Instrumental Variable Estimation of Structural VAR Models Robust to Possible Non-Stationarity

Xu Cheng* Xu Han † Atsushi Inoue‡

April 18, 2019

Abstract

This paper considers the estimation of dynamic causal effects using external instruments and a structural vector-autoregressive model with possibly non-stationary regressors. We provide general conditions under which the asymptotic normal approximation remains valid. In this case, the asymptotic variance depends on the persistence property of each series. We further provide a consistent asymptotic covariance matrix estimator that requires neither such knowledge nor pre-tests for nonstationarity. The proposed consistent covariance matrix estimator is robust and is easy to implement in practice.

Keywords: external instruments, non-stationarity, robust inference, structural VAR
JEL Codes: C32, C36

*University of Pennsylvania, Department of Economics. Email: xucheng@upenn.edu
†City University of Hong Kong, Department of Economics and Finance. Email: xuhan25@cityu.edu.hk
‡Vanderbilt University, Department of Economics. Email: atsushi.inoue@vanderbilt.edu

1 Introduction

To study the dynamic causal effects of macroeconomic shocks, it has become increasingly popular to use external instruments for the identification and estimation of structural vector autoregressive (SVAR) models, following Stock and Watson (2012) and Mertens and Ravn (2013). These instruments are constructed with information outside of the system and their correlation with the structural shocks is used for identification of dynamic casual effects – structural impulse response functions (IRF). Different from identification restrictions within the SVAR system, these external instruments can be viewed as external sources of variation that provide quasi-experiments to identify causal effects (Stock and Watson, 2018). These analytical frameworks typically assume that the external instruments, the structural shocks, and all the variables in the SVAR system are stationary and conduct inference with stationary time series.

This paper is concerned with possible non-stationarity and its impact on the external IV estimation of SVAR models (SVAR-IV). For example, Gertler and Karadi (2015) use the SVAR-IV approach to estimate the dynamic causal effects of a monetary policy, where the baseline SVAR model includes log industrial production, log consumer price index, one year government bond rate, and a credit spread. The first two variables are often regarded as non-stationary. Vector autoregressions (VAR) with nonstationary processes typically involve non-standard inference (Park and Phillips, 1989a, 1989b; Sims, Stock, and Watson, 1990; Toda and Phillips,1993). This leads us to the following questions. Is the standard inference still valid if we conduct the SVAR-IV estimation directly with these possibly non-stationary time series as in Gertler and Karadi (2015)? Do we need to first transform them to stationary time series before the analysis, as is done by Stock and Watson (2018) when they revisit this application? What if some variables are cointegrated with an unknown rank and some variables are highly persistent but not exactly unit root?

To answer the above questions, we provide several robust results for SVAR-IV estimation with possible nonstationary variables in the vector autoregression system. The system may contain unit roots, local-to-unity processes, cointegration, or only stationary variables (Phillips, 1987, 1988; Engle and Granger, 1987). These robust results do not require knowing the persistence property of any series or knowing the cointegration relationship. Therefore, we avoid the pre-test or post-model-selection bias (Leeb and Pötscher, 2005). Such bias could be particularly prominent in the presence of local-to-unity variables (Elliott, 1998).

First, we show that the SVAR-IV estimator of the IRF has an asymptotic normal distribution as long as the system contains some stationary variables or cointegration, or its lag order is larger than one. The asymptotic variance, however, depends on the persistence property, including the classification of stationary variables and the cointegration relationship among nonstationary variables. Second, we provide a consistent estimator of the asymptotic covariance matrix, without using the persistence property. Thus, the t and Wald statistics based on this consistent covariance estimator have standard asymptotic distributions. Third, we show that the optimal weighting matrix under over-identification depends on the persistence property. Nevertheless, we again provide a consistent estimator of the optimal weighting matrix without using the persistence property. We maintain the assumption that the external instruments and the structural shocks are stationary, which is well justified in relevant applications. We also assume that the instruments are strong, and we do not allow for weak instruments as in Montiel Olea, Stock, and Watson (2018).

The robust results in this paper stem from the fact that coefficient estimates of the nonstationary regressors converge at a faster rate and its influence is asymptotically negligible when compared to that from the stationary regressors. This phenomenon has been studied and utilized extensively in the nonstationary VAR literature. Sims, Stock, and Watson (1990) show that a normal approximation is valid in VAR with unit roots as long as the parameter of interest can be written as coefficients of stationary regressors. Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) propose to do standard inference in nonstationary vector autoregression by adding extra lag variables. For autoregression with a unit root, Inoue and Kilian (2002) establish the validity of the residual-based bootstrap by exploring the asymptotic normality of the least squares estimator of the slope parameter. Inoue and Kilian (2019) consider uniform inference on impulse responses of autoregressive processes.

Our work contributes to the nonstationary VAR literature by obtaining the asymptotic normality of structural IRF estimated using external instruments. The SVAR-IV estimation has a generated-regressor issue, where the unobserved errors are replaced by the estimation residuals based on possibly nonstationary regressors. We show that the generated-regressor issue has an impact on the asymptotic distribution and leads to an optimal weighting matrix different from the standard two-stage least squares (2SLS) weighting matrix even in the conditional homoskedastic context.

The present paper is also related to the literature about inference on structural IRF. Confidence bands for IRF with exact and local-to-unity processes are considered by Phillips (1998), Wright (2000), Gospodinov(2004), Pesavento and Rossi (2006) and Mikusheva (2012) among others. We focus on inference on IRF for a given horizon, and unlike the nonstandard inference in these papers, inference based on the asymptotic normality is valid in the present context. Although we focus on IRF for a single horizon, our results provide a basis for joint inference over multiple horizons considered by Inoue and Kilian (2016) and Montiel Olea and Plagborg-Møller (2019).

The rest of the paper is organized as follows. Section 2 presents the model and the estimation procedure. Section 3 studies the asymptotic distribution of the contemporaneous and dynamic IRF based on the SVAR-IV estimation. Section 4 provides a robust consistent covariance matrix estimator without requiring the knowledge of persistence properties. Section 5 proposes an optimal weighting matrix and provides results when the optimal weighting matrix is used. Section 6 presents Monte Carlo simulation results and Section 7 concludes.

2 Model and Estimation

Let $\{Y_t : t = -p+1, ..., T\}$ be an $r \times 1$ vector of observed variables that follows a structural VAR model

$$Y_t = d + \sum_{j=1}^{p} \Phi_j Y_{t-j} + \eta_t \text{ and } \eta_t = H\varepsilon_t, \tag{2.1}$$

where Φ_j is an $r \times r$ coefficient matrix for $j = 1, ..., p$, η_t is the reduced-form error, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{rt})'$ is the vector of structural shocks, and H is an $r \times r$ invertible matrix.[1] Suppose the structural shock of interest is $\varepsilon_{\iota t}$ for some $\iota = 1, ..., r$. To study the IRF with respect to $\varepsilon_{\iota t}$, we study the estimation of the ι^{th} column of H, denoted by $h = (h_1, ..., h_r)'$, with external instruments $\{Z_t = (z_{1t}, ..., z_{kt})' \in R^k : t = 1, ..., T\}$. These external instruments are assumed to satisfy (i) $E(Z_t \varepsilon_{\iota t}) = \alpha \neq 0_k$ and (ii) $E(Z_t \varepsilon_{jt}) = 0_k$ for $j \neq \iota$. Condition (i) ensures that the instruments are relevant and condition (ii) requires that the instruments are orthogonal to other structural shocks. Under these conditions, the instruments satisfy

$$E(\eta_t Z_t') = E(H\varepsilon_t Z_t') = h\alpha' \in R^{r \times k}. \tag{2.2}$$

We allow the series in Y_t to display different degrees of persistence. From a practical perspective, one does not have to know the persistence level of any series to conduct the estimation and inference procedure proposed in this paper. For our theoretical analysis, we write $Y_t = [Y_{1t}', Y_{2t}', Y_{3t}']'$ and assume that Y_{1t} and Y_{2t} follow a local-to-unity vector

[1] We focus on the model without a linear time trend. The presence of a linear time trend does not change our results qualitatively, however.

process and may be cointegrated while Y_{3t} follows a stationary vector process, as specified in (3.1) below. The literature typically assumes that the shock of interest is the first shock ε_{1t} without loss of generality in a stationary VAR system. Here, we denote the shock of interest by $\varepsilon_{\iota t}$ to make it clear that it could be the shock associated with either the nonstationary or the stationary series.

Given that α is unknown and $\alpha \neq 0$, the moment conditions in (2.2) only identify h up to a scale constant. We normalize the ι^{th} element of h to be 1, i.e., $h_\iota = 1$. This normalization pins down the scale of the IRFs by standardizing the contemporaneous effect of the target shock (e.g., an oil price shock) on the corresponding variable (e.g., the oil price). In the existing literature, ε_{1t} is often assumed to be the structural shock of interest, and the first element of h is normalized to be 1.

Removing the constant 1 from h, we define the parameter

$$\theta = [h_1, ..., h_{\iota-1}, h_{\iota+1}, ..., h_r]' \in R^{r-1}. \tag{2.3}$$

Using $h_\iota = 1$, (2.2) is equivalent to the moment conditions

$$E[(\eta_{-\iota,t} - \theta\eta_{\iota,t}) \otimes Z_t] = 0 \in R^{(r-1)k}, \tag{2.4}$$

where $\eta_{\iota,t}$ is the ι^{th} element of η_t and $\eta_{-\iota,t} = [\eta_{1,t}, ..., \eta_{\iota-1,t}, \eta_{\iota+1,t}, ..., \eta_{r,t}]'$ is the rest of η_t with $\eta_{\iota,t}$ removed. Below we study the estimation of θ based on the moments in (2.4).

Because η_t is unobserved, we estimate the level VAR model in (2.1) by OLS and use the residual $\widetilde{\eta}_t = (\widetilde{\eta}_{1t}, ..., \widetilde{\eta}_{rt})'$ to construct the sample moment conditions. Let $\widetilde{\eta}_{-\iota,t}$ and $\widetilde{\eta}_{-\iota,t}$ denote the counterpart of $\eta_{-\iota,t}$ and $\eta_{\iota,t}$, respectively. We estimate θ by minimizing the GMM criterion

$$\begin{aligned} \mathcal{Q}_T(\theta) &= \overline{g}_T(\theta)' W_T \overline{g}_T(\theta), \text{ where} \\ \overline{g}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} [(\widetilde{\eta}_{-\iota,t} - \theta\widetilde{\eta}_{\iota,t}) \otimes Z_t] \end{aligned} \tag{2.5}$$

and W_T is the weighting matrix. The first order condition gives the GMM estimator

$$\begin{aligned} \widehat{\theta} &= \left(\mathcal{A}_T W_T \mathcal{A}_T'\right)^{-1} \mathcal{A}_T W_T \mathcal{G}_T, \text{ where} \\ \mathcal{A}_T &= I_{r-1} \otimes (T^{-1}\sum_{t=1}^{T} \widetilde{\eta}_{\iota,t} Z_t') \text{ and } \mathcal{G}_T = T^{-1}\sum_{t=1}^{T} (\widetilde{\eta}_{-\iota,t} \otimes Z_t). \end{aligned} \tag{2.6}$$

If $W_T = I_{r-1} \otimes (T^{-1}\sum_{t=1}^{T} Z_t Z_t')^{-1}$, $\widehat{\theta}$ is the equation-by-equation 2SLS estimator. We provide the optimal weighting matrix in Section 5 below.

