

# Online Appendix for: Instrumental Variable Identification of Dynamic Variance Decompositions

Mikkel Plagborg-Møller      Christian K. Wolf

July 11, 2021

This online appendix contains supplemental material for the article “Instrumental Variable Identification of Dynamic Variance Decompositions”. We provide (i) bounds on other notions of variance decompositions, (ii) extensions of the identification analysis to multiple instruments correlated with a single or multiple shocks, (iii) characterizations of the bias of SVAR-IV (or “proxy SVAR”) procedures under noninvertibility, (iv) an illustration of our method using a quantitative structural macro model, (v) supplementary results for the monetary shock application, (vi) a second set of empirical results on the importance of oil news shocks, and (vii) asymptotic theory on the nonparametric validity of our sieve VAR inference strategy. The end of this appendix contains proofs and auxiliary lemmas.

**Any references to equations, figures, tables, assumptions, propositions, lemmas, or sections that are not preceded by “B.” refer to the main article.**

## Contents

<b>B.1 Identification and estimation of other variance decomposition concepts</b>	<b>3</b>
<b>B.2 Multiple instruments correlated with one shock</b>	<b>7</b>
<b>B.3 Instruments correlated with multiple shocks</b>	<b>9</b>
<b>B.4 Invertibility and SVAR-IV</b>	<b>13</b>
<b>B.5 Illustration in a structural macro model</b>	<b>15</b>

B.5.1	Preliminaries . . . . .	15
B.5.2	Information content of several observables . . . . .	16
B.5.3	Dynamic information content . . . . .	17
B.5.4	Non-invertibility and news shocks . . . . .	18
B.5.5	Other observables . . . . .	20
<b>B.6</b>	<b>Monetary policy shock application: further results</b>	<b>23</b>
<b>B.7</b>	<b>Oil news shock application</b>	<b>24</b>
<b>B.8</b>	<b>Nonparametric sieve VAR inference</b>	<b>30</b>
B.8.1	Assumptions, parameters of interest, and estimator . . . . .	30
B.8.2	Main convergence results . . . . .	33
<b>B.9</b>	<b>Additional proofs and auxiliary lemmas</b>	<b>36</b>
B.9.1	Proof of Lemma 1 . . . . .	36
B.9.2	Auxiliary lemma for proof of Proposition B.1 . . . . .	36
B.9.3	Proof of Proposition B.1 . . . . .	37
B.9.4	Proof of Proposition B.2 . . . . .	38
B.9.5	Proof of Proposition B.3 . . . . .	39
B.9.6	Auxiliary lemmas for sieve VAR results . . . . .	40
B.9.7	Proof of Lemma B.3 . . . . .	42
B.9.8	Proof of Lemma B.4 . . . . .	42
B.9.9	Proof of Lemma B.6 . . . . .	44
B.9.10	Proof of Lemma B.7 . . . . .	45
B.9.11	Proof of Lemma B.8 . . . . .	47
B.9.12	Proof of Lemma B.9 . . . . .	49
B.9.13	Proof of Lemma B.10 . . . . .	49
B.9.14	Proof of Lemma B.11 . . . . .	50
B.9.15	Proof of Proposition B.4 . . . . .	53
B.9.16	Proof of Proposition B.5 . . . . .	53
<b>References</b>		<b>57</b>

## B.1 Identification and estimation of other variance decomposition concepts

Our main analysis focuses on forecast variance ratios as a measure of shock importance, defined in [Section 2.2](#). This appendix defines two additional concepts – forecast variance decompositions (FVD) and unconditional frequency-specific variance decompositions (VD) – and discusses the identification and estimation of both.

DEFINITIONS. The *forecast variance decomposition* (FVD) for variable  $i$  at horizon  $\ell$  is defined as

$$FVD_{i,\ell} \equiv 1 - \frac{\text{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty < \tau \leq t}, \{\varepsilon_{1,\tau}\}_{t < \tau < \infty})}{\text{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty < \tau \leq t})} = \frac{\sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2}{\sum_{j=1}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2}. \quad (\text{B.1})$$

The FVD measures the reduction in forecast variance that arises from learning the path of future realizations of the shock of interest, supposing that we already had the history of past structural shocks  $\varepsilon_t$  available when forming our forecast. Because the econometrician generally does not observe the structural shocks directly, the FVD is best thought of as reflecting forecasts of economic agents who observe the underlying shocks. The FVD always lies between 0 and 1, purely reflects fundamental forecasting uncertainty, and equals 1 if the first shock is the only shock driving variable  $i$  in equation (1). The software package Dynare reports FVDs after having estimated a DSGE model.

While the FVR and FVD concepts generally differ, they coincide in the case where all shocks are invertible, since in that case the information set  $\{y_\tau\}_{-\infty < \tau \leq t}$  equals the information set  $\{\varepsilon_\tau\}_{-\infty < \tau \leq t}$ . This explains why the SVAR literature has not made the distinction between the two concepts.<sup>B.1</sup>

Our second additional concept is the frequency-specific *unconditional* variance decomposition (VD) of [Forni et al. \(2019, Sec. 3.4\)](#). The VD for variable  $i$  over the frequency band  $[\omega_1, \omega_2]$  is given by

$$VD_i(\omega_1, \omega_2) \equiv \frac{\int_{\omega_1}^{\omega_2} |\Theta_{i,1}(e^{-i\omega})|^2 d\omega}{\sum_{j=1}^{n_\varepsilon} \int_{\omega_1}^{\omega_2} |\Theta_{i,j}(e^{-i\omega})|^2 d\omega}, \quad 0 \leq \omega_1 < \omega_2 \leq \pi, \quad (\text{B.2})$$

where  $\Theta_{i,j}(L)$  is the  $(i, j)$  element of the lag polynomial  $\Theta(L)$ .  $VD_i(\omega_1, \omega_2)$  is the percentage

---

<sup>B.1</sup>[Forni et al. \(2019\)](#) point out the bias caused by noninvertibility when estimating the FVD using SVARs.

reduction in the variance of  $y_{i,t}$  – after passing the data through a bandpass filter that retains only cyclical frequencies  $[\omega_1, \omega_2]$  – caused by entirely “shutting off” the shock of interest  $\varepsilon_{1,t}$ . The software package Dynare automatically reports  $VD_i(0, \pi)$  after solving a DSGE model.

**IDENTIFICATION AND ESTIMATION: VD.** Identification of the VD is completely analogous to our analysis of the FVR. By definition,

$$VD_i(\omega_1, \omega_2) = \frac{1}{\alpha^2} \times \frac{\int_{\omega_1}^{\omega_2} |s_{y_i\bar{z}}(\omega)|^2 d\omega}{\int_{\omega_1}^{\omega_2} s_{y_i}(\omega) d\omega},$$

where  $s_{y_i\bar{z}}(\omega) = \alpha\Theta_{i,1}(e^{-i\omega})$  is the  $i$ -th element of  $s_{y\bar{z}}(\omega)$ , cf. equation (B.2). Since the last fraction on the right-hand side is point-identified, our identified set for  $\alpha^2$  immediately maps into an identified for the VD.

We estimate the bounds as

$$\left[ \frac{1}{\hat{\alpha}^2} \times \frac{\int_{\omega_1}^{\omega_2} |\hat{s}_{y_i\bar{z}}(\omega)|^2 d\omega}{\int_{\omega_1}^{\omega_2} \hat{s}_{y_i}(\omega) d\omega}, \frac{1}{\underline{\alpha}^2} \times \frac{\int_{\omega_1}^{\omega_2} |\hat{s}_{y_i\bar{z}}(\omega)|^2 d\omega}{\int_{\omega_1}^{\omega_2} \hat{s}_{y_i}(\omega) d\omega} \right]. \quad (\text{B.3})$$

The integrals are computed numerically. The spectral densities required to compute (B.3) are functions of the estimated reduced-form VAR parameters (see Appendix A.1). Specifically,

$$\begin{aligned} \hat{s}_y(\omega) &= \frac{1}{2\pi} \hat{B}_y(e^{-i\omega}) \hat{B}_y(e^{-i\omega})^*, \\ \hat{s}_{y\bar{z}}(\omega) &= \frac{1}{2\pi} \sum_{\ell=0}^{\infty} \hat{\Sigma}_{y,\bar{z},\ell} e^{-i\omega\ell}, \end{aligned}$$

with

$$\hat{B}_y(e^{-i\omega}) \equiv \sum_{\ell=0}^{\infty} \hat{B}_{y,\ell} e^{-i\omega\ell}, \quad \hat{\Sigma}_{y,\bar{z},\ell} \equiv \widehat{\text{Cov}}(y_t, \bar{z}_{t-\ell}) = \hat{B}_{y,\ell} \hat{B}'_{\bar{z}}.$$

In practice, we truncate the infinite sums at a large lag.

**IDENTIFICATION AND ESTIMATION: FVD.** Bounding the FVD requires more work. Intuitively, the reason that identification of the FVD is more challenging than for the FVR is that, even if we knew  $\alpha$ , the IV  $z_t$  provides no information about the other structural shocks  $\varepsilon_{j,t}$ ,  $j \neq 1$ . This matters because the definition (B.1) of the FVD, unlike that of the FVR, conditions on knowing all past shocks, rather than all past macro observables. **Proposition B.1** formally characterizes the resulting identified set.

**Proposition B.1.** *Let there be given a joint spectral density for  $w_t = (y'_t, \tilde{z}'_t)'$  satisfying the assumptions in [Proposition 1](#). Given knowledge of  $\alpha \in (\alpha_{LB}, \alpha_{UB}]$ , the largest possible value of the forecast variance decomposition  $FVD_{i,\ell}$  is 1 (the trivial bound), while the smallest possible value is given by*

$$\frac{\sum_{m=0}^{\ell-1} \text{Cov}(y_{i,t}, \tilde{z}_{t-m})^2}{\sum_{m=0}^{\ell-1} \text{Cov}(y_{i,t}, \tilde{z}_{t-m})^2 + \alpha^2 \text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty < \tau \leq t})}. \quad (\text{B.4})$$

Here  $\tilde{y}_t^{(\alpha)} = (\tilde{y}_{1,t}^{(\alpha)}, \dots, \tilde{y}_{n_y,t}^{(\alpha)})'$  denotes a stationary Gaussian time series with spectral density  $s_{\tilde{y}^{(\alpha)}}(\omega) = s_y(\omega) - \frac{2\pi}{\alpha^2} s_{y\tilde{z}}(\omega) s_{y\tilde{z}}(\omega)^*$ ,  $\omega \in [0, 2\pi]$ . Expression (B.4) is monotonically decreasing in  $\alpha$ , so the overall lower bound on  $FVD_{i,\ell}$  is attained by  $\alpha = \alpha_{UB}$ ; in this boundary case we can represent  $\tilde{y}_t^{(\alpha_{UB})} = y_t - E(y_t \mid \{\tilde{z}_\tau\}_{-\infty < \tau \leq t})$ .

The upper bound on the FVD always equals the trivial bound of 1, for any  $\ell \geq 1$ . This upper bound is achieved by a model in which all shocks, except the first one, only affect  $y_t$  after an  $\ell$ -period delay. The lower bound in contrast is nontrivial and informative. The argument is as follows: Even if  $\alpha$  is known, the denominator  $\text{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty < \tau \leq t})$  of the FVD is not identified due to the lack of information about shocks other than the first. Although we can upper-bound this conditional variance by the denominator of the FVR, this upper bound is not sharp. Instead, to maximize the denominator, as much forecasting noise as possible should be of the pure forecasting variety, and not related to noninvertibility. For all shocks except for  $\varepsilon_{1,t}$ , this is achievable through a Wold decomposition construction ([Hannan, 1970](#), Thm. 2'', p. 158). Given  $\alpha$ , we know the contribution of the first shock to  $y_t$ ; the residual after removing this contribution has the distribution of  $\tilde{y}_t^{(\alpha)}$ , as defined in the proposition. If  $\alpha$  is not known, the smallest possible value of the lower bound (B.4) is attained at the largest possible value of  $\alpha$ , namely  $\alpha_{UB}$ , for which  $\varepsilon_{1,t}$  contributes the least to forecasts of  $y_t$ .

We estimate the bounds in [Proposition B.1](#) as

$$\left[ \frac{\sum_{m=0}^{\ell-1} \widehat{\text{Cov}}(y_{i,t}, \tilde{z}_{t-m})^2}{\sum_{m=0}^{\ell-1} \widehat{\text{Cov}}(y_{i,t}, \tilde{z}_{t-m})^2 + \hat{\alpha}^2 \widehat{\text{Var}}(\tilde{y}_{i,t+\ell}^{(\hat{\alpha})} \mid \{\tilde{y}_\tau^{(\hat{\alpha})}\}_{-\infty < \tau \leq t})}, 1 \right]. \quad (\text{B.5})$$

To approximate the conditional variance in the denominator, we proceed as in [Appendix A.1](#). First, we replace the infinite conditioning set with the finite set  $\{\tilde{y}_\tau^{(\hat{\alpha})}\}_{\tau-M \leq \tau \leq t}$ . Second, we compute the conditional variance using the standard projection formula, where the autocor-

variances of the process  $\{\tilde{y}_\tau^{(\bar{\alpha})}\}$  are estimated as

$$\widehat{\text{Cov}}(\tilde{y}_{t+\ell}^{(\bar{\alpha})}, \tilde{y}_t^{(\bar{\alpha})}) = \widehat{\text{Cov}}(y_{t+\ell}, y_t) - \frac{1}{\hat{\alpha}^2} \sum_{m=0}^{\infty} \widehat{\text{Cov}}(y_t, \tilde{z}_{t-m-\ell}) \widehat{\text{Cov}}(y_t, \tilde{z}_{t-m})'.$$

In practice, we truncate the infinite sum at a large lag.

## B.2 Multiple instruments correlated with one shock

Here we show that the multiple-IV model in [Assumptions 1 to 3](#) is testable, but if it is consistent with the data, then identification analysis can be reduced to the single-IV case.

Define the IV residual vector  $\tilde{z}_t$  as in equation (18). The multiple-IV model in [Assumptions 1 and 2](#) implies the following cross-spectrum between  $y_t$  and  $\tilde{z}_t$ :

$$s_{y\tilde{z}}(\omega) = \frac{\alpha}{2\pi} \Theta(e^{-i\omega}) e_1 \lambda', \quad \omega \in [0, 2\pi]. \quad (\text{B.6})$$

Thus, the cross-spectrum has rank-1 factor structure: It equals a nonconstant column vector times a constant row vector. This testable property turns out to be exactly what characterizes the multiple-IV model.

**Proposition B.2.** *Let a spectrum  $s_w(\omega)$  for  $w_t = (y'_t, \tilde{z}'_t)'$  be given, satisfying the assumptions of [Proposition 1](#). There exists a model of the form in [Assumptions 1 and 2](#) which generates the spectrum  $s_w(\omega)$  if and only if there exist  $n_y$ -dimensional real vectors  $\zeta_\ell$ ,  $\ell \geq 0$ , and an  $n_z$ -dimensional constant real vector  $\eta$  of unit length such that*

$$s_{y\tilde{z}}(\omega) = \zeta(e^{-i\omega}) \eta', \quad \omega \in [0, 2\pi], \quad (\text{B.7})$$

where  $\zeta(L) = \sum_{\ell=0}^{\infty} \zeta_\ell L^\ell$ .

Assuming henceforth that the factor structure obtains, we now show that identification in the multiple-IV model reduces to the single-IV case. It is convenient first to reparametrize the model slightly, by setting  $\Sigma_v = \Sigma_{\tilde{z}} - \alpha^2 \lambda \lambda'$  and treating  $\Sigma_{\tilde{z}}$  as a basic model parameter instead of  $\Sigma_v$ . We then impose the requirement that  $\Sigma_{\tilde{z}} - \alpha^2 \lambda \lambda'$  be positive semidefinite. Clearly,  $\Sigma_{\tilde{z}} = \text{Var}(\tilde{z}_t)$  is point-identified. Next, note from (B.6) that  $\lambda$  is point-identified and equal to the  $\eta$  vector in equation (B.7). This is because any rank-1 factorization of a matrix is identified up to sign and scale, and we have normalized  $\eta$  to have length 1. Let  $\Xi$  be any  $(n_z - 1) \times n_z$  matrix such that  $\Xi \Sigma_{\tilde{z}}^{-1/2} \lambda = 0$ . Define the  $n_z \times n_z$  matrix

$$Q \equiv \begin{pmatrix} \frac{1}{\lambda' \Sigma_{\tilde{z}}^{-1} \lambda} \lambda' \Sigma_{\tilde{z}}^{-1} \\ \Xi \Sigma_{\tilde{z}}^{-1/2} \end{pmatrix}.$$

Since  $Q$  is point-identified (given a choice of  $\Xi$ ), it is without loss of generality to perform

identification analysis based on the linearly transformed IV residuals

$$Q\tilde{z}_t = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_{1,t} + \tilde{v}_t, \quad \tilde{v}_t \sim N \left( 0, \begin{pmatrix} \frac{1}{\lambda' \Sigma_{\tilde{z}}^{-1} \lambda} - \alpha^2 & 0 \\ 0 & \Xi \Xi' \end{pmatrix} \right).$$

Notice, however, that  $\alpha$  only enters into the equation for the first element of  $Q\tilde{z}_t$ , and the  $(n_z - 1)$  last elements of  $Q\tilde{z}_t$  are independent of the first element (and independent of  $y_t$  at all leads and lags). Hence, it is without loss of generality to limit attention to the first element of  $Q\tilde{z}_t$  when performing identification analysis for  $\Theta_{i,j,\ell}$  and  $\alpha$ . The first element of  $Q\tilde{z}_t$  equals  $\check{z}_t$  as defined in equation (19) in the main text.<sup>B.2</sup>

Additional restrictions on the IVs can ensure point identification. In particular, if  $n_z \geq 2$  and the researcher is willing to restrict  $\Sigma_v$  to be diagonal, then  $\alpha$  is point-identified from any off-diagonal element of  $\text{Var}(\tilde{z}_t) = \Sigma_v + \alpha^2 \lambda \lambda'$ , since  $\lambda$  is point-identified.

---

<sup>B.2</sup>The above display implies that we must have  $\alpha^2 \leq (\lambda' \text{Var}(\tilde{z}_t)^{-1} \lambda)^{-1}$ , which is precisely what the upper bound for  $\alpha^2$  yields when applied to  $\check{z}_t$ .

### B.3 Instruments correlated with multiple shocks

In this section, we ask how much can be said about forecast variance ratios if the researcher is only willing to assume that the observed set of external instruments  $z_t$  is correlated with at most  $n_{\varepsilon_x}$  shocks, collected in the vector  $\varepsilon_{x,t}$ . Hence, in this section we do not impose the exclusion restriction that only the first shock  $\varepsilon_{1,t}$  be correlated with the IV(s).

EXTENDED MODEL AND FVR. Without loss of generality, suppose the  $n_z$  IVs are correlated with the first  $n_{\varepsilon_x}$  of the  $n_\varepsilon$  shocks. Denote this sub-vector of shocks by  $\varepsilon_{x,t}$ . For now,  $n_{\varepsilon_x}$  need not be known to the econometrician. We define the extended SVMA-IV model as

$$y_t = \Theta(L)\varepsilon_t, \quad \Theta(L) \equiv \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \quad (\text{B.8})$$

$$z_t = \sum_{\ell=1}^{\infty} (\Psi_\ell z_{t-\ell} + \Lambda_\ell y_{t-\ell}) + \underbrace{\Gamma \varepsilon_{x,t} + \Sigma_v^{1/2} v_t}_{\tilde{z}_t}, \quad (\text{B.9})$$

where  $\Gamma$  is  $n_z \times n_{\varepsilon_x}$ . We continue to impose i.i.d. normality of the shocks, cf. [Assumption 3](#). Our object of interest is the forecast variance ratio with respect to the  $n_z$  particular linear combinations of shocks that enter into the IV equations,  $\Gamma \varepsilon_{x,t}$ :

$$\begin{aligned} FVR_{i,\ell} &\equiv 1 - \frac{\text{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty < \tau \leq t}, \{\Gamma \varepsilon_{x,\tau}\}_{t < \tau < \infty})}{\text{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty < \tau \leq t})} \\ &= \frac{\sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})(\Gamma \Gamma')^{-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})'}{\text{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty < \tau \leq t})}. \end{aligned} \quad (\text{B.10})$$

In the following we provide upper and lower bounds on this object. Given  $\Gamma \Gamma'$ , the FVR is point-identified, so we need to derive the identified set for  $\Gamma \Gamma'$ . At the end of this section we discuss how the FVR with respect to  $\Gamma \varepsilon_{x,t}$  relates to other objects of interest.

