Asymptotic Distributions of Impulse Response Functions in Short Panel Vector Autoregressions

Bolong Cao and Yixiao Sun*

Department of Economics,
University of California, San Diego

May 2006

Abstract

This paper establishes the limiting distributions of orthogonalized and nonorthogonalized impulse response functions in panel vector autoregressions with a fixed time dimension. 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. We find that the GMM estimator of the autoregressive coefficients depends on the estimator of error variance, even in large samples with a large cross sectional dimension. The asymptotic dependence leads to additional terms in the asymptotic variance of the orthogonalized impulse response function that are not present in the time series literature. Simulation results show that the asymptotic distribution of the orthogonalized impulse response function that takes the dependence into account is more accurate than the one that does not.

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

JEL Classification Numbers: C33

*Email: bcao@dss.ucsd.edu and yisun@ucsd.edu. Tel: 858-534-4692 (Y. Sun). Correspondnece to: Yixiao Sun, University of California, San Diego, Department of Economics, 9500 Gilman Drive, La Jolla, CA 92093-0508

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 (1982, hereafter AH) estimator or the Arellano and 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) pointed 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 covariance estimator while for short panel VARs the AH or AB estimator of the slope coefficients depends on the error covariance estimator even in the limit as $N \to \infty$ for a fixed T. Due to the presence of the fixed individual effects, the regressors in the short panel data VAR model are correlated with the regression error. This nonzero correlation leads to the asymptotic dependence between the slope coefficient estimator and the error covariance estimator.

Second, for time series VARs, the error covariance estimator based on the estimated OLS residual is asymptotically equivalent to that based on the true error term. For short panel VARs, the error covariance estimator has different asymptotic distributions, depending on whether the error term is known or is based on consistent estimates of the slope coefficients. In other words, the estimation uncertainty of the slope coefficients affects the asymptotic distribution of the error covariance estimator.

These two differences imply that the usual asymptotic results for orthogonalized impulse responses are not applicable to short panel VARs. The main contribution of the paper is to derive the asymptotic distributions of the orthogonalized IRFs for short panel VARs. 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 covariance matrix. As a result, the reported confidence bands are more narrow than they should be. This may lead to the finding of statistical significance that does not actually exist.

The rest of the paper is organized as follows. Section 2 describes the vector autoregressions model for panel data and presents the GMM estimators of the slope coefficients and analysis-of-variance-type estimator of the error covariance matrix. This section also establishes the joint asymptotic distribution of the slope coefficients estimator and the error covariance estimator. Using the asymptotic results in section 2, section 3 derives the asymptotic distributions of the orthogonalized and non-orthogonalized impulse response functions. The next section provides some simulation evidence and the final section concludes. Proofs 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_{mn} is defined such that, for any $(m \times n)$ matrix G, $K_{mn}vec(G) = vec(G')$, 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,

$vech\,(F) = L_m vec(F)$.

2 The Model and Estimation

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

$$w_{i,t} = c + A_1 w_{i,t-1} + ... + A_p w_{i,t-p} + e_{i,t} \quad (1)$$

$$e_{i,t} = \mu_i + \lambda_t + \varepsilon_{i,t} \quad (2)$$

for $t = T_{i0}, T_{i0}+1, ..., T_i$ and $i = 1, 2, ..., N$ where $w_{i,t} = (w_{1,it}, ..., w_{m,it})'$, A_j are $(m \times m)$ coefficient matrices, μ_i, λ_t and $\varepsilon_{i,t}$ have zero means and are independent among themselves and with each other. We allow for unbalanced panel data sets. For individual i, the time series starts at period T_{i0} and ends at period T_i. We assume that

$$E(\varepsilon_{i,t}|w_{i,t-1}, ..., w_{i,T_{i0}}) = 0 \text{ for } T_{i0} \leq t \leq T_i \quad (3)$$

and for $T_{i0} \leq t \leq s$,

$$E\left(\varepsilon_{i,t}\varepsilon_{j,s}'|w_{i,t-1}, ..., w_{i,T_{i0}}\right) = \begin{cases} \Sigma, & i = j \text{ and } t = s \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where Σ is a positive definite matrix.

The model is the same as that considered by Binder, Hsiao, and Pesaran (2005). Here we do not specify the initial conditions for the VAR model as the asymptotic properties of the GMM estimator used in this paper do not depend on how the process is initialized. We focus on the GMM type estimator because it is widely used in empirical applications, see, for example, Love and Zicchino (2002) and Gilchrist, Himmelberg, and Huberman (2005).

Let

$$y_{i,t} = w_{i,t} - w_{\cdot,t},\ c_i = \mu_i - \frac{1}{N}\sum_{j=1}^{N}\mu_j,\ u_{i,t} = \varepsilon_{i,t} - \varepsilon_{\cdot,t} \quad (5)$$

where $w_{\cdot,t}$ $\varepsilon_{\cdot,t}$ are the cross sectional averages of $w_{i,t}$ and $\varepsilon_{i,t}$, respectively. Due to the possible unbalancedness, the cross sectional averages may be taken over different numbers of observations for different periods. Writing the VAR system in terms of $y_{i,t}$, we have

$$y_{i,t} = c + A_1 y_{i,t-1} + ... + A_p y_{i,t-p} + c_i + u_{i,t}, t = T_{i0} + p, ..., T_i. \quad (6)$$

It is now well known that, due to the correlation between the fixed effect c_i and the regressors, the OLS estimator of A_i based on equation (6) is inconsistent when T is small. To remove the fixed individual effect, we take the first difference of equation (6), leading to

$$\Delta y_{i,t} = A_1 \Delta y_{i,t-1} + ... + A_p \Delta y_{i,t-p} + \Delta u_{i,t},\ t = T_{i0} + p + 1, ..., T_i \tag{7}$$

The OLS estimator based on the above 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 $\Delta u_{i,t}$. For the AH estimator, the underlying moment conditions are

$$E\left(\sum_{t=T_{i0}+p+1}^{T_i} \Delta u_{i,t} y'_{i,t-1-\ell} \right) = 0 \text{ for } \ell = 1, 2, ..., \ p; \tag{8}$$

while for the AB estimator, the moment conditions are

$$E\left(\Delta u_{i,t} y'_{i,t-1-\ell}\right) = 0 \text{ for } \ell\ = 1, 2, ..., t-1;\ t = T_{i0} + p + 1, ..., T. \tag{9}$$

To write the equations in the vector form, we let

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

and

$$\underbrace{\Delta y_i}_{(T_i - T_{i0} - p)\times m} = \begin{pmatrix} \Delta y'_{i,T_{i0}+p+1} \\ \Delta y'_{i,T_{i0}+p+2} \\ ... \\ \Delta y'_{i,T_i} \end{pmatrix}, \quad \underbrace{\Delta u_i}_{(T_i - T_{i0} - p)\times m} = \begin{pmatrix} \Delta u'_{i,T_{i0}+p+1} \\ \Delta u'_{i,T_{i0}+p+2} \\ ... \\ \Delta u'_{i,T_i} \end{pmatrix}, \tag{10}$$

$$\underbrace{\Delta X_{i,t}}_{mp\times 1} = \begin{pmatrix} \Delta y_{i,t-1} \\ \Delta y_{i,t-2} \\ ... \\ \Delta y_{i,t-p} \end{pmatrix}, \quad \underbrace{\Delta X_i}_{(T_i - T_{i0} - p)\times mp} = \begin{pmatrix} \Delta X'_{i,T_{i0}+p+1} \\ \Delta X'_{i,T_{i0}+p+2} \\ ... \\ \Delta X'_{i,T_i} \end{pmatrix}. \tag{11}$$

then

$$vec(\Delta y_i) = vec((\Delta X_i)\, A) + vec(\Delta u_i). \tag{12}$$

In other words,

$$\Delta \tilde{y}_i = \Delta \tilde{X}_i \alpha + v_i, \tag{13}$$

where

$$\Delta\tilde{y}_i = vec(\Delta y_i), \quad \underbrace{\Delta\tilde{X}_i}_{m(T_i-T_{i0}-p)\times m^2p} = (I_m \otimes \Delta X_i), \ \alpha = vec(A), v_i = vec(\Delta u_i). \tag{14}$$