3 Asymptotic Results

In this section, we study the asymptotic property of the GMM estimator $\widehat{\theta}$ and the IRF. Write $Y_t = [Y_{1t}' \; Y_{2t}' \; Y_{3t}']'$, where Y_{1t}, Y_{2t} and Y_{3t} are $r_1 \times 1$, $r_2 \times 1$ and $r_3 \times 1$ vectors, respectively, $r_1 \geq 0$, $r_2 \geq 0$, $r_3 \geq 0$ and $r_1 + r_2 + r_3 = r$. We assume that Y_t follows

$$\begin{aligned} Y_t &= c + Y_t^*, \\ Y_{1t}^* &= \left(I_{r_1} + \frac{1}{T} C \right) Y_{1,t-1}^* + u_{1t}, \\ Y_{2t}^* &= Q Y_{1t}^* + u_{2t}, \\ Y_{3t}^* &= u_{3t}, \\ \Psi(L) u_t &= e_t, \end{aligned} \tag{3.1}$$

where $c = [c_1', c_2', c_3']'$, $Y_t^* = [Y_{1t}^{*\prime}, Y_{2t}^{*\prime}, Y_{3t}^{*\prime}]'$, C is an $r_1 \times r_1$ diagonal matrix with nonpositive diagonal elements, Q is an $r_2 \times r_1$ matrix, $\Psi(L) = I_r - \Psi_1 L - \cdots - \Psi_{p-1} L^{p-1}$ is a $(p-1)$th-order lag polynomial, $u_t = [u_{1t}', u_{2t}', u_{3t}']'$, and $e_t = [e_{1t}', e_{2t}', e_{3t}']'$.[2]

We also assume that Z_t follows a linear process

$$Z_t = \mu_Z + \Xi(L) v_t, \text{ where } \Xi(L) = \sum_{j=0}^{\infty} \Xi_j L^j. \tag{3.2}$$

Assumption LP. (i) The roots of $\Psi(L)$ are all outside the unit circle.
(ii) $\Xi_0 = I_k$, $\Xi(1)$ has full rank, $\sum_{j=0}^{\infty} j^2 ||\Xi_j||^2 < \infty$.
(iii) $\overline{e}_t = [e_t', v_t']'$ is a i.i.d. $(r+k) \times 1$ vector with mean zero, $E(\overline{e}_t \overline{e}_t') = \Sigma$ is positive definite, fourth moment of $\overline{e}_t$ is finite, and e_t is homoskedastic conditional on v_t

The model in (3.1) has a VAR representation as in (2.1). We show in the appendix that the model in (3.1) can also be rearranged and written in an error correction form:

$$\Delta Y_t = A_1 (Y_{1,t-1} - c_1) + A_2 + A_3 D_t + \eta_t, \tag{3.3}$$

where $Y_{1,t-1}$ is the non-stationary lag,

$$D_t = [(Y_{2,t-1} - c_2 - Q(Y_{1,t-1} - c_1))', (Y_{3,t-1} - c_3)', \Delta Y_{t-1}', ..., \Delta Y_{t-p+1}']' \in R^{rp - r_1} \tag{3.4}$$

is a collection of all zero-mean stationary lags, A_1, A_2 A_3 are coefficient matrices, and

$$\eta_t = P e_t \text{ for } P = \begin{bmatrix} I_{r_1} & 0 & 0 \\ Q & I_{r_2} & 0 \\ 0 & 0 & I_{r_3} \end{bmatrix}. \tag{3.5}$$

[2]We assume that $r_2 = 0$ if $r_1 = 0$ and that Q does not have a row of zeros.

Let

$$x_t = [(Y_{1,t-1} - c_1)', 1, D_t']' \tag{3.6}$$

denote the regressors in (3.3). This equivalent representation implies that the least square residual $\widetilde{\eta}_t$ obtained from the VAR model in (2.1) is numerically equivalent to that obtained from regressing ΔY_t on x_t. Therefore, we study the regression in (3.3) in order to study the OLS residual $\widetilde{\eta}_t$.

We have the following weak convergence results following Phillips and Solo (1992).

Lemma 1 *Suppose Assumption LP holds. Then,*
(i)

$$\begin{bmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} e_t \\ T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} v_t \end{bmatrix} \Rightarrow \begin{bmatrix} B_e(s) \\ B_v(s) \end{bmatrix} = \Sigma^{1/2} \begin{bmatrix} W_e(s) \\ W_v(s) \end{bmatrix},$$

where $W_e(s)$ and $W_v(s)$ are $r \times 1$ and $k \times 1$ standard Brownian motions, respectively, and they are independent of each other.
(ii) $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} u_t \Rightarrow B_u(s) = [\Psi(1)]^{-1} B_e(s)$.
(iii) $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \eta_t \Rightarrow B_\eta(s) = P B_e(s)$.
(iv) $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} (Z_t - \mu_Z) \Rightarrow B_z(s) = \Xi(1) B_v(s)$.

Define

$$\begin{aligned}
\Gamma_{DD} &= \lim_{T\to\infty} E(D_t D_t'), \\
\Gamma_{DZ} &= \lim_{T\to\infty} E[D_t (Z_t - \mu_Z)'], \\
\Gamma_{ZZ} &= E[(Z_t - \mu_Z)(Z_t - \mu_Z)'],\ \Gamma = [\Gamma_{DZ} : \Gamma_{DD}], \\
\Gamma_{\eta Z} &= E(\eta_{\iota,t} Z_t'),\ \Lambda_{1Z} = \sum_{h=1}^{\infty} E(u_{1,t} Z_{t+h}'), \\
\gamma &= E(\eta_t \otimes Z_t), \\
\Sigma_\eta &= E[\eta_t \eta_t'], \\
\Omega &= \begin{bmatrix} \Sigma_\eta \otimes \Gamma_{DD} & \Sigma_\eta \otimes \Gamma_{DZ} \\ \Sigma_\eta \otimes \Gamma_{DZ}' & \Sigma_\eta \otimes \Gamma_{ZZ} - \gamma\gamma' \end{bmatrix}.
\end{aligned} \tag{3.7}$$

In some of these definitions, we have $T \to \infty$ built in because D_t is triangular arrays due to the local-to-unity process.

Assumption R1. (i) D_t is non-empty, i.e, $r_2 > 0$, or $r_3 > 0$, or $p \geq 2$.
(ii) The matrices $\Gamma, \Gamma_{\eta Z}, \Omega$ all have full rank.

Let $J_c(s)$ denote an $r_1 \times 1$ vector Ornstein Uhlenbeck process, such that

$$dJ_c(s) = CJ_c(s)ds + dB_{u,1}(s), \tag{3.8}$$

where $B_{u,1}(s)$ is the first $r_1 \times 1$ subvector of $B_u(s)$. Let

$$\Upsilon_T = \begin{bmatrix} T^{\frac{1}{2}} I_{r_1} & 0 \\ 0 & I_{pr-r_1+1} \end{bmatrix}. \tag{3.9}$$

The following results follow from Phillips (1987) and Phillips and Solo (1992).

Lemma 2 *Suppose Assumptions LP and R1 hold. Then,*
(a)

$$T^{-1}\sum_{t=1}^{T}\Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1} \to_d \mathbb{V} = \begin{bmatrix} \int_0^1 J_c(s)J_c(s)'ds & \int_0^1 J_c(s)ds & 0 \\ \int_0^1 J_c(s)'ds & 1 & 0 \\ 0 & 0 & \Gamma_{DD} \end{bmatrix}.$$

(b)

$$\begin{bmatrix} T^{-1}\sum_{t=1}^{T} Y_{1,t-1}^{*}(Z_t-\mu_Z)' \\ T^{-1}\sum_{t=1}^{T}(Z_t-\mu_Z)' \\ T^{-1}\sum_{t=1}^{T} D_t(Z_t-\mu_Z)' \\ T^{-1}\sum_{t=1}^{T}(Z_t-\mu_Z)(Z_t-\mu_Z)' \end{bmatrix} \to_d \begin{bmatrix} \int_0^1 B_{u,1}(s)dB_z(s)' + \Lambda_{1Z} \\ 0_{1\times k} \\ \Gamma_{DZ} \\ \Gamma_{ZZ} \end{bmatrix}.$$

(c)

$$\begin{bmatrix} T^{-\frac{3}{2}}\sum_{t=1}^{T} Y_{1,t-1}^{*} \\ T^{-1}\sum_{t=1}^{T} Y_{1,t-1}^{*}\eta_t' \end{bmatrix} \to_d \begin{bmatrix} \int_0^1 J_c(s)ds \\ \int_0^1 J_c(s)dB_\eta(s)' \end{bmatrix}.$$

(d)

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}\begin{pmatrix} \eta_t \otimes D_t \\ \eta_t \otimes (Z_t-\mu_Z) - E[\eta_t \otimes (Z_t-\mu_Z)] \end{pmatrix} \to_d \begin{pmatrix} \xi_D \\ \xi_Z \end{pmatrix} \sim N(0,\Omega).$$

Let $\mathbb{S}_\theta$ be a $(r-1)\times r$ matrix such that

$$\mathbb{S}_\theta \eta_t \;=\; \eta_{-\iota,t} - \theta\eta_{\iota,t}. \tag{3.10}$$

By definition, it takes the form

$$\mathbb{S}_\theta = [I_{r-1}(1:\iota-1) : -\theta : I_{r-1}(\iota : r-1)], \tag{3.11}$$

where $I_{r-1}(1:\iota-1)$ is the first $(\iota-1)$ columns of I_{r-1} and $I_{r-1}(\iota : r-1)$ is the last $(r-\iota)$ matrix of I_{r-1}.

Theorem 1 *Suppose Assumptions LP and R1 hold and $W_T \to_p W$. Then,*
(a)

$$T^{\frac{1}{2}}\left(\widehat{\theta}-\theta\right) \to_d \left(\mathcal{A}W\mathcal{A}'\right)^{-1}\mathcal{A}W \cdot \left[-\left(\mathbb{S}_\theta \otimes \mathcal{K}\right)\xi_D + \left(\mathbb{S}_\theta \otimes I_k\right)\xi_Z\right],$$

where $\mathcal{A} = I_{r-1} \otimes \Gamma_{\eta Z}$, $\mathcal{K} = \Gamma'_{DZ}\Gamma_{DD}^{-1}$.
(b) The optimal choice of the weighting matrix is V^{-1}, where

$$V = \mathcal{B}\Omega\mathcal{B}' \text{ and } \mathcal{B} = [-\mathbb{S}_\theta \otimes \mathcal{K} : \mathbb{S}_\theta \otimes I_k].$$

Because Γ_{DZ} is nonzero in general, replacing η with $\widetilde{\eta}$ affects the asymptotic distribution of $\widehat{\theta}$.

Next, we study the asymptotic distribution of the IRF. We start with the moving average (MA) coefficients Θ_s in the vector moving average (VMA) representation of (2.1), i.e.,

$$Y_t = d + \eta_t + \sum_{s=1}^{\infty} \Theta_s \eta_{t-s}. \tag{3.12}$$

By definition, $\Theta_s = I_r$ for $s = 0$. Define the companion matrix for the VAR presentation in (2.1) as

$$\mathbb{F} = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p \\ I_r & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & I_r & 0 \end{bmatrix}. \tag{3.13}$$

Then $\Theta_s = \mathbb{M}'\mathbb{F}^s\mathbb{M}$ for $\mathbb{M}' = [I_r, 0, ..., 0]$. We estimate Θ_s by $\widetilde{\Theta}_s = \mathbb{M}'\widetilde{\mathbb{F}}^s\mathbb{M}$, where $\widetilde{\mathbb{F}}$ is defined analogously to $\mathbb{F}$ but with Φ_j for $j = 1, ..., p$ replaced by their OLS estimator $\widetilde{\Phi}_j$ based on (2.1).