Similar to [Appendix B.2](#), the testable restriction of the model (B.8)–(B.9) is that the joint spectrum of  $y_t$  and  $\tilde{z}_t$  has a rank- $n_{\varepsilon_x}$  factor structure. If this assumption is not rejected, we can reduce the instrument vector to dimension  $\min(n_z, n_{\varepsilon_x})$  without affecting the identification of  $FVR_{i,\ell}$ .<sup>B.3</sup> In particular, we may assume that  $\Gamma$  has full row rank, which we do from now on, thus justifying the second equality in (B.10).

---

<sup>B.3</sup>The argument is very similar to the one-shock case in [Appendix B.2](#) and is available upon request.

IDENTIFIED SET FOR  $\Gamma$ . Define  $\Sigma_{\tilde{z}} \equiv \text{Var}(\tilde{z}_t)$ . Proceeding similarly to the proof of [Proposition 1](#), we can show that a given  $\Gamma$  is consistent with the joint spectral density of the data if and only if  $\Gamma\Gamma'$  has full row rank,

$$\Sigma_{\tilde{z}} - \Gamma\Gamma' \geq 0, \quad (\text{B.11})$$

and

$$\Gamma\Gamma' - 2\pi s_{z\ddagger}(\omega) \geq 0, \quad \forall \omega \in [0, \pi], \quad (\text{B.12})$$

where  $s_{z\ddagger}(\omega) = s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega)$  and we use the notation  $A \geq B$  if  $A - B$  is Hermitian positive semi-definite (and similarly for  $\leq$ ). Sharp bounds on  $FVR_{i,\ell}$  thus follow from minimizing/maximizing [\(B.10\)](#) over the space of  $n_z \times n_z$  symmetric positive definite matrices  $\Gamma\Gamma'$  subject to constraints [\(B.11\)](#)–[\(B.12\)](#).

LOWER BOUND ON FVR. We now establish a sharp lower bound on the numerator in the definition [\(B.10\)](#) of the FVR (the denominator is point-identified). Observe that

$$\begin{aligned} & \sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})(\Gamma\Gamma')^{-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})' \\ &= \sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})\Sigma_{\tilde{z}}^{-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})' + \sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})\{(\Gamma\Gamma')^{-1} - \Sigma_{\tilde{z}}^{-1}\} \text{Cov}(y_{it}, \tilde{z}_{t-m})' \\ &\geq \sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})\Sigma_{\tilde{z}}^{-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})', \end{aligned}$$

where the inequality uses the constraint [\(B.11\)](#). The above lower bound is sharp: It is attained in a model where  $\Sigma_v = 0_{n_z \times n_z}$  and  $\Gamma\Gamma' = \Sigma_{\tilde{z}}$ , i.e., when all IVs are perfect.<sup>[B.4](#)</sup>

UPPER BOUND ON FVR. While we have not been able to derive a closed-form expression for the sharp upper bound on the FVR, it is straight-forward to numerically compute it. Let  $\mathcal{S}_n$  denote the space of  $n \times n$  real symmetric positive definite matrices, and let  $\text{tr}(A)$  denote the trace of a matrix  $A$ . The sharp upper bound on the numerator in the definition [\(B.10\)](#)

---

<sup>B.4</sup>Note that  $\Gamma\Gamma' = \Sigma_{\tilde{z}} = 2\pi s_{z\ddagger}(\omega)$  satisfies constraint [\(B.12\)](#) by the Schur complement formula and the positive semidefiniteness of the spectrum of  $(y'_t, \tilde{z}'_t)'$ .

of the FVR is given by the value of the program

$$\begin{aligned} \max_{X \in \mathcal{S}_{n_z}} \text{tr}(XC) + \text{tr}(AC) \\ X \leq B(\omega), \quad \omega \in [0, \pi]. \end{aligned} \tag{B.13}$$

Here  $X$  is a stand-in for  $(\Gamma\Gamma')^{-1} - \Sigma_{\tilde{z}}^{-1}$ ,  $C \equiv \sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})' \text{Cov}(y_{it}, \tilde{z}_{t-m})$ ,  $A \equiv \Sigma_{\tilde{z}}^{-1}$ , and  $B(\omega) \equiv \frac{1}{2\pi} s_{\tilde{z}^\dagger}(\omega)^{-1} - \Sigma_{\tilde{z}}^{-1}$ . We can solve the above program to arbitrary accuracy by casting it as a (convex) semi-definite program with a finite number of constraints. Partition the interval  $[0, \pi]$  into  $N$  equal-length pieces, and consider the relaxed constraint set

$$X \leq \tilde{B}_m, \quad m \in \{1, 2, \dots, N\}, \tag{B.14}$$

where  $\tilde{B}_m \equiv \frac{N}{\pi} \times \int_{\frac{(m-1)\pi}{N}}^{\frac{m\pi}{N}} B(\omega) d\omega$ . As  $N \rightarrow \infty$ , this constraint set approximates that of the original problem arbitrarily well, but for any finite  $N$  the value of the discretized program provides an upper bound on the numerator in (B.10). Efficient numerical algorithms to compute the solution to semidefinite programs of the form (B.13)–(B.14) are available in Matlab and other environments.<sup>B.5</sup>

Alternatively, we can derive non-sharp upper bounds on the FVR numerator (B.10) in closed form. For example, one conservative upper bound is obtained by maximizing  $\text{tr}(XC) + \text{tr}(AC)$  subject to  $X + \tilde{\Sigma}_{\tilde{z}}^{-1} \leq \left( \int_{-\pi}^{\pi} s_{\tilde{z}^\dagger}(\omega) d\omega \right)^{-1} = \text{Var}(\tilde{z}_t^\dagger)^{-1}$ . This yields the upper bound

$$\sum_{m=0}^{\ell-1} \text{Cov}(y_{it}, \tilde{z}_{t-m}) \text{Var}(\tilde{z}_t^\dagger)^{-1} \text{Cov}(y_{it}, \tilde{z}_{t-m})',$$

which binds if the shocks  $\varepsilon_{x,t}$  are all recoverable, but is otherwise not sharp. A less conservative – but still generally suboptimal – upper bound is given by

$$\text{tr}(AC) + \sum_{m=1}^{n_z} \inf_{\omega \in [0, \pi]} \check{B}_{mm}(\omega),$$

where  $\check{B}_{mm}(\omega)$  is the  $(m, m)$  element of  $\check{B}(\omega) \equiv C^{1/2'} B(\omega) C^{1/2}$ , and  $C = C^{1/2} C^{1/2'}$ . This latter upper bound is sharp when  $n_z = 1$ , in which case the lower and upper bounds in this section reduce to the FVR bound expressions derived in Section 3.

---

<sup>B.5</sup>See for example <http://cvxr.com/cvx/doc/sdp.html>. To transform our constraints into ones involving real matrices, note that a Hermitian matrix with real part  $A$  and imaginary part  $B$  is positive semi-definite if and only if the real symmetric matrix  $\begin{pmatrix} A & B \\ B & A \end{pmatrix}$  is positive semi-definite.

INTERPRETATION. We highlight two special cases where the FVR with respect to  $\Gamma\varepsilon_{x,t}$  (which we partially identified above) is of interest.

First, as in [Mertens & Ravn \(2013\)](#), one may assume that the  $n_z$  instruments are correlated with the same number  $n_{\varepsilon_x} = n_z$  of structural shocks. In that case  $\Gamma$  is square and nonsingular, so the FVR with respect to  $\Gamma\varepsilon_{x,t}$  is the same as the FVR with respect to the shocks  $\varepsilon_{x,t}$  themselves. Moreover, if we further assume that all included shocks  $\varepsilon_{x,t}$  are recoverable, then  $\tilde{z}_t^\dagger \equiv E(\tilde{z}_t | \{y_\tau\}_{-\infty < \tau < \infty}) = \Gamma\varepsilon_{x,t}$ , so the historical decomposition of  $y_t$  with respect to  $\varepsilon_{x,t}$  is point-identified as  $E(y_t | \{\varepsilon_{x,t}\}_{-\infty < \tau \leq t}) = E(y_t | \{\tilde{z}_\tau^\dagger\}_{-\infty < \tau \leq t})$ .

Second, consider the case with a single IV but possibly several included shocks,  $n_{\varepsilon_x} > 1 = n_z$ . The above analysis shows that, even though the IV exclusion restrictions in the baseline model (3) fail, the data are informative about the FVR with respect to the particular linear combination  $\Gamma\varepsilon_{x,t}$  of shocks that enters the IV equation. The FVR with respect to this particular linear combination of shocks is evidently a lower bound for the FVR with respect to the full vector  $\varepsilon_{x,t}$  of shocks that are correlated with the IV.

## B.4 Invertibility and SVAR-IV

In this section we characterize the bias of SVAR-IV methods when shocks may be noninvertible. Throughout we assume the validity of the SVMA-IV model in [Assumptions 1 to 3](#). Our analysis builds on results by [Lippi & Reichlin \(1994\)](#) and [Forni et al. \(2019\)](#), who do not consider identification using external instruments.

The SVAR-IV (or “proxy SVAR”) strategy identifies structural shocks by using the external IV to rotate the forecast errors from a reduced-form VAR ([Stock, 2008](#); [Stock & Watson, 2012](#); [Mertens & Ravn, 2013](#); [Gertler & Karadi, 2015](#); [Ramey, 2016](#)). For analytical clarity, we work with a VAR( $\infty$ ) model with forecast errors  $u_t \equiv y_t - E(y_t | \{y_\tau\}_{-\infty < \tau < t})$ . Suppose the single residualized IV  $\tilde{z}_t = \alpha\varepsilon_{1,t} + \sigma_v v_t$  is used by the econometrician. Under the SVAR-IV assumption of  $n_\varepsilon = n_y$  and invertibility of all shocks, we would have  $u_t = \Theta_0 \varepsilon_t$ , with  $\Theta_0$  square and nonsingular. Then the shock of interest would be identified as  $\varepsilon_{1,t} = \gamma' u_t$ , where  $\gamma \equiv (\Sigma'_{u\tilde{z}} \Sigma_u^{-1} \Sigma_{u\tilde{z}})^{-1/2} \Sigma_u^{-1} \Sigma_{u\tilde{z}}$ ,  $\Sigma_{u\tilde{z}} \equiv \text{Cov}(u_t, \tilde{z}_t)$ , and  $\Sigma_u \equiv \text{Var}(u_t)$ .

We now ask what happens to the outputs of the SVAR-IV procedure if the invertibility assumption does not hold and  $n_\varepsilon \geq n_y$ .

**Proposition B.3.** *Assume the SVMA-IV model in [Assumptions 1 to 3](#). The shock that is (mis)identified by SVAR-IV is given by*

$$\tilde{\varepsilon}_{1,t} \equiv \gamma' u_t = \sum_{j=1}^{n_\varepsilon} \sum_{\ell=0}^{\infty} a_{j,\ell} \varepsilon_{j,t-\ell}, \quad (\text{B.15})$$

where the scalar coefficients  $\{a_{j,\ell}\}$  satisfy  $\sum_{j=1}^{n_\varepsilon} \sum_{\ell=0}^{\infty} a_{j,\ell}^2 = 1$  and  $a_{1,0} = \sqrt{R_0^2}$ . The associated SVAR-IV impulse responses are given by<sup>B.6</sup>

$$\tilde{\Theta}_{\bullet,1,\ell} \equiv \text{Cov}(y_t, \tilde{\varepsilon}_{1,t-\ell}) = \sum_{j=1}^{n_\varepsilon} \sum_{m=0}^{\infty} a_{j,m} \Theta_{\bullet,1,\ell+m}, \quad \ell = 0, 1, 2, \dots,$$

and the impact impulse responses satisfy

$$\tilde{\Theta}_{\bullet,1,0} = \frac{1}{\sqrt{R_0^2}} \Theta_{\bullet,1,0}.$$

Under noninvertibility, SVAR-IV mis-identifies the shock as a distributed lag of all the

---

<sup>B.6</sup>In any SVAR( $\infty$ ) model, the impulse responses implied by the model must equal the local projections of the outcomes on the identified shock(s). This follows from the Wold representation.

shocks in the underlying model, with the coefficient on the true shock of interest  $\varepsilon_{1,t}$  equal to  $\sqrt{R_0^2}$  (the square root of the degree of invertibility, cf. [Section 2.1](#)). This causes impulse responses to be conflated across horizons and shocks. At the impact horizon, SVAR-IV overstates the magnitudes of the true impulse responses  $\Theta_{\bullet,1,0}$  (to a one standard deviation shock) by a factor of  $1/\sqrt{R_0^2}$ . Thus, the SVAR-IV-implied one-step-ahead forecast variance decompositions for the first shock overstate the true one-step-ahead FVRs (as defined in [Section 2.1](#)) by a factor of  $1/R_0^2$ . The bias of SVAR-IV-implied *multi-step* forecast variance decompositions depends in more complicated ways on the sequence of true impulse responses.

In summary, while SVAR-IV analysis solves the familiar “rotation problem” in SVAR analysis, it does not solve the invertibility problem. The issue is not that the IV selects a suboptimal linear combination  $\gamma$  of the forecast residuals  $u_t$  under noninvertibility, since it can be verified that  $\gamma' u_t \propto E(\varepsilon_{1,t} | u_t)$  regardless of invertibility.<sup>B.7</sup> Rather, SVAR methods fail because they assume that the time- $t$  forecast residuals suffice to recover  $\varepsilon_{1,t}$  ([Lippi & Reichlin, 1994](#)). Only under invertibility (i.e.,  $R_0^2 = 1$ ) do we have  $a_{j,\ell} = 0$  for all  $(j, \ell) \neq (1, 0)$ , so that the SVAR-IV shock  $\tilde{\varepsilon}_{1,t}$  equals the true shock  $\varepsilon_{1,t}$ . The higher the degree of invertibility  $R_0^2$ , the smaller is the extent of the SVAR-IV bias, as discussed by [Sims & Zha \(2006\)](#), [Forni et al. \(2019\)](#), and [Wolf \(2020\)](#). An explicit illustration of SVAR-IV mis-identification is provided in [Appendix B.5](#).

---

<sup>B.7</sup>In particular, no other linear combination  $\gamma$  can yield a representation ([B.15](#)) where the weight  $a_{1,0}$  exceeds  $\sqrt{R_0^2}$  (subject to  $\text{Var}(\tilde{\varepsilon}_{1,t}) = 1$ ). Thus, the IV handles the identification problem as well as possible subject to the constraints imposed by the (erroneous) invertibility assumption. As discussed in [Section 2.1](#), *dynamic rotations* circumvent this issue by obtaining the shock  $\tilde{\varepsilon}_{1,t}$  as a function of current *and future* reduced-form residuals  $\{u_\tau\}_{\tau \geq t}$ . An argument similar to that in the proof of [Proposition B.3](#) shows that, with such dynamic rotations, the weight on the true shock of interest is bounded above by  $\sqrt{R_\infty^2}$ . Dynamic rotations can thus solve the identification problem if and only if the shock of interest is recoverable.

## B.5 Illustration in a structural macro model

In [Section 6](#) we use several simple analytical examples to illustrate how our upper bound works. In this section we complement those simple examples with a quantitative exercise.

The nature of our exercise is as follows. We consider an econometrician observing (i) a small set of macroeconomic aggregates generated from the model of [Smets & Wouters \(2007\)](#) and (ii) noisy measures of some of the model’s true underlying structural shocks (i.e., valid external instruments). For clarity, we abstract from any sampling uncertainty and assume that the econometrician observes an infinite amount of data, so the joint spectral density of observed macro aggregates and external IVs is perfectly known to her. Given this spectral density, she uses our bounds to draw conclusions about variance decompositions and the degree of invertibility, without exploiting the underlying structure of the model. Overall, the point of this exercise is to show that our conclusions on likely tightness of the upper bound are not an artifact of the particular stylized environments considered in [Section 6](#), but similarly obtain in quantitatively relevant, dynamic structural macro models, for exactly the same economic reasons.

### B.5.1 Preliminaries

We employ the [Smets & Wouters \(2007\)](#) model. Throughout, we parametrize the model according to the posterior mode estimates of [Smets & Wouters \(2007\)](#).<sup>B.8</sup> Following the canonical trivariate VAR in the empirical literature on monetary policy shock transmission, our baseline specification assumes the econometrician observes aggregate output, inflation, and the short-term policy interest rate; we consider additional observables below. These macro aggregates are all stationary in the model, so they should be viewed as deviations from trend. The model features seven unobserved shocks, so not all shocks can be invertible in the baseline specification.

The econometrician observes a single external instrument  $z_t$  for the shock of interest  $\varepsilon_{1,t}$ :

$$z_t = \alpha \varepsilon_{1,t} + \sigma_v v_t.$$

We normalize  $\alpha = 1$  throughout and compute identified sets for two different degrees of informativeness of the external instrument,  $\frac{1}{1+\sigma_v^2} \in \{0.25, 0.5\}$ . We do not attach any specific

---

<sup>B.8</sup>Our implementation of the Smets-Wouters model is based on Dynare replication code kindly provided by Johannes Pfeifer. The code is available at <https://sites.google.com/site/pfeiferecon/dynare>.

economic interpretation to the IV in the context of the model.

We separately consider three different shocks of interest: a monetary shock, a technology shock, and a forward guidance shock. The conventional monetary shock as well as the technology shock are already included in the original model of [Smets & Wouters](#). For the forward guidance shock, we depart from their model by assuming that monetary policy shocks are known two quarters in advance; that is, we change the model’s Taylor rule to

$$r_t = \rho_r r_{t-1} + (1 - \rho_r) \times (\phi_\pi \pi_t + \phi_y \hat{y}_t + \phi_{dy} (\hat{y}_t - \hat{y}_{t-1})) + \varepsilon_{t-2}^m,$$

where  $r_t$  denotes the nominal interest rate,  $\pi_t$  denotes the inflation,  $\hat{y}_t$  is the output gap, and  $\varepsilon_t$  is the monetary shock.<sup>B.9</sup> Overall, these three shocks are chosen in line with our simple analytical illustrations in [Section 6](#), and identification will be subject to the same economic intuition as the small-scale examples discussed there.

We emphasize that our results in the remainder of this section should *not* be taken to imply that conventional monetary shocks are robustly near-invertible, or that forward guidance and technology shocks are never invertible – clearly, our statements are always conditional on a certain set of observables. Instead, the only purpose of this section is to document that the simple economic intuition of [Section 6](#) still plays out in a quantitative macro model with rich dynamics. To further clarify this point, we finish this section by briefly discussing how our results change with alternative sets of observables.

## B.5.2 Information content of several observables

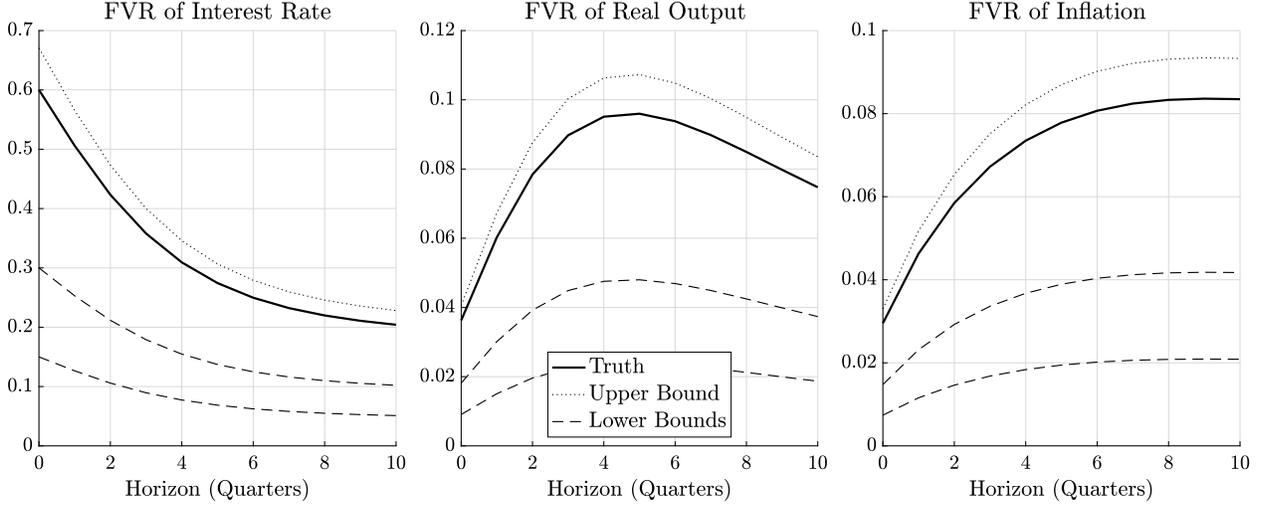
We first consider identification of monetary policy shocks, i.e., shocks to the serially correlated disturbance in the model’s Taylor rule.

