Online Appendix for "Doubly Robust Local Projections and Some Unpleasant VARithmetic"

José Luis Montiel Olea Mikkel Plagborg-Møller
Eric Qian Christian K. Wolf

November 11, 2025

Appendix C Further theoretical results

C.1 Heteroskedasticity

The conclusions of Propositions 3.1 and 3.2 do not require shock independence, either cross-sectionally or across time. In particular, our proofs of these propositions and the auxiliary lemmas in Supplemental Appendix E replace Assumption 2.1(i) by the following:[C.1]

Assumption C.1. *ε_t is a strictly stationary martingale difference sequence with respect to its natural filtration; $\operatorname{Var}(\varepsilon_t) = D \equiv \operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2)$; $\sigma_j > 0$ for all $j = 1, \dots, m$; $E[\|\varepsilon_t\|^4] < \infty$; and $\sum_{\ell=1}^{\infty} \sum_{\tau=1}^{\infty} \|\operatorname{Cov}(\varepsilon_t \otimes \varepsilon_t, \varepsilon_{t-\ell} \otimes \varepsilon_{t-\tau})\| < \infty$.*

Assumption C.1 strictly weakens Assumption 2.1(i) by allowing the shocks to be conditionally heteroskedastic and by weakening mutual shock independence to orthogonality. The last part of Assumption C.1 is a vector version of the fourth-order cumulant summability condition of Kuersteiner (2001, Assumption A1). It restricts the higher-order dependence of the shocks, consistent with many stationary models of conditional heteroskedasticity.[C.2]

C.2 External instruments and proxies

Our framework can accommodate identification via an external instrument (also known as a proxy) by a simple reparametrization. To see this, let the proxy be ordered first in y_t. Set

[C.1] It is an interesting topic for future research to investigate whether the other results in Sections 3 and 4 also hold under the weaker condition.

[C.2] One sufficient condition is finite dependence, i.e., there exists an integer K such that $\{\varepsilon_s\}_{s \geq t+K}$ is independent of $\{\varepsilon_s\}_{s \leq t}$. Another set of sufficient conditions is that $E(\varepsilon_t \mid \{\varepsilon_s\}_{s \neq t}) = 0$ and $\{\varepsilon_t \otimes \varepsilon_t\}$ has absolutely summable autocovariance function. See also Kuersteiner (2001, Remark 2, p. 362).

$j^* = 1$, and replace Assumption 2.1(iii) with the following:

Assumption C.2. *The first column of A consists of zeros, except possibly the first element; the first row of H equals $(1, 0, 0, \dots, 0, 1)$; the last column of H consists of zeros, except the first element; and the last column of $\alpha(L)$ consists of zeros.*

This assumption imposes the following restrictions:

- The proxy $y_{1,t}$ equals

$$y_{1,t} = A_{1,\bullet} y_{t-1} + \varepsilon_{1,t} + \varepsilon_{m,t} + T^{-\zeta}[\alpha_{1,\bullet}(L) + \alpha_{m,\bullet}(L)]\varepsilon_t,$$

  which generalizes Assumption 4 in Plagborg-Møller and Wolf (2021) to allow for local contamination by lagged shocks. The last shock $\varepsilon_{m,t}$ is viewed as measurement error or noise (v_t in the notation of Plagborg-Møller and Wolf, 2021).

- The dynamics of $(y_{2,t}, \dots, y_{n,t})$ (i.e., with the proxy excluded) follow a VAR(1), up to the local misspecification in the form of lags of $(\varepsilon_{1,t}, \dots, \varepsilon_{m-1,t})$.

- The proxy measurement error $\varepsilon_{m,t}$ is orthogonal to all leads and lags of $(y_{2,t}, \dots, y_{n,t})$.

Now transform the shocks from ε_t to $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1,t}, \dots, \tilde{\varepsilon}_{m,t})'$ as follows:

$$\tilde{\varepsilon}_{1,t} \equiv \varepsilon_{1,t} + \varepsilon_{m,t}; \quad \tilde{\varepsilon}_{j,t} \equiv \varepsilon_{j,t} \text{ for } j = 2, \dots, m-1; \quad \tilde{\varepsilon}_{m,t} \equiv \varepsilon_{m,t} - \frac{\sigma_m^2}{\sigma_1^2 + \sigma_m^2}(\varepsilon_{1,t} + \varepsilon_{m,t}).$$

By construction, the elements of $\tilde{\varepsilon}_t$ are mutually orthogonal. Note that

$$\varepsilon_{1,t} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_m^2}\tilde{\varepsilon}_{1,t} - \tilde{\varepsilon}_{m,t}; \quad \varepsilon_{j,t} = \tilde{\varepsilon}_{j,t} \text{ for } j = 2, \dots, m-1; \quad \varepsilon_{m,t} = \frac{\sigma_m^2}{\sigma_1^2 + \sigma_m^2}\tilde{\varepsilon}_{1,t} + \tilde{\varepsilon}_{m,t};$$

write Q for the $m \times m$ matrix such that $\varepsilon_t = Q\tilde{\varepsilon}_t$. We can then re-express the VARMA$(1, \infty)$ model for y_t in Equation (2.1) as

$$y_t = Ay_{t-1} + \tilde{H}[I + T^{-\zeta}\tilde{\alpha}(L)]\tilde{\varepsilon}_t, \tag{C.1}$$

where

$$\tilde{H} \equiv HQ = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H_{2,1} & H_{2,2} & H_{2,3} & \cdots & H_{2,m-1} & -H_{2,1} \\ \vdots & & & & & \vdots \\ \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H_{n,1} & H_{n,2} & H_{n,3} & \cdots & H_{n,m-1} & -H_{n,1} \end{pmatrix},$$

and $\tilde{\alpha}(L) \equiv Q^{-1}\alpha(L)Q$. Under Assumption C.2 above, the impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\tilde{\varepsilon}_{1,t}$ in the reparametrized system (C.1) equal $\sigma_1^2/(\sigma_1^2 + \sigma_m^2)$ times the impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\varepsilon_{1,t}$ in the original parametrization in Equation (2.1). This follows by inspection of the elements $\tilde{H}_{i,1}$ for $i \geq 2$, the assumption that $A_{i,1} = 0$ for $i \geq 2$, and the fact that the first column of $\tilde{H}\tilde{\alpha}(L)$ equals $H\alpha(L)Q_{\bullet,1} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H\alpha_{\bullet,1}(L)$ (here we use the assumption that the last column of $\alpha(L)$ is zero).

If the original shocks (and measurement error) ε_t satisfy Assumption C.1, then the transformed shocks $\tilde{\varepsilon}_t$ do also, provided that the fourth-order cumulant condition holds for the transformed shocks (recall that Assumption C.1 only requires the shocks to be mutually orthogonal, not independent). The transformed system (C.1) therefore satisfies Assumption C.1 and Assumption 2.1(ii)–(v), with $\tilde{H}$ and $\tilde{\alpha}(L)$ in place of H and $\alpha(L)$. Hence, Propositions 3.1 and 3.2 apply. We conclude that LPs on the proxy $y_{1,t}$—and recursive VARs with the proxy ordered first—consistently estimate the true impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\varepsilon_{1,t}$, up to the scale factor $\frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2}$. This scale factor, which is the same across all response variables i^* and horizons h, reflects attenuation bias caused by the measurement error in the proxy. It gets canceled out if one reports relative (i.e., unit-effect-normalized) impulse responses, as explained by Plagborg-Møller and Wolf (2021, Section 3.3). Of course, though the VAR estimator is consistent, it suffers from asymptotic bias of order $T^{-\zeta}$, while the LP estimator has bias of the smaller order $T^{-2\zeta}$.

In summary, the main results in the paper carry over to identification via an external instrument/proxy. A caveat is that the instrument/proxy must be strong, in the sense that $\sigma_1^2 = \mathrm{Var}(\varepsilon_{1,t})$ is not close to zero; otherwise, we will end up dividing by a number close to zero when computing relative impulse responses.

C.3 Data-dependent lag selection

Here we argue that local projection inference is more robust to data-dependent lag selection errors than conventional VAR inference. The following proposition establishes the properties of conventional information criteria in our class of DGPs.

Proposition C.1. *Assume that the $\check{n}$-dimensional process $\{\check{y}_t\}$ is a stationary solution of the local-to-SVAR(p_0) model in Equation (2.2). Assume also that $\check{A}_{p_0} \neq 0$ so that p_0 is the minimal true autoregressive lag order, $\check{H}D\check{H}'$ is positive definite, and the process (written in companion form as on p. 7 of the main paper) satisfies Assumption 2.1.*

Let $\hat{\check{\Sigma}}(p)$ denote the $\check{n} \times \check{n}$ sample residual variance-covariance matrix from a least-squares VAR(p) regression on the data $(\check{y}_1, \dots, \check{y}_T)$. Let $\{g_T\}_T$ be a deterministic scalar sequence. Fix *a maximal lag length $K \geq p_0$. Suppose we select the lag length by minimizing an information criterion:*

$$\hat{p} \equiv \underset{0 \leq p \leq K}{\operatorname{argmin}} \left\{ \log \det \left(\hat{\check{\Sigma}}(p) \right) + p \times g_T \right\}.$$

Then the following statements hold:

i) If $g_T \to 0$, then $P(\hat{p} < p_0) \to 0$ as $T \to \infty$.

ii) If $T g_T \to \infty$ and $\zeta \geq 1/2$, then $P(\hat{p} > p_0) \to 0$ as $T \to \infty$.

The proposition implies that, when applied to a VAR in the data $\{\check{y}_t\}$, both the BIC ($g_T = \check{n}^2(\log T)/T$) and AIC ($g_T = 2\check{n}^2/T$) select p_0 or more lags with high probability in large samples, and in fact the BIC selects exactly p_0 lags asymptotically when $\zeta \geq 1/2$.[C.3] Hence, in the latter case, it follows from Corollary 3.1 that it is pointwise asymptotically valid to report a local projection confidence interval that controls for $\hat{p}$ lags of the data, where $\hat{p}$ is selected by applying the BIC to auxiliary VAR regressions as defined above.

