Bias-corrected estimation of panel vector autoregressions

GEERT DHAENE[1] and KOEN JOCHMANS[2]

[1] *KU Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium*
E-mail: geert.dhaene@kuleuven.be

[2] *Sciences Po, Département d'économie,, 28 rue des Saints Pères, 75007 Paris, France*
E-mail: koen.jochmans@sciencespo.fr

First version: June 12, 2015. This version: June 4, 2016

We derive a bias-corrected least-squares estimator for panel vector autoregressions with fixed effects. The estimator is straightforward to implement and is asymptotically unbiased under asymptotics where the number of time series observations and the number of cross-sectional observations grow at the same rate. This makes the estimator particularly well suited for most macroeconomic data sets.

JEL classification: C33

Keywords: bias correction, fixed effects, panel data, vector autoregression.

1. INTRODUCTION

Vector autoregressions are a standard tool in macroeconometrics since the work of Sims (1972, 1980). A growing literature exploits the availability of large longitudinal data sets to fit panel versions of vector autoregressive models; see, e.g., Canova and Ciccarelli (2013). In addition, panel vector autoregressions also find application in microeconomics; examples include the estimation of wage equations in Holtz-Eakin et al. (1988) and Alonso-Borrego and Arellano (1999).[1]

We consider the estimation of vector autoregressions from panel data on N units and T (effective) time periods. While it is well-known that least-squares estimators of vector autoregressions that feature fixed effects are heavily biased in short panels, the fact that they are also asymptotically biased as $N, T \to \infty$ unless $N/T \to 0$ (Phillips and Moon 1999; Hahn and Kuersteiner 2002) seems to be largely neglected in the empirical literature. At the same time, the generalized method-of-moment estimator of Holtz-Eakin

[1] Further examples of the use of panel vector autoregressions include studies of fiscal policy and investment (Alesina et al. 2002), fiscal policy and consumption (Tagkalakis, 2008), financial intermediation and growth (Rousseau and Wachtel 2000), risk-sharing (Asdrubali and Kim 2004), savings (Loayza et al. 2000, Attanasio et al. 2000), and crime (Fajnzylber et al. 2002).

et al. (1988), which performs well under fixed-T asymptotics, is asymptotically biased under asymptotics where $N, T \to \infty$ unless $N/T \to \infty$ (Alvarez and Arellano 2003). In addition, such estimators are well known to have poor finite-sample properties when the data are persistent, and their performance is sensitive to the distribution of the fixed effects. Here we consider estimation under asymptotics where N/T converges to a (positive) constant. Such an asymptotic approximation represents a middle ground between the two sampling schemes discussed above. It is well suited for most macroeconomic data sets, where T typically cannot reasonably be considered small relative to N.

We extend the bias-correction approach of Hahn and Kuersteiner (2002) from first-order vector autoregressions to models with higher-order dynamics. This is important, not in the least because the lag length is generally unknown. Bias-corrected estimation under lag-length misspecification can exacerbate the bias, thus worsening inference; see Lee (2012) for theoretical and simulation results. We also derive the corresponding bias correction for impulse-response functions.

2. VECTOR AUTOREGRESSION FOR PANEL DATA

Consider panel data on N units observed for $T + P$ consecutive time periods. For each unit i we observe M outcome variables $y_{it1}, \dots, y_{itM}$, where t ranges from $1-P$ to T. The behavior of $\boldsymbol{y}_{it} = (y_{it1}, y_{it2}, \dots, y_{itM})'$ is described by the Pth order vector autoregression

$$\boldsymbol{y}_{it} = \boldsymbol{\Gamma}_1 \boldsymbol{y}_{it-1} + \boldsymbol{\Gamma}_2 \boldsymbol{y}_{it-2} + \cdots + \boldsymbol{\Gamma}_P \boldsymbol{y}_{it-P} + \boldsymbol{\epsilon}_{it}, \tag{2.1}$$

where $\boldsymbol{y}_{it-p} = L^p\, \boldsymbol{y}_{it}$ is the pth lag of $\boldsymbol{y}_{it}$, $\boldsymbol{\Gamma}_p$ is the associated $M \times M$ coefficient matrix, and $\boldsymbol{\epsilon}_{it}$ is an M-dimensional error term. We assume that

$$\boldsymbol{\epsilon}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\upsilon}_{it}$$

for a fixed effect $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{iM})'$ and an error vector $\boldsymbol{\upsilon}_{it} = (\upsilon_{it1}, \dots, \upsilon_{itM})'$. We normalize $E[\boldsymbol{\upsilon}_{it}] = 0$ and let $\boldsymbol{\Omega} = E[\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}']$.

