Online Appendix for: Local Projections and VARs Estimate the Same Impulse Responses

Mikkel Plagborg-Møller   Christian K. Wolf

October 6, 2020

This online appendix contains supplemental material for the article "Local Projections and VARs Estimate the Same Impulse Responses".

Any references to equations or sections that are not preceded by "B." refer to the main article.

Contents

B.1 Asymptotic equivalence of LP and VAR impulse response estimators **2**
B.1.1 Proof of Proposition B.1 3

B.2 General system identification **9**

B.3 Narrative identification **12**

B.4 Estimands in non-linear models **13**
B.4.1 Technical details 14

References **15**

B.1 Asymptotic equivalence of LP and VAR impulse response estimators

Here we prove that local projections and recursively identified VARs estimate nearly the same impulse response functions *in sample*, provided the lag lengths used in the specifications are large enough. Assume we observe the data $w_1, w_2, \dots, w_T$ (recall the notation in Section 2.1). For all lag lengths $p \le T$, define the following:

- Let $\hat{x}_t(p)$ be the residual from a regression of x_t on an intercept, r_t, and $w_{t-1}, \dots, w_{t-p}$.
- Let $\hat{\beta}_h(p)$ denote the OLS estimator of the local projection parameter β_h in the sample version of regression equation (1), where we include p lags of w_t on the right-hand side instead of the infeasible infinite distibuted lag. By the Frisch-Waugh theorem,

$$\hat{\beta}_h(p) = \frac{\sum_{t=p+1}^{T-h} y_{t+h}\hat{x}_t(p)}{\sum_{t=p+1}^{T-h} \hat{x}_t(p)^2}.$$

- Let $\hat{\theta}_h(p)$ denote the horizon-h impulse response of y_t to an innovation in x_t in a Cholesky-identified VAR(p) model (with intercept) estimated by least squares on the data points $t = p+1, p+2, \dots, T$.

In detail, the VAR estimator $\hat{\theta}_h(p)$ is defined as follows. Let $\hat{A}_\ell(p)$ denote the usual least-squares VAR(p) coefficient matrix estimator at lag ℓ, and let $\hat{c}(p)$ denote the corresponding intercept vector estimator. Let $\hat{u}_t(p)$ denote the residual vector. Define the innovation covariance matrix estimator $\hat{\Sigma}(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{u}_t(p)\hat{u}_t(p)'$ and let $\hat{\Sigma}(p) = \hat{B}(p)\hat{B}(p)'$ denote its lower triangular Cholesky decomposition. Define the reduced-form impulse response matrices by $\hat{C}_0(p) = I_{n_w}$ and $\hat{C}_m(p) = \sum_{\ell=1}^{m} \hat{A}_\ell(p)\hat{C}_{m-\ell}(p)$ for $m = 1, \dots, h$. Then $\hat{\theta}_h(p)$ is given by the (n_r+2, n_r+1) element of $\hat{C}_h(p)\hat{B}(p)$.

Note that the VAR(p) residuals

$$\hat{u}_t(p) \equiv w_t - \hat{c}(p) - \sum_{\ell=1}^{p} \hat{A}_\ell(p) w_{t-\ell}, \quad t = p+1, p+2, \dots, T,$$

satisfy

$$\sum_{t=p+1}^{T} \hat{u}_t(p) = 0_{n_w \times 1}, \quad \sum_{t=p+1}^{T} \hat{u}_t(p) w_{t-\ell} = 0_{n_w \times n_w}, \;\; \ell = 1, 2, \dots, p. \tag{B.1}$$

We adopt the convention that $\hat{u}_t(p) \equiv 0$ whenever $t \le p$.

We are now ready to state the near-equivalence result for LP and VAR impulse response estimators. Let $\|\cdot\|$ denote the Frobenius norm.

Proposition B.1. *In the following, the lag length $p = p(T)$ used for estimation is implicitly a function of T. Assume the following:*

i) $\{w_t\}$ is covariance stationary and has a VAR(∞) representation (3), where $\sum_{\ell=1}^{\infty} \|A_\ell\| < \infty$, and the Wold innovations u_t have finite and positive definite covariance matrix Σ. (We do not assume that the innovations are necessarily Gaussian.)

ii) $\|\hat{c}(p) - c\| = o_p(1)$, $\|\hat{A}(p) - A(p)\| = o_p(1)$, and $\|\hat{\Sigma}(p) - \Sigma\| = o_p(1)$, where we have defined $\hat{A}(p) \equiv (\hat{A}_1(p), \dots, \hat{A}_p(p))$ and $A(p) \equiv (A_1, \dots, A_p)$.

