Online Appendix for: Instrumental Variable Identification of Dynamic Variance Decompositions

Mikkel Plagborg-Møller   Christian K. Wolf

June 9, 2019

This online appendix contains supplemental material for the article "Instrumental Variable Identification of Dynamic Variance Decompositions". We provide (i) bounds on the FVD, (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 analytical example to illustrate the degree of invertibility concept, (v) supplementary results on the structural macro model and our empirical application, (vi) details on the partial identification robust confidence interval for a parameter, (vii) asymptotic theory on the nonparametric validity of our sieve VAR inference strategy, and (viii) simulation evidence on the finite-sample performance of the inference method. 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.1 Identification of forecast variance decomposition

We here state and discuss a proposition characterizing the sharp identified set for the FVD. See also the discussion in Section 3.

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*

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

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.1) is monotonically decreasing in α, 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 expression for the lower bound (B.1) has a simple interpretation. Even if α is known, the denominator $\operatorname{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t})$ 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 α, 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 α is not known, the smallest possible value of the lower bound (B.1) is attained at the largest possible value of α, namely α_{UB}, for which $\varepsilon_{1,t}$ contributes the least to forecasts of y_t.

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 (15). The multiple-IV model in Assumptions 1 and 2 implies the following cross-spectrum between y_t and $\tilde{z}_t$:

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

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 ζ_ℓ, $\ell \geq 0$, and an n_z-dimensional constant real vector η of unit length such that*

$$s_{y\tilde{z}}(\omega) = \zeta(e^{-i\omega})\eta', \quad \omega \in [0, 2\pi], \tag{B.3}$$

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 Σ_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.2) that λ is point-identified and equal to the η vector in equation (B.3). This is because any rank-1 factorization of a matrix is identified up to sign and scale, and we have normalized η to have length 1. Let Ξ 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 Ξ), 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 α 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 α. The first element of $Q\tilde{z}_t$ equals $\breve{z}_t$ as defined in equation (16) in the main text.[B.1]

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

[B.1]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 α^2 yields when applied to $\breve{z}_t$.

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_{ε_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_{ε_x} of the n_ε shocks. Denote this sub-vector of shocks by $\varepsilon_{x,t}$. For now, n_{ε_x} need not be known to the econometrician. We define the extended LP-IV model as

$$y_t = \Theta(L)\varepsilon_t, \quad \Theta(L) \equiv \sum_{\ell=0}^{\infty} \Theta_\ell L^\ell, \tag{B.4}$$

$$z_t = \sum_{\ell=1}^{\infty} (\Psi_\ell z_{t-\ell} + \Lambda_\ell y_{t-\ell}) + \underbrace{\Gamma\varepsilon_{x,t} + \Sigma_v^{1/2} v_t}_{\tilde{z}_t}, \tag{B.5}$$

where Γ 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, $\Gamma\varepsilon_{x,t}$:

$$\begin{aligned} FVR_{i,\ell} &\equiv 1 - \frac{\mathrm{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t}, \{\Gamma\varepsilon_{x,\tau}\}_{t<\tau<\infty})}{\mathrm{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})} \\ &= \frac{\sum_{m=0}^{\ell-1} \mathrm{Cov}(y_{it}, \tilde{z}_{t-m})(\Gamma\Gamma')^{-1}\,\mathrm{Cov}(y_{it}, \tilde{z}_{t-m})'}{\mathrm{Var}(y_{i,t+\ell} \mid \{y_\tau\}_{-\infty<\tau\leq t})}. \end{aligned} \tag{B.6}$$

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 Section B.2, the testable restriction of the model (B.4)–(B.5) is that the joint spectrum of y_t and $\tilde{z}_t$ has a rank-n_{ε_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 $FVR_{i,\ell}$.[B.2] In particular, we may assume that Γ has full row rank, which we do from now on, thus justifying the second equality in (B.6).

[B.2] The argument is very similar to the one-shock case in Section B.2 and is available upon request.

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

$$\Sigma_{\tilde{z}} - \Gamma\Gamma' \geq 0, \tag{B.7}$$

and

$$\Gamma\Gamma' - 2\pi s_{\tilde{z}^\dagger}(\omega) \geq 0, \quad \forall\, \omega \in [0, \pi], \tag{B.8}$$

where $s_{\tilde{z}^\dagger}(\omega) = s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega)$ 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.6) over the space of $n_z \times n_z$ symmetric positive definite matrices $\Gamma\Gamma'$ subject to constraints (B.7)–(B.8).

Lower bound on FVR. We now establish a sharp lower bound on the numerator in the definition (B.6) 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.7). 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.[B.3]

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.6) of

[B.3]Note that $\Gamma\Gamma' = \Sigma_{\tilde{z}} = 2\pi s_{\tilde{z}}(\omega)$ satisfies constraint (B.8) by the Schur complement formula and the positive semidefiniteness of the spectrum of $(y_t', \tilde{z}_t')'$.

the FVR is given by the value of the program

$$\max_{X \in \mathcal{S}_{n_z}} \operatorname{tr}(XC) + \operatorname{tr}(AC) \tag{B.9}$$
$$X \leq B(\omega), \quad \omega \in [0, \pi].$$

Here X is a stand-in for $(\Gamma\Gamma')^{-1} - \Sigma_{\tilde{z}}^{-1}$, $C \equiv \sum_{m=0}^{\ell-1} \operatorname{Cov}(y_{it}, \tilde{z}_{t-m})' \operatorname{Cov}(y_{it}, \tilde{z}_{t-m})$, $A \equiv \Sigma_{\tilde{z}}^{-1}$, 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

$$X \leq \tilde{B}_m, \quad m \in \{1, 2, \dots, N\}, \tag{B.10}$$

where $\tilde{B}_m \equiv \frac{N}{\pi} \times \int_{(m-1)\frac{\pi}{N}}^{m\frac{\pi}{N}} B(\omega)\, d\omega$. As $N \to \infty$, 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.6). Efficient numerical algorithms to compute the solution to semidefinite programs of the form (B.9)–(B.10) are available in Matlab and other environments.[B.4]

Alternatively, we can derive non-sharp upper bounds on the FVR numerator (B.6) in closed form. For example, one conservative upper bound is obtained by maximizing $\operatorname{tr}(XC) + \operatorname{tr}(AC)$ subject to $X + \tilde{\Sigma}_{\tilde{z}}^{-1} \leq \left(\int_{-\pi}^{\pi} s_{\tilde{z}^\dagger}(\omega)\, d\omega\right)^{-1} = \operatorname{Var}(\tilde{z}_t^\dagger)^{-1}$. This yields the upper bound

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

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

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

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

[B.4] 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 $\left(\begin{smallmatrix} A & B' \\ B & A \end{smallmatrix}\right)$ is positive semi-definite.

section reduce to the FVR bound expressions derived in Section 3.2.