Unlike LP inference, VAR inference is sensitive to minor model selection errors. Leeb and Pötscher (2005) show that VAR confidence intervals with lag length selected by BIC or AIC fail to control coverage uniformly over the VAR parameter space. The breakdown in performance happens for (a sequence of) DGPs that satisfy a VAR($\tilde{p}_0$) model where the first p_0 lags have large coefficients and the remaining $\tilde{p}_0 - p_0$ lags have small coefficients of order $T^{-1/2}$; this implies a VARMA(p_0, ∞) representation where the moving average coefficients are of order $T^{-1/2}$. Because the BIC or AIC cannot reliably detect the small coefficients, they will tend to select fewer than $\tilde{p}_0$ lags (cf. Proposition C.1). The small coefficients on the omitted lags impart an asymptotic bias in the VAR estimator that can cause large coverage distortions for the associated confidence interval, consistent with Corollary 3.1. However, if in this context the BIC-selected lag length is instead used for *local projection* inference, the bias imparted by the omitted lags is much smaller than for the VAR estimator and is in fact asymptotically negligible: this is precisely the message of Proposition 3.1. Though our arguments fall short of proving that local projection inference with data-dependent lag length is *uniformly* valid over some parameter space, they nevertheless show that LP inference remains valid under the types of drifting parameter sequences that Leeb and Pötscher (2005)

[C.3] For additional intuition, consider the critical case $\zeta = 1/2$ where the moving average coefficients are of order $T^{-1/2}$. Under smoothness assumptions on the density of the shocks ε_t, the local-to-SVAR(p_0) DGP is then contiguous to an exact SVAR(p_0) DGP (for formal results, see Hallin and Puri, 1988; Hallin, Ingenbleek, and Puri, 1989). Since the BIC is consistent for p_0 in latter DGP, contiguity implies that the BIC also selects p_0 lags with probability approaching 1 in the former DGP.

show are responsible for the non-uniformity of VAR inference.

Proof of Proposition C.1. The proof follows standard arguments for exact VAR models, see Lütkepohl (2005, Chapter 4.3.2) and references therein. We merely show that extra terms induced by the vanishing moving average process are asymptotically negligible under our assumptions. Let $\|\cdot\|$ denote the Frobenius norm.

Lemma E.6 implies that $T^{-1}\sum_{t=1}^T \check{y}_t\check{y}_{t-\ell}' = T^{-1}\sum_{t=1}^T \tilde{\check{y}}_t\tilde{\check{y}}_{t-\ell}' + O_p(T^{-\zeta})$ for any $\ell \geq 0$, where $\{\tilde{\check{y}}_t\}$ is a process that satisfies the VAR(p_0) model with no moving average term ($\alpha(L) = 0$), see also Corollary 3.2. It follows from least-squares algebra that the probability limit of $\hat{\check{\Sigma}}(p)$ for any fixed p is the same as it would be in an exact VAR(p_0) DGP with no moving average term. Statement (i) of the proposition then follows immediately from standard arguments for lag length estimation in exact VAR models (Lütkepohl, 2005, Proposition 4.2).

To prove statement (ii), assume $p \geq p_0$. It suffices to show that $\hat{\check{\Sigma}}(p) = \hat{\check{\Sigma}}_0 + O_p(T^{-1})$ for some data-dependent matrix $\hat{\check{\Sigma}}_0$ independent of p, since this implies[C.4]

$$P\left(\log\det\left(\hat{\check{\Sigma}}(p)\right) + pg_T > \log\det\left(\hat{\check{\Sigma}}(p_0)\right) + p_0 g_T\right) = P\Big((p-p_0)Tg_T > O_p(1)\Big) \to 1$$

when $p > p_0$. Let y_t denote the ($\check{n}p$)-dimensional companion form vector obtained by stacking p lags of $\check{y}_t$, as in Equation (2.2). Then $\hat{\check{\Sigma}}(p)$ is the upper left $\check{n} \times \check{n}$ block of the $(\check{n}p) \times (\check{n}p)$ matrix $\hat{\Sigma} = T^{-1}\sum_{t=1}^T \hat{u}_t\hat{u}_t'$ defined in Section 2.2. To finish the proof, we show that $T^{-1}\sum_{t=1}^T \|\hat{u}_t - H\varepsilon_t\|^2 = O_p(T^{-1})$, which by Cauchy-Schwarz implies $\hat{\Sigma} = T^{-1}\sum_{t=1}^T H\varepsilon_t\varepsilon_t'H' + O_p(T^{-1})$, as needed.

Note that when the estimation lag length p weakly exceeds the true autoregressive order p_0, all results in our paper apply, as noted in the discussion surrounding Equation (2.2). Hence, using the definition of u_t in Lemma E.6,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \|\hat{u}_t - H\varepsilon_t\|^2 &\leq \frac{2}{T}\sum_{t=1}^T \|\underbrace{\hat{u}_t - u_t}_{(A-\hat{A})y_{t-1}}\|^2 + \frac{2}{T}\sum_{t=1}^T \|\underbrace{u_t - H\varepsilon_t}_{T^{-\zeta}H\alpha(L)\varepsilon_t}\|^2 \\
&\leq 2\underbrace{\|\hat{A}-A\|^2}_{O_p((T^{-1/2}+T^{-\zeta})^2)}\underbrace{\frac{1}{T}\sum_{t=1}^T \|y_{t-1}\|^2}_{O_p(1)} + 2T^{-2\zeta}\|H\|^2\underbrace{\frac{1}{T}\sum_{t=1}^T \|\alpha(L)\varepsilon_t\|^2}_{O_p(1)} \\
&= O_p(T^{-1} + T^{-2\zeta}),
\end{aligned}$$

[C.4] Actually, in order to apply the delta method to the log determinant, we also use that plim $\hat{\check{\Sigma}}_0 = \check{H}D\check{H}'$, a matrix that is non-singular by assumption.

where the three $O_p(\cdot)$ statements in the penultimate line rely on Lemmas E.2 and E.8 and Assumption 2.1(v), respectively. When $\zeta \geq 1/2$, the right-hand side above is $O_p(T^{-1})$. □

C.4 Inference on multiple impulse responses

This subsection generalizes the worst-case bias formula in Proposition 4.1 to the multi-dimensional case and derives the worst-case coverage of the Wald confidence ellipsoid.

SET-UP. We consider inference on any combination of impulse responses for various horizons h, response variables i^*, and shocks j^*. When referring to impulse responses and estimators of these, we need to make the response variable and shock explicit in the notation. Thus, we write $\theta_{i^*,j^*,h,T}$, $\hat{\beta}_{i^*,j^*,h}$, and $\hat{\delta}_{i^*,j^*,h}$, with the definitions being the same as in Section 2. Let k denote the total number of impulse responses of interest. We refer to the list of impulse responses by the collection of triples $\{(i_a^*, j_a^*, h_a)\}_{a=1}^k$ indexing the response variable, shock variable, and horizon, respectively. Define the k-dimensional vectors of true impulse responses and LP and VAR estimators:

$$\boldsymbol{\theta}_T \equiv \begin{pmatrix} \theta_{i_1^*,j_1^*,h_1,T} \\ \vdots \\ \theta_{i_k^*,j_k^*,h_k,T} \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} \equiv \begin{pmatrix} \hat{\beta}_{i_1^*,j_1^*,h_1} \\ \vdots \\ \hat{\beta}_{i_k^*,j_k^*,h_k} \end{pmatrix}, \quad \hat{\boldsymbol{\delta}} \equiv \begin{pmatrix} \hat{\delta}_{i_1^*,j_1^*,h_1} \\ \vdots \\ \hat{\delta}_{i_k^*,j_k^*,h_k} \end{pmatrix}.$$

It follows from Propositions 3.1 and 3.2 that, when $\zeta = 1/2$,

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\theta}_T \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\theta}_T \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0_{k\times 1} \\ \mathrm{aBias}(\hat{\boldsymbol{\delta}}) \end{pmatrix}, \begin{pmatrix} \mathrm{aVar}(\hat{\boldsymbol{\beta}}) & \mathrm{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) \\ \mathrm{aCov}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) & \mathrm{aVar}(\hat{\boldsymbol{\delta}}) \end{pmatrix} \right), \tag{C.2}$$

for a k-dimensional vector $\mathrm{aBias}(\hat{\boldsymbol{\delta}})$ (defined in the proof of Proposition C.2 below) and $k \times k$ matrices $\mathrm{aVar}(\hat{\boldsymbol{\beta}})$, $\mathrm{aVar}(\hat{\boldsymbol{\delta}})$, and $\mathrm{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$ given in Corollary A.2 in Appendix A.3. This corollary also implies that the difference $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\delta}}$ is asymptotically independent of $\hat{\boldsymbol{\delta}}$, which is not surprising given the general arguments of Hausman (1978) and the facts that (i) the asymptotic variances of the estimators are the same as in the model with $\alpha(L) = 0$ and (ii) the VAR estimator is the quasi-MLE in such a model. It follows that $\mathrm{aVar}(\hat{\boldsymbol{\beta}}) \geq \mathrm{aVar}(\hat{\boldsymbol{\delta}})$ in the positive semidefinite sense.

WORST-CASE BIAS. The following result generalizes the univariate worst-case bias formula in Proposition 4.1.

Proposition C.2. *Impose Assumption 2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$, and let $\zeta = 1/2$. Let R be a constant matrix with k columns. Then*

$$\max_{\alpha(L)\colon \|\alpha(L)\| \leq M} \|R \,\mathrm{aBias}(\hat{\boldsymbol{\delta}})\|^2 = M^2 \lambda_{\max}\Big(R[\mathrm{aVar}(\hat{\boldsymbol{\beta}}) - \mathrm{aVar}(\hat{\boldsymbol{\delta}})]R'\Big),$$

where $\lambda_{\max}(B)$ denotes the largest eigenvalue of the matrix B.

