1$. Furthermore, $\sum_{t=1}^{T} t^{-1} - 1 \leq \ln T$ for $T \geq 1$ (see,e.g., Havil, 2003, p. 47). The foregoing implies that, for $T \geq 3$,

$$
|E\left(b_{n,T,k}\right)| \leq 2M \frac{q_T^*}{\sqrt{nT}} \ln T \qquad (k = 1, \ldots, K). \tag{A.11}
$$

Now suppose $q_T^* = O(T^\delta)$. This assumption and (A.11) imply

$$E\left(b_{n,T,k}\right) = O\left(n^{-1/2}T^{\delta-1/2}\ln T\right) \qquad (k = 1,\ldots,K).$$

A.2 Theorem 2 proof:

Note that

$$\sqrt{nT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) - \boldsymbol{A}_{n,T}^{-1}E\left(\boldsymbol{b}_{n,T}\right) = \boldsymbol{A}_{n,T}^{-1}\left[\boldsymbol{b}_{n,T} - E\left(\boldsymbol{b}_{n,T}\right)\right].$$

If $q_T^* = O(T^\delta)$, with $0 \le \delta < 1/2$, it follows from Theorem 1 that $\boldsymbol{b}_{n,T} - E(\boldsymbol{b}_{n,T}) = \boldsymbol{b}_{n,T} + o(1)$ for all sequences of n and T—for example, for $n \to \infty$ with T fixed or $T \to \infty$. Consequently, we can ignore $E(\boldsymbol{b}_{n,T})$ when considering the asymptotic behavior of $\boldsymbol{b}_{n,T}$ and that of $\sqrt{nT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

Note that Assumption A3* implies $E(\boldsymbol{z}_{1,t}\ddot{v}_{1,t}) = \boldsymbol{0}$. Moreover, from Lemma A.2, it follows that $\mathrm{Var}(\boldsymbol{z}_{1,t}\ddot{v}_{1,t}) = \sigma^2\boldsymbol{Q}_t$. Therefore, by the central limit theorem for i.i.d. random vectors, we have $\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t/\sqrt{n} \xrightarrow{d} N(\boldsymbol{0}, \sigma^2\boldsymbol{Q}_t)$, as $n \to \infty$. Hence, $\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t/\sqrt{n} = O_p(1)$ for each t. Moreover, A1, A2, A6, and the law of large numbers implies $\ddot{\boldsymbol{X}}_t'\boldsymbol{Z}_t\left(\boldsymbol{Z}_t'\boldsymbol{Z}_t\right)^{-1} \xrightarrow{p} \boldsymbol{R}_t'\boldsymbol{Q}_t^{-1}$, as $n \to \infty$, for each t. It follows from the preceding that

$$\begin{aligned}\boldsymbol{b}_{n,T} - \frac{1}{\sqrt{nT}}\sum_{t=1}^{T-1}\boldsymbol{R}_t'\boldsymbol{Q}_t^{-1}\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\ddot{\boldsymbol{X}}_t'\boldsymbol{Z}_t\left(\boldsymbol{Z}_t'\boldsymbol{Z}_t\right)^{-1} - \boldsymbol{R}_t'\boldsymbol{Q}_t^{-1}\right)\frac{1}{\sqrt{n}}\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t \\ &= o_p(1) \qquad (n \to \infty).\end{aligned}$$

Therefore, the asymptotic distribution of $\boldsymbol{b}_{n,T}$, as $n \to \infty$, is the same as the asymptotic distribution of $\sum_{t=1}^{T-1}\boldsymbol{R}_t'\boldsymbol{Q}_t\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t/\sqrt{nT}$, as $n \to \infty$.