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 Γ 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 $\tilde{z}_t^\dagger \equiv E(\tilde{z}_t \mid \{y_\tau\}_{-\infty<\tau<\infty}) = \Gamma\varepsilon_{x,t}$, so the historical decomposition of y_t with respect to $\varepsilon_{x,t}$ is point-identified as $E(y_t \mid \{\varepsilon_{x,t}\}_{-\infty<\tau\leq t}) = E(y_t \mid \{\tilde{z}_\tau^\dagger\}_{-\infty<\tau\leq t})$.

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 LP-IV model in Assumptions 1 to 3. Our analysis builds on results by Lippi & Reichlin (1994), Forni et al. (2018), and Wolf (2018), 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(∞) model with forecast errors $u_t \equiv y_t - E(y_t \mid \{y_\tau\}_{-\infty<\tau<t})$. 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 Θ_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 LP-IV model in Assumptions 1 to 3. The shock that is (mis)identified by SVAR-IV is given by*

$$\tilde{\varepsilon}_{1,t} \equiv \gamma' u_t = \sum_{j=1}^{n_\varepsilon}\sum_{\ell=0}^{\infty} a_{j,\ell}\varepsilon_{j,t-\ell}, \tag{B.11}$$

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[B.5]

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

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

[B.5] In any SVAR(∞) 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). 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) 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 γ of the forecast residuals u_t under noninvertibility, since it can be verified that $\gamma' u_t \propto E(\varepsilon_{1,t} \mid u_t)$ regardless of invertibility.[B.6] 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. (2018), and Wolf (2018). An explicit illustration of SVAR-IV mis-identification is provided in Section B.6.

[B.6] In particular, no other linear combination γ can yield a representation (B.11) 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, *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 Analytical illustration of the degree of invertibility

Here we use an analytical example to illustrate the degree of invertibility concept defined in Section 2. Consider the univariate MA(1) model

$$y_t = \varepsilon_{1,t} + \Theta_1 \varepsilon_{1,t-1}, \quad \varepsilon_{1,t} \overset{i.i.d.}{\sim} N(0,1),$$

where $\Theta_1 \in \mathbb{R}$. The shock $\varepsilon_{1,t}$ is invertible (in the sense $E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t}) = \varepsilon_{1,t}$) if and only if $|\Theta_1| \leq 1$ (Brockwell & Davis, 1991, Remark 6, p. 88). If instead $|\Theta_1| > 1$ (a prototypical "news shock" process), standard calculations give

$$u_t \equiv y_t - E(y_t \mid \{y_\tau\}_{-\infty<\tau<t}) = \varepsilon_{1,t} + (1-\Theta_1^2)\sum_{\ell=1}^{\infty}\left(-\frac{1}{\Theta_1}\right)^{\ell} \varepsilon_{1,t-\ell},$$

so that $\text{Var}(u_t) = \Theta_1^2$. Conversely,

$$\varepsilon_{1,t} = -\frac{1}{\Theta_1^2}u_t - \left(1 - \frac{1}{\Theta_1^2}\right)\sum_{\ell=1}^{\infty}\left(-\frac{1}{\Theta_1}\right)^{\ell} u_{t+\ell}.$$

Hence, in the noninvertible case $|\Theta_1| > 1$, the degree of invertibility up to time $t+\ell$ equals

$$\begin{aligned} R_\ell^2 &= \text{Var}(E(\varepsilon_{1,t} \mid \{y_\tau\}_{-\infty<\tau\leq t+\ell})) \\ &= \text{Var}(E(\varepsilon_{1,t} \mid \{u_\tau\}_{-\infty<\tau\leq t+\ell})) \\ &= \text{Var}(E(\varepsilon_{1,t} \mid \{u_\tau\}_{t\leq\tau\leq t+\ell})) \\ &= 1 - (\Theta_1^2 - 1)\Theta_1^{-2(\ell+1)}. \end{aligned}$$

This expression is of course increasing in ℓ. As $\ell \to \infty$, the expression converges to the degree of recoverability $R_\infty^2 = 1$ (note that $n_\varepsilon = n_y$ implies $R_\infty^2 = 1$). The speed of convergence is governed by $|\Theta_1|$. The degree of invertibility equals $R_0^2 = \min\{\Theta_1^{-2}, 1\}$. The continuity of R_0^2 in the impulse response parameter Θ_1 reflects the fact that, for small $\epsilon > 0$, a noninvertible process with $\Theta_1 = 1 + \epsilon$ has similar properties to an invertible process with $\Theta_1 = 1 - \epsilon$.

Structural illustration: Degree of invertibility/recoverability

	Monetary shock		Forw. guid. shock		Technology shock	
Macro observables	R_0^2	R_∞^2	R_0^2	R_∞^2	R_0^2	R_∞^2
Baseline	0.8705	0.8767	0.0792	0.8813	0.2007	0.2209
+ investm. + consum.	0.9419	0.9511	0.0997	0.9497	0.2151	0.2421
+ hours	0.9277	0.9290	0.0800	0.9336	0.9835	0.9849
All observables	1	1	0.1110	1	1	1

Table B.1: Degree of invertibility R_0^2 and degree of recoverability R_∞^2 in Smets-Wouters model in Section 4, 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).

B.6 Supplementary results for the structural macro model

We now show how our illustration using the Smets & Wouters (2007) model changes when we expand the list of observables used by the econometrician. We stress, however, that quantitative realism is not the primary goal of the exercise in Section 4; we merely wish to illustrate how the tightness of the identification bounds depends on the underlying model. Additionally, we use the model to illustrate the potential bias of SVAR-IV methods.