To construct the instrument matrix, we normalize $\min(T_{i0})$ to be zero and denote $T = \max(T_i)$. For the AH estimator, we let $\tilde{Z}_i = I_m \otimes Z_i$ where

$$Z_i = \begin{pmatrix} y'_{i,0}, ..., y'_{i,T_{i0}+p-1} \\ y'_{i,1}, ..., y'_{i,T_{i0}+p} \\ ... \\ y'_{i,T_i-p-1}, ..., y'_{i,T_i-2} \end{pmatrix}, \tag{15}$$

is $m(T_i - T_{i0} - p) \times pm$ matrix with all missing values replaced by zeros. For the AB estimator, we first let

$$Z_i^0 = \begin{pmatrix} \left(y'_{i,0}, ..., y'_{i,p-1}\right) & 0 & ... & 0 \\ 0 & \left(y'_{i,0}, ..., y'_{i,p}\right) & 0 & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & \left(y'_{i,0}, ..., y'_{i,T-2}\right) \end{pmatrix} \tag{16}$$

which is an $m(T-p)\times m\,(T+p-1)\,(T-p)\,/2$ matrix. Again the missing values of $y_{i,t}$ in Z_i^0 are replaced by zeros. Next, we keep the rows of Z_i^0 that correspond to periods $T_{i0}+p+1, ..., T_i$ and remove the rest of the rows. We denote the resulting matrix as Z_i, an $m(T_i - T_{i0} - p) \times m\,(T+p-1)\,(T-p)\,/2$ matrix, and define $\tilde{Z}_i = I_m \otimes Z_i$. With these definitions of $\tilde{Z}_i$, the moment conditions in (8) and (9) can be written as $E\tilde{Z}_i' v_i = 0$.

The GMM estimator of α is now given by

$$\hat{\alpha} = \left(S'_{ZX} W_N S_{ZX}\right)^{-1} \left(S_{ZX} W_N S_{ZY}\right) \tag{17}$$

where

$$S_{ZX} = \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i' \Delta\tilde{X}_i, \ S_{ZY} = \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i' \Delta\tilde{y}_i \tag{18}$$

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 e_{it} in (1) is observable, then an analysis-of-variance type estimator of Σ is given by

$$\tilde{\Sigma} = \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{T_i - T_{i0} - p} \sum_{t=T_{i0}+p}^{T_i} \left(e_{i,t} - \bar{e}_{\cdot,t} - \bar{e}_{i,\cdot} + \bar{e}_{\cdot,\cdot}\right)\left(e_{i,t} - \bar{e}_{\cdot,t} - \bar{e}_{i,\cdot} + \bar{e}_{\cdot,\cdot}\right)' \tag{19}$$

where a dot in the subscript indicates the average over that subscript. Under the assumption that all three components in $e_{i,t}$ are normal, it can be shown that $\tilde{\Sigma}$ is the best quadratic unbiased estimator. Since $e_{i,t}$ is not observable, however, we must replace it by some estimator of it. Given the estimate $\hat{\alpha}$, it is natural to estimate $e_{i,t}$ (up to an additive constant) by

$$\hat{e}_{i,t} = w_{i,t} - \left(\hat{A}_1 w_{i,t-1} + ... + \hat{A}_p w_{i,t-p}\right). \tag{20}$$

The resulting estimate of Σ is then given by

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{T_i - T_{i0} - p} \sum_{t=T_{i0}+p}^{T_i} \left(\hat{e}_{i,t} - \bar{\hat{e}}_{\cdot,t} - \bar{\hat{e}}_{i,\cdot} + \bar{\hat{e}}_{\cdot,\cdot}\right)\left(e_{i,t} - \bar{\hat{e}}_{\cdot,t} - \bar{\hat{e}}_{i,\cdot} + \bar{\hat{e}}_{\cdot,\cdot}\right)' \\
&= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{T_i - T_{i0} - p} \sum_{t=T_{i0}+p}^{T_i} \hat{u}_{i,t}\hat{u}_{i,t}', \qquad (21)
\end{aligned}$$

where

$$\hat{u}_{i,t} = \left(y_{i,t} - \bar{y}_{i,\cdot}\right) - \hat{A}'\left(X_{i,t} - \bar{X}_{i,\cdot}\right), \text{ for } t = T_{i0} + p, ..., T_i \tag{22}$$

We now consider the large N asymptotics for a fixed T. First, under the assumption of cross sectional independence and $E\left\|Z_i'\Delta X_i\right\| < \infty$, we have

$$\begin{aligned}
p\lim_{N\to\infty} S_{ZX} &= p\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(I_m \otimes Z_i'\right)\left(I_m \otimes \Delta X_i\right) \\
&= I_m \otimes \left(E Z_i' \Delta X_i\right) = C \text{ for some matrix } C, \qquad (23)
\end{aligned}$$

Second, we consider the asymptotic distribution of $1/\sqrt{N}\sum_{i=1}^{N} \tilde{Z}_i' v_i$. Let

$$D_i = \begin{pmatrix} 1 & -1 & 0 & ... & ... & 0 \\ 0 & 1 & -1 & 0 & ... & ... \\ ... & 0 & 1 & -1 & 0 & ... \\ ... & ... & 0 & ... & ... & ... \\ 0 & ... & ... & 0 & 1 & -1 \end{pmatrix}_{(T_i - p)\times(T_i - T_{i0} - p + 1)}, \tag{24}$$

then

$$v_i = vec\left(\Delta u_i\right) = \left(I_m \otimes D_i\right) vec \begin{pmatrix} u'_{i,T_{i0}+p} \\ u'_{i,T_{i0}+p+1} \\ \cdots \\ u'_{i,T_i} \end{pmatrix}, \tag{25}$$

which implies that

$$Ev_iv_i' = \left(\frac{N-1}{N}\right)\left(I_m \otimes D_i\right)\left(\Sigma \otimes I_{(T_i - T_{i0} - p+1)}\right)\left(I_m \otimes D_i'\right) = \left(\frac{N-1}{N}\right)\Sigma \otimes D_iD_i' \tag{26}$$

and

$$V = E\tilde{Z}_i'v_iv_i'\tilde{Z}_i = \left(\frac{N-1}{N}\right)\Sigma \otimes \left(EZ_i'D_iD_i'Z_i\right). \tag{27}$$

Assuming that $E\left\|\tilde{Z}_i'v_i\right\|^{2+\delta} < \infty$ for some $\delta > 0$ and invoking Lyapunov's central limit theorem gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{Z}_i'v_i \rightarrow_d N(0,V) \text{ with } V = \Sigma \otimes \left(EZ_i'D_iD_i'Z_i\right). \tag{28}$$

Combining (23) and (28), we get

$$\sqrt{N}(\widehat{\alpha} - \alpha) \rightarrow \left(C'WC\right)^{-1} CW\xi := N(0, \Omega_{\alpha\alpha}) \tag{29}$$

where

$$\Omega_{\alpha\alpha} = \left(C'WC\right)^{-1} C'WVWC\left(C'WC\right)^{-1}. \tag{30}$$

To minimize the asymptotic variance of the GMM estimator, we choose the weighting matrix W_N such that its limit $W = V^{-1}$ (see, Hansen 1982). In this case,

$$\begin{aligned} \Omega_{\alpha\alpha} &= \left(C'V^{-1}C\right)^{-1} \\ &= \Sigma \otimes \left(\left(E\Delta X_i'Z_i\right)\left(EZ_i'D_iD_i'Z_i\right)^{-1}\left(EZ_i'\Delta X_i\right)\right)^{-1} := \Sigma \otimes Q^{-1} \end{aligned} \tag{31}$$

The above asymptotic variance can be also achieved by letting

$$W_N = I_m \otimes \left(\frac{1}{N}\sum_{i=1}^{N} Z_i'D_iD_i'Z_i\right)^{-1}, \tag{32}$$

in which case $W = p\lim_{N\rightarrow\infty} W_N = I_m \otimes \left(EZ_i'D_iD_i'Z_i\right)^{-1}$. To see this, note that for this choice of the weighting matrix, we have

$$C'WVWC = \Sigma \otimes \left\{\left(E\Delta X_i'Z_i\right)\left(EZ_i'D_iD_i'Z_i\right)^{-1}\left(EZ_i'\Delta X_i\right)\right\} \tag{33}$$

and

$$\left(C'WC\right)^{-1} = I_m \otimes \left\{ \left(E\Delta X_i' Z_i\right) \left(EZ_i' D_i D_i' Z_i\right)^{-1} \left(EZ_i' \Delta X_i\right) \right\}. \tag{34}$$

Therefore

$$\Omega_{\alpha\alpha} = \Sigma \otimes \left(\left(E\Delta X_i' Z_i\right) \left(EZ_i' D_i D_i' Z_i\right)^{-1} \left(EZ_i' \Delta X_i\right) \right)^{-1}, \tag{35}$$

which is identical to the asymptotic variance given in (31).