Equation (2.1) can be extended to include time dummies, one for each period. Time dummies change the least-squares estimator but they leave the bias adjustment and the asymptotic approximation developed below unchanged (see Hahn and Moon 2006 for a discussion on this).

We complete the model by imposing the following conditions.

Assumption 1 (Stationarity condition). *The roots of the determinantal equation*

$$\det\left(\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 z - \cdots - \boldsymbol{\Gamma}_P z^P\right) = 0$$

lie outside the unit circle.

Assumption 1 implies that the vector autoregressive process is stable.

Throughout, we treat the $\boldsymbol{\alpha}_i$ as fixed, that is, we condition on them. We also condition on the initial observations, $\boldsymbol{y}_{i(1-P)}, \ldots, \boldsymbol{y}_{i0}$. This allows the initial observations not to be generated from the corresponding stationary distribution, and so, does not require the time series processes to have started in the distant past.

Assumption 2 (Regularity conditions). $\boldsymbol{v}_{it}$ *has finite eight-order moments and, as* $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\alpha}_i\|^2 = O(1), \qquad \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{y}_{i1-p}\|^2 = O(1),$$

for $p = 1, \ldots, P$.

Assumption 2 ensures regular asymptotic behavior of the least-squares estimator.

For simplicity, we will assume that the errors $\boldsymbol{v}_{it}$ are independent and identically distributed.

Assumption 3 (Errors). $\boldsymbol{v}_{it}$ *is independent and identically distributed across* i *and* t, *and* $E[\boldsymbol{v}_{it}] = 0$.

Independence across time can be relaxed to allow for dependence between $\boldsymbol{v}_{it}$ and $\boldsymbol{v}_{it-p}$ through their higher-order moments. This would come at the cost of more complicated regularity conditions, parallelling Hahn and Kuersteiner (2002, Conditions 1 and 2), but would leave our bias calculations unchanged.

Under these conditions, as $t \to \infty$, $\boldsymbol{y}_{it}$ has the moving-average representation

$$\boldsymbol{y}_{it} = \sum_{k=0}^{\infty}\boldsymbol{\Phi}_k\,\boldsymbol{\epsilon}_{it-k} = \boldsymbol{\mu}_i + \sum_{k=0}^{\infty}\boldsymbol{\Phi}_k\,\boldsymbol{v}_{it-k},$$

where $\boldsymbol{\mu}_i = \sum_{k=0}^{\infty} \boldsymbol{\Phi}_k \, \boldsymbol{\alpha}_i$ and the matrices $\boldsymbol{\Phi}_k$ are defined by

$$\left(\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 L - \cdots - \boldsymbol{\Gamma}_P L^P\right)^{-1} = \sum_{k=0}^{\infty} \boldsymbol{\Phi}_k L^k. \tag{2.2}$$

This representation is important as it implies $\partial \boldsymbol{y}_{it+h} / \partial \boldsymbol{\upsilon}_{it}' = \boldsymbol{\Phi}_h$, which quantifies the impact on $\boldsymbol{y}_{it+h}$ of a unit increase in the elements of $\boldsymbol{\upsilon}_{it}$. As a function of h, this defines the impulse-response functions, which are key parameters of interest (see, e.g., Hamilton 1994, Section 11.4).

3. BIAS-CORRECTED ESTIMATION

Define the $M \times MP$ matrix $\boldsymbol{\Gamma}' = (\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_P)$ and $\boldsymbol{x}_{it} = (\boldsymbol{y}_{it-1}', \boldsymbol{y}_{it-2}', \ldots, \boldsymbol{y}_{it-P}')'$ to write (2.1) as

$$\boldsymbol{y}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\Gamma}' \boldsymbol{x}_{it} + \boldsymbol{\upsilon}_{it}.$$

Collect all time-series observations and error terms for unit i in the matrices

$$\boldsymbol{Y}_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iT}), \qquad \boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}), \qquad \boldsymbol{\Upsilon}_i = (\boldsymbol{\upsilon}_{i1}, \ldots, \boldsymbol{\upsilon}_{iT}),$$

and let $\boldsymbol{\iota}_T$ denote the T-dimensional vector of ones to define $\boldsymbol{A}_i = \boldsymbol{\iota}_T' \otimes \boldsymbol{\alpha}_i$. We may then write

$$\boldsymbol{Y}_i = \boldsymbol{A}_i + \boldsymbol{\Gamma}' \boldsymbol{X}_i + \boldsymbol{\Upsilon}_i.$$