Table B.1 lists the R_0^2 and R_∞^2 measures for each of our three shocks of interest, across different sets of observable macro aggregates. Since the Smets-Wouters model features seven structural shocks, a judicious choice of seven observed macro aggregates would be guaranteed to yield recoverability of all shocks. In the baseline variant of the model – i.e., with a conventional monetary policy shock, not a forward guidance shock – the seven macro aggregates explicitly considered in the estimation of Smets & Wouters (2007) are in fact sufficient to yield invertibility, not just recoverability. With the forward guidance shock, in contrast, the SVAR bias remains significant even for this expanded set of observables, with an R_0^2 of 0.111, meaning that SVAR-based estimates of the one-period-ahead forecast variance ratio are biased upward by a factor of around 9. Figure B.1 shows R_ℓ^2 for the forward guidance shock as a function of ℓ, either with our baseline set of observables or with the full set of seven macro aggregates. In both cases, the R_ℓ^2 spikes at horizon 2, consistent with the intu-

Forward guidance shock: Degree of invertibility at time $t+\ell$

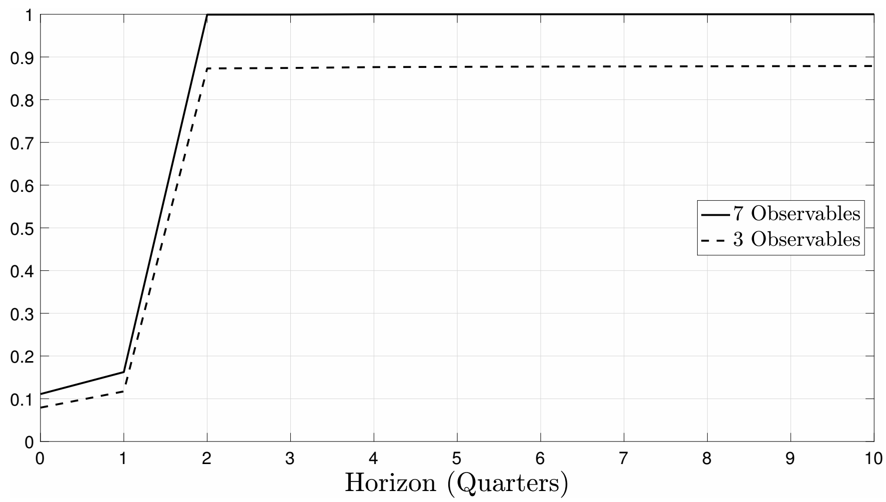

Figure B.1: Population R^2_ℓ for the forward guidance shock, with three observables (output, inflation, interest rate) and seven observables (the full set in Smets & Wouters, 2007).

ition given in Section 4, and in fact already at that point reaches 1 in the large model (so for recoverability the far future is not needed).

For the productivity shock, our conclusions hinge on whether some notion of aggregate labor is included in the set of observables. In the posterior mode parameterization of the Smets-Wouters model, productivity shocks are an important driver of fluctuations in aggregate hours, even at high frequencies. Thus, as soon as aggregate hours is included among the set of observables, the spectral density $s_{\varepsilon_1^\dagger}(\omega)$ of the best two-sided linear predictor is close to $s_{\varepsilon_1}(\omega) = \frac{1}{2\pi}$ everywhere. Inclusion of any other conventional observables – and in particular of investment and consumption – in contrast leaves the conclusions of Section 4 almost unaffected.

Finally, we illustrate the potential bias if SVAR-IV methods are used to identify forecast variance ratios and decompositions through a valid external instrument, as discussed in Section B.4. For particularly stark results, we consider the case of a forward guidance shock. Figure B.2 shows that the contributions of this shock to macro fluctuations are dramatically overstated in the baseline specification with three macro observables. On impact, the ratio of (mis)identified to true FVR is $1/0.0792 \approx 13$. The dynamics, while not obeying a similarly simple relation, remain substantially distorted, as the SVAR does not capture the hump shapes in the contributions of the shock to the responses of the interest rate and output; moreover, the long-run FVRs of the interest rate and inflation are highly upward biased.

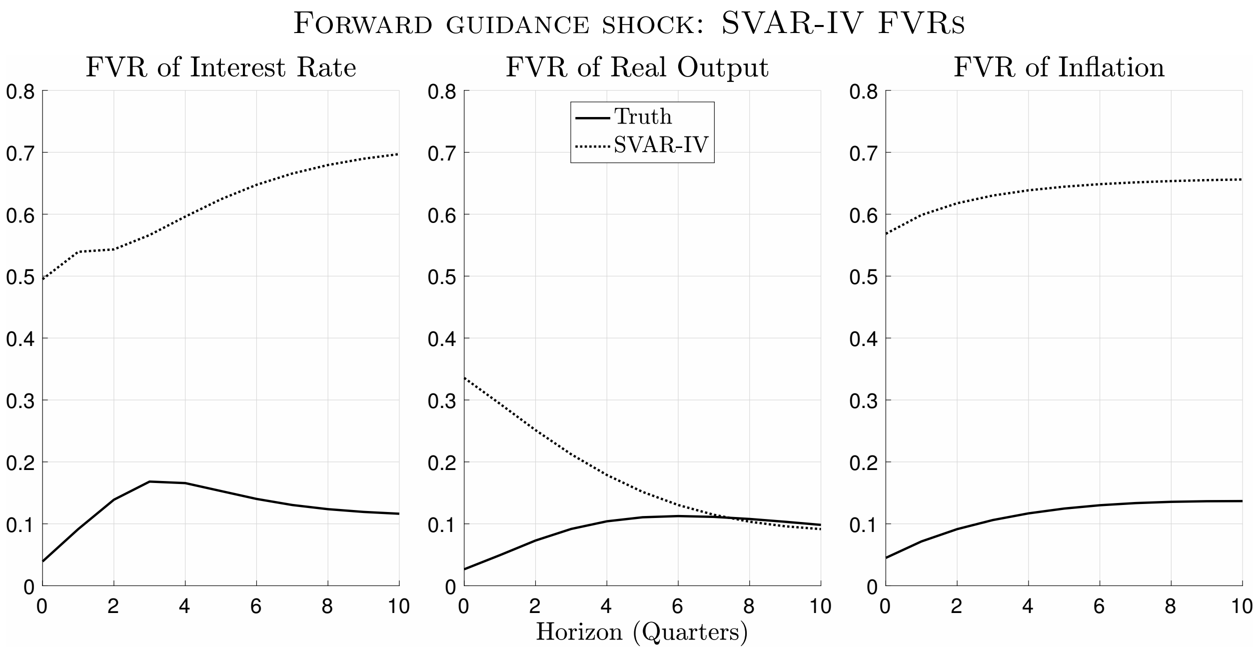

Figure B.2: FVRs for a forward guidance shock in the Smets-Wouters model, true values and SVAR-IV-implied values (population limit). Baseline set of three observables.

Empirical application: Granger causality tests

Equation	χ^2-stat.	d.f.	p-value
FFR	21.52	6	0.001
IP growth	13.07	6	0.042
CPI growth	4.65	6	0.590
EBP	18.31	6	0.006
All	58.94	24	0.000

Table B.2: Tests of Granger non-causality of the external IV. First four rows show tests for each y_t equation separately. Last row shows joint test.

B.7 Supplementary empirical results

Here we present supplementary results for the empirical application in Section 6.

