Asymptotic Distributions of Impulse Response Functions in Short Panel Vector Autoregressions

Bolong Cao* and Yixiao Sun†

Abstract

This paper establishes the asymptotic distributions of the impulse response functions in panel vector autoregressions with a fixed time dimension. It also proves the asymptotic validity of a bootstrap approximation to their sampling distributions. The autoregressive parameters are estimated using the GMM estimators based on the first differenced equations and the error variance is estimated using an extended analysis-of-variance type estimator. Contrary to the time series setting, we find that the GMM estimator of the autoregressive coefficients is not asymptotically independent of the error variance estimator. The asymptotic dependence calls for variance correction for the orthogonalized impulse response functions. Simulation results show that the variance correction improves the coverage accuracy of both the asymptotic confidence band and the studentized bootstrap confidence band for the orthogonalized impulse response functions.

Keywords: Asymptotic Distribution, Bootstrap, Nonorthogonalized Impulse Response Function, Orthogonalized Impulse Response Function, Panel Data, Vector Autoregressions.

JEL Classification Codes: C33, C53

*Email: caob@ohio.edu. Department of Economic, Ohio University, Athens, OH 45701.

†Correspondence to: Yixiao Sun, yisun@ucsd.edu, University of California, San Diego, Department of Economics, 9500 Gilman Drive, La Jolla, CA 92093-0508. We thank Cheng Hsiao, the coeditor, an associate editor and two anonymous referees for helpful comments that lead to considerable improvement of the paper. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-0752443.

1 Introduction

In this paper, we consider the panel vector autoregressions (VARs) where the cross sectional dimension (N) is large and the time series dimension (T) is short (typically less than 10). Panel VARs with a short T have been investigated, for examples, by Holtz-Eakin, Newey, and Rosen (1988) and Binder, Hsiao, and Pesaran (2005). While these papers focus on the estimation of the slope coefficients, our focus here is on the estimation of the impulse response functions (IRFs) and their confidence bands. Following the traditional panel data literature, we assume that the slope coefficients are the same across different cross sectional units and there is no cross sectional dependence after controlling for the fixed time effects. These two assumptions allow us to make good long-horizon forecasts, especially when the forecasting horizon is comparable to the time series length. This argument is consistent with the view of Binder, Hsiao, and Pesaran (2005) who use short panel VARs to infer the long run properties of the underlying time series.

For time series data, VAR models are typically estimated using the equation-by-equation OLS as it is asymptotically equivalent to the full system-of-equations estimator. For panel data VARs, the OLS estimator is inconsistent for a fixed T as $N \to \infty$. In this case, the VAR models are typically estimated using the Anderson-Hsiao (1981, 1982, hereafter AH) estimator or the Arellano-Bond (1991, hereafter AB) estimator. These estimators can be applied to each equation in the VAR system or the full system of equations. Holtz-Eakin, Newey, and Rosen (1988) and Arellano (2003, p.120) point out that it may be possible to improve the efficiency by estimating the system of equations jointly. We show that, under the model specification given below, the equation-by-equation AH or AB estimator is asymptotically equivalent to the corresponding system-of-equations estimator.

Impulse response analysis in the time series setting has been examined by Baillie (1987), Lütkepohl (1989, 1990), among others. However, there are two important differences between the time series case and the short panel case considered in this paper.

First, for time series VARs, the OLS estimator of the slope coefficients is asymptotically independent of the error variance estimator while for short panel VARs the AH or AB estimator of the slope coefficients depends on the error variance estimator even in the limit as $N \to \infty$ for a fixed T. Since the regressors are only sequentially exogenous, the demeaned regressors in the short panel VARs are correlated with the demeaned regression error. This nonzero correlation leads to the asymptotic dependence between the slope coefficient estimator and the error variance estimator.

Second, for time series VARs, the error variance estimator based on the estimated OLS residual is asymptotically equivalent to that based on the true error term. For short

panel VARs, the error variance estimator has different asymptotic distributions, depending on whether the error term is known or is based on estimated slope coefficients and fixed effects. In other words, the estimation uncertainty of the slope coefficients and the fixed effects affects the asymptotic distribution of the error variance estimator.

These two differences imply that the usual asymptotic results for orthogonalized impulse responses are not applicable to short panel VARs. One of the main contributions of the paper is to derive the asymptotic distributions of the orthogonalized IRFs for short panel VARs. The asymptotic distributions are obtained under the asymptotic specification that $N \to \infty$ with T fixed. Based on our asymptotic result, confidence bands for the IRFs can be easily constructed. Although impulse response analyses using short panels have been employed in the empirical applications, to the best of our knowledge, no study has reported confidence bands for orthogonalized IRFs that account for the estimation uncertainty of the error variance matrix. As a result, the reported confidence bands are often more narrow than they should be. This may lead to the finding of statistical significance that does not actually exist.

A further contribution of the paper is to establish the asymptotic validity of bootstrap confidence bands. Our simulation results show that bootstrap confidence bands usually provide more accurate coverage than the asymptotic analytical bands. In addition, the percentile-t bootstrap band that takes the dependence between the autoregressive coefficient estimator and the error variance estimator into account performs better than those that do not.

The rest of the paper is organized as follows. Section 2 describes the vector autoregression model for panel data and presents the standard GMM estimator of the slope coefficients and analysis-of-variance-type estimator of the error variance matrix. This section also establishes the joint asymptotic distribution of the slope coefficients estimator and the error variance estimator. Using these asymptotic results, we derive in Section 3 the asymptotic distributions of the orthogonalized and non-orthogonalized impulse response functions. We also prove the asymptotic validity of various bootstrap confidence bands. Section 4 provides some simulation evidence. The final section concludes. Proofs and a technical lemma are collected in the Appendix.

Throughout the paper, *vec* denotes the column stacking operator and *vech* is the corresponding operator that stacks only the elements on and below the main diagonal. As usual, the Kronecker product is denoted by $\otimes$, the commutation matrix $K_{m,n}$ is defined such that, for any $(m \times n)$ matrix G, $K_{m,n}vec(G) = vec(G')$, and the $m^2 \times (m(m+1))/2$ duplication matrix D_m is defined such that $D_m vech(F) = vec(F)$ for a symmetric $(m \times m)$ matrix F. Furthermore, $D_m^+ = (D_m'D_m)^{-1}D_m'$ and L_m is the $(m(m+1))/2 \times m^2$ elimination matrix

defined such that, for any $(m \times m)$ matrix F, $vech\,(F) = L_m vec(F)$. For matrix A, $||A||$ is the Euclidian norm of A. "$\Rightarrow$" denotes weak convergence and "$\equiv$" denotes distributional equivalence.

2 The Model and GMM Estimation

2.1 The Model

We consider an m-dimensional panel VAR(p) process:

$$y_{i,t} = \mu + A_1 y_{i,t-1} + ... + A_p y_{i,t-p} + \mu_i + u_{i,t} \tag{1}$$

for $t = 0, ..., T$ and $i = 1, 2, ..., N$ where $y_{i,t} = (y_{1,it}, ..., y_{m,it})'$, A_j are $(m \times m)$ coefficient matrices, μ_i is an $m \times 1$ vector of individual fixed effects, μ is an $m \times 1$ vector of intercepts, and $u_{i,t}$ is the error term. To simplify the discussion, we focus on balanced panel data sets. For each individual i, the time series starts at period 0 and ends at period T. Without the loss of generality, we assume that the initial values $y_{i,-1}, ..., y_{i,-p}$ are observed. We make the following assumption.

Assumption 1. $u_{i,t}$ is independently and identically distributed across i and t with

$$E(u_{i,t}|y_{i,t-1}, ..., y_{i,-p}) = 0 \text{ for } 0 \leq t \leq T$$

and

$$E\left(u_{i,t}u'_{j,s}|y_{i,t-1}, ..., y_{i,-p}\right) = \begin{cases} \Sigma, & i = j \text{ and } t = s \\ 0, & \text{otherwise} \end{cases}, \text{ for } 0 \leq t \leq s \leq T \tag{2}$$

where Σ is a positive definite matrix.

The model is the same as that considered by Binder, Hsiao, and Pesaran (2005). We do not parametrize the initial conditions for the VAR model as the asymptotic properties of the GMM estimators used in this paper do not reply on any parametric specification of $y_{i,-1}, ..., y_{i,-p}$. This is an advantage of the GMM estimators as compared to the quasi maximum likelihood (QML) estimator in Binder, Hsiao, and Pesaran (2005). In their fixed effects specification, they assume that the initial observations $y_{i,-1}, ..., y_{i,-p}$ are generated according to

$$y_{i,t} = (I - A_1 - ... - A_p)^{-1}(\mu + \mu_i) + \xi_{i,t} \text{ for } -p \leq t \leq -1$$

where the initial deviations $\xi_{i,t}$ are iid across i and t with zero mean and constant variance matrix (see their assumption G3). This homogeneity assumption may help improve the asymptotic efficiency of the QML estimator but will lead to inconsistency when it is violated.

Here we maintain minimal assumptions without parametrizing the initial conditions and focus only on the GMM type estimators. The standard GMM estimator considered below is widely used in empirical applications, see, for example, Love and Zicchino (2002) and Gilchrist, Himmelberg and Huberman (2005).

We can include the time fixed effects in the model so that

$$y_{i,t} = \mu + A_1 y_{i,t-1} + ... + A_p y_{i,t-p} + \mu_i + \lambda_t + u_{i,t}. \tag{3}$$

We have done so in a previous version of this paper. In this case, we can remove λ_t by taking out the cross sectional average. All of our results remain valid for the above model. To simplify the notation, we focus on the VAR model in (1).

2.2 Panel GMM Estimator and its Asymptotic Distribution

It is well-known that, due to the correlation between the fixed effect μ_i and the regressors, the OLS estimator of A_j based on equation (1) is inconsistent when T is small. To remove the fixed individual effect, we take the first difference of equation (1), leading to

$$\Delta y_{i,t} = A_1 \Delta y_{i,t-1} + ... + A_p \Delta y_{i,t-p} + \Delta u_{i,t},\ t = 1, ..., T.$$

The OLS estimator based on the first differenced equation is still inconsistent because $\Delta u_{i,t}$ is correlated with $\Delta y_{i,t-1}$. The standard GMM estimators of AH and AB employ instruments that are orthogonal to $\Delta u_{i,t}$. Additional nonlinear moment conditions implied by the homoscedasticity assumption in (2) are considered in Ahn and Schmidt (1995) and Binder, Hsiao, and Pesaran (2005). We provide similar results to this paper for the Ahn and Schmidt estimator in a previous version of this paper. In what follows, we will mainly focus on the AB estimator since it is easy to implement as the underlying moment conditions are linear in parameters. It has also been a standard practice to employ the AB estimator in empirical studies. In addition, for dynamic panel data models, the latest version of STATA contains only GMM estimators with linear moment conditions, the leading case of which is the AB estimator. Furthermore, in his seminal monograph, Hsiao (2003) discusses only the AB estimator for panel vector autoregressive models. Our results can be extended straightforwardly to the AH estimator.

The moment conditions for the AB estimator are

$$E\left(\Delta u_{i,t} y'_{i,t-1-\ell}\right) = 0 \text{ for } \ell = 1, 2, ..., t + p - 1;\ t = 1, ..., T. \tag{4}$$

To write the equations in the vector form, we let

$$\underbrace{A'}_{m \times mp} = (A_1, A_2, ..., A_p),$$

and define the first-differenced variables:

$$\underbrace{\Delta y_i}_{T\times m} = \begin{pmatrix} \Delta y'_{i,1} \\ \Delta y'_{i,2} \\ ... \\ \Delta y'_{i,T} \end{pmatrix}, \ \underbrace{\Delta u_i}_{T\times m} = \begin{pmatrix} \Delta u'_{i,1} \\ \Delta u'_{i,2} \\ ... \\ \Delta u'_{i,T} \end{pmatrix},$$

$$\underbrace{\Delta X_{i,t}}_{mp\times 1} = \begin{pmatrix} \Delta y_{i,t-1} \\ \Delta y_{i,t-2} \\ ... \\ \Delta y_{i,t-p} \end{pmatrix}, \ \underbrace{\Delta X_i}_{T\times mp} = \begin{pmatrix} \Delta X'_{i,1} \\ \Delta X'_{i,2} \\ ... \\ \Delta X'_{i,T} \end{pmatrix}.$$

We define the level variables y_i, u_i and X_i similarly except that they have $(T+1)$ rows. Then

$$vec(\Delta y_i) = (I_m \otimes \Delta X_i)\, vec\,(A) + vec(\Delta u_i). \tag{5}$$

To construct the instrument matrix, we let

$$Z_i = \begin{pmatrix} \left(y'_{i,-p}, ..., y'_{i,-1}\right) & 0 & ... & 0 \\ 0 & \left(y'_{i,-p}, ..., y'_{i,0}\right) & 0 & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & \left(y'_{i,-p}, ..., y'_{i,T-2}\right) \end{pmatrix} := \begin{pmatrix} Z'_{i,1} \\ Z'_{i,2} \\ ... \\ Z'_{i,T} \end{pmatrix} \tag{6}$$

which is a $T \times m\,[pT + (T-1)\,T/2]$ matrix. Then the moment conditions in (4) can be written as $E\left[(I_m \otimes Z_i)'\, vec(\Delta u_i)\right] = 0$.

The GMM estimator of $\alpha = vec\,(A)$ is now given by

$$\begin{aligned} \hat{\alpha}_{GMM} &= vec(\hat{A}) := vec((\hat{A}_1, \hat{A}_2, ..., \hat{A}_p)') \\ &= \left[\left(I_m \otimes S'_{ZX}\right) W_N \left(I_m \otimes S_{ZX}\right)\right]^{-1} \left[\left(I_m \otimes S'_{ZX}\right) W_N vec\,(S_{ZY})\right] \end{aligned} \tag{7}$$

where

$$S_{ZX} = \frac{1}{N}\sum_{i=1}^{N} Z'_i \Delta X_i, \ S_{ZY} = \frac{1}{N}\sum_{i=1}^{N} Z'_i \Delta y_i$$

and W_N is a weighting matrix that converges to W, a positive definite matrix as $N \to \infty$.