The proposition shows that the worst-case squared norm of the bias of the VAR estimator $R\hat{\boldsymbol{\delta}}$ of $R\boldsymbol{\theta}_T$ is a function of two simple quantities: the bound M on misspecification, and the largest eigenvalue of the difference $\mathrm{aVar}(R\hat{\boldsymbol{\beta}}) - \mathrm{aVar}(R\hat{\boldsymbol{\delta}})$ between the variance-covariance matrices for the LP and VAR estimators. The latter eigenvalue equals $\max_{\|\varsigma\|=1}\{\mathrm{aVar}(\varsigma' R\hat{\boldsymbol{\beta}}) - \mathrm{aVar}(\varsigma' R\hat{\boldsymbol{\delta}})\}$, i.e., the *largest* efficiency gain for VAR over LP across all linear combinations (with norm 1) of the estimated parameters. Consequently, the worst-case bias is non-negligible if the VAR offers efficiency gains for *any* linear combination of the parameters of interest, echoing our univariate results. When R is a row vector, then the proposition implies that our conclusions from Section 4.2 extend to inference on any *linear combination* of impulse responses.

Worst-case coverage of confidence ellipsoid. We next derive the coverage of the conventional Wald confidence ellipsoid based on the VAR estimator. The level-$(1-a)$ confidence ellipsoid is given by

$$\mathrm{CE}(\hat{\boldsymbol{\delta}}) \equiv \left\{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^k \colon T(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}})' \,\mathrm{aVar}(\hat{\boldsymbol{\delta}})^{-1}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}}) \leq \chi^2_{1-a,k}\right\},$$

where $\chi^2_{1-a,k}$ is the $1-a$ quantile of the χ^2 distribution with k degrees of freedom.

Corollary C.1. *Impose Assumption 2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$, and let $\zeta = 1/2$. Assume also that $\mathrm{aVar}(\hat{\boldsymbol{\delta}})$ is non-singular. Then*

$$\min_{\alpha(L)\colon \|\alpha(L)\| \leq M} \lim_{T\to\infty} P(\boldsymbol{\theta}_T \in \mathrm{CE}(\hat{\boldsymbol{\delta}})) = F_k\left(\chi^2_{1-a,k}; M^2\left[\lambda_{\max}(\mathrm{aVar}(\hat{\boldsymbol{\beta}})\,\mathrm{aVar}(\hat{\boldsymbol{\delta}})^{-1}) - 1\right]\right),$$

where $F_k(x;c)$ is the cumulative distribution function, evaluated at point $x \geq 0$, of a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter $c \geq 0$.

The worst-case coverage probability of the VAR confidence ellipsoid depends on three scalars: the bound M on misspecification, the dimension k of the ellipsoid, and the "multivariate relative standard error”

$$\sqrt{\lambda_{\min}(\mathrm{aVar}(\hat{\boldsymbol{\delta}})\,\mathrm{aVar}(\hat{\boldsymbol{\beta}})^{-1})} = [\lambda_{\max}(\mathrm{aVar}(\hat{\boldsymbol{\beta}})\,\mathrm{aVar}(\hat{\boldsymbol{\delta}})^{-1})]^{-1/2} = \min_{\varsigma\in\mathbb{R}^k}\sqrt{\mathrm{aVar}(\varsigma'\hat{\boldsymbol{\delta}})/\,\mathrm{aVar}(\varsigma'\hat{\boldsymbol{\beta}})}.$$

Again, the worst-case coverage distortion is an increasing function of the *largest* efficiency gain for VAR over LP across *all* linear combinations of the impulse responses. Since VAR impulse response estimates are often highly correlated across horizons, this suggests that the VAR undercoverage can in fact be particularly severe in the multivariate case. Numerical calculations (available upon request from the authors) show that the coverage distortions can be severe even when $M = 1$, regardless of the dimension k.

Proof of Proposition C.2. Define $\tilde{\alpha}_\ell = D^{-1/2}\alpha_\ell D^{1/2}$ for all $\ell \geq 1$. Notice that $\|\alpha(L)\|^2 = \sum_{\ell=1}^{\infty}\|\tilde{\alpha}_\ell\|^2$.

By Proposition 3.2, we have $\mathrm{aBias}(\hat{\delta}_{i^*,j^*,h}) = \sum_{\ell=1}^{\infty}\mathrm{trace}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell)$, where

$$\Xi_{i^*,j^*,h,\ell} \equiv D^{1/2}H'(A')^{\ell-1}S^{-1}\Psi_{i^*,j^*,h}HD^{1/2} - \mathbb{1}(\ell\leq h)D^{-1/2}e_{j^*,m}e_{i^*,n}'A^{h-\ell}HD^{1/2}.$$

Since $\mathrm{trace}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell) = \mathrm{vec}(\Xi_{i^*,j^*,h,\ell})'\,\mathrm{vec}(\tilde{\alpha}_\ell')$, we can write

$$\mathrm{aBias}(\hat{\boldsymbol{\delta}}) = \sum_{\ell=1}^{\infty}\Upsilon_\ell\,\mathrm{vec}(\tilde{\alpha}_\ell'),$$

where

$$\Upsilon_\ell \equiv \Big(\mathrm{vec}(\Xi_{i_1^*,j_1^*,h_1,\ell}),\ldots,\mathrm{vec}(\Xi_{i_k^*,j_k^*,h_k,\ell})\Big)' \in \mathbb{R}^{k\times m^2}.$$

Hence,

$$\max_{\alpha(L)\colon\,\|\alpha(L)\|\leq M}\|R\,\mathrm{aBias}(\hat{\boldsymbol{\delta}})\|^2 = \max_{\{\tilde{\alpha}_\ell\}_{\ell=1}^{\infty}\colon\,\sum_{\ell=1}^{\infty}\|\tilde{\alpha}_\ell'\|^2\leq M^2}\left\|\sum_{\ell=1}^{\infty}R\Upsilon_\ell\,\mathrm{vec}(\tilde{\alpha}_\ell')\right\|^2.$$

Lemma C.1 below shows that the final expression above equals $M^2\lambda_{\max}(\sum_{\ell=1}^{\infty}R\Upsilon_\ell\Upsilon_\ell'R')$ (the lemma only explicitly considers the case $M = 1$, but the general case then follows from the homogeneity of degree 1 of the norm). Finally, Lemma C.2 below shows that $\sum_{\ell=1}^{\infty}\Upsilon_\ell\Upsilon_\ell' = \mathrm{aVar}(\hat{\boldsymbol{\beta}}) - \mathrm{aVar}(\hat{\boldsymbol{\delta}})$. This completes the proof of the proposition. The proof of Lemma C.1 shows that the maximum above is achieved when $\mathrm{vec}(\tilde{\alpha}_\ell') \propto \Upsilon_\ell'v$ (with the constant of proportionality being independent of ℓ and chosen to satisfy the norm constraint), where v is the eigenvector corresponding to the largest eigenvalue of $R[\mathrm{aVar}(\hat{\boldsymbol{\beta}})-\mathrm{aVar}(\hat{\boldsymbol{\delta}})]R'$.

In the univariate case $k = 1$, this reduces to expression (4.1) in Section 4.2. $\square$

Lemma C.1. *Let $\mathcal{X}$ denote the set of sequences $\{x_\ell\}_{\ell=1}^\infty$ of $m \times m$ matrices x_ℓ satisfying $\sum_{\ell=1}^\infty \|x_\ell\|^2 \leq 1$. Let $\{L_\ell\}_{\ell=1}^\infty$ be a sequence of $r \times m^2$ matrices L_ℓ satisfying $\sum_{\ell=1}^\infty \|L_\ell\|^2 < \infty$. Then*

$$\max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=1}^\infty L_\ell \operatorname{vec}(x_\ell) \right\|^2 = \lambda_{\max}\left(\sum_{\ell=1}^\infty L_\ell L_\ell' \right). \tag{C.3}$$

Proof. A short proof using abstract functional analysis is available upon request from the authors. Below we provide a more elementary proof.

The statement of the lemma is obvious if $\sum_{\ell=1}^\infty \|L_\ell\|^2 = 0$, in which case both sides of the above display equal 0. Hence, we may assume that the series $V \equiv \sum_{\ell=1}^\infty L_\ell L_\ell'$ converges to a non-zero matrix. Let v be the unit-length eigenvector corresponding to the largest eigenvalue $\lambda \equiv \lambda_{\max}(V) \in (0, \infty)$ of V.

The purported maximum (C.3) is achieved by the sequence $\{x_\ell^*\}$ given by $\operatorname{vec}(x_\ell^*) = \lambda^{-1/2} L_\ell' v$:

$$\left\| \sum_{\ell=1}^\infty L_\ell \operatorname{vec}(x_\ell^*) \right\|^2 = \left\| \lambda^{-1/2} \sum_{\ell=1}^\infty L_\ell L_\ell' v \right\|^2 = \lambda^{-1} \|Vv\|^2 = \lambda^{-1} \|\lambda v\|^2 = \lambda \|v\|^2 = \lambda,$$

and

$$\sum_{\ell=1}^\infty \|x_\ell^*\|^2 = \sum_{\ell=1}^\infty \operatorname{vec}(x_\ell^*)' \operatorname{vec}(x_\ell^*) = \lambda^{-1} v' \sum_{\ell=1}^\infty L_\ell L_\ell' v = \lambda^{-1} v' V v = \lambda^{-1} \lambda = 1.$$

We complete the proof by showing that the left-hand side of (C.3) is bounded above by the right-hand side. Let K be an arbitrary positive integer. Then

$$\max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=1}^\infty L_\ell \operatorname{vec}(x_\ell) \right\| \leq \max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=1}^K L_\ell \operatorname{vec}(x_\ell) \right\| + \max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=K+1}^\infty L_\ell \operatorname{vec}(x_\ell) \right\|.$$

The second term on the right-hand side is bounded above by $(\sum_{\ell=K+1}^\infty \|L_\ell\|^2)^{1/2}$ by Cauchy-Schwarz. As for the first term, standard results for the eigenvalues of finite-dimensional matrices yield

$$\max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=1}^K L_\ell \operatorname{vec}(x_\ell) \right\|^2 = \max_{x \in \mathbb{R}^{Km^2} \colon \|x\| \leq 1} \left\| \begin{pmatrix} L_1 & L_2 & \cdots & L_K \end{pmatrix} x \right\|^2$$

$$= \lambda_{\max}\left(\begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix}' \begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix} \right) = \lambda_{\max}\left(\begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix} \begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix}' \right) = \lambda_{\max}\left(\sum_{\ell=1}^K L_\ell L_\ell' \right).$$