With the weighting matrix given in (32), the GMM estimator of $\hat{\alpha}$ reduces to

$$\begin{aligned} \hat{\alpha} \quad = \quad & vec \left\{ \left[\left(\sum_{i=1}^{N} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{N} Z_i' D_i D_i' Z_i \right)^{-1} \left(\sum_{i=1}^{N} \left(Z_i' \Delta X_i \right) \right) \right]^{-1} \right. \\ & \left. \left(\sum_{i=1}^{N} \Delta X_i' Z_i \right) \left(\sum_{i=1}^{N} Z_i' D_i D_i' Z_i \right)^{-1} \sum_{i=1}^{N} (Z_i' \Delta y_i) \right\}. \end{aligned} \tag{36}$$

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) pointed 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 (4), there is no efficiency gain from joint estimation.

In the rest of the paper, we focus on the equation-by-equation GMM estimator given in (36). Let

$$W_{i,t} = \left(w_{i,t-1}',\ w_{i,t-2}', ..., w_{i,t-p}'\right)', \tag{37}$$

$$L_i = \left(T_i - T_{i0} - p\right), \ M_i = \left(T_i - T_{i0} - p\right)\left(T_i - T_{i0} - p + 1\right), \tag{38}$$

$$B = -\text{Plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} E \left(\sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \right) \left(\sum_{t=T_{i0}+p}^{T_i} W_{i,t}' \right), \tag{39}$$

$$\bar{L} = \lim_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{L_i} \right)^{-1}, \tag{40}$$

and

$$\bar{M} = \lim_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\left(T_i - T_{i0} - p\right)\left(T_i - T_{i0} - p + 1\right)} \right)^{-1}. \tag{41}$$

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

Theorem 1 *Assume that $\varepsilon_{it} \sim iidN(0,\Sigma)$. Then*

$$\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}-\alpha\right) \\ \sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\sigma}' \\ \Omega_{\alpha\sigma} & \Omega_{\sigma\sigma} \end{pmatrix}\right) \tag{42}$$

where

$$\Omega_{\alpha\alpha} = \Sigma \otimes Q^{-1} \tag{43}$$

$$\Omega_{\alpha\sigma} = -D_m^+(I_m \otimes B)\left(\Sigma \otimes Q^{-1}\right) - D_m^+ K_{m,m}(I_m \otimes B)\left(\Sigma \otimes Q^{-1}\right) \tag{44}$$

$$\begin{aligned} \Omega_{\sigma\sigma} \quad = \quad & \left(\frac{2}{\bar{L}} + \frac{1}{\bar{M}}\right) D_m^+\left(\Sigma \otimes \Sigma\right)\left(D_m^+\right)' \\ & + D_m^+\left(\Sigma \otimes BQ^{-1}B'\right)\left(D_m^+\right)' + D_m^+\left(BQ^{-1}B' \otimes \Sigma\right)\left(D_m^+\right)' \\ & + D_m^+\left(\Sigma \otimes BQ^{-1}B\right)K_{m,m}'\left(D_m^+\right)' + D_m^+ K_{m,m}\left(\Sigma \otimes BQ^{-1}B'\right)\left(D_m^+\right)'. \end{aligned} \tag{45}$$

Remark 2 *For a time series VAR model with Gaussian innovations, the MLEs of α and Σ have the following limiting distribution (c.f. Proposition 11.2 in Hamilton (1994))*

$$\begin{pmatrix} \sqrt{T}\left(\hat{\alpha}_{MLE}-\alpha\right) \\ \sqrt{T}vech\left(\hat{\Sigma}_{MLE}-\Sigma\right) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Sigma \otimes \left(EX_t'X_t\right)^{-1} & 0 \\ 0 & 2D_m^+\left(\Sigma \otimes \Sigma\right)\left(D_m^+\right)' \end{pmatrix}\right). \tag{46}$$

Comparing this with the limiting distribution in Theorem 1, we see that $\sqrt{N}\left(\hat{\alpha}-\alpha\right)$ is not asymptotically independent of $\sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right)$ while $\sqrt{T}\left(\hat{\alpha}_{MLE}-\alpha\right)$ is asymptotically independent of $\sqrt{T}vech\left(\hat{\Sigma}_{MLE}-\Sigma\right)$. To construct valid confidence bands for the IRFs from short panel VARs, we have to take the asymptotic dependence between $\sqrt{N}\left(\hat{\alpha}-\alpha\right)$ and $\sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right)$ into account.

Remark 3 *It follows from the proof of the theorem in the Appendix that the infeasible estimator $\tilde{\Sigma}$ satisfies*

$$\sqrt{N}vech\left(\tilde{\Sigma}-\Sigma\right) \to_d N\left(0, \left(\frac{2}{\bar{L}} + \frac{1}{\bar{M}}\right) D_m^+\left(\Sigma \otimes \Sigma\right)\left(D_m^+\right)'\right). \tag{47}$$

So the asymptotic variance of $\sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right)$ differs from that of $\sqrt{N}vech\left(\tilde{\Sigma}-\Sigma\right)$ by a few extra terms. These extra terms capture the estimation uncertainty of the slope coefficients.

Remark 4 *When $T \to \infty$, we have $B \to 0$. So*

$$\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}-\alpha\right) \\ \sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Sigma \otimes Q^{-1} & 0 \\ 0 & 2D_m^+\left(\Sigma\otimes\Sigma\right)\left(D_m^+\right)' \end{pmatrix}\right). \tag{48}$$

Therefore, the asymptotic dependence between $\sqrt{N}\left(\hat{\alpha}-\alpha\right)$ and $\sqrt{N}vech\left(\hat{\Sigma}-\Sigma\right)$ diminishes and the effect of estimation uncertainty dies out, as $T \to \infty$.

Remark 5 *Note that the least squared dummy variable (LSDV) estimator is biased and inconsistent because $B \neq 0$. The asymptotic bias of the LSDV estimator depends on the magnitude of B. If we make a bias correction to the LSDV estimator, then the bias corrected LSDV estimator is no longer asymptotically independent of $\hat{\Sigma}$. In this case, we should employ the limiting distribution similar to that given in Theorem 1 instead of (48).*

Remark 6 *Let $P, \hat{P}$ be lower triangular matrices such that $PP' = \Sigma$ and $\hat{P}\hat{P}' = \hat{\Sigma}$. Then it follows from Theorem 1 and*

$$\frac{\partial vech(P)}{\partial vech\left(\Sigma\right)'} = \left\{L_m\left(I_{m^2}+K_{mm}\right)\left(P\otimes I_m\right)L_m'\right\}^{-1} =: H_0$$

that

$$\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}-\alpha\right) \\ \sqrt{N}vech\left(\hat{P}-P\right) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\sigma}'H_0' \\ H_0\Omega_{\alpha\sigma} & H_0\Omega_{\sigma\sigma}H_0' \end{pmatrix}\right). \tag{49}$$

This joint distribution can be employed to derive the asymptotic distribution of a statistic that is a function of $\hat{\alpha}$ and $\hat{P}$. Either the delta method or a simulation-based method can be used.

3 Asymptotic Distributions of the IRFs

Since the impulse response function does not depend on the index i and deterministic variables in the system, without the loss of generality we consider the model

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t. \tag{50}$$

Given the estimators $\hat{\alpha}$ and $\hat{\Sigma}$, our objective is to compute the IRFs and the associated confidence bands.