To estimate the orthogonalized impulse response function, we need to estimate the covariance matrix Σ. If the error term $u_{i,t}$ in (1) is observable, then an analysis-of-variance type estimator of Σ is given by

$$\tilde{\Sigma} = \frac{1}{N(T+1)} \sum_{i=1}^{N} \sum_{t=0}^{T} u_{i,t} u'_{i,t}.$$

Under the assumption that $u_{i,t}$ is normal, it can be shown that $\tilde{\Sigma}$ is the best quadratic unbiased estimator. It is also asymptotically equivalent to the QML estimator when a normal likelihood function is used. Since the error term is not observable, however, we have to replace it by some estimate. Given the estimate $\hat{\alpha}$, it is natural to estimate $u_{i,t}$ by

$$\hat{u}_{i,t} = (y_{i,t} - \bar{y}_{i,\cdot}) - \hat{A}' \left(X_{i,t} - \bar{X}_{i,\cdot} \right) \text{ for } t = 0, ..., T.$$

Here and hereafter, a dot in the subscript indicates the average over that subscript. The resulting estimator of Σ is then given by

$$\hat{\Sigma}_{GMM} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} \hat{u}_{i,t} \hat{u}'_{i,t}. \tag{8}$$

The subscript on $\hat{\Sigma}_{GMM}$ indicates that the estimator is based on the GMM estimator of A.

We now consider the large N asymptotics for a fixed T. Under some regularity conditions, we have

$$\mathrm{P} \lim_{N \to \infty} S_{ZX} = S_{ZX}^{\infty} \tag{9}$$

for some constant matrix S_{ZX}^{∞} and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} vec \left(Z_i' \Delta u_i \right) \Rightarrow N(0, \Sigma \otimes S_{ZZ}^{\infty}), \tag{10}$$

where

$$S_{ZZ}^{\infty} = \mathrm{P} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Z_i' G Z_i, \tag{11}$$

and G is the $T \times T$ symmetric tridiagonal matrix with the main diagonal elements equal to 2 and the sub-diagonal elements equal -1. Combining (9) and (10), we get

$$\sqrt{N}(\hat{\alpha}_{GMM} - \alpha) \Rightarrow N(0, \Omega_{\alpha\alpha})$$

for some variance matrix $\Omega_{\alpha\alpha}$.

To minimize the asymptotic variance of the GMM estimator, we choose the weighting matrix W_N such that its limit is $W = (\Sigma \otimes S_{ZZ}^{\infty})^{-1}$ (see, Hansen 1982). With the optimal weighting matrix, we have

$$\Omega_{\alpha\alpha} = \left\{ \left[I_m \otimes (S_{ZX}^{\infty})' \right] (\Sigma \otimes S_{ZZ}^{\infty})^{-1} \left[I_m \otimes S_{ZX}^{\infty} \right] \right\}^{-1} := \Sigma \otimes Q^{-1}, \tag{12}$$

where

$$Q = (S_{ZX}^{\infty})' (S_{ZZ}^{\infty})^{-1} S_{ZX}^{\infty}.$$

The above asymptotic variance can be also achieved by letting

$$W_N = I_m \otimes \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' G Z_i \right)^{-1}, \tag{13}$$

in which case $W = \text{Plim}_{N\to\infty} W_N = I_m \otimes (S_{ZZ}^{\infty})^{-1}$. To see this, note that for this choice of the weighting matrix, we have

$$(I_m \otimes S_{ZX}^{\infty})\, W\, (I_m \otimes S_{ZX}^{\infty}) = I_m \otimes Q,$$

and

$$var\left[\left(I_m \otimes S_{ZX}' \right) W_N vec\left(S_{ZY} \right) \right] \to \Sigma \otimes Q.$$

Therefore

$$\Omega_{\alpha\alpha} = (I_m \otimes Q)^{-1} (\Sigma \otimes Q) (I_m \otimes Q)^{-1} = \Sigma \otimes Q^{-1},$$

which is identical to the asymptotic variance given in (12).

With the weighting matrix given in (13), the GMM estimator of $\hat{\alpha}$ reduces to

$$\begin{aligned} \hat{\alpha}_{GMM} &= vec \left\{ \left[\left(\sum_{i=1}^{N} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{N} Z_i' G Z_i \right)^{-1} \left(\sum_{i=1}^{N} \left(Z_i' \Delta X_i \right) \right) \right]^{-1} \right. \\ &\quad \left. \times \left(\sum_{i=1}^{N} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{N} Z_i' G Z_i \right)^{-1} \sum_{i=1}^{N} (Z_i' \Delta y_i) \right\}. \end{aligned} \tag{14}$$

This is the equation-by-equation GMM estimator. Therefore, we have shown that the equation-by-equation GMM estimator is asymptotically as efficient as the system GMM estimator. This result is analogous to the asymptotic efficiency of the equation-by-equation OLS in an ordinary VAR system. Holtz-Eakin, Newey, and Rosen (1988) and Arellano (2003) both point out the possibility of improving the efficiency by jointly estimating all equations in the VAR system. Our result shows that, under the assumption of conditional homoskedasticity given in (2), there is no efficiency gain from joint estimation.

We now focus on the equation-by-equation GMM estimator given in (14) and the associated variance estimator defined in (8). To establish their joint limiting distribution, we maintain Assumption 2 below.

Assumption 2. The following hold for some $\delta > 0$:
(i) $E(\|u_{i,t}\|^{4+2\delta}) < \infty$,
(ii) $\max_i E\|X_{i,0}\|^{4+2\delta} < \infty$, $\max_i E\|\mu_i\|^{4+2\delta}) < \infty$,
(iii) S_{ZZ}^{∞} and S_{ZX}^{∞} have full rank,
(iv) $E(u_{i,t}u_{i,t}' \otimes u_{i,t} | y_{i,t-1}, ..., y_{i,-p}) = 0$,

where $X_{i,0} = (y'_{i,-1},\, y'_{i,-2}, ..., y'_{i,-p})'$.

Some comments on Assumption 2 are in order. Assumption 2(i) is a standard moment condition on $u_{i,t}$. Assumption 2(ii) assumes that the fixed effect μ_i and the initial values $y_{i,-1}, ..., y_{i,-p}$ have uniformly bounded $4+2\delta$ moments. Together with Assumption 2(i), Assumption 2(ii) ensures that individual contributions to cross sectional averages do not play a dominating role so that LLN and CLT hold. The moment conditions are not necessary but sufficient for our results. We point out in passing that while we do not parametrize the initial conditions, we still need to control their cross sectional heterogeneity by assuming that $\max_i E \left\| X_{i,0} \right\|^{4+2\delta} < \infty$.

To avoid the weak instrument problem, we impose Assumption 2(iii). If the initial observations are generated from the stationary distribution of the process, the full rank assumption rules out unit roots in the system, see for example, Binder, Hsiao and Pesaran (2005). It should be pointed out that the presence of a unit root does not necessarily lead to the weak instrument problem as the fixed individual effects combined with unrestricted initialization can ensure that S^{∞}_{ZX} is of full rank. When the initial values do not follow the stationary distribution of the VAR process, both ΔX_i and Z_i are affected by the fixed effect μ_i. As a result, Z_i can help predict ΔX_i not only because of the presence of time series dynamics but also because of the presence of the fixed effects.

We maintain the technical condition in Assumption 2(iv) in order to simplify the asymptotic variance. Under this assumption, the infeasible estimator

$$\hat{\Sigma}_0 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} \left(u_{i,t} - \bar{u}_{i,.}\right) \left(u_{i,t} - \bar{u}_{i,.}\right)'$$

is asymptotically independent of $\sqrt{N}(\hat{A} - A)$. Otherwise, there will be extra terms in the asymptotic variance that reflect the skewness of $u_{i,t}$. This assumption is also needed to ensure the asymptotic independence of the variance estimator and the slope estimator in time series VARs. A sufficient condition for Assumption 2(iv) to hold is that $u_{i,t}$ follows an elliptical distribution, which includes normal distributions as special cases.

Let

$$X_{i,t} = \left(y'_{i,t-1},\ y'_{i,t-2}, ..., y'_{i,t-p}\right)', \tag{15}$$

and

$$B = -\mathrm{P} \lim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} \left(u_{i,t} - \bar{u}_{i,.}\right) \left(X_{i,t} - \bar{X}_{i,.}\right)'. \tag{16}$$

The following theorem establishes the asymptotic distributions of $\hat{\alpha}_{GMM}$ and $\hat{\Sigma}_{GMM}$ when $N \to \infty$ for a fixed T.

Theorem 1 *Let Assumptions 1 and 2 hold. Then*

$$\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}_{GMM}-\alpha\right) \\ \sqrt{N}vech(\hat{\Sigma}_{GMM}-\Sigma) \end{pmatrix} \Rightarrow N\left(0,\Omega\right),\ \Omega=\begin{pmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\sigma}' \\ \Omega_{\alpha\sigma} & \Omega_{\sigma\sigma} \end{pmatrix} \tag{17}$$

where

$$\Omega_{\alpha\alpha}=\Sigma\otimes Q^{-1}, \tag{18}$$

$$\Omega_{\alpha\sigma}=-D_m^+(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right)-D_m^+K_{m,m}(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right), \tag{19}$$

$$\begin{aligned}\Omega_{\sigma\sigma} &= D_m^+\left[\frac{1}{T+1}\Lambda_{m^2}+\frac{1}{T\left(T+1\right)}\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\right]\left(D_m^+\right)' \\ &\quad +D_m^+\left(\Sigma\otimes BQ^{-1}B'\right)\left(D_m^+\right)'+D_m^+\left(BQ^{-1}B'\otimes\Sigma\right)\left(D_m^+\right)' \\ &\quad +D_m^+\left(\Sigma\otimes BQ^{-1}B\right)K_{m,m}'\left(D_m^+\right)'+D_m^+K_{m,m}\left(\Sigma\otimes BQ^{-1}B'\right)\left(D_m^+\right)',\end{aligned} \tag{20}$$

and $\Lambda_{m^2}=var(vec(u_{i,t}u_{i,t}'))$ *is an* $m^2\times m^2$ *matrix.*

Remark 1 *Theorem 1 shows that* $\sqrt{N}\left(\hat{\alpha}_{GMM}-\alpha\right)$ *is not asymptotically independent of* $\sqrt{N}vech(\hat{\Sigma}_{GMM}-\Sigma)$*. This is in sharp contrast with the time series case. For a time series VAR model with Gaussian innovations, the MLEs of* α *and* Σ *are asymptotically independent, see Hamilton (1994, Proposition 11.2). To construct valid confidence bands for the orthogonalized IRFs from short panel VARs, we have to take the asymptotic dependence between* $\sqrt{N}\left(\hat{\alpha}_{GMM}-\alpha\right)$ *and* $\sqrt{N}vech(\hat{\Sigma}_{GMM}-\Sigma)$ *into account.*

Remark 2 *Equation (41) in the proof states that*

$$\sqrt{N}vech(\hat{\Sigma}_0-\Sigma)\rightarrow D_m^+\left[\left(T+1\right)^{-1}\Lambda_{m^2}+T^{-1}\left(T+1\right)^{-1}\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\right]\left(D_m^+\right)'.$$

So the asymptotic variance of $\sqrt{N}vech(\hat{\Sigma}_0-\Sigma)$ *contains two terms. The second term* $T^{-1}\left(T+1\right)^{-1}D_m^+\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\left(D_m^+\right)'$ *reflects the estimation uncertainty of the fixed effects. When* $T\rightarrow\infty$*, the second term is of smaller order than the first term* $\left(T+1\right)^{-1}D_m^+\Lambda_{m^2}\left(D_m^+\right)'$ *and disappears asymptotically. However, when* T *is assumed to be fixed, both terms contribute to the asymptotic variance. This is different from the time series asymptotics. The difference highlights the risk of naively extending time series results to short panels.*

Remark 3 *It is precisely because* $B\neq 0$ *that the fixed effects estimator or the least squared dummy variable (LSDV) estimator is asymptotically biased. If the asymptotic bias of the fixed effects estimator is nonnegligible, then it is likely that the asymptotic dependence between* $\hat{\alpha}_{GMM}$ *and* $\hat{\Sigma}_{GMM}$ *is also nonnegligible. This paper complements the papers by Anderson and Hsiao (1982) and Arellano and Bond (1991) in that they investigate the consequence of demeaning for the short panels on the slope estimation while we examine the consequence on the variance estimation.*

Remark 4 *In general, the variance matrix Λ_{m^2} depends on the fourth-order multivariate cumulants of $u_{i,t}$. The relation between product-moments and multivariate cumulants is rather technical. See Bilodeau and Brenner (1999, Appendix B). However, if we make some distributional assumptions on $u_{i,t}$, Λ_{m^2} may be simplified. For example, if we assume that $u_{i,t}$ follows an elliptical distribution, then*

$$\Lambda_{m^2} = \frac{1}{3}\kappa\left(\Sigma \otimes \Sigma\right)\left(I_{m^2} + K_{m,m}\right) + \left(\frac{\kappa}{3} - 1\right) vec\left(\Sigma\right)\left[vec\left(\Sigma\right)\right]'$$

where κ is the kurtosis of any element of the standardized error $\tilde{u}_{i,t} = \Sigma^{-1/2}u_{i,t}$. That is $\kappa = E([\tilde{u}_{i,t}^{(j)}]^4)/\{E([\tilde{u}_{i,t}^{(j)}]^2)\}^2$. For a proof of this result, see Bilodeau and Brenner (1999, example 13.6). As a special case, when $u_{i,t} \sim iidN(0,\Sigma)$, we have $\kappa = 3$ and

$$\Lambda_{m^2} = \left(\Sigma \otimes \Sigma\right)\left(I_{m^2} + K_{m,m}\right).$$

We will use this formula in our simulation study.

3 Asymptotic Approximation to the Distribution of IRFs

In this section, we first define the IRFs for reduced-form VARs and structural VARs. We then consider the large sample approximation and bootstrap approximation to the sampling distribution of the IRFs.

3.1 IRFs for Reduced-form and Structural VARs

Since the impulse response function does not depend on the index i and fixed effects in the system, we omit the subscript i and consider the reduced-form VAR model:

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t \text{ for } t = 0, ..., T. \tag{21}$$

The impulse response matrix is defined to be

$$\Phi_j = \frac{\partial y_{t+j}}{\partial u_t'}.$$

The (k,ℓ)-*th* element of Φ_j describes the response of k-th element of y_{t+j} to one unit impulse in ℓ-th element of y_t with all other variables dated t or earlier held constant. The plot of the (k,ℓ)-*th* element of Φ_j as a function of j is called the non-orthogonalized impulse-response function.

To compute Φ_j, we let

$$Y_t = \underset{mp\times 1}{\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{pmatrix}}, F = \underset{mp\times mp}{\begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_m & 0 & \cdots & 0 & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_m & 0 \end{pmatrix}}, U_t = \underset{mp\times 1}{\begin{pmatrix} u_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}.$$

Then

$$Y_t = \sum_{j=0}^{t} F^j U_{t-j} + F^{t+1} Y_{-1},$$

so $\partial Y_{t+j}/\partial U_t' = F^j$. By definition, Φ_j is the first $m \times m$ block of F^j.