We have shown

$$\max_{\{x_\ell\}_{\ell=1}^{\infty} \in \mathcal{X}} \left\| \sum_{\ell=1}^{\infty} L_\ell \operatorname{vec}(x_\ell) \right\| \leq \left(\lambda_{\max} \left(\sum_{\ell=1}^{K} L_\ell L_\ell' \right) \right)^{1/2} + \left(\sum_{\ell=K+1}^{\infty} \|L_\ell\|^2 \right)^{1/2}.$$

Now let $K \to \infty$. Since $\sum_{\ell=1}^{\infty} L_\ell L_\ell'$ is a convergent series, the first term on the right-hand side above converges to $\lambda^{1/2}$ by continuity of eigenvalues, while the second term converges to 0. This establishes the required bound. □

Lemma C.2. *Under the assumptions of Proposition C.2, and using the notation in the proof of that proposition, we have*

$$\sum_{\ell=1}^{\infty} \Upsilon_\ell \Upsilon_\ell' = \operatorname{aVar}(\hat{\boldsymbol{\beta}}) - \operatorname{aVar}(\hat{\boldsymbol{\delta}}).$$

Proof. By definition of Υ_ℓ, it suffices to show that, for any indices $a, b \in \{1, \dots, k\}$,

$$\sum_{\ell=1}^{\infty} \operatorname{vec}(\Xi_{i_a^*, j_a^*, h_a, \ell})' \operatorname{vec}(\Xi_{i_b^*, j_b^*, h_b, \ell}) = \operatorname{aCov}(\hat{\beta}_{i_a^*, j_a^*, h_a}, \hat{\beta}_{i_b^*, j_b^*, h_b}) - \operatorname{aCov}(\hat{\delta}_{i_a^*, j_a^*, h_a}, \hat{\delta}_{i_b^*, j_b^*, h_b}). \tag{C.4}$$

Multiplying out terms, we find that the left-hand side above equals

$$\begin{aligned}
\sum_{\ell=1}^{\infty} \operatorname{trace}(\Xi_{i_a^*, j_a^*, h_a, \ell}' \Xi_{i_b^*, j_b^*, h_b, \ell}) = & \sum_{\ell=1}^{\infty} \operatorname{trace}\left(A^{\ell-1} \Sigma (A')^{\ell-1} S^{-1} \Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) \\
& - \sum_{\ell=1}^{h_b} \operatorname{trace}\left(A^{\ell-1} H_{\bullet, j_b^*} e_{i_b^*, n}' A^{h_b - \ell} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) \\
& - \sum_{\ell=1}^{h_a} \operatorname{trace}\left(A^{\ell-1} H_{\bullet, j_a^*} e_{i_a^*, n}' A^{h_a - \ell} \Sigma \Psi_{i_b^*, j_b^*, h_b}' S^{-1} \right) \\
& + \mathbb{1}(j_a^* = j_b^*) \sigma_{j_a^*}^{-2} \sum_{\ell=1}^{\min\{h_a, h_b\}} e_{i_b^*, n}' A^{h_b - \ell} \Sigma (A')^{h_a - \ell} e_{i_a^*, n}.
\end{aligned}$$

We now evaluate each of the four terms on the right-hand side above. The *first* term equals

$$\operatorname{trace}\left(\underbrace{\sum_{\ell=1}^{\infty} A^{\ell-1} \Sigma (A')^{\ell-1}}_{=S} S^{-1} \Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) = \operatorname{trace}\left(\Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right).$$

The *second* term (in the earlier display) equals

$$-\operatorname{trace}\left(\underbrace{\sum_{\ell=1}^{h_b} A^{\ell-1}H_{\bullet,j_b^*}e'_{i_b^*,n}A^{h_b-\ell}}_{=\sum_{\ell=1}^{h_b} A^{h_b-\ell}H_{\bullet,j_b^*}e'_{i_b^*,n}A^{\ell-1}=\Psi_{i_b^*,j_b^*,h_b}} \Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right) = -\operatorname{trace}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right),$$

and the *third* term (in the earlier display) also equals this quantity by a symmetric calculation. In conclusion, we have shown

$$\begin{aligned}&\sum_{\ell=1}^{\infty}\operatorname{trace}(\Xi'_{i_a^*,j_a^*,h_a,\ell}\Xi_{i_b^*,j_b^*,h_b,\ell})\\ &= \mathbb{1}(j_a^*=j_b^*)\sigma_{j_a^*}^{-2}\sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_b^*,n}A^{h_b-\ell}\Sigma(A')^{h_a-\ell}e_{i_a^*,n} - \operatorname{trace}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a}S^{-1}\right).\end{aligned}$$

The desired result (C.4) now follows from Corollary A.2. □

Proof of Corollary C.1. The result follows straightforwardly from (C.2) if we can show that the maximal non-centrality parameter equals

$$\max_{\alpha(L)\colon \|\alpha(L)\|\le M} \operatorname{aBias}(\hat{\boldsymbol{\delta}})'\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}\operatorname{aBias}(\hat{\boldsymbol{\delta}}) = M^2\left[\lambda_{\max}(\operatorname{aVar}(\hat{\boldsymbol{\beta}})\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1})-1\right].$$

But this follows from applying Proposition C.2 with $R=\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}$, since

$$\begin{aligned}&\lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}[\operatorname{aVar}(\hat{\boldsymbol{\beta}})-\operatorname{aVar}(\hat{\boldsymbol{\delta}})]\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime}\right)\\ &=\lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}\operatorname{aVar}(\hat{\boldsymbol{\beta}})\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime}-I_k\right)\\ &=\lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\beta}})\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}\right)-1. \quad \square\end{aligned}$$

Appendix D Simulation details and further results

We here provide supplementary details for the simulation study reported in Section 5.2.

Implementation details. All inference procedures (correctly) assume homoskedastic shocks. The VAR is estimated with an intercept, and confidence intervals are constructed either using the delta method or the recursive residual bootstrap; following the recommendation of Inoue and Kilian (2020), we report the Efron bootstrap confidence interval. For LPs, we include the shock measure as the only contemporaneous regressor, control for an intercept and the same p lags of all observables as in the VAR, and report the OLS coefficient on the shock. Confidence intervals are constructed either using homoskedastic OLS standard errors or by bootstrapping an auxiliary VAR as in Montiel Olea and Plagborg-Møller (2021), but using a recursive residual bootstrap instead of a wild bootstrap; we follow the latter paper and report the percentile-t bootstrap confidence interval. We use 1,000 bootstrap draws, and the maximal lag length considered for the AIC is 24.

Further results. Figure D.1 shows coverage probabilities and median confidence interval lengths for longer estimation lag lengths $p \in \{15, 18\}$. The results are consistent with the asymptotic theory: the longer the estimation lag length, the less severe VAR undercoverage, with correct coverage ensured through equivalence with LP. Of course, since the true DGP is a VAR(18), the $p = 18$ VAR estimator is more efficient than LP at longer horizons (middle panel). The bottom panel of the figure shows the root MSE for VAR and LP, with lag length selected by AIC. By this measure, VAR outperforms LP, indicating that while the VAR bias is large enough to seriously compromise inference (as shown earlier), it is not so large as to threaten the VAR estimator's status as a useful point estimator.

Appendix E Further proofs

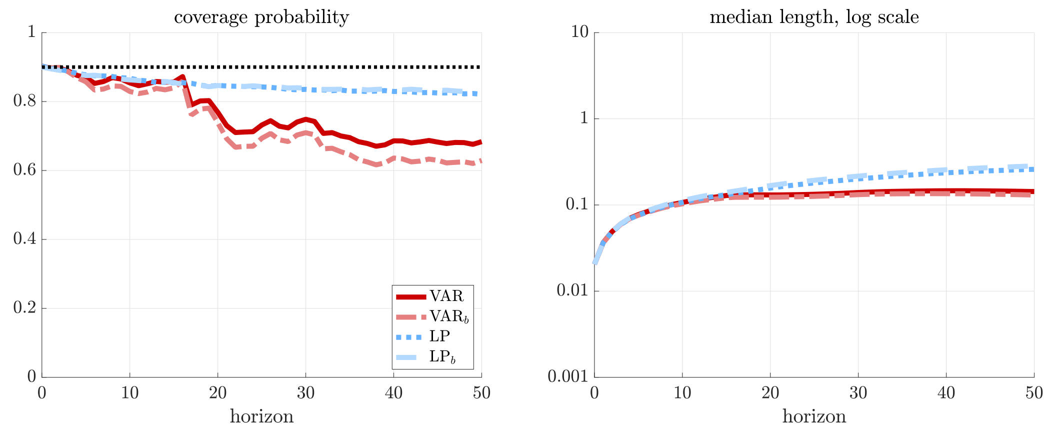
We impose Assumption C.1 and Assumption 2.1(ii)–(v) throughout; as discussed in Supplemental Appendix C.1, none of the proofs below require Assumption 2.1(i). Let $\|B\|$ denote the Frobenius norm of any matrix B. It is well known that this norm is sub-multiplicative: $\|BC\| \leq \|B\| \cdot \|C\|$. Let I_n denote the $n \times n$ identity matrix, $0_{m\times n}$ the $m \times n$ matrix of zeros, and $e_{i,n}$ the n-dimensional unit vector with a 1 in the i-th position. Recall from Assumption 2.1 the definitions $D \equiv \mathrm{Var}(\varepsilon_t) = \mathrm{diag}(\sigma_1^2, \dots, \sigma_m^2)$, $\tilde{y}_t \equiv (I_n - AL)^{-1} H\varepsilon_t = \sum_{s=0}^{\infty} A^s H \varepsilon_{t-s}$, and $S \equiv \mathrm{Var}(\tilde{y}_t)$.