The within-group least-squares (WG-OLS) estimator is

$$\widehat{\boldsymbol{\Gamma}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{Y}_i'\right),$$

where $\boldsymbol{M} = \boldsymbol{I}_T - \boldsymbol{\iota}_T \boldsymbol{\iota}_T'$. We consider asymptotically unbiased estimation of $\boldsymbol{\Gamma}$ under asymptotics where $N, T \to \infty$ with $N/T \to \rho^2$ and $0 < \rho < \infty$. Under this asymptotic scheme, WG-OLS is asymptotically normal, but biased. Let

$$\begin{aligned} \boldsymbol{\Sigma} &= \operatorname{plim} \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{X}_i' = \lim_{t \to \infty} \operatorname{Var}(\boldsymbol{x}_{it}), \\ \boldsymbol{B} &= -\left(\boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_P)^{-1}\right) \boldsymbol{\Omega}, \end{aligned} \tag{3.1}$$

and $\boldsymbol{b} = \operatorname{vec} \boldsymbol{B} = -(\boldsymbol{I}_M \otimes \boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_P)^{-1}) \operatorname{vec} \boldsymbol{\Omega}$.

Lemma 1 (Within-group least-squares estimator). *Let Assumptions 1–3 hold. Then*

$$\sqrt{NT}\,\mathrm{vec}\,(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}) \xrightarrow{d} \mathcal{N}\left(\rho\,(\boldsymbol{I}_M \otimes \boldsymbol{\Sigma}^{-1})\,\boldsymbol{b}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1}\right) \tag{3.2}$$

as $N, T \to \infty$ *with* $N/T \to \rho^2$.

(Proofs are given in the Appendix.)

Lemma 1 suggests the bias-corrected WG-OLS (BC-WG-OLS) estimator

$$\widetilde{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}} - \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{B}}}{T}, \tag{3.3}$$

where

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{X}_i', \qquad \widehat{\boldsymbol{B}} = -(\boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \widehat{\boldsymbol{\Gamma}}_1 - \cdots - \widehat{\boldsymbol{\Gamma}}_P)^{-1})\,\widehat{\boldsymbol{\Omega}}, \tag{3.4}$$

with

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{NT}\sum_{i=1}^{N}\left(\boldsymbol{Y}_i - \widehat{\boldsymbol{\Gamma}}'\boldsymbol{X}_i\right)\boldsymbol{M}\left(\boldsymbol{Y}_i - \widehat{\boldsymbol{\Gamma}}'\boldsymbol{X}_i\right)'. \tag{3.5}$$

This estimator is asymptotically unbiased. Lemma 1 immediately implies the following.

Theorem 1 (Bias-corrected least-squares estimator). *Let Assumptions 1–3 hold. Then*

$$\sqrt{NT}\,\mathrm{vec}\,(\widetilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1}\right)$$

as $N, T \to \infty$ *with* $N/T \to \rho^2$.

When $M = 1$ the model in (2.1) reduces to

$$y_{it} = \alpha_i + \gamma_1 y_{it-1} + \cdots + \gamma_P y_{it-P} + \upsilon_{it}, \qquad \upsilon_{it} \sim \text{i.i.d.}\,(0, \omega^2). \tag{3.6}$$

In this case a simpler, yet asymptotically equivalent, bias correction may be performed. Write (3.6) as

$$y_{it} = \alpha_i + \boldsymbol{\gamma}'\boldsymbol{x}_{it} + \upsilon_{it},$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_P)'$ and $\boldsymbol{x}_{it} = (y_{it-1}, \ldots, y_{it-P})'$. The WG-OLS estimator of $\boldsymbol{\gamma}$ is

$$\widehat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \boldsymbol{x}_{i\cdot})(\boldsymbol{x}_{it} - \boldsymbol{x}_{i\cdot})'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \boldsymbol{x}_{i\cdot})(y_{it} - y_{i\cdot})\right).$$