The monetary shock is nearly invertible. In the model of [Smets & Wouters](#), monetary policy shocks are the only shock to contemporaneously move inflation and nominal rates in opposite directions ([Uhlig, 2005](#)). Given this unique conditional co-movement, the intuition offered in [Section 6.1](#) suggests that the degree of invertibility should be high, and indeed it equals  $R_0^2 = 0.8702$ , in spite of the limited importance of monetary shocks for aggregate fluctuations in this model ([Wolf, 2020](#)). Looking forward in time does not sharpen identification much further ( $R_\infty^2 = 0.8763$ ), and neither does looking across the spectrum

---

<sup>B.9</sup>This is the notion of forward guidance discussed, for example, in [Del Negro et al. \(2012\)](#).

## MONETARY SHOCK: IDENTIFIED SET OF FVRS



**Figure B.1:** Horizon-by-horizon identified sets for FVRs up to 10 quarters. The two lower bounds correspond to an IV with  $\frac{1}{1+\sigma_v^2} = 0.25$  (lower dashed line) and an IV with  $\frac{1}{1+\sigma_v^2} = 0.5$ .

frequency-by-frequency ( $\alpha_{LB}^2 = 0.8947$ ).<sup>B.10</sup>

Figure B.1 shows that the upper bounds on the forecast variance ratios are close to the true values. By construction, the upper and lower bounds are proportional to the true FVRs. The lower bound scales one-for-one with instrument informativeness, while the upper bound scales one-for-one with the maximal informativeness of the data for the shock across frequencies. Thus, near-invertibility immediately implies that the upper bounds are throughout close to the true FVR. In contrast, the informativeness of the lower bounds depends entirely on the strength of the IV.

### B.5.3 Dynamic information content

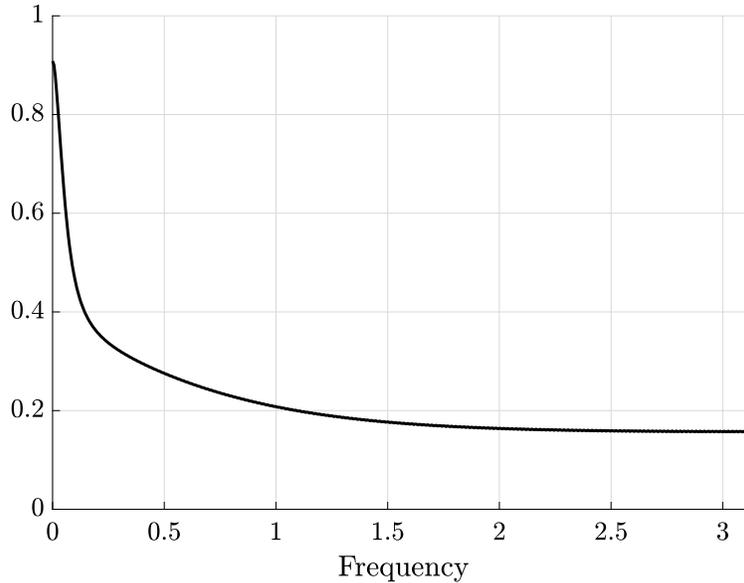
Next, we consider the identification of technology shocks, i.e., innovations to the autoregressive process of total factor productivity. This type of shock illustrates how our sharp upper bound leverages information across frequencies.

In our baseline trivariate specification, the macro aggregates are informative about only the longest cycles of the technology shock. Figure B.2 reports the spectral density of the best two-sided linear predictor of the technology shock. Strikingly, this spectral density is

---

<sup>B.10</sup>Formally, the scaled spectral density  $2\pi s_{\varepsilon_1^\dagger}(\cdot)$  of the best two-sided linear predictor of the monetary shock is nearly flat at around 0.9.

TECHNOLOGY SHOCK: SPECTRAL DENSITY OF BEST 2-SIDED LINEAR PREDICTOR



**Figure B.2:** Scaled spectral density  $2\pi s_{\varepsilon_1^*}(\cdot)$  of the best two-sided linear predictor of the technology shock. A frequency  $\omega$  corresponds to a cycle of length  $\frac{2\pi}{\omega}$  quarters.

small at business-cycle frequencies, but close to 1 for long-run fluctuations, with a peak of  $\alpha_{LB}^2 = 0.9084$ . Intuitively, as in our simple example in [Section 6.2](#), technology shocks here account for most of the long-run fluctuations, and so macro aggregates are highly informative about the IV at low frequencies; as a result, our sharp upper bound on shock importance will again be tight. In contrast, averaging across frequencies, the shock is neither near-invertible ( $R_0^2 = 0.1977$ ) nor near-recoverable ( $R_\infty^2 = 0.2166$ ).<sup>B.11</sup>

Consequently, [Figure B.3](#) shows that the sharp upper bounds on FVRs are tight. As always, the tightness of the lower bound is entirely governed by the strength of the IV.

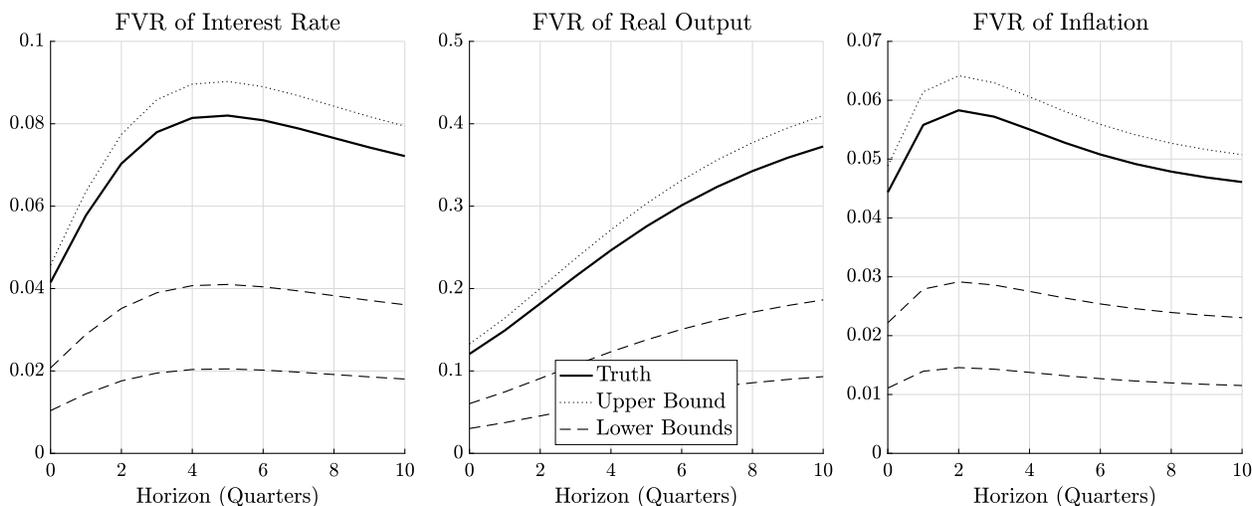
### B.5.4 Non-invertibility and news shocks

For our third example, we modify the model to include forward guidance shocks, a type of news shock. As discussed above, a forward guidance shock is identical to a monetary shock, except that it is anticipated two quarters in advance by economic agents. This third example

---

<sup>B.11</sup>The issue is that, at short horizons, other shocks – notably the price and wage mark-up shocks – also push inflation and output in opposite directions. However, the technology shock becomes nearly invertible with a judicious choice of further observables; in particular, including either the level of TFP or hours worked leads to a nearly invertible representation.

### TECHNOLOGY SHOCK: IDENTIFIED SET OF FVRs



**Figure B.3:** Horizon-by-horizon identified sets for FVRs up to 10 quarters. The two lower bounds correspond to an IV with  $\frac{1}{1+\sigma_v^2} = 0.25$  (lower dashed line) and an IV with  $\frac{1}{1+\sigma_v^2} = 0.5$ .

illustrates the robustness of our method to non-invertibility induced by news shocks.

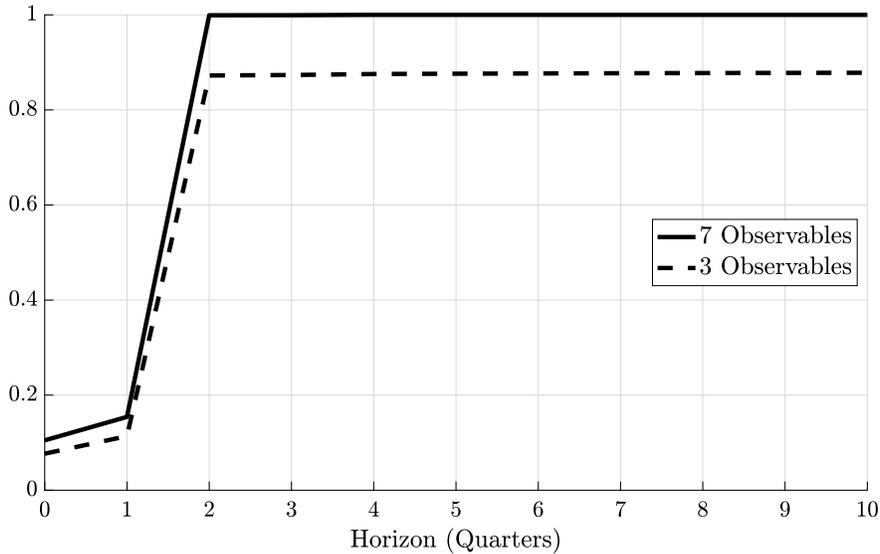
The forward guidance shock is highly non-invertible, though it is nearly recoverable. [Figure B.4](#) shows the degree of invertibility  $R_\ell^2$  up to time  $t + \ell$  of the forward guidance shock. The figure considers horizons from  $\ell = 0$  (the degree of invertibility) up to  $\ell = 10$  (close to the degree of recoverability). Contemporaneous informativeness is limited, with  $R_0^2 = 0.0768$ ; intuitively, when the forward guidance is announced, all macro aggregates move in the same direction, suggesting to the econometrician that the economy was probably buffeted by an ordinary demand shock.<sup>B.12</sup> At  $\ell = 2$ , however, the corresponding  $R^2$  jumps to  $R_2^2 = 0.8724$ , reaching  $R_\infty^2 = 0.8807$  as the overall degree of recoverability. The economic intuition is again simple: Two quarters from now, when the anticipated innovation finally directly hits the Taylor rule, the interest rate response suddenly switches sign, sending a strong signal that in fact a monetary policy shock – and not some other kind of demand shock – had occurred. By the same logic as in [Appendix B.5.2](#), the forward guidance shock is nearly recoverable.

Because of non-invertibility, [Figure B.5](#) shows that the conventional SVAR-IV approach dramatically overstates the forward guidance FVRs. In particular, as revealed by our analysis

---

<sup>B.12</sup>This belief is further reinforced by the movement of nominal interest rates: Since output and inflation have both increased (for an expansionary forward guidance shock), nominal interest rates initially increase (as dictated by the Taylor rule), before declining two periods later.

FORWARD GUIDANCE SHOCK: DEGREE OF INVERTIBILITY AT TIME  $t + \ell$



**Figure B.4:** Population  $R_\ell^2$  for the forward guidance shock, with three observables (output, inflation, interest rate) and seven observables (the full set in Smets & Wouters, 2007).

in Appendix B.4, the (impact) FVR is biased upward by a factor of  $1/R_0^2 \approx 13$  (!).<sup>B.13</sup>

Figure B.6 shows that our method instead achieves a tight upper bound on the FVR, irrespective of the degree of invertibility. Since the forward guidance shock is nearly recoverable, the upper bounds of our identified sets for the different FVRs are again close to the truth, similar to the conventional (near-invertible) monetary shock studied in Appendix B.5.2.

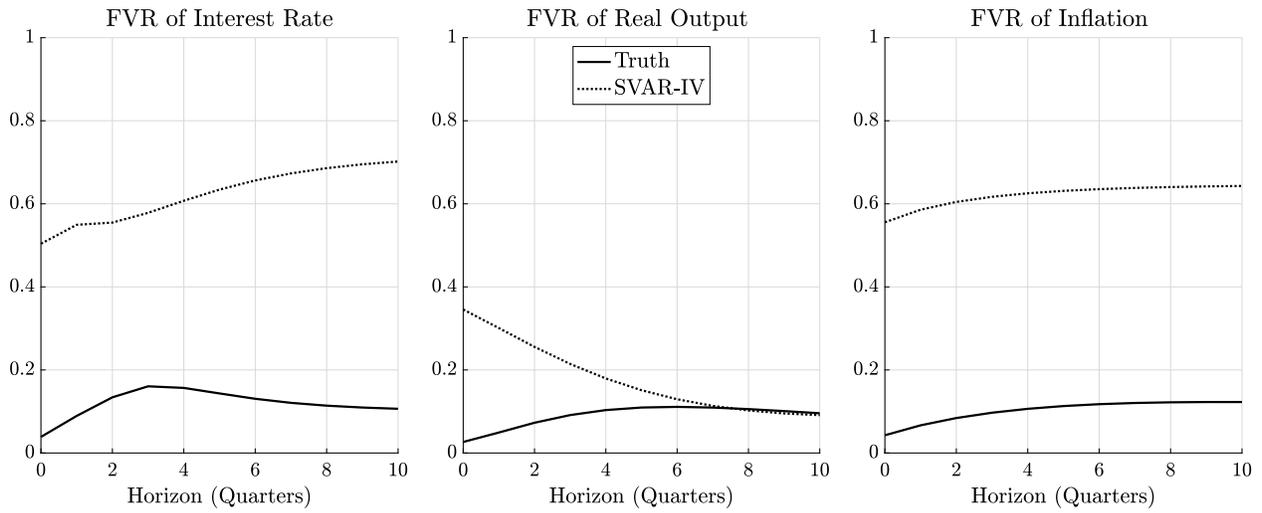
### B.5.5 Other observables

The results in the preceding sections are designed to illustrate the economic logic of our method. They should not, however, be interpreted as offering generally valid conclusions on the invertibility (or lack thereof) of different structural shocks – such statements are invariably sensitive to the choice of observables. To further emphasize this point, we in Table B.1 compute the degrees of invertibility and recoverability for each shock, for different sets of macro observables.

The degrees of invertibility and recoverability are by definition increasing in the number of macroeconomic observables. For the baseline monetary shock, the degree of invertibility is

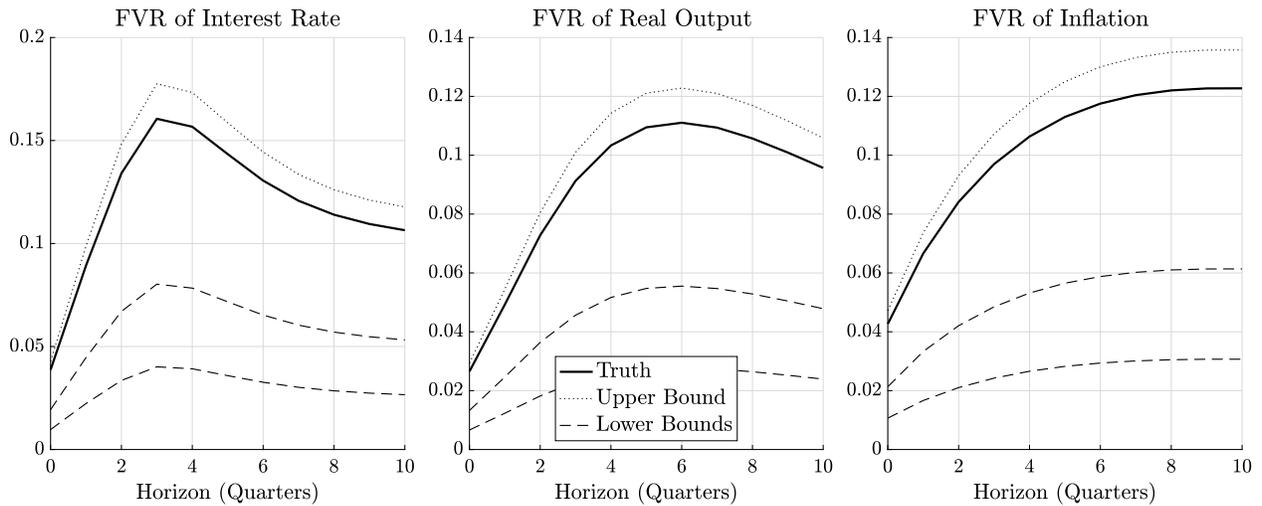
<sup>B.13</sup>Of course, for suitably chosen sets of observables (e.g., expectations about future interest rates), the non-invertibility problem would disappear (Leeper et al., 2013). Our method is robust in the sense that it does not require such a judicious choice of further observables – it works even under extreme non-invertibility.

### FORWARD GUIDANCE SHOCK: SVAR-IV FVRs



**Figure B.5:** FVRs for a forward guidance shock in the Smets-Wouters model, true values and SVAR-IV-estimated values (population limit). Baseline set of three observables.

### FORWARD GUIDANCE SHOCK: IDENTIFIED SET OF FVRs



**Figure B.6:** Horizon-by-horizon identified sets for FVRs up to 10 quarters. The two lower bounds correspond to an IV with  $\frac{1}{1+\sigma_v^2} = 0.25$  (lower dashed line) and an IV with  $\frac{1}{1+\sigma_v^2} = 0.5$ .

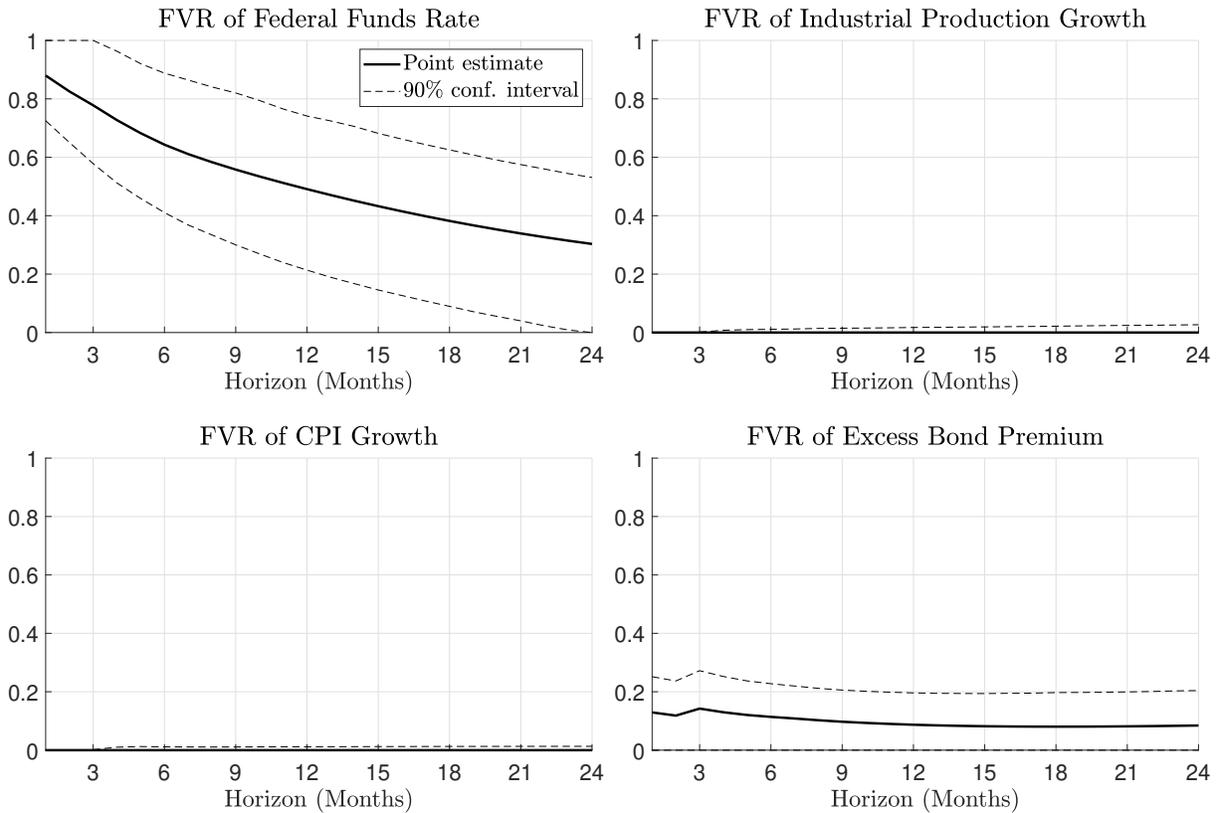
STRUCTURAL ILLUSTRATION: DEGREE OF INVERTIBILITY/RECOVERABILITY

Macro observables	Monetary shock		Technology shock		Forw. guid. shock	
	$R_0^2$	$R_\infty^2$	$R_0^2$	$R_\infty^2$	$R_0^2$	$R_\infty^2$
Baseline	0.8702	0.8763	0.1977	0.2166	0.0768	0.8807
+ investm. + consum.	0.9415	0.9507	0.2128	0.2384	0.0980	0.9492
+ hours	0.9272	0.9286	0.9799	0.9816	0.0774	0.9331
All observables	1	1	1	1	0.1049	1

**Table B.1:** Degree of invertibility  $R_0^2$  and degree of recoverability  $R_\infty^2$  in Smets-Wouters model, given three different sets of macro observables  $y_t$ . “Baseline” is the 3-variable specification with output, inflation, and short-term interest rate. The second and third rows add either (i) investment and consumption or (ii) hours to the baseline observables. The last row has the full set of observables considered in [Smets & Wouters \(2007\)](#).

high as soon as the researcher observes both nominal interest and inflation; with the full set of observables, the shock becomes perfectly invertible. For the technology shock, the degree of invertibility jumps to almost 1 as soon as hours worked become observable; intuitively, this is so because, at the posterior mode of the Smets-Wouters model, most *high- and low-frequency* variation of hours worked is driven by technology shocks. Finally, because none of the observables included in the estimation exercise of [Smets & Wouters](#) are forward-looking measures of nominal interest rates, the forward guidance shock remains highly non-invertible regardless of the choice of observables.