Assuming that the process is stationary, we can write the model in the $MA(\infty)$ form:

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} \tag{51}$$

where $\Phi_0 = I_m$ and

$$\Phi_j = \sum_{\ell=1}^{p} A_\ell \Phi_{j-\ell}, \; j = 1, 2, ... \tag{52}$$

with $\Phi_s = 0$ for $s < 0$. The matrix Φ_j has the interpretation: $\Phi_j = \partial y_{t+j}/\partial u_t$. The plot of (k,ℓ)-*th* element of Φ_j as a function of j is called the non-orthogonalized impulse-response function. It 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.

To estimate the non-orthogonalized impulse response, we plug the estimate $\hat{A}$ into (52) and get $\hat{\Phi}_j = \sum_{\ell=1}^{p} \hat{A}_\ell \hat{\Phi}_{j-\ell}$. The limiting distribution of $\hat{\Phi}_j$ can be derived using the delta method. More specifically, taking transposes of (52) and differentiating the resulting equation with respect to α_q, the q-th element of α yields:

$$\begin{aligned} \frac{\partial \Phi_j'}{\partial \alpha_q} &= \sum_{\ell=1}^{p} \frac{\partial \Phi_{j-\ell}'}{\partial \alpha_q} A_\ell' + \sum_{\ell=1}^{p} \Phi_{j-\ell}' \frac{\partial A_\ell'}{\partial \alpha_q} \\ &= \sum_{\ell=1}^{p} \frac{\partial \Phi_{j-\ell}'}{\partial \alpha_q} A_\ell' + [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}'] \frac{\partial A}{\partial \alpha_q}. \end{aligned} \tag{53}$$

Consequently,

$$vec\left(\frac{\partial \Phi_j'}{\partial \alpha_q}\right) = \sum_{\ell=1}^{p} (A_\ell \otimes I_m) vec\left(\frac{\partial \Phi_{j-\ell}'}{\partial \alpha_q}\right) + \left(I_m \otimes [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}']\right) \frac{\partial \alpha}{\partial \alpha_q}. \tag{54}$$

Stacking the above equations gives

$$\frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'} = \sum_{\ell=1}^{p} (A_\ell \otimes I_m) \left(\frac{\partial vec\left(\Phi_{j-\ell}'\right)}{\partial \alpha'}\right) + \left(I_m \otimes [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}']\right). \tag{55}$$

Let

$$\underset{(m^2 \times pm^2)}{G_0} = 0 \text{ and } \underset{(m^2 \times pm^2)}{G_j} = \frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'}, \; j = 1, 2, ... \tag{56}$$

then

$$G_j = \sum_{\ell=1}^{p} (A_\ell \otimes I_m) G_{j-\ell} + (I_m \otimes [\Phi_{j-1}', \Phi_{j-2}', ..., \Phi_{j-p}']) \tag{57}$$

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

$$G_j = \sum_{\ell=0}^{j-1} \Phi_\ell \otimes [\Phi_{j-\ell-1}', \Phi_{j-\ell-2}', ..., \Phi_{j-\ell-p}'] = \sum_{\ell=0}^{j-1} \Phi_\ell \otimes J \left(F'\right)^{j-\ell-1}, \tag{58}$$

where $J = [I_m, 0, ...0]$ is an $m \times mp$ matrix and

$$F = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_m & 0 & \cdots & 0 & 0 \\ 0 & I_m & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_m & 0 \end{pmatrix}. \tag{59}$$

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

$$\sqrt{N} vec\left(\hat{\Phi}_j' - \Phi_j'\right) \xrightarrow{d} N(0, G_j \Omega_{\alpha\alpha} G_j'). \tag{60}$$

where $\Omega_{\alpha\alpha}$ can be consistently estimated by

$$\hat{\Omega}_{\alpha\alpha} = \hat{\Sigma} \otimes \left(\left(\frac{1}{N} \sum_{i=1}^{N} \Delta X_i' Z_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' D_i D_i' Z_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \Delta X_i \right) \right)^{-1}. \tag{61}$$

In empirical applications, it is a standard practice to report the orthogonalized impulse response function. Let $PP' = \Sigma$ where P is a lower triangular matrix with positive diagonal elements, then

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j} = \sum_{j=0}^{\infty} \Phi_j P \left(P^{-1} u_{t-j}\right) = \sum_{j=0}^{\infty} \Theta_j \epsilon_{t-j} \tag{62}$$

where $\Theta_j = \Phi_j P$ and $\epsilon_{t-j} = P^{-1} u_{t-j}$ with $var(\epsilon_{t-j}) = I_m$. The orthogonalized impulse response function is the plot of (k, ℓ)-th element of Θ_j as a function of j.

The orthogonalized IRF can be estimated by plugging the estimates $\hat{\Phi}_j$ and $\hat{\Sigma}$ into its definition. To derive the limiting distribution of $\hat{\Theta}_j$, we use the delta method again. Let

$$H = L_m' \left\{ L_m \left(I_{m^2} + K_{mm}\right) \left(P \otimes I_m\right) L_m' \right\}^{-1}, \tag{63}$$

and

$$C_j = \left(I_m \otimes P'\right) G_j, \ \bar{C}_j = \left(\Phi_j' \otimes I_m\right) K_{mm} H, \tag{64}$$

then

$$\frac{\partial vec\left(\Theta_j'\right)}{\partial \alpha'} = \frac{\partial vec\left(P' \Phi_j'\right)}{\partial \alpha'} = \left(I_m \otimes P'\right) \frac{\partial vec\left(\Phi_j'\right)}{\partial \alpha'} = C_j \tag{65}$$

and

$$\begin{aligned}
\frac{\partial vec\left(\Theta_j'\right)}{\partial\left[vech(\Sigma)\right]'} &= \frac{\partial vec\left(P'\Phi_j'\right)}{\partial\left[vech(\Sigma)\right]'} = \left(\Phi_j \otimes I_m\right)\frac{\partial vec\left(P'\right)}{\partial\left[vech(\Sigma)\right]'} \\
&= \left(\Phi_j \otimes I_m\right)\frac{K_{mm}\partial vec\left(P\right)}{\partial\left[vech(\Sigma)\right]'} = \left(\Phi_j \otimes I_m\right)\frac{K_{mm}L_m'\partial vech\left(P\right)}{\partial\left[vech(\Sigma)\right]'} \\
&= \left(\Phi_j \otimes I_m\right)K_{mm}H = \bar{C}_j \qquad (66)
\end{aligned}$$

where the second last equality follows from Lemma 1 in Lütkepohl (1989). Therefore

$$\sqrt{N}vec\left(\hat{\Theta}_j' - \Theta_j'\right) \xrightarrow{d} N\left(0, \Sigma_j^{\theta}\right), \qquad (67)$$

where

$$\begin{aligned}
\Sigma_j^{\theta} &= \left[C_j, \bar{C}_j\right]\begin{pmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\sigma}' \\ \Omega_{\alpha\sigma} & \Omega_{\sigma\sigma} \end{pmatrix}\begin{pmatrix} C_j' \\ \bar{C}_j' \end{pmatrix} \qquad (68) \\
&= C_j\Omega_{\alpha\alpha}C_j' + \bar{C}_j\Omega_{\sigma\sigma}\bar{C}_j' + \bar{C}_j\Omega_{\alpha\sigma}C_j' + C_j\Omega_{\alpha\sigma}'\bar{C}_j'.
\end{aligned}$$

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 $\bar{C}_j\Omega_{\alpha\sigma}C_j'$, $C_j\Omega_{\alpha\sigma}'\bar{C}_j'$ are not present in the asymptotic variance of the orthogonalized IRF. In contrast, $\Omega_{\alpha\sigma} \neq 0$ for short panel VARs. In this case, it is important to include these cross product terms in computing the asymptotic variance, especially when T is small.