Write $\boldsymbol{\Sigma} = \omega^2 \boldsymbol{V}$, where $\boldsymbol{V}$ is the variance matrix of $\omega^{-1}\boldsymbol{x}_{it}$ in the limit $t \to \infty$. By combining Lemma 1 with the expression for $\boldsymbol{V}^{-1}$ in Galbraith and Galbraith (1974), we show in the Appendix that

$$\sqrt{NT}\,(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(\rho \boldsymbol{c}, \boldsymbol{V}^{-1}), \tag{3.7}$$

where $\boldsymbol{c} = (c_1, c_2, \ldots, c_P)'$ and $c_p = c_{P-p+1} = -(1-\gamma_1-\gamma_2-\cdots-\gamma_{p-1}+\gamma_{P-p+1}+\cdots+\gamma_p)$ for $p = 1, \ldots, \lceil P/2 \rceil$. (This expression is also implicit in the result of Lee 2012.) Thus, a bias-corrected estimator is

$$\widetilde{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}} - \frac{\widehat{\boldsymbol{c}}}{T}, \tag{3.8}$$

which does not require estimating ω^2.

Corollary 1 (Single-equation bias correction). *Let Assumptions 1–3 hold. Then*

$$\sqrt{NT}\,(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{V}^{-1}\right)$$

as $N, T \to \infty$ *with* $N/T \to \rho^2$.

When $P = 1$, (3.8) reduces to the well-known correction (Nickell 1981, p. 1422)

$$\widetilde{\gamma} = \widehat{\gamma} + \frac{1 + \widehat{\gamma}}{T},$$

and Corrolary 1 yields

$$\sqrt{NT}\left(\widehat{\gamma} + \frac{1 + \widehat{\gamma}}{T} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, 1 - \gamma^2\right),$$

which agrees with Hahn and Kuersteiner (2002, p. 1645).

Finally, besides $\boldsymbol{\Gamma}$, estimation of the impulse-response functions is of interest. Recall that

$$\boldsymbol{\Phi}_h = \boldsymbol{\Gamma}_1 \boldsymbol{\Phi}_{h-1} + \boldsymbol{\Gamma}_2 \boldsymbol{\Phi}_{h-2} + \cdots + \boldsymbol{\Gamma}_P \boldsymbol{\Phi}_{h-P},$$

with $\boldsymbol{\Phi}_h = \boldsymbol{0}$ if $h < 0$ and $\boldsymbol{\Phi}_0 = \boldsymbol{I}_M$ (see, e.g., Hamilton 1994, Eq. 10.1.19). Clearly, a plug-in estimator of $\boldsymbol{\Phi}_h$ based on the WG-OLS estimator will suffer from asymptotic bias. However, Theorem 1 directly implies that the bias-corrected estimator based on the recursion

$$\widetilde{\boldsymbol{\Phi}}_h = \widetilde{\boldsymbol{\Gamma}}_1 \widetilde{\boldsymbol{\Phi}}_{h-1} + \widetilde{\boldsymbol{\Gamma}}_2 \widetilde{\boldsymbol{\Phi}}_{h-2} + \cdots + \widetilde{\boldsymbol{\Gamma}}_P \widetilde{\boldsymbol{\Phi}}_{h-P}$$

will be asymptotically unbiased if $N, T \to \infty$ with $N/T \to \rho^2$. Asymptotically valid inference on the impulse-responses can then be performed. To state the result, let

$$\boldsymbol{G}_h = \sum_{s=1}^{h} \boldsymbol{\Phi}_{s-1} \otimes (\boldsymbol{\Phi}'_{h-s},\ \boldsymbol{\Phi}'_{h-s-1},\ \cdots,\ \boldsymbol{\Phi}'_{h-s-P+1}).$$

THEOREM 2 (Bias-corrected impulse-response functions). *Let Assumptions 1–3 hold. Then*

$$\sqrt{NT}\,\mathrm{vec}\,(\widetilde{\boldsymbol{\Phi}}'_h - \boldsymbol{\Phi}'_h) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{G}_h(\boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{G}'_h\right)$$

as $N, T \to \infty$ *with* $N/T \to \rho^2$.

4. SIMULATIONS

We present simulation results for a two-equation two-lag autoregressive model with

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} .75 & -.20 \\ .20 & .25 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_2 = \begin{pmatrix} .20 & -.10 \\ .10 & .05 \end{pmatrix}, \qquad \boldsymbol{\Omega} = \begin{pmatrix} 1 & .2 \\ .2 & 1 \end{pmatrix},$$

errors generated as $\boldsymbol{v}_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, and various panel sizes. We start the time series processes in the distant past, so the initial observations are drawn from their steady-state distribution. The results are invariant to the choice of $\boldsymbol{\alpha}_i$.