Differentiating both sides of (21) yields:

$$\frac{\partial y_{t+j}}{\partial u_t'} = A_1 \frac{\partial y_{t+j-1}}{\partial u_t'} + ... + A_p \frac{\partial y_{t+j-p}}{\partial u_t'}.$$

That is, Φ_j satisfies the recursive relationship:

$$\Phi_j = \sum_{\ell=1}^{p} A_\ell \Phi_{j-\ell}, \ j = 1, 2, ... \tag{22}$$

with $\Phi_j = 0$ for $j < 0$ and $\Phi_0 = I_m$.

To estimate the non-orthogonalized impulse response, we plug the estimate $\hat{A}$ into (22) and get $\hat{\Phi}_j = \sum_{\ell=1}^{p} \hat{A}_\ell \hat{\Phi}_{j-\ell}$ with the initialization $\hat{\Phi}_j = 0$ for $j < 0$ and $\hat{\Phi}_0 = I_m$.

In empirical applications, it is a standard practice to report the orthogonalized impulse response function. Let $P_r P_r' = \Sigma$ be the Cholesky decomposition of Σ, where P_r is a lower triangular matrix with positive diagonal elements. Then

$$P_r^{-1} y_t = P_r^{-1} A_1 y_{t-1} + ... + P_r^{-1} A_p y_{t-p} + u_t^r \text{ for } t = 0, ..., T \tag{23}$$

where $u_t^r = P_r^{-1} u_t$ has mean zero and variance I_m. The orthogonalized impulse response matrix is defined to be

$$\Theta_j^r = \frac{\partial y_{t+j}}{\partial \left(u_t^r\right)'} = \frac{\partial y_{t+j}}{\partial u_t'} \frac{\partial u_t}{\partial \left(u_t^r\right)'} = \Phi_j P_r.$$

In general, the model in (23) does not have a structural interpretation. One exception is the recursive structural VAR model defined by

$$\mathcal{A} y_t = A_1^s y_{t-1} + ... + A_p^s y_{t-p} + u_t^s \text{ for } t = 0, ..., T \tag{24}$$

where $\mathcal{A}$ is a lower triangular matrix, and u_t^s is a vector white noise process with variance matrix I_m. To obtain the recursive structure, we may have to re-order the variables in y_t according to some economic theory. The structural model coincides with the reduced-form model in (23) if we let $P_r^{-1} = \mathcal{A}$, $P_r^{-1}A_i = A_i^s$ and $u_t^r = u_t^s$. In this case, the IRFs based on the Cholesky decomposition have structural interpretations.

In the absence of a recursive structure, we decompose Σ into $\Sigma = P^s (P^s)'$ where $(P^s)^{-1}$ satisfies the same identification restrictions imposed on the matrix $\mathcal{A}$. This decomposition is different from the Cholesky decomposition, as it is dictated by the identification restrictions. To identify $\mathcal{A}$, it is necessary to impose $m(m-1)/2$ restrictions. Following Lütkepohl (2005, p. 360), we consider the identification restrictions: $\mathbb{S}_{\mathcal{A}} vec(\mathcal{A}) = s_{\mathcal{A}}$ where $\mathbb{S}_{\mathcal{A}}$ is a $m(m-1)/2 \times m^2$ selection matrix and $s_{\mathcal{A}}$ is a suitable $m(m-1)/2 \times 1$ fixed vector. Under these conditions, we solve for P^s such that $\Sigma = P^s (P^s)'$ and $\mathbb{S}_{\mathcal{A}} vec[(P^s)^{-1}] = s_{\mathcal{A}}$.

The reduced-form Cholesky decomposition and the structural decomposition can be represented in a unified framework. Given the covariance matrix Σ, we want to find P such that

$$\Sigma = PP' \text{and } \mathbb{S} vec(P^{-1}) = s \tag{25}$$

for some selection matrix $\mathbb{S}$ and constant vector s. For the structural VAR, $(\mathbb{S}, s)$ is based on economic theory. For the reduced-form Cholesky decomposition, $\mathbb{S}$ is the matrix that selects the upper triangular elements of P^{-1} and s is a vector of zeros. Given the matrix P, the orthogonalized IRFs are $\Theta_j = \Phi_j P$.

The orthogonalized IRF can be estimated by plugging the estimates $\hat{\Phi}_j$ and $\hat{\Sigma}$ into its definition. More specifically, let $\hat{P}$ be the plug-in estimator of P such that $\hat{\Sigma} = \hat{P}\hat{P}'$ and $\mathbb{S} vec(\hat{P}^{-1}) = s$. Then the orthogonalized IRF can be estimated by $\hat{\Theta}_j = \hat{\Phi}_j \hat{P}$.

3.2 Large Sample Approximation

The limiting distribution of $\hat{\Phi}_j$ can be derived using the delta method. More specifically, taking transposes of (22) and differentiating the resulting equation with respect to α_q, the q-th element of α, yields:

$$\begin{aligned}
\frac{\partial \Phi_j'}{\partial \alpha_q} &= \sum_{\ell=1}^{p} \frac{\partial \Phi_{j-\ell}'}{\partial \alpha_q} A_\ell' + \sum_{\ell=1}^{p} \Phi_{j-\ell}' \frac{\partial A_\ell'}{\partial \alpha_q} \\
&= \sum_{\ell=1}^{p} \frac{\partial \Phi_{j-\ell}'}{\partial \alpha_q} A_\ell' + [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}'] \frac{\partial A}{\partial \alpha_q}.
\end{aligned}$$

Consequently,

$$\frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'} = \sum_{\ell=1}^{p} (A_\ell \otimes I_m) \left(\frac{\partial vec\left(\Phi_{j-\ell}'\right)}{\partial \alpha'} \right) + \left(I_m \otimes [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}'] \right).$$

Let

$$\underset{(m^2\times pm^2)}{G_0} = 0 \text{ and } \underset{(m^2\times pm^2)}{G_j} = \frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'},\ j = 1, 2, ...$$

then

$$G_j = \sum_{\ell=1}^{p} (A_\ell \otimes I_m) G_{j-\ell} + (I_m \otimes [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}'])$$

with $G_j = 0$ for $j < 0$. A closed-form solution for G_j is

$$G_j = \sum_{\ell=0}^{j-1} \Phi_\ell \otimes [\Phi_{j-\ell-1}', \Phi_{j-\ell-2}', ..., \Phi_{j-\ell-p}'] = \sum_{\ell=0}^{j-1} \Phi_\ell \otimes J\left(F'\right)^{j-\ell-1},$$

where $J = [I_m, \mathbf{0}_m, ..., \mathbf{0}_m]$ is an $m \times mp$ matrix.

A consistent estimator of G_j can be obtained by plugging $\hat{A}$ and $\hat{\Phi}_j$ into the above equation. The asymptotic distribution of the non-orthogonalized impulse response function is

$$\sqrt{N} vec\left(\hat{\Phi}_j' - \Phi_j'\right) \xrightarrow{d} N(0, \Omega_{\Phi j}) \text{ for } \Omega_{\Phi j} = G_j \Omega_{\alpha\alpha} G_j'. \tag{26}$$

We can estimate $\Omega_{\Phi j}$ by

$$\hat{\Omega}_{\Phi j} = \hat{G}_j \hat{\Omega}_{\alpha\alpha} \hat{G}_j' \tag{27}$$

where

$$\hat{G}_j = \sum_{\ell=0}^{j-1} \hat{\Phi}_\ell \otimes [\hat{\Phi}_{j-\ell-1}', \hat{\Phi}_{j-\ell-2}', ..., \hat{\Phi}_{j-\ell-p}'],\ \hat{\Omega}_{\alpha\alpha} = \hat{\Sigma}_{GMM} \otimes \hat{Q}^{-1},$$

and

$$\hat{Q} = \left(\frac{1}{N}\sum_{i=1}^{N} \Delta X_i' Z_i\right)\left(\frac{1}{N}\sum_{i=1}^{N} Z_i' G Z_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} Z_i' \Delta X_i\right).$$

To derive the limiting distribution of $\hat{\Theta}_j$, we use the delta method again. We compute $C_{\alpha j} = \partial vec(\Theta_j')/\partial \alpha'$ and $C_{\sigma j} = \partial vec(\Theta_j')/\partial \left[vech(\Sigma)\right]'$ as follows:

$$C_{\alpha j} = \frac{\partial vec\left(P'\Phi_j'\right)}{\partial \alpha'} = \left(I_m \otimes P'\right)\frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'} = \left[I_m \otimes P'\right] G_j,$$

$$C_{\sigma j} = \frac{\partial vec\left(P'\Phi_j'\right)}{\partial \left[vech(\Sigma)\right]'} = (\Phi_j \otimes I_m)\frac{\partial vec\left(P'\right)}{\partial \left[vech(\Sigma)\right]'} = (\Phi_j \otimes I_m)\frac{K_{m,m}\partial vec\left(P\right)}{\partial \left[vech(\Sigma)\right]'}.$$

To obtain a closed form expression for $\partial vec\left(P\right)/\partial\left[vech(\Sigma)\right]'$, we differentiate both sides of the equations in (25) and get

$$\begin{bmatrix} 2D_m^+\left(P \otimes I_m\right) \\ \mathbb{S}[(P^{-1})' \otimes P^{-1}] \end{bmatrix} vec\left(dP\right) = \begin{bmatrix} vech\left(d\Sigma\right) \\ 0 \end{bmatrix}.$$

Let O_m be the first $m(m+1)/2$ column of the matrix on the left hand side, that is

$$\left[\begin{array}{c} 2D_m^+(P\otimes I_m) \\ \mathbb{S}[(P^{-1})'\otimes P^{-1}] \end{array}\right]^{-1} = [O_m, \Delta]$$

for some $m^2 \times m(m+1)/2$ matrix O_m and $m^2 \times m(m-1)/2$ matrix Δ. Then $vec(dP) = O_m vech(d\Sigma)$, and so

$$C_{\sigma j} = (\Phi_j \otimes I_m) K_{m,m} O_m.$$

The following theorem gives that asymptotic distribution of the orthogonalized IRF $\hat{\Theta}_j$.

Theorem 2 *Let Assumptions 1 and 2 hold. Assume that the matrix*

$$\left[\begin{array}{c} 2D_m^+(P\otimes I_m) \\ \mathbb{S}[(P^{-1})'\otimes P^{-1}] \end{array}\right]$$

is of full rank. Then

$$\sqrt{N} vec[(\hat{\Theta}_j - \Theta_j)'] \xrightarrow{d} N(0, \Omega_{\Theta j}), \tag{28}$$

where

$$\Omega_{\Theta j} = [C_{\alpha j}, C_{\sigma j}] \left(\begin{array}{cc} \Omega_{\alpha\alpha} & \Omega_{\alpha\sigma} \\ \Omega_{\sigma\alpha} & \Omega_{\sigma\sigma} \end{array}\right) \left(\begin{array}{c} C'_{\alpha j} \\ C'_{\sigma j} \end{array}\right), \tag{29}$$

and

$$C_{\alpha j} = \left(I_m \otimes P'\right) G_j, C_{\sigma j} = (\Phi_j \otimes I_m) K_{m,m} O_m.$$

In the time series setting, the matrix $\Omega_{\alpha\sigma} = 0$ (e.g. Proposition 1 in Lütkepohl (1990)). As a result, the cross product terms $C_{\sigma j}\Omega_{\sigma\alpha}C'_{\alpha j}$, $C_{\alpha j}\Omega_{\alpha\sigma}C'_{\sigma j}$ are not present in the asymptotic variance of the orthogonalized IRF. In contrast, for short panel VARs, $\Omega_{\alpha\sigma} \neq 0$, and the cross product terms can not be ignored. In addition, compared to the time series cases, the asymptotic variance $\Omega_{\sigma\sigma}$ contains a few extra terms, reflecting the estimation uncertainty of the slope coefficients and the fixed effects. So for short panel VARs, it is important to include the cross product terms $C_{\sigma j}\Omega_{\sigma\alpha}C'_{\alpha j}$, $C_{\alpha j}\Omega_{\alpha\sigma}C'_{\sigma j}$ and extra terms in $\Omega_{\sigma\sigma}$ in computing the asymptotic variance, especially when T is small.

The matrix O_m can be simplified if P is restricted to be a lower triangular matrix. In this case

$$O_m = \frac{\partial vec(P)}{\partial [vech(\Sigma)]'} = L'_m \frac{\partial vech(P)}{\partial [vech(\Sigma)]'}.$$

But it follows from $\Sigma = PP'$ that $vec(d\Sigma) = (I_{m^2} + K_{mm})(P \otimes I_m) vec(dP)$ and so

$$\frac{\partial vech(P)}{\partial [vech(\Sigma)]'} = \left[L_m (I_{m^2} + K_{mm})(P \otimes I_m) L'_m\right]^{-1} = \left[2D_m^+ (P \otimes I_m) L'_m\right]^{-1}.$$

As a result

$$O_m = L_m' \left[2D_m^+ \left(P \otimes I_m \right) L_m' \right]^{-1}.$$

The structural VAR model in (24) is referred to as the A-model by Lütkepohl (2005, p. 358). Another class of structural VAR models, the so-called B-model, is defined to be

$$y_t = A_1^s y_{t-1} + ... + A_p^s y_{t-p} + \mathcal{B} u_t^s \text{ for } t = 0, ..., T$$

where u_t^s is normalized to have mean zero and variance I_m. To identify the structural parameter $\mathcal{B}$, we impose $m(m-1)/2$ restrictions of the form $\mathbb{S}_{\mathcal{B}} vec(\mathcal{B}) = s_{\mathcal{B}}$. In this case, the map between the reduced form variance Σ and the structural matrix $\mathcal{B}$ is defined by $\Sigma = \mathcal{B}\mathcal{B}'$ and $\mathbb{S}_{\mathcal{B}} vec(\mathcal{B}) = s_{\mathcal{B}}$. In other words, we solve for P such that $\Sigma = PP'$ such that $\mathbb{S}_{\mathcal{B}} vec(P) = s_{\mathcal{B}}$. Theorem 2 continues to hold but O_m is now the first $m\left(m+1\right)/2$ column of the matrix

$$\begin{bmatrix} 2D_m^+ \left(P \otimes I_m \right) \\ \mathbb{S}_{\mathcal{B}} \end{bmatrix}^{-1}.$$