Then

$$\hat{\theta}_h(p) = \frac{\frac{1}{T-p}\sum_{t=p+1}^{T-h} y_{t+h}\hat{x}_t(p)}{\left(\frac{1}{T-p}\sum_{t=p+1}^{T} \hat{x}_t(p)^2\right)^{1/2}} + O_p(\hat{R}(p)),$$

where

$$\hat{R}(p) \equiv \frac{\max\{1, \sup_{1\le t\le T}\|w_t\|\}^2}{T-p} + \left(\sum_{\ell=p-h+1}^{p} \|\hat{A}_\ell(p)\|^2\right)^{1/2}.$$

Thus, the VAR impulse response estimator $\hat{\theta}_h(p)$ approximately equals the LP impulse response estimator $\hat{\beta}_h(p)$ up to a scale factor that does not depend on the horizon h. The approximation error is of an order $O_p(\hat{R}(p))$ that is likely to be small unless the data is so persistent that the estimated VAR coefficients at the very longest lags are non-negligible.

Assumptions (i) and (ii) of the proposition are easily satisfied under standard nonparametric regularity conditions on the data generating process and a restriction on how quickly the lag length p can grow with T. See for example Lewis & Reinsel (1985) and Gonçalves & Kilian (2007).

B.1.1 Proof of Proposition B.1

We split the proof into several steps.

STEP 1. We will show that $\sum_{\ell=1}^{p} \|\hat{A}_\ell(p)\| = O_p(1)$. The statement follows from

$$\sum_{\ell=1}^{p} \|\hat{A}_\ell(p)\| \le \sum_{\ell=1}^{p} \|A_\ell\| + \sum_{\ell=1}^{p} \|\hat{A}_\ell(p) - A_\ell\| \le \sum_{\ell=1}^{\infty} \|A_\ell\| + \|\hat{A}(p) - A(p)\|$$

and then exploiting assumptions (i) and (ii).

STEP 2. We will show that $\sup_{p+1\le t\le T}\|\hat{u}_t(p)\| = \sup_{1\le t\le T}\|w_t\| \times O_p(1)$. Observe that

$$\begin{aligned}\sup_{p+1\le t\le T}\|\hat{u}_t(p)\| &= \sup_{p+1\le t\le T}\left\| w_t - \sum_{\ell=1}^{p}\hat{A}_\ell(p)w_{t-\ell}\right\| \\ &\le \left(\sup_{1\le t\le T}\|w_t\|\right)\left(1+\sum_{\ell=1}^{p}\|\hat{A}_\ell(p)\|\right).\end{aligned}$$

Step 1 then gives the desired result.

STEP 3. We will show that, for any $m = 0, 1, \ldots, h$,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p) = O_p\left(\frac{\sup_t\|w_t\|}{T-p}\right).$$

We have

$$\sum_{t=p+1}^{T}\hat{u}_{t-m}(p) = \sum_{t=p+1}^{T}\hat{u}_t(p) - \sum_{t=T-m+1}^{T}\hat{u}_t(p).$$

The first sum on the right-hand side is exactly zero by the orthogonality conditions (B.1). The second sum consists of m terms, each of which is $O_p(\sup_t\|w_t\|)$ by Step 2.

STEP 4. We will show that, for any $m = 1, 2, \ldots, h$,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)' = O_p\left(\left(\sum_{\ell=p-h+1}^{p}\|\hat{A}_\ell(p)\|^2\right)^{1/2}\right).$$

$\hat{u}_{t-m}(p)$ is a linear function of $w_{t-m}, w_{t-1-m}, \ldots, w_{t-p-m}$. By the orthogonality conditions (B.1), $\hat{u}_t(p)$ is orthogonal to $w_{t-m}, w_{t-1-m}, \ldots, w_{t-p}$ (and a constant). Thus,

$$\begin{aligned}\left\|\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)'\right\| &= \left\|\frac{1}{T-p}\sum_{t=p+m+1}^{T}\hat{u}_t(p)\sum_{\ell=p-m+1}^{p}w'_{t-m-\ell}\hat{A}_\ell(p)'\right\| \\ &\le \left(\sum_{\ell=p-m+1}^{p}\|\hat{A}_\ell(p)\|^2\right)^{1/2}\left(\sum_{\ell=p-m+1}^{p}\left\|\frac{1}{T-p}\sum_{t=p+m+1}^{T}\hat{u}_t(p)w'_{t-m-\ell}\right\|^2\right)^{1/2}.\end{aligned}$$