To evaluate the latter, note that if we set $\boldsymbol{\epsilon}_{i,t} := \boldsymbol{R}_t'\boldsymbol{Q}_t^{-1}\boldsymbol{z}_{i,t}\ddot{v}_{i,t}$, then $\boldsymbol{R}_t'\boldsymbol{Q}_t^{-1}\boldsymbol{Z}_t'\ddot{\boldsymbol{v}}_t = \sum_{i=1}^n \boldsymbol{\epsilon}_{i,t}$. Next, define the $K(T-1)\times 1$ vectors $\boldsymbol{\epsilon}_i := (\boldsymbol{\epsilon}_{i1}', \ldots, \boldsymbol{\epsilon}_{i,T-1}')'$ $(i = 1,\ldots,n)$. Assumption A3* implies $E(\boldsymbol{\epsilon}_1) = \boldsymbol{0}$. Moreover, the conclusion of Lemma A.2 implies the entries in $\boldsymbol{\epsilon}_{1,s}$ are uncorrelated with the entries in $\boldsymbol{\epsilon}_{1,t}$ for $s \neq t$. Hence, the variance-covariance matrix of $\boldsymbol{\epsilon}_1$ is block diagonal. In particular, $\mathrm{Var}(\boldsymbol{\epsilon}_1) = \sigma^2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega} := \mathrm{diag}(\boldsymbol{\Omega}_1,\ldots,\boldsymbol{\Omega}_{T-1})$. Moreover, Assumptions A2 and A7 imply $\boldsymbol{\Omega}$ is positive definite. And, A1 implies the $\boldsymbol{\epsilon}_i$s are i.i.d across i. It follows from the central limit theorem for i.i.d. random vectors that, as $n \to \infty$, we have $\sum_{i=1}^n \boldsymbol{\epsilon}_i/\sqrt{n} \xrightarrow{d} \boldsymbol{\epsilon}^*$, with $\boldsymbol{\epsilon}^* \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega})$.

Partition $\boldsymbol{\epsilon}^*$ so that $\boldsymbol{\epsilon}^* = (\boldsymbol{\epsilon}_1^{*\prime}, \ldots, \boldsymbol{\epsilon}_{T-1}^{*\prime})'$, where $\boldsymbol{\epsilon}_1^*, \ldots, \boldsymbol{\epsilon}_{T-1}^*$ are $K \times 1$ vectors, and recall that $\boldsymbol{\Omega}$ is block diagonal; that is, the entries of $\boldsymbol{\epsilon}_s^*$ are uncorrelated with the entries in $\boldsymbol{\epsilon}_t^*$ for $s \neq t$. Moreover, as shown in the last paragraph, the entries in $\boldsymbol{\epsilon}^*$ are multivariate normal. Hence, $\boldsymbol{\epsilon}_s^*$ is independent of $\boldsymbol{\epsilon}_t^*$ for $s \neq t$.

Let $\boldsymbol{a} := (a_1,\ldots,a_K)'$ be a vector of constants with $\boldsymbol{a} \neq \boldsymbol{0}$, and define $\mathcal{Y}_t = \boldsymbol{a}'\boldsymbol{\epsilon}_t^*$ $(t = 1,\ldots,T-1)$. The variable $\mathcal{Y}_t$ is independent across t, with mean $E(\mathcal{Y}_t) = 0$ and variance $\sigma_{\mathcal{Y},t}^2 = \sigma^2\boldsymbol{a}'\boldsymbol{\Omega}_t\boldsymbol{a}$. Let $s_{T-1}^2 = \sum_{t=1}^{T-1}\sigma_{\mathcal{Y},t}^2$ and $\mathcal{Z}_{T-1} = \sum_{t=1}^{T-1}\mathcal{Y}_t/s_{T-1}$.

Then, by Lyapunov's central limit theorem, $\mathcal{Z}_{T-1} \xrightarrow{d} N(0,1)$, as $T \to \infty$, if $s_{T-1}^{-3} \sum_{t=1}^{T-1} E\left|\mathcal{Y}_t\right|^3 \to 0$, as $T \to \infty$ (see, e.g., Sen and Singer, 1993, Theorem 3.3.2). But $\mathcal{Y}_t$ is normal, and thus $E\left|\mathcal{Y}_t\right|^3 = \sigma_{\mathcal{Y},t}^3 2^{3/2}\Gamma(2)/\sqrt{\pi} < \infty$ (Winkelbauer, 2014). Thus, $s_{T-1}^{-3} \sum_{t=1}^{T-1} E\left|\mathcal{Y}_t\right|^3 \to 0$, as $T \to \infty$, if

$$\frac{1}{s_{T-1}^3} \sum_{t=1}^{T-1} \sigma_{\mathcal{Y},t}^3 \to 0, \qquad T \to \infty. \tag{A.12}$$

In order to verify (A.12) note that

$$\frac{1}{s_{T-1}^3} \sum_{t=1}^{T-1} \sigma_{\mathcal{Y},t}^3 = \frac{1}{\left(\boldsymbol{a}'\overline{\boldsymbol{\Omega}}_T\boldsymbol{a}\right)^{3/2}} \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} \left(\boldsymbol{a}'\boldsymbol{\Omega}_t\boldsymbol{a}\right)^{3/2}. \tag{A.13}$$