To derive the distribution of $\widetilde{\Theta}_s$, we first define a matrix $\mathbb{L}$ that transforms the regressors in (2.1), denoted by X_t, to those in (3.3), denoted by x_t, i.e.,

$$x_t \equiv \begin{bmatrix} Y_{1,t-1} - c_1 \\ 1 \\ D_t \end{bmatrix} = \mathbb{L}\begin{bmatrix} 1 \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix} \equiv \mathbb{L}X_t. \tag{3.14}$$

By the definition of D_t in (3.4), we have

$$\mathbb{L} = \begin{bmatrix} -c_1 & I_{r_1} & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ Qc_1 - c_2 & -Q & I_{r_2} & 0 & 0 & \cdots & 0 \\ -c_3 & 0 & 0 & I_{r_3} & 0 & \cdots & 0 \\ & I_{r_1} & 0 & 0 & & & \\ 0 & 0 & I_{r_2} & 0 & -I_r & \cdots & 0 \\ & 0 & 0 & I_{r_3} & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & I_r & -I_r \end{bmatrix}. \tag{3.15}$$

Because the VAR model in (2.1) can be equivalently written as in (3.3), the OLS estimators of the coefficients in (2.1) and those in (3.3) satisfy

$$\left[\widetilde{d} : \ \widetilde{\Phi}_1 - I_r : \ \widetilde{\Phi}_2 : \cdots : \ \widetilde{\Phi}_p\right] \ = \ \left[\widetilde{A}_1 : \ \widetilde{A}_2 : \ \widetilde{A}_3\right] \mathbb{L}. \tag{3.16}$$

Thus, we can study the distribution of $\widetilde{\Theta}_s$ using the equivalent representation in (3.3) and the asymptotic results in Lemma 2.

Define

$$\overline{\mathbb{L}} = \begin{bmatrix} -Q & 0 & I_{r_1} & 0 & 0 & \cdots & 0 \\ I_{r_2} & 0 & 0 & I_{r_2} & 0 & \cdots & 0 \\ 0 & I_{r_3} & 0 & 0 & I_{r_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & -I_r \end{bmatrix}, \tag{3.17}$$

which is a $rp \times (rp - r_1)$ lower-right submatrix of $\mathbb{L}'$ used in the transformation above. Define

$$\mathcal{R} = \sum_{i=0}^{s-1} \Theta_{s-1-i} \otimes \left(\mathbb{M}'\mathbb{F}^{i\prime}\right) \text{ and } \mathcal{J} = \overline{\mathbb{L}}\Gamma_{DD}^{-1}. \tag{3.18}$$

Assumption R2. The matrix $\mathcal{R}$ has full rank.

Theorem 2 *Suppose Assumptions LP and R1-R2 hold. Then for $s \geq 1$*

$$T^{\frac{1}{2}}\text{vec}(\widetilde{\Theta}_s' - \Theta_s') \ \to_d \ \mathcal{R}(I_r \otimes \mathcal{J})\xi_D.$$

Next, we consider the asymptotic distribution of the IRFs. For a fixed horizon $s \geq 1$, the IRF is defined as

$$\beta_s = \frac{\partial Y_{t+s}}{\partial \varepsilon_{\iota,t}} = \Theta_s h = \Theta_s \mathbb{S}_\iota \begin{bmatrix} 1 \\ \theta \end{bmatrix}, \tag{3.19}$$

where

$$\mathbb{S}_\iota \equiv \begin{bmatrix} 0 & I_{\iota-1} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & I_{r-\iota} \end{bmatrix} \in R^{r\times r} \tag{3.20}$$

rearranges the elements of $(1, \theta')'$ such that it becomes h. Let $\overline{\mathbb{S}}_\iota$ denote the last $r-1$ columns of $\mathbb{S}_\iota$. The estimator of β_s is

$$\widehat{\beta}_s = \widetilde{\Theta}_s \widehat{h} = \widetilde{\Theta}_s \mathbb{S}_\iota \begin{bmatrix} 1 \\ \widehat{\theta} \end{bmatrix}. \tag{3.21}$$

Define

$$\begin{aligned} G_{1s} &= \Theta_s \overline{\mathbb{S}}_\iota \left[\mathcal{A}W\mathcal{A}'\right]^{-1} \mathcal{A}W[-\mathbb{S}_\theta \otimes \mathcal{K} : \mathbb{S}_\theta \otimes I_k], \\ G_{2s} &= \left[(I_r \otimes h')\mathcal{R}(I_r \otimes \mathcal{J}) : 0_{r\times rk}\right]. \end{aligned} \tag{3.22}$$

Assumption R3. $G_{1s} + G_{2s}$ has rank r.

Theorem 3 *Suppose that Assumptions LP and R1-R3 hold and* $W_T \to_p W$, *then*

$$T^{\frac{1}{2}}(\widehat{\beta}_s - \beta_s) \to_d \ N(0, (G_{1s} + G_{2s})\Omega(G_{1s} + G_{2s})').$$

In the asymptotic distribution in Theorem 3, the first part associated with G_{1s} comes from the estimation of the contemporaneous IRF $\widehat{h}$, whose random elements are $\widehat{\theta}$, and the second part associated with G_{2s} comes from the estimation of the MA parameter $\widetilde{\Theta}_s$ for the dynamic response.

Constructing the asymptotic variance of $\widehat{\beta}_s$ using sample analogs of G_{1s}, G_{2s}, and Ω requires one to distinguish stationary and nonstationary series in Y_t and specify the cointegration relationship among the nonstationary series. To see this, note that $\mathcal{K}$, $\mathcal{J}$, and Ω by definition are all constructed with D_t defined in (3.4). Using model selection procedures or pre-tests to specify D_t may result in model selection errors and undesirable consequences for subsequent inference. Below we provide a consistent covariance matrix estimator that avoids this specification problem.

4 Consistent Covariance Matrix Estimator

In this section, we propose a robust consistent estimator of the asymptotic variance of $\widehat{\beta}_s$. The key feature is that it does not require knowing the persistence property of any series or any cointegrating relationship. It is constructed with the whole vector Y_t in the VAR system, instead of the stationary regressors only. We show that it is consistent under all different forms of nonstationarity allowed in this paper.

In the estimation of the covariance matrix, the main challenge comes from the estimation of $\mathcal{K}$, $\mathcal{J}$, and Ω, all of which are defined with the stationary regressors only. Without distinguishing the stationary regressors from the nonstationary ones, we use $X_t = [1, Y_{t-1}, ..., Y_{t-p}]'$ and propose to estimate $\mathcal{K}$ and $\mathcal{J}$, respectively, by

$$\begin{aligned}
\widehat{\mathcal{K}} &= \widehat{\Gamma}_{ZX}\widehat{\Gamma}_{XX}^{-1} \text{ and } \widehat{\mathcal{J}} = \mathbb{S}_2\widehat{\Gamma}_{XX}^{-1}, \text{ where} \\
\widehat{\Gamma}_{XX} &= \frac{1}{T}\sum_{t=1}^{T} X_t X_t', \\
\widehat{\Gamma}_{ZX} &= \frac{1}{T}\sum_{t=1}^{T}(Z_t - \overline{Z}_T)X_t', \ \overline{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t, \qquad (4.1)
\end{aligned}$$

and $\mathbb{S}_2 = [0_{rp\times 1} : I_{rp}]$ is a selector matrix. Lemma 3 below shows that some proper rotation with the matrix $\mathbb{L}$ and rescaling using the matrix Υ_T lead to the limits of $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{J}}$ that contain $\mathcal{K}$ and $\mathcal{J}$ as subvectors, respectively.

Using $\widehat{\mathcal{K}}$ and $\widehat{\mathcal{J}}$, we construct

$$\begin{aligned}
\widehat{G}_{1s} &= \widetilde{\Theta}_s\overline{S}_\iota(\mathcal{A}_T W_T \mathcal{A}_T')^{-1}\mathcal{A}_T W_T[-S_{\widehat{\theta}} \otimes \widehat{\mathcal{K}} : S_{\widehat{\theta}} \otimes I_k], \\
\widehat{G}_{2s} &= \left[(I_r \otimes \widehat{h}')\widehat{\mathcal{R}}\left(I_r \otimes \widehat{\mathcal{J}}\right) : 0_{r\times rk}\right], \qquad (4.2)
\end{aligned}$$

where $S_{\widehat{\theta}}$ and $\widehat{h}$ are defined as S_θ and h with θ replaced by $\widehat{\theta}$, respectively, $\widehat{\mathcal{R}} = \sum_{i=0}^{s-1}\widetilde{\Theta}_{s-1-i}\otimes (\mathbb{M}'\widetilde{\mathbb{F}}^{i\prime})$, and $\widetilde{\mathbb{F}}$ is defined as $\mathbb{F}$ with $\Phi_1, ..., \Phi_p$ replaced by $\widetilde{\Phi}_1, ..., \widetilde{\Phi}_p$.

Let

$$\mathbb{P} = \begin{bmatrix} I_r \otimes \mathbb{L}^{-1}\Upsilon_T & 0 \\ 0 & I_{rk} \end{bmatrix}. \qquad (4.3)$$

Lemma 3 *Suppose Assumptions LP and R1 hold. Then,*

$$\begin{aligned}
\widehat{\mathcal{K}}\mathbb{L}^{-1}\Upsilon_T &\to_p [0_{k\times(r_1+1)} : \mathcal{K}], \\
\widehat{\mathcal{J}}\mathbb{L}^{-1}\Upsilon_T &\to_p [0_{rp\times(r_1+1)} : \mathcal{J}],
\end{aligned}$$

and

$$\widehat{G}_{1s}\mathbb{P} \to_p \Theta_s\overline{S}_\iota(\mathcal{A}W\mathcal{A}')^{-1}\mathcal{A}W[-S_\theta \otimes [0_{k\times(r_1+1)} : \mathcal{K}] : S_\theta \otimes I_k],$$
$$\widehat{G}_{2s}\mathbb{P} \to_p \left[(I_r \otimes h')\mathcal{R}\left(I_r \otimes [0_{rp\times(r_1+1)} : \mathcal{J}]\right) : 0_{r\times rk}\right].$$

The limits of $\widehat{G}_{1s}\mathbb{P}$ and $\widehat{G}_{2s}\mathbb{P}$ are analogous to G_{1s} and G_{2s} but with $\mathcal{K}$ and $\mathcal{J}$ augmented with r_1+1 columns of zeros in the front. This shows that even if we do not know which series are non-stationary, their effects are asymptotically negligible after the rotation and rescaling by the matrix $\mathbb{P}$.

Next, we consider estimation of the covariance Ω. Using $X_t, Z_t, \widetilde{\eta}_t$, we propose to estimate Ω by

$$\widehat{\Omega} = \begin{bmatrix} \widehat{\Sigma}_\eta \otimes \widehat{\Gamma}_{XX} & \widehat{\Sigma}_\eta \otimes \widehat{\Gamma}'_{ZX} \\ \widehat{\Sigma}_\eta \otimes \widehat{\Gamma}_{ZX} & \widehat{\Sigma}_\eta \otimes \widehat{\Gamma}_{ZZ} - \widehat{\gamma}\widehat{\gamma}' \end{bmatrix}, \tag{4.4}$$

where

$$\begin{aligned} \widehat{\Sigma}_\eta &= T^{-1}\sum_{t=1}^{T}\widetilde{\eta}_t\widetilde{\eta}_t' \\ \widehat{\Gamma}_{ZZ} &= \frac{1}{T}\sum_{t=1}^{T}(Z_t-\overline{Z}_T)(Z_t-\overline{Z}_T)', \\ \widehat{\gamma} &= T^{-1}\sum_{t=1}^{T}\widetilde{\eta}_t \otimes Z_t. \end{aligned} \tag{4.5}$$

Define

$$\gamma_{xZ} = \begin{bmatrix} 0_{(r_1+1)\times k} \\ \Gamma_{DZ} \end{bmatrix}. \tag{4.6}$$

Lemma 4 *Suppose Assumptions LP and R1 hold. Then,*

$$\mathbb{P}^{-1}\widehat{\Omega}\mathbb{P}^{-1\prime} \to_d \begin{bmatrix} \Sigma_\eta \otimes \mathbb{V} & \Sigma_\eta \otimes \gamma_{xZ} \\ \Sigma_\eta \otimes \gamma'_{xZ} & \Sigma_\eta \otimes \Gamma_{ZZ} - \gamma\gamma' \end{bmatrix}.$$

Comparing the limit of $\mathbb{P}^{-1}\widehat{\Omega}\mathbb{P}^{-1\prime}$ and Ω, we see that $\mathbb{V}$ and γ_{xZ} contain Γ_{DD} and Γ_{DZ} as submatrices. Theorem 4 below shows that the covariance matrix estimator is consistent, because the extra r_1+1 columns of zeros in Lemma 3 reduce $\mathbb{V}$ and γ_{xZ} to Γ_{DD} and Γ_{DZ}, respectively.