## EMPIRICAL APPLICATION: FORECAST VARIANCE RATIOS, SVAR-IV



**Figure B.7:** Point estimates and 90% confidence intervals for the identified sets of forecast variance ratios/decompositions estimated by SVAR-IV, across different variables and forecast horizons. For visual clarity, we force bias-corrected estimates/bounds to lie in  $[0, 1]$ . The interest rate variable is the Federal Funds Rate.

## B.6 Monetary policy shock application: further results

Complementing the empirical results in [Section 5](#), [Figure B.7](#) shows the forecast variance ratios/decompositions estimated by a conventional SVAR-IV procedure, with bootstrap confidence intervals. The conclusions about the irrelevance of monetary shocks for output growth and inflation are even starker in this figure than in the main paper. Our bounds estimates in [Section 5](#) show that the irrelevance of monetary shocks is not merely an artifact of assuming invertibility, but is instead a robust implication of the empirical covariances of the macro aggregates  $y_t$  with the monetary shock instrument  $z_t$  of [Gertler & Karadi \(2015\)](#).

## B.7 Oil news shock application

As a second application of our method, we study the importance of news about oil supply for aggregate business-cycle fluctuations. We use changes in oil supply expectations as an IV for oil news shocks and obtain two main results. First, we find that such oil supply news shocks are highly non-invertible, invalidating the standard SVAR-IV approach. Second, we show that our invertibility-robust method again yields sharp inference: Over the past three decades, oil news shocks have mattered little for U.S. inflation and played almost no role in U.S. output fluctuations.

BACKGROUND. [Känzig \(2021\)](#) constructs a measure of shocks to oil supply expectations from high-frequency changes in asset prices around press releases of the Organization of the Petroleum Exporting Countries (OPEC). Since news shocks are often non-invertible, as discussed in [Section 6.3](#), this setting is a second promising laboratory to showcase the appeal of our SVMA-IV approach.<sup>B.14</sup>

MODEL. We consider the same set of endogenous macro observables  $y_t$  as in [Känzig \(2021, Section III.E\)](#): log real oil prices, log oil production, log oil inventories, log world industrial production (IP), log broad U.S. nominal effective exchange rate index (NEER), log U.S. industrial production (IP), log U.S. consumer price index (CPI), federal funds rate (FFR), log VXO uncertainty index, and log U.S. terms of trade (TOT). Since our method requires stationary data, we transform the following variables to log growth rates (rather than log levels): oil production, oil inventories, world and U.S. industrial production, the nominal effective exchange rate, and U.S. CPI. Data are monthly from April 1983 to December 2017, as dictated by the availability of the IV. We include  $p = 12$  lags and a constant in the reduced-form VAR (following [Känzig](#)), and use 1,000 bootstrap draws from a homoskedastic recursive residual VAR bootstrap.

We estimate the importance of oil shocks using both our invertibility-robust SVMA-IV framework as well as the conventional SVAR-IV approach. Note that, due to the differences in sample, data transformations, and bootstrap procedure, our SVAR-IV results are not directly comparable to those reported in [Känzig \(2021, Table 2\)](#).

---

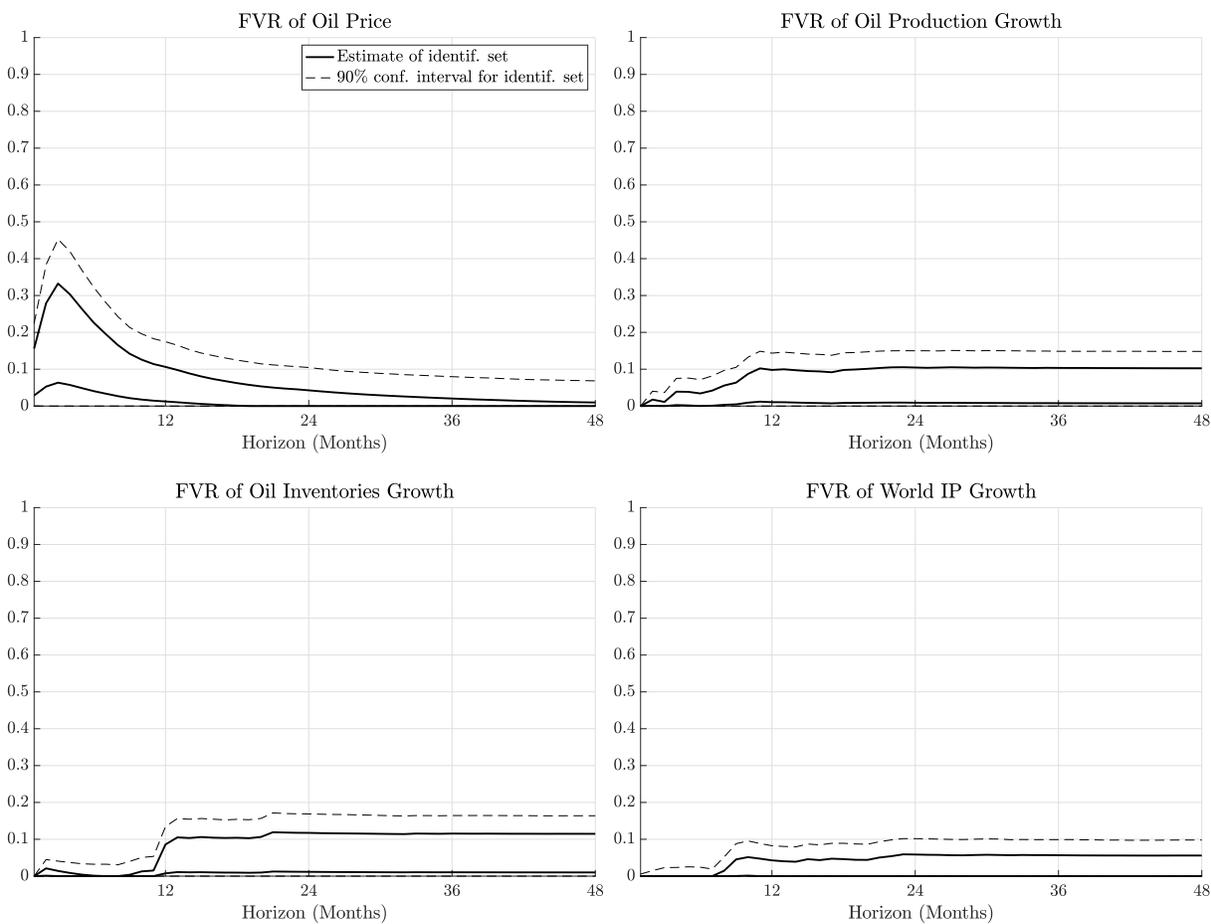
<sup>B.14</sup>We thank an anonymous referee for suggesting this application of our method.

RESULTS. Our first main finding is that invertibility is strongly rejected: The 90% confidence interval for the identified set of the degree of invertibility is  $[0, 0.190]$ , and the p-value for the Granger causality pre-test of invertibility in [Section 4](#) is 0.0064. We emphasize that invertibility is strongly rejected even though the VAR includes several forward-looking financial variables (e.g., oil prices, exchange rates, and the VIX). While these variables are known to respond quickly to oil supply news, such fast responses are apparently not sufficient to ensure invertibility; intuitively, financial variables are likely to respond quickly to many other nuisance shocks as well, which all things equal lowers the degree of invertibility of our shock of interest. We conclude that the analysis of oil supply news shocks calls for invertibility-robust approaches, such as our SVMA-IV procedure.

[Figures B.8](#) and [B.9](#) show our partial identification robust confidence intervals for the forecast variance ratio of the endogenous macro variables with respect to the oil news shocks. We report point estimates and confidence intervals for the identified sets at each horizon separately. As in [Känzig \(2021\)](#), we find that the data are consistent with the oil news shock explaining a sizable portion (but not the majority) of short-run oil price fluctuations. In spite of our weak identifying assumptions, we can furthermore conclude that – on our post-1983 sample – oil supply news have played a limited role in U.S. consumer price fluctuations, and have been essentially irrelevant for world and (in particular) U.S. real economic activity. Consistent with these results, oil supply news are estimated to have an at most moderate short-term effect on the U.S. terms of trade.

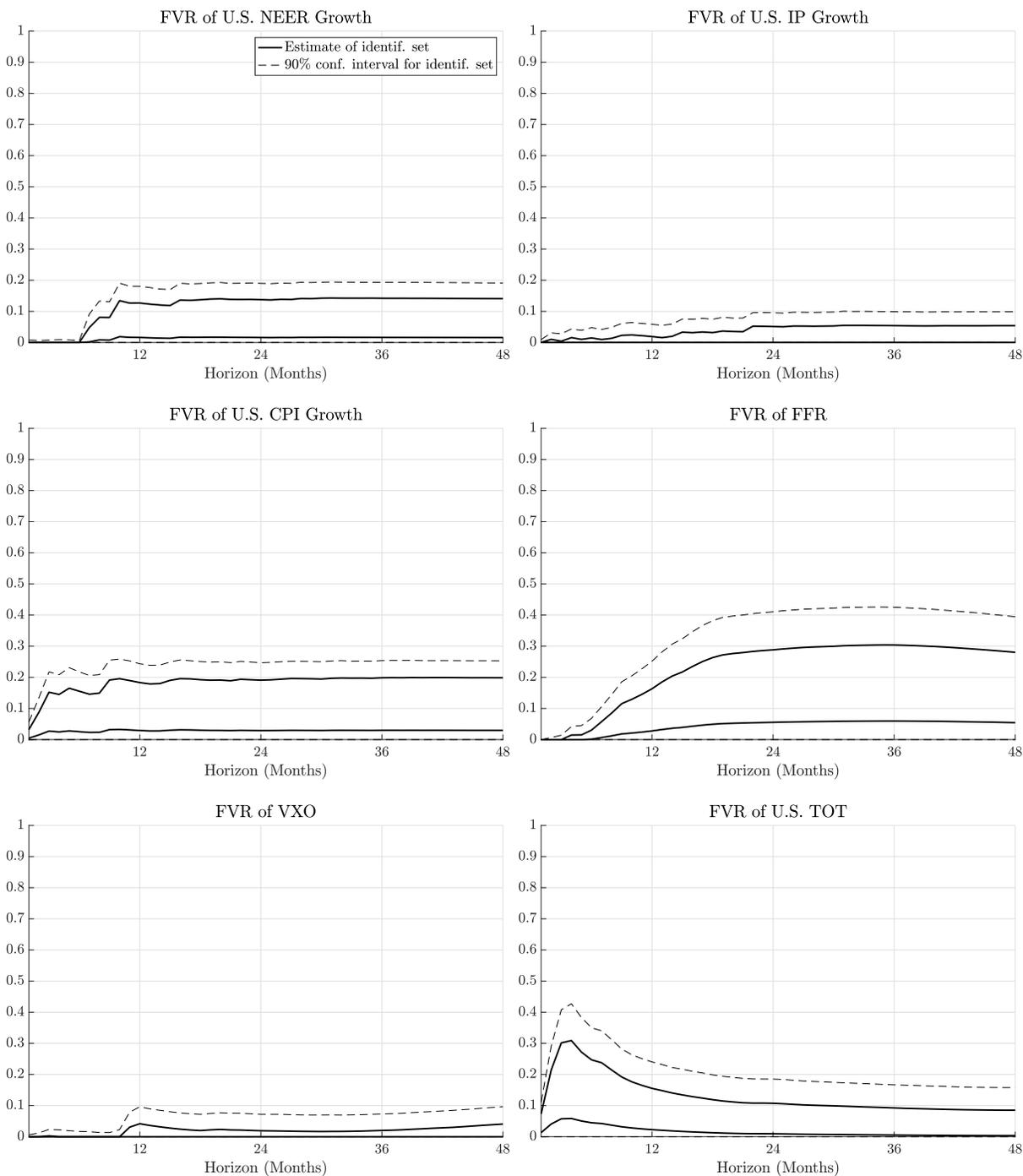
For comparison, [Figures B.10](#) and [B.11](#) show the analogous confidence intervals obtained from a conventional SVAR-IV procedure. The impact FVR estimates are substantially over-estimated relative to the invertibility-robust results, consistent with our finding of substantial non-invertibility and with [Proposition B.3](#). Our invertibility-robust results show that the following conclusions from the non-robust SVAR-IV analysis are spurious: (i) Oil supply news shocks are the overwhelming driver of oil price fluctuations in the short run; (ii) they are important drivers of medium-run fluctuations in U.S. inflation and real activity; and (iii) they are the dominant driver of the U.S. terms of trade at essentially all horizons.

## OIL NEWS SHOCK: SVMA-IV FVRs, GLOBAL VARIABLES



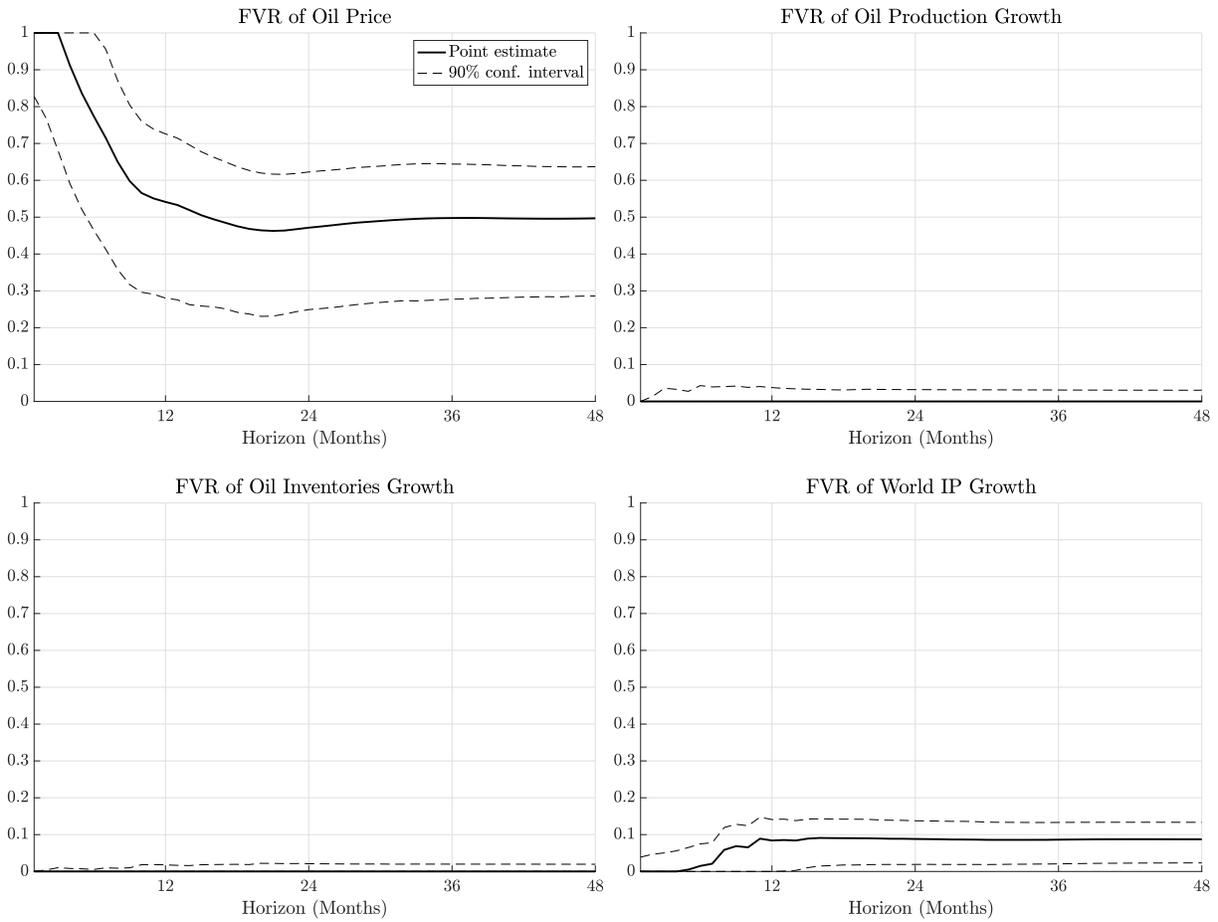
**Figure B.8:** Point estimates and 90% confidence intervals for the identified set of oil news shock FVRs, across different variables and forecast horizons, produced using our invertibility-robust SVMA-IV approach. For visual clarity, we force bias-corrected estimates/bounds to lie in  $[0, 1]$ .

## OIL NEWS SHOCK: SVMA-IV FVRs, U.S. VARIABLES



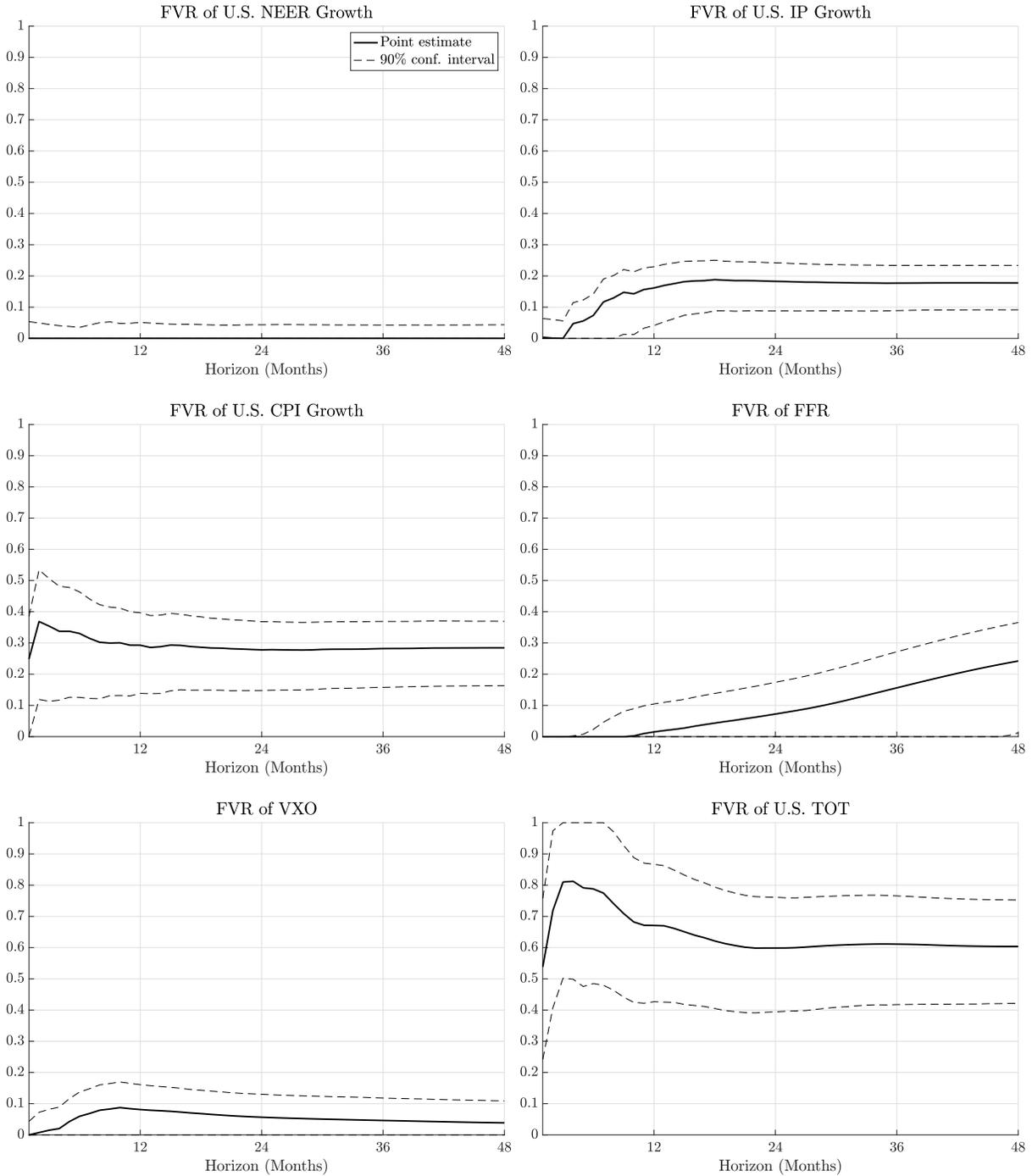
**Figure B.9:** Point estimates and 90% confidence intervals for the identified set of oil news shock FVRs, across different variables and forecast horizons, produced using our invertibility-robust SVMA-IV approach. For visual clarity, we force bias-corrected estimates/bounds to lie in  $[0, 1]$ .