E.1 Main lemmas

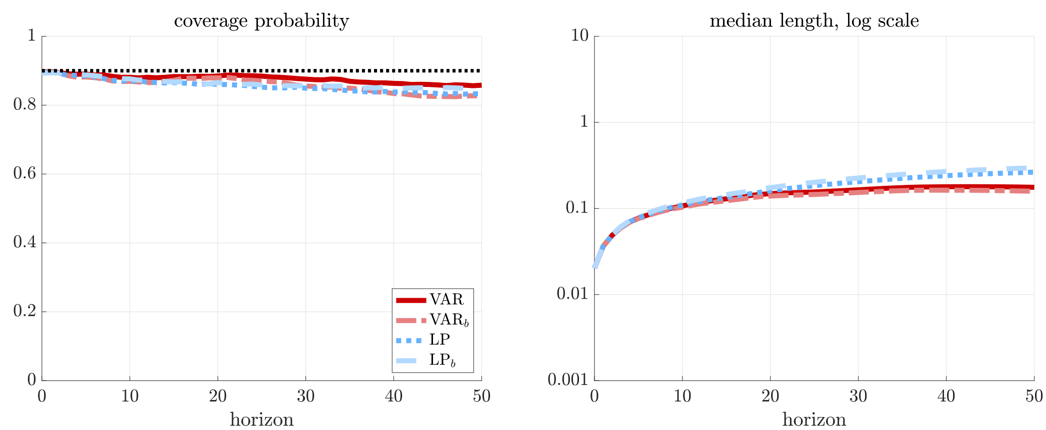
Lemma E.1. *For any $i^* \in \{1, \dots, n\}$ and $j^* \in \{1, \dots, m\}$, we have*

$$y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,i^*,j^*}\underline{y}_{j^*,t} + B'_{h,i^*,j^*} y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t,$$

Coverage and length for lag length $p = 15$



Coverage and length for lag length $p = 18$



RMSE for lag length selected by AIC

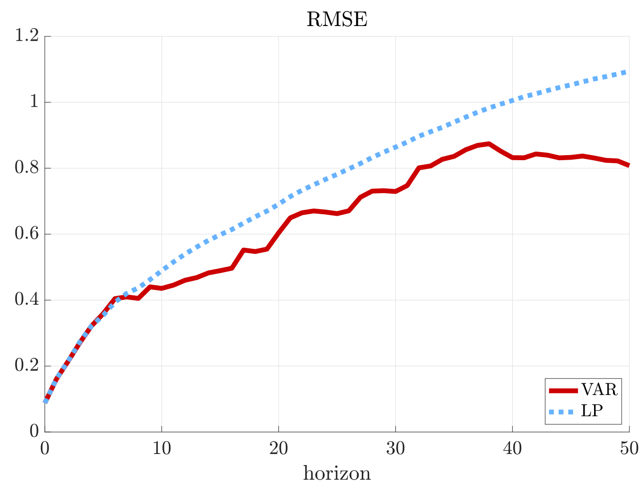


Figure D.1: Top two panels: coverage probability (left) and median length (right) for VAR (red) and LP (blue) 90% confidence intervals computed via the delta method or bootstrap (the latter are indicated with subscript "b" in the legends). Bottom panel: root MSE of estimators. Lag length: $p = 15$ in the top panel, $p = 18$ in the middle panel, and p selected by AIC in the bottom panel.

where

$$\begin{aligned}
\theta_{h,T} &\equiv e'_{i^*,n}(A^h H + T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell}H\alpha_\ell)e_{j^*,m}, \\
\underline{B}'_{h,i^*,j^*} &\equiv e'_{i^*,n}A^h \underline{H}_{j^*}H_{11}^{-1}, \\
B'_{h,i^*,j^*} &\equiv e'_{i^*,n}\left[A^{h+1} - A^h \underline{H}_{j^*}H_{11}^{-1}\underline{I}_{j^*}A\right], \\
\xi_{i^*,h,t} &\equiv e'_{i^*,n}A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h} e'_{i^*,n}A^{h-\ell}H\varepsilon_{t+\ell},
\end{aligned}$$

and $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty}\Theta_{h,\ell}L^\ell$ *is an absolutely summable,* $1\times n$ *two-sided lag polynomial with the* j^**-th element of* $\Theta_{h,0}$ *equal to zero. Moreover,*

$$T^{-1}\sum_{t=1}^{T-h}(\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} = O_p(T^{-1/2}).$$

Proof. Iteration on the model in Equation (2.1) yields

$$y_{t+h} = A^{h+1}y_{t-1} + \sum_{\ell=0}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}). \tag{E.1}$$

As in Section 2.2, let $\underline{y}_{j^*,t} \equiv (y_{1,t},\dots,y_{j^*-1,t})'$ denote the variables ordered before $y_{j^*,t}$ (if any). Analogously, let $\overline{y}_{j^*,t} \equiv (y_{j^*+1,t},\dots y_{n,t})'$ denote the variables ordered after $y_{j^*,t}$.

Using Assumption 2.1(iii), partition

$$H = (\underline{H}_{j^*}, H_{\bullet,j^*}, \overline{H}_{j^*}) = \begin{pmatrix} H_{11} & 0 & 0 \\ H_{21} & H_{22} & 0 \\ H_{31} & H_{32} & H_{33} \end{pmatrix}$$

conformably with the vector $y_t = (\underline{y}'_{j^*,t}, y_{j^*,t}, \overline{y}'_{j^*,t})'$. Let $\underline{I}_{j^*}$ denote the first j^*-1 rows of the $n\times n$ identity matrix. Using the definition of y_t in Equation (2.1),

$$\underline{y}_{j^*,t} = \underline{I}_{j^*}Ay_{t-1} + H_{11}\underline{\varepsilon}_{j^*,t} + T^{-\zeta}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t,$$

where $\underline{\varepsilon}_{j^*,t} = \underline{I}_{j^*}\varepsilon_t$. Using the previous equation to solve for $\underline{\varepsilon}_{j^*,t}$ we get

$$\underline{\varepsilon}_{j^*,t} = H_{11}^{-1}(\underline{y}_{j^*,t} - \underline{I}_{j^*}Ay_{t-1} - T^{-\zeta}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t). \tag{E.2}$$

Expanding the terms in (E.1) we get:

$$
\begin{aligned}
y_{t+h} &= A^{h+1}y_{t-1} + A^h H\varepsilon_t + T^{-\zeta}A^h H\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}) \\
&= A^{h+1}y_{t-1} + \left(A^h \underline{H}_{j^*}\underline{\varepsilon}_{j^*,t} + A^h H_{\bullet,j^*}\varepsilon_{j^*,t} + A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t}\right) \\
&\quad + T^{-\zeta}A^h H\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}) \\
&= A^{h+1}y_{t-1} + A^h\underline{H}_{j^*}H_{11}^{-1}(\underline{y}_{j^*,t} - \underline{I}_{j^*}Ay_{t-1} - T^{-\zeta}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t) + A^h H_{\bullet,j^*}\varepsilon_{j^*,t} + A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t} \\
&\quad + T^{-\zeta}A^h H\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}),
\end{aligned}
$$

where the last equality follows from substituting (E.2). Re-arranging terms we get

$$
\begin{aligned}
y_{i^*,t+h} &= \left(e'_{i^*,n}A^h H_{\bullet,j^*}\right)\varepsilon_{j^*,t} + \underbrace{\left(e'_{i^*,n}A^h\underline{H}_{j^*}H_{11}^{-1}\right)}_{\equiv\underline{B}'_{h,i^*,j^*}}\underline{y}_{j^*,t} + \underbrace{\left(e'_{i^*,n}\left[A^{h+1} - A^h\underline{H}_{j^*}H_{11}^{-1}\underline{I}_{j^*}A\right]\right)}_{\equiv B'_{h,i^*,j^*}}y_{t-1} \\
&\quad + \underbrace{e'_{i^*,n}\left(A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h}A^{h-\ell}H\varepsilon_{t+\ell}\right)}_{=\xi_{i^*,h,t}} + T^{-\zeta}e'_{i^*,n}\left(-A^h\underline{H}_{j^*}H_{11}^{-1}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t + \sum_{\ell=0}^{h}A^{h-\ell}H\alpha(L)\varepsilon_{t+\ell}\right),
\end{aligned}
$$

Using the definition of $\theta_{h,T} \equiv e'_{i^*,n}(A^h H + T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell}H\alpha_\ell)e_{j^*,m}$ and adding and subtracting $e'_{i^*,n}\left(T^{-\zeta}\sum_{\ell=1}^{h}A^{h-\ell}H\alpha_\ell\right)e_{j^*,m}\varepsilon_{j^*,t}$ in the display above, we obtain a representation of the form

$$
y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,i^*,j^*}\underline{y}_{j^*,t} + B'_{h,i^*,j^*}y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\tilde{u}_t, \tag{E.3}
$$

where

$$
\tilde{u}_t \equiv e'_{i^*,n}\left(-A^h\underline{H}_{j^*}\underline{I}_{j^*}\alpha(L)\varepsilon_t + \sum_{\ell=0}^{h}A^{h-\ell}H\alpha(L)\varepsilon_{t+\ell} - \left(\sum_{\ell=1}^{h}A^{h-\ell}H\alpha_\ell e_{j^*,m}e'_{j^*,m}\right)\varepsilon_t\right). \tag{E.4}
$$

Algebra shows that $\tilde{u}_t$ can be written as a two-sided lag polynomial, $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty}\Theta_{h,\ell}L^\ell$, with coefficients of dimension $1\times n$ given by the following formulae:

$$
\Theta_{h,\ell} = \begin{cases}
-e'_{i^*,n}A^h\underline{H}_{j^*}\underline{I}_{j^*}\alpha_\ell + \sum_{s=0}^{h}e'_{i^*,n}A^{h-s}H\alpha_{\ell+s} & \text{for } \ell \geq 1, \\
\sum_{s=1}^{h}e'_{i^*,n}A^{h-s}H\alpha_s - \sum_{s=1}^{h}e'_{i^*,n}A^{h-s}H\alpha_s e_{j^*,m}e'_{j^*,m} & \text{for } \ell = 0, \\
\sum_{s=1}^{h+\ell}e'_{i^*,n}A^{h-s+\ell}H\alpha_s & \text{for } \ell \in \{-(h-1),\dots,-1\}, \\
0_{1\times n} & \text{for } \ell \leq -h.
\end{cases}
$$

In particular, $\Theta_{h,0,j^*} \equiv \Theta_{h,0} e_{j^*,m} = 0$.