Table B.2 shows standard Granger causality tests of the invertibility assumption, conditional on the reduced-form VAR. Granger non-causality of the external IV is rejected at the 5% level in each y_t equation except for inflation. We can also reject the joint hypothesis of Granger non-causality in all equations at the 0.1% level.

Figure B.3 displays forecast variance decompositions with respect to the monetary shock, as implied by an SVAR-IV model estimated on the same data as our LP-IV analysis in Section 6. The SVAR-IV confidence intervals are much narrower than the LP-IV ones in Figure 5, since the SVAR-IV model imposes the point-identifying invertibility assumption.[B.7] We emphasize, however, that this assumption is rejected in the data, and without the use of the tools provided in this paper, it would not be clear how distorted SVAR inference could be. Our SVAR specification follows Gertler & Karadi (2015) closely, except that we use only the 1990–2012 sample and choose $p = 6$ lags for consistency with the reduced-form VAR in the LP-IV analysis.[B.8] We use 10,000 homoskedastic recursive residual bootstrap draws. See Caldara & Herbst (2019) for a thorough Bayesian analysis of SVAR-IV-implied FVDs in a closely related empirical specification.

[B.7] The upward bias emphasized in Section B.4 is present at the one-step forecast horizon, and very visible for the federal funds rate. For the other variables, impact FVRs are close to 0, so the bias is harder to see.

[B.8] Hence, the only difference from the spectrum estimated in Section 6 is that here we impose that the IV does not Granger-cause the other variables.

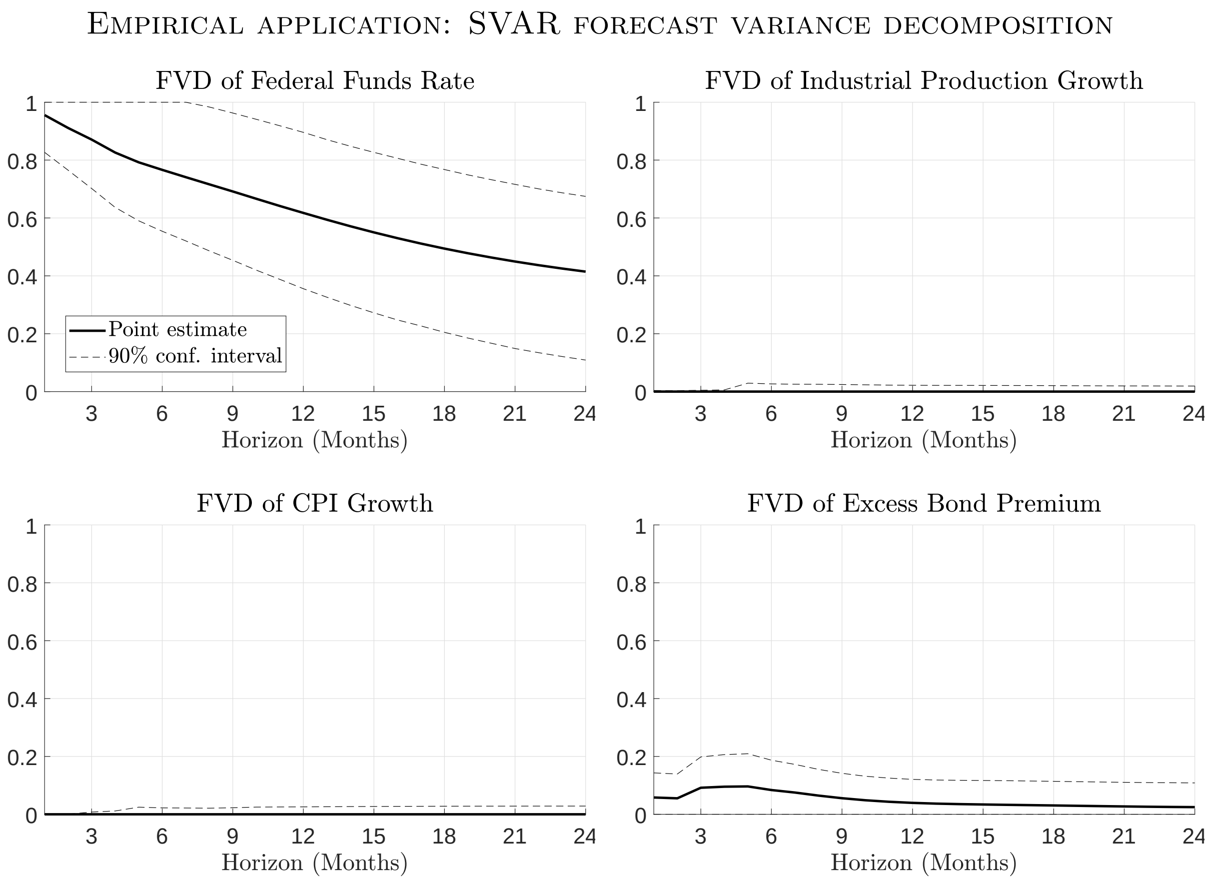

Figure B.3: Point estimates and 90% confidence intervals for forecast variance decompositions implied by an SVAR-IV model, across different variables and forecast horizons. For visual clarity, we force bias-corrected estimates/bounds to lie in $[0, 1]$.

B.8 Confidence interval for parameter

Here we describe how to compute a confidence interval for a parameter of interest, as mentioned in Section 5. As in Stoye (2009, p. 1305), define the two scalars $\underline{\hat{c}}, \hat{\overline{c}}$ as the minimizers of the objective function

$$\underline{\hat{\sigma}} \times \underline{\hat{c}} + \hat{\overline{\sigma}} \times \hat{\overline{c}},$$

subject to the two constraints

$$\Pr\left(-\underline{\hat{c}} \leq U_1,\ \hat{\rho} U_1 \leq \hat{\overline{c}} + \frac{\hat{\Delta}}{\hat{\overline{\sigma}}} + \sqrt{1-\hat{\rho}^2} \times U_2\right) \geq 1-\beta,$$
$$\Pr\left(-\underline{\hat{c}} - \frac{\hat{\Delta}}{\underline{\hat{\sigma}}} - \sqrt{1-\hat{\rho}^2} \times U_2 \leq \hat{\rho} U_1,\ U_1 \leq \hat{\overline{c}}\right) \geq 1-\beta.$$

Here the probabilities are taken solely over the distribution of $(U_1, U_2)'$, which is bivariate standard normal. The above minimization problem is easy to solve numerically, cf. Stoye (2009, Appendix B). Given these definitions, the interval

$$\left[\underline{h}(\hat{\vartheta}) - \underline{\hat{c}} \times \underline{\hat{\sigma}}\,,\ \overline{h}(\hat{\vartheta}) + \hat{\overline{c}} \times \hat{\overline{\sigma}}\right]$$

is a (pointwise) asymptotically valid level-$(1-\beta)$ confidence interval for the true parameter. This result follows from the delta method and the results in Stoye (2009), who builds on Imbens & Manski (2004).[B.9]

[B.9] We do not require that the VAR parameters are *uniformly* asymptotically normal, since we only develop pointwise valid confidence intervals, as discussed in Section 5.