## OIL NEWS SHOCK: SVAR-IV FVRs, GLOBAL VARIABLES



**Figure B.10:** Point estimates and 90% confidence intervals for oil news shock FVRs, across different variables and forecast horizons, produced using a conventional SVAR-IV approach. For visual clarity, we force bias-corrected estimates/bounds to lie in  $[0, 1]$ .

## OIL NEWS SHOCK: SVAR-IV FVRs, U.S. VARIABLES



**Figure B.11:** Point estimates and 90% confidence intervals for oil news shock FVRs, across different variables and forecast horizons, produced using a conventional SVAR-IV approach. For visual clarity, we force bias-corrected estimates/bounds to lie in  $[0, 1]$ .

## B.8 Nonparametric sieve VAR inference

In this appendix section we show that the bound estimates proposed in [Section 4](#) are jointly asymptotically normal under nonparametric conditions on the DGP, as long as the VAR lag length is chosen to increase with the sample size at an appropriate rate. The nonparametric viewpoint does not change the practical steps necessary to implement the inference strategy; it only provides regularity conditions under which it is asymptotically innocuous to approximate the true VAR( $\infty$ ) data generating process by a finite-lag VAR. We utilize the classic sieve VAR results of [Lewis & Reinsel \(1985\)](#) (who build on the univariate results of [Berk, 1974](#)) to prove asymptotic normality of those nonlinear functionals of the estimated VAR spectrum that appear in our bounds. Our main result below is similar in spirit to the abstract theorem in [Saikkonen & Lutkepohl \(2000, Thm. 2\)](#), although our regularity conditions are more easily verifiable as they are tailored to our parameters of interest (however, unlike [Saikkonen & Lutkepohl](#), we only consider stationary data).

The purpose of this section is merely to demonstrate that existing sieve VAR theory implies that empirical SVMA-IV analysis can be carried out in a nonparametric fashion. We do not claim to provide conceptually new insights into sieve VAR econometrics. Although here we only prove the validity of the sieve VAR strategy for delta method inference, we expect that similar results could be established for bootstrap sieve VAR inference in the SVMA-IV model, see [Gonçalves & Kilian \(2007\)](#), [Meyer & Kreiss \(2015\)](#), and references therein.

### B.8.1 Assumptions, parameters of interest, and estimator

We first define the general class of parameters of interest for empirical SVMA-IV analysis, and we place assumptions on the DGP and VAR lag length. Our goal is to stay close to the set-up in [Lewis & Reinsel \(1985\)](#), so as to demonstrate how existing asymptotic results can be readily adapted to study sieve VAR estimators for SVMA-IV purposes.

We assume that the data are generated by a reduced-form VAR( $\infty$ ) model with i.i.d. innovations. The observations are denoted by  $W_t \equiv (y'_t, z'_t)' \in \mathbb{R}^{n_W}$ ,  $t = 1, 2, \dots, T$ , where  $n_W \equiv n_y + 1$ . In order to make clear the connection with [Lewis & Reinsel \(1985\)](#), we assume that the data are known to have mean zero. It is straight-forward to extend all results to allow for non-zero means by including an intercept in the estimated VAR. Let  $\|B\| \equiv (\text{tr}(B'B))^{1/2}$  denote the Frobenius norm.

**Assumption B.1.** *The process  $\{W_t\}$  is generated by the mean-zero stationary VAR( $\infty$ ) model*

$$A(L)W_t = e_t.$$

Here  $A(z) \equiv I_{n_W} - \sum_{\ell=1}^{\infty} A_\ell z^\ell$  for  $z \in \mathbb{C}$ , and  $A_\ell \in \mathbb{R}^{n_W \times n_W}$  for all  $\ell$ . We impose the following conditions:

- i)  $\det(A(z)) \neq 0$  for all  $|z| \leq 1$ , and  $\sum_{\ell=1}^{\infty} \|A_\ell\| < \infty$ .
- ii)  $\{e_t\}$  is an  $n_W$ -dimensional i.i.d. process with  $E(e_t) = 0_{n_W \times 1}$ ,  $\Sigma \equiv \text{Var}(e_t)$  is positive definite, and  $E\|e_t\|^8 < \infty$ .

These conditions are the same as in [Lewis & Reinsel \(1985\)](#), except that we here assume that  $e_t$  has 8 moments instead of just 4.<sup>B.15</sup> [Meyer & Kreiss \(2015\)](#) discuss the generality of assuming a reduced-form VAR( $\infty$ ) with i.i.d. disturbances, see also [Kreiss et al. \(2011\)](#) for more details in the univariate case. Assuming the SVMA-IV model (1)–(3) holds, the i.i.d. assumption on the one-step-ahead reduced-form forecast errors  $e_t$  is automatically satisfied, provided that the structural shocks  $(\varepsilon'_t, v_t)'$  are themselves i.i.d. and either (i) invertible, or (ii) Gaussian (regardless of invertibility). Although we are here deliberately aiming at conceptual clarity rather than full generality, we expect it would be straight-forward to weaken the i.i.d.-ness assumption on  $e_t$  by appealing to a suitable multivariate version of the sieve VAR result of [Gonçalves & Kilian \(2007\)](#), who assume heteroskedastic martingale difference innovations.

Next, we define the class of parameters of interest for empirical SVMA-IV analysis. Define the two matrix-valued functions

$$A_{\cos}(\omega) \equiv \sum_{\ell=1}^{\infty} A_\ell \cos(\omega\ell), \quad A_{\sin}(\omega) \equiv \sum_{\ell=1}^{\infty} A_\ell \sin(\omega\ell), \quad \omega \in [0, 2\pi].$$

The parameter of interest is of the form

$$\psi \equiv \int_0^{2\pi} h(\omega)' g(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) d\omega,$$

where we define the functions  $h: [0, 2\pi] \rightarrow \mathbb{R}^K$  and  $g: \mathcal{A}_\delta \times \mathbb{S}_{n_W} \rightarrow \mathbb{R}^K$ , the set  $\mathcal{A}_\delta = \{(B_1, B_2) \in \mathbb{R}^{n_W \times n_W} \times \mathbb{R}^{n_W \times n_W} : |\det(I_{n_W} - B_1 - iB_2)| \geq \delta\}$ , and the fixed number  $\delta > 0$  that

---

<sup>B.15</sup>We only use more than four moments in the proofs of [Lemmas B.6 to B.8](#) below, where the extra moments make the arguments more transparent.

is strictly smaller than  $\inf_{\omega \in [0, 2\pi]} |\det(A(e^{i\omega}))|$ .  $\mathbb{S}_{n_W}$  denotes the set of  $n_W \times n_W$  symmetric positive definite matrices.

For appropriate choices of  $h(\cdot)$  and  $g(\cdot)$ , the above class of parameters includes almost all the parameters/bounds in SVMA-IV analysis.<sup>B.16</sup> For example, the class contains (i) elements  $\Sigma_{ij}$  of  $\Sigma$ , (ii) the degree of recoverability  $R_\infty^2 = \int_0^{2\pi} s_{y\bar{z}}(\omega)^* s_y(\omega)^{-1} s_{y\bar{z}}(\omega) d\omega$ , and (iii) autocovariances  $E(w_{i,t} w_{j,t-\ell}) = \int_0^{2\pi} e^{i\omega\ell} s_{w,ij}(\omega) d\omega$ . Here  $w_t \equiv (y'_t, \tilde{z}'_t)'$ , and for all  $\omega \in [0, 2\pi]$ ,

$$s_w(\omega) = \begin{pmatrix} s_y(\omega) & s_{y\bar{z}}(\omega) \\ s_{\bar{z}y}(\omega) & s_{\bar{z}}(\omega) \end{pmatrix} = \frac{1}{2\pi} \begin{pmatrix} (I_{n_y}, 0_{n_y \times 1})(A_{\cos}(\omega) + iA_{\sin}(\omega))^{-1} & 0_{n_y \times 1} \\ 0_{1 \times n_y} & 1 \end{pmatrix} \\ \times \Sigma \begin{pmatrix} (A_{\cos}(\omega) - iA_{\sin}(\omega))^{-1'}(I_{n_y}, 0_{n_y \times 1})' & 0_{n_y \times 1} \\ 0_{1 \times n_y} & 1 \end{pmatrix}.$$

Other SVMA-IV parameters can be constructed as nonlinear transformations of a finite number of autocovariances. By the Cramér-Wold device, it is without loss of generality to consider vector-valued (rather than matrix-valued) functions  $h(\cdot)$  and  $g(\cdot)$ . In the following, we further assume  $K = 1$  so that both  $h(\cdot)$  and  $g(\cdot)$  are scalar. This eases the notation without sacrificing essential generality, as should be clear from the proofs.

We place certain smoothness conditions on the parameter of interest, thus permitting a delta method argument.

**Assumption B.2.** *The function  $h(\cdot)$  is continuous on  $[0, 2\pi]$ . On any non-empty, compact subset of the domain  $\mathcal{A}_\delta \times \mathbb{S}_{n_W}$ , the function  $g(\cdot, \cdot, \cdot)$  is twice continuously differentiable. Denote the partial derivatives by  $g_1(B_1, B_2, S) \equiv \frac{\partial g(B_1, B_2, S)}{\partial \text{vec}(B_1)}$ ,  $g_2(B_1, B_2, S) \equiv \frac{\partial g(B_1, B_2, S)}{\partial \text{vec}(B_2)}$ , and  $g_3(B_1, B_2, S) \equiv \frac{\partial g(B_1, B_2, S)}{\partial \text{vec}(S)}$ . At the true VAR parameters  $\{A_\ell\}$  and  $\Sigma$ , each of the functions  $\omega \mapsto g_j(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)$ ,  $j = 1, 2, 3$ , belongs to  $L_2(0, 2\pi)$  (element-wise).*

The smoothness conditions in [Assumption B.2](#) are easily verified for all parameters of interest in SVMA-IV analysis, since [Assumption B.1](#) ensures that the true VAR spectrum is non-singular.

Finally, we define a sieve VAR estimator as the sample analogue of the population parameter of interest. For any  $p \in \mathbb{N}$ , define  $X_t(p) \equiv (W'_{t-1}, \dots, W'_{t-p})' \in \mathbb{R}^{n_W p}$  and the

---

<sup>B.16</sup>The only exception is the parameter  $\sup_{\omega \in [0, 2\pi]} s_{\bar{z}^\dagger}(\omega)$ , which is discussed in the main text.

least-squares VAR estimator

$$\hat{\beta}(p) \equiv \left( \hat{A}_1(p), \dots, \hat{A}_p(p) \right) \equiv \left( \sum_{t=p+1}^T W_t(p) X_t(p)' \right) \left( \sum_{t=p+1}^T X_t(p) X_t(p)' \right)^{-1}.$$

Let  $\hat{\Sigma}(p) \equiv (T-p)^{-1} \sum_{t=p+1}^T \hat{e}_t(p) \hat{e}_t(p)'$ , where  $\hat{e}_t(p) \equiv W_t - \hat{\beta}(p) X_t(p)$ . Define also

$$\hat{A}_{\cos}(\omega; p) \equiv \sum_{\ell=1}^p \hat{A}_\ell \cos(\omega \ell), \quad \hat{A}_{\sin}(\omega; p) \equiv \sum_{\ell=1}^p \hat{A}_\ell \sin(\omega \ell), \quad \omega \in [0, 2\pi].$$

The VAR( $p$ ) estimator of the parameter of interest  $\psi$  is then

$$\hat{\psi}(p) \equiv \int_0^{2\pi} h(\omega)' g(\hat{A}_{\cos}(\omega; p), \hat{A}_{\sin}(\omega; p), \hat{\Sigma}) d\omega.$$

The VAR lag length  $p = p_T$  must be chosen to grow with the sample size  $T$  at an appropriate rate, unless the true DGP is a finite-order VAR.

**Assumption B.3.**  $p_T \in \mathbb{N}$  is a deterministic function of the sample size  $T$  such that  $p_T^3/T \rightarrow 0$  and  $T^{1/2} \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| \rightarrow 0$  as  $T \rightarrow \infty$ .

These conditions are adopted from [Lewis & Reinsel \(1985, Thm. 2\)](#), see also [Berk \(1974\)](#). The last condition in [Assumption B.3](#) amounts to oversmoothing (i.e., choosing the lag length  $p$  so large that the variance dominates the mean square error), which ensures that the nonparametric bias does not show up in asymptotic limiting distributions. If the partial autocorrelations of the data decay exponentially fast with the lag length, [Assumption B.3](#) is satisfied by choosing  $p_T \propto T^\phi$  for any  $\phi \in (0, 1/3)$ . If the true DGP is a finite-order VAR, we may select  $p_T$  to be any constant greater than the true lag length.

## B.8.2 Main convergence results

We now state our main results on the asymptotic normality of the sieve VAR estimator and the consistency of the asymptotic variance estimator.

In preparation for stating our results, define for all  $T$  the vector  $\nu_T = (\nu'_{1,T}, \dots, \nu'_{p_T,T})' \in \mathbb{R}^{n_w^2 p_T}$ , where

$$\nu_{\ell,T} \equiv \int_0^{2\pi} h(\omega) \{g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \cos(\omega \ell) + g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \sin(\omega \ell)\} d\omega \in \mathbb{R}^{n_w^2}$$

for  $\ell = 1, 2, \dots, p_T$ . Define also

$$\xi \equiv \int_0^{2\pi} h(\omega) g_3(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) d\omega \in \mathbb{R}^{n_W^2}.$$

We also define the estimators  $\hat{\nu}_T$  and  $\hat{\xi}(p_T)$  of  $\nu_T$  and  $\xi$  obtained by substituting  $A_{\cos}(\cdot)$  and  $A_{\sin}(\cdot)$  with  $\hat{A}_{\cos}(\cdot; p_T)$  and  $\hat{A}_{\sin}(\cdot; p_T)$  in the above formulas. Finally, we define  $\Gamma(p) \equiv E(X_t(p)X_t(p)')$  for all  $p \in \mathbb{N}$  and the sample analogue  $\hat{\Gamma}(p) \equiv (T - p)^{-1} \sum_{t=p+1}^T X_t(p)X_t(p)'$ . In the rest of this section, all convergence statements are understood to be taken as  $T \rightarrow \infty$ .

Our first main proposition states that the sieve VAR estimator of the parameter of interest is asymptotically normal under our nonparametric conditions on the data generating process, the conditions on the estimated VAR lag order, and the regularity conditions on the parameter of interest.

**Proposition B.4.** *Let Assumptions B.1 to B.3 hold. Assume  $\sigma_\psi^2 \equiv \lim_{T \rightarrow \infty} \nu_T'(\Gamma(p_T)^{-1} \otimes \Sigma)\nu_T + \xi' \text{Var}(e_t \otimes e_t)\xi$  is strictly positive and that the limit exists. Then*

$$(T - p_T)^{1/2}(\hat{\psi}(p_T) - \psi) \xrightarrow{d} N(0, \sigma_\psi^2).$$

Under our regularity conditions on the parameter of interest, the convergence rate of the sieve VAR estimator  $\hat{\psi}(p_T)$  is  $(T - p_T)^{-1/2} = O(T^{-1/2})$ . The condition that  $\sigma_\psi^2$  exists and is nonzero rules out degenerate parameters that can be estimated super-consistently. This condition could for example be violated if the true parameter of interest is on the boundary of its parameter space (e.g., if the true FVD is 0, or the true degree of invertibility is 1). Such issues are not unique to SVMA-IV and could similarly arise in SVAR inference.

Our second main proposition states that the usual delta method standard errors for a VAR( $p_T$ ) model are valid asymptotically.

**Proposition B.5.** *Let the assumptions of Proposition B.4 hold. Let  $\hat{\sigma}_\psi^2(p_T) \equiv \hat{\nu}_T'(\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T))\hat{\nu}_T + \hat{\xi}(p_T)'\hat{\Xi}(p_T)\hat{\xi}(p_T)$ , where  $\hat{\chi}_t(p_T) \equiv \text{vec}(\hat{e}_t(p_T)\hat{e}_t(p_T)' - \hat{\Sigma}(p_T))$  and  $\hat{\Xi}(p_T) \equiv (T - p_T)^{-1} \sum_{t=p_T+1}^T \hat{\chi}_t(p_T)\hat{\chi}_t(p_T)'$ . Then*

$$\hat{\sigma}_\psi^2(p_T) \xrightarrow{p} \sigma_\psi^2.$$

Observe that  $\hat{\sigma}_\psi^2(p_T)$  is precisely the asymptotic variance estimator for  $\hat{\psi}(p_T)$  that one would compute from the delta method formula based on a VAR( $p_T$ ) model for the data.

To summarize, Propositions B.4 and B.5 imply that delta method inference based on

the estimated  $\text{VAR}(p_T)$  process is valid asymptotically even if the true DGP is a  $\text{VAR}(\infty)$ . Hence, the partial identification robust confidence intervals proposed in [Section 4](#) are valid under our regularity conditions. This conclusion is consistent with the finite-sample simulation evidence presented in [Section 7](#).

## B.9 Additional proofs and auxiliary lemmas

Here we prove [Lemma 1](#) and all additional results stated in this appendix. We first prove results related to the SVMA-IV identification analysis. Then we address the sieve VAR convergence results.

### B.9.1 Proof of [Lemma 1](#)

We focus on the semidefiniteness statement. Decompose  $B = B^{1/2}B^{1/2*}$  and define  $\tilde{b} = B^{-1/2}b$ . The statement of the lemma is equivalent with the statement that  $I_n - x^{-1}\tilde{b}\tilde{b}^*$  is positive semidefinite if and only if  $x \geq b^*b$ . Let  $\nu$  be an arbitrary  $n$ -dimensional complex vector satisfying  $\nu^*\nu = 1$ . Then

$$\nu^* \left( I_n - x^{-1}\tilde{b}\tilde{b}^* \right) \nu = 1 - \frac{\tilde{b}^*\tilde{b}}{x} \cos^2 \left( \theta(\nu, \tilde{b}) \right),$$

where  $\theta(\nu, \tilde{b})$  is the angle between  $\nu$  and  $\tilde{b}$ . Evidently,  $x^{-1}\tilde{b}^*\tilde{b} \leq 1$  is precisely the condition needed to ensure that the above display is nonnegative for every choice of  $\nu$ .  $\square$

### B.9.2 Auxiliary lemma for proof of [Proposition B.1](#)

**Lemma B.1.** *Let  $x_t$  and  $\tilde{x}_t$  be two stationary  $n$ -dimensional Gaussian time series whose spectral densities  $s_x(\omega)$  and  $s_{\tilde{x}}(\omega)$  are such that  $s_{\tilde{x}}(\omega) - s_x(\omega)$  is positive semidefinite for all  $\omega \in [0, 2\pi]$ . Then  $\text{Var}(\mu'x_{t+\ell} \mid \{x_\tau\}_{-\infty < \tau \leq t}) \leq \text{Var}(\mu'\tilde{x}_{t+\ell} \mid \{\tilde{x}_\tau\}_{-\infty < \tau \leq t})$  for all  $\ell = 1, 2, \dots$  and all constant vectors  $\mu \in \mathbb{R}^n$ .*

*Proof.* We may define an  $n$ -dimensional stationary Gaussian process  $\nu_t$  with spectral density  $s_\nu(\omega) = s_{\tilde{x}}(\omega) - s_x(\omega)$ ,  $\omega \in [0, 2\pi]$ , and such that the  $\nu_t$  process is independent of the  $x_t$  process. Then the process  $\check{x}_t = x_t + \nu_t$  has the same distribution as the  $\tilde{x}_t$  process. Hence,

$$\begin{aligned} \text{Var}(\mu'\tilde{x}_{t+\ell} \mid \{\tilde{x}_\tau\}_{-\infty < \tau \leq t}) &= \text{Var}(\mu'\check{x}_{t+\ell} \mid \{\check{x}_\tau\}_{-\infty < \tau \leq t}) \\ &\geq \text{Var}(\mu'\check{x}_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty < \tau \leq t}) \\ &= \text{Var}(\mu'x_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty < \tau \leq t}) + \text{Var}(\mu'\nu_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty < \tau \leq t}) \\ &\geq \text{Var}(\mu'x_{t+\ell} \mid \{x_\tau, \nu_t\}_{-\infty < \tau \leq t}) \\ &= \text{Var}(\mu'x_{t+\ell} \mid \{x_\tau\}_{-\infty < \tau \leq t}). \end{aligned}$$

The second equality above uses that the independence of the  $x_t$  and  $\nu_t$  processes implies that  $x_{t+\ell}$  and  $\nu_{t+\ell}$  are independent also conditional on  $\{x_\tau, \nu_t\}_{-\infty < \tau \leq t}$ .  $\square$

### B.9.3 Proof of Proposition B.1

The proof proceeds in two steps. First, for a given known  $\alpha$ , we show that  $FVD_{i,\ell}$  is sharply bounded above by 1 and below by (B.4). Second, we show that the lower bound is monotonically decreasing in  $\alpha$ , so that the overall lower bound is attained by  $\alpha_{UB}$ .