To consistently estimate the asymptotic variance Σ_j^{θ}, we plug consistent estimators of $C_j, \bar{C}_j, \Omega_{\alpha\alpha}, \Omega_{\alpha\sigma}$ and $\Omega_{\sigma\sigma}$ into (68), leading to

$$\hat{\Sigma}_j^{\theta} = \hat{C}_j\hat{\Omega}_{\alpha\alpha}\hat{C}_j' + \tilde{C}_j\hat{\Omega}_{\sigma\sigma}\tilde{C}_j' + \tilde{C}_j'\hat{\Omega}_{\alpha\sigma}\hat{C}_j' + \hat{C}_j\hat{\Omega}_{\alpha\sigma}'\tilde{C}_j', \qquad (69)$$

where

$$\hat{C}_j = \left(I_m \otimes \hat{\Sigma}^{-1/2\prime}\right)\hat{G}_j, \quad \tilde{C}_j = \left(\hat{\Phi}_j' \otimes I_m\right)K_{mm}H, \qquad (70)$$

and $\hat{\Omega}$ are defined in (43)–(45) with Q, B, Σ replaced by

$$\hat{Q} = \left(\frac{1}{N}\sum_{i=1}^{N}\Delta X_i'Z_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}Z_i'D_iD_i'Z_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i'\Delta X_i\right), \qquad (71)$$

$$\hat{B} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{L_i}\sum_{t=T_{i0}+p}^{T_i}\left(\hat{u}_{i,t} - \hat{u}_{i,\cdot}\right)\left(X_{i,t} - X_{i,\cdot}\right)', \qquad (72)$$

$$\hat{\Sigma} = \frac{1}{(N-1)}\sum_{i=1}^{N}\frac{1}{T_i - T_{i0} - p}\sum_{t=T_{i0}+p}^{T_i}\hat{u}_{i,t}\hat{u}_{i,t}', \qquad (73)$$

respectively.

4 Simulation Evidence

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

We consider the panel VAR model defined in equations (1) – (4). To evaluate the empirical relevance of the asymptotic distribution, we set the model parameters equal to the estimates from the empirical study by Cao (2006). Using a panel VAR(4) model, Cao (2006) studies how financial pressure affects agency costs for corporations over time. There are four variables in the panel VAR system: a measurement of financial pressure, a proxy for future profit opportunities, and two proxies for agency cost. In the benchmark model, these four variables are: asset turnover, market to book ratio, adjusted interest expense to net fixed assets ratio and selling, general and administrative expenses to net fixed assets ratio. They will be labelled as var1, var2, var3, var4 in the rest of this section. For more details on the construction of these variables, see Cao (2006). For our purposes here, it suffices to describe the estimated model parameters. They are given in Table 1.

Table 1: AR Parameters for the Panel VAR(4) Model

A_1				A_2			
0.4959	0.0694	0.0133	-0.0004	0.0741	-0.0228	0.0420	0.0077
-0.1225	0.4235	-0.0833	0.0046	-0.1371	0.0427	0.0236	0.0368
-0.0780	-0.1079	0.4187	0.0318	-0.0304	0.0175	-0.2036	-0.0519
-0.1247	0.0381	-0.4184	0.2290	0.0447	-0.0254	0.1875	0.0156
A_3				A_4			
0.0795	-0.0057	-0.0028	-0.0036	0.0204	-0.0032	0.0207	0.0103
-0.0076	-0.0589	-0.0248	-0.0112	0.0117	0.0397	0.0238	0.0275
-0.0634	-0.0052	0.0526	0.0080	-0.0834	-0.0048	-0.1117	-0.0403
0.0054	-0.0035	0.0626	0.0084	0.0494	0.0059	0.1959	0.0871

The variance covariance matrix of the error term is estimated to be PP' where

$$P = \begin{pmatrix} 0.2755 & 0 & 0 & 0 \\ 0.1379 & 1.3589 & 0 & 0 \\ -0.2613 & -0.0345 & 0.5999 & 0 \\ 0.1799 & 0.0053 & -0.1808 & 0.9724 \end{pmatrix}.$$

Given A_1, A_2, A_3, A_4 and P, we generate the data according to

$$w_{i,t} = c + A_1 w_{i,t-1} + A_2 w_{i,t-2} + A_3 w_{i,t-4} + A_4 w_{i,t-4} + e_{i,t} \tag{74}$$

$$e_{i,t} = \mu_i + \lambda_t + P\varepsilon_{i,t} \tag{75}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ where $c = (0, 0, 0, 0)'$, $\lambda_t \sim iid\ N(0, 1)$, $\varepsilon_{i,t} \sim iid\ N(0, I_4)$ and μ_i is randomly drawn from the estimated fixed effects. For each given sample size T, we set the initial values of the process $\{w_{it}\}$ to be zero and generate a 4-dimensional time series of length $T + 20$. We drop the first 20 observations to obtain the simulated sample.

Figure 1 presents the impulse response function for the first variable (denoted as var1) in response to one standard deviation (SD) shock to each variable in the VAR system. The figure is reported here to illustrate the order of magnitude of the impulse responses. Since P is a lower triangular matrix, the impulse responses in Figure 1(b)-(d) start from zero. The impulse responses reported here represent a rich class of dynamics embodied in the VAR system.

We consider different N and T combinations, i.e. $N = 200, 300, 400, 500$ and $T = 10, 20$. For each (N, T) combination, we estimate the model using the AB method outlined in Section 2. To avoid the weak instrument program, we do not use the lagged dependent variable 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 4. Our simulation results remain more or less the same when we set the maximum number of lags to be 1, 2 and 3. Given the estimated parameters, we construct the orthogonalized impulse response functions and the corresponding 95% confidence bands based on the asymptotic distribution in (67). As a comparison, we also construct the 95% confidence bands when $\Omega_{\alpha\sigma}$ and $\Omega_{\sigma\sigma}$ are set to be

$$\begin{aligned} \Omega_{\alpha\sigma} &= 0, \\ \Omega_{\sigma\sigma} &= \left(\frac{2}{\bar{L}} + \frac{1}{\bar{M}}\right) D_m^+ (\Sigma \otimes \Sigma) \left(D_m^+\right)'. \end{aligned}$$

In this case, the asymptotic dependence between $\sqrt{N}(\hat{\alpha} - \alpha)$ and $\sqrt{N}(vech\hat{\Sigma} - \Sigma)$ and the additional randomness of $\sqrt{N} vech(\hat{\Sigma} - \Sigma)$ are ignored. We call the resulting confidence band the naive confidence band and the one based on (67) the new confidence band. For each confidence band, we compute its empirical coverage and average length based 10000 simulations.

Figure 2 graphs the empirical coverages of each confidence band against the forecasting horizons when $N = 100$ and $T = 10$. To save space, we only report the change in the first variable in response to one standard deviation shock to each variable in the VAR system. The qualitative results are similar for other cases. It is clear from the figure that the empirical coverage of the new confidence band is closer to the nominal coverage probability than the naive confidence band. For some scenarios, the new confidence band dominates the naive confidence band by a large margin. The largest gain in the empirical coverage occurs when we consider the impulse response of a variable to its own shock. To sum up, the figure provides strong evidence that the new asymptotic approximation is more accurate than the naive asymptotic approximation.

Figure 3 reports the average length of the 95% confidence interval at each forecasting horizon. The figure reveals that the new confidence band is slightly wider than the naive confidence band. This result, combined with the empirical coverage in Figure 2, shows that the naive asymptotic variance under-estimates the sampling variability of the impulse response. A direct implication is that inferences based on the naive asymptotic variance may lead to the finding a statistically significant relationship that does not actually exist.

We now briefly comment on the figures not reported here. The qualitative results remain valid for other values of N with $T = 10$. When $T = 20$, the advantage of the new confidence band decays. This is consistent with the asymptotic theory. When T goes to ∞, the difference between the new confidence band and naive confidence band disappears.

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 adjustment for the asymptotic variance of the orthogonalized impulse response function. The failure of making such an adjustment may lead to the spurious finding of a statistically significant result.

In practice, Monte Carlo methods are sometimes used to evaluate the sampling variability of the orthogonalized impulse response function. The standard practice in the time series literature is to randomly draw autoregressive coefficients from the asymptotic distribution of the estimator of the autoregressive coefficients, conditioning on the estimated error variance. The use of conditioning is innocuous in large samples when the estimator of the error variance is asymptotically independent of those of the autoregressive coefficients. However, the asymptotic independence does not hold for short Panel VARs. In this case, the standard Monte Carlo methods are expected to give rise to confidence bands that are more narrow than they should be.