B.9 Nonparametric sieve VAR inference

In this appendix section we show that the VAR(p) delta method inference strategy proposed in the main paper is valid 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(∞) 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 LP-IV 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 LP-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 LP-IV model, see Gonçalves & Kilian (2007), Meyer & Kreiss (2015), and references therein. We also remark that this section is only concerned with asymptotics that are pointwise in the true DGP. Issues related to uniform inference are outside the scope of this paper.

B.9.1 Assumptions, parameters of interest, and estimator

We first define the general class of parameters of interest for empirical LP-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 LP-IV purposes.

We assume that the data are generated by a reduced-form VAR(∞) 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 is 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 $\|B\| \equiv (\mathrm{tr}(B'B))^{1/2}$ denote the Frobenius norm.

Assumption B.1. *The process* $\{W_t\}$ *is generated by the mean-zero stationary VAR(*∞*) 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* ℓ*. 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* $E(e_t) = 0_{n_W \times 1}$*,* $\Sigma \equiv \mathrm{Var}(e_t)$ *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.[B.10] Meyer & Kreiss (2015) discuss the generality of assuming a reduced-form VAR(∞) with i.i.d. disturbances, see also Kreiss et al. (2011) for more details in the univariate case. Assuming the LP-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 LP-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 $h\colon [0, 2\pi] \to \mathbb{R}^K$ and $g\colon \mathcal{A}_\delta \times \mathbb{S}_{n_W} \to \mathbb{R}^K$, 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

[B.10] 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 LP-IV analysis.[B.11] For example, the class contains (i) elements Σ_{ij} of Σ, (ii) the degree of recoverability $R^2_\infty = \int_0^{2\pi} s_{y\tilde{z}}(\omega)^* s_y(\omega)^{-1} s_{y\tilde{z}}(\omega)\, d\omega$, 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 $\omega \in [0, 2\pi]$,

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

Other LP-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 $g_1(B_1, B_2, S) \equiv \frac{\partial g(B_1,B_2,S)}{\partial \operatorname{vec}(B_1)}$, $g_2(B_1, B_2, S) \equiv \frac{\partial g(B_1,B_2,S)}{\partial \operatorname{vec}(B_2)}$, and $g_3(B_1, B_2, S) \equiv \frac{\partial g(B_1,B_2,S)}{\partial \operatorname{vec}(S)}$. At the true VAR parameters $\{A_\ell\}$ and Σ, 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 LP-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 least-squares VAR estimator

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

[B.11] The only exception is the parameter $\sup_{\omega\in[0,2\pi]} s_{\tilde{z}^\dagger}(\omega)$, which is discussed in the main text.

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 ψ 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 $p_T^3/T \to 0$ and $T^{1/2} \sum_{\ell=p_T+1}^{\infty} \|A_\ell\| \to 0$ as $T \to \infty$.*

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.9.2 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 $\nu_T = (\nu_{1,T}', \ldots, \nu_{p_T,T}')' \in \mathbb{R}^{n_W^2 p_T}$, where

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

for $\ell = 1, 2, \ldots, p_T$. 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 ν_T and ξ 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 $T \to \infty$.

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* $\sigma_\psi^2 \equiv \lim_{T\to\infty} \nu_T'(\Gamma(p_T)^{-1} \otimes \Sigma)\nu_T + \xi' \operatorname{Var}(e_t \otimes e_t)\xi$ *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 σ_ψ^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 LP-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* $\hat{\chi}_t(p_T) \equiv \operatorname{vec}(\hat{e}_t(p_T)\hat{e}_t(p_T)' - \hat{\Sigma}(p_T))$ *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 VAR(p_T) process is valid asymptotically even if the true DGP is a VAR(∞). Hence, the partial identification robust confidence intervals proposed in Section 5 are valid under our regularity conditions. This conclusion is consistent with the finite-sample simulation evidence presented in Section B.10.

B.10 Simulation study of inference procedure

In this section we examine the finite-sample performance of our identification and inference procedures through simulations. We assume that the macro aggregates y_t follow a structural VARMA(p,1) model:

$$y_t = \sum_{\ell=1}^{p} \Xi_\ell y_{t-\ell} + \Theta_0 \varepsilon_t + \Theta_1 \varepsilon_{t-1}.$$

We adopt a variant of the DGP in Kilian & Kim (2011). We consider $n_y = 2$ macro variables, $p = 1$ autoregressive lag (with one exception discussed below), and set

$$\Xi_1 = \begin{pmatrix} \rho_y & 0 \\ 0.5 & 0.5 \end{pmatrix}.$$

For the MA part, we consider $n_\varepsilon = 2$ shocks and set

$$\Theta_0 = \text{chol}\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

"chol" denotes the lower triangular Cholesky decomposition. We set

$$\Theta_1 = \theta \times \Theta_0,$$

where θ is a scalar parameter that governs the degree of invertibility, with $\theta > 1$ implying non-invertibility. We then add an external instrument for the shock of interest $\varepsilon_{1,t}$, with

$$z_t = \rho_z z_{t-1} + \rho_{zy}(y_{1,t-1} + y_{2,t-1}) + \varepsilon_{1,t} + \sigma_v v_t.$$

Relative to the notation in the main text, we have normalized $\alpha = 1$. Finally, measurement error and structural shocks satisfy the same assumptions as in our main analysis:

$$(\varepsilon_t', v_t')' \overset{i.i.d.}{\sim} N(0, I_3).$$

We run Monte Carlo experiments for a range of parameterizations of the above DGP. We conduct $5,000$ Monte Carlo repetitions per DGP, and construct confidence intervals at the 90% level using $1,000$ bootstrap draws per simulation. Our objects of interest throughout are R_0^2 and the FVR for variable $y_{2,t}$ at horizons 1 and 4. We construct confidence intervals for parameter and identified set using our LP-IV procedure, with coverage rates referring to the true parameter and the entire population-implied identified set, respectively. For the FVR, we also construct confidence intervals using a SVAR-IV bootstrap procedure, still evaluating coverage relative to the true population FVR. As in the empirical application in Section 6, the reduced-form VAR lag length is selected using AIC, and we use Hall's asymmetric bootstrap confidence interval.

