Bias-corrected estimation of panel vector autoregressions

Geert Dhaene, Koen Jochmans

SciencesPo
DEPARTMENT OF ECONOMICS

Discussion paper 2015-06

BIAS-CORRECTED ESTIMATION OF PANEL VECTOR AUTOREGRESSIONS

Geert Dhaene
Koen Jochmans

Sciences Po Economics Discussion Papers

Bias-corrected estimation of panel vector autoregressions

GEERT DHAENE* AND KOEN JOCHMANS†

Draft: June 12, 2015

We derive bias-corrected least-squares estimators of panel vector autoregressions with fixed effects. The correction is straightforward to implement and yields an estimator that is asymptotically unbiased under asymptotics where the number of time series observations grows at the same rate as the number of cross-sectional observations. This makes the estimator well suited for most macroeconomic data sets. Simulation results show that the estimator yields substantial improvements over within-group least-squares estimation. We illustrate the bias correction in a study of the relation between the unemployment rate and the economic growth rate at the U.S. state level.
JEL Classification: C33
Keywords: bias correction, fixed effects, panel data, vector autoregression

Introduction

Vector autoregressions are a standard tool in macroeconometrics since the work of Sims (1972, 1980). Stock and Watson (2001) provide a survey and critical assessment. A growing literature exploits the availability of large longitudinal data sets to fit panel versions of vector autoregressive models. A recent overview of this literature is available in Canova and Ciccarelli (2013).

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

* University of Leuven, Department of Economics, Naamsestraat 69, B-3000 Leuven, Belgium; geert.dhaene@kuleuven.be.
† Sciences Po, Département d'économie, 28 rue des Saints Pères, 75007 Paris, France; koen.jochmans@sciencespo.fr.

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 neglected in the empirical literature. At the same time, the GMM estimator of Holtz-Eakin, Newey and Rosen (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; Hsiao and Zhang 2015).

In this paper we consider bias-corrected estimation under asymptotics where N/T converges to a constant. Such an asymptotic approximation is well suited for many macroeconomic data sets, where T cannot reasonably be considered small relative to N. Our estimator consists of a simple additive correction to the (within-group) least-squares estimator and extends the work of Hahn and Kuersteiner (2002). The correction removes the asymptotic bias from the limit distribution of the least-squares estimator while leaving its asymptotic variance unchanged. The resulting estimator is straightforward to implement. We also discuss bias-corrected estimation of impulse-reponse functions, which are often the ultimate parameters of interest.

In Section I we present the model of interest. In Section II we show that the limit distribution of the within-group least-squares estimator is incorrectly centered and derive the bias. The bias-corrected estimator is then obtained by subtracting an estimator of this bias from the within-group least-squares estimator. It is asymptotically unbiased. We also show that the correction immediately yields an estimator of the impulse-reponse function that is asymptotically unbiased. In Section III we present Monte Carlo simulations. In Section IV we investigate the relationship between the unemployment rate and the economic growth rate at the U.S. state level. The appendix contains technical details.

I. 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 $y_{it1}, \ldots, y_{itM}$, where t ranges from

$1-P$ to T. The behavior of $\boldsymbol{y}_{it} = (y_{it1}, y_{it2}, \ldots, y_{itM})'$ is described by the Pth order vector autoregression

$$\boldsymbol{y}_{it} = \boldsymbol{\Gamma}_1 \boldsymbol{y}_{it-1} + \boldsymbol{\Gamma}_2 \boldsymbol{y}_{it-2} + \cdots + \boldsymbol{\Gamma}_P \boldsymbol{y}_{it-P} + \boldsymbol{\epsilon}_{it}, \tag{1.1}$$

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

Equation (1.1) can be extended to include time dummies, one for each period. The presence of time dummies changes the least-squares estimator but leaves the bias correction 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*

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

all 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)**.** $\boldsymbol{\upsilon}_{it}$ *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 $p = 1, \ldots, P$.

The moment conditions in Assumption 2 ensure regular asymptotic behavior of the least-squares estimator and are standard.

For simplicity, we will maintain the assumption that the error vectors $\boldsymbol{\upsilon}_{it}$ are independent and identically distributed.