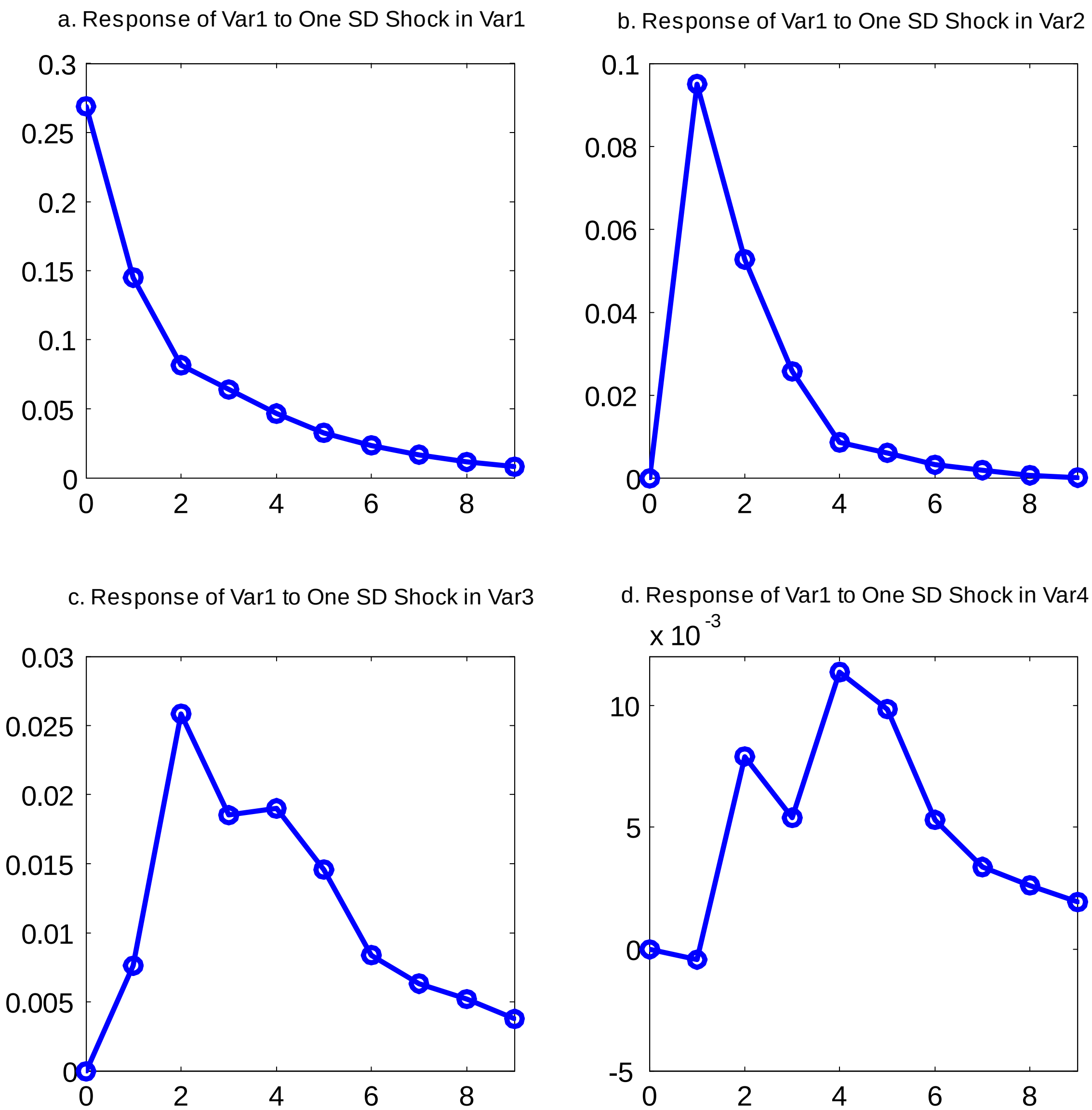

Figure 1: Change in the First Variable in Response to One SD Shocks to the VAR System

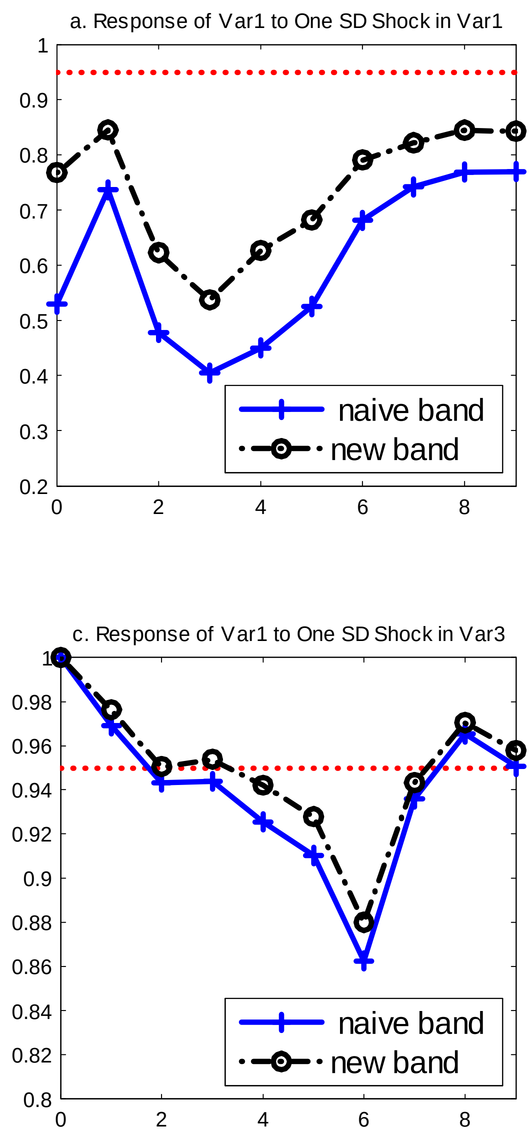

Figure 2: The Empirical Coverages of Different Confidence Bands for N=100 and T=10