To consistently estimate the asymptotic variance $\Omega_{\Theta j}$, we plug consistent estimates of $C_{\alpha j}, C_{\sigma j}, \Omega_{\alpha\alpha}, \Omega_{\alpha\sigma}$ and $\Omega_{\sigma\sigma}$ into (29), leading to

$$\hat{\Omega}_{\Theta j} = \hat{C}_{\alpha j} \hat{\Omega}_{\alpha\alpha} \hat{C}_{\alpha j}' + \hat{C}_{\sigma j} \hat{\Omega}_{\sigma\sigma} C_{\sigma j}' + \hat{C}_{\sigma j} \hat{\Omega}_{\sigma\alpha} \hat{C}_{\alpha j}' + \hat{C}_{\alpha j} \hat{\Omega}_{\alpha\sigma} \hat{C}_{\sigma j}', \tag{30}$$

where

$$\hat{C}_{\alpha j} = (I_m \otimes \hat{P}') \hat{G}_j \text{ and } \hat{C}_{\sigma j} = (\hat{\Phi}_j \otimes I_m) K_{m,m} \hat{O}_m.$$

Here $\hat{\Omega}$'s are defined in (18)–(20) with B, Q, Σ replaced by $\hat{B}, \hat{Q}, \hat{\Sigma}_{GMM}$, where

$$\hat{B} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=0}^{T} \hat{u}_{i,t} \left(X_{i,t} - \bar{X}_{i,\cdot} \right)'.$$

$\hat{O}_m$ is the first $m\left(m+1\right)/2$ column of

$$\begin{bmatrix} 2D_m^+ \left(\hat{P} \otimes I_m \right) \\ \mathbb{S}_{\mathcal{A}}[(\hat{P}^{-1})' \otimes \hat{P}^{-1}] \end{bmatrix}^{-1} \text{ or } \begin{bmatrix} 2D_m^+ \left(\hat{P} \otimes I_m \right) \\ \mathbb{S}_{\mathcal{B}} \end{bmatrix}^{-1},$$

respectively for the A-model and B-model. For recursive structural VARs, we can take

$$\hat{O}_m = L_m' \left[2D_m^+ \left(\hat{P} \otimes I_m \right) L_m' \right]^{-1}.$$

3.3 Bootstrap Approximation

The large sample approximation in the previous subsection is based on the delta method. In finite samples, the approximation may not capture the finite sample distribution very well. In this subsection, we consider the bootstrap approximation to the IRFs.

We use the nonparametric iid bootstrap along the cross sectional dimension. Let $\{y_i = (y_{i,-p}, ..., y_{i,T})', i = 1, 2, ..., N\}$ denote the original sample and $\mathbb{P}$ denote the true distribution that generates the sample $\{y_1, ..., y_N\}$. The bootstrap sample is denoted by $\{y_i^* = (y_{i,-p}^*, ..., y_{i,T}^*)', i = 1, 2, ..., N\}$ where y_i^* follows the bootstrap distribution $\mathbb{P}^*$, a discrete distribution that places probability mass $1/N$ at each point in the sample $\{y_1, ..., y_N\}$. By definition, the bootstrap sample satisfies

$$\Delta y_{i,t}^* = \hat{A}_1 \Delta y_{i,t-1}^* + ... + \hat{A}_p \Delta y_{i,t-p}^* + \Delta u_{i,t}^*, t = 1, ..., T.$$

where $\{u_i^* = (u_{i,-p,}^*..., u_{i,T}^*)', i = 1, 2, ..., N\}$ are simple random draws from the estimated errors $\{\hat{u}_i = (\hat{u}_{i,-p,}..., \hat{u}_{i,T})', i = 1, 2, ..., N\}$.

The GMM estimator of α based on the bootstrap sample is

$$\begin{aligned}
\hat{\alpha}_{GMM}^* &= vec((\hat{A}_1^*, ..., \hat{A}_p^*)') \\
&= vec\left\{ \left[\left(\sum_{i=1}^N (\Delta X_i^*)' Z_i^* \right) \left(\sum_{i=1}^N (Z_i^*)' G Z_i^* \right)^{-1} \left(\sum_{i=1}^N (Z_i^*)' \Delta X_i^* \right) \right]^{-1} \right. \\
&\quad \left. \times \left(\sum_{i=1}^N (\Delta X_i^*)' Z_i^* \right) \left(\sum_{i=1}^N (Z_i^*)' G Z_i^* \right)^{-1} \sum_{i=1}^N (Z_i^*)' \Delta y_i^* \right\},
\end{aligned}$$

where the variables X_i^*, Z_i^* and y_i^* are defined in the same way as X_i, Z_i and y_i except that they are based on the bootstrap sample. Similarly, the bootstrap estimator of Σ is

$$\hat{\Sigma}_{GMM}^* = \frac{1}{NT} \sum_{i=1}^N \sum_{t=0}^T \hat{u}_{i,t}^* \hat{u}_{i,t}^{*\prime},$$

where

$$\hat{u}_{i,t}^* = \left(y_{i,t}^* - \bar{y}_{i,\cdot}^* \right) - (\hat{A}^*)' \left(X_{i,t}^* - \bar{X}_{i,\cdot}^* \right), \text{ for } t = 0, ..., T.$$

Given $\hat{\alpha}_{GMM}^*$ and $\hat{\Sigma}_{GMM}^*$, we can compute the IRFs $\hat{\Phi}_j^*$ and the orthogonalized IRFs $\hat{\Theta}_j^*$ using the same procedure as that for $\hat{\Phi}_j$ and $\hat{\Theta}_j$. In addition, we can compute $\hat{\Omega}_{\Phi j}^*$ and $\hat{\Omega}_{\Theta j}^*$ in exactly the same way as $\hat{\Omega}_{\Phi j}$ and $\hat{\Omega}_{\Theta j}$ defined in (27) and (30) but use the bootstrap sample.

Theorem 3 *Let Assumption 1, Assumptions 2(iii) and (iv) hold. In addition, assume that* $E(\|u_{i,t}\|^{16+8\delta})$, $\max_i E\|X_{i,0}\|^{16+8\delta}$, $\max_i E\|\mu_i\|^{16+8\delta}$ *are finite for some* $\delta > 0$. *Then for any conformable vector* c, *the following hold uniformly over* $x \in \mathbb{R}$:

(a) $\mathbb{P}\left(c'\sqrt{N}vec[(\hat{\Phi}_j^* - \hat{\Phi}_j)'] < x\right) = \mathbb{P}\left(c'\sqrt{N}vec[(\hat{\Phi}_j - \Phi_j)'] < x\right) + o(1)$,
(b) $\mathbb{P}\left(c'\sqrt{N}vec[(\hat{\Theta}_j^* - \hat{\Theta}_j)'] < x\right) = \mathbb{P}\left(c'\sqrt{N}vec[(\hat{\Theta}_j - \Theta_j)'] < x\right) + o(1)$,
(c)

$$\mathbb{P}\left(\frac{c'\sqrt{N}vec[(\hat{\Phi}_j^* - \hat{\Phi}_j)']}{\sqrt{c'\hat{\Omega}_{\Phi j}^* c}} < x\right) = \mathbb{P}\left(\frac{c'\sqrt{N}vec[(\hat{\Phi}_j - \Phi_j)']}{\sqrt{c'\hat{\Omega}_{\Phi j} c}} < x\right) + o(1),$$

(d)

$$\mathbb{P}\left(\frac{c'\sqrt{N}vec[(\hat{\Theta}_j^* - \hat{\Theta}_j)']}{\sqrt{c'\hat{\Omega}_{\Theta j}^* c}} < x\right) = \mathbb{P}\left(\frac{c'\sqrt{N}vec[(\hat{\Theta}_j - \Theta_j)']}{\sqrt{c'\hat{\Omega}_{\Theta j} c}} < x\right) + o(1).$$

To evaluate the probabilities in Theorem 3, we first condition on the original sample under which the average statistics of interest converge almost surely. The moment conditions in Theorem 3 ensure that the sample we condition on occurs with probability one. Consequently, the conditional convergence results can then be converted into unconditional results. The moment conditions are likely to be stronger than necessary but they facilitate the proof.

To prove Theorem 3, we first show that $[\sqrt{N}\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right),\ \sqrt{N}vech(\hat{\Sigma}_{GMM}^* - \hat{\Sigma}_{GMM})]$ has the same joint limiting distribution as $[\sqrt{N}\left(\hat{\alpha}_{GMM} - \alpha\right),\ \sqrt{N}vech(\hat{\Sigma}_{GMM} - \Sigma)]$ in Lemma 1 given in the appendix. We then invoke a delta-type method for bootstrap approximation. For the standard delta method, it is sufficient to assume that the function of interest is continuous. Here we require the function to be continuously differentiable. The delta method for bootstrap approximation is likely to be of independent interest.

A direct implication of Theorem 3 is that bootstrap percentile confidence band and bootstrap percentile-t confidence band are asymptotically valid to the first order. Higher order refinement of the bootstrap approximation may require stronger moment conditions and some adjustment of the bootstrap GMM estimator when the model is overidentified (c.f. Horowitz, 1997). This is beyond the scope of the present paper.

4 Simulation Evidence

In this section, we provide some simulation evidence on the accuracy of the asymptotic approximation and the bootstrap approximation to the sampling variability of the orthogonalized impulse response functions.

We consider the panel VAR model with two variables. The data generating process is

$$y_{i,t} = \mu + Ay_{i,t-1} + \mu_i + Pe_{i,t}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ where $\mu = (0, 0)'$, $e_{i,t} \sim iid\ N(0, I_2)$ and $\mu_i \sim iidN(0, I_2)$. For each given T, we set the initial value of the process $\{y_{i,t}\}$ to be zero and generate a 2-dimensional time series of length $T + T_d$. We drop the first T_d observations to obtain the simulated sample.

We specify matrices A and P as follows. Let R and S be 2×2 random matrices whose elements are iid from the uniform $[0, 1]$ distribution. Let b_1 and b_2 be the eigenvectors of $R'R$. We set A to be

$$A = \lambda_1 b_1 b_1' + \lambda_2 b_2 b_2'$$

for λ_1 and $\lambda_2 \in (0.10, 0.35, 0.60, 0.85)$, $\lambda_1 \neq \lambda_2$, and set P to be the low triangle matrix such that $P'P = SS'$. Since the performance of the AB estimator may be sensitive to the specifications of A and P (e.g. Bun and Kiviet, 2006), we have experimented with different random matrices R and S. We have also experimented with

$$P = \begin{pmatrix} 1 & 0 \\ \rho & 1 \end{pmatrix}$$

for different values of ρ. The qualitative information from our simulation results remains more or less the same.

We consider different N and T combinations, i.e. $N = 100, 200$ and $T = 5, 10, 20$. We set T_d to be $1, 5, 10, 50$. For each (N, T) combination, we estimate the model using the AB estimator. To avoid the weak instrument problem, we do not use the lagged dependent variables dated too early as instruments. Instead, we set the maximum number of lags of the dependent variable that can be used as instruments to be 3. Our simulation results change only slightly when we set the maximum number of lags to be 1, 2 and 4. The number of simulation replications is 5000.

4.1 The Asymptotic and Bootstrap Confidence Bands

Given the estimated autoregressive and variance parameters, we construct the orthogonalized impulse response functions and the corresponding 95% confidence band based on the asymptotic distribution in (28). As a comparison, we also construct the 95% confidence band when $\Omega_{\alpha\sigma}$ and $\Omega_{\sigma\sigma}$ are set to be

$$\Omega_{\alpha\sigma} = 0,\ \Omega_{\sigma\sigma} = (T+1)^{-1} D_m^+ (\Sigma \otimes \Sigma) (I_{m^2} + K_{m,m}) \left(D_m^+\right)'. \tag{31}$$

In this case, the asymptotic dependence between $\sqrt{N}(\hat{\alpha}_{GMM} - \alpha)$ and $\sqrt{N} vech(\hat{\Sigma}_{GMM} - \Sigma)$ and the additional randomness of $\sqrt{N} vech(\hat{\Sigma}_{GMM} - \Sigma)$ are ignored. Both confidence bands are of the form

$$\left[\hat{\Theta}_j(k,\ell) - 1.96 \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}, \hat{\Theta}_j(k,\ell) + 1.96 \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}\right].$$

For convenience, we call the confidence band with the naive asymptotic variance given in (31) the naive CLT confidence band and the one based on (26) the variance-corrected CLT confidence band.

In the simulation experiment, we also consider the finite sample performances of bootstrap confidence bands. Hall (1992) discuses three types of bootstrap confidence bands. See Lütkepohl (2005) for VARs. The first is Hall's percentile confidence band. In our case, the 95% bootstrap confidence band is

$$\left[\hat{\Theta}_j(k,\ell) - CV_U^{\hat{\Theta}_j^*(k,\ell)}, \hat{\Theta}_j(k,\ell) - CV_L^{\hat{\Theta}_j^*(k,\ell)}\right], \tag{32}$$

where $CV_U^{\hat{\Theta}_j^*(k,\ell)}$ and $CV_L^{\hat{\Theta}_j^*(k,\ell)}$ are the 97.5% and 2.5% quantiles of $\hat{\Theta}_j^*(k,\ell) - \hat{\Theta}_j(k,\ell)$, respectively.

The other two types of bootstrap bands are based on the t-statistic. For the original sample, the t-statistic is

$$t = \frac{\sqrt{N}\left[\hat{\Theta}_j(k,\ell) - \Theta_j(k,\ell)\right]}{\hat{\Omega}_{\Theta j}(k,\ell)} \tag{33}$$

while for the bootstrap sample the t-statistic is

$$t^* = \frac{\sqrt{N}\left[\hat{\Theta}_j^*(k,\ell) - \hat{\Theta}_j(k,\ell)\right]}{\hat{\Omega}_{\Theta j}^*(k,\ell)}. \tag{34}$$

Then the equal-tailed percentile-t confidence band is

$$\left[\hat{\Theta}_j(k,\ell) - CV_U^{t^*} \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}, \hat{\Theta}_j(k,\ell) - CV_L^{t^*} \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}\right], \tag{35}$$

where $CV_U^{t^*}$ and $CV_L^{t^*}$ are the 97.5% and 2.5% quantiles of t^* respectively. If instead of t^*, the quantiles are calculated based on $|t^*|$, then we have the symmetric percentile-t confidence band. Let $CV^{|t^*|}$ be the 95% quantile of the bootstrap distribution of $|t^*|$, the symmetric bootstrap confidence band is:

$$\left[\hat{\Theta}_j(k,\ell) - CV^{|t^*|} \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}, \hat{\Theta}_j(k,\ell) + CV^{|t^*|} \times \hat{\Omega}_{\Theta j}(k,\ell)/\sqrt{N}\right]. \tag{36}$$

General bootstrap theory suggests that the symmetric percentile-t confidence band has the most accurate coverage probability among the three bootstrap confidence bands considered here. See Hall (1992). An extensive simulation study in a previous version of this paper supports this qualitative observation. So we focus on the symmetric percentile-t confidence band hereafter. Depending on how $\hat{\Omega}_{\Theta j}(k,\ell)$ and its bootstrap version are computed, we obtain two different symmetric percentile-t bootstrap confidence bands: the naive bootstrap confidence band and the variance-corrected bootstrap confidence band.