Assumption 3 (Errors)**.** $\boldsymbol{\upsilon}_{it}$ *is independent and identically distributed across* i *and* t*,* $E[\boldsymbol{\upsilon}_{it}] = 0$*, and* $\boldsymbol{\Omega} = E[\boldsymbol{\upsilon}_{it}\boldsymbol{\upsilon}_{it}'] < +\infty$*.*

Independence across time can be relaxed to allow for dependence between $\boldsymbol{\upsilon}_{it}$ and $\boldsymbol{\upsilon}_{it-p}$ 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. Any intertemporal dependence would come into play only in the asymptotic variance of the within-group least-squares estimator.

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

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

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

$$\left(\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 L - \cdots - \boldsymbol{\Gamma}_P L^P\right)^{-1} = \sum_{k=0}^{\infty}\boldsymbol{\Phi}_k L^k. \tag{1.2}$$

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).

II. Bias-corrected estimation

A. *Least-squares estimator*

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 (1.1) as

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

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

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

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 solves the normal equations

$$\sum_{i=1}^{N} \left(\boldsymbol{Y}_i - \boldsymbol{\Gamma}' \boldsymbol{X}_i \right) \boldsymbol{M} \boldsymbol{X}'_i = \boldsymbol{0},$$

where $\boldsymbol{M} = \boldsymbol{I}_T - \boldsymbol{\iota}_T \boldsymbol{\iota}'_T$ is the usual demeaning matrix that sweeps out the fixed effects.

The WG-OLS estimator is

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

Given $\widehat{\boldsymbol{\Gamma}}$ we may construct an estimator of the $\boldsymbol{\Phi}_k$ defined in (1.2), and with it an estimator of the impulse-response functions.

It is well-known that $\widehat{\boldsymbol{\Gamma}}$ is inconsistent under asymptotics where T is treated as fixed (see, e.g., Nickell 1981). Holtz-Eakin, Newey and Rosen (1988) proposed GMM estimators that are consistent and asymptotically normal under fixed-T asymptotics. However, these estimators are known to be asymptotically biased when T is allowed to grow at the same rate as N (Alvarez and Arellano 2003; Hsiao and Zhang 2015), and so are not attractive in situations where T/N is not close to zero.

On the other hand, even though $\widehat{\boldsymbol{\Gamma}}$ is consistent under asymptotics where $N, T \to \infty$ jointly, its limit distribution has a non-zero asymptotic bias unless $N/T \to 0$ (see Hahn and Kuersteiner 2002 and also below). This suggests that inference based on the WG-OLS estimator will tend to be unreliable unless T is substantially larger than N.

B. *Bias-corrected least-squares estimator*

We consider asymptotically unbiased estimation of $\boldsymbol{\Gamma}$ under asymptotics where $N/T \to \rho^2 < \infty$ as $N, T \to \infty$, extending the applicability of the approach of Hahn and Kuersteiner (2002).

To derive the bias of the WG-OLS estimator, start from the sampling-error representation

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

For any vector $\boldsymbol{a}_{it}$, let $\boldsymbol{a}_{i\cdot} = T^{-1}\sum_{t=1}^{T} \boldsymbol{a}_{it}$. Then

$$\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{X}_i' \xrightarrow{p} \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E[(\boldsymbol{x}_{it} - \boldsymbol{x}_{i\cdot})\,(\boldsymbol{x}_{it} - \boldsymbol{x}_{i\cdot})'] = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is the variance matrix of $\boldsymbol{x}_{it}$ in the limit $t \to \infty$. Further,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathrm{vec}\,(\boldsymbol{X}_i \boldsymbol{M} \boldsymbol{\Upsilon}_i' - E[\boldsymbol{X}_i \boldsymbol{M} \boldsymbol{\Upsilon}_i']) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}).$$

However, the within transformation introduces a bias in the normal equations. In particular, $E\left[\boldsymbol{X}_i \boldsymbol{M} \boldsymbol{\Upsilon}_i'\right] = O(1)$, which implies that the bias of $\widehat{\boldsymbol{\Gamma}}$ is $O(T^{-1})$. In the Appendix we show that, as $T \to \infty$,

$$E\left[\boldsymbol{X}_i \boldsymbol{M} \boldsymbol{\Upsilon}_i'\right] = \boldsymbol{B} + o(1)$$

where

$$\boldsymbol{B} = -\left(\boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_P)^{-1}\right) \boldsymbol{\Omega}. \tag{2.1}$$

Therefore, letting $\boldsymbol{b} = \operatorname{vec} \boldsymbol{B} = -\left(\boldsymbol{I}_M \otimes \boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_P)^{-1}\right) \operatorname{vec} \boldsymbol{\Omega}$, we have the following result.