We computed point estimates and confidence intervals for the elements of the coefficient matrices, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, and for the impulse-response functions

$$\phi_{mn}(h) = \frac{\partial y_{i(t+h)m}}{\partial v_{itn}},$$

using WG-OLS and BC-WG-OLS estimates. For brevity we do not consider generalized method-of-moment estimators. They are not designed for problems where T/N is not very close to zero, and so are not well suited for most macroeconomic data sets; see also Juessen and Linnemann (2010, 2012).

Table 1 reports the biases, standard deviations, and coverage rates of 95% confidence intervals centered at point estimates (WG-OLS in the top half of the table, BC-WG-OLS in the bottom half), computed from 10,000 Monte Carlo replications. The (m, n)th element of $\boldsymbol{\Gamma}_p$ is denoted as γ_{pmn}. Clearly, the bias of the within-group estimator varies substantially across the coefficients and is non-negligible relative to the standard deviation. Consequently, the confidence intervals centered at the within-group estimator suffer

Table 1. Coefficient matrices

N	T	γ_{111}	γ_{121}	γ_{112}	γ_{122}	γ_{211}	γ_{221}	γ_{212}	γ_{222}
					WG-OLS: BIAS				
25	25	$-.0557$	.0006	$-.0230$	$-.0256$	.0089	$-.0459$	.0376	$-.0367$
50	50	$-.0237$	$-.0012$	$-.0101$	$-.0137$	.0040	$-.0211$	.0186	$-.0171$
100	100	$-.0109$	$-.0008$	$-.0046$	$-.0070$	.0019	$-.0102$	.0093	$-.0083$
200	200	$-.0052$	$-.0004$	$-.0023$	$-.0034$	.0009	$-.0051$	.0045	$-.0040$
					WG-OLS: STD				
25	25	.0432	.0416	.0443	.0405	.0420	.0426	.0446	.0408
50	50	.0210	.0205	.0220	.0203	.0204	.0208	.0220	.0201
100	100	.0102	.0101	.0111	.0100	.0100	.0102	.0109	.0100
200	200	.0051	.0051	.0054	.0050	.0050	.0051	.0054	.0050
					WG-OLS: COVERAGE				
25	25	.6991	.9391	.9131	.8968	.9338	.7825	.8578	.8429
50	50	.7734	.9434	.9222	.8905	.9428	.8143	.8606	.8588
100	100	.8084	.9490	.9252	.8939	.9433	.8234	.8635	.8683
200	200	.8183	.9466	.9331	.8919	.9474	.8281	.8678	.8740
					BC-WG-OLS: BIAS				
25	25	$-.0175$	.0048	$-.0047$	.0008	.0024	$-.0078$	.0034	$-.0054$
50	50	$-.0043$	.0010	$-.0012$	.0001	.0006	$-.0017$	.0009	$-.0013$
100	100	$-.0011$	.0004	$-.0002$	.0000	.0002	$-.0004$	.0003	$-.0004$
200	200	$-.0003$	.0001	$-.0001$	.0001	.0000	$-.0002$	.0000	.0000
					BC-WG-OLS: STD				
25	25	.0430	.0417	.0445	.0413	.0421	.0426	.0458	.0417
50	50	.0209	.0206	.0220	.0205	.0204	.0208	.0223	.0203
100	100	.0102	.0101	.0111	.0101	.0100	.0102	.0110	.0101
200	200	.0051	.0051	.0054	.0050	.0050	.0051	.0055	.0050
					BC-WG-OLS: COVERAGE				
25	25	.8719	.9059	.9071	.9072	.9380	.9331	.9369	.9374
50	50	.9125	.9265	.9245	.9247	.9483	.9428	.9454	.9483
100	100	.9322	.9408	.9357	.9380	.9476	.9469	.9459	.9478
200	200	.9416	.9437	.9443	.9450	.9521	.9516	.9498	.9530

from substantial undercoverage and the distortion persists as the sample size grows. The bottom half of the table shows that much of the bias of the within-group estimator is successfully removed by the bias correction. Furthermore, the correction has very little effect on the variance of the estimator: the standard deviation of the corrected estimator is almost identical to that of the uncorrected estimator. The bias correction leads to confidence intervals with substantially improved coverage rates. Therefore, the conclusions from Table 1 support our theoretical findings summarized in Lemma 1 and Theorem 1.