We next show that $\Theta_h(L)$ is absolutely summable, that is $\sum_{\ell=-\infty}^{\infty} \|\Theta_{h,l}\| < \infty$. To do this, it suffices to show that $\sum_{\ell=1}^{\infty} \|\Theta_{h,l}\| < \infty$, since all the coefficients with index $\ell \leq -h$ are 0. Note that, by definition,

$$\sum_{\ell=1}^{\infty} \|\Theta_{h,\ell}\| \leq \|A^h\| \|\underline{H}_{j^*} \underline{I}_{j^*}\| \sum_{\ell=1}^{\infty} \|\alpha_\ell\| + \|H\| \sum_{\ell=1}^{\infty} \sum_{s=0}^{h} \|A^{h-s}\| \|\alpha_{\ell+s}\|.$$

Let $\lambda \in [0,1)$ and $C > 0$ be chosen such that $\|A^\ell\| \leq C\lambda^\ell$ for all $\ell \geq 0$ (such constants exists by Assumption 2.1(ii)). Then

$$\sum_{\ell=1}^{\infty} \sum_{s=0}^{h} \|A^{h-s}\| \|\alpha_{\ell+s}\| \leq C \sum_{\ell=1}^{\infty} \sum_{s=1}^{h} \lambda^{h-s} \|\alpha_{\ell+s}\| \leq C \sum_{\ell=1}^{\infty} \sum_{s=1}^{h} \|\alpha_{\ell+s}\| \leq Ch \sum_{\ell=1}^{\infty} \|\alpha_\ell\| < \infty,$$

where the last inequality holds because the coefficients of $\alpha(L)$ are summable. We thus conclude that

$$y_{i^*,t+h} = \theta_{h,T} \varepsilon_{j^*,t} + \underline{B}'_{h,y} \underline{y}_{j^*,t} + B'_{h,y} y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta} \Theta_h(L) \varepsilon_t,$$

where $\Theta_h(L)$ is a two-sided lag-polynomial with summable coefficients.

Finally, we show that

$$T^{-1} \sum_{t=1}^{T-h} (\Theta_h(L)\varepsilon_t) \varepsilon_{j^*,t} = O_p(T^{-1/2}). \tag{E.5}$$

We can decompose $(\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t}$ into finitely many terms of the form $(\psi_j(L)\varepsilon_{j,t})\varepsilon_{j^*,t}$ for some two-sided, absolutely summable lag polynomial $\psi_j(L)$ and integer j. Since $\Theta_{h,0,j^*} = 0$, each of these terms has mean zero by shock orthogonality. By Lemma E.11, each term also has absolutely summable autocovariances. Hence, Brockwell and Davis (1991, Thm. 7.1.1) and Chebyshev's inequality imply that the sample average of each term is $O_p(T^{-1/2})$. $\square$

Lemma E.2.

$$\hat{A} - A = T^{-\zeta} H \sum_{\ell=1}^{\infty} \alpha_\ell D H' (A')^{\ell-1} S^{-1} + T^{-1} \sum_{t=1}^{T} H \varepsilon_t \tilde{y}'_{t-1} S^{-1} + o_p(T^{-\zeta}).$$

In particular, $\hat{A} - A = O_p(T^{-\zeta} + T^{-1/2})$.

Proof. Since,

$$\hat{A} - A = \left(T^{-1}\sum_{t=1}^{T-h} u_t y_{t-1}'\right)\left(T^{-1}\sum_{t=1}^{T-h} y_{t-1}y_{t-1}'\right)^{-1},$$

the result follows from Lemmas E.7 and E.8. $\square$

Lemma E.3.

$$\hat{\nu} - H_{\bullet,j^*} = \frac{1}{\sigma_{j^*}^2} T^{-1}\sum_{t=1}^{T} \xi_{0,t}\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2}).$$

Proof. By Lemma E.5, $\hat{\nu} = (0_{1\times(j^*-1)}, 1, \hat{\bar{\nu}}')$, where the j-th element of $\hat{\bar{\nu}}$ equals the on-impact local projection of $y_{i^*+j,t}$ on $y_{j^*,t}$, controlling for $\underline{y}_{j^*,t}$ and y_{t-1}. The statement of the lemma is therefore a direct consequence of Proposition 3.1 and the fact that (by definition) $\xi_{0,i,t} = 0$ for $i \leq j^*$. $\square$

Lemma E.4. *Fix $h \geq 0$. Consider the regression of $y_{j^*,t}$ on $q_{j^*,t} \equiv (\underline{y}_{j^*,t}', y_{t-1}')'$, using the observations $t = 1, 2, \ldots, T-h$:*

$$y_{j^*,t} = \hat{\vartheta}_h' q_{j^*,t} + \hat{x}_{h,t}.$$

Note that the residuals $\hat{x}_{h,t}$ are consistent with the earlier definition in the proof of Proposition 3.1. Let $\underline{\lambda}_{j^*}'$ be the row vector containing the first j^*-1 elements of the last row of $-\tilde{H}^{-1}$ (where $\tilde{H}$ is defined in Assumption 2.1(iii)). Let $\lambda_{j^*}' \equiv (-\underline{\lambda}_{j^*}', 1, 0_{1\times(n-j^*)})$ and $\vartheta \equiv (\underline{\lambda}_{j^*}', (\lambda_{j^*}'A))'$. Then:

i) $\hat{\vartheta}_h - \vartheta = O_p(T^{-\zeta} + T^{-1/2})$.

ii) For $j \geq j^*$, $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{j,t} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

iii) For $\ell \geq 1$, $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{t+\ell} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

iv) $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\hat{x}_{h,t} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

v) $T^{-1}\sum_{t=1}^{T-h}\hat{x}_{h,t}^2 \xrightarrow{p} \sigma_{j^*}^2$.

vi) For any absolutely summable two-sided lag polynomial $B(L)$, $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})B(L)\varepsilon_t = O_p(T^{-\zeta} + T^{-1/2})$.

Proof. By Equation (2.1), the outcome variables in the model satisfy

$$y_t = Ay_{t-1} + H[I_m + T^{-\zeta}\alpha(L)]\varepsilon_t,\ t = 1, 2, \ldots, T.$$

By Assumption 2.1(iii), the first j^* rows of the matrix H above are of the form $(\tilde{H}, 0_{j^*\times(j^*-m)})$, where m is the number of shocks and $\tilde{H}$ is a $j^* \times j^*$ lower triangular matrix with 1's on the diagonal.

We can premultiply the first j^* equations of (2.1) by $\tilde{H}^{-1}$ to obtain:

$$[\tilde{H}^{-1}, 0_{j^*\times(n-j^*)}]y_t = [\tilde{H}^{-1}, 0_{j^*\times(n-j^*)}]Ay_{t-1} + [I_{j^*}, 0_{j^*\times(m-j^*)}][I_m + T^{-\zeta}\alpha(L)]\varepsilon_t.$$

By definition, $-\underline{\lambda}'_{j^*}$ is the row vector containing the first $j^* - 1$ elements of the last row of $\tilde{H}^{-1}$ and $\lambda'_{j^*} \equiv (-\underline{\lambda}'_{j^*}, 1, 0_{1\times(n-j^*)})$. Thus, we can re-write the j^*-th equation above as

$$[-\underline{\lambda}'_{j^*}, 1, 0_{j^*\times(n-j^*)}]y_t = \lambda'_{j^*}Ay_{t-1} + \varepsilon_{j^*,t} + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t,$$

where $\alpha_{j^*}(L)$ is the j^*-th row of $\alpha(L)$. Re-arranging terms we get

$$y_{j^*,t} = \vartheta' q_{j^*,t} + \varepsilon_{j^*,t} + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t,$$

where $\vartheta \equiv (\underline{\lambda}'_{j^*}, (\lambda'_{j^*}A))'$ and $q_{j^*,t} \equiv (\underline{y}'_{j^*,t}, y'_{t-1})'$. In a slight abuse of notation, and for notational simplicity, we henceforth replace $q_{j^*,t}$ by q_t.

Statement (i) follows from standard OLS algebra if we can show that a) $T^{-1}\sum_{t=1}^{T-h} q_t\varepsilon_{j^*,t} = O_p(T^{-\zeta} + T^{-1/2})$, b) $(T^{-1}\sum_{t=1}^{T-h} q_t q_t')^{-1} = O_p(1)$, and c) $T^{-\zeta-1}\sum_{t=1}^{T-h} q_t(\alpha_{j^*}(L)\varepsilon_t) = O_p(T^{-\zeta})$. Lemma E.9 establishes these results.

The proofs of statements (ii) and (iii) are similar, so we focus on the latter. By definition of $\hat{x}_{h,t}$, we have $\hat{x}_{h,t} - \varepsilon_{j^*,t} = (\vartheta - \hat{\vartheta}_h)'q_t + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t$. Let $\tilde{q}_t \equiv (\underline{\tilde{y}}'_{j^*,t}, \tilde{y}'_{t-1})'$ and $\Delta_t \equiv q_t - \tilde{q}_t$. Then

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{t+\ell} = (\vartheta - \hat{\vartheta}_h)'\left(\frac{1}{T}\sum_{t=1}^{T-h}\Delta_t\varepsilon_{t+\ell}\right) \tag{E.6}$$

$$+ (\vartheta - \hat{\vartheta}_h)'\left(\frac{1}{T}\sum_{t=1}^{T-h}\tilde{q}_t\varepsilon_{t+\ell}\right) \tag{E.7}$$

$$+ \frac{1}{T^\zeta}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\alpha_{j^*}(L)\varepsilon_t\right)\varepsilon_{t+\ell}\right). \tag{E.8}$$

By (i), $(\vartheta - \hat{\vartheta}_h) = O_p(T^{-\zeta} + T^{-1/2})$. Lemma E.6, Assumption C.1, and Cauchy-Schwarz imply that the sample average in parenthesis in (E.6) is $O_p\left(T^{-\zeta}\right)$. The two sample averages in parentheses in lines (E.7)–(E.8) have mean zero, since the shocks are white noise and mutually orthogonal, so Lemma E.11 and Brockwell and Davis (1991, Thm. 7.1.1) imply that they are each $O_p(T^{-1/2})$. It follows that

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{t+\ell}=O_p(T^{-2\zeta})+o_p(T^{-1/2}).$$

For statement (iv), note that

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\hat{x}_{h,t}=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2+T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{j^*,t}.$$

Lemma E.10 shows that $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2=O_p(T^{-2\zeta})+o_p(T^{-1/2})$. This result, combined with (ii), implies that statement (iv) holds.