Theorem 4 *Suppose Assumptions LP and R1 hold and $W_T \to_p W$. Then*

$$(\widehat{G}_{1s}+\widehat{G}_{2s})\widehat{\Omega}(\widehat{G}_{1s}+\widehat{G}_{2s})' \overset{p}{\to} (G_{1s}+G_{2s})\Omega(G_{1s}+G_{2s})'.$$

5 Optimal GMM Estimation

Following Theorem 1(b), the optimal GMM estimation uses the weighting matrix $W_T = \widehat{V}^{-1}$, where $\widehat{V}$ is a consistent estimator of $V = \mathcal{B}\Omega\mathcal{B}'$. Note that because of the generated regressor the optimal weighting matrix is different from the weighting matrix implicit for the 2SLS estimator even in the absence of conditional heteroskedasticity.

We estimate V by

$$\widehat{V} = \widehat{\mathcal{B}}\widehat{\Omega}\widehat{\mathcal{B}}', \text{ where } \widehat{\mathcal{B}} = [-\mathbb{S}_{\widetilde{\theta}} \otimes \widehat{\mathcal{K}} : \mathbb{S}_{\widetilde{\theta}} \otimes I_k], \tag{5.1}$$

where $\widetilde{\theta}$ is a preliminary consistent estimator of θ. The consistency of $\widehat{V}$ follows the same arguments used to show Theorem 4.

Let $\widehat{\theta}^o$ denote the two-step GMM estimator. In the first step, we use either $I_{(r-1)k}$ or $I_{r-1} \otimes (T^{-1}\sum_{t=1}^{T} Z_t Z_t')^{-1}$ as the weighting matrix and compute the GMM estimator $\widetilde{\theta}$ following (2.6). In the second step, we compute $\widehat{V}$ with $\widetilde{\theta}$ and obtain the GMM estimator $\widehat{\theta}_o$ with weighting matrix $\widehat{V}^{-1}$. Let $\widehat{\beta}_s^o$ denote the IRF calculated with $\widehat{\theta}_o$.

The following theorem summarizes the properties of the optimal GMM estimator.

Theorem 5 *Suppose Assumptions LP,R1-R3 hold. Then*
(a) $\widehat{V} \to_p V$.
(b) $T^{\frac{1}{2}}(\widehat{\theta}^o - \theta) \to_d N(0, [\mathcal{A}V^{-1}\mathcal{A}']^{-1})$.
(c) $T^{\frac{1}{2}}(\widehat{\beta}_s^o - \beta_s) \to_d N(0, (G_{1s}^o + G_{2s})\Omega(G_{1s}^o + G_{2s})')$, *where* G_{1s}^o *is defined as* G_{1s} *but with* W *replaced with* V^{-1}.

The asymptotic covariance of $\widehat{\theta}^o$ can be consistently estimated with $\mathcal{A}$ and V replaced by $\mathcal{A}_T$ and $\widehat{V}$, respectively. The asymptotic covariance of $\widehat{\beta}_s^o$ can be consistently estimated following Theorem 4 with W_T replaced by $\widehat{V}^{-1}$.

The t statistic and the Wald statistic based on the consistent covariance estimator have asymptotic normal or chi-square distribution, respectively.

Finally, it is worth mentioning that, although all the results are robust to the presence of nonstationary time series, neither the estimators nor their consistent covariance estimators require practitioners to specify which series are stationary. The robustness condition holds as long as Y_t contains stationary or cointegrated regressors or the VAR order is larger than one.

6 Simulations

To study the finite-sample performance of inference based on the asymptotic distributions derived above, we consider the following data generating processes (DGPs):

$$u_t = \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + e_t, \tag{6.1}$$

where $u_t = [u_{1t}, u_{2t}, u_{3t}]' \in R^3$, $\Psi_1 = 0.5 I_3$,

$$\Psi_2 = \begin{bmatrix} 0 & 0 & 0.2 \\ 0 & 0.2 & 0 \\ 0.2 & 0 & 0 \end{bmatrix}, \tag{6.2}$$

and $e_t \sim N(0, I_3)$ is i.i.d. over t. The matrix Ψ_2 allows for spillover effects between u_{1t} and u_{3t}. We generate y^*_{1t}, y^*_{2t} and y^*_{3t} using two DGPs below.

DGP1: y^*_{2t} is cointegrated with y^*_{1t}.

$$\begin{aligned} y^*_{1t} &= \left(1 - \frac{C_1}{T}\right) y^*_{1t-1} + u_{1t}, \\ y^*_{2t} &= 2y^*_{1t} + u_{2t}, \\ y^*_{3t} &= u_{3t}, \end{aligned} \tag{6.3}$$

so y^*_{1t}, y^*_{2t}, and y^*_{2t} correspond to Y^*_{1t}, Y^*_{2t}, and Y^*_{3t} in (3.1), respectively. The model under DGP1 contains only one root equal or close to unity. We set the drift parameter $C_1 \in [0, -2, -5, -10]$ following Stock (1991).

DGP2: y^*_{2t} is not cointegrated with y^*_{1t}.

$$\begin{aligned} y^*_{1t} &= y^*_{1t-1} + u_{1t}, \\ y^*_{2t} &= \left(1 - \frac{C_2}{T}\right) y^*_{2t-1} + u_{2t}, \\ y^*_{3t} &= u_{3t}. \end{aligned} \tag{6.4}$$

We set $C_2 \in [-2, -5, -10, -0.5T]$. For $C_2 = -2$, -5, or -10, the model under DGP2 has one unit root and a local-to-unity root. For $C_2 = -0.5T$, y^*_{2t} is stationary and the model has only one unit root.

Given y^*_t, we generate the observed data $Y_t = c + y^*_t$, where $y^*_t = [y^*_{1t},\ y^*_{2t}, y^*_{3t}]'$ and $c = [1, 0.5, -1]'$. The reduced-form errors are given by

$$\eta_t = Pe_t, \text{ where } P = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.5}$$

for DGP1 and $P = I_3$ for DGP2 following the derivation of P in (A.2). Next, we specify the matrix H. Since $\eta_t = H\varepsilon_t$ and $e_t \sim N(0, I_3)$ and they are i.i.d. over t, the normalization $E(\varepsilon_t \varepsilon_t') = I_3$ implies $E(\eta_t \eta_t') = PP' = HH'$. We set H equal to the positive definite square root of PP'. Given H and η_t, we can obtain the true values of the structural shocks by computing $\varepsilon_t = H^{-1}\eta_t$. Finally, we generate the instrument Z_{jt} that is correlated with ι-th structural shock at time t by

$$Z_{jt} = \sqrt{1-a^2} w_{jt} + a\varepsilon_{\iota t} + \varepsilon_{3t-1} \text{ for } j = 1, ..., k,$$

where w_{jt}'s are i.i.d. standard normal random variables, and a is set equal to $\sqrt{2}/2$ so that the correlation between Z_{jt} and $\varepsilon_{\iota t}$ is equal to 0.5. The instruments are correlated with the lags of the structural shock, so Γ_{ZD} is not zero and the estimation uncertainty in $\widehat{\eta}$ affects the asymptotic distribution of $\widehat{\theta}$. We set $k = 2$ in the simulation.

With the observed data Y_t, we run an OLS estimation to fit a VAR(3) model with an intercept. The MA coefficients $\widetilde{\Theta}_s$ are computed using these OLS estimates. The residuals $\widetilde{\eta}_t$ and the instruments are used for the estimation of θ. We implement a two-step GMM estimation. In the first step, we obtain a consistent estimator of θ using the weighting matrix $I_r \otimes (T^{-1}\sum_{t=1}^{T} Z_t Z_t')^{-1}$. Then we re-estimate θ using the optimal weighting matrix $\widehat{V}^{-1}$, with $\widehat{V}$ given by (5.1). The confidence intervals for the structural IRFs are computed based on the asymptotic normal distribution in Theorem 5 and the proposed consistent estimator of the covariance matrix. Our inferential results do not require the knowledge about the cointegrating relationship or the (non)stationarity of a particular series.

Tables 1A-1C and 2A-2C report the finite-sample coverage rates of confidence intervals for DGP1 and DGP2, respectively. The nominal level is 95%. The number of simulation replications is 5000. The notation "-" denotes that the corresponding contemporaneous IRF is normalized to be one. Under DGP1, the coverage rates of confidence intervals of the IRFs to a shock in ε_{jt} for $j = 1, 2, 3$ are summarized by Tables 1A, 1B and 1C, respectively. The three rows correspond to the same value of C_1 in Tables 1A-1C report the results for Y_{1t}, Y_{2t}, and Y_{3t}, respectively.

The results in Tables 1A-1C show several patterns. First, the coverage rates are close to the nominal level for short horizons. Even in small samples with $T = 200$, for example, the effective coverage rates of the confidence intervals are always between 93.4% and 94.7% for horizon $s = 0$ and between 91.8% and 94.6% for horizon $s = 2$. Second, as the horizon increases, the effective coverage rates decrease. This is not surprising because (1) the estimators $\widetilde{\Theta}_s$ based on the unrestricted OLS estimation are inconsistent for long horizons and its limiting distribution is also nonstandard in the presence of roots equal or close to unity (Phillips, 1998; Gospodinov, 2004; Pesavento and Rossi, 2006)[3]; and (2) the asymptotic normal approximation based on the Delta method can perform poorly in small samples even in stationary VARs since the IRFs are highly nonlinear functions of the VAR coefficients (Kilian, 1999). Finally, the effective coverage rates are improving as the sample size increases, which confirms our asymptotic theory.

The main patterns in Tables 2A-2C are similar to those in Tables 1A-1C. Compared to the results under DGP1, the effective coverage rates in Table 2 tend to have larger downward biases, especially for shocks to ε_{1t} and ε_{2t} with $C_2 \in \{-2, -5, -10\}$. This could be caused by the fact that the system under DGP2 has one more large root than DGP1 for $C_2 \in \{-2, -5, -10\}$. Under DGP2, the asymptotic distribution tends to require larger samples to generate good approximations. Additional simulation results (not included in Table 2) confirm that the effective coverage rates are much closer to the nominal level under DGP2 for all IRFs with $s \leq 12$ when T increases to 5000.

7 Conclusion

This paper shows that for SVAR estimation with external instruments, standard asymptotic normal inference remains valid under a general form of nonstationarity in the vector autoregression system. In the presence of stationary regressors, cointegration relatioships, or more than one lag variables in the vector autoregression, the estimation error from the nonstationary component is asymptotically negligible. The asymptotic variance of the IRF only depends on the stationary component, but a consistent covariance matrix estimator is available even without knowing which series are stationary. This robust and simple covariance matrix estimator is particularly appealing for practical applications.

The robustness result is rather general by allowing for local-to-unity processes. However, the theoretical result is not uniform over the entire parameter space of the roots of the autoregssive model, as those studied in Mikusheva (2007, 2012), Andrews and Guggenberger (2010), and Phillips (2014). Establishing uniform inference for the SVAR-IV estimation is an interesting direction for future research.