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

$$\sqrt{NT} \operatorname{vec}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}) \xrightarrow{d} \mathcal{N}\left(\rho\,(\boldsymbol{I}_M \otimes \boldsymbol{\Sigma}^{-1})\,\boldsymbol{b}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}^{-1}\right) \tag{2.2}$$

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

In Theorem 1, $\boldsymbol{b}$ represents the non-negligible part of the bias in the normal equations that causes conventional confidence intervals, constructed around the WG-OLS estimator, to be incorrectly centered.

Equations (2.1)–(2.2) suggest correcting $\widehat{\boldsymbol{\Gamma}}$ by subtracting an estimate of its bias, which may be formed as a plug-in estimate using $\widehat{\boldsymbol{\Gamma}}$ itself. This yields the bias-corrected WG-OLS (BC-WG-OLS) estimator

$$\widetilde{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}} - \frac{\widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{B}}}{T}, \tag{2.3}$$

where

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{M} \boldsymbol{X}_i', \qquad \widehat{\boldsymbol{B}} = -(\boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \widehat{\boldsymbol{\Gamma}}_1 - \cdots - \widehat{\boldsymbol{\Gamma}}_P)^{-1})\,\widehat{\boldsymbol{\Omega}}, \tag{2.4}$$

with

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{NT}\sum_{i=1}^{N} \widehat{\boldsymbol{\Upsilon}}_i \boldsymbol{M} \widehat{\boldsymbol{\Upsilon}}_i', \qquad \widehat{\boldsymbol{\Upsilon}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{\Gamma}}' \boldsymbol{X}_i. \tag{2.5}$$

This estimator is asymptotically unbiased.

Theorem 2 (Bias-corrected least-squares estimator)**.** *Let Assumptions 1–3 hold. Then*

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

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

Theorem 2 shows that the asymptotic distribution of $\widetilde{\boldsymbol{\Gamma}}$ is correctly centered and has the same variance as that of $\widehat{\boldsymbol{\Gamma}}$. This result extends Hahn and Kuersteiner (2002, Theorem 2). One may expect that, under the additional asumption that $\boldsymbol{v}_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, in which case the within-group least-squares estimator coincides with the maximum-likelihood estimator, the bias-corrected estimator in (2.3) is asymptotically efficient, as in Hahn and Kuersteiner (2002, Corollary 1), although we do not pursue a formal proof here.

The estimator in (2.3)–(2.5) is constructed using a plug-in estimator of the bias based on the WG-OLS estimator. We may consider repeating the correction by iterating until convergence on

$$\widetilde{\boldsymbol{\Gamma}}^{(1)} = \widetilde{\boldsymbol{\Gamma}}, \qquad \widetilde{\boldsymbol{\Gamma}}^{(j+1)} = \widehat{\boldsymbol{\Gamma}} - \frac{\widehat{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{B}}^{(j)}}{T}, \qquad j = 1, 2, \ldots,$$

where $\widetilde{\boldsymbol{B}}^{(j)}$ is constructed as in (2.4)–(2.5) but with $\widehat{\boldsymbol{\Gamma}}$ replaced by $\widetilde{\boldsymbol{\Gamma}}^{(j)}$. Iterating the correction does not yield an asymptotic improvement over Theorem 2, but it may reduce the bias in finite samples since now $\boldsymbol{B}$ itself is estimated using an asymptotically unbiased plug-in estimator of $\boldsymbol{\Gamma}$.

C. Impulse-response functions

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 $\boldsymbol{\Phi}_h = \mathbf{0}$ 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 2 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

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

in the following theorem.

Theorem 3 (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 $N/T \to \rho^2$.

D. Single-equation correction

When $M = 1$ the model in (1.1) reduces to

$$y_{it} = \alpha_i + \gamma_1 y_{it-1} + \cdots + \gamma_P y_{it-P} + \upsilon_{it}, \qquad \upsilon_{it} \sim \text{i.i.d.}\,(0, \omega^2). \tag{2.6}$$

In this case a simpler, yet asymptotically equivalent, bias correction may be performed.

Write (2.6) 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 $t \to \infty$. Combining Theorem 1 with the expression for $\boldsymbol{V}^{-1}$ in Galbraith and Galbraith (1974, p. 70) we find

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

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 $p = 1, \ldots, \lceil P/2 \rceil$. Therefore, a bias-corrected estimator can be constructed as

$$\widetilde{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}} - \frac{\widehat{\boldsymbol{c}}}{T}, \tag{2.7}$$

which does not require estimating ω^2. We have the following result.