4.2 Simulation Results

Figure 1 graphs the empirical coverage of different confidence bands against the forecasting horizons when $N = 100$, $T = 5, T_d = 1$. The confidence bands considered are: the naive CLT band, the variance-corrected CLT band, the naive symmetric percentile-t bootstrap band, and the variance-corrected symmetric percentile-t bootstrap band. Here we focus on the case $(\lambda_1, \lambda_2) = (0.1, 0.6)$, which is representative of other cases. We will refer to this parameterization as our base case hereafter. The empirical coverage of the CLT confidence bands is the coverage rate based on 5000 rounds of Monte Carlo simulations. For the bootstrap confidence bands, we compute their empirical coverage based on 999 bootstrap replications in each of 5000 simulation replications.

Several patterns emerge from the figure. First, the empirical coverage of the variance-corrected CLT is closer to the nominal coverage probability than the naive CLT confidence band. For some scenarios, the variance-corrected CLT confidence band dominates the naive CLT confidence band by a large margin. Across all the subplots in the figure, we find that the empirical coverage of the naive CLT band is considerately lower than the nominal coverage. A direct implication is that the naive asymptotic variance under-estimates the sampling variability of the impulse response. As a result, inferences based on the naive asymptotic variance may lead to the finding a statistically significant relationship that does not actually exist. Second, similar to the findings for the CLT bands, the advantage in coverage for the variance-corrected bootstrap band over the naive bootstrap band is visible, although the margin of improvement is smaller than the CLT case. Third, the bootstrap confidence band has a more accurate coverage than the corresponding CLT confidence band. The larger coverage error of the CLT bands may reflect the limitation of the delta method in capturing the finite sample distribution for the IRFs. On the other hand, the coverage of the bootstrap band is very close to the nominal level for all forecasting horizons. This superior performance suggests that the bootstrap approximation may provide a high order refinement to the first order normal approximation. This is an interesting theoretical question for future research.

Figure 2 shows the median widths of the confidence bands reported in Figure 1. It is clear that the variance-corrected confidence bands, for both the asymptotic and bootstrap ones, are wider than the corresponding naive band. From this figure, we can see that the widths of the confidence bands ranked from high to low are: variance-corrected bootstrap band, naive bootstrap band, variance-corrected CLT band, and naive CLT band. This is generally true for all the other cases that we consider. The next figure, Figure 3, shows the average and median of the relative biases of the IRFs. The biases are measured as

percentages of the true IRFs. We can see that the biases of the one-period-ahead IRF can be either positive or negative, ranging from 8% to almost 50% of the true IRFs. For longer horizon IRFs, the mean biases fall within the 10% range.

We do not report the figures for other parameter configurations but summarize the main results here. We have considered a DGP with higher persistence. In our high persistence case, the eigenvalues λ_1 and λ_2 of A are set to be 0.1 and 0.85. Since λ_2 is closer to 1, the process is more persistent. The rest of the parameters, N, T and T_d, remain the same as in the base case. Compared to the base case, the coverage of the bootstrap band is in similar range but the coverage of the CLT bands improves for this case. A contributing factor to this improvement is that the median bias for this high persistence case is much lower than the base case. When the process becomes more persistent, the signal-to-noise ratios, as measured by $var(y_{m,it})/var(u_{m,it})$, become higher but the instruments become weaker. These two offsetting forces have opposite impacts on the coverage accuracy. When T_d is small and the variance of the fixed effects $var(\mu_i)$ is relatively large, the instruments remain relatively strong. The effect of higher signal-to-noise ratios dominates that of weaker instruments, leading to improved coverage accuracy. Simulations show that for large T_d and small $var(\mu_i)$, the coverage accuracy may deteriorate as the process becomes more persistent.

The next case we consider contains more observations with $N = 200$, the only deviation from the base case. The coverage probability for both the CLT bands and the bootstrap bands increase under larger N. In particular, the coverage of the bootstrap bands closely tracks the 95% nominal coverage probability for all four IRFs and for all forecasting horizons. As one would expect, the mean and median of the relative biases of the IRFs remain the same as in the base case, which confirms that the biases come from the time series dimension rather than the cross sectional dimension.

We also examine a case in which the time series are close to be stationary. The parameter configuration is the same as the base case except that T_d is now equal to 50. The margin of improvement from using the variance-corrected confidence bands shrinks a little comparing with the base case but remains positive and visible in the omitted figure.

The basic qualitative observations are the same for other (N, T) combinations and initialization schemes and for 90% confidence bands. In an overall sense, the bootstrap bands have smaller coverage errors than the corresponding CLT bands, and variance correction is effective in reducing the coverage errors.

5 Conclusion

The paper establishes the asymptotic distribution of the orthogonalized impulse response function for short panel VARs. Due to the correlation between the demeaned regressors and the demeaned error term, the estimator of the autoregressive coefficients and that of the error variance are not independent, even in large samples with a fixed time series dimension. The dependence calls for correction for the asymptotic variance of the orthogonalized impulse response function. In this paper, we have developed the corrected asymptotic formula for both reduced form VAR and structural VAR for short panels. We also have proved the asymptotic validity of the bootstrapped confidence bands in this context.

Our simulation analysis shows that the proposed variance correction leads to confidence bands that have smaller coverage errors. In practical applications, we recommend using the corrected variance to studentize the t-statistic and employ the bootstrap approximation to construct the confidence bands.

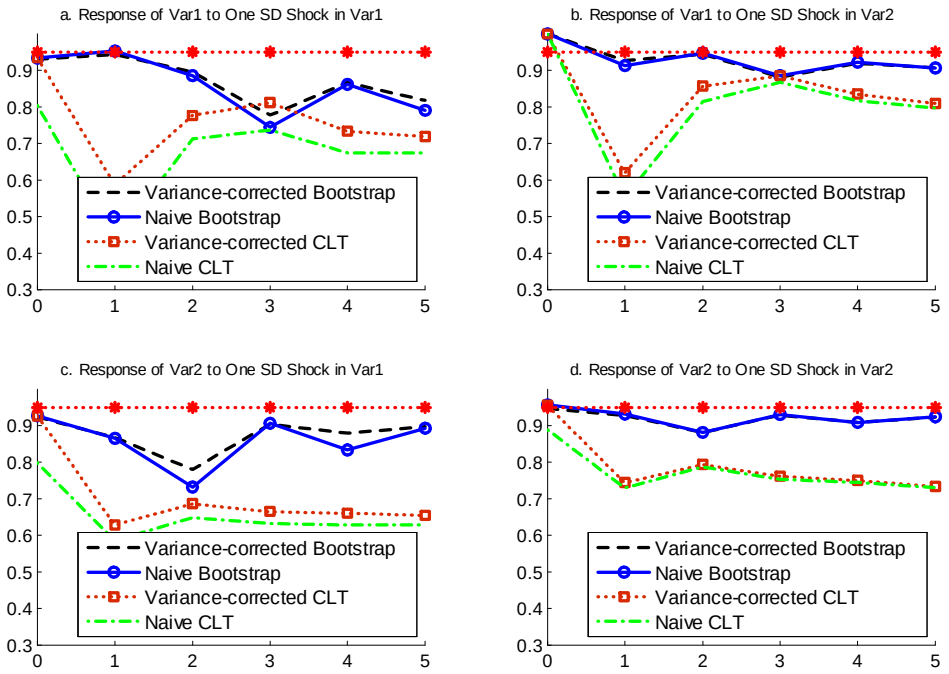

Figure 1: The Empirical Coverage of Different 95% Confidence Bands of the Orthogonalized IRFs for the Base Case with $N = 100, T = 5$, and $T_d = 1$

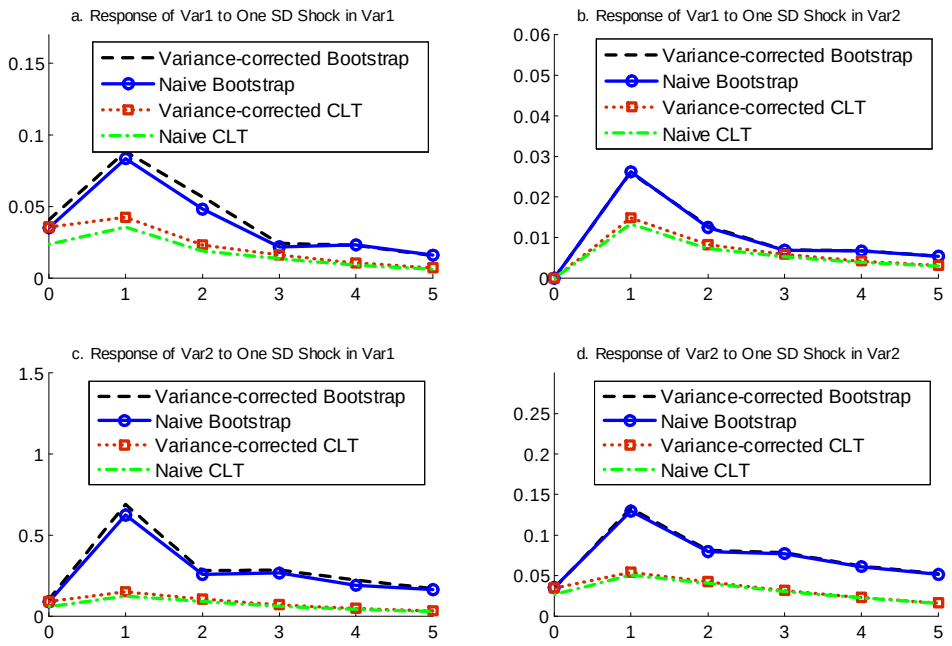

Figure 2: The Median Length Different 95% Confidence Bands of the Orthogonalized IRFs for the Base Case with $N = 100, T = 5$, and $T_d = 1$

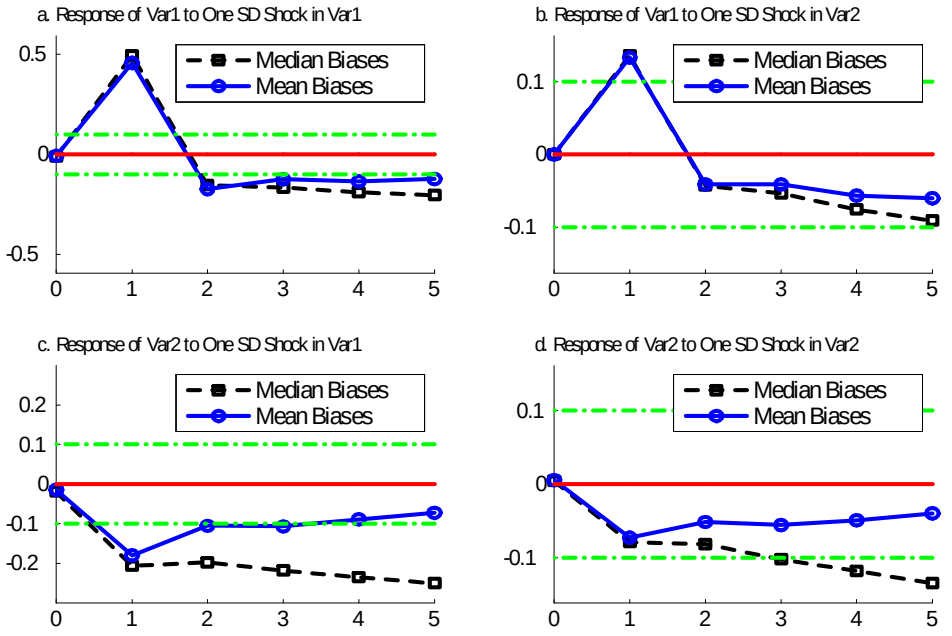

Figure 3: The Mean and Median of the Relative Biases of the Orthogonalized IRFs for the Base Case with $N = 100, T = 5$, and $T_d = 1$

6 Appendix

Proof of Theorem 1. To establish the asymptotic distribution of $\hat{\alpha}_{GMM}$, we only need to verify the conditions for the LLN in (9), (11), and the CLT in (10). Under the cross sectional independence, a sufficient condition for the LLN in (9) is

$$\max_i E\left[\left\|Z_i'\Delta X_i\right\|^2\right] < \infty.$$

But for a generic constant $\mathcal{C}$, which may be different for different occurrences,

$$\begin{aligned}
\max_i E\left[\left\|Z_i'\Delta X_i\right\|^2\right] &\leq \mathcal{C}\max_i \max_{t=1,...,T}\max_{s=1,...,t} E\left\|\Delta X_{i,t} y_{i,t-s}'\right\|^2 \\
&\leq \mathcal{C}\max_i \max_{t=1,...,T}\max_{s=1,...,t}\max_{\tau=1,...,p} E\left\|\Delta y_{i,t-\tau} y_{i,t-s}'\right\|^2 \\
&\leq \mathcal{C}\max_i \max_{t=0,...,T}(E\left\|y_{i,t}\right\|^4) \leq \mathcal{C}.
\end{aligned}$$

The last inequality holds by Assumptions 2(i) and 2(ii). Similarly, we can show that the LLN in (11) holds under Assumptions 2(i) and 2(ii).

To verify the Lyapunov condition for the CLT, we use the Cramer-Wold theorem. Let $Z_{ui} = K'vec\left(Z_i'\Delta u_i\right)$ for any fixed vector K and $\sigma_i^2 = var\left(Z_{ui}\right)$. Under Assumptions 2(i) and 2(ii), we have

$$\begin{aligned}
\max_i E(|Z_{ui}|^{2+\delta}) &= \max_i E(\left|K'vec\left(Z_i'\Delta u_i\right)\right|^{2+\delta}) \\
&\leq \mathcal{C}\max_i E\left\|\left(Z_i'\Delta u_i\right)\right\|^{2+\delta} \\
&\leq \mathcal{C}\max_i \left(E\left\|Z_i\right\|^{4+2\delta}\right)^{1/2}\left(E\left\|\Delta u_i\right\|^{4+2\delta}\right)^{1/2} \leq \mathcal{C}.
\end{aligned}$$

So $\sum_{i=1}^N E(|Z_{ui}|^{2+\delta}) = O(N)$. Under Assumption 2(iii), we have

$$\left(\sum_{i=1}^N \sigma_i^2\right) = NK'E\left[\Sigma \otimes \frac{1}{N}\sum_{i=1}^N Z_i'GZ_i\right]K \geq \mathcal{C}N$$

for some constant $\mathcal{C}$ and large enough N, as $\operatorname{plim} N^{-1}\sum_{i=1}^N Z_i'GZ_i$ is of full rank. This implies that $\left(\sum_{i=1}^N \sigma_i^2\right)^{-1-\delta/2} = o\left(N^{-1-\delta/2}\right)$. As a result,

$$\lim_{N\to\infty}\left(\sum_{i=1}^N \sigma_i^2\right)^{-1-\delta/2}\sum_{i=1}^N E(|Z_{ui}|^{2+\delta}) = 0.$$

That is, the Lyapunov condition holds.