1. Given  $\alpha \in (\alpha_{LB}, \alpha_{UB}]$ , the numerator of  $FVD_{i,\ell}$  is point-identified (see below), so we need only concern ourselves with the denominator. We can write the denominator as

$$\begin{aligned} \text{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty < \tau \leq t}) &= \sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2 + \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2 \\ &= \frac{1}{\alpha^2} \sum_{m=0}^{\ell-1} \text{Cov}(y_{i,t}, \tilde{z}_{t-m})^2 + \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2. \end{aligned} \quad (\text{B.16})$$

Given  $\alpha$ , the first term in (B.16) is point-identified (note that it equals the numerator of the FVD), while the second is not. To upper-bound  $FVD_{i,\ell}$ , we seek to make that second term as small as possible. In fact, we can always set it to 0. To see this, let  $\{\Theta_{\bullet,j,m}\}_{2 \leq j \leq n_\varepsilon, 0 \leq m < \infty}$  denote some sequence of impulse responses for the structural shocks  $j \neq 1$  that is consistent with the second-moment properties of the data. Since  $\alpha \in (\alpha_{LB}, \alpha_{UB}]$ , such a sequence exists by Proposition 1. Now, for a given forecast horizon  $\ell$ , instead consider the new sequence  $\{\check{\Theta}_{\bullet,j,m}\}_{2 \leq j \leq n_\varepsilon, 0 \leq m < \infty}$ , defined via

$$\check{\Theta}_{\bullet,j,m} = \begin{cases} 0_{n_y \times 1} & \text{if } m \leq \ell - 1, \\ \Theta_{\bullet,j,m-\ell} & \text{if } m > \ell - 1. \end{cases}$$

Then the stochastic process induced by  $\{\check{\Theta}_{\bullet,j,m}\}_{2 \leq j \leq n_\varepsilon, 0 \leq m < \infty}$  has the exact same second-moment properties as the (by assumption admissible) stochastic process induced by  $\{\Theta_{\bullet,j,m}\}_{2 \leq j \leq n_\varepsilon, 0 \leq m < \infty}$ . However, by construction, we now have  $FVD_{i,\ell} = 1$ , as claimed.

For the lower bound, we want to make the second term in (B.16) as large as possible. Given a known  $\alpha \in (\alpha_{LB}, \alpha_{UB}]$ , define

$$\tilde{y}_t^{(\alpha)} = (\tilde{y}_{1,t}^{(\alpha)}, \dots, \tilde{y}_{n_y,t}^{(\alpha)})' \equiv y_t - \frac{1}{\alpha} \sum_{\ell=0}^{\infty} \text{Cov}(y_t, \tilde{z}_{t-\ell}) \varepsilon_{1,t-\ell} = \sum_{j=2}^{n_\varepsilon} \sum_{\ell=0}^{\infty} \Theta_{\bullet,j,\ell} \varepsilon_{j,t-\ell},$$

whose spectral density is given by the expression stated in the proposition. We have

$$\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty < \tau \leq t}) \geq \text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\varepsilon_{j,\tau}\}_{2 \leq j \leq n_\varepsilon, -\infty < \tau \leq t}) = \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2,$$

so the second term in (B.16) has an point-identified upper bound. Thus, given  $\alpha$ ,  $FVD_{i,\ell}$  is bounded below by the expression (B.4).

We now argue that the lower bound (B.4) is attained by an admissible model with the given  $\alpha$ . To that end, consider the Wold decomposition of  $\tilde{y}_t^{(\alpha)} = \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \tilde{\varepsilon}_{t-\ell}$ , where the  $\tilde{\Theta}_\ell$  matrices are  $n_y \times n_y$ , and  $\tilde{\varepsilon}_t$  is  $n_y$ -dimensional i.i.d. standard normal and spanned by  $\{\tilde{y}_\tau^{(\alpha)}\}_{-\infty < \tau \leq t}$ .<sup>B.17</sup> Then  $\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty < \tau \leq t}) = \sum_{j=2}^{n_\varepsilon} \sum_{m=0}^{\ell-1} \tilde{\Theta}_{i,j,m}^2$ , so the following model attains the lower bound (B.4) and is consistent with the given spectrum  $s_w(\cdot)$ :

$$\begin{aligned} y_t &= \frac{1}{\alpha} \sum_{\ell=0}^{\infty} \text{Cov}(y_t, \tilde{z}_{t-\ell}) \bar{\varepsilon}_{1,t} + \sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \tilde{\varepsilon}_{t-\ell}, \\ \tilde{z}_t &= \alpha \bar{\varepsilon}_{1,t} + \sqrt{\text{Var}(\tilde{z}_t) - \alpha^2} \times \bar{v}_t, \\ (\bar{\varepsilon}_{1,t}, \tilde{\varepsilon}_t', \bar{v}_t)' &\stackrel{i.i.d.}{\sim} N(0, I_{n_y+2}). \end{aligned} \tag{B.17}$$

2. **Lemma B.1** implies that  $\text{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty < \tau \leq t})$  is increasing in  $\alpha$ . Hence, the expression (B.4) is decreasing in  $\alpha$ , as claimed. At  $\alpha = \alpha_{UB}$ , the representation (B.17) has  $\tilde{z}_t = \alpha_{UB} \bar{\varepsilon}_{1,t}$ , so we can represent  $\tilde{y}_t^{(\alpha_{UB})} = y_t - E(y_t \mid \{\bar{\varepsilon}_{1,\tau}\}_{-\infty < \tau \leq t}) = y_t - E(y_t \mid \{\tilde{z}_\tau\}_{-\infty < \tau \leq t})$ .  $\square$

## B.9.4 Proof of Proposition B.2

The ‘‘only if’’ part was proved already in the text of [Appendix B.2](#). For the ‘‘if’’ part, assume that the cross-spectrum has the given factor structure. Since  $\tilde{z}_t$  is serially uncorrelated, we can write  $s_{\tilde{z}}(\cdot) = s_{\tilde{z}}$ . Because  $s_w(\omega)$  is positive definite, the Schur complement

$$s_{\tilde{z}} - s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega) = s_{\tilde{z}} - \eta \zeta(\omega)^* s_y(\omega)^{-1} \zeta(\omega) \eta'$$

---

<sup>B.17</sup>Since  $\alpha > \alpha_{LB}$ , the Wold decomposition has no deterministic term, cf. the proof of [Proposition 1](#).

is also positive definite. Pre-multiplying the above expression by  $\eta' s_{\bar{z}}^{-1}$ , post-multiplying by  $s_{\bar{z}}^{-1} \eta$ , and rearranging the positive definiteness condition, we obtain the implication that

$$2\pi \zeta(\omega)^* s_y(\omega)^{-1} \zeta(\omega) < \frac{2\pi}{\eta' s_{\bar{z}}^{-1} \eta}, \quad \omega \in [0, 2\pi].$$

Now choose any  $\bar{\alpha} \geq 0$  such that  $\bar{\alpha}^2$  lies strictly between the left- and right-hand sides in the above inequality. The matrix

$$\bar{\Sigma}_v \equiv 2\pi s_{\bar{z}} - \bar{\alpha}^2 \eta \eta'$$

is then positive definite by [Lemma 1](#). Moreover, the same lemma implies that

$$s_y(\omega) - \frac{2\pi}{\bar{\alpha}^2} \zeta(\omega) \zeta(\omega)^*$$

is positive definite for all  $\omega \in [0, 2\pi]$ . If we set  $\bar{\Theta}_{\bullet,1}(L) = (2\pi/\bar{\alpha})\zeta(L)$ , the same arguments as in the proof of [Proposition 1](#) show that there exists an  $n_y \times n_y$  matrix polynomial  $\tilde{\Theta}(L)$  such that the following model achieves the desired spectrum  $s_w(\omega)$ :

$$\begin{aligned} y_t &= \bar{\Theta}_{\bullet,1}(L) \bar{\varepsilon}_{1,t} + \tilde{\Theta}(L) \tilde{\varepsilon}_t, \\ \tilde{z}_t &= \bar{\alpha} \eta \bar{\varepsilon}_{1,t} + \bar{\Sigma}_v^{1/2} \bar{v}_t, \\ (\bar{\varepsilon}_{1,t}, \tilde{\varepsilon}'_t, \bar{v}'_t)' &\stackrel{i.i.d.}{\sim} N(0, I_{n_y+n_z+1}). \end{aligned}$$

Note that  $\eta$  assumes the role of  $\lambda$ . □

### B.9.5 Proof of [Proposition B.3](#)

According to the model [\(1\)](#), we can write

$$u_t = \sum_{\ell=0}^{\infty} M_{\ell} \varepsilon_{t-\ell},$$

for some  $n_y \times n_{\varepsilon}$  matrices  $\{M_{\ell}\}$ . Let  $M_{\bullet,j,\ell}$  denote the  $j$ -th column of  $M_{\ell}$ . Then

$$\tilde{\varepsilon}_{1,t} = \gamma' u_t = \sum_{j=1}^{n_{\varepsilon}} \sum_{\ell=0}^{\infty} a_{j,\ell} \varepsilon_{j,t-\ell},$$

where  $a_{j,\ell} = \gamma' M_{\bullet,j,\ell}$ . We have  $\text{Var}(\tilde{\varepsilon}_{1,t}) = 1$  by construction of  $\gamma$ , so  $\sum_{j=1}^{n_\varepsilon} \sum_{\ell=0}^{\infty} a_{j,\ell}^2 = 1$ . The expression for  $\tilde{\Theta}_{\bullet,1,\ell}$  in the proposition also immediately follows from the above display and the fact  $\text{Cov}(y_t, \varepsilon_{j,t-\ell}) = \Theta_{\bullet,j,\ell}$ . Next, observe that

$$\begin{aligned} R_0^2 &= \text{Var}(E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty < \tau \leq t})) \\ &= \text{Var}(E(\varepsilon_{1,t} \mid \{u_\tau\}_{-\infty < \tau \leq t})) \\ &= \text{Var}(E(\varepsilon_{1,t} \mid u_t)) \\ &= \text{Cov}(u_t, \varepsilon_{1,t})' \Sigma_u^{-1} \text{Cov}(u_t, \varepsilon_{1,t}) \\ &= M_{\bullet,1,0}' \Sigma_u^{-1} M_{\bullet,1,0}. \end{aligned}$$

Since  $\Sigma_{u\tilde{z}} = \sum_{\ell=0}^{\infty} M_\ell \text{Cov}(\varepsilon_{t-\ell}, \tilde{z}_t) = \alpha M_{\bullet,1,0}$ , we therefore have

$$\gamma = \frac{1}{\sqrt{\Sigma_{u\tilde{z}}' \Sigma_u^{-1} \Sigma_{u\tilde{z}}}} \Sigma_u^{-1} \Sigma_{u\tilde{z}} = \frac{1}{\sqrt{M_{\bullet,1,0}' \Sigma_u^{-1} M_{\bullet,1,0}}} \Sigma_u^{-1} M_{\bullet,1,0} = \frac{1}{\sqrt{R_0^2}} \Sigma_u^{-1} M_{\bullet,1,0}.$$

This implies

$$a_{1,0} = \gamma' M_{\bullet,1,0} = \sqrt{M_{\bullet,1,0}' \Sigma_u^{-1} M_{\bullet,1,0}} = \sqrt{R_0^2}.$$

Finally,

$$\tilde{\Theta}_{\bullet,1,0} = \text{Cov}(y_t, \tilde{\varepsilon}_{1,t}) = \text{Cov}(y_t, u_t) \gamma = \text{Cov}(u_t, u_t) \gamma = \Sigma_u \gamma = \frac{1}{\sqrt{R_0^2}} M_{\bullet,1,0},$$

and  $M_{\bullet,1,0} = \text{Cov}(u_t, \varepsilon_{1,t}) = \text{Cov}(y_t, \varepsilon_{1,t}) = \Theta_{\bullet,1,0}$ . □

### B.9.6 Auxiliary lemmas for sieve VAR results

Here we define notation and state auxiliary lemmas used to prove the propositions in [Appendix B.8](#). The lemmas are proved below. For any matrix  $B$ , let  $\|B\|_1$  denote the largest singular value of  $B$ . Recall that  $\|B\|_1 \leq \|B\|$  and  $\|BC\| \leq \|B\| \|C\|_1$  for conformable matrices  $B$  and  $C$ . Let  $e_t(p) \equiv W_t - \beta(p) X_t(p)$  for all  $t$  and  $p$ . Finally, define

$$A_{\cos}(\omega; p) \equiv \sum_{\ell=1}^p A_\ell \cos(\omega \ell), \quad A_{\sin}(\omega; p) \equiv \sum_{\ell=1}^p A_\ell \sin(\omega \ell), \quad \omega \in [0, 2\pi], \quad p \in \mathbb{N}.$$

**Lemma B.2** (Lewis & Reinsel, 1985, p. 397). *Let Assumptions B.1 and B.3 hold. Then  $E(\|\hat{\Gamma}(p_T) - \Gamma(p_T)\|^2) = O(p_T^2/T)$ .*

**Lemma B.3.** *Let Assumptions B.1 and B.3 hold. Then  $\|\hat{\beta}(p_T) - \beta(p_T)\| = O_p((p_T/T)^{1/2})$ .*

**Lemma B.4.** *Let Assumptions B.1 and B.3 hold. Then  $\hat{\Sigma}(p_T) - (T - p_T)^{-1} \sum_{t=p_T+1}^T e_t e_t' = o_p(T^{-1/2})$ .*

**Lemma B.5** (Lewis & Reinsel, 1985, Thm. 2). *Let Assumptions B.1 and B.3 hold. Let  $\tilde{\nu}_T \in \mathbb{R}^{n_W^{p_T}}$  be a deterministic sequence of vectors such that  $\|\tilde{\nu}_T\|^2 \leq M < \infty$  for all  $T$ . Define*

$$\zeta_T \equiv (T - p_T)^{-1/2} \sum_{t=p_T+1}^T \tilde{\nu}_T' (\Gamma(p_T)^{-1} X_t(p_T) \otimes e_t).$$

Then

$$(T - p_T)^{1/2} \tilde{\nu}_T' \text{vec}(\hat{\beta}(p_T) - \beta(p_T)) - \zeta_T \xrightarrow{p} 0.$$

**Lemma B.6.** *Let Assumption B.1 hold. Then for all  $j_1, j_2, j_3, j_4 \in \{1, 2, \dots, n_W\}$ , all  $p, T \in \mathbb{N}$  such that  $p < T$ , and all  $m_1, m_2, m_3, m_4 \in \mathbb{Z}$  we have*

$$\frac{1}{T - p} \sum_{t=p+1}^T \sum_{s=p+1}^T |\text{Cov}(e_{j_1, t+m_1} e_{j_2, t+m_2} e_{j_3, t} e_{j_4, t}, e_{j_1, s+m_3} e_{j_2, s+m_4} e_{j_3, s} e_{j_4, s})| \leq 9E\|e_t\|^8.$$

**Lemma B.7.** *Let Assumptions B.1 and B.3 hold. Then*

$$\left\| \frac{1}{T - p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')' - E[\text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')'] \right\|^2 = O_p(p_T^2/T),$$

and

$$\left\| \frac{1}{T - p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t e_t' - \Sigma)' \right\|^2 = O_p(p_T/T).$$

**Lemma B.8.** *Let Assumptions B.1 and B.3 hold. Define a sequence  $\tilde{\nu}_T$  as in Lemma B.5, and assume  $v_\zeta \equiv \lim_{T \rightarrow \infty} \tilde{\nu}_T' (\Gamma(p_T)^{-1} \otimes \Sigma) \tilde{\nu}_T$  exists. Then*

$$(T - p_T)^{1/2} \tilde{\nu}_T' \text{vec}(\hat{\beta}(p_T) - \beta(p_T)) \xrightarrow{d} N(0, v_\zeta),$$

$$(T - p_T)^{1/2} \text{vec}(\hat{\Sigma}(p_T) - \Sigma) \xrightarrow{d} N(0, \text{Var}(e_t \otimes e_t)),$$

and these two random vectors are asymptotically independent.

**Lemma B.9.** *Let Assumptions B.1 and B.3 hold. Then*

$$\sup_{\omega \in [0, 2\pi]} \left( \|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T)\|^2 + \|\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega; p_T)\|^2 \right) = O_p(p_T/T).$$

**Lemma B.10.** *Let Assumptions B.1 and B.3 hold. For  $M > 0$ , define  $\mathcal{A}_M^0 \equiv \{(B_1, B_2) \in \mathcal{A}_\delta \times \mathbb{R}^{n_W \times n_W} : \|B_j - \sum_{\ell=1}^{\infty} A_\ell\| \leq M, j = 1, 2\}$  and  $\mathcal{S}_M^0 = \{\tilde{\Sigma} \in \mathbb{S}_{n_W} : \|\tilde{\Sigma} - \Sigma\| \leq M\}$ . Then there exists an  $M < \infty$  such that*

$$P \left( (\hat{A}_{\cos}(\omega; p_T), \hat{A}_{\sin}(\omega; p_T)) \in \mathcal{A}_M^0 \text{ for all } \omega \in [0, 2\pi], \hat{\Sigma}(p_T) \in \mathcal{S}_M^0 \right) \rightarrow 1.$$

**Lemma B.11.** *Let Assumptions B.1 to B.3 hold. Define  $\nu_T$  and  $\xi$  as in Appendix B.8.2. Then*

$$(T - p_T)^{1/2} \left\{ (\hat{\psi}(p_T) - \psi) - \nu_T' \text{vec}(\hat{\beta}(p_T) - \beta(p_T)) - \xi' \text{vec}(\hat{\Sigma} - \Sigma) \right\} \xrightarrow{p} 0.$$

### B.9.7 Proof of Lemma B.3

The result follows almost directly from the proof of Thm. 1 in Lewis & Reinsel (1985). As in that proof, define

$$U_{1,T} \equiv \frac{1}{T - p_T} \sum_{t=p_T+1}^T (e_t(p_T) - e_t) X_t(p_T)', \quad U_{2,T} \equiv \frac{1}{T - p_T} \sum_{t=p_T+1}^T e_t X_t(p_T)'$$

Lewis & Reinsel's arguments show that  $\|U_{1,T}\| = O_p(p_T^{1/2} \sum_{\ell=p_T+1}^{\infty} \|A_\ell\|) = o_p((p_T/T)^{1/2})$  and  $\|U_{2,T}\| = O_p((p_T/T)^{1/2})$  under Assumptions B.1 and B.3. The rest of the arguments in Lewis & Reinsel's proof now yields the desired convergence rate of  $\hat{\beta}(p_T)$ .  $\square$

### B.9.8 Proof of Lemma B.4

Recall the notation  $U_{1,T}$  and  $U_{2,T}$  in the proof of Lemma B.3. Since

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{T - p_T} \sum_{t=p_T+1}^T e_t e_t' + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t) e_t' \\ &\quad + \frac{1}{T - p_T} \sum_{t=p_T+1}^T e_t (\hat{e}_t - e_t)' + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t) (\hat{e}_t - e_t)' \end{aligned}$$

$$\equiv \frac{1}{T - p_T} \sum_{t=p_T+1}^T e_t e_t' + R_{1,T} + R_{1,T}' + R_{2,T},$$

we need to show  $R_{1,T} = o_p(T^{-1/2})$  and  $R_{2,T} = o_p(T^{-1/2})$ .