Theorem 4 (Single-equation bias correction)**.** *Let Assumptions 1–3 hold. Then*

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

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

When $P=1$, (2.7) reduces to the well-known correction (Nickell 1981, p. 1422)

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

and Theorem 4 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).

III. 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. Note that 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

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

Table 1 gives the bias, the standard deviation, and the coverage rate of 95% confidence intervals centered at the WG-OLS estimator, computed from $10,000$ Monte Carlo replications. Table 2 gives the corresponding results for the bias-corrected estimator. In the tables, we denote the (m,n)th element of $\boldsymbol{\Gamma}_p$ as γ_{mnp}.

TABLE 1—WG-OLS

N	T	γ_{111}	γ_{121}	γ_{112}	γ_{122}	γ_{211}	γ_{221}	γ_{212}	γ_{222}
					BIAS				
25	25	−.0557	.0006	−.0230	−.0256	.0089	−.0459	.0376	−.0367
50	50	−.0237	−.0012	−.0101	−.0137	.0040	−.0211	.0186	−.0171
75	75	−.0151	−.0010	−.0064	−.0091	.0026	−.0139	.0124	−.0113
100	100	−.0109	−.0008	−.0046	−.0070	.0019	−.0102	.0093	−.0083
200	200	−.0052	−.0004	−.0023	−.0034	.0009	−.0051	.0045	−.0040
					STD				
25	25	.0432	.0416	.0443	.0405	.0420	.0426	.0446	.0408
50	50	.0210	.0205	.0220	.0203	.0204	.0208	.0220	.0201
75	75	.0137	.0137	.0147	.0133	.0136	.0134	.0146	.0135
100	100	.0102	.0101	.0111	.0100	.0100	.0102	.0109	.0100
200	200	.0051	.0051	.0054	.0050	.0050	.0051	.0054	.0050
					COVERAGE				
25	25	.6991	.9391	.9131	.8968	.9338	.7825	.8578	.8429
50	50	.7734	.9434	.9222	.8905	.9428	.8143	.8606	.8588
75	75	.7918	.9478	.9252	.8933	.9443	.8248	.8604	.8669
100	100	.8084	.9490	.9252	.8939	.9433	.8234	.8635	.8683
200	200	.8183	.9466	.9331	.8919	.9474	.8281	.8678	.8740

Table 1 shows that the magnitude of the bias in the within-group estimator varies quite substantially across the coefficients. The results also illustrate that the bias is not negligible relative to the standard deviation. Indeed, the confidence intervals suffer from substantial undercoverage and the distortion persists as the sample size grows.

Table 2 shows that much of the bias in the within-group estimator is succesfully removed by the bias correction. Furthermore, the correction has very little effect on the variance of the estimator. Indeed, the standard deviation of the corrected estimator is almost identical to that of the uncorrected estimator. Also, correcting for the bias leads to confidence intervals with substantially improved coverage rates. Therefore, the conclusions from Table 2 support our theoretical findings summarized in Theorem 2.