It remains to establish the asymptotic distribution of $\hat{\Sigma}_{GMM}$ and its relationship with $\hat{\alpha}_{GMM}$. Writing $\hat{\Sigma}_{GMM}$ in terms of the unobserved error term, we have

$$\begin{aligned}
\hat{\Sigma}_{GMM} &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=0}^T\left[\left(y_{i,t}-\bar{y}_{i,\cdot}\right) - A'\left(X_{i,t}-\bar{X}_{i,\cdot}\right) - \left(\hat{A}-A\right)'\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\right] \\
&\quad \times\left[\left(y_{i,t}-\bar{y}_{i,\cdot}\right) - A'\left(X_{i,t}-\bar{X}_{i,\cdot}\right) - \left(\hat{A}-A\right)'\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\right]', \\
&= \hat{\Sigma}_0 + I_1 + I_2 + I_3
\end{aligned}$$

where

$$\begin{aligned}
\hat{\Sigma}_0 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(u_{i,t}-\bar{u}_{i,\cdot}\right)', \\
I_1 &= \left(\hat{A}-A\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\left(\hat{A}-A\right), \\
I_2 &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\left(\hat{A}-A\right), \\
I_3 &= -\left(\hat{A}-A\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'.
\end{aligned}$$

In view of $\hat{A}-A=O_p(1/\sqrt{N})$, we have $\sqrt{N}I_1=o_p(1)$. As a result,

$$\sqrt{N}(\hat{\Sigma}_{GMM}-\Sigma)=\sqrt{N}(\hat{\Sigma}_0-\Sigma)+\sqrt{N}I_2+\sqrt{N}I_3+o_p(1). \tag{37}$$

To evaluate I_2 and I_3, we note that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'=B+o_p(1),$$

where the $o_p(1)$ term follows from Assumptions 2(i) and (ii). Therefore,

$$\sqrt{N}I_2=-B\sqrt{N}\left(\hat{A}-A\right)+o_p(1),\ \sqrt{N}I_3=-\sqrt{N}\left(\hat{A}-A\right)'B'+o_p(1). \tag{38}$$

Combining (37) with (38) yields

$$\sqrt{N}\left(\hat{\Sigma}_{GMM}-\Sigma\right)=\sqrt{N}(\hat{\Sigma}_0-\Sigma)-B\sqrt{N}\left(\hat{A}-A\right)-\sqrt{N}\left(\hat{A}-A\right)'B'+o_p(1).$$

To derive the limiting distribution of $\sqrt{N}(\hat{\Sigma}_{GMM}-\Sigma)$, we consider each of the three terms. First,

$$\begin{aligned}
\sqrt{N}(\hat{\Sigma}_0-\Sigma) &= \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\left[\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'-\Sigma\right] \\
&= \frac{1}{\sqrt{N}\left(T+1\right)}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}u_{i,t}'-\Sigma\right)-\frac{1}{\sqrt{N}T\left(T+1\right)}\sum_{i=1}^{N}\sum_{s,t=0,s\neq t}^{T}u_{i,t}u_{i,s}'
\end{aligned} \tag{39}$$

Using the Lyapunov CLT, we have

$$\begin{aligned}
\sqrt{N}vech(\hat{\Sigma}-\Sigma) &= vech\left(\frac{1}{\sqrt{N}\left(T+1\right)}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}u_{i,t}'-\Sigma\right)\right) \\
&\Rightarrow N\left(0,\frac{1}{T+1}D_m^{+}\Lambda_{m^2}\left(D_m^{+}\right)'\right),
\end{aligned} \tag{40}$$

where $\Lambda_{m^2} = var[vec(u_{i,t}u'_{i,t})]$. It is easy to see that the Lyapunov condition holds under Assumption 2(i). Similarly,

$$\begin{aligned} & vech\left(\frac{1}{\sqrt{N}T\left(T+1\right)}\sum_{i=1}^{N}\sum_{t=0}^{T}\sum_{s=0,s\neq t}^{T}u_{i,t}u'_{i,s}\right) \\ \Rightarrow\quad & N\left(0,\frac{1}{T\left(T+1\right)}D_m^+\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\left(D_m^+\right)'\right). \end{aligned}$$

As a result,

$$\begin{aligned} & \sqrt{N}vech\left(\hat{\Sigma}_0-\Sigma\right) \\ \Rightarrow\quad & N\left(0,D_m^+\left[\frac{1}{T+1}\Lambda_{m^2}+\frac{1}{T\left(T+1\right)}\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\right]\left(D_m^+\right)'\right) \end{aligned} \tag{41}$$

where we have used the asymptotic independence between the two terms in (39).

Next, using the properties of the commutation matrix: $K_{m,m}\left(\Sigma\otimes BQ^{-1}B'\right)=\left(BQ^{-1}B'\otimes\Sigma\right)K_{m,m}$ and $K_{m,m}K'_{m,m}=I_{m^2}$, we can show that

$$vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right]\Rightarrow N\left(0,D_m^+\left(\Sigma\otimes BQ^{-1}B'\right)\left(D_m^+\right)'\right),$$

and

$$vech\left[\sqrt{N}\left(\hat{A}-A\right)'B'\right]\Rightarrow N\left(0,D_m^+\left(BQ^{-1}B'\otimes\Sigma\right)\left(D_m^+\right)'\right).$$

In addition,

$$\begin{aligned} & cov\left(vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right],vech\left[\sqrt{N}\left(\hat{A}-A\right)'B'\right]\right) \\ = \quad & D_m^+\left(\Sigma\otimes BQ^{-1}B'\right)K'_{m,m}\left(D_m^+\right)'. \end{aligned}$$

Therefore,

$$B\sqrt{N}\left(\hat{A}-A\right)+\sqrt{N}\left(\hat{A}-A\right)'B'\Rightarrow N(0,V_{AB}),$$

where

$$\begin{aligned} V_{AB} \quad = \quad & D_m^+(\Sigma\otimes BQ^{-1}B')\left(D_m^+\right)'+D_m^+\left(BQ^{-1}B'\otimes\Sigma\right)\left(D_m^+\right)' \\ & +D_m^+(\Sigma\otimes BQ^{-1}B)K'_{m,m}\left(D_m^+\right)'+D_m^+K_{m,m}\left(\Sigma\otimes BQ^{-1}B'\right)\left(D_m^+\right)'. \end{aligned} \tag{42}$$

We proceed to prove that $\sqrt{N}(\hat{\Sigma}_0-\Sigma)$ and $\sqrt{N}(\hat{A}-A)$ are asymptotically independent under Assumption 2(iv). We write

$$\sqrt{N}vec(\hat{A}-A)=M\frac{1}{\sqrt{N}}\sum_{i=1}^{N}vec\left(\sum_{t=1}^{T}Z'_{i,t-2}\left(u_{i,t}-u_{i,t-1}\right)\right)+o_p\left(1\right)$$

for some matrix M. Note that

$$\begin{aligned}
& E\sum_{t=0}^{T} vec\left(u_{i,t}u_{i,t}'-\Sigma\right)\sum_{t=1}^{T}\left[vec\left(Z_{i,t-2}'u_{i,t}\right)\right]' \\
= \; & E\sum_{t=0}^{T} vec\left(u_{i,t}u_{i,t}'\right)\sum_{t=1}^{T}\left[vec\left(Z_{i,t-2}'u_{i,t}\right)\right]' = E\sum_{t=0}^{T}\sum_{s=1}^{T}\left(u_{i,t}\otimes u_{i,t}\right)\left(u_{i,s}'\otimes Z_{is-2}\right) \\
= \; & E\sum_{t=0}^{T}\sum_{s=1}^{T}\left(u_{i,t}u_{i,s}'\otimes u_{i,t}Z_{is-2}\right) = E\sum_{t=1}^{T}\sum_{s=1}^{T}\left(u_{i,t}u_{i,s}'\otimes u_{i,t}Z_{is-2}\right) \\
= \; & E\sum_{s=1}^{T}\sum_{t\neq s}\left(u_{i,t}u_{i,s}'\otimes u_{i,t}Z_{is-2}\right)+E\sum_{t=1}^{T}\left(u_{i,t}u_{i,t}'\otimes u_{i,t}Z_{i,t-2}\right) \\
= \; & E\sum_{t=1}^{T}\left(u_{i,t}u_{i,t}'\otimes u_{i,t}Z_{i,t-2}\right) = E\sum_{t=1}^{T}\left[\left(u_{i,t}u_{i,t}'\otimes u_{i,t}\right)\otimes Z_{i,t-2}\right] = 0
\end{aligned}$$

by Assumption 2(iv). Similarly,

$$E\sum_{t=0}^{T} vec\left(u_{i,t}u_{i,t}'-\Sigma\right)\sum_{t=1}^{T}\left[vec\left(Z_{i,t-2}'u_{i,t-1}\right)\right]' = 0.$$

Hence the first term in (39) is asymptotically independent of $\sqrt{N}(\hat{A}-A)$. It is easy to see that the asymptotic independence also holds for the second term in (39).

As a result,

$$\sqrt{N}vech\left(\hat{\Sigma}_{GMM}-\Sigma\right)\rightarrow N(0,\Omega_{\sigma\sigma}), \tag{43}$$

where $\Omega_{\sigma\sigma}$ is defined in the theorem.

Finally, we examine the asymptotic covariance between $vech[\sqrt{N}(\hat{\Sigma}_{GMM}-\Sigma)]$ and $vec[\sqrt{N}(\hat{A}-A)]$. Since $\sqrt{N}(\hat{\Sigma}_0-\Sigma)$ is asymptotically independent of $\sqrt{N}(\hat{A}-A)$, the asymptotic covariance is given by

$$\begin{aligned}
\Omega_{\alpha,\sigma} &= -cov\left(vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right], vec\left[\sqrt{N}\left(\hat{A}-A\right)\right]\right) \\
&\quad -cov\left(vech\left[\sqrt{N}\left(\hat{A}-A\right)'B'\right], vec\left[\sqrt{N}\left(\hat{A}-A\right)\right]\right) \\
&= -D_m^+(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right)-D_m^+K_{m,m}(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right).
\end{aligned} \tag{44}$$

Combining (43), (44) and $\sqrt{N}\left(\hat{\alpha}_{GMM}-\alpha\right)\Rightarrow N(0,\Omega_{\alpha\alpha})$ completes the proof of the theorem. ■

Lemma 1 *Let the assumptions in Theorem 3 hold, then for any conformable vector c :*

$$\begin{aligned}
& \mathbb{P}\left(c'\begin{pmatrix}\sqrt{N}\left(\hat{\alpha}_{GMM}^*-\hat{\alpha}_{GMM}\right)\\ \sqrt{N}vech(\hat{\Sigma}_{GMM}^*-\hat{\Sigma}_{GMM})\end{pmatrix}<x\right) \\
= \; & \mathbb{P}\left(c'\begin{pmatrix}\sqrt{N}\left(\hat{\alpha}_{GMM}-\alpha\right)\\ \sqrt{N}vech(\hat{\Sigma}_{GMM}-\Sigma)\end{pmatrix}<x\right)+o\left(1\right)
\end{aligned}$$

uniformly over $x\in\mathbb{R}$ as $N\rightarrow\infty$ for a fixed T.

Proof of Lemma 1. The result is on the joint convergence of $\sqrt{N}\left(\hat{\alpha}^*_{GMM} - \hat{\alpha}_{GMM}\right)$ and $\sqrt{N}(\hat{\Sigma}^*_{GMM} - \hat{\Sigma}_{GMM})$. We prove only the convergence of the marginal distributions, as the joint convergence follows easily from the same argument we present here. The proof consists of two parts.

Part (a) Proof of $\sqrt{N}\left(\hat{\alpha}^*_{GMM} - \hat{\alpha}_{GMM}\right) \Rightarrow N(0, \Omega_{\alpha\alpha})$.