We consider DGPs that are deviations from a baseline parametrization. In our benchmark, we set $\rho_y = 0.5$, $\rho_z = \rho_{zy} = 0$, $\theta = 0$, $\sigma_\nu = 1$, and sample size $T = 250$. We then consider variations with more autoregressive persistence (either $\rho_y = 0.9$, or $\rho_z = 0.8$ and $\rho_{zy} = 0.3$), an invertible MA component ($\theta = 0.5$), a non-invertible MA component ($\theta = 2$), a weaker instrument ($\sigma_\nu = 2$), and different sample sizes ($T = 100$, $T = 500$). Finally, we allow for richer dynamics, with $p = 4$ and $\Xi_j = \frac{1}{j^2}\Xi_1$ for $j = 2, 3, 4$.

Table B.3 shows that the LP-IV confidence sets achieve coverage rates close to or exceeding the desired level of 90% throughout. Several features of the table are noteworthy. First, the LP-IV coverage rates are almost always closer to the nominal level than the SVAR-IV coverage. The LP-IV procedure controls coverage even in the non-invertible case ($\theta = 2$), whereas the SVAR-IV procedure does not. Second, coverage deteriorates somewhat with noisier instruments, as expected. Third, we face some well-known parameter-at-the-boundary issues. For most experiments, $R_0^2 = 1$. This explains the over-coverage of confidence intervals for the parameter and, less so, for the overall identified set. Similar problems would arise if the true FVR were close to 0. Fourth, for more persistent DGPs, the AIC tends to select an insufficient number of lags, resulting in moderate under-coverage, in particular for the FVRs at horizon 4. For example, in the experiment with $p = 4$ autoregressive lags, the AIC selects an average lag length of 2.2. In unreported simulations with fixed higher lag lengths, we find that coverage improves substantially.

Monte Carlo study: Coverage rates of confidence intervals

Experiment	R_0^2			$FVR_{1,2}$				$FVR_{4,2}$			
	Truth	Param.	IS	Truth	Param.	IS	SVAR	Truth	Param.	IS	SVAR
Baseline	1	0.993	0.931	0.640	0.907	0.897	0.869	0.819	0.897	0.873	0.841
$\rho_y = 0.9$	1	0.995	0.923	0.640	0.904	0.900	0.878	0.862	0.903	0.880	0.820
$\rho_z = 0.8,\ \rho_{zy} = 0.3$	1	0.994	0.928	0.640	0.906	0.899	0.877	0.819	0.903	0.884	0.843
$\theta = 0.5$	1	0.979	0.939	0.640	0.882	0.897	0.878	0.839	0.875	0.859	0.835
$\theta = 2$	0.25	0.880	0.886	0.160	0.859	0.856	0.143	0.806	0.890	0.860	0.674
$\sigma_v = 2$	1	0.976	0.939	0.640	0.880	0.885	0.821	0.819	0.862	0.854	0.772
$T = 100$	1	0.985	0.912	0.640	0.897	0.870	0.831	0.819	0.873	0.847	0.789
$T = 500$	1	0.994	0.934	0.640	0.900	0.900	0.883	0.819	0.892	0.889	0.868
$p = 4$	1	0.986	0.934	0.640	0.891	0.895	0.870	0.855	0.784	0.822	0.841

Table B.3: Coverage rates of 90% confidence intervals, constructed as in Section 5, using 1,000 bootstrap for each Monte Carlo experiment, and 5,000 Monte Carlo experiments per DGP. The DGPs (along rows) are described in the text. “Truth”: true parameter value. “Param.”: coverage rate of LP-IV confidence interval for parameter. “IS”: coverage rate of LP-IV confidence interval for identified set. “SVAR”: coverage rate of SVAR-IV confidence interval for parameter.

B.11 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 LP-IV identification analysis. Then we address the sieve VAR convergence results.

B.11.1 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 ν 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 ν 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 ν. □

B.11.2 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 ν_t with spectral density $s_\nu(\omega) = s_{\tilde{x}}(\omega) - s_x(\omega)$, $\omega \in [0, 2\pi]$, and such that the ν_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 ν_t processes implies that $x_{t+\ell}$ and $\nu_{t+\ell}$ are independent also conditional on $\{x_\tau, \nu_t\}_{-\infty<\tau\leq t}$. □

B.11.3 Proof of Proposition B.1

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

$$\begin{aligned}\mathrm{Var}(y_{i,t+\ell} \mid \{\varepsilon_\tau\}_{-\infty<\tau\leq t}) &= \sum_{m=0}^{\ell-1} \Theta_{i,1,m}^2 + \sum_{j=2}^{n_\varepsilon}\sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2 \\ &= \frac{1}{\alpha^2}\sum_{m=0}^{\ell-1} \mathrm{Cov}(y_{i,t}, \tilde{z}_{t-m})^2 + \sum_{j=2}^{n_\varepsilon}\sum_{m=0}^{\ell-1} \Theta_{i,j,m}^2. \end{aligned} \tag{B.12}$$

   Given α, the first term in (B.12) 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 ℓ, instead consider the new sequence $\{\breve{\Theta}_{\bullet,j,m}\}_{2\leq j\leq n_\varepsilon, 0\leq m<\infty}$, defined via

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

   Then the stochastic process induced by $\{\breve{\Theta}_{\bullet,j,m}\}_{2\leq j\leq n_\varepsilon, 0\leq m<\infty}$ 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.12) as large as possible. Given a known $\alpha \in (\alpha_{LB}, \alpha_{UB}]$, define

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

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

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

so the second term in (B.12) has an point-identified upper bound. Thus, given α, $FVD_{i,\ell}$ is bounded below by the expression (B.1).

We now argue that the lower bound (B.1) is attained by an admissible model with the given α. 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 $\{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t}$.[B.12] Then $\mathrm{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t}) = \sum_{j=2}^{n_\varepsilon}\sum_{m=0}^{\ell-1}\tilde{\Theta}_{i,j,m}^2$, so the following model attains the lower bound (B.1) and is consistent with the given spectrum $s_w(\cdot)$:

$$\begin{aligned} y_t &= \frac{1}{\alpha}\sum_{\ell=0}^\infty \mathrm{Cov}(y_t, \tilde{z}_{t-\ell})\bar{\varepsilon}_{1,t} + \sum_{\ell=0}^\infty \tilde{\Theta}_\ell \tilde{\varepsilon}_{t-\ell}, \\ \tilde{z}_t &= \alpha\bar{\varepsilon}_{1,t} + \sqrt{\mathrm{Var}(\tilde{z}_t) - \alpha^2} \times \bar{v}_t, \\ (\bar{\varepsilon}_{1,t}, \tilde{\varepsilon}_t', \bar{v}_t)' &\overset{i.i.d.}{\sim} N(0, I_{n_y+2}). \end{aligned} \tag{B.13}$$