[3] In this paper, our asymptotic theory is based on the assumption that the horizon s is a fixed number as $T \to \infty$. Under the alternative asymptotic setup where s grows at the same rate as T, the asymptotic distribution of the IRFs will be different and nonstandard when the VAR model has large roots.

A Appendix

Below we first show the representation in (3.3). Following (3.1), we can write

$$\Delta Y_t = M(Y_{t-1} - c) + P\Psi(L)^{-1}e_t, \text{ where} \tag{A.1}$$

$$\begin{aligned} M &= \begin{bmatrix} T^{-1}C & 0 & 0 \\ Q\left(I_{r_1} + T^{-1}C\right) & -I_{r_2} & 0 \\ 0 & 0 & -I_{r_3} \end{bmatrix} \text{ and} \\ P &= \begin{bmatrix} I_{r_1} & 0 & 0 \\ Q & I_{r_2} & 0 \\ 0 & 0 & I_{r_3} \end{bmatrix}. \end{aligned} \tag{A.2}$$

Multiplying both sides of (A.1) by $P\Psi(L)P^{-1}$, we obtain

$$P\Psi(L)P^{-1}\Delta Y_t = P\Psi(L)P^{-1}M(Y_{t-1} - c) + Pe_t. \tag{A.3}$$

Define $\pi_1 = \Psi_1 + \cdots + \Psi_{p-1}$, $\pi_2 = \Psi_2 + \cdots + \Psi_{p-1}$, ..., and $\pi_{p-1} = \Psi_{p-1}$. We can write

$$\begin{aligned} \Psi(L) &= 1 - \Psi_1 L - \cdots - \Psi_{p-1}L^{p-1} \\ &= \Psi(1) + \pi_1(1 - L) + \pi_2(L - L^2) + \cdots + \pi_{p-1}(L^{p-2} - L^{p-1}). \end{aligned} \tag{A.4}$$

Plugging (A.4) in (A.3), we obtain

$$\begin{aligned} & P\left[1 - \Psi_1 L - \cdots - \Psi_{p-1}L^{p-1}\right] P^{-1}\Delta Y_t \qquad \text{(A.5)} \\ = \; & P\left[\Psi(1) + \pi_1(1 - L) + \pi_2(L - L^2) + \cdots + \pi_{p-1}(L^{p-2} - L^{p-1})\right] P^{-1}M(Y_{t-1} - c) + Pe_t, \end{aligned}$$

and an rearrangement gives

$$\begin{aligned} \Delta Y_t = \; & P\Psi(1)P^{-1}M(Y_{t-1} - c) + \qquad \text{(A.6)} \\ & (P\Psi_1 P^{-1} + P\pi_1 P^{-1}M)\Delta Y_{t-1} + \cdots + (P\Psi_{p-1}P^{-1} + P\pi_{p-1}P^{-1}M)\Delta Y_{t-p+1} + Pe_t. \end{aligned}$$

Define

$$\Pi(L) = \Pi_1 + \Pi_2 L + \cdots + \Pi_{p-1}L^{p-2}, \tag{A.7}$$

where $\Pi_1 = P\Psi_1 P^{-1} + P\pi_1 P^{-1}M, ..., \Pi_{p-1} = P\Psi_{p-1}P^{-1} + P\pi_{p-1}P^{-1}M$. Then, we can write the model as

$$\Delta Y_t = (P\Psi(1)P^{-1}M)(Y_{t-1} - c) + \Pi(L)\Delta Y_{t-1} + \eta_t, \text{ where } \eta_t = Pe_t. \tag{A.8}$$

Using the definition of M and D_t, (A.8) can be equivalently written as

$$\begin{aligned} \Delta Y_t &= Ax_t + \eta_t \\ &= A_1(Y_{1,t-1} - c_1) + A_2 + A_3 D_t + \eta_t, \end{aligned} \tag{A.9}$$

where

$$\begin{aligned} A &= [A_1 : A_2 : A_3],\ x_t = [(Y_{1,t-1} - c_1)', 1, D_t']', \\ A_1 &= P\Psi(1)P^{-1}\begin{bmatrix} T^{-1}C \\ T^{-1}QC \\ 0 \end{bmatrix},\ A_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ A_3 &= \left[P\Psi(1)P^{-1}\begin{bmatrix} 0 & 0 \\ -I_{r_2} & 0 \\ 0 & -I_{r_3} \end{bmatrix}, \Pi_1, \cdots, \Pi_{p-1} \right]. \end{aligned} \tag{A.10}$$

□

Proof of Lemma 1. The results follow from Theorem 3.4 of Phillips and Solo (1992). The summability condition for the linear processes are satisfied by Assumption LP(i) and LP(ii).

□

Proof of Lemma 2. Parts (a)-(c) follow from the convergence of the near unit root process in Phillips (1987) and convergence of linear processes as in Phillips and Solo (1992).

Parts (d) applies the CLT for second moments of linear processes as in Theorem 3.8 of Phillips and Solo (1992). The liming asymptotic distribution is $\sum_{j=-\infty}^{\infty} \Omega_j$, where

$$\begin{aligned} \Omega_j &= \begin{bmatrix} \Omega_{DD,j} & \Omega_{DZ,j} \\ \Omega_{ZD,j} & \Omega_{ZZ,j} \end{bmatrix}, \\ \Omega_{DD,j} &= \lim_{T\to\infty} E\left[(\eta_t \otimes D_t)(\eta_{t-j} \otimes D_{t-j})' \right], \\ \Omega_{DZ,j} &= \lim_{T\to\infty} E\left[(\eta_t \otimes D_t)(\eta_{t-j} \otimes (Z_{t-j} - \mu_Z) - \gamma)' \right], \\ \Omega_{ZD,j} &= \lim_{T\to\infty} E\left[(\eta_t \otimes (Z_t - \mu_Z) - \gamma)(\eta_{t-j} \otimes D_{t-j})' \right], \\ \Omega_{ZZ,j} &= E\left[(\eta_t \otimes (Z_t - \mu_Z) - \gamma)(\eta_{t-j} \otimes (Z_{t-j} - \mu_Z) - \gamma)' \right]. \end{aligned} \tag{A.11}$$

Below, we show $\Omega_j = 0$ for $j \neq 0$ and $\Omega_0 = \Omega$.

Let $\mathcal{F}_{t-1}$ denote the information set at $t-1$ generated by $\{\eta_{t-1}, Z_{t-1}, \eta_{t-2}, Z_{t-2}, ...\}$. Note that in the VAR model, D_t is a function of η_{t-1} and its lags. Because η_t and Z_t are both linear processes and the errors are i.i.d. by Assumption LP, we have (i)$E[\eta_t|\mathcal{F}_{t-1}] = 0$, and

$E[\eta_t \otimes Z_t|\mathcal{F}_{t-1}] = \gamma_T$ and $E[\eta_t\eta_t'|\mathcal{F}_{t-1}] = \Sigma_\eta$ are constant for any T, and (ii) $\lim_{T\to\infty} E[\eta_t \otimes Z_t|\mathcal{F}_{t-1}] = \gamma$ and $E[\eta_t\eta_t'|\mathcal{F}_{t-1}] = \Sigma_\eta$. Therefore, all the auto-covariance Ω_j with $j \neq 0$ are zero by the law of iterated expectation (LIE).

For $j = 0$, we show $\Omega_0 = \Omega$, i.e., the matrices have the Kronecker product structure. To this end, note that

$$\begin{aligned}\Omega_{DD,0} &= E\left[(\eta_t\eta_t') \otimes (D_tD_t')\right] \\ &= E\left[E\left(\eta_t\eta_t'|\mathcal{F}_{t-1}\right) \otimes (D_tD_t')\right] = \Sigma_\eta \otimes \Gamma_{DD}, \end{aligned} \tag{A.12}$$

by LIE and $E\left(\eta_t\eta_t'|\mathcal{F}_{t-1}\right) = \Sigma_\eta$. Next,

$$\begin{aligned}\Omega_{DZ,0} &= \lim_{T\to\infty} E\left[(\eta_t\eta_t') \otimes D_tZ_t'\right] \\ &= \lim_{T\to\infty} E\left[E\left(\eta_t\eta_t'|\mathcal{F}_{t-1}, v_t\right) \otimes (D_tZ_t')\right] \\ &= \Sigma_\eta \otimes \Gamma_{DZ}, \end{aligned} \tag{A.13}$$

because $\eta_t = Pe_t$ and e_t is homoskedastic conditional on $\mathcal{F}_{t-1}$ and v_t. We have $\Omega_{ZD,0} = \Omega_{DZ,0}'$. Finally,

$$\begin{aligned}\Omega_{ZZ,0} &= \lim_{T\to\infty} E\left[\eta_t\eta_t' \otimes (Z_t - \mu_Z)(Z_t - \mu_Z)'\right] - \gamma\gamma' \\ &= \lim_{T\to\infty} E\left[E\left(\eta_t\eta_t'|\mathcal{F}_{t-1}, v_t\right) \otimes (Z_t - \mu_Z)(Z_t - \mu_Z)'\right] - \gamma\gamma' \\ &= \Sigma_\eta \otimes \Gamma_{ZZ} - \gamma\gamma'. \end{aligned} \tag{A.14}$$

□

Proof of Theorem 1. Applying the formula for $\widehat{\theta}$ in (2.6), we have

$$\begin{aligned}T^{\frac{1}{2}}\left(\widehat{\theta} - \theta\right) &= \left(\mathcal{A}_T W_T \mathcal{A}_T'\right)^{-1} \mathcal{A}_T W_T B, \text{ where} \\ \mathcal{A}_T &= I_{r-1} \otimes \left(T^{-1}\sum_{t=1}^{T} \widetilde{\eta}_{\iota,t} Z_t'\right) \text{ and} \\ B &= T^{-\frac{1}{2}}\left[\sum_{t=1}^{T} \widetilde{\eta}_{-\iota,t} \otimes Z_t - \left(I_{r-1} \otimes \sum_{t=1}^{T} Z_t \widetilde{\eta}_{\iota,t}\right)\theta\right] \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} \left(\widetilde{\eta}_{-\iota,t} - \theta\widetilde{\eta}_{\iota,t}\right) \otimes Z_t \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} \left(\mathbb{S}_\theta \widetilde{\eta}_t\right) \otimes (Z_t - \mu_Z), \end{aligned} \tag{A.15}$$

where the second last equality uses $(I_{r-1} \otimes Z_t \widetilde{\eta}_{\iota,t})\theta = \theta\widetilde{\eta}_{\iota,t} \otimes Z_t$ because $\widetilde{\eta}_{\iota,t}$ is a scalar, and the last equality holds because $S_\theta \eta_t = \eta_{-\iota,t} - \theta\eta_{\iota,t}$ and the fitted model (2.1) includes an intercept and thus the OLS residuals have sample mean equal to 0.