Figure 1. Impulse-response functions

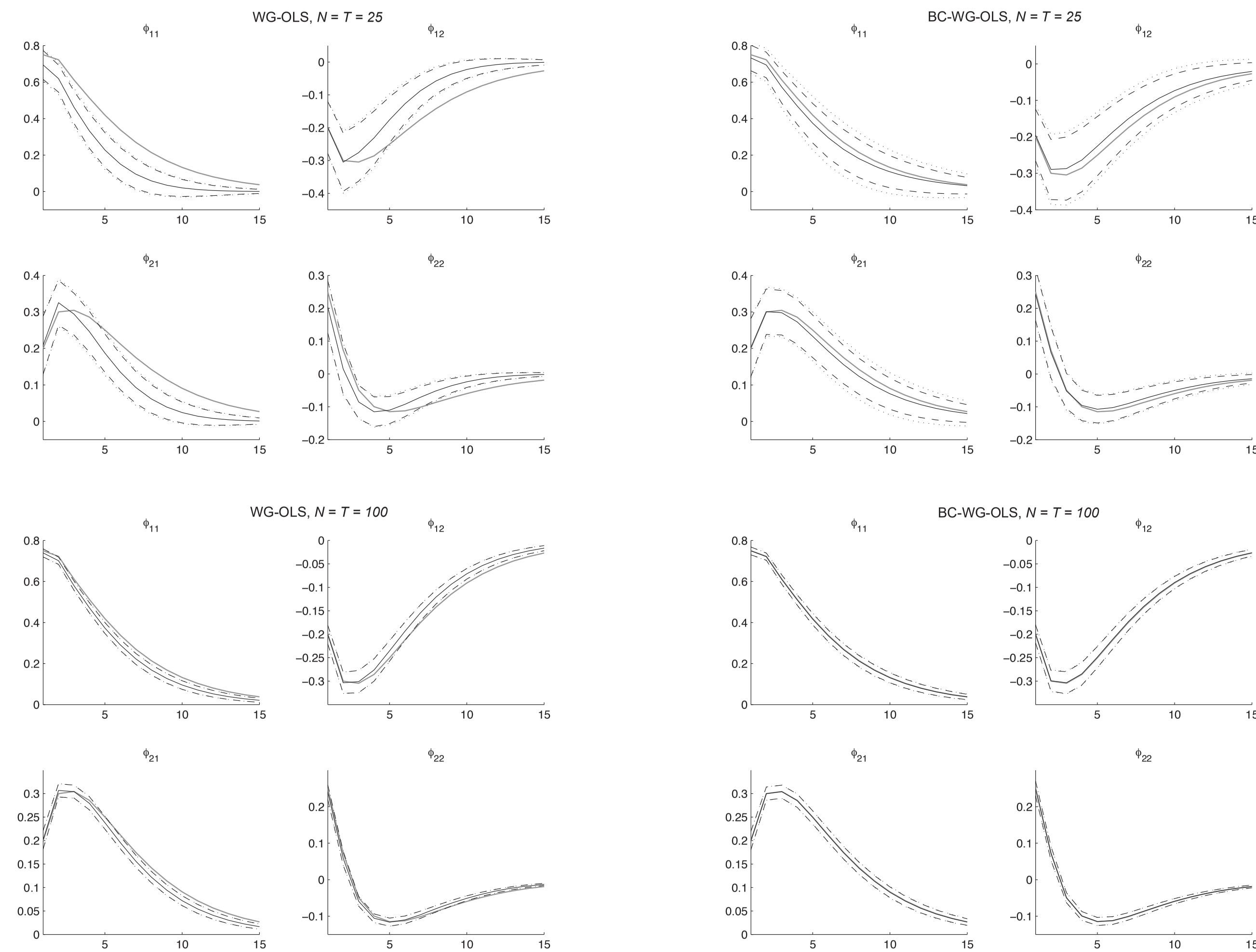

Figure 1 summarizes the simulation results for the impulse-response function estimates for $N = T = 25$ (the plots in the upper half of the figure) and $N = T = 100$ (bottom half), based on the within-group estimator (left half) and on the bias-corrected estimator (right half). Each plot corresponds to a particular impulse-response function, ϕ_{mn}, as indicated, and displays the true ϕ_{mn} (solid grey line), the average (across the Monte Carlo replications) of the estimated ϕ_{mn} (solid black line), and pointwise 95%-confidence bands constructed from the Monte Carlo standard deviation (dotted black lines) and from the estimated standard error based on Theorem 2 (dashed black lines). The four plots in the upper left part of the figure, corresponding to $N = T = 25$, show that least-squares may introduce substantial bias in the impulse-response function estimates, of the same order as the width of the confidence bands. As evidenced by the four plots in the upper right part of the figure, the bias-corrected impulse-response function estimates suffer from much less bias. The remaining bias is small relative to the standard error. These findings are in line with Theorem 2. Moving to the bottom half of the figure, where $N = T = 100$, we see that the ratio of bias to standard error of the least-squares estimator of ϕ_{mn} persists, with confidence bounds settling around the wrong curve. In contrast, as shown by the four plots in the bottom right part of the figure, the corrected impulse-response function estimates are asymptotically unbiased.

5. CONCLUSION

We derived an approximate bias correction for panel vector autoregressions with fixed effects. It extends the correction of Hahn and Kuersteiner (2002) to higher-order panel vector autoregressions and should be useful for macro-panel data sets, where the number of time periods is often of the same order as the number of cross-sectional units (e.g., countries or states). The correction is given in closed form and, therefore, straightforward to implement.

APPENDIX