Moreover,

$$\left(\boldsymbol{a}'\boldsymbol{\Omega}_t\boldsymbol{a}\right)^{3/2} \leq K \sum_{j=1}^{K}\sum_{k=1}^{K} \left|a_j a_k \omega_{t,j,k}\right|^{3/2} \leq \nu_t^{3/2} K \sum_{j=1}^{K}\sum_{k=1}^{K} \left|a_j a_k\right|^{3/2}, \tag{A.14}$$

where the first inequality in (A.14) follows from the c$_r$-inequality (see, e.g., Sen and Singer, 1993, p. 21) and the second from the definition of ν_t. It follows from (A.13), (A.14), Assumption A8, which implies $\lim_{T\to\infty} \boldsymbol{a}'\overline{\boldsymbol{\Omega}}_T\boldsymbol{a} = \boldsymbol{a}'\overline{\boldsymbol{\Omega}}\boldsymbol{a} > 0$, and Assumption A9 that

$$\frac{1}{s_{T-1}^3} \sum_{t=1}^{T-1} \sigma_{\mathcal{Y},t}^3 \leq \frac{1}{\left(\boldsymbol{a}'\overline{\boldsymbol{\Omega}}_T\boldsymbol{a}\right)^{3/2}} K \sum_{j=1}^{K}\sum_{k=1}^{K} \left|a_j a_k\right|^{3/2} \frac{1}{T^{3/2}} \sum_{t=1}^{T-1} \nu_t^{3/2} \to 0, \qquad T \to \infty.$$

In other words, (A.12) is satisfied, and therefore $\mathcal{Z}_{T-1} \xrightarrow{d} N(0,1)$, as $T \to \infty$. Upon applying the Cramér-Wold Theorem (Sen and Singer, 1993, Theorem 3.2.4), it follows that $\sum_{t=1}^{T-1} \boldsymbol{\epsilon}_t^*/\sqrt{T} \xrightarrow{d} N(\mathbf{0}, \sigma^2\overline{\boldsymbol{\Omega}})$, as $T \to \infty$.

The preceding verifies $\boldsymbol{b}_{n,T} \xrightarrow{d} N(\mathbf{0}, \sigma^2\overline{\boldsymbol{\Omega}})$ $(n, T \to \infty)_{\text{seq}}$. Moreover, $\boldsymbol{A}_{n,T} \xrightarrow{p} \overline{\boldsymbol{\Omega}}$ $(n, T \to \infty)_{\text{seq}}$. Hence, $\sqrt{nT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{A}_{n,T}^{-1}\boldsymbol{b}_{n,T} \xrightarrow{d} N(\mathbf{0}, \sigma^2\overline{\boldsymbol{\Omega}}^{-1})$ $(n, T \to \infty)_{\text{seq}}$.

References

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

Alvarez, J., & Arellano, M. (2004). Robust likelihood estimation of dynamic panel data models. CEMFI Working Paper 0421.

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

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

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

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

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

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

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

Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, 141, 597-620.

Havil, J. (2003). *Gamma: Exploring Euler's Constant*. Princeton University Press, Princeton, New Jersey.

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

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

Hsiao, C. & Zhou, Q. (2017). First difference or forward demeaning: Implications for the method of moments estimators. *Econometric Reviews 36*, 883–897.

Kruiniger, H. (2013). Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of Econometrics 173*, 175–188.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics 31*, 451–472.

Phillips, R. F. (2010). Iterated feasible generalized least-squares estimation of augmented dynamic panel data models. *Journal of Business & Economic Statistics 28*, 410–422.

Phillips, R. F. (2018). Quasi maximum likelihood estimation of dynamic panel data models. *Communications in Statistics – Theory and Methods 47*, 3970–3986.

Phillips, R. F. (2019). A numerical equivalence result for generalized method of moments. *Economics Letters 179*, 13–15.

Phillips, R. F. (2020). Quantifying the advantages of forward orthogonal deviations for long time series. *Computational Economics 55*, 653–672.

Phillips, R. F. (2022). Supplement to "Forward orthogonal deviations GMM and the absence of large sample bias".

Sen, P. K., & Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Boca Raton, FL: Chapman & Hall/CRC.

Winkelbauer, A. (2014). Moments and absolute moments of the normal distribution. arXiv:1209.4340v2 [math.ST].