To study the asymptotic distribution of B, note that

$$\begin{aligned} B &= B_1 + B_2, \text{ where} \\ B_1 &= T^{-\frac{1}{2}} \sum_{t=1}^{T} (\mathbb{S}_\theta \eta_t) \otimes (Z_t - \mu_Z), \\ B_2 &= T^{-\frac{1}{2}} \sum_{t=1}^{T} [\mathbb{S}_\theta(\widetilde{\eta}_t - \eta_t)] \otimes (Z_t - \mu_Z). \end{aligned} \tag{A.16}$$

In B_1, note that $E\left[\mathbb{S}_\theta \eta_t \otimes (Z_t - \mu_Z)\right] = 0$ following the moment condition for the estimation of θ and $E[\eta_t] = 0$. Then it follows from Lemma 2(d) that

$$\begin{aligned} B_1 &= T^{-\frac{1}{2}} \sum_{t=1}^{T} \left\{(\mathbb{S}_\theta \eta_t) \otimes (Z_t - \mu_Z) - E\left[(\mathbb{S}_\theta \eta_t) \otimes (Z_t - \mu_Z)\right]\right\} \\ &= (\mathbb{S}_\theta \otimes I_k) \left\{ T^{-1/2} \sum_{t=1}^{T} \eta_t \otimes (Z_t - \mu_Z) - E\left[\eta_t \otimes (Z_t - \mu_Z)\right] \right\} \\ &\to_d (\mathbb{S}_\theta \otimes I_k)\, \xi_Z. \end{aligned} \tag{A.17}$$

Let X denote the matrix of x_t and η denote the vector of η_t. To study the distribution of B_2, note that

$$\begin{aligned} B_2 &= T^{-\frac{1}{2}} \sum_{t=1}^{T} [\mathbb{S}_\theta(\widetilde{\eta}_t - \eta_t)] \otimes (Z_t - \mu_Z) \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} -[\mathbb{S}_\theta \eta' X (X'X)^{-1} x_t] \otimes (Z_t - \mu_Z) \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} -[\mathbb{S}_\theta \eta' X \Upsilon_T^{-1} (\Upsilon_T^{-1} X' X \Upsilon_T^{-1})^{-1} \Upsilon_T^{-1} x_t] \otimes (Z_t - \mu_Z) \\ &= T^{-\frac{1}{2}} \sum_{t=1}^{T} -\mathrm{vec}((Z_t - \mu_Z) x_t' \Upsilon_T^{-1} (\Upsilon_T^{-1} X' X \Upsilon_T^{-1})^{-1} \Upsilon_T^{-1} X' \eta \mathbb{S}_\theta') \\ &= -\mathrm{vec}[(T^{-1}(Z - \ell_T \mu_Z')' X \Upsilon_T^{-1})(T^{-1} \Upsilon_T^{-1} X' X \Upsilon_T^{-1})^{-1} (T^{-1/2} \Upsilon_T^{-1} X' \eta \mathbb{S}_\theta')] \end{aligned} \tag{A.18}$$

where the second equality follows from the least square estimation of η_t, Υ_T in the third equality is defined in (3.9), the fourth equality uses $vec(AXB) = (B' \otimes A) vec(X)$ and ℓ_T

denotes the $(T-p)\times 1$ vector of ones. In the last equality above, the first term satisfies

$$\begin{aligned}
& T^{-1}(Z-\ell_T\mu_Z')'X\Upsilon_T^{-1} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.19)}\\
= & \left[T^{-\frac{3}{2}}\sum_{t=1}^{T}(Z_t-\mu_Z)Y_{1t-1}^{*\prime} : T^{-1}\sum_{t=1}^{T}(Z_t-\mu_Z) : T^{-1}\sum_{t=1}^{T}(Z_t-\mu_Z)D_t'\right]\\
\to_p & \left[0_{k\times(r_1+1)},\Gamma_{DZ}'\right],
\end{aligned}$$

by Lemma 2(b). Using the block-diagonality of $\mathbb{V}$ in Lemma 2(a), we can reduce (A.18) to

$$\begin{aligned}
& -vec[\Gamma_{DZ}'\cdot\Gamma_{DD}{}^{-1}\cdot(T^{-\frac{1}{2}}\sum_{t=1}^{T}D_t\eta_t')\mathbb{S}_\theta'] + o_p(1)\\
= & -\left(\mathbb{S}_\theta\otimes\left[\Gamma_{DZ}'\cdot\Gamma_{DD}{}^{-1}\right]\right)\mathrm{vec}[T^{-\frac{1}{2}}\sum_{t=1}^{T}D_t\eta_t'] + o_p(1)\\
= & -\left(\mathbb{S}_\theta\otimes\left[\Gamma_{DZ}'\cdot\Gamma_{DD}{}^{-1}\right]\right)T^{-\frac{1}{2}}\sum_{t=1}^{T}\eta_t\otimes D_t + o_p(1)\\
\to_d & -\left(\mathbb{S}_\theta\otimes\mathcal{K}\right)\xi_D, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.20)}
\end{aligned}$$

where both equalities use $\mathrm{vec}(AXB)=(B'\otimes A)\mathrm{vec}(X)$ and the convergence follows from Lemma 2(d).

To derive the limit of $\mathcal{A}_T$, we have

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\widetilde{\eta}_{\iota,t}Z_t' &= T^{-1}\sum_{t=1}^{T}\eta_{\iota,t}Z_t' + T^{-1}\sum_{t=1}^{T}\left(\widetilde{\eta}_{\iota,t}-\eta_{\iota,t}\right)Z_t'\\
&\to_p E(\eta_{\iota,t}Z_t'), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.21)}
\end{aligned}$$

where $T^{-1}\sum_{t=1}^{T}\left(\widetilde{\eta}_{\iota,t}-\eta_{\iota,t}\right)Z_t'\to_p 0$ follows from arguments similar to those in used to study B_2. Hence,

$$\mathcal{A}_T\to_p\mathcal{A}=I_{r-1}\otimes E(\eta_{\iota,t}Z_t'). \tag{A.22}$$

Under Assumption R, V is a full rank matrix and thus is invertible. The optimal choice of the weighting matrix follows standard arguments for GMM estimators. □

Proof of Theorem 2. Because the s-step-ahead VMA coefficient matrix is given by

$\Theta_s = M'F^sM$, we have

$$\begin{aligned}
\mathbf{d}\Theta_s &= \sum_{i=0}^{s-1} \mathbb{M}'\mathbb{F}^{s-1-i}\mathbf{d}\mathbb{F}\mathbb{F}^i\mathbb{M} \\
&= \sum_{i=0}^{s-1} \mathbb{M}'\mathbb{F}^{s-1-i}\mathbb{M}[\mathbf{d}\Phi_1 \ \cdots \ \mathbf{d}\Phi_p]\mathbb{F}^i\mathbb{M} \\
&= \sum_{i=0}^{s-1} \Theta_{s-1-i}[\mathbf{d}\Phi_1 \ \cdots \ \mathbf{d}\Phi_p]\mathbb{F}^i\mathbb{M} \\
&= \sum_{i=0}^{s-1} \Theta_{s-1-i}[\mathbf{d}d \ d\Phi_1 \ \cdots \ \mathbf{d}\Phi_p]\mathbb{S}_2'\mathbb{F}^i\mathbb{M} \qquad \text{(A.23)}
\end{aligned}$$

and

$$\operatorname{vec}(\mathbf{d}\Theta_s') = \sum_{i=0}^{s-1} \left[\Theta_{s-1-i} \otimes \left(\mathbb{M}'\mathbb{F}^{i\prime}\right)\right] \operatorname{vec}(\mathbb{S}_2[\mathbf{d}d \ \mathbf{d}\Phi_1 \ \cdots \ \mathbf{d}\Phi_p]'), \qquad \text{(A.24)}$$

where $\mathbb{S}_2 = [0_{rp\times 1} : I_{rp}]$ is the selector matrix that removes the first row.

Recall that $\mathbb{L}$ is defined in (3.15) such that

$$x_t = \begin{bmatrix} Y_{1,t-1} - c_1 \\ 1 \\ Y_{2,t-1} - c_2 - Q(Y_{1,t-1} - c_1) \\ Y_{3,t-1} - c_3 \\ \Delta Y_{t-1} \\ \vdots \\ \Delta Y_{t-p+1} \end{bmatrix} = \mathbb{L} \begin{bmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix} = \mathbb{L}X_t, \qquad \text{(A.25)}$$

which is a transformation between regressors in the model in (2.1) and the regressors in its equivalent representation in (3.3). For the OLS estimators, this implies that the OLS regression coefficients of model in (2.1) and the OLS coefficients of that in (3.3) satisfy

$$\left[\widetilde{d} : \widetilde{\Phi}_1 - I_r : \ \widetilde{\Phi}_2, \ \ldots, \ \widetilde{\Phi}_p\right] = \left[\widetilde{A}_1 : \ \widetilde{A}_2 : \ \widetilde{A}_3\right]\mathbb{L}. \qquad \text{(A.26)}$$

It follows from (A.24) and (A.26) that

$$\begin{aligned}
&T^{\frac{1}{2}}\operatorname{vec}(\widetilde{\Theta}_s' - \Theta_s') \\
= \ &\mathcal{R}\operatorname{vec}\left(\mathbb{S}_2\mathbb{L}'T^{\frac{1}{2}}(\sum_{t=p+1}^{T} x_t x_t')^{-1} \sum_{t=p+1}^{T} x_t\eta_t'\right) + o_p(1) \\
= \ &\mathcal{R}\operatorname{vec}\left(\mathbb{S}_2\mathbb{L}'\Upsilon_T^{-1}(T^{-1}\sum_{t=p+1}^{T} \Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}T^{-\frac{1}{2}}\sum_{t=p+1}^{T} x_t\eta_t'\right) + o_p(1). \quad \text{(A.27)}
\end{aligned}$$

Because $\mathbb{V}$ is block-diagonal that consists of the $(r_1+1)\times(r_1+1)$ upper-left submatrix and the $(rp-r_1)\times(rp-r_1)$ lower-right submatrix, $\Upsilon_T^{-1} = diag(0_{r_1\times r_1}, I_{rp-r_1+1}) + o(1)$, and the (r_1+1)st column of $S_2\mathbb{L}'$ consists of zeros. Thus,

$$\begin{aligned}
& \mathbb{S}_2\mathbb{L}'\Upsilon_T^{-1}(T^{-1}\sum_{t=1}^{T}\Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{T}x_t\eta_t' \\
= \quad & \overline{\mathbb{L}}\Gamma_{DD}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{T}D_t\eta_t' + o_p(1), \qquad \text{(A.28)}
\end{aligned}$$

where $\overline{\mathbb{L}}$ is the $rp\times(rp-r_1)$ lower-right submatrix of $\mathbb{L}'$ defined in (3.17) and it has full rank by construction.

Combining (A.27) and (A.28) yields

$$\begin{aligned}
T^{\frac{1}{2}}\mathrm{vec}(\widetilde{\Theta}_s' - \Theta_s') \quad &= \quad \mathcal{R}\left[I_r\otimes(\overline{\mathbb{L}}\Gamma_{DD}^{-1})\right]\mathrm{vec}\left(T^{-\frac{1}{2}}\sum_{t=p+1}^{T}D_t\eta_t'\right) + o_p(1) \\
&\to_d \mathcal{R}\left[I_r\otimes\mathcal{J}\right]\xi_D. \qquad \text{(A.29)}
\end{aligned}$$

□

Proof of Theorem 3. For $s\geq 1$,

$$\begin{aligned}
T^{\frac{1}{2}}(\widehat{\beta}_s - \beta_s) &= \widetilde{\Theta}_s T^{\frac{1}{2}}(\widehat{h} - h) + T^{\frac{1}{2}}(\widetilde{\Theta}_s - \Theta_s)h \\
&= \widetilde{\Theta}_s T^{\frac{1}{2}}(\widehat{h} - h) + (I_r\otimes h')T^{\frac{1}{2}}\mathrm{vec}(\widetilde{\Theta}_s' - \Theta_s'). \qquad \text{(A.30)}
\end{aligned}$$

The first term in (A.30) can be rewritten as

$$\begin{aligned}
\widetilde{\Theta}_s T^{\frac{1}{2}}(\widehat{h} - h) &= \Theta_s T^{\frac{1}{2}}\mathbb{S}_\iota\begin{bmatrix} 0 \\ \widehat{\theta} - \theta\end{bmatrix} + o_p(1) \\
&= \Theta_s\overline{\mathbb{S}}_\iota T^{\frac{1}{2}}(\widehat{\theta} - \theta) + o_p(1) \\
&\to_d \ \Theta_s\overline{\mathbb{S}}_\iota\left[\mathcal{A}W\mathcal{A}'\right]^{-1}\mathcal{A}W\cdot\left[-\mathbb{S}_\theta\otimes K : \mathbb{S}_\theta\otimes I_k\right]\begin{bmatrix}\xi_D \\ \xi_Z\end{bmatrix}, \qquad \text{(A.31)}
\end{aligned}$$

where $\overline{\mathbb{S}}_\iota$ denotes the last $r-1$ columns of S_ι, and the convergence follows from Theorem 1.