Table 2—BC-WG-OLS

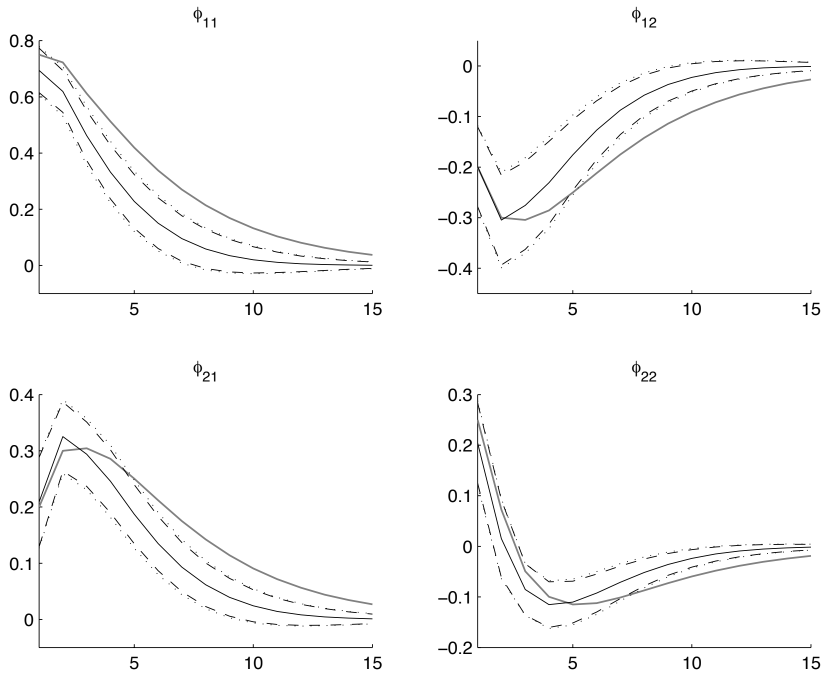
N	T	γ_{111}	γ_{121}	γ_{112}	γ_{122}	γ_{211}	γ_{221}	γ_{212}	γ_{222}
					BIAS				
25	25	$-.0175$	.0048	$-.0047$	.0008	.0024	$-.0078$	.0034	$-.0054$
50	50	$-.0043$	.0010	$-.0012$	.0001	.0006	$-.0017$	.0009	$-.0013$
75	75	$-.0020$	.0005	$-.0005$	.0001	.0004	$-.0009$	.0005	$-.0007$
100	100	$-.0011$	.0004	$-.0002$	.0000	.0002	$-.0004$	.0003	$-.0004$
200	200	$-.0003$	.0001	$-.0001$	.0001	.0000	$-.0002$	.0000	.0000
					STD				
25	25	.0430	.0417	.0445	.0413	.0421	.0426	.0458	.0417
50	50	.0209	.0206	.0220	.0205	.0204	.0208	.0223	.0203
75	75	.0136	.0137	.0147	.0134	.0136	.0134	.0148	.0136
100	100	.0102	.0101	.0111	.0101	.0100	.0102	.0110	.0101
200	200	.0051	.0051	.0054	.0050	.0050	.0051	.0055	.0050
					COVERAGE				
25	25	.8719	.9059	.9071	.9072	.9380	.9331	.9369	.9374
50	50	.9125	.9265	.9245	.9247	.9483	.9428	.9454	.9483
75	75	.9319	.9356	.9338	.9391	.9470	.9515	.9492	.9463
100	100	.9322	.9408	.9357	.9380	.9476	.9469	.9459	.9478
200	200	.9416	.9437	.9443	.9450	.9521	.9516	.9498	.9530

Figures 1 and 2 summarize the simulation results for the WG-OLS estimator and the BC-WG-OLS estimator, respectively, of the impulse-response functions for $N = T = 25$. The (m, n) plot of each figure contains the (true) impulse-response function ϕ_{mn} (solid grey line), the average (across the Monte Carlo replications) of the estimated functions (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 3 (dashed black lines).

The plots in Figure 1 demonstrate that estimating impulse-response functions using least-squares coefficient estimates introduces substantial bias. This bias is not negligible, as is evident from the confidence bands. As the plots in Figure 2 show, using bias-corrected estimates of the coefficient matrices yields estimated impulse-response functions that suffer from much less bias. The remaining bias is also small relative to the standard error. These findings are in line with the conclusions of Theorem 3

To show that the bias in the least-squares estimator persists Figures 3 and 4 provide plots of the impulse-response functions analogous to those in Figures 1