Proof of Lemma 1. A standard argument (Phillips and Moon 1999; Hahn and Kuersteiner 2000) yields the asymptotic-normality result of the within-group least-squares estimator centered around its probability limit. It therefore suffices to calculate the bias term in the limit distribution. Given the sampling-error representation

$$\sqrt{NT}\,\mathrm{vec}\big(\widehat{\boldsymbol{\varGamma}}-\boldsymbol{\varGamma}\big)=\left(\boldsymbol{I}_M\otimes\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i\boldsymbol{M}\boldsymbol{X}_i'\right)^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathrm{vec}(\boldsymbol{X}_i\boldsymbol{M}\boldsymbol{\varUpsilon}_i')\right),$$

the bias follows as $\rho\,(\boldsymbol{I}_M\otimes\boldsymbol{\varSigma}^{-1})\,\mathrm{vec}\boldsymbol{B}$ where $\boldsymbol{B}=\lim_{T\to\infty}E\,[\boldsymbol{X}_i\boldsymbol{M}\boldsymbol{\varUpsilon}_i']$ is the large-T approximation to the bias in the normal equations, as in Hahn and Kuersteiner (2002). $\boldsymbol{B}$ consists of the $M\times M$ matrices $\boldsymbol{B}_1,\ldots,\boldsymbol{B}_P$, where

$$\begin{aligned}\boldsymbol{B}_j&=\lim_{T\to\infty}E\left[\sum_{t=1}^{T}\left(\boldsymbol{y}_{it-j}-\frac{1}{T}\sum_{t'=1}^{T}\boldsymbol{y}_{it'-j}\right)\boldsymbol{\upsilon}_{it}'\right]=-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\boldsymbol{y}_{it'-j}\boldsymbol{\upsilon}_{it}'\right]\\&=-\lim_{T\to\infty}E\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{k=0}^{\infty}\boldsymbol{\varPhi}_k\boldsymbol{\upsilon}_{it'-j-k}\boldsymbol{\upsilon}_{it}'\right]=-\sum_{k=0}^{\infty}\boldsymbol{\varPhi}_k\,\boldsymbol{\varOmega}\\&=-(\boldsymbol{I}_M-\boldsymbol{\varGamma}_1-\cdots-\boldsymbol{\varGamma}_P)^{-1}\,\boldsymbol{\varOmega},\end{aligned}$$

which does not depend on j. Hence, $\boldsymbol{B}=-(\boldsymbol{\iota}_P\otimes(\boldsymbol{I}_M-\boldsymbol{\varGamma}_1-\cdots-\boldsymbol{\varGamma}_P)^{-1})\,\boldsymbol{\varOmega}$, as defined in the main text. □

Proof of (3.7). In the single-equation case, Lemma 1 and Theorem 1 hold with

$$\boldsymbol{\varSigma}^{-1}\boldsymbol{b}=-\frac{\boldsymbol{V}^{-1}\,\boldsymbol{\iota}_P}{1-\gamma_1-\cdots-\gamma_P}.$$