By Theorem 2, the second term in (A.30) can be rewritten as

$$(I_r\otimes h')T^{\frac{1}{2}}vec(\widetilde{\Theta}_s' - \Theta_s') \to_d \ (I_r\otimes h')\mathcal{R}\left[I_r\otimes\mathcal{J}\right]\xi_D. \qquad \text{(A.32)}$$

Thus, combining (A.31) and (A.32) yields Theorem 3.

Proof of Lemma 3. The estimator $\widehat{\mathcal{K}}$ satisfies

$$
\begin{aligned}
& \widehat{\mathcal{K}}\mathbb{L}^{-1}\Upsilon_T \\
= \;& \widehat{\Gamma}_{ZX}\widehat{\Gamma}_{XX}\mathbb{L}^{-1}\Upsilon_T \\
= \;& \left(\sum_{t=p+1}^{T}(Z_t-\overline{Z}_T)x_t'\Upsilon_T^{-1}\right)\left(\sum_{t=p+1}^{T}\Upsilon_T^{-1}x_tx_t'\Upsilon_T^{-1}\right)^{-1} \\
= \;& \left(T^{-1}\sum_{t=p+1}^{T}(Z_t-\mu_Z)x_t'\Upsilon_T^{-1}\right)\left(T^{-1}\sum_{t=p+1}^{T}\Upsilon_T^{-1}x_tx_t'\Upsilon_T^{-1}\right)^{-1} + o_p(1) \\
\to_p \;& [0_{k\times(r_1+1)} : \Gamma_{DZ}'\Gamma_{DD}^{-1}] \;=\; [0_{k\times(r_1+1)} : \mathcal{K}],
\end{aligned} \tag{A.33}
$$

where the first equality holds by definition, the second equality follows from $x_t = \mathbb{L}X_t$ by (A.25), the third equality uses

$$
\left[T^{-\frac{3}{2}}\sum_{t=1}^{T}(\overline{Z}_T-\mu_Z)Y_{1t-1}^{*\prime} : T^{-1}\sum_{t=1}^{T}(\overline{Z}_T-\mu_Z) : T^{-1}\sum_{t=1}^{T}(\overline{Z}_T-\mu_Z)D_t'\right] \;=\; o_p(1), \tag{A.34}
$$

which further follows from Lemma 2(b), (c), and the convergence in probability follows from Lemma 2(a) and (b), in particular the block diagonal structure of $\mathbb{V}$ in Lemma 2(a).

The estimator $\widehat{\mathcal{J}}$ satisfies that

$$
\begin{aligned}
& \widehat{\mathcal{J}}\mathbb{L}^{-1}\Upsilon_T \\
= \;& \mathbb{S}_2\widehat{\Gamma}_{XX}^{-1}\mathbb{L}^{-1}\Upsilon_T \\
= \;& \mathbb{S}_2\mathbb{L}'\Upsilon_T^{-1}\left(T^{-1}\sum_{t=1}^{T}\Upsilon_T^{-1}x_tx_t'\Upsilon_T^{-1}\right)^{-1} \\
\to_p \;& [0_{rp\times(r_1+1)} : \overline{\mathbb{L}}\Gamma_{DD}^{-1}] = [0_{rp\times(r_1+1)} : \mathcal{J}],
\end{aligned} \tag{A.35}
$$

where the first equality holds by definition, the second equality uses $x_t = \mathbb{L}X_t$, and the convergence in probability follows from Lemma 2(a) with $\mathbb{V}$ being block diagonal, the first r_1 diagonal elements of Υ_T^{-1} are $o(1)$, the $(r_1+1)^{th}$ column of $\mathbb{S}_2\mathbb{L}'$ is 0, and the remaining columns of $\mathbb{S}_2\mathbb{L}'$ are denoted by $\overline{\mathbb{L}}$ by definition.

Given the definition of $\mathbb{P}$, (A.33), (A.35), the consistency of $\widehat{\theta}$, and the continuous mapping theorem give

$$
\widehat{G}_{1s}\mathbb{P} \to_p \Theta_s\overline{S}_\iota(\mathcal{A}W\mathcal{A}')^{-1}\mathcal{A}W[-S_\theta\otimes[0_{k\times(r_1+1)}:\mathcal{K}] : S_\theta\otimes I_k], \tag{A.36}
$$

$$
\widehat{G}_{2s}\mathbb{P} \to_p \left[(I_r\otimes h')\mathcal{R}\left(I_r\otimes[0_{rp\times(r_1+1)}:\mathcal{J}]\right) : 0_{r\times rk}\right]. \tag{A.37}
$$

Proof of Lemma 4. Note that

$$\begin{aligned}
& \mathbb{P}^{-1}\widehat{\Omega}\mathbb{P}^{-1\prime} && \text{(A.38)} \\
= & \begin{bmatrix} \widehat{\Sigma}_\eta \otimes \left[T^{-1}\sum_{t=1}^T \Upsilon_T^{-1}x_t x_t' \Upsilon_T^{-1}\right] & \widehat{\Sigma}_\eta \otimes \left[T^{-1}\sum_{t=1}^T \Upsilon_T^{-1}x_t (Z_t-\overline{Z}_T)'\right] \\ \widehat{\Sigma}_\eta \otimes \left[T^{-1}\sum_{t=1}^T (Z_t-\overline{Z}_T) x_t' \Upsilon_T^{-1}\right] & \widehat{\Sigma}_\eta \otimes \left[T^{-1}\sum_{t=1}^T (Z_t-\overline{Z}_T)(Z_t-\overline{Z}_T)'\right] - \widehat{\gamma}\widehat{\gamma}' \end{bmatrix},
\end{aligned}$$

where all X_t is transformed to $\Upsilon_T^{-1}x_t$ and Z_t. Let $\widehat{\beta}_T = (\sum_{t=1}^T x_t x_t')^{-1}\sum_{t=1}^T x_t \Delta Y_t'$ denote the OLS regression coefficients that yields the residual $\widetilde{\eta}_t$. Note that

$$\Delta = -\Upsilon_T(\widehat{\beta}_T - \beta) = -\left(T^{-1}\sum_{t=1}^T \Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1}\right)^{-1} T^{-1}\sum_{t=1}^T \Upsilon_T^{-1}x_t\eta_t' = O_p(T^{-\frac{1}{2}}) \quad \text{(A.39)}$$

by Lemma 2(a),(c),(d). To investigate $\widehat{\Sigma}_\eta$, note that

$$\begin{aligned}
\widetilde{\eta}_t\widetilde{\eta}_t' &= (\eta_t - (\widehat{\beta}_T-\beta)'x_t)(\eta_t - (\widehat{\beta}_T-\beta)'x_t)' \\
&= \eta_t\eta_t' - \Delta'\Upsilon_T^{-1}x_t\eta_t' - \eta_t x_t'\Upsilon_T^{-1}\Delta' + \Delta'\Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1}\Delta'. \quad \text{(A.40)}
\end{aligned}$$

Applying (A.40) to $\widehat{\Sigma}_\eta$, we obtain

$$\begin{aligned}
\widehat{\Sigma}_\eta &= \frac{1}{T}\sum_{t=1}^T \eta_t\eta_t' - \Delta'\frac{1}{T}\sum_{t=1}^T \Upsilon_T^{-1}x_t\eta_t' - \frac{1}{T}\sum_{t=1}^T \eta_t x_t'\Upsilon_T^{-1}\Delta' - \Delta'\frac{1}{T}\sum_{t=1}^T \Upsilon_T^{-1}x_t x_t'\Upsilon_T^{-1}\Delta \\
&\to_p \Sigma_\eta \quad \text{(A.41)}
\end{aligned}$$

by Lemma 2(a),(c),(d) and (A.39).

To study the upper-right submatrix of (A.38), note that

$$\frac{1}{T}\sum_{t=1}^T \Upsilon_T^{-1}x_t(Z_t-\overline{Z}_T)' \to_p \left[0_{k\times(r_1+1)}, \Gamma_{DZ}'\right]' \quad \text{(A.42)}$$

following (A.19) and Lemma 2.

Finally, the first moment satisfies

$$\begin{aligned}
\widehat{\gamma} - \gamma &= \frac{1}{T}\sum_{t=1}^T \widetilde{\eta}_t \otimes Z_t - \gamma \\
&= \left(\frac{1}{T}\sum_{t=1}^T \eta_t \otimes Z_t - \gamma\right) - \Delta'\left(\frac{1}{T}\sum_{t=1}^T \Upsilon_T^{-1}x_t \otimes Z_t\right) \\
&= O_p(T^{-\frac{1}{2}}) \quad \text{(A.43)}
\end{aligned}$$

by Lemma 2(d) and (A.39).

The desirable result follow from (A.38), (A.41), (A.42), (A.43), and Lemma 2(a), (b). □

Proof of Theorem 4. Combining Lemma 3 and Lemma 4, the zero matrices in the limit of $(\widehat{G}_{1s}+\widehat{G}_{2s})\mathbb{P}$ and its transpose reduce $\mathbb{V}$ and γ_{xZ} in the limit of $\mathbb{P}^{-1}\widehat{\Omega}\mathbb{P}^{-1\prime}$ to their submatrices Γ_{DD} and Γ_{DZ}, respectively. This removes all nondeterministic elements in $\mathbb{V}$ and γ_{xZ} and the consistency result follows immediately. □

Proof of Theorem 5. Part (a) follows from the arguments used for Theorem 4. Parts (b) and (c) follow from Theorems 1 and 3. □

References

Andrews, D. W. K., P. Guggenberger (2010) Asymptotic size and a problem with subsampling and with the m out of n boostrap. *Econometric Theory* 26, 426–468.

Dolado, J. J., H. Lütkepohl (1996) Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* 15, 369–386.

Elliott, G. (1998) On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, 66, 149–158.

Engle, R. F., W. J. Granger (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251–276.

Gertler, M. and P. Karadi (2015) Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics* 7, 44–76.

Gospodinov, N. (2004) Asymptotic confidence intervals for impulse responses of near-integrated processes. *Econometrics Journal* 7, 505–527.

Inoue, A., and L. Kilian (2002) Bootstrapping autoregressive processes with possible unit roots. *Econometrica* 70, 377–391.

Inoue, A., and L. Kilian (2016) Joint confidence sets for structural impulse responses. *Journal of Econometrics* 192, 421–432.

Inoue, A., and L. Kilian (2019) The uniform validity of impulse response inference in autoregressions. Working paper, Vanderbilt University and Federal Reserve Bank of Dallas.

Kilian, L., (1999) Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. *Review of Economics and Statistics* 81, 652–660.

Leeb, H., B. M. Pötscher (2005) Selection and inference: facts and fiction. *Econometric Theory* 21, 21–59.

Mertens, K. and M. O. Ravn (2013) The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review* 103, 1212–1247.

Mikusheva, A. (2007) Uniform inference in autoregressive models. *Econometrica* 75, 1411–1452.

Mikusheva, A. (2012) One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica* 80, 173–212.

Montiel Olea, J.L., and M. Plagborg-Møller (2019) Simultaneous confidence bands: theory, implementation, and an application to SVARs. *Journal of Applied Econometrics* 34, 1–17.

Montiel Olea, J.L., J.H. Stock and M.W. Watson (2018) Inference in structural vector autoregressions identified with an external instrument. Working paper, Columbia University, Harvard University and Princeton University.

Park, J.Y., and P.C.B. Phillips (1989a) Statistical inference in regressions with integrated processes: part 1. *Econometric Theory* 4, 468–497.

Park, J.Y., and P.C.B. Phillips (1989b) Statistical inference in regressions with integrated processes: part 2. *Econometric Theory* 5, 95–131.