For statement (v), note that

$$\begin{aligned}T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t})^2&=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t}+\varepsilon_{j^*,t})^2\\&=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2-2T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{j^*,t}+T^{-1}\sum_{t=1}^{T-h}\varepsilon_{j^*,t}^2.\end{aligned}$$

Lemma E.10 and statement (ii) imply that the first two terms converge in probability to zero. Since $T^{-1}\sum_{t=1}^{T-h}\varepsilon_{j^*,t}^2\xrightarrow{p}\sigma_{j^*}^2$ by Lemma E.11 and Brockwell and Davis (1991, Thm. 7.1.1), statement (v) holds.

Finally, statement (vi) obtains by decomposing

$$\begin{aligned}T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t(\hat{x}_{h,t}-\varepsilon_{j^*,t})&=T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t q_t'(\vartheta-\hat{\vartheta}_h)+T^{-\zeta}T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t[\alpha_{j^*}(L)\varepsilon_t]'\\&=O_p(1)\times O_p(T^{-\zeta}+T^{-1/2})+T^{-\zeta}\times O_p(1),\end{aligned}$$

where the last line follows from statement (i), Lemma E.6, and Lemma E.11. □

E.2 Auxiliary numerical lemma

Lemma E.5. *Define $\bar{y}_{i,t}\equiv(y_{i+1,t},y_{i+2,t},\ldots,y_{nt})'$ to be the (possibly empty) vector of variables that are ordered after $y_{i,t}$ in y_t. Partition*

$$\hat{\Sigma}=\begin{pmatrix}\hat{\Sigma}_{11}&\hat{\Sigma}_{12}&\hat{\Sigma}_{13}\\\hat{\Sigma}_{21}&\hat{\Sigma}_{22}&\hat{\Sigma}_{23}\\\hat{\Sigma}_{31}&\hat{\Sigma}_{32}&\hat{\Sigma}_{33}\end{pmatrix},\quad\hat{C}=\begin{pmatrix}\hat{C}_{11}&0&0\\\hat{C}_{21}&\hat{C}_{22}&0\\\hat{C}_{31}&\hat{C}_{32}&\hat{C}_{33}\end{pmatrix},$$

conformably with $y_t = (\underline{y}'_{j^*,t}, y_{j^*,t}, \overline{y}'_{j^*,t})'$, *where* $\hat{\Sigma} = \hat{C}\hat{C}'$ *(in particular,* $\hat{C}_{22} = \hat{C}_{j^*,j^*}$*). Then*

$$(\hat{\Sigma}_{31}, \hat{\Sigma}_{32}) \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} = \hat{C}_{22}^{-1}\hat{C}_{32}. \tag{E.9}$$

Note that the lemma implies $\hat{\beta}_0 = \hat{\delta}_0$: If $i^* < j^*$ or $i^* = j^*$, then both estimators equal 0 or 1 (by definition), respectively; if $i^* > j^*$, then $\hat{\beta}_0$ is defined as the $i^* - j^*$ element of the left-hand side of (E.9) (by Frisch-Waugh), while $\hat{\delta}_0$ is defined as the $i^* - j^*$ element of the right-hand side of (E.9).

Proof. From the relationship $\hat{\Sigma} = \hat{C}\hat{C}'$, we get

$$\begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \\ \hat{\Sigma}_{31} & \hat{\Sigma}_{32} \end{pmatrix} = \begin{pmatrix} \hat{C}_{11}\hat{C}'_{11} & \hat{C}_{11}\hat{C}'_{21} \\ \hat{C}_{21}\hat{C}'_{11} & \hat{C}_{21}\hat{C}'_{21} + \hat{C}_{22}^2 \\ \hat{C}_{31}\hat{C}'_{11} & \hat{C}_{31}\hat{C}'_{21} + \hat{C}_{32}\hat{C}_{22} \end{pmatrix}.$$

The partitioned inverse formula implies

$$\begin{aligned} \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} &= \frac{1}{\hat{C}_{21}\hat{C}'_{21} + \hat{C}_{22}^2 - \hat{C}_{21}\hat{C}'_{11}(\hat{C}_{11}\hat{C}'_{11})^{-1}\hat{C}_{11}\hat{C}'_{21}} \begin{pmatrix} -(\hat{C}_{11}\hat{C}'_{11})^{-1}\hat{C}_{11}\hat{C}'_{21} \\ 1 \end{pmatrix} \\ &= \frac{1}{\hat{C}_{22}^2} \begin{pmatrix} -\hat{C}_{11}^{-1\prime}\hat{C}'_{21} \\ 1 \end{pmatrix}, \end{aligned}$$

so

$$(\hat{\Sigma}_{31}, \hat{\Sigma}_{32}) \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} = \frac{1}{\hat{C}_{22}^2}\left(-\hat{C}_{31}\hat{C}'_{11}\hat{C}_{11}^{-1\prime}\hat{C}'_{21} + \hat{C}_{31}\hat{C}'_{21} + \hat{C}_{32}\hat{C}_{22}\right) = \frac{1}{\hat{C}_{22}}\hat{C}_{32}. \quad \square$$

E.3 Auxiliary asymptotic lemmas

Lemma E.6. $T^{-1}\sum_{t=1}^T \|y_t - \tilde{y}_t\|^2 = O_p(T^{-2\zeta})$ *and* $T^{-1}\sum_{t=1}^T u_t(y_{t-1} - \tilde{y}_{t-1})' = O_p(T^{-2\zeta} + T^{-\zeta-1/2})$*, where* $u_t \equiv y_t - Ay_{t-1}$*.*

Proof. Using Equation (2.1) and the definition $\tilde{y}_t \equiv (I_n - AL)^{-1}H\varepsilon_t$, we have

$$y_t - \tilde{y}_t = T^{-\zeta}(I_n - AL)^{-1}H\alpha(L)\varepsilon_t.$$

The lag polynomial $(I_n - AL)^{-1}H\alpha(L)$ is absolutely summable by virtue of being a product of absolutely summable polynomials, see Assumption 2.1(ii) and (v).

Brockwell and Davis (1991, Prop. 3.1.1) implies $E[T^{2\zeta}\|y_t - \tilde{y}_t\|^2] < \infty$. The first statement of the lemma then follows from Markov's inequality.

In order to establish the second part of the lemma, note that

$$\frac{1}{T}\sum_{t=1}^{T} u_t\left(y_{t-1} - \tilde{y}_{t-1}\right)' = H\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(y_{t-1} - \tilde{y}_{t-1}\right)'\right) + T^{-\zeta}H\left(\frac{1}{T}\sum_{t=1}^{T}\alpha(L)\varepsilon_t\left(y_{t-1} - \tilde{y}_{t-1}\right)'\right).$$

The product process $T^{\zeta}\varepsilon_t \otimes (y_{t-1} - \tilde{y}_{t-1})$ is a martingale difference sequence with finite variance by Lemma E.11 (note that this is a standard stochastic process and not a triangular array). Hence, the first sample average in parenthesis on the right-hand side above is $O_p(T^{-\zeta-1/2})$ by Chebyshev's inequality. The second sample average in parenthesis on the right-hand side above is $O_p(T^{-\zeta})$ by Lemma E.11. Hence, the entire right-hand side in the above display is $O_p(T^{-\zeta-1/2} + T^{-2\zeta})$, as claimed. $\square$

Lemma E.7.

$$T^{-1}\sum_{t=1}^{T} u_t y_{t-1}' = T^{-\zeta}H\sum_{\ell=1}^{\infty}\alpha_\ell D H'(A')^{\ell-1} + T^{-1}\sum_{t=1}^{T} H\varepsilon_t\tilde{y}_{t-1}' + o_p(T^{-\zeta}).$$

Proof.

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} u_t y_{t-1}' &= T^{-1}\sum_{t=1}^{T} u_t\tilde{y}_{t-1}' + \underbrace{T^{-1}\sum_{t=1}^{T} u_t(y_{t-1} - \tilde{y}_{t-1})'}_{=o_p(T^{-\zeta}) \text{ by Lemma E.6}} \\
&= T^{-1}\sum_{t=1}^{T} H\varepsilon_t\tilde{y}_{t-1}' + T^{-\zeta-1}\sum_{t=1}^{T} H\alpha(L)\varepsilon_t\tilde{y}_{t-1}' + o_p(T^{-\zeta}) \\
&= T^{-1}\sum_{t=1}^{T} H\varepsilon_t\tilde{y}_{t-1}' + T^{-\zeta}H\left(E[\alpha(L)\varepsilon_t\tilde{y}_{t-1}'] + o_p(1)\right) + o_p(T^{-\zeta}).
\end{aligned}$$

The last line invokes a law of large numbers, which applies because $\tilde{y}_t$ is an absolutely summable linear filter of the shocks, so we can apply Lemma E.11 in conjunction with Brockwell and Davis (1991, Thm. 7.1.1) and Chebyshev's inequality. Finally, note that

$$E[\alpha(L)\varepsilon_t\tilde{y}_{t-1}'] = \sum_{\ell=1}^{\infty}\sum_{s=0}^{\infty}\alpha_\ell E[\varepsilon_{t-\ell}\varepsilon_{t-s-1}']H'(A')^s = \sum_{\ell=1}^{\infty}\alpha_\ell DH'(A')^{\ell-1}. \qquad \square$$

Lemma E.8. $T^{-1}\sum_{t=1}^{T} y_{t-1}y_{t-1}' \xrightarrow{p} S$.