FIGURE 1. IMPULSE-RESPONSE FUNCTIONS, $N = T = 25$, WG-OLS

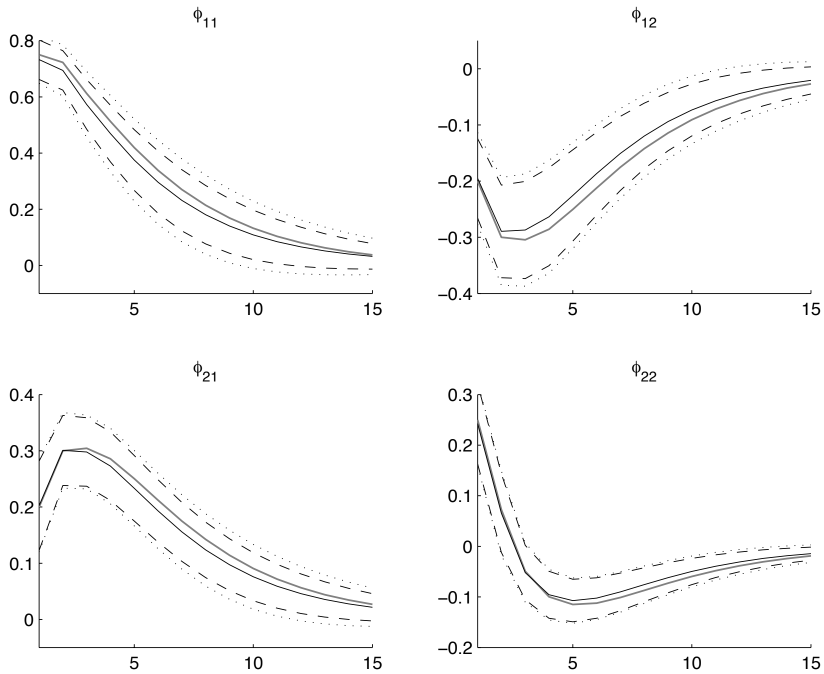


and 2, but now for panels with $N = T = 100$. The plots in Figure 3 show that the least-squares based confidence bounds settle around the wrong curve. In contrast, as evidenced by the plots in Figure 4, the corrected impulse-response functions are asymptotically unbiased.

IV. Empirical application

As an empirical illustration we investigate the relation between the unemployment rate and GDP growth by means of a vector autoregressive model estimated from a panel on U.S. states. The data span the period $2003 - 2013$, corresponding to 11 time-series observations for 51 cross-sectional units (all U.S. states and the District of Columbia). The data on the unemployment rate were taken from the U.S. Bureau of Labor Statistics and matched with those on gross state product

Figure 2. Impulse-response functions, $N = T = 25$, BC-WG-OLS



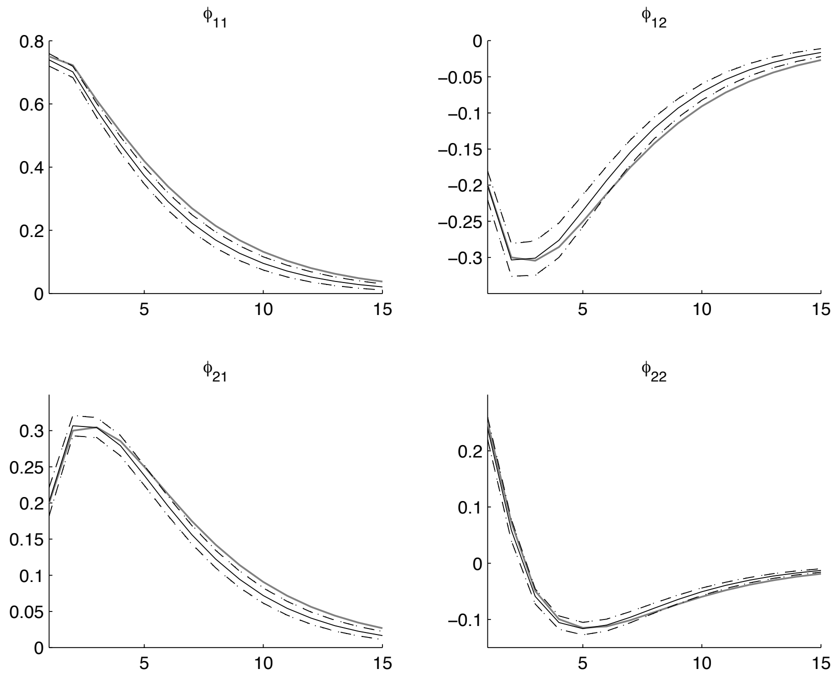
(in current U.S. dollars) from the U.S. Bureau of Economic Analysis.

We select the number of lags in the specification by information criteria. Table 3 contains the values of the Akaike (AIC), Schwarz (SIC), and Hannan-Quinn (HQC) information criteria (based on the bias-corrected point estimates). All three criteria point to a first-order autoregressive model.