By construction, we have

$$\begin{aligned}
&\hat{\alpha}^*_{GMM} - \hat{\alpha}_{GMM} \\
= \;& vec\left\{\left[\left(\sum_{i=1}^{N}(\Delta X_i^*)' Z_i^*\right)\left(\sum_{i=1}^{N}(Z_i^*)' G Z_i^*\right)^{-1}\left(\sum_{i=1}^{N}(Z_i^*)' \Delta X_i^*\right)\right]^{-1}\right. \\
&\left.\times\left(\sum_{i=1}^{N}(\Delta X_i^*)' Z_i^*\right)\left(\sum_{i=1}^{N}(Z_i^*)' G Z_i^*\right)^{-1}\sum_{i=1}^{N}(Z_i^*)' \Delta u_i^*\right\}. \qquad (45)
\end{aligned}$$

Define the set of samples $\mathcal{E}_\alpha$ as

$$\begin{aligned}
\mathcal{E}_\alpha = \Big\{&[y_1(\omega), ..., y_N(\omega)]: \\
&(i)\ N^{-1}\textstyle\sum_{i=1}^{N}\Delta X_i' Z_i \to S_{XZ}^{\infty}, \\
&(ii)\ N^{-1}\textstyle\sum_{i=1}^{N} Z_i' G Z_i \to S_{ZZ}^{\infty}, \\
&(iii)\ N^{-2}\textstyle\sum_{i=1}^{N} vec\left[(\Delta X_i)' Z_i\right]\left\{vec\left[(\Delta X_i)' Z_i\right]\right\}' \to 0, \\
&(iv)\ N^{-2}\textstyle\sum_{i=1}^{N} vec\left(Z_i' G Z_i\right)\left\{vec\left[Z_i' G Z_i\right]\right\}' \to 0, \\
&(v)\ \left[\textstyle\sum_{i=1}^{N}|Z_{ui}|^2\right]^{-1-\delta/2}\textstyle\sum_{i=1}^{N}|Z_{ui}|^{2+\delta} \to 0 \\
&(vi)\ \hat{A} \to A\Big\}
\end{aligned}$$

In the above definition, $S_{XZ}^{\infty} = (S_{ZX}^{\infty})'$ and $Z_{ui} = K' vec\left(Z_i' \Delta u_i\right)$ for any fixed vector K. Under the moment conditions in the theorem, we have

$$\begin{aligned}
&\sum_{N=1}^{\infty}\mathbb{P}\left(\left\|N^{-1}\sum_{i=1}^{N}\left(\Delta X_i' Z_i - S_{XZ}^{\infty}\right)\right\| > \varepsilon\right) \\
\leq\; &\sum_{N=1}^{\infty}\frac{E\left[\left\|\sum_{i=1}^{N}\left(\Delta X_i' Z_i - S_{XZ}^{\infty}\right)\right\|^4\right]}{N^4\varepsilon^4} = O\left(\sum_{N=1}^{\infty}\frac{1}{N^2\varepsilon^4}\right) < \infty \qquad (46)
\end{aligned}$$

It follows from the Borel-Cantelli lemma that $N^{-1}\sum_{i=1}^{N}\Delta X_i' Z_i \to S_{XZ}^{\infty}$ almost surely. So condition (i) in the definition of $\mathcal{E}_\alpha$ holds almost surely. Similarly, we can show conditions (ii)-(iv) and (vi) hold almost surely. To show that condition (v) holds almost surely, we let

$\bar{Z}_{ui} = |Z_{ui}|^{2+\delta} - E\left(|Z_{ui}|^{2+\delta}\right)$ and note that

$$\begin{aligned}
\sum_{N=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N} \bar{Z}_{ui}\right| > \varepsilon\right) &\leq \sum_{N=1}^{\infty} \frac{E\left(\sum_{i=1}^{N} \bar{Z}_{ui}\right)^4}{N^4\varepsilon^4} \\
&= \sum_{N=1}^{\infty} \frac{E\left(\sum_{i=1}^{N} \bar{Z}_{ui}^4\right)}{N^4\varepsilon^4} + \sum_{N=1}^{\infty} \frac{\left[\left(\sum_{i=1}^{N} E\bar{Z}_{ui}^2\right)\right]^2}{N^4\varepsilon^4} \\
&\leq \sum_{N=1}^{\infty} \frac{E\left(\sum_{i=1}^{N} \bar{Z}_{ui}^4\right)}{N^4\varepsilon^4} + \sum_{N=1}^{\infty} \frac{\left[\left(\sum_{i=1}^{N} E\bar{Z}_{ui}^2\right)\right]^2}{N^4\varepsilon^4} < \infty
\end{aligned}$$

where the last line follows because

$$E\left(\bar{Z}_{ui}^4\right) = E\left(|Z_{ui}|^{2+\delta} - E|Z_{ui}|^{2+\delta}\right)^4 = O\left(E|Z_{ui}|^{8+4\delta}\right) = O\left(E\|Z_i\|^{16+8\delta} E\|u_i\|^{16+8\delta}\right) = O(1)$$

using the moment conditions in the theorem. So $\sum_{i=1}^{N} |Z_{ui}|^{2+\delta} = O_{a.s.}(N)$. Similarly, we can show that $\left[\sum_{i=1}^{N} |Z_{ui}|^2\right]^{-1-\delta/2} = O_{a.s.}\left(N^{-1-\delta/2}\right)$. Consequently, condition (v) indeed holds almost surely. We can conclude that $\mathbb{P}(\mathcal{E}_\alpha) = 1$.

Conditional on the sample in $\mathcal{E}_\alpha$, $(\Delta X_i^*)' Z_i^*$ is a triangular array of rowwise iid random variables and a law of large numbers gives

$$\frac{1}{N}\sum_{i=1}^{N} (\Delta X_i^*)' Z_i^* - E^*\left[(\Delta X_i^*)' Z_i^*\right] = o_{p^*}(1) \tag{47}$$

where E^* denotes the expectation with respect to the bootstrap distribution $\mathbb{P}^*$, and $o_{p^*}(1)$ denotes a sequence of random variables that converges to zero in probability under $\mathbb{P}^*$. By definition,

$$E^*\left[(\Delta X_i^*)' Z_i^*\right] = \frac{1}{N}\sum_{i=1}^{N} \Delta X_i' Z_i \to S_{XZ}^{\infty}.$$

Combining this with (47) yields

$$\frac{1}{N}\sum_{i=1}^{N} (\Delta X_i^*)' Z_i^* = S_{XZ}^{\infty} + o_p^*(1). \tag{48}$$

By the same argument, we have, conditional on $\mathcal{E}_\alpha$,

$$\frac{1}{N}\sum_{i=1}^{N} (Z_i^*)' G Z_i^* = S_{ZZ}^{\infty} + o_{p^*}(1). \tag{49}$$

Therefore,

$$\left[\left(\frac{1}{N}\sum_{i=1}^{N} (\Delta X_i^*)' Z_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N} (Z_i^*)' G Z_i^*\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} (Z_i^*)' \Delta X_i^*\right)\right]^{-1} = Q + o_p^*(1). \tag{50}$$

Next

$$\left(\frac{1}{N}\sum_{i=1}^{N}\left(\Delta X_i^*\right)' Z_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(Z_i^*\right)' G Z_i^*\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i^*\right)'\Delta u_i^* := I_4 + I_5$$

where

$$\begin{aligned} I_4 &= \left(\frac{1}{N}\sum_{i=1}^{N}\left(\Delta X_i^*\right)' Z_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(Z_i^*\right)' G Z_i^*\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E^*\left[\left(Z_i^*\right)'\Delta u_i^*\right], \\ I_5 &= \left(\frac{1}{N}\sum_{i=1}^{N}\left(\Delta X_i^*\right)' Z_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(Z_i^*\right)' G Z_i^*\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(Z_i^*\right)'\Delta u_i^* - E^*\left[\left(Z_i^*\right)'\Delta u_i^*\right]\right\}. \end{aligned}$$

For I_4, we use (48) and (49) to obtain:

$$\begin{aligned} I_4 &= \left(\frac{1}{N}\sum_{i=1}^{N}\Delta X_i' Z_i + o_p^*(1)\right)\left(\frac{1}{N}\sum_{i=1}^{N}Z_i' G Z_i + o_p^*(1)\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i'\Delta\hat{u}_i\right) \\ &= o_p^*(1) \end{aligned}$$

where the second equality holds because

$$\frac{1}{N}\sum_{i=1}^{N}\Delta X_i' Z_i\left(\frac{1}{N}\sum_{i=1}^{N}Z_i' G Z_i\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i'\Delta\hat{u}_i = 0,$$

by the definition of the GMM estimator $\hat{\alpha}_{GMM}$.

For I_5, we have, using (48) and (49) again:

$$I_5 = \left[S_{XZ}^{\infty} + o_p^*(1)\right]\left[S_{ZZ}^{\infty} + o_p^*(1)\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(Z_i^*\right)'\Delta u_i^* - E^*\left[\left(Z_i^*\right)'\Delta u_i^*\right]\right\}.$$

By the Lyapunov CLT for triangular arrays, we have

$$\begin{aligned} & vec\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\left(Z_i^*\right)'\Delta u_i^* - E^*\left[\left(Z_i^*\right)'\Delta u_i^*\right]\right)\right\} \\ = \;& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left[I_m\otimes\left(Z_i^*\right)'\right]vec(\Delta u_i^*) - E^*\left(\left[I_m\otimes\left(Z_i^*\right)'\right]vec(\Delta u_i^*)\right)\right\} \\ \Rightarrow \;& N(0,\Sigma\otimes S_{ZZ}^{\infty}). \end{aligned}$$

Define $Z_{ui}^* = K' vec\left[\left(Z_i^*\right)'\Delta u_i^*\right]$ for any conformable constant vector K. Then the Lyapunov condition is

$$\frac{1}{\left[\sum_{i=1}^{N}E^*\left(\left|Z_{ui}^*\right|^2\right)\right]^{1+\delta/2}}\sum_{i=1}^{N}E^*\left|Z_{ui}^*\right|^{2+\delta}\to 0 \text{ for some } \delta>0.$$

That is

$$\frac{1}{\left[\sum_{i=1}^{N}\left|Z_{ui}\right|^2\right]^{1+\delta/2}}\sum_{i=1}^{N}\left|Z_{ui}\right|^{2+\delta}\to 0 \text{ for some } \delta>0.$$

This condition holds conditional on the sample in $\mathcal{E}_\alpha$. Therefore

$$I_5 \Rightarrow N(0, S_{XZ}^{\infty}\left[S_{ZZ}^{\infty}\right]^{-1}\left[S_{XZ}^{\infty}\right]') \equiv N(0, Q).$$

As a result,

$$\begin{aligned} & vec\left[\left(\frac{1}{N}\sum_{i=1}^{N}\left(\Delta X_i^*\right)' Z_i^*\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(Z_i^*\right)' G Z_i^*\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(Z_i^*\right)'\Delta u_i^*\right] \\ \Rightarrow \quad & N\left(0, \left[I_m \otimes S_{XZ}^{\infty}\left(S_{ZZ}^{\infty}\right)^{-1}\right](\Sigma \otimes S_{ZZ}^{\infty})\left[I_m \otimes \left(S_{ZZ}^{\infty}\right)^{-1}\left(S_{XZ}^{\infty}\right)'\right]\right) \\ \equiv \quad & N(0, \Sigma \otimes Q). \end{aligned}$$

Together with (50), this leads to

$$\begin{aligned} \sqrt{N}\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) \quad & \Rightarrow \quad N\left[0, \left(I_m \otimes Q^{-1}\right)(\Sigma \otimes Q)\left(I_m \otimes Q^{-1}\right)\right] \\ & \equiv \quad N\left(0, \Sigma \otimes Q^{-1}\right) \end{aligned}$$

conditional on $\mathcal{E}_\alpha$. Equivalently, for any conformable vector c_α,

$$\mathbb{P}^*\left(\sqrt{N}c_\alpha'\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) < x | \mathcal{E}_\alpha\right) \to \Phi\left(\frac{x}{c_\alpha'\left[\Sigma \otimes Q^{-1}\right]c_\alpha}\right).$$

Since $\mathbb{P}(\mathcal{E}_\alpha) = 1$, the above conditional result implies the following unconditional almost sure convergence result:

$$\mathbb{P}^*\left(\sqrt{N}c_\alpha'\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) < x\right) \to \Phi\left(\frac{x}{c_\alpha'\left[\Sigma \otimes Q^{-1}\right]c_\alpha}\right) \text{ a.s. as } N \to \infty$$

for all $x \in \mathbb{R}$. This and the dominated convergence theorem imply

$$\begin{aligned} \mathbb{P}\left(\sqrt{N}c_\alpha'\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) < x\right) \quad & = \quad EP^*\left(\sqrt{N}c_\alpha'\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) < x\right) \\ & \to \quad \Phi\left(\frac{x}{c_\alpha'\left[\Sigma \otimes Q^{-1}\right]c_\alpha}\right) \text{ for all } x \in \mathbb{R}, \end{aligned}$$

or equivalently

$$\sqrt{N}\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right) \to N\left(0, \Sigma \otimes Q^{-1}\right).$$

Thus, the unconditional asymptotic distribution of the bootstrap statistic $\sqrt{N}\left(\hat{\alpha}_{GMM}^* - \hat{\alpha}_{GMM}\right)$ is $N\left(0, \Sigma \otimes Q^{-1}\right)$, which is the same as that of the original-sample statistic $\sqrt{N}\left(\hat{\alpha}_{GMM} - \alpha\right)$.

Part (b): Proof of $\sqrt{N}(\hat{\Sigma}_{GMM}^* - \hat{\Sigma}_{GMM}) \Rightarrow N(0, \Omega_{\sigma\sigma})$.

Note that

$$\begin{aligned} \hat{u}_{i,t}^* \quad & = \quad \left(y_{i,t}^* - \bar{y}_{i,\cdot}^*\right) - \left(\hat{A}^*\right)'\left(X_{i,t}^* - \bar{X}_{i,\cdot}^*\right) \\ & = \quad \left(y_{i,t}^* - \bar{y}_{i,\cdot}^*\right) - \hat{A}\left(X_{i,t}^* - \bar{X}_{i,\cdot}^*\right) - \left(\hat{A}^* - \hat{A}\right)'\left(X_{i,t}^* - \bar{X}_{i,\cdot}^*\right) \\ & = \quad u_{i,t}^* - \bar{u}_{i,\cdot}^* - \left(\hat{A}^* - \hat{A}\right)'\left(X_{i,t}^* - \bar{X}_{i,\cdot}^*\right), \end{aligned}$$

we have

$$\hat{\Sigma}_{GMM}^* = \hat{\Sigma}_0^* + I_1^* + I_2^* + I_3^*,$$

where

$$\begin{aligned}
\hat{\Sigma}_0^* &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}^*-\bar{u}_{i,\cdot}^*\right)\left(u_{i,t}^*-\bar{u}_{i,\cdot}^*\right)', \\
I_1^* &= \left(\hat{A}^*-\hat{A}\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(X_{i,t}^*-\bar{X}_{i,\cdot}^*\right)\left(X_{i,t}^*-\bar{X}_{i,\cdot}^*\right)'\left(\hat{A}^*-\hat{A}\right), \\
I_2^* &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}^*-\bar{u}_{i,\cdot}^*\right)\left(X_{i,t}^*-\bar{X}_{i,\cdot}^*\right)'\left(\hat{A}^*-\hat{A}\right), \\
I_3^* &= -\left(\hat{A}^*-\hat{A}\right)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(X_{i,t}^*-\bar{X}_{i,\cdot}^*\right)\left(u_{i,t}^*-\bar{u}_{i,\cdot}^*\right)'.
\end{aligned}$$

As in the proof of part (a), we first derive the asymptotic distribution conditional on an event that occurs with probability one. Define $\mathcal{E}=\mathcal{E}_\alpha\cap\mathcal{E}_\sigma$ where

$$\begin{aligned}
\mathcal{E}_\sigma= &\Big\{[y_1(\omega),...,y_N(\omega)]: \\
&(i)\ (NT)^{-1}\textstyle\sum_{i=1}^{N}\sum_{t=0}^{T}\left(X_{i,t}-\bar{X}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\to S_{XX}^{\infty}, \\
&(ii)\ (NT)^{-1}\textstyle\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'\to B, \\
&(iii)\ (NT)^{-1}\textstyle\sum_{i=1}^{N}\sum_{t=0}^{T}\left[\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\right]\to\Sigma, \\
&(iv)\ V_N\to D_m^+\left[T^{-1}\Lambda_{m^2}+\left[T\left(T+1\right)\right]^{-1}\left(\Sigma\otimes\Sigma\right)\left(I_{m^2}+K_{m,m}\right)\right]\left(D_m^+\right)', \\
&(v)\ N^{-1}\textstyle\sum_{i=1}^{N}\sum_{t=0}^{T}vech\left[\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\right]vec(Z_i\Delta u_i)\to 0\Big\}
\end{aligned}$$

and

$$\begin{aligned}
V_N &= \frac{1}{T^2}\frac{1}{N}\sum_{i=1}^{N}vech\left[\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\right] \\
&\quad\times\left\{vech\left[\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\right]\right\}' \\
&\quad-\frac{1}{T^2N^2}\sum_{j=1}^{N}vech\left[\sum_{t=0}^{T}\left(u_{j,t}-\bar{u}_{i,\cdot}\right)'\left(u_{j,t}-\bar{u}_{i,\cdot}\right)\right] \\
&\quad\times\sum_{i=1}^{N}vech\left[\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\right]'.
\end{aligned}$$

Under the moment conditions in the theorem, a strong law of large number implies that $\mathbb{P}\left(\mathcal{E}_\sigma\right)=1$.