Proof. By Lemma E.6 and Cauchy-Schwarz, $T^{-1}\sum_{t=1}^{T} y_{t-1}y_{t-1}' = T^{-1}\sum_{t=1}^{T} \tilde{y}_{t-1}\tilde{y}_{t-1}' + o_p(1)$. Lemma E.11, Brockwell and Davis (1991, Thm. 7.1.1), and Chebyshev's inequality imply that the law of large numbers holds for $\{\tilde{y}_{t-1}\tilde{y}_{t-1}'\}$. $\square$

Lemma E.9. *Omitting the subscript j^* in a slight abuse of notation, let $q_t \equiv (\underline{y}_{j^*,t}', y_{t-1}')'$. Then*

i) $T^{-1}\sum_{t=1}^{T-h} q_t\varepsilon_{j^*,t} = O_p(T^{-\zeta} + T^{-1/2})$,

ii) $(T^{-1}\sum_{t=1}^{T-h} q_t q_t')^{-1} = O_p(1)$,

iii) $T^{-1}\sum_{t=1}^{T-h} q_t(\alpha_{j^*}(L)\varepsilon_t) = O_p(1)$,

where $\alpha_{j^*}(L)$ is the j^*-th row of $\alpha(L)$.

Proof. Let $\tilde{q}_t \equiv (\tilde{\underline{y}}_{j^*,t}', \tilde{y}_{t-1}')'$ and $\Delta_t \equiv q_t - \tilde{q}_t$. Note that

$$T^{-1}\sum_{t=1}^{T-h} q_t\varepsilon_{j^*,t} = T^{-1}\sum_{t=1}^{T-h} \Delta_t\varepsilon_{j^*,t} + T^{-1}\sum_{t=1}^{T-h} \tilde{q}_t\varepsilon_{j^*,t}. \tag{E.10}$$

Cauchy-Schwarz implies

$$\left\| T^{-1}\sum_{t=1}^{T-h} \Delta_t\varepsilon_{j^*,t} \right\| \leq \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\Delta_t\|^2\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon_{j^*,t}^2\right)^{1/2} = O_p(T^{-\zeta}) \times O_p(1),$$

using Lemma E.6 and Assumption C.1. The summands in the last sample average in (E.10) have mean zero due to shock orthogonality, so this sample average is $O_p\left(T^{-1/2}\right)$ by Lemma E.11 and Brockwell and Davis (1991, Thm. 7.1.1). This establishes part (i) of the lemma.

For part (ii) of the lemma, note that

$$\frac{1}{T}\sum_{t=1}^{T-h} q_t q_t' = \frac{1}{T}\sum_{t=1}^{T-h} \Delta_t\Delta_t' + \frac{1}{T}\sum_{t=1}^{T-h} \tilde{q}_t\Delta_t' + \frac{1}{T}\sum_{t=1}^{T-h} \Delta_t\tilde{q}_t' + \frac{1}{T}\sum_{t=1}^{T-h} \tilde{q}_t\tilde{q}_t'. \tag{E.11}$$

Lemma E.6 implies that the first term is $O_p\left(T^{-2\zeta}\right)$. Cauchy-Schwarz, along with Lemmas E.6 and E.8, imply that the second and third terms are $O_p(T^{-\zeta})$. The last term converges in probability to $\mathrm{Var}(\tilde{q}_t)$, as in the proof of Lemma E.8. This matrix is non-singular, since $\tilde{q}_t = (\underline{\tilde{y}}'_{j^*,t}, \tilde{y}'_{t-1})'$, where $\mathrm{Var}(\tilde{y}_{t-1}) = S$ is non-singular by Assumption 2.1(iv), and Assumption 2.1(iii) implies that $\underline{\tilde{y}}_{j^*,t}$ equals a linear transformation of $\tilde{y}_{t-1}$ plus a non-singular orthogonal noise term.

Part (iii) follows from Cauchy-Schwarz, Lemma E.8, and Assumption 2.1(v). □

Lemma E.10. *Use the same notation as Lemma E.9, and let*

$$\hat{x}_{h,t} \equiv (\vartheta - \hat{\vartheta}_h)' q_t + \varepsilon_{j^*,t} + T^{-\zeta} \alpha_{j^*}(L) \varepsilon_t.$$

Then

$$T^{-1} \sum_{t=1}^{T-h} (\hat{x}_{h,t} - \varepsilon_{j^*,t})^2 = O_p(T^{-2\zeta}) + o_p(T^{-1/2}). \tag{E.12}$$

Proof. It suffices by the c_r-inequality to show that

a) $T^{-1} \sum_{t=1}^{T-h} \left((\vartheta - \hat{\vartheta}_h)' q_t \right)^2 = O_p(T^{-2\zeta}) + o_p\left(T^{-1/2}\right)$,

b) $T^{-1} \sum_{t=1}^{T-h} \left(\alpha_{j^*}(L) \varepsilon_t\right)^2 = O_p\left(1\right)$.

To establish (a), note that Cauchy-Schwarz implies

$$\frac{1}{T} \sum_{t=1}^{T-h} \left((\vartheta - \hat{\vartheta}_h)' q_t \right)^2 \leq \left\| \vartheta - \hat{\vartheta}_h \right\|^2 \left(\frac{1}{T} \sum_{t=1}^{T-h} \|q_t\|^2 \right) = O_p\left(\left(T^{-\zeta} + T^{-1/2} \right)^2 \right) \times O_p(1),$$

by Lemmas E.8 and E.9. Hence, the right-hand side is $O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

Statement (b) follows from Assumption 2.1(v) and Markov's inequality. □

Lemma E.11. *Let $\psi(L)$ and $\varphi(L)$ be two absolutely summable, univariate, two-sided lag polynomials. Then for any $j, k \in \{1, \ldots, m\}$, the product process $z_t \equiv [\psi(L)\varepsilon_{j,t}] \times [\varphi(L)\varepsilon_{k,t}]$ has absolutely summable autocovariance function.*

Proof. Bound $\sum_{\ell=-\infty}^{\infty} |\operatorname{Cov}(z_t, z_{t+\ell})|$ by

$$\begin{aligned} &\sum_{\ell=-\infty}^{\infty} \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} \sum_{\tau_4=-\infty}^{\infty} |\psi_{\tau_1}||\varphi_{\tau_2}||\psi_{\tau_3}||\varphi_{\tau_4}| |\operatorname{Cov}(\varepsilon_{j,t+\tau_1}\varepsilon_{k,t+\tau_2}, \varepsilon_{j,t+\tau_3+\ell}\varepsilon_{k,t+\tau_4+\ell})| \\ &\leq \left(\sum_{\tau=-\infty}^{\infty} |\psi_\tau| \right)^2 \left(\sum_{\tau=-\infty}^{\infty} |\varphi_\tau| \right)^2 \left(\sup_{\tau_1,\ldots,\tau_4} \sum_{\ell=-\infty}^{\infty} |\operatorname{Cov}(\varepsilon_{j,t+\tau_1}\varepsilon_{k,t+\tau_2}, \varepsilon_{j,t+\tau_3+\ell}\varepsilon_{k,t+\tau_4+\ell})| \right). \end{aligned} \tag{E.13}$$

To finish the proof, we show that the supremum above is finite. By stationarity, the sum inside the supremum can be written as

$$\sum_{r=-\infty}^{\infty} |\operatorname{Cov}(\varepsilon_{j,0}\varepsilon_{k,s}, \varepsilon_{j,r}\varepsilon_{k,r+\tau})|, \tag{E.14}$$

where we have substituted $s = \tau_2 - \tau_1$, $\tau = \tau_4 - \tau_3$, and $r = \ell + \tau_3 - \tau_1$. Now fix s and τ. Let $N \colon \mathbb{Z} \to \{1, 2, 3, 4\}$ denote the function that assigns each integer r to the number of times the maximum value of the tuple $(0, s, r, r + \tau)$ appears in the tuple. We group the terms indexed by r in the sum (E.14) according to their value of $N(r)$. First, since $\{\varepsilon_t\}$ is a martingale difference sequence, all terms r with $N(r) = 1$ yield a covariance of 0 and so do not contribute to the sum. Second, a simple enumeration of cases shows that there is at most 1 value of r with $N(r) = 3$ and at most one with $N(r) = 4$. Finally, consider terms r with $N(r) = 2$. If $s \neq 0$ and $\tau \neq 0$, then there are at most 4 such terms (since this requires $r \in \{0, s, -\tau, s - \tau\}$). If $s = 0$ and/or $\tau = 0$, terms with $N(r) = 2$ must be of the form $|\operatorname{Cov}(\varepsilon_{j,0}\varepsilon_{k,0}, \varepsilon_{j,r}\varepsilon_{k,r+\tau})|$ (with $r, r + \tau < 0$) and/or $|\operatorname{Cov}(\varepsilon_{j,0}\varepsilon_{k,0}, \varepsilon_{j,-r}\varepsilon_{k,s-r})|$ (with $r, r - s > 0$). The preceding arguments show that the sum (E.14) is bounded by

$$6E[\|\varepsilon_t\|^4] + 2\sum_{r=1}^{\infty}\sum_{b=1}^{\infty} \|\operatorname{Cov}(\varepsilon_0 \otimes \varepsilon_0, \varepsilon_{-r} \otimes \varepsilon_{-b})\|,$$

which is finite by Assumption C.1 and does not depend on s or τ. Thus, the supremum in (E.13) is finite. □

References

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*, Springer Series in Statistics. Springer, 2nd edn.

Hallin, M., J.-F. Ingenbleek, and M. L. Puri (1989): "Asymptotically most powerful rank tests for multivariate randomness against serial dependence," *Journal of Multivariate Analysis*, 30(1), 34–71.

Hallin, M., and M. L. Puri (1988): "Optimal Rank-Based Procedures for Time Series Analysis: Testing an ARMA Model Against Other ARMA Models," *The Annals of Statistics*, 16(1), 402–432.

Hausman, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.

Inoue, A., and L. Kilian (2020): "The uniform validity of impulse response inference in autoregressions," *Journal of Econometrics*, 215(2), 450–472.

Kuersteiner, G. M. (2001): "Optimal instrumental variables estimation for ARMA models," *Journal of Econometrics*, 104(2), 359–405.

Leeb, H., and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21(01), 21–59.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer.

Montiel Olea, J. L., and M. Plagborg-Møller (2021): "Local Projection Inference Is Simpler and More Robust Than You Think," *Econometrica*, 89(4), 1789–1823.

Plagborg-Møller, M., and C. K. Wolf (2021): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89(2), 955–980.