Table 3—Information criteria

P	AIC	SIC	HQC
1	−16.0327	−15.9116	−16.1655
2	−14.9970	−14.8217	−15.3754
3	−13.7702	−13.6510	−14.5739
4	−12.3148	−12.4384	−13.8429
5	−10.5011	−11.1952	−13.2798

Figure 3. Impulse-response functions, $N = T = 100$, WG-OLS

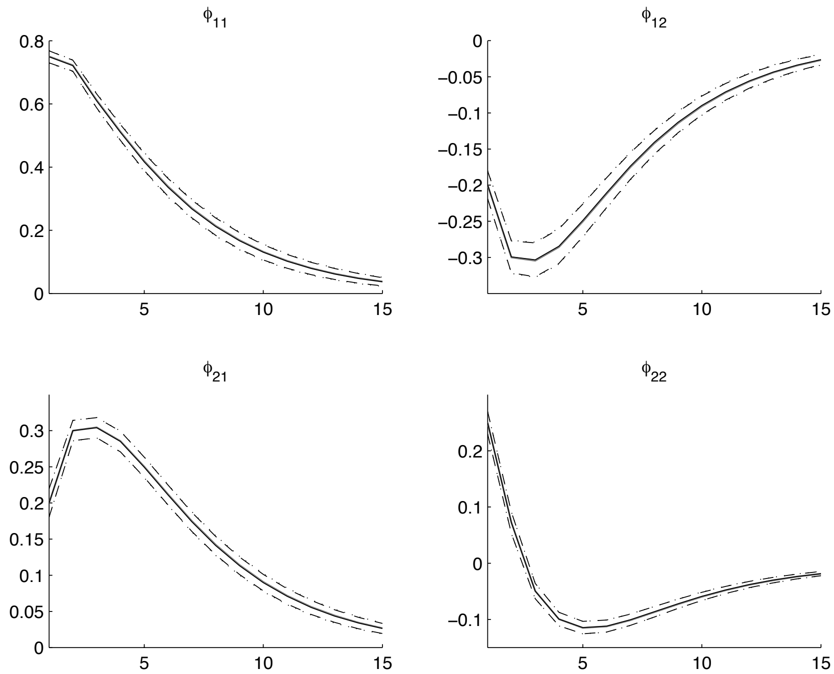


We thus estimate

$$\begin{pmatrix} \text{urate}_{it} \\ \text{grate}_{it} \end{pmatrix} = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} \text{urate}_{it-1} \\ \text{grate}_{it-1} \end{pmatrix} + \begin{pmatrix} \upsilon_{it1} \\ \upsilon_{it2} \end{pmatrix}.$$

Table 4 contains the point estimates of the coefficient matrix. The signs on the own autoregressive parameters, $\gamma_{11} > 0, \gamma_{22} > 0$, are as expected. The corrected point estimates are larger than those of WG-OLS, revealing greater persistence in the dynamics of both the unemployment rate and the growth rate of the gross state product. The coefficient γ_{12}, which relates to the impact of grate_{it-1} on urate_{it} is found to be negative and is similarly estimated by WG-OLS and BC-WG-OLS. In contrast, where WG-OLS picks up a large positive effect of urate_{it-1} on grate_{it}, bias-corrected estimation yields a small negative point estimate that

FIGURE 4. IMPULSE-RESPONSE FUNCTIONS, $N = T = 100$, BC-WG-OLS



is statistically insignificant.