2. Lemma B.1 implies that $\mathrm{Var}(\tilde{y}_{i,t+\ell}^{(\alpha)} \mid \{\tilde{y}_\tau^{(\alpha)}\}_{-\infty<\tau\le t})$ is increasing in α. Hence, the expression (B.1) is decreasing in α, as claimed. At $\alpha = \alpha_{UB}$, the representation (B.13) has $\tilde{z}_t = \alpha_{UB}\bar{\varepsilon}_{1,t}$, so we can represent $\tilde{y}_t^{(\alpha_{UB})} = y_t - E(y_t \mid \{\bar{\varepsilon}_{1,\tau}\}_{-\infty<\tau\le t}) = y_t - E(y_t \mid \{\tilde{z}_\tau\}_{-\infty<\tau\le t})$. □

B.11.4 Proof of Proposition B.2

The "only if" part was proved already in the text of Section 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

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

[B.12] 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 $\eta' s_{\tilde{z}}^{-1}$, post-multiplying by $s_{\tilde{z}}^{-1}\eta$, and rearranging the positive definiteness condition, we obtain the implication that

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

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

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

is then positive definite by Lemma 1. Moreover, the same lemma implies that

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

is positive definite for all $\omega \in [0, 2\pi]$. If we set $\overline{\Theta}_{\bullet,1}(L) = (2\pi/\overline{\alpha})\zeta(L)$, 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 $s_w(\omega)$:

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

Note that η assumes the role of λ. $\square$

B.11.5 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_ℓ. 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 γ, 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.11.6 Auxiliary lemmas for sieve VAR results

Here we define notation and state auxiliary lemmas used to prove the propositions in Section B.9. 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

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

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* $\tilde{\nu}_T \in \mathbb{R}^{n_W^2 p_T}$ *be a deterministic sequence of vectors such that* $\|\tilde{\nu}_T\|^2 \leq M < \infty$ *for all* T. *Define*

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

Then

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

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*

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

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

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

and

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

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

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

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

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

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

Lemma B.11. *Let Assumptions B.1 to B.3 hold. Define ν_T and ξ as in Section B.9.2. Then*

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

B.11.7 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

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

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 $\hat{\beta}(p_T)$. □

B.11.8 Proof of Lemma B.4

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

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

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

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

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

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

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

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,

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

B.11.9 Proof of Lemma B.6

By stationarity,

$$\begin{aligned}
&\frac{1}{T-p} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} \left| \operatorname{Cov}(e_{j_1,t+m_1} e_{j_2,t+m_2} e_{j_3,t} e_{j_4,t}, e_{j_1,s+m_3} e_{j_2,s+m_4} e_{j_3,s} e_{j_4,s}) \right| \\
&= \sum_{\ell=-(T-p-1)}^{T-p-1} \left(1 - \frac{|\ell|}{T-p}\right) \left| \operatorname{Cov}(e_{j_1,\ell+m_1} e_{j_2,\ell+m_2} e_{j_3,\ell} e_{j_4,\ell}, e_{j_1,m_3} e_{j_2,m_4} e_{j_3,0} e_{j_4,0}) \right|. \qquad \text{(B.14)}
\end{aligned}$$

We first argue that each term in the sum (B.14) is bounded. This follows from Cauchy-Schwarz:

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

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

Next, we show that at most 9 of the terms in the sum (B.14) are nonzero. Consider the term corresponding to a given index ℓ 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 ℓ have this property.

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

B.11.10 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_ℓ are absolutely summable across ℓ (Brockwell & Davis, 1991, p. 418).

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

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

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 $\operatorname{Var}(R_{1,T,i,j}) = O(T^{-1})$ uniformly in i, j. The typical element $R_{1,T,i,j}$ has the form

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

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

$$\frac{1}{(T-p_T)^2} \sum_{t=p_T+1}^{T} \sum_{s=p_T+1}^{T} \operatorname{Cov}(e_{j_1,t} W_{j_2,t-m_1} e_{j_3,t} W_{j_4,t-m_2}, e_{j_1,s} W_{j_2,s-m_1} e_{j_3,s} W_{j_4,s-m_2}). \quad \text{(B.15)}$$

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.15) equals

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

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

$$\frac{1}{T-p_T}\sum_{\ell_1,\ell_2,\ell_3,\ell_4=0}^{\infty}\sum_{b_1,b_2,b_3,b_4=1}^{n_W} |B_{j_2,b_1,\ell_1}B_{j_4,b_2,\ell_2}B_{j_2,b_3,\ell_3}B_{j_4,b_4,\ell_4}|\times 9E\|e_t\|^8 = O(T^{-1}), \quad \text{(B.16)}$$

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

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

Decompose it as

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

Since $\{\operatorname{vec}(e_t X_t(p_T)')\}$ 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

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

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

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

$$= \frac{1}{T-p_T} \sum_{\ell_1,\ell_2=0}^{\infty} \sum_{b_1,b_2=1}^{n_W} B_{j_2,b_1,\ell_1} B_{j_2,b_2,\ell_2}$$
$$\times \frac{1}{T-p_T} \sum_{s,t=p_T+1}^{T} \operatorname{Cov}\left(e_{j_1,t} e_{b_1,t-m-\ell_1} e_{j_3,t} e_{j_4,t},\, e_{j_1,s} e_{b_2,s-m-\ell_2} e_{j_3,s} e_{j_4,s}\right).$$

This expression is of order $O(T^{-1})$, for the same reason as (B.16) above. □

B.11.11 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 $v_\zeta \equiv \lim_{T\to\infty} v_{\zeta,T} = 0$, it is easy to show that $(T-p_T)^{1/2}\tilde{\nu}_T' \operatorname{vec}(\hat{\beta}(p_T) - \beta(p_T)) = o_p(1)$ 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

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

Since $\{e_t\}$ is i.i.d., e_t is independent of $X_t(p_T)$, so $\{J_{t,T}\}_{p_T+1\le t\le T}$ 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

$$\textstyle\sum_{t=p_T+1}^{T} J_{t,T}^2 \xrightarrow{p} 1 \tag{B.17}$$

and

$$\max_{p_T+1\le t\le T} |J_{t,T}| \xrightarrow{p} 0. \tag{B.18}$$

We first prove (B.17), following the univariate argument in Gonçalves & Kilian (2007, pp. 633–636). Decompose

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

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

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,

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

where the last line follows from Lemma B.7 and Assumptions B.1 and B.3. Second, we analogously have

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

again using Lemma B.7 and Assumptions B.1 and B.3. This concludes the proof of (B.17).