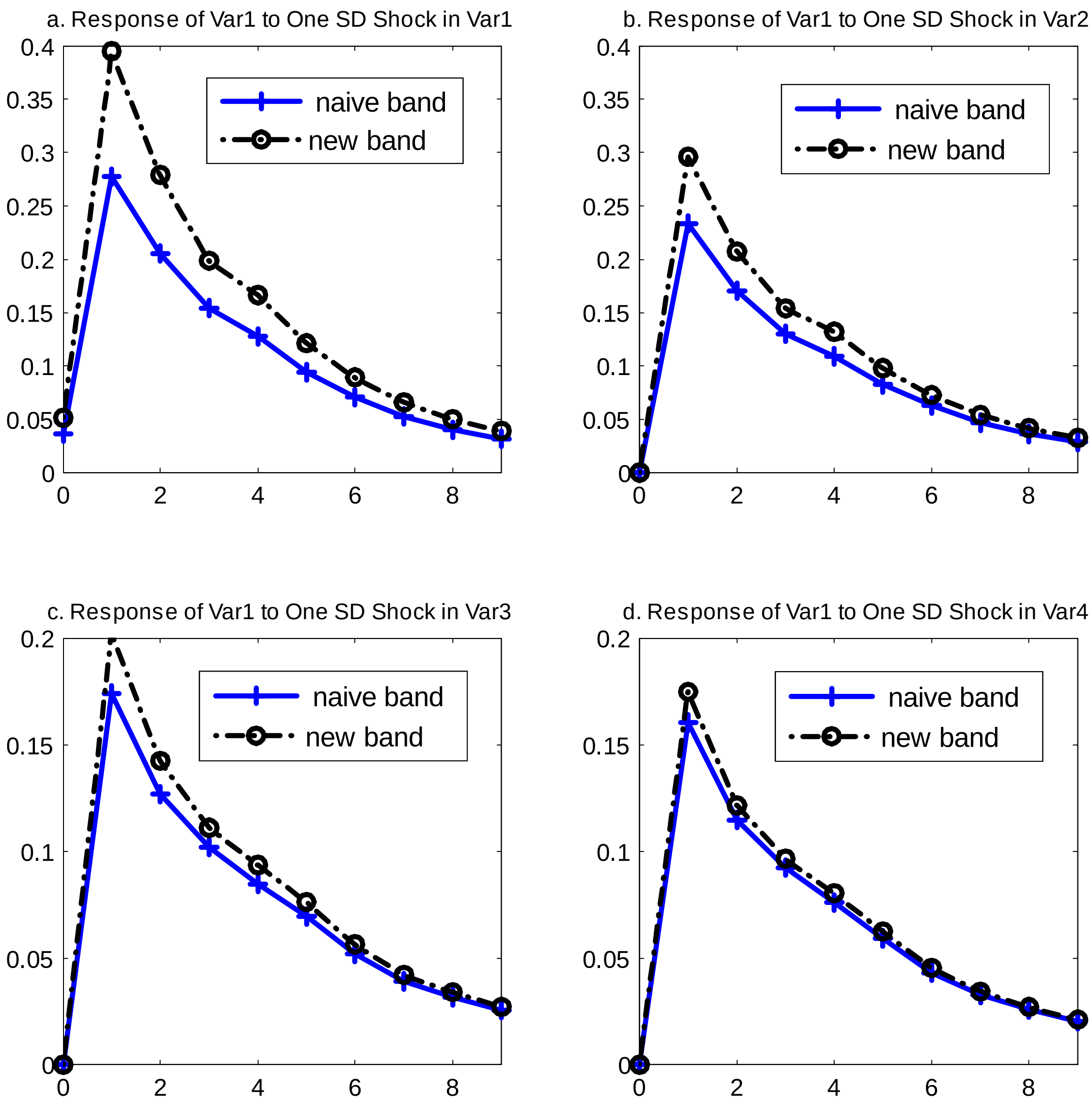

Figure 3: The Average Widths of Different Confidence Bands for N=100 and T=10

6 Appendix

Proof of Theorem 1. We have established the asymptotic distribution of $\hat{\alpha}$ in the main part of the paper. It remains to establish the asymptotic distribution of $\hat{\Sigma}$ and its relationship with $\hat{\alpha}$. Let $L_i = T_i - T_{i0} - p$, we write $\hat{\Sigma}$ in terms of the unobserved error term:

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left[(y_{i,t} - \bar{y}_{i,\cdot}) - A' \left(X_{i,t} - \bar{X}_{i,\cdot} \right) \right. \\
&\quad \left. - \left(\hat{A} - A \right)' \left(X_{i,t} - \bar{X}_{i,\cdot} \right) \right] \left[(y_{i,t} - \bar{y}_{i,\cdot}) - A' \left(X_{i,t} - \bar{X}_{i,\cdot} \right) - \left(\hat{A} - A \right)' \left(X_{i,t} - \bar{X}_{i,\cdot} \right) \right]', \\
&= \hat{\Sigma}^* + I_1 + I_2 + I_3 \qquad (76)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\Sigma}^* &= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} (u_{i,t} - \bar{u}_{i,\cdot}) (u_{i,t} - \bar{u}_{i,\cdot})', \\
I_1 &= \left(\hat{A} - A \right)' \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left(X_{i,t} - \bar{X}_{i,\cdot} \right) \left(X_{i,t} - \bar{X}_{i,\cdot} \right)' \left(\hat{A} - A \right), \\
I_2 &= -\frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} (u_{i,t} - \bar{u}_{i,\cdot}) \left(X_{i,t} - \bar{X}_{i,\cdot} \right)' \left(\hat{A} - A \right), \\
I_3 &= -\left(\hat{A} - A \right)' \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left(X_{i,t} - \bar{X}_{i,\cdot} \right) (u_{i,t} - \bar{u}_{i,\cdot})'.
\end{aligned}$$

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

$$\sqrt{N}\left(\hat{\Sigma} - \Sigma\right) = \sqrt{N}\left(\hat{\Sigma}^* - \Sigma\right) + \sqrt{N} I_2 + \sqrt{N} I_3 + o_p(1). \qquad (77)$$

To evaluate I_2 and I_3, we note that

$$\begin{aligned}
&\frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} (u_{i,t} - \bar{u}_{i,\cdot}) \left(X_{i,t} - \bar{X}_{i,\cdot} \right)' \\
&= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} (\varepsilon_{i,t} - \bar{\varepsilon}_{i,\cdot} - \bar{\varepsilon}_{\cdot,t} + \bar{\varepsilon}_{\cdot,\cdot}) \left(W_{i,t} - \bar{W}_{i,\cdot} - \bar{W}_{\cdot,t} + \bar{W}_{\cdot,\cdot} \right)' \\
&= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} (\varepsilon_{i,t} - \bar{\varepsilon}_{\cdot,t}) \left(W_{i,t} - \bar{W}_{i,\cdot} - \bar{W}_{\cdot,t} + \bar{W}_{\cdot,\cdot} \right)' \\
&= \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \left(W_{i,t} - \bar{W}_{i,\cdot} - \bar{W}_{\cdot,t} + \bar{W}_{\cdot,\cdot} \right)' + o_p(1). \qquad (78)
\end{aligned}$$

The first quantity in the above expression can be written as

$$
\begin{aligned}
& \frac{1}{(N-1)} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} W_{i,t}' - \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\frac{L_i+1}{L_i} \right) \bar{\varepsilon}_{i,\cdot} \bar{W}_{i,\cdot}' \\
& - \frac{1}{(N-1)} \sum_{i=1}^{N} \sum_{t=T_{i0}+p}^{T_i} \frac{1}{L_i} \varepsilon_{i,t} \bar{W}_{\cdot,t}' + \frac{1}{(N-1)} \sum_{i=1}^{N} \left(\frac{L_i+1}{L_i} \right) \bar{\varepsilon}_{i,\cdot} \bar{W}_{\cdot,\cdot}' \\
= & - \frac{1}{(N-1)} \sum_{i=1}^{N} E \left(\left(\frac{L_i+1}{L_i} \right) \bar{\varepsilon}_{i,\cdot} \bar{W}_{i,\cdot}' \right) + o_p(1) \\
= & B_{m \times mp} + o_p(1) \qquad (79)
\end{aligned}
$$

where

$$
B = -\mathrm{P} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{M_i} E \left(\sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \right) \left(\sum_{t=T_{i0}+p}^{T_i} W_{i,t}' \right) \qquad (80)
$$

a constant matrix and $M_i = L_i (L_i + 1)$. Therefore,

$$
\sqrt{N} I_2 = -B \sqrt{N} \left(\hat{A} - A \right) + o_p(1), \;\; \sqrt{N} I_3 = -\sqrt{N} \left(\hat{A} - A \right)' B' + o_p(1). \qquad (81)
$$

Combining (77) with (81) yields

$$
\sqrt{N} \left(\hat{\Sigma} - \Sigma \right) = \sqrt{N} \left(\hat{\Sigma}^* - \Sigma \right) - B \sqrt{N} \left(\hat{A} - A \right) - \sqrt{N} \left(\hat{A} - A \right)' B' + o_p(1). \qquad (82)
$$

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

$$
\begin{aligned}
& \sqrt{N} \left(\hat{\Sigma}^* - \Sigma \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left(u_{i,t} - \bar{u}_{i,\cdot} \right) \left(u_{i,t} - \bar{u}_{i,\cdot} \right)' - \Sigma \right] + o_p(1) \\
= & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left(\varepsilon_{i,t} - \bar{\varepsilon}_{i,\cdot} - \bar{\varepsilon}_{\cdot,t} - \bar{\varepsilon}_{\cdot,\cdot} \right) \left(\varepsilon_{i,t} - \bar{\varepsilon}_{i,\cdot} - \bar{\varepsilon}_{\cdot,t} - \bar{\varepsilon}_{\cdot,\cdot} \right)' - \Sigma \right] + o_p(1) \\
= & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \varepsilon_{i,t}' - \frac{1}{M_i} \left(\sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \right) \left(\sum_{t=T_{i0}+p}^{T_i} \varepsilon_{i,t} \right)' - \Sigma \right] + o_p(1) \\
= & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\frac{1}{L_i+1} \sum_{t=T_{i0}+p}^{T_i} \left(\varepsilon_{i,t} \varepsilon_{i,t}' - \Sigma \right) - \frac{1}{M_i} \sum_{s,\; t=T_{i0}+p, s \neq t}^{T_i} \varepsilon_{i,t} \varepsilon_{i,s}' \right] + o_p(1). \qquad (83)
\end{aligned}
$$