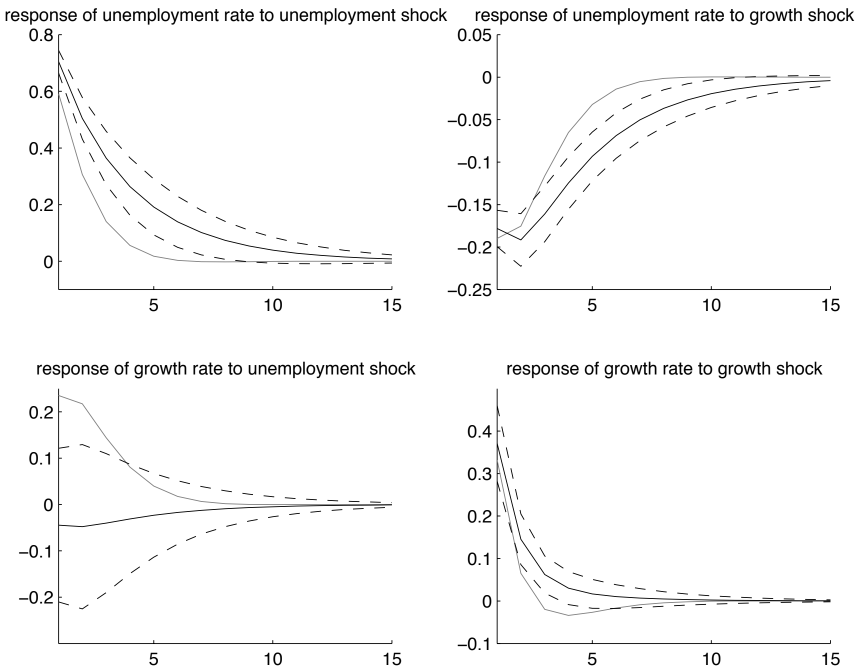
TABLE 4—COEFFICIENT ESTIMATES

	γ_{11}	γ_{12}	γ_{21}	γ_{22}
WG-OLS	0.5921	−0.1898	0.2351	0.3316
	(0.0280)	(0.0150)	(0.0873)	(0.0468)
BC-WG-OLS	0.7048	−0.1782	−0.0446	0.3708
	(0.0204)	(0.0109)	(0.0845)	(0.0453)

The shocks $\upsilon_{it1}, \upsilon_{it2}$ are found to be negatively correlated with an estimated correlation coefficient of $-.4275$. The standard deviations of the unemployment and growth shocks are estimated at .0077 and .0318, respectively.

Figure 5 contains the estimated impulse-response functions. The bias-corrected estimates (solid black line) are given along with 95%-confidence bands. The

FIGURE 5. IMPULSE-RESPONSE FUNCTIONS



uncorrected (least-squares) estimates (solid grey line) are also provided but we do not report confidence bands for clarity of the plots.

The two upper plots suggest that WG-OLS underestimates the persistence of the impact of shocks on the unemployment rate and the growth rate on future unemployment rates, with the point estimates lying almost completely outside the bias-corrected confidence bands. The difference between the least-squares estimates and the biased-corrected estimates is also economically important. For example, where least-squares predicts the unemployment rate to fully recover from an unemployment shock in about five years, the bias-corrected impulse-response function suggests it would take almost three times as long for the shock to die out. A similar qualitative difference is observed for the response of unemployment to growth rate shocks.

Turning to the two plots in the bottom of Figure 5, we observe that WG-OLS also tend to underestimate the duration of impact of a growth rate shock on future growth rates, although the underestimation is less severe here. Starting out in the positive range, the WG-OLS point estimate quickly becomes negative as the forecast horizon increases, eventually converging to zero after roughly ten years. The associated 95% confidence bands would not allow to statistically discriminate the negative point estimates from zero, however. On the other hand, failure to bias-correct leads to the largely spurious finding that positive shocks to the unemployment rate have a positive impact on the future growth rate. Even though the bias-corrected confidence interval is somewhat wide for the first few periods, it does not cover the impulse-response function estimated by least-squares in that period.

References

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

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

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.

Hsiao, C., and J. Zhang. 2015. "IV, GMM or likelihood approach to estimate dynamic panel models when N or T or both are large." *Journal of Econometrics*, 187: 312–322.

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.

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.

Stock, J. H., and M. A. Watson. 2001. "Vector autoregressions." *The Journal of Economic Perspectives*, 15: 101–115.

Appendix

Proof of Theorem 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.

As in Hahn and Kuersteiner (2002), $\boldsymbol{B} = \lim_{T\to\infty} E\left[\boldsymbol{X}_i \boldsymbol{M} \boldsymbol{\Upsilon}_i'\right]$ is the large-T approximation to the bias in the normal equations. Note that $\boldsymbol{B}$ consists of the $M \times M$ matrices $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_P$, where

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

which does not depend on j. Hence,

$$\boldsymbol{B} = -(\boldsymbol{\iota}_P \otimes (\boldsymbol{I}_M - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_P)^{-1})\,\boldsymbol{\Omega},$$

as claimed. This concludes the proof. □

Proof of Theorem 2. The result readily follows from the previous theorem. □

Proof of Theorem 3. The result follows along the same lines as in, e.g., Hamilton (1994, Section 11.7) from Theorem 2 by an application of the delta method. □

Proof of Theorem 4. In the single-equation case, Theorems 1 and 2 hold with

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

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,

$$\boldsymbol{A}'\boldsymbol{\iota}_P = \boldsymbol{a} + (1 - \gamma_1 - \cdots - \gamma_P)\boldsymbol{\iota}_P, \qquad \boldsymbol{H}\boldsymbol{\iota}_P = \boldsymbol{h} + (1 - \gamma_1 - \cdots - \gamma_P)\boldsymbol{\iota}_P,$$

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

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{b} = -(\boldsymbol{A} - \boldsymbol{H}')\,\boldsymbol{\iota}_P = \boldsymbol{c}.$$

This concludes the proof. $\square$