Decompose  $R_{1,T}$  as

$$R_{1,T} = \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t(p_T)) e_t' + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (e_t(p_T) - e_t) e_t' \equiv \tilde{R}_{1,T} + \tilde{R}_{2,T}.$$

Since  $\hat{e}_t(p_T) - e_t(p_T) = (\hat{\beta}(p_T) - \beta(p_T)) X_t(p_T)$ , we have

$$\|\tilde{R}_{1,T}\| \leq \|\hat{\beta}(p_T) - \beta(p_T)\| \|U_{2,T}\| = O_p((p_T/T)^{1/2}) O_p((p_T/T)^{1/2}) = o(T^{-1/2}).$$

Moreover, since  $e_t - e_t(p_T) = \sum_{\ell=p_T+1}^{\infty} A_\ell W_{t-\ell}$ ,

$$\begin{aligned} E\|\tilde{R}_{2,T}\| &\leq \frac{1}{T - p_T} \sum_{t=p_T+1}^T \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| E(\|W_{t-\ell} e_t'\|) \\ &\leq \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| (E\|W_t\|^2 E\|e_t\|^2)^{1/2} \\ &= \text{constant} \times \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| \\ &= o_p(T^{-1/2}). \end{aligned}$$

Now decompose  $R_{2,T}$  as

$$\begin{aligned} \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t)(\hat{e}_t - e_t)' &= \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t(p_T))(\hat{e}_t - e_t(p_T))' \\ &\quad + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (\hat{e}_t - e_t(p_T))(e_t(p) - e_t)' \\ &\quad + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (e_t(p) - e_t)(\hat{e}_t - e_t(p_T))' \\ &\quad + \frac{1}{T - p_T} \sum_{t=p_T+1}^T (e_t(p_T) - e_t)(e_t(p_T) - e_t)' \end{aligned}$$

$$\equiv \hat{R}_{1,T} + \hat{R}_{2,T} + \hat{R}'_{2,T} + \hat{R}_{3,T}.$$

We have

$$\begin{aligned} \|\hat{R}_{1,T}\| &\leq \|\hat{\beta}(p_T) - \beta(p_T)\|^2 \|\hat{\Gamma}(p_T)\|_1 \\ &\leq \|\hat{\beta}(p_T) - \beta(p_T)\|^2 (\|\hat{\Gamma}(p_T) - \Gamma(p_T)\|_1 + \|\Gamma(p_T)\|_1) \\ &= O_p(p_T/T), \end{aligned}$$

using [Lemma B.2](#) and [Lemma B.3](#). Further,

$$\|\hat{R}_{2,T}\| \leq \|\hat{\beta}(p_T) - \beta(p_T)\| \|U_{1,T}\| = O_p((p_T/T)^{1/2}) o_p((p/T)^{1/2}) = o_p(T^{-1/2}).$$

Finally,

$$\begin{aligned} E\|\hat{R}_{3,T}\| &\leq E\|e_t(p_T) - e_t\|^2 \\ &\leq \sum_{\ell=p_T+1}^{\infty} \sum_{m=p_T+1}^{\infty} \|A_\ell\| \|A_m\| E(\|W_{t-\ell}\| \|W_{t-m}\|) \\ &\leq \text{constant} \times \left( \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| \right)^2 \\ &= o(T^{-1}). \quad \square \end{aligned}$$

### B.9.9 Proof of [Lemma B.6](#)

By stationarity,

$$\begin{aligned} &\frac{1}{T-p} \sum_{t=p+1}^T \sum_{s=p+1}^T \left| \text{Cov}(e_{j_1,t+m_1} e_{j_2,t+m_2} e_{j_3,t} e_{j_4,t}, e_{j_1,s+m_3} e_{j_2,s+m_4} e_{j_3,s} e_{j_4,s}) \right| \\ &= \sum_{\ell=-(T-p-1)}^{T-p-1} \left( 1 - \frac{|\ell|}{T-p} \right) \left| \text{Cov}(e_{j_1,\ell+m_1} e_{j_2,\ell+m_2} e_{j_3,\ell} e_{j_4,\ell}, e_{j_1,m_3} e_{j_2,m_4} e_{j_3,0} e_{j_4,0}) \right|. \end{aligned} \quad (\text{B.18})$$

We first argue that each term in the sum [\(B.18\)](#) is bounded. This follows from Cauchy-Schwarz:

$$\begin{aligned} &\left| \text{Cov}(e_{j_1,\ell+m_1} e_{j_2,\ell+m_2} e_{j_3,\ell} e_{j_4,\ell}, e_{j_1,m_3} e_{j_2,m_4} e_{j_3,0} e_{j_4,0}) \right| \\ &\leq (\text{Var}(e_{j_1,\ell+m_1} e_{j_2,\ell+m_2} e_{j_3,\ell} e_{j_4,\ell}) \text{Var}(e_{j_1,m_3} e_{j_2,m_4} e_{j_3,0} e_{j_4,0}))^{1/2} \\ &\leq \max_{1 \leq j \leq n_W} E(e_{j,t}^8) \end{aligned}$$

$$\leq E\|e_t\|^8.$$

Next, we show that at most 9 of the terms in the sum (B.18) are nonzero. Consider the term corresponding to a given index  $\ell$  in the sum. For the covariance in the term to be nonzero, it must be the case that  $\{\ell + m_1, \ell + m_2, \ell\} \cap \{m_3, m_4, 0\} \neq \emptyset$  (otherwise the two variables in the covariance would be independent). At most 9 values of  $\ell$  have this property.

Putting the preceding two results together, we obtain the statement of the lemma.  $\square$

### B.9.10 Proof of Lemma B.7

We first remark that Assumption B.1 implies  $\{W_t\}$  is a strictly non-deterministic time series with Wold innovation  $e_t$ . Thus, the Wold representation  $W_t = B(L)e_t$  has  $B(L) = \sum_{\ell=0}^{\infty} B_\ell L^\ell = A(L)^{-1}$ , and so for fixed  $i, j$ , the elements  $B_{i,j,\ell}$  of  $B_\ell$  are absolutely summable across  $\ell$  (Brockwell & Davis, 1991, p. 418).

Define the  $n_W^2 p_T \times n_W^2 p_T$  matrix

$$R_{1,T} \equiv \frac{1}{T - p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')' - E[\text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')']$$

with elements  $R_{1,T,i,j}$ . Then  $\|R_{1,T}\|^2 = \sum_{i,j=1}^{n_W^2 p_T} R_{1,T,i,j}^2$ , and the first statement of the lemma follows if we can show that  $E(R_{1,T,i,j}^2) = O(T^{-1})$  uniformly in  $i, j$ . Since  $E(R_{1,T,i,j}) = 0$  for all  $i, j$ , we need to show that  $\text{Var}(R_{1,T,i,j}) = O(T^{-1})$  uniformly in  $i, j$ . The typical element  $R_{1,T,i,j}$  has the form

$$\frac{1}{T - p_T} \sum_{t=p_T+1}^T e_{j_1,t} W_{j_2,t-m_1} e_{j_3,t} W_{j_4,t-m_2} - E[e_{j_1,t} W_{j_2,t-m_1} e_{j_3,t} W_{j_4,t-m_2}]$$

for appropriate  $j_1, j_2, j_3, j_4, m_1, m_2 \in \mathbb{N}$ . Here  $W_{j,t}$  is the  $j$ -th element of  $W_t$ , and similarly for  $e_t$ . The variance of the above expression is given by

$$\frac{1}{(T - p_T)^2} \sum_{t=p_T+1}^T \sum_{s=p_T+1}^T \text{Cov}(e_{j_1,t} W_{j_2,t-m_1} e_{j_3,t} W_{j_4,t-m_2}, e_{j_1,s} W_{j_2,s-m_1} e_{j_3,s} W_{j_4,s-m_2}). \quad (\text{B.19})$$

Using the above-mentioned Wold decomposition of  $\{W_t\}$ , we can write

$$W_{j_2,t-m_1} = \sum_{b_1=1}^{n_W} \sum_{\ell_1=0}^{\infty} B_{j_2,b_1,\ell_1} e_{b_1,t-m_1-\ell_1},$$

say. Hence, the expression (B.19) equals

$$\begin{aligned} & \frac{1}{T-p_T} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=0}^{\infty} \sum_{b_1, b_2, b_3, b_4=1}^{n_W} B_{j_2, b_1, \ell_1} B_{j_4, b_2, \ell_2} B_{j_2, b_3, \ell_3} B_{j_4, b_3, \ell_4} \\ & \times \frac{1}{T-p_T} \sum_{s, t=p_T+1}^T \text{Cov}(e_{j_1, t} e_{b_1, t-m_1-\ell_1} e_{j_3, t} e_{b_2, t-m_2-\ell_2}, e_{j_1, s} e_{b_3, s-m_1-\ell_3} e_{j_3, s} e_{b_4, s-m_2-\ell_4}). \end{aligned}$$

According to Lemma B.6, the above display is bounded by

$$\frac{1}{T-p_T} \sum_{\ell_1, \ell_2, \ell_3, \ell_4=0}^{\infty} \sum_{b_1, b_2, b_3, b_4=1}^{n_W} |B_{j_2, b_1, \ell_1} B_{j_4, b_2, \ell_2} B_{j_2, b_3, \ell_3} B_{j_4, b_4, \ell_4}| \times 9E\|e_t\|^8 = O(T^{-1}), \quad (\text{B.20})$$

where the equality uses the previously-mentioned absolute summability of  $\{B_\ell\}$ . This concludes the proof of the first statement of the lemma.

We prove the second statement of the lemma in a similar fashion. Define the  $n_W^2 p_T \times n_W^2 p_T$  matrix

$$R_{2,T} \equiv \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t e_t' - \Sigma)'$$

Decompose it as

$$\begin{aligned} R_{2,T} &= \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t e_t')' - \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(\Sigma)' \\ &\equiv \tilde{R}_{1,T} - \tilde{R}_{2,T}. \end{aligned}$$

Since  $\{\text{vec}(e_t X_t(p_T)')\}$  is a serially uncorrelated  $(n_W p_T)$ -dimensional sequence, it is easy to show that  $E\|\tilde{R}_{2,T}\|^2 = O_p(p_T/T)$ . Consider now the matrix  $\tilde{R}_{1,T}$ . Its typical element

$$(T-p_T)^{-1} \sum_{t=p_T+1}^T e_{j_1, t} W_{j_2, t-m} e_{j_3, t} e_{j_4, t}$$

has mean zero due to the independence of  $e_t$  and  $W_{t-m}$  for  $m \geq 1$ . We need to show that it has variance of order  $O(T^{-1})$ . Said variance equals

$$\frac{1}{(T-p_T)^2} \sum_{s, t=p_T+1}^T \text{Cov}(e_{j_1, t} W_{j_2, t-m} e_{j_3, t} e_{j_4, t}, e_{j_1, s} W_{j_2, s-m} e_{j_3, s} e_{j_4, s})$$

$$\begin{aligned}
&= \frac{1}{T - p_T} \sum_{\ell_1, \ell_2=0}^{\infty} \sum_{b_1, b_2=1}^{n_W} B_{j_2, b_1, \ell_1} B_{j_2, b_2, \ell_2} \\
&\quad \times \frac{1}{T - p_T} \sum_{s, t=p_T+1}^T \text{Cov}(e_{j_1, t} e_{b_1, t-m-\ell_1} e_{j_3, t} e_{j_4, t}, e_{j_1, s} e_{b_2, s-m-\ell_2} e_{j_3, s} e_{j_4, s}).
\end{aligned}$$

This expression is of order  $O(T^{-1})$ , for the same reason as (B.20) above.  $\square$

### B.9.11 Proof of Lemma B.8

This result is very similar to Thm. 2 in Lewis & Reinsel (1985), with the twist that we here deal also with the convergence of  $\hat{\Sigma}$ . Define  $v_{\zeta, T} \equiv \tilde{\nu}'_T (\Gamma(p_T)^{-1} \otimes \Sigma) \tilde{\nu}_T$  for all  $T$ . If  $v_{\zeta} \equiv \lim_{T \rightarrow \infty} v_{\zeta, T} = 0$ , it is easy to show that  $(T - p_T)^{1/2} \tilde{\nu}'_T \text{vec}(\hat{\beta}(p_T) - \beta(p_T)) = o_p(1)$  using Lemma B.5 and an mean-square bound, so in the following we assume  $v_{\zeta} > 0$ . By Lemma B.5 and the Cramér-Wold device, we need to show that, for any  $\lambda \in \mathbb{R}^{n_W^2}$ ,

$$\sum_{t=p_T+1}^T J_{t, T} \xrightarrow{d} N(0, 1),$$

where we define the triangular array

$$J_{t, T} \equiv \frac{\tilde{\nu}'_T (\Gamma(p_T)^{-1} X_t(p_T) \otimes e_t) + \lambda' \text{vec}(e_t e_t' - \Sigma)}{(T - p_T)^{1/2} (v_{\zeta, T} + \lambda' \text{Var}(e_t \otimes e_t) \lambda)^{1/2}}, \quad t = p_T + 1, \dots, T, \quad T \in \mathbb{N}.$$

Since  $\{e_t\}$  is i.i.d.,  $e_t$  is independent of  $X_t(p_T)$ , so  $\{J_{t, T}\}_{p_T+1 \leq t \leq T}$  is a martingale difference sequence with respect to the filtration generated by  $\{e_t\}$ . Also, since  $E[X_t(p_T)] = 0$ , we have  $E(J_{t, T}^2) = (T - p_T)^{-1}$ . The statement of the lemma then follows from Davidson (1994, Thm. 24.3) if we can show

$$\sum_{t=p_T+1}^T J_{t, T}^2 \xrightarrow{p} 1 \tag{B.21}$$

and

$$\max_{p_T+1 \leq t \leq T} |J_{t, T}| \xrightarrow{p} 0. \tag{B.22}$$

We first prove (B.21), following the univariate argument in Gonçalves & Kilian (2007, pp. 633–636). Decompose

$$\sum_{t=p_T+1}^T J_{t, T}^2 - 1 = \{v_{\zeta, T} + \lambda' \text{Var}(e_t \otimes e_t) \lambda\}^{-1}$$

$$\begin{aligned}
& \times \left\{ \frac{1}{T-p_T} \sum_{t=p_T+1}^T \left[ (\tilde{\nu}'_T(\Gamma(p_T)^{-1}X_t(p_T) \otimes e_t))^2 - v_{\zeta,T} \right] \right. \\
& \quad + \frac{2}{T-p_T} \sum_{t=p_T+1}^T \tilde{\nu}'_T(\Gamma(p_T)^{-1}X_t(p_T) \otimes e_t) \text{vec}(e_t e'_t - \Sigma)' \lambda \\
& \quad \left. + \frac{1}{T-p_T} \sum_{t=p_T+1}^T \left[ (\lambda' \text{vec}(e_t e'_t - \Sigma))^2 - \lambda' \text{Var}(e_t \otimes e_t) \lambda \right] \right\} \\
& \equiv \{v_{\zeta,T} + \lambda' \text{Var}(e_t \otimes e_t) \lambda\}^{-1} \{R_{1,T} + 2R_{2,T} + R_{3,T}\}.
\end{aligned}$$

The i.i.d. law of large numbers implies that  $R_{3,T} = o_p(1)$ . We now show that also  $R_{1,T} = o_p(1)$  and  $R_{2,T} = o_p(1)$ . First,

$$\begin{aligned}
& |R_{1,T}| \\
& = \left| \tilde{\nu}'_T(\Gamma(p_T)^{-1} \otimes I_{n_w}) \left\{ \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')' \right. \right. \\
& \quad \left. \left. - E[\text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')'] \right\} (\Gamma(p_T)^{-1} \otimes I_{n_w}) \tilde{\nu}_T \right| \\
& \leq \|\tilde{\nu}_T\|^2 \|\Gamma(p_T)^{-1}\|_1^2 \\
& \quad \times \left\| \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')' - E[\text{vec}(e_t X_t(p_T)') \text{vec}(e_t X_t(p_T)')'] \right\| \\
& = o_p(1),
\end{aligned}$$

where the last line follows from [Lemma B.7](#) and [Assumptions B.1](#) and [B.3](#). Second, we analogously have

$$\begin{aligned}
|R_{2,T}| & \leq \|\tilde{\nu}_T\| \|\lambda\| \|\Gamma(p_T)^{-1}\|_1 \left\| \frac{1}{T-p_T} \sum_{t=p_T+1}^T \text{vec}(e_t X_t(p_T)') \text{vec}(e_t e_t - \Sigma)' \right\| \\
& = o_p(1),
\end{aligned}$$

again using [Lemma B.7](#) and [Assumptions B.1](#) and [B.3](#). This concludes the proof of [\(B.21\)](#).

To prove [\(B.22\)](#), first note that since  $E\|e_t\|^{4+\epsilon} < \infty$  for some  $\epsilon > 0$ , a standard argument for i.i.d. variables gives that  $(T-p_T)^{-1/2} \max_{p_T+1 \leq t \leq T} |\lambda' \text{vec}(e_t e'_t - \Sigma)| = o_p(1)$ . Next, the

same calculations as in equation (2.12) in [Lewis & Reinsel \(1985, p. 401\)](#) yield

$$\begin{aligned}
& P \left( \max_{p_T+1 \leq t \leq T} \frac{(\tilde{\nu}'_T (\Gamma(p_T)^{-1} X_t(p_T) \otimes e_t))^2}{T - p_T} \geq \tilde{\epsilon} \right) \\
& \leq \frac{1}{\tilde{\epsilon}^2} \frac{p_T^2}{(T - p_T)} \|\tilde{\nu}_T\|^4 \|\Gamma(p_T)^{-1}\|_1^4 E\|e_t\|^4 E\|W_t\|^4 \\
& \rightarrow 0
\end{aligned}$$

for any  $\tilde{\epsilon} > 0$ . Putting the previous two facts together, we obtain [\(B.22\)](#).  $\square$

### B.9.12 Proof of [Lemma B.9](#)

For any  $\omega \in [0, 2\pi]$ ,

$$\begin{aligned}
\|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T)\|^2 &= \sum_{\ell=1}^{p_T} \|\hat{A}_\ell - A_\ell\|^2 \cos^2(\omega\ell) \\
&\leq \sum_{\ell=1}^{p_T} \|\hat{A}_\ell - A_\ell\|^2 \\
&= \|\hat{\beta}(p_T) - \beta(p_T)\|^2 \\
&= O(p_T/T),
\end{aligned}$$

using [Lemma B.3](#). The argument for  $A_{\sin}$  is identical.  $\square$

### B.9.13 Proof of [Lemma B.10](#)

We start off by showing that the estimated VAR spectrum is nonsingular, asymptotically. Extend the definition of the Frobenius norm to complex matrices, so  $\|B\|^2 \equiv \text{tr}(B^*B)$ . The matrix perturbation bound  $|\det(B) - \det(C)| \leq n\|C - B\| \max\{\|B\|, \|C\|\}^{n-1}$  for  $n \times n$  complex matrices  $B$  and  $C$  ([Bhatia, 1997](#), Problem I.6.11, p. 22) implies

$$\begin{aligned}
& \left| \det(A(e^{i\omega})) - \det(I_{n_W} - \hat{A}_{\cos}(\omega; p_T) - i\hat{A}_{\sin}(\omega; p_T)) \right| \\
& \leq n_W \left\| \sum_{\ell=p_T+1}^{\infty} A_\ell e^{i\omega} - \sum_{\ell=1}^{p_T} (\hat{A}_\ell - A_\ell) e^{i\omega} \right\| \max \left\{ \left\| \sum_{\ell=1}^{\infty} A_\ell e^{i\omega} \right\|, \left\| \sum_{\ell=1}^{p_T} \hat{A}_\ell e^{i\omega} \right\| \right\}^{n_W-1}. \quad (\text{B.23})
\end{aligned}$$

[Lemma B.9](#) implies

$$\sup_{\omega \in [0, 2\pi]} \left\| \sum_{\ell=1}^{p_T} (\hat{A}_\ell - A_\ell) e^{i\omega} \right\| = o_p(1).$$

By [Assumptions B.1](#) and [B.3](#), the right-hand side of [\(B.23\)](#) therefore tends to 0 in probability uniformly in  $\omega$ , implying

$$\inf_{\omega \in [0, 2\pi]} \left| \det(I_{n_W} - \hat{A}_{\cos}(\omega; p_T) - i\hat{A}_{\sin}(\omega; p_T)) \right| = \inf_{\omega \in [0, 2\pi]} |\det(A(e^{i\omega}))| + o_p(1) > \delta + o_p(1).$$