Galbraith and Galbraith (1974, p. 70) showed that $\boldsymbol{V}^{-1}=\boldsymbol{A}\boldsymbol{A}'-\boldsymbol{H}'\boldsymbol{H}$ where

$$\boldsymbol{A}=-\begin{pmatrix}\gamma_0&0&\cdots&0\\\gamma_1&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\\gamma_{P-1}&\cdots&\gamma_1&\gamma_0\end{pmatrix},\qquad\boldsymbol{H}=-\begin{pmatrix}\gamma_P&\gamma_{P-1}&\cdots&\gamma_1\\0&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&\gamma_{P-1}\\0&\cdots&0&\gamma_P\end{pmatrix},$$

and $\gamma_0=-1$. Now, $\boldsymbol{A}'\boldsymbol{\iota}_P=\boldsymbol{a}+(1-\gamma_1-\cdots-\gamma_P)\boldsymbol{\iota}_P$ and $\boldsymbol{H}\boldsymbol{\iota}_P=\boldsymbol{h}+(1-\gamma_1-\cdots-\gamma_P)\boldsymbol{\iota}_P$, where

$$\boldsymbol{a}=\begin{pmatrix}\gamma_P\\\gamma_{P-1}+\gamma_P\\\vdots\\\gamma_1+\cdots+\gamma_P\end{pmatrix},\qquad\boldsymbol{h}=\begin{pmatrix}\gamma_0\\\gamma_0+\gamma_1\\\vdots\\\gamma_0+\cdots+\gamma_{P-1}\end{pmatrix}.$$

Noting that $\boldsymbol{A}\boldsymbol{a}-\boldsymbol{H}'\boldsymbol{h}=\boldsymbol{0}$, we find $\boldsymbol{\varSigma}^{-1}\boldsymbol{b}=-(\boldsymbol{A}-\boldsymbol{H}')\,\boldsymbol{\iota}_P=\boldsymbol{c}$. □

Proof of Theorem 2. The result follows along the same lines as in, e.g., Hamilton (1994, Section 11.7) from Theorem 1 by an application of the delta method. □

REFERENCES

Alesina, A., S. Ardagna, R. Perotti, and F. Schiantarelli (2002). Fiscal policy, profits, and investment. *American Economic Review 92*, 571–589.

Alonso-Borrego, C. and M. Arellano (1999). Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business and Economic Statistics 17*, 36–49.

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

Asdrubali, P. and S. Kim (2004). Dynamic risk-sharing in the United States and Europe. *Journal of Monetary Economics 51*, 809–836.

Attanasio, O. P., L. Picci, and A. E. Scorcu (2000). Savings, growth, and investment: A macroeconomic analysis using a panel of countries. *Review of Economics and Statistics 82*, 182–211.

Canova, F. and M. Ciccarelli (2013). Panel vector autoregressive models: A survey. In T. B. Fomby, L. Kilian, and A. Murphy (Eds.), *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*, Volume 32 of *Advances in Econometrics*, pp. 205–246. Emerald Group Publishing Limited.

Fajnzylber, P., D. Lederman, and N. Loayza (2002). What causes violent crime? *European Economic Review 46*, 1323–1357.

Galbraith, R. F. and J. I. Galbraith (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability 11*, 63–71.

Hahn, J. and G. Kuersteiner (2000). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. MIT Department of Economics Working Paper 01-17.

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

Hahn, J. and H. R. Moon (2006). Reducing bias of MLE in a dynamic panel model. *Econometric Theory 22*, 499–512.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

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

Juessen, F. and L. Linnemann (2010). Estimating panel VARs from macroeconomic data: Some Monte Carlo evidence and an application to OECD public spending shocks. SFB Discussion Paper 823.

Juessen, F. and L. Linnemann (2012). Markups and fiscal transmission in a panel of OECD countries. *Journal of Macroeconomics 34*, 674–686.

Lee, Y. (2012). Bias in dynamic panel models under time series misspecification. *Journal of Econometrics 169*, 54–60.

Loayza, N., K. Schmidt-Hebbel, and L. Servén (2000). What drives private saving across the world? *Review of Economics and Statistics 82*, 166–181.

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

Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*, 1057–1111.

Rousseau, P. and P. Wachtel (2000). Equity markets and growth: Cross-country evidence on timing and outcomes. *Journal of Banking and Finance 24*, 1933–1957.

Sims, C. A. (1972). Money, income, and causality. *American Economic Review 62*(4), 540–552.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Tagkalakis, A. (2008). The effects of fiscal policy on consumption in recessions and expansions. *Journal of Public Economics 92*, 1486–1508.