To prove (B.18), first note that since $E\|e_t\|^{4+\epsilon} < \infty$ for some $\epsilon > 0$, a standard argument for i.i.d. variables gives that $(T-p_T)^{-1/2} \max_{p_T+1 \leq t \leq T} |\lambda' \operatorname{vec}(e_t e_t' - \Sigma)| = o_p(1)$. Next, the same calculations as in equation (2.12) in Lewis & Reinsel (1985, p. 401) yield

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

for any $\tilde{\epsilon}>0$. Putting the previous two facts together, we obtain (B.18). □

B.11.12 Proof of Lemma B.9

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

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

using Lemma B.3. The argument for $A_{\sin}$ is identical. □

B.11.13 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 $\|B\|^2\equiv\operatorname{tr}(B^*B)$. The matrix perturbation bound $|\det(B)-\det(C)|\le n\|C-B\|\max\{\|B\|,\|C\|\}^{n-1}$ for $n\times n$ complex matrices B and C (Bhatia, 1997, Problem I.6.11, p. 22) implies

$$\begin{aligned}
&\left|\det(A(e^{i\omega}))-\det(I_{n_W}-\hat{A}_{\cos}(\omega;p_T)-i\hat{A}_{\sin}(\omega;p_T))\right|\\
&\le n_W\left\|\sum_{\ell=p_T+1}^{\infty}A_\ell e^{i\omega}-\sum_{\ell=1}^{p_T}(\hat{A}_\ell-A_\ell)e^{i\omega}\right\|\max\left\{\left\|\sum_{\ell=1}^{\infty}A_\ell e^{i\omega}\right\|,\left\|\sum_{\ell=1}^{p_T}\hat{A}_\ell e^{i\omega}\right\|\right\}^{n_W-1}. \qquad \text{(B.19)}
\end{aligned}$$

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.19) therefore tends to 0 in probability uniformly in ω, 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

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

uniformly in ω 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. □

B.11.14 Proof of Lemma B.11

We start out by applying a first-order Taylor expansion to the parameter of interest ψ. By Lemma B.10 and Assumption B.2, we can write

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

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

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

for all ω, with probability approaching 1. Since

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

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

$$\begin{aligned}
\hat{\psi}(p_T) - \psi = &\int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(\hat{A}_{\cos}(\omega;p_T) - A_{\cos}(\omega))\, d\omega \\
&+ \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(\hat{A}_{\sin}(\omega;p_T) - A_{\sin}(\omega))\, d\omega \\
&+ \int_0^{2\pi} h(\omega) g_3(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(\hat{\Sigma} - \Sigma)\, d\omega \\
&+ O_p(p_T/T) \\
= &\int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(\hat{A}_{\cos}(\omega;p_T) - A_{\cos}(\omega;p_T))\, d\omega \\
&+ \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(\hat{A}_{\sin}(\omega;p_T) - A_{\sin}(\omega;p_T))\, d\omega \\
&+ \int_0^{2\pi} h(\omega) g_1(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(A_{\cos}(\omega;p_T) - A_{\cos}(\omega))\, d\omega \qquad \text{(B.20)} \\
&+ \int_0^{2\pi} h(\omega) g_2(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)' \operatorname{vec}(A_{\sin}(\omega;p_T) - A_{\sin}(\omega))\, d\omega \qquad \text{(B.21)} \\
&+ \xi' \operatorname{vec}(\hat{\Sigma} - \Sigma) \\
&+ O_p(p_T/T).
\end{aligned}$$

We now bound the nonparametric bias term (B.20); the argument for (B.21) is similar. Note that $h(\cdot)$ is bounded, and

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

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.20)–(B.21) are each $o(T^{-1/2})$.

To complete the proof, observe that

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

In conclusion,

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

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

B.11.15 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

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

The sum

$$\sum_{\ell=1}^{p_T}\left(\frac{1}{2\pi}\int_0^{2\pi} g_{1,i}(A_{\cos}(\omega), A_{\sin}(\omega), \Sigma)\cos(\omega\ell)\,d\omega\right)^2 \tag{B.22}$$

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 $\{\omega \mapsto \cos(\omega\ell)\}_{1\le\ell\le p_T}$. Bessel's inequality therefore states that (B.22) 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.22) 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,

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

using obvious notation for the L_2 norms. These norms are finite by Assumption B.2. □

B.11.16 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 $\hat{\sigma}^2_\psi(p_T)$ 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 $\Xi\equiv\mathrm{Var}(e_t\otimes e_t)$ 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 $\chi_t \equiv \mathrm{vec}(e_t e_t' - \Sigma)$, 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

$$
\begin{aligned}
\|\hat{\Xi}(p_T) - \Xi\| &\le \frac{1}{T - p_T} \sum_{t=p_T+1}^{T} \|\hat{\chi}_t \hat{\chi}_t' - \chi_t \chi_t'\| \\
&\le \frac{1}{T - p_T} \sum_{t=p_T+1}^{T} \|\hat{\chi}_t - \chi_t\|^2 + \frac{2}{T - p_T} \sum_{t=p_T+1}^{T} \|\hat{\chi}_t - \chi_t\| \, \|\chi_t\| \\
&\le \frac{1}{T - p_T} \sum_{t=p_T+1}^{T} \|\hat{\chi}_t - \chi_t\|^2 + 2\Bigg(\frac{1}{T - p_T} \sum_{t=p_T+1}^{T} \|\hat{\chi}_t - \chi_t\|^2 \\
&\qquad \times \frac{1}{T - p_T} \sum_{t=p_T+1}^{T} \|\chi_t\|^2 \Bigg)^{1/2}
\end{aligned}
$$

by Cauchy-Schwarz, we just need to show that

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

Since

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

we have

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

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

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

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,

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

since

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

Second,

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

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.

Caldara, D. & Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars. *American Economic Journal: Macroeconomics*, *11*(1), 157–92.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. Oxford University Press.

Forni, M., Gambetti, L., & Sala, L. (2018). Structural VARs and Non-invertible Macroeconomic Models. *Journal of Applied Econometrics*. Forthcoming.

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 nootstrap inference for AR(∞) 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.

Imbens, G. W. & Manski, C. F. (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica*, *72*(6), 1845–1857.

Kilian, L. & Kim, Y. J. (2011). How Reliable Are Local Projection Estimators of Impulse Responses? *Review of Economics and Statistics*, *93*(4), 1460–1466.

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.

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.

Stoye, J. (2009). More on Confidence Intervals for Partially Identified Parameters. *Econometrica*, *77*(4), 1299–1315.

Wolf, C. K. (2018). SVAR (Mis)Identification and the Real Effects of Monetary Policy. Manuscript, Princeton University.