Thus, with probability approaching 1,

$$(\hat{A}_{\cos}(\omega; p_T), \hat{A}_{\sin}(\omega; p_T)) \in \mathcal{A}_\delta \quad \text{for all } \omega \in [0, 2\pi].$$

We now show that, asymptotically, the estimated VAR spectrum lies in a region where  $g(\cdot)$  is smooth. Let  $M \equiv \max\{2 \sum_{\ell=1}^{\infty} \|A_\ell\|, \|\Sigma\|\} + 1$ . By [Assumption B.2](#),  $g(\cdot, \cdot, \cdot)$  is continuously differentiable on  $\mathcal{A}_M^0 \times \mathcal{S}_M^0$ . Since

$$\left\| \hat{A}_{\cos}(\omega; p_T) - \sum_{\ell=1}^{\infty} A_\ell \right\| \leq \left\| \hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T) \right\| + 2 \sum_{\ell=1}^{\infty} \|A_\ell\| = 2 \sum_{\ell=1}^{\infty} \|A_\ell\| + o_p(1)$$

uniformly in  $\omega$  by [Lemma B.9](#) and [Assumption B.3](#) (and similarly for sin instead of cos), it follows that, with probability approaching 1,

$$(\hat{A}_{\cos}(\omega; p_T), \hat{A}_{\sin}(\omega; p_T)) \in \mathcal{A}_M^0 \quad \text{for all } \omega \in [0, 2\pi].$$

Moreover, by the law of large numbers for i.i.d. variables and [Lemma B.4](#), we also have  $\hat{\Sigma}(p_T) \in \mathcal{S}_M^0$  with probability approaching 1.  $\square$

### B.9.14 Proof of [Lemma B.11](#)

We start out by applying a first-order Taylor expansion to the parameter of interest  $\psi$ . By [Lemma B.10](#) and [Assumption B.2](#), we can write

$$\begin{aligned} & g(\hat{A}_{\cos}(\omega; p_T), \hat{A}_{\sin}(\omega; p_T), \hat{\Sigma}) - g(A_{\cos}(\omega; p_T), A_{\sin}(\omega; p_T), \Sigma) \\ &= g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)) \\ & \quad + g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega)) \\ & \quad + g_3(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{\Sigma} - \Sigma) \\ & \quad + \hat{R}_T(\omega), \end{aligned}$$

where the fact that  $g(\cdot, \cdot, \cdot)$  is twice continuously differentiable implies that there exists a  $C > 0$  such that the remainder satisfies

$$|\hat{R}_T(\omega)| \leq C \left( \|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)\|^2 + \|\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega)\|^2 + \|\hat{\Sigma} - \Sigma\|^2 \right)$$

for all  $\omega$ , with probability approaching 1. Since

$$\|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)\| \leq \sum_{\ell=p_T+1}^{\infty} \|A_{\ell}\| + \|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T)\| = O_p((p_T/T)^{1/2})$$

by [Lemma B.9](#) and [Assumption B.3](#) (and similarly with sin instead of cos), and since  $\|\hat{\Sigma} - \Sigma\| = O_p(T^{-1/2})$  by [Lemma B.4](#), we obtain

$$\int_0^{2\pi} |\hat{R}_T(\omega)| d\omega = O_p(p_T/T).$$

Using the continuity and thus boundedness of  $h(\cdot)$ , we therefore get

$$\begin{aligned} \hat{\psi}(p_T) - \psi &= \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)) d\omega \\ &\quad + \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega)) d\omega \\ &\quad + \int_0^{2\pi} h(\omega) g_3(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{\Sigma} - \Sigma) d\omega \\ &\quad + O_p(p_T/T) \\ &= \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T)) d\omega \\ &\quad + \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega; p_T)) d\omega \\ &\quad + \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(A_{\cos}(\omega; p_T) - A_{\cos}(\omega)) d\omega \quad (\text{B.24}) \end{aligned}$$

$$\begin{aligned} &\quad + \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(A_{\sin}(\omega; p_T) - A_{\sin}(\omega)) d\omega \quad (\text{B.25}) \\ &\quad + \xi' \text{vec}(\hat{\Sigma} - \Sigma) \\ &\quad + O_p(p_T/T). \end{aligned}$$

We now bound the nonparametric bias term (B.24); the argument for (B.25) is similar. Note that  $h(\cdot)$  is bounded, and

$$\begin{aligned}
& \int_0^{2\pi} \|g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(A_{\cos}(\omega; p_T) - A_{\cos}(\omega))\| d\omega \\
& \leq \int_0^{2\pi} \|g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)\| d\omega \times \sup_{\omega \in [0, 2\pi]} \|A_{\cos}(\omega; p_T) - A_{\cos}(\omega)\| \\
& \leq \int_0^{2\pi} \|g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)\| d\omega \times \sum_{\ell=p_T+1}^{\infty} \|A_{\ell}\| \\
& = o(T^{-1/2}),
\end{aligned}$$

by **Assumption B.3**. We also used that **Assumption B.2** implies  $\omega \mapsto \|g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)\|$  is in  $L_2(0, 2\pi)$ , implying that this function is integrable. Thus, the terms (B.24)–(B.25) are each  $o(T^{-1/2})$ .

To complete the proof, observe that

$$\begin{aligned}
& \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega; p_T)) d\omega \\
& \quad + \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \text{vec}(\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega; p_T)) d\omega \\
& = \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \sum_{\ell=1}^{p_T} \text{vec}(\hat{A}_{\ell} - A_{\ell}) \cos(\omega\ell) d\omega \\
& \quad + \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \sum_{\ell=1}^{p_T} \text{vec}(\hat{A}_{\ell} - A_{\ell}) \sin(\omega\ell) d\omega \\
& = \sum_{\ell=1}^{p_T} \nu'_{\ell, T} \text{vec}(\hat{A}_{\ell} - A_{\ell}).
\end{aligned}$$

In conclusion,

$$\hat{\psi}(p_T) - \psi = \nu'_T \text{vec}(\hat{\beta}(p_T) - \beta(p_T)) + \xi' \text{vec}(\hat{\Sigma} - \Sigma) + o(T^{-1/2}) + O_p(p_T/T).$$

The above remainder terms are both  $o_p((T - p_T)^{-1/2})$  by **Assumption B.3**. □

### B.9.15 Proof of Proposition B.4

The proposition follows immediately from Lemmas B.8 and B.11 if we can show that  $\|\nu_T\|^2$  is bounded asymptotically. Let  $g_{j,i}(\cdot, \cdot, \cdot)$  denote the  $i$ -th element of  $g_j(\cdot, \cdot, \cdot)$ ,  $j = 1, 2$ ,  $i = 1, 2, \dots, n_W^2$ . Let  $M \equiv \sup_{\omega \in [0, 2\pi]} |h(\omega)| < \infty$ . Then

$$\begin{aligned} \|\nu_T\|^2 &= \sum_{i=1}^{n_W^2} \sum_{\ell=1}^{p_T} \left( \int_0^{2\pi} h(\omega) \{g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \cos(\omega\ell) \right. \\ &\quad \left. + g_{2,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \sin(\omega\ell)\} d\omega \right)^2 \\ &\leq 2M^2 \sum_{i=1}^{n_W^2} \sum_{\ell=1}^{p_T} \left\{ \left( \int_0^{2\pi} g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \cos(\omega\ell) d\omega \right)^2 \right. \\ &\quad \left. + \left( \int_0^{2\pi} g_{2,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \sin(\omega\ell) d\omega \right)^2 \right\}. \end{aligned}$$

The sum

$$\sum_{\ell=1}^{p_T} \left( \frac{1}{2\pi} \int_0^{2\pi} g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \cos(\omega\ell) d\omega \right)^2 \quad (\text{B.26})$$

equals the  $L_2(0, 2\pi)$  norm of the projection of the function  $\omega \mapsto g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)$  onto the space of orthonormal functions  $\{\omega \mapsto \cos(\omega\ell)\}_{1 \leq \ell \leq p_T}$ . Bessel's inequality therefore states that (B.26) is bounded above by the squared  $L_2(0, 2\pi)$  norm of the function  $\omega \mapsto g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)$ . We can similarly bound the expression (B.26) with  $g_{2,i}(\cdot, \cdot, \cdot)$  in place of  $g_{1,i}(\cdot, \cdot, \cdot)$  and with  $\sin(\omega\ell)$  in place of  $\cos(\omega\ell)$ . Hence,

$$\|\nu_T\|^2 \leq 8\pi^2 M^2 \sum_{i=1}^{n_W^2} \left( \|g_{1,i}(A_{\cos}(\cdot), A_{\sin}(\cdot), \Sigma)\|_{L_2(0, 2\pi)}^2 + \|g_{2,i}(A_{\cos}(\cdot), A_{\sin}(\cdot), \Sigma)\|_{L_2(0, 2\pi)}^2 \right),$$

using obvious notation for the  $L_2$  norms. These norms are finite by Assumption B.2.  $\square$

### B.9.16 Proof of Proposition B.5

We start by showing that  $\|\hat{\nu}_T - \nu_T\| = o_p(1)$  and  $\|\hat{\xi}(p_T) - \xi\| = o_p(1)$ . By Lemma B.10, and the twice continuous differentiability assumed in Assumption B.2, there exists a constant  $C < \infty$  such that, with probability approaching one,

$$\|g_j(\hat{A}_{\cos}(\omega; p_T), \hat{A}_{\sin}(\omega; p_T), \hat{\Sigma}) - g_j(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)\|$$

$$\leq C \left( \|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)\| + \|\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega)\| + \|\hat{\Sigma}(p_T) - \Sigma\| \right)$$

for  $j = 1, 2, 3$ . By [Lemma B.4](#) and the i.i.d. central limit theorem, we have  $\|\hat{\Sigma}(p_T) - \Sigma\| = O_p(T^{-1/2})$ . Using additionally [Lemma B.9](#), we then have, for example, that

$$\begin{aligned} & \sum_{\ell=1}^{p_T} \left\| \int_0^{2\pi} h(\omega) \left[ g_1(\hat{A}_{\cos}(\omega), \hat{A}_{\sin}(\omega), \hat{\Sigma}) - g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma) \right] \cos(\omega\ell) d\omega \right\| \\ & \leq \tilde{C} p_T \sup_{\omega \in [0, 2\pi]} \left( \|\hat{A}_{\cos}(\omega; p_T) - A_{\cos}(\omega)\| + \|\hat{A}_{\sin}(\omega; p_T) - A_{\sin}(\omega)\| + \|\hat{\Sigma} - \Sigma(p_T)\| \right) \\ & = O_p((p_T^3/T)^{1/2}) \\ & = o_p(1), \end{aligned}$$

where  $\tilde{C}$  is some constant. This type of calculation implies  $\|\hat{\nu}_T - \nu_T\| = o_p(1)$  and  $\|\hat{\xi}(p_T) - \xi\| = o_p(1)$ .

We now deal with the consistency of the two terms in  $\hat{\sigma}_\psi^2(p_T)$  one at a time. First, decompose

$$\begin{aligned} \hat{\nu}'_T (\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T)) \hat{\nu}_T &= \nu'_T \left( (\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T)) - (\Gamma(p_T)^{-1} \otimes \Sigma) \right) \nu_T \\ &\quad + (\hat{\nu}_T - \nu_T)' (\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T)) (\hat{\nu}_T - \nu_T) \\ &\quad + 2(\hat{\nu}_T - \nu_T)' (\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T)) \nu_T \\ &\equiv R_{1,T} + R_{2,T} + 2R_{3,T}. \end{aligned}$$

Using [Lemma B.2](#), we find

$$\begin{aligned} |R_{1,T}| &\leq \|\nu_T\|^2 \left\| (\hat{\Gamma}(p_T)^{-1} \otimes \hat{\Sigma}(p_T)) - (\Gamma(p_T)^{-1} \otimes \Sigma) \right\|_1 \\ &\leq M \left( \|\hat{\Gamma}(p_T)^{-1} - \Gamma(p_T)^{-1}\|_1 \|\hat{\Sigma}(p_T)\|_1 + \|\Gamma(p_T)^{-1}\|_1 \|\hat{\Sigma}(p_T) - \Sigma\|_1 \right) \\ &\leq M \left( \|\hat{\Gamma}(p_T) - \Gamma(p_T)\| \|\Gamma(p_T)^{-1}\|_1 \|\hat{\Gamma}(p_T)^{-1}\|_1 \|\hat{\Sigma}(p_T)\| + \|\Gamma(p_T)^{-1}\|_1 \|\hat{\Sigma}(p_T) - \Sigma\| \right) \\ &= o_p(1). \end{aligned}$$

Similar calculations, along with the fact  $\|\hat{\nu}_T - \nu_T\| = o_p(1)$ , can be used to show that  $R_{2,T} = o_p(1)$  and  $R_{3,T} = o_p(1)$ .

Second, define  $\Xi \equiv \text{Var}(e_t \otimes e_t)$  and decompose

$$\hat{\xi}(p_T)' \hat{\Xi}(p_T) \hat{\xi}(p_T) - \xi' \Xi \xi = \xi' (\hat{\Xi}(p_T) - \Xi) \xi$$

$$\begin{aligned}
& + (\hat{\xi}(p_T) - \xi)' \hat{\Xi}(p_T) (\hat{\xi}(p_T) - \xi) \\
& + 2(\hat{\xi}(p_T) - \xi)' \hat{\Xi}(p_T) \xi \\
& \equiv \tilde{R}_{1,T} + \tilde{R}_{2,T} + 2\tilde{R}_{3,T}.
\end{aligned}$$

Since  $\|\hat{\xi}(p_T) - \xi\| = o_p(1)$ , the statement of the proposition follows if we can show  $\|\hat{\Xi}(p_T) - \Xi\| = o_p(1)$ . Define  $\chi_t \equiv \text{vec}(e_t e_t' - \Sigma)$ , and note that  $(T - p_T)^{-1} \sum_{t=p_T+1}^T \chi_t \chi_t' \xrightarrow{P} \Xi$  by the usual law of large numbers for i.i.d. variables. Because

$$\begin{aligned}
\|\hat{\Xi}(p_T) - \Xi\| & \leq \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t \hat{\chi}_t' - \chi_t \chi_t'\| \\
& \leq \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\|^2 + \frac{2}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\| \|\chi_t\| \\
& \leq \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\|^2 + 2 \left( \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\|^2 \right. \\
& \quad \left. \times \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\chi_t\|^2 \right)^{1/2}
\end{aligned}$$

by Cauchy-Schwarz, we just need to show that

$$(T - p_T)^{-1} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\|^2 = o_p(1).$$

Since

$$\begin{aligned}
\|\hat{\chi}_t - \chi_t\| & = \|\hat{e}_t(p_T) \hat{e}_t(p_T)' - e_t e_t'\| \\
& \leq \|\hat{e}_t(p_T) - e_t\|^2 + 2\|\hat{e}_t(p_T) - e_t\| \|e_t\|,
\end{aligned}$$

we have

$$\begin{aligned}
\frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{\chi}_t - \chi_t\|^2 & \leq \frac{2}{T - p_T} \sum_{t=p_T+1}^T \|\hat{e}_t - e_t\|^4 \\
& + 4 \left( \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|\hat{e}_t - e_t\|^4 \frac{1}{T - p_T} \sum_{t=p_T+1}^T \|e_t\|^4 \right)^{1/2}.
\end{aligned}$$

The i.i.d. law of large numbers gives  $(T - p_T)^{-1} \sum_{t=p_T+1}^T \|e_t\|^4 = O_p(1)$ . To complete the

proof, we bound

$$\begin{aligned} \frac{1}{T-p_T} \sum_{t=p_T+1}^T \|\hat{e}_t - e_t\|^4 &\leq \frac{8}{T-p_T} \sum_{t=p_T+1}^T \|\hat{e}_t - e_t(p_T)\|^4 + \frac{8}{T-p_T} \sum_{t=p_T+1}^T \|e_t - e_t(p_T)\|^4 \\ &\equiv 8(\hat{R}_{1,T} + \hat{R}_{2,T}) \end{aligned}$$

and show that the two terms on the right-hand side tend to zero, using similar arguments as in the proof of [Lemma B.4](#). First,

$$\hat{R}_{1,T} \leq \|\hat{\beta}(p_T) - \beta(p_T)\|^4 \frac{1}{T-p_T} \sum_{t=p_T+1}^T \|X_t(p_T)\|^4 = O_p((p_T/T)^2) O_p(p_T^2) = o_p(1),$$

since

$$E\|X_t(p_T)\|^4 = E(\sum_{\ell=1}^{p_T} \|W_{t-\ell}\|^2)^2 = \sum_{\ell=1}^{p_T} \sum_{m=1}^{p_T} E(\|W_{t-\ell}\|^2 \|W_{t-m}\|^2) = O(p_T^2).$$

Second,

$$\begin{aligned} E(\hat{R}_{2,T}) &= E\|e_t - e_t(p_T)\|^4 \\ &\leq E\left(\sum_{\ell=p_T+1}^{\infty} \|A_{\ell}\| \|W_{t-\ell}\|\right)^4 \\ &= \sum_{\ell_1, \ell_2, \ell_3, \ell_4=p_T+1}^{\infty} \|A_{\ell_1}\| \|A_{\ell_2}\| \|A_{\ell_3}\| \|A_{\ell_4}\| E(\|W_{t-\ell_1}\| \|W_{t-\ell_2}\| \|W_{t-\ell_3}\| \|W_{t-\ell_4}\|) \\ &\leq \text{constant} \times \left(\sum_{\ell=p_T+1}^{\infty} \|A_{\ell}\|\right)^4 \\ &= o(1). \quad \square \end{aligned}$$

## References

- Berk, K. (1974). Consistent Autoregressive Spectral Estimates. *Annals of Statistics*, 2(3), 489–502.
- Bhatia, R. (1997). *Matrix Analysis*. Graduate Texts in Mathematics. Springer.
- Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods* (2nd ed.). Springer Series in Statistics. Springer.
- Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford University Press.
- Del Negro, M., Giannoni, M. P., & Patterson, C. (2012). The Forward Guidance Puzzle. Federal Reserve Bank of New York Staff Report No. 574.
- Forni, M., Gambetti, L., & Sala, L. (2019). Structural VARs and noninvertible macroeconomic models. *Journal of Applied Econometrics*, 34(2), 221–246.
- Gertler, M. & Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1), 44–76.
- Gonçalves, S. & Kilian, L. (2007). Asymptotic and bootstrap inference for  $AR(\infty)$  processes with conditional heteroskedasticity. *Econometric Reviews*, 26(6), 609–641.
- Hannan, E. (1970). *Multiple Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons.
- Känzig, D. R. (2021). The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements. *American Economic Review*, 111(4), 1092–1125.
- Kreiss, J.-P., Paparoditis, E., & Politis, D. N. (2011). On the Range of Validity of the Autoregressive Sieve Bootstrap. *Annals of Statistics*, 39(4), 2103–2130.
- Leeper, E. M., Walker, T. B., & Yang, S.-C. S. (2013). Fiscal Foresight and Information Flows. *Econometrica*, 81(3), 1115–1145.
- Lewis, R. & Reinsel, G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis*, 16(3), 393–411.

- Lippi, M. & Reichlin, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *Journal of Econometrics*, *63*(1), 307–325.
- Mertens, K. & Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, *103*(4), 1212–1247.
- Meyer, M. & Kreiss, J.-P. (2015). On the Vector Autoregressive Sieve Bootstrap. *Journal of Time Series Analysis*, *36*(3), 377–397.
- Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2 chapter 2, (pp. 71–162). Elsevier.
- Saikkonen, P. & Lutkepohl, H. (2000). Asymptotic Inference on Nonlinear Functions of the Coefficients of Infinite Order Cointegrated VAR Processes. In W. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim, & A. Würtz (Eds.), *Nonlinear Econometric Modeling in Time Series Analysis* (pp. 165–201). Cambridge University Press.
- Sims, C. A. & Zha, T. (2006). Does Monetary Policy Generate Recessions? *Macroeconomic Dynamics*, *10*(02), 231–272.
- Smets, F. & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, *97*(3), 586–606.
- Stock, J. H. (2008). What’s New in Econometrics: Time Series, Lecture 7. Lecture slides, NBER Summer Institute.
- Stock, J. H. & Watson, M. W. (2012). Disentangling the Channels of the 2007–09 Recession. *Brookings Papers on Economic Activity*, *2012*(1), 81–135.
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419.
- Wolf, C. K. (2020). SVAR (Mis)identification and the Real Effects of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, *12*(4), 1–32.