Pesavento, E., and B. Rossi (2007) Impulse response confidence intervals for persistent data: what have we learned? *Journal of Economic Dynamics and Control* 31, 2398–2412.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.

Phillips, P.C.B. (1998) Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics* 83, 21–56.

Phillips, P.C.B. (2014) On confidence interval for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.

Sims, C.A., J.H. Stock and M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.

Stock, J.H., and M.W. Watson (2012) Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity* Spring 2012, 81–135.

Stock, J.H., and M.W. Watson (2018) Identification and estimation of dynamic causal effects in macroeconomics. *Economic Journal* 128, 917–948.

Toda, H.Y., and Phillips, P.C.B. (1993) The spurious effect of unit roots on vector autoregressions. *Journal of Econometrics* 59, 229–255.

Toda, H.Y., and T. Yamamoto (1995) Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66, 225–250.

Wright, J. (2000) Confidence intervals for univariate impulse responses with a near unit root. *Journal of Business & Economic Statistics* 18, 368–373.

Table 1A: Coverage rates of the 95% level confidence intervals under DGP1

A shock in ε_{1t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_1	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	-	93.2	91.9	90.7	86.2	81.2	75.3	-	95.1	94.0	93.8	91.9	90.2	87.5
0	94.0	94.5	93.4	91.9	87.0	81.5	75.6	94.2	95.0	94.8	94.4	92.4	90.5	87.8
	94.7	94.4	94.0	94.0	88.0	82.3	79.1	95.7	94.7	94.8	94.9	92.0	88.3	83.1
	-	94.1	92.9	91.0	86.8	81.6	76.7	-	94.7	94.0	93.5	91.4	89.2	86.6
-2	93.4	94.3	93.8	92.4	87.7	81.9	76.9	94.8	94.9	94.7	94.2	91.9	89.6	86.9
	94.4	94.5	94.1	94.2	89.3	86.2	84.6	94.6	95.0	94.6	94.4	91.4	88.8	86.3
	-	94.2	92.8	91.8	88.7	85.5	81.3	-	94.2	94.1	93.7	92.4	90.7	88.7
-5	94.5	94.4	93.9	92.6	89.4	85.7	81.5	94.5	94.9	94.7	94.1	92.5	90.9	88.9
	94.7	94.2	94.6	94.2	90.6	89.0	89.4	95.6	95.0	94.9	94.9	92.7	91.1	89.6
	-	93.9	93.0	92.7	90.1	87.1	83.9	-	95.0	94.7	94.0	92.8	91.1	89.0
-10	93.6	94.2	94.0	93.7	90.6	87.6	84.4	94.2	95.0	95.3	94.3	93.2	91.2	89.2
	94.7	94.1	94.5	93.9	92.2	91.9	91.9	95.2	94.5	94.6	94.5	92.8	91.7	91.5

Table 1B: Coverage rates of the 95% level confidence intervals under DGP1

A shock in ε_{2t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_1	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	94.4	94.1	92.7	91.2	87.2	81.7	76.2	94.1	94.4	94.2	93.9	92.1	90.0	86.8
0	-	93.4	91.8	90.0	86.5	81.3	76.2	-	94.0	93.7	93.1	91.7	89.7	86.6
	94.4	94.3	94.1	93.2	86.8	81.9	78.7	95.6	94.9	95.1	94.2	91.5	87.5	82.6
	93.9	93.8	93.0	91.4	87.9	82.9	78.0	94.8	94.9	93.9	93.3	91.7	89.6	87.2
-2	-	93.6	92.1	90.6	87.5	82.7	77.9	-	94.6	93.6	93.1	91.3	89.5	87.3
	94.1	94.8	94.6	93.9	88.5	85.8	85.2	95.5	95.2	95.5	94.7	91.4	88.4	85.4
	93.9	93.9	93.3	91.6	88.0	84.4	80.6	94.8	94.8	94.2	94.4	93.0	91.6	89.7
-5	-	93.6	92.1	90.6	87.4	84.3	80.5	-	94.9	94.3	94.0	92.6	91.4	89.6
	94.4	94.2	94.4	93.5	89.5	88.7	88.7	94.5	94.2	94.9	94.7	92.9	91.6	89.9
	94.6	94.3	93.5	92.5	90.0	87.6	84.9	94.3	94.5	94.3	94.4	93.3	92.0	90.0
-10	-	94.2	93.2	91.5	89.4	86.7	84.5	-	94.6	94.1	94.1	93.2	91.9	89.9
	94.6	94.3	94.6	93.4	91.7	91.5	92.2	94.6	94.5	95.2	94.8	92.9	92.2	91.7

Table 1C: Coverage rates of the 95% level confidence intervals under DGP1

A shock in ε_{3t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_1	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	94.3	93.7	93.6	93.1	92.2	88.6	84.6	94.6	95.2	95.4	95.0	94.4	93.3	92.0
0	93.7	93.5	93.1	93.0	91.9	88.5	84.8	94.7	95.1	95.6	95.1	94.4	93.3	92.0
	-	94.0	92.7	92.1	88.7	78.6	76.6	-	94.6	94.2	93.6	91.5	85.1	79.1
	94.5	93.5	93.8	93.7	92.1	89.7	86.1	94.4	94.7	95.0	94.9	94.5	93.2	91.7
-2	93.9	93.8	93.8	93.7	92.0	89.7	86.0	94.3	94.3	94.6	94.8	94.6	93.2	91.6
	-	93.1	92.6	92.6	90.8	81.9	82.7	-	94.7	93.8	93.8	91.6	85.6	81.3
	94.0	93.8	94.0	93.3	92.3	89.7	86.3	94.9	94.7	94.7	94.4	94.3	94.0	92.8
-5	93.7	94.0	93.6	93.4	92.6	89.7	86.1	94.7	94.4	94.5	94.1	94.3	94.0	92.9
	-	93.5	93.1	92.7	91.2	84.6	87.2	-	94.6	93.5	93.4	93.0	88.6	85.6
	94.1	93.7	94.1	93.9	92.9	90.9	87.4	94.8	94.6	94.4	94.6	94.5	93.6	92.3
-10	94.6	94.0	94.2	94.3	92.9	90.8	87.5	95.1	94.6	94.3	94.7	94.7	93.6	92.3
	-	93.7	92.8	92.6	91.2	87.2	90.1	-	93.9	94.7	94.1	92.1	87.7	86.8

Table 2A: Coverage rates of the 95% level confidence intervals under DGP2

A shock in ε_{1t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_2	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	-	90.9	86.0	82.1	69.5	58.8	51.9	-	93.2	91.0	89.3	82.9	76.3	70.0
-2	94.5	93.4	93.5	93.1	91.8	89.7	86.8	94.0	94.1	93.8	93.8	93.4	93.1	92.3
	94.3	94.3	93.9	91.7	71.5	71.1	73.6	94.5	94.9	94.3	93.6	80.2	73.9	72.8
	-	91.4	87.0	82.9	71.8	61.7	54.2	-	93.2	91.7	90.2	84.8	78.1	71.4
-5	94.0	93.6	93.3	93.3	92.8	90.3	87.7	94.6	94.0	93.8	93.9	93.8	93.1	92.4
	94.1	94.1	93.3	91.9	72.0	72.5	75.1	94.4	94.4	94.0	93.5	81.6	77.4	76.0
	-	91.6	87.8	84.2	73.7	64.8	57.0	-	93.6	92.4	91.0	85.4	78.9	72.7
-10	95.1	94.7	94.4	94.0	93.3	91.9	89.9	94.7	94.7	94.6	94.6	94.4	94.1	93.3
	94.6	93.9	93.3	92.2	73.9	73.5	76.1	95.0	94.6	94.9	93.6	81.5	76.5	75.5
	-	91.7	88.8	86.0	76.9	68.1	60.9	-	93.7	92.5	91.2	86.1	80.8	76.4
$-0.5T$	94.5	94.3	93.9	94.1	94.4	94.1	93.2	94.9	94.5	94.5	94.9	94.6	95.0	94.6
	94.3	94.0	94.2	92.1	77.7	77.2	77.6	94.6	94.5	94.4	93.8	83.7	79.9	78.5

Table 2B: Coverage rates of the 95% level confidence intervals under DGP2

A shock in ε_{2t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_2	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	94.1	94.0	93.2	92.9	91.1	90.0	88.0	94.3	94.4	93.7	93.7	93.4	92.9	91.8
-2	-	91.3	87.2	84.0	74.8	65.7	58.2	-	93.1	91.3	90.2	85.8	80.1	73.9
	94.0	93.8	93.0	93.2	92.8	90.5	90.5	95.2	94.4	94.6	94.7	94.6	93.4	91.6
	93.9	94.1	93.7	92.9	91.8	89.8	88.1	94.5	94.5	94.9	94.4	93.8	93.3	92.1
-5	-	91.9	89.0	86.7	78.2	70.1	63.6	-	94.4	92.9	91.8	87.7	82.6	77.2
	94.3	93.4	93.4	93.4	92.9	91.3	91.3	94.7	94.7	94.4	94.6	94.4	92.8	91.6
	94.4	94.0	93.4	92.9	91.9	90.2	88.8	94.8	95.1	94.4	94.1	93.9	93.3	92.5
-10	-	92.5	90.1	88.0	79.5	72.6	68.7	-	94.4	93.7	92.4	88.8	84.5	80.0
	94.4	93.7	94.0	94.2	93.4	91.7	92.5	95.1	94.3	94.4	94.2	94.3	93.2	92.1
	94.5	93.9	93.7	93.6	93.3	93.3	93.8	94.9	94.5	95.0	95.1	94.7	94.0	94.0
$-0.5T$	-	93.6	91.4	88.1	86.9	85.0	82.4	-	93.4	93.7	91.8	91.3	90.6	88.0
	94.4	94.6	93.9	93.9	94.1	96.8	99.4	94.3	94.8	94.8	94.6	94.4	94.9	97.2

Table 2C: Coverage rates of the 95% level confidence intervals under DGP2

A shock in ε_{3t}

	IRFs at horizon s for $T = 200$							IRFs at horizon s for $T = 500$						
C_2	0	1	2	3	6	9	12	0	1	2	3	6	9	12
	94.4	93.6	93.5	93.0	89.8	85.4	79.7	94.5	94.5	94.8	94.7	93.4	91.0	88.8
-2	93.8	94.0	94.0	93.2	93.3	93.4	93.9	95.3	95.0	94.7	94.5	94.2	94.0	94.0
	-	92.1	91.2	89.9	85.9	75.2	81.1	-	94.3	93.3	92.7	89.1	80.2	74.3
	94.0	93.8	93.4	93.0	89.7	85.4	79.9	94.7	94.8	94.5	94.3	93.7	92.3	89.9
-5	94.4	94.0	93.8	93.5	94.0	94.4	95.3	94.3	94.3	94.2	94.4	94.6	94.8	94.8
	-	92.8	91.4	90.4	85.7	75.3	82.7	-	94.5	93.1	92.3	90.5	82.5	77.7
	93.8	93.6	93.6	93.3	90.5	86.7	82.0	94.8	94.7	94.1	94.3	93.8	92.1	90.4
-10	93.8	93.8	93.5	93.8	93.9	94.4	95.4	94.7	94.7	94.5	94.6	94.6	94.8	94.9
	-	92.9	91.5	90.3	87.1	77.8	83.8	-	93.9	94.3	93.4	89.5	81.5	76.8
	94.5	93.6	93.3	93.4	92.0	89.1	85.0	94.6	94.6	94.4	94.3	93.7	92.4	90.9
$-0.5T$	94.8	94.5	94.4	93.0	93.6	97.8	99.1	94.6	94.7	94.5	93.8	94.6	96.1	98.1
	-	92.9	93.1	92.7	89.2	81.1	82.1	-	94.4	93.7	93.5	91.8	85.5	80.2