It follows from a proof similar to that of Proposition 11.2 in Hamilton (1994) that

$$
vech \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{L_i} \sum_{t=T_{i0}+p}^{T_i} \left(\varepsilon_{i,t} \varepsilon_{i,t}' - \Sigma \right) \right) \to N \left(0, \frac{2}{\bar{L}} D_m^+ \left(\Sigma \otimes \Sigma \right) D_m^+ \right). \qquad (84)
$$

where

$$\bar{L}=\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{L_i}\right)^{-1}=\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i-T_{i0}-p}\right)^{-1} \tag{85}$$

is the limit of the harmonic mean of the length of each time series. A standard application of a central limit theorem gives

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{M_i}\sum_{t=T_{i0}+p}^{T_i}\sum_{s=T_{i0}+p,s\neq t}^{T_i}\varepsilon_{i,t}\varepsilon_{i,s}'\\ \to\ & {}_dN\left(0,\frac{1}{\bar{M}}D_m^{+}\left(\Sigma\otimes\Sigma\right)D_m^{+}\right),\end{aligned} \tag{86}$$

where

$$\bar{M}=\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i}\right)^{-1}=\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\left(T_i-T_{i0}-p\right)\left(T_i-T_{i0}-p+1\right)}\right)^{-1}. \tag{87}$$

The asymptotic variance is of the above form because

$$\begin{aligned}&\sum_{t=T_{i0}+p}^{T_i}\sum_{s=T_{i0}+p,s\neq t}^{T_i}\sum_{l=T_{i0}+p}^{T_i}\sum_{m=T_{i0}+p,l\neq m}^{T_i}Evec\left(\varepsilon_{i,t}\varepsilon_{i,s}'\right)vec\left(\varepsilon_{i,l}\varepsilon_{i,m}'\right)'\\
=\ &\sum_{t=T_{i0}+p}^{T_i}\sum_{s=T_{i0}+p,s\neq t}^{T_i}\sum_{l=T_{i0}+p}^{T_i}\sum_{m=T_{i0}+p,l\neq m}^{T_i}E\left(\varepsilon_{i,s}\otimes\varepsilon_{i,t}\right)\left(\varepsilon_{i,m}'\otimes\varepsilon_{i,l}'\right)\\
=\ &\sum_{t=T_{i0}+p}^{T_i}\sum_{s=T_{i0}+p,s\neq t}^{T_i}\sum_{l=T_{i0}+p}^{T_i}\sum_{m=T_{i0}+p,l\neq m}^{T_i}E\left(\varepsilon_{i,s}\varepsilon_{i,m}'\otimes\varepsilon_{i,t}\varepsilon_{i,l}'\right)\\
=\ &\sum_{t=T_{i0}+p}^{T_i}\sum_{s=T_{i0}+p,s\neq t}^{T_i}\Sigma\otimes\Sigma\\
=\ &\left(\left(T_i-T_{i0}-p+1\right)^2-\left(T_i-T_{i0}-p+1\right)\right)\Sigma\otimes\Sigma\\
=\ &M_i\left(\Sigma\otimes\Sigma\right).\end{aligned} \tag{88}$$

As a result

$$\sqrt{N}vech\left(\hat{\Sigma}^{*}-\Sigma\right)\to_d N\left(0,\left(\frac{2}{\bar{L}}+\frac{1}{\bar{M}}\right)D_m^{+}\left(\Sigma\otimes\Sigma\right)\left(D_m^{+}\right)'\right) \tag{89}$$

where we have used the asymptotic independence between the two terms in (83).

Next,

$$\begin{aligned} vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right] &= D_m^+ vec\left(B\sqrt{N}\left(\hat{A}-A\right)\right) \\ &= D_m^+(I_m \otimes B) vec\left(\sqrt{N}\left(\hat{A}-A\right)\right) \\ &\to {}_d N(0, D_m^+(I_m \otimes B)\left(\Sigma \otimes Q^{-1}\right)\left(I_m \otimes B'\right)\left(D_m^+\right)') \\ &= {}_d N(0, D_m^+(\Sigma \otimes B) Q^{-1} B' \left(D_m^+\right)'), \end{aligned} \tag{90}$$

$$\begin{aligned} vech\left[\sqrt{N}\left(\hat{A}-A\right)' B'\right] &= D_m^+ vec\left(\sqrt{N}\left(\hat{A}-A\right)' B'\right) \\ &= D_m^+ K_{m,m} vec\left(B\sqrt{N}\left(\hat{A}-A\right)\right) \\ &\to {}_d N(\left[0, D_m^+ K_{m,m}\left(\Sigma \otimes BQ^{-1}B'\right) K'_{m,m}\left(D_m^+\right)'\right], \\ &: = N(\left[0, D_m^+\left(BQ^{-1}B' \otimes \Sigma\right)\left(D_m^+\right)'\right] \end{aligned} \tag{91}$$

where we have used 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 $K_{m,m}K'_{m,m} = I_{m^2}$. In addition,

$$\begin{aligned} & cov\left(vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right], vech\left[\sqrt{N}\left(\hat{A}-A\right)' B'\right]\right) \\ = \; & D_m^+\left(\Sigma \otimes BQ^{-1}B'\right) K'_{m,m}\left(D_m^+\right)'. \end{aligned} \tag{92}$$

Therefore,

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

where

$$\begin{aligned} V_{AB} &= D_m^+(\Sigma \otimes BQ^{-1}B')\left(D_m^+\right)' + D_m^+\left(BQ^{-1}B' \otimes \Sigma\right)\left(D_m^+\right)' \\ &+ D_m^+(\Sigma \otimes BQ^{-1}B) K'_{m,m}\left(D_m^+\right)' + D_m^+ K_{m,m}\left(\Sigma \otimes BQ^{-1}B'\right)\left(D_m^+\right)'. \end{aligned} \tag{93}$$

It is easy to show that $\sqrt{N}\left(\hat{\Sigma}^* - \Sigma\right)$ and $\sqrt{N}(\hat{A}-A)$ are asymptotically independent. As a result,

$$\sqrt{N} vech\left(\hat{\Sigma} - \Sigma\right) \to N(0, \Omega_{\sigma\sigma}), \tag{94}$$

where

$$\begin{aligned} \Omega_{\sigma\sigma} &= \left(\frac{2}{\bar{T}} + \frac{1}{\bar{M}}\right) D_m^+\left(\Sigma \otimes \Sigma\right)\left(D_m^+\right)' \\ &+ D_m^+(\Sigma \otimes BQ^{-1}B'\left(D_m^+\right)' + D_m^+\left(BQ^{-1}B' \otimes \Sigma\right)\left(D_m^+\right)' \\ &+ D_m^+(\Sigma \otimes BQ^{-1}BK'_{m,m}\left(D_m^+\right)' + D_m^+ K_{m,m}((\Sigma \otimes BQ^{-1}B')\left(D_m^+\right)'. \end{aligned} \tag{95}$$

Finally, we examine the asymptotic covariance between $vech\left\{\sqrt{N}\left(\hat{\Sigma}-\Sigma\right)\right\}$ and $vec\left(\sqrt{N}\left(\hat{A}-A\right)\right)$. Since $\sqrt{N}\left(\hat{\Sigma}^{*}-\Sigma\right)$ is asymptotically independent of $\sqrt{N}(\hat{A}-A)$, the asymptotic covariance is given by

$$\begin{aligned}
\Omega_{\alpha,\sigma} &= -cov\left(vech\left[B\sqrt{N}\left(\hat{A}-A\right)\right], vec\left[\sqrt{N}\left(\hat{A}-A\right)\right]\right) \\
&\quad -cov\left(vech\left[\sqrt{N}\left(\hat{A}-A\right)' B'\right], vec\left[\sqrt{N}\left(\hat{A}-A\right)\right]\right) \\
&= -D_m^{+}(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right) - D_m^{+}K_{m,m}(I_m\otimes B)\left(\Sigma\otimes Q^{-1}\right) \qquad (96)
\end{aligned}$$

Combining (94), (96) and $\sqrt{N}\left(\hat{\alpha}-\alpha\right)\to_d N(0,\Omega_{\alpha\alpha})$ completes the proof of the theorem. ■

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., 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.

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

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

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

[7] Cao, B. (2006): Debt Financing and the Dynamics of Agency Costs. Department of Economics, UCSD.

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

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

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

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

[12] Love I. and L. Zicchino (2002): Financial Development and Dynamic Investment Behavior: Evidence from Panel Vector Autoregression, *Policy Research Working Paper Series* No 2913, The World Bank

[13] 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.

[14] 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.