In view of part (a), we have $I_1^*=o_p^*(1/\sqrt{N})$. So

$$\sqrt{N}(\hat{\Sigma}_{GMM}^*-\hat{\Sigma}_{GMM})=\sqrt{N}(\hat{\Sigma}_0^*-\hat{\Sigma}_{GMM})+\sqrt{N}I_2^*+\sqrt{N}I_3^*+o_p^*\left(1\right). \tag{51}$$

Conditional on $\mathcal{E}$,

$$\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)\left(X_{i,t}^{*}-\bar{X}_{i,\cdot}^{*}\right)' \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\hat{u}_{i,t}\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'+o_{p}^{*}\left(1\right) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\left(X_{i,t}-\bar{X}_{i,\cdot}\right)'+o_{p}^{*}\left(1\right)=B+o_{p}^{*}\left(1\right).
\end{aligned}$$

Therefore,

$$\sqrt{N}I_{2}^{*}=-B\sqrt{N}\left(\hat{A}^{*}-\hat{A}\right)+o_{p}^{*}\left(1\right),\ \sqrt{N}I_{3}^{*}=-\sqrt{N}\left(\hat{A}^{*}-\hat{A}\right)'B'+o_{p}^{*}\left(1\right). \tag{52}$$

Combining (51) with (52) yields

$$\sqrt{N}(\hat{\Sigma}_{GMM}^{*}-\hat{\Sigma}_{GMM})=\sqrt{N}(\hat{\Sigma}_{0}^{*}-\hat{\Sigma}_{GMM})-B\sqrt{N}\left(\hat{A}^{*}-\hat{A}\right)-\sqrt{N}\left(\hat{A}^{*}-\hat{A}\right)'B'+o_{p}^{*}\left(1\right).$$

Now

$$\begin{aligned}
& \sqrt{N}(\hat{\Sigma}_{0}^{*}-\hat{\Sigma}_{GMM}) \\
= & \frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=0}^{T}\left\{\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)'-E^{*}\left[\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)'\right]\right\}+o_{p}^{*}\left(1\right).
\end{aligned}$$

Conditional on $\mathcal{E}$, the first term is a normalized sum of iid random variables with mean zero. By the Lyapunov CLT for triangular arrays, we have

$$vech\left[\sqrt{N}\left(\hat{\Sigma}_{0}^{*}-\hat{\Sigma}_{GMM}\right)\right]\rightarrow N(0,V)$$

where

$$\begin{aligned}
V = & \lim_{N\rightarrow\infty}\frac{1}{T^{2}}var^{*}\left\{\sum_{t=0}^{T}vech[\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)\left(u_{i,t}^{*}-\bar{u}_{i,\cdot}^{*}\right)']\right\} \\
= & \lim_{N\rightarrow\infty}\frac{1}{T^{2}}\frac{1}{N}\sum_{i=1}^{N}vech\left[\sum_{t=0}^{T}\hat{u}_{i,t}\hat{u}_{i,t}'\right]\left\{vech\left[\sum_{t=0}^{T}\hat{u}_{i,t}\hat{u}_{i,t}'\right]\right\}' \\
& -\lim_{N\rightarrow\infty}\frac{1}{T^{2}N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}vech\left[\sum_{t=0}^{T}\hat{u}_{i,t}\hat{u}_{i,t}'\right]\left\{vech\left[\sum_{t=0}^{T}\hat{u}_{j,t}\hat{u}_{j,t}'\right]\right\}'
\end{aligned}$$

and var^{*} is the variance operator under $\mathbb{P}^{*}$. By definition,

$$\hat{u}_{i,t}=u_{i,t}-\bar{u}_{i,\cdot}-\left(X_{i,t}-\bar{X}_{i,\cdot}\right)(\hat{A}-A),$$

so conditional on $\mathcal{E}$, we have

$$vech\left[\sum_{t=0}^{T}\hat{u}_{i,t}\hat{u}_{i,t}'\right]=vech\left[\sum_{t=0}^{T}\left(u_{i,t}-\bar{u}_{i,\cdot}\right)'\left(u_{i,t}-\bar{u}_{i,\cdot}\right)\right]+o(1).$$

Consequently,

$$V = \lim_{N\to\infty} V_N = D_m^+ \left[\frac{1}{T+1}\Lambda_{m^2} + \frac{1}{T(T+1)}(\Sigma\otimes\Sigma)(I_{m^2}+K_{m,m})\right](D_m^+)'.$$

We have therefore shown that, conditional on $\mathcal{E}$,

$$\begin{aligned} & vech\left[\sqrt{N}\left(\hat{\Sigma}_0^* - \hat{\Sigma}_{GMM}\right)\right] \\ \Rightarrow \quad & N\left(0, D_m^+\left[\frac{1}{T+1}\Lambda_{m^2} + \frac{1}{T(T+1)}(\Sigma\otimes\Sigma)(I_{m^2}+K_{m,m})\right](D_m^+)'\right). \end{aligned}$$

Next, using part (a), we can show that

$$B\sqrt{N}\left(\hat{A}^* - \hat{A}\right) + \sqrt{N}\left(\hat{A}^* - \hat{A}\right)' B' \Rightarrow N(0, V_{AB}),$$

where V_{AB} is defined in (42).

Finally, we prove that $\sqrt{N}(\hat{\Sigma}_0^* - \hat{\Sigma}_{GMM})$ and $\sqrt{N}(\hat{A}^* - \hat{A})$ are asymptotically independent conditional on $\mathcal{E}$. That is, we need to show

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=0}^{T} vech\left[\left(u_{i,t}^* - \bar{u}_{i,\cdot}^*\right)\left(u_{i,t}^* - \bar{u}_{i,\cdot}^*\right)'\right]$$

is asymptotically independent of

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} vec\left\{(Z_i^*)'\Delta u_i^* - E^*\left[(Z_i^*)'\Delta u_i^*\right]\right\}.$$

Their covariance is

$$\begin{aligned} & cov^*\left(\sum_{t=0}^{T} vech\left[\left(u_{i,t}^* - \bar{u}_{i,\cdot}^*\right)\left(u_{i,t}^* - \bar{u}_{i,\cdot}^*\right)'\right], vec\left[(Z_i^*)'\Delta u_i^*\right]\right) \\ = \quad & \frac{1}{N}\sum_{i=1}^{N} vech(\hat{u}_i'\hat{u}_i) vec(Z_i\Delta\hat{u}_i) - \left[\frac{1}{N}\sum_{i=1}^{N} vec(\hat{u}_i'\hat{u}_i)\right]\left[\frac{1}{N}\sum_{i=1}^{N} vec(Z_i\Delta\hat{u}_i)\right] \\ = \quad & \frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T} vech\left[(u_{i,t} - \bar{u}_{i,\cdot})(u_{i,t} - \bar{u}_{i,\cdot})'\right] vec(Z_i\Delta u_i) \\ & - \left[\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T}(u_{i,t} - \bar{u}_{i,\cdot})(u_{i,t} - \bar{u}_{i,\cdot})'\right]\left[\frac{1}{N}\sum_{i=1}^{N} vec(Z_i\Delta u_i)\right] + o(1) \\ \to \quad & 0. \end{aligned}$$

As a result, conditional on $\mathcal{E}$, we have

$$\sqrt{N} vech\left(\hat{\Sigma}_{GMM}^* - \hat{\Sigma}_{GMM}\right) \Rightarrow N(0, \Omega_{\sigma\sigma}). \tag{53}$$

Using the same argument as in part (a), we can show that this conditional convergence result implies unconditional convergence. That is, for any conformable vector c_σ,

$$\mathbb{P}\left[\sqrt{N}c_\sigma' vech\left(\hat{\Sigma}_{GMM}^* - \hat{\Sigma}_{GMM}\right) < x\right] \to \Phi\left(\frac{x}{c_\sigma'\Omega_{\sigma\sigma}c_\sigma}\right).$$

In this result, the randomness with respect to both the bootstrap sample and the original sample is taken into account.

Part (c): Proof of Uniformity over $x \in \mathbb{R}$.

It follows from parts (a) and (b) that

$$\begin{aligned} & \mathbb{P}\left(c'\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}^*_{GMM} - \hat{\alpha}_{GMM}\right) \\ \sqrt{N}vech(\hat{\Sigma}^*_{GMM} - \hat{\Sigma}_{GMM}) \end{pmatrix} < x\right) \\ = \; & \mathbb{P}\left(c'\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}_{GMM} - \alpha_{GMM}\right) \\ \sqrt{N}vech(\hat{\Sigma}_{GMM} - \Sigma_{GMM}) \end{pmatrix} < x\right) + o\left(1\right) \end{aligned}$$

for any given $x \in \mathbb{R}$. By Polya's theorem (see DasGupta, 2008, p. 3), the above pointwise result holds uniformly over $x \in \mathbb{R}$. ■

Proof of Theorem 3. Part (a) For any conformable vector c, we can write

$$c'vec[(\hat{\Phi}^*_j)'] = F_j(\hat{A}^*) \text{ and } c'vec[(\hat{\Phi}_j)'] = F_j(\hat{A})$$

where F_j is a continuously differentiable function. By the Delta method, we have

$$\mathbb{P}\{c'vec[\sqrt{N}(\hat{\Phi}^*_j - \hat{\Phi}_j)'] < x\} = \mathbb{P}\{F^{(1)}_j(\hat{A})vec[\sqrt{N}(\hat{A}^* - \hat{A})] < x\} + o\left(1\right)$$

and

$$\mathbb{P}\{c'vec[\sqrt{N}(\hat{\Phi}_j - \Phi_j)'] < x\} = \mathbb{P}\{F^{(1)}_j(A)vec[\sqrt{N}(\hat{A} - A)] < x\} + o\left(1\right)$$

where $F^{(1)}_j\left(A\right) = \partial F_j\left(A\right)/\partial vec(A)'$. Combining the above two equations with

$$\frac{\partial F_j\left(A\right)}{\partial vec(A)'}|_{A=\hat{A}} = \frac{\partial F_j\left(A\right)}{\partial vec(A)'} + o_p\left(1\right),$$

which holds by the consistency of $\hat{A}$ and the continuity of $F^{(1)}_j\left(\cdot\right)$, yields

$$\mathbb{P}\{c'vec[\sqrt{N}(\hat{\Phi}^*_j - \hat{\Phi}_j)'] < x\} = \mathbb{P}\{c'vec[\sqrt{N}(\hat{\Phi}_j - \Phi_j)'] < x\} + o\left(1\right).$$

Invoking the Polya's theorem gives the desired result.

Part (b). The same argument for part (a) applies. Details are omitted.

Parts (c)(d). It is sufficient to show that $\hat{\Omega}_{\Phi j} = \Omega_{\Phi j} + o_p\left(1\right)$, $\hat{\Omega}_{\Theta j} = \Omega_{\Theta j} + o_p\left(1\right)$, $\hat{\Omega}^*_{\Phi j} = \Omega_{\Phi j} + o_p\left(1\right)$, and $\hat{\Omega}^*_{\Phi j} = \Omega_{\Phi j} + o_p\left(1\right)$. The proof is straightforward and details are omitted here. ■

References

[1] Anderson, T.W., and C. Hsiao (1981): Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association,* 76, 598–606.

[2] Anderson, T.W., and C. Hsiao (1982): Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 47–82.

[3] Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press, USA.

[4] Arellano, M., and S.R. Bond (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies,* 58, 277–297.

[5] Baillie, R.T. (1987): Inference in Dynamic Models Containing 'Surprise' Variables, *Journal of Econometrics* 35, 101–117.

[6] Bilodeau, M. and D. Brenner, (1999): *Theory of Multivariate Statistics,* Springer.

[7] Binder, M., C. Hsiao, and M.H. Pesaran (2005): Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, *Econometric Theory,* 21, 795–837.

[8] Blundell, R., and S.R. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115–143.

[9] Bun, M.J.G. and J.F. Kiviet (2006): The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models, *Journal of Econometrics,* Vol. 132(2), 409–444.

[10] DasGupta, A. (2008): *Asymptotic Theory of Statistics and Probability*, Springer.

[11] Gilchrist, S., C.P. Himmelberg, and G. Huberman (2005): Do Stock Price Bubbles Influence Corporate Investment? *Journal of Monetary Economics,* 52(4), 805–827.

[12] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag New York, Inc.

[13] Hamilton, J. (1994). *Time Series Analysis*, Princeton University Press.

[14] Hansen, L.P. (1982): Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, 1029–1054.

[15] Holtz-Eakin, D., W.K. Newey, and H.S. Rosen (1988): Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371–1395.

[16] Horowitz, J.L. (1997): Bootstrap Methods in Econometrics: Theory and Numerical Performance, in *Advances in Economics and Econometrics: Theory and Applications*, Vol. III, Econometric Society Monographs No. 28, ed. by D.M. Kreps and K.F. Wallis, Cambridge University Press.

[17] Hsiao, C. (2003): Analysis of Panel Data (Econometric Society Monographs), *2nd ed.*, Cambridge University Press.

[18] Love I. and L. Zicchino (2002): Financial Development and Dynamic Investment Behavior: Evidence from Panel Vector Autoregression, *The Quarterly Review of Economics and Finance*, 46, 190-210.

[19] Lütkepohl, H. (1989): A Note on the Asymptotic Distribution of Impulse Response Functions of Estimated VAR Models with Orthogonal Residuals, *Journal of Econometrics*, 42, 371–376.

[20] Lütkepohl, H. (1990): Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models, *Review of Economics and Statistics*, 72, 116–125.

[21] Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*, Springer-Verlag Berlin Heidelberg New York .