Note that $\sum_{\ell=p-m+1}^{p} \|\hat{A}_\ell(p)\|^2 \leq \sum_{\ell=p-h+1}^{p} \|\hat{A}_\ell(p)\|^2$ since $m \leq h$. Finally,

$$\begin{aligned}
\left\| \frac{1}{T-p} \sum_{t=p+m+1}^{T} \hat{u}_t(p) w'_{t-m-\ell} \right\| &\leq \left(\frac{1}{T-p} \sum_{t=p+m+1}^{T} \|\hat{u}_t(p)\|^2 \right)^{1/2} \left(\frac{1}{T-p} \sum_{t=p+m+1}^{T} \|w_{t-m-l}\|^2 \right)^{1/2} \\
&\leq \left(\frac{1}{T-p} \sum_{t=p+1}^{T} \|\hat{u}_t(p)\|^2 \right)^{1/2} \left(\frac{1}{T-p} \sum_{t=1}^{T} \|w_t\|^2 \right)^{1/2} \\
&\leq \|\hat{\Sigma}(p)\| \left(\frac{1}{T-p} \sum_{t=1}^{T} \|w_t\|^2 \right)^{1/2}.
\end{aligned}$$

The first factor on the right-hand side above is $O_p(1)$ by assumption (ii), while the second factor is $O_p(1)$ since $E\|w_t\|^2 < \infty$.

Step 5. Let $\ell, m \geq 0$ satisfy $m \leq h$ and $\ell \leq p$. If $m \leq \ell$, then

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w'_{t-\ell} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_t(p) w'_{t-(\ell-m)} + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right), \tag{B.2}$$

while if $m > \ell$, then

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w'_{t-\ell} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-(m-\ell)}(p) w'_t + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right), \tag{B.3}$$

where the $O_p(\cdot)$ terms are uniform in ℓ and m. Claim (B.3) is proven in the same way as (B.2), so we only prove the latter. Simply note that

$$\sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w'_{t-\ell} = \sum_{t=p+1}^{T} \hat{u}_t(p) w'_{t-(\ell-m)} - \sum_{t=T-m+1}^{T} \hat{u}_t(p) w'_{t-(\ell-m)},$$

and the second sum consists of m terms, each of which is $O_p(\sup_t \|w_t\|^2)$ by Step 2.

Step 6. We will show that, for any $\ell, m \geq 0$ such that $m \leq h$ and $m < \ell \leq p$,

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_{t-m}(p) w'_{t-\ell} = O_p\left(\frac{\sup_t \|w_t\|^2}{T-p} \right),$$

where the $O_p(\cdot)$ term is uniform in ℓ and m. By Step 5,

$$\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p)w'_{t-\ell} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)w'_{t-(\ell-m)} + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right).$$

Since $1 \le \ell - m \le p$, the sum on the right-hand side is precisely zero by the orthogonality conditions (B.1).

Step 7. Define for all $m = 0, 1, \dots, h$ the matrix $\hat{H}_m(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} w_t\hat{u}_{t-m}(p)'$. We will show that

$$\hat{H}_m(p) = \sum_{\ell=1}^{m}\hat{A}_\ell(p)\hat{H}_{m-\ell}(p) + O_p(\hat{R}(p)), \quad m = 1, 2, \dots, h.$$

Let $m = 1, \dots, h$ be arbitrary. Since

$$w_t = \hat{c}(p) + \sum_{\ell=1}^{p}\hat{A}_\ell(p)w_{t-\ell} + \hat{u}_t(p),$$

we obtain

$$\begin{aligned}
\hat{H}_m(p) &= \frac{1}{T-p}\sum_{t=p+1}^{T} w_t\hat{u}_{t-m}(p)' \\
&= \sum_{\ell=1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' \\
&\quad + \hat{c}(p)\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-m}(p)' \\
&\quad + \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_t(p)\hat{u}_{t-m}(p)'.
\end{aligned}$$

By Step 3, the second term above is $O_p(\frac{1}{T-p}\sup_t \|w_t\|)$. By Step 4, the third term is $O_p((\sum_{\ell=p-h+1}^{p}\|\hat{A}_\ell(p)\|^2)^{1/2})$. As for the first term above, we split it up as follows:

$$\begin{aligned}
\sum_{\ell=1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' &= \sum_{\ell=1}^{m}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)' \\
&\quad + \sum_{\ell=m+1}^{p}\hat{A}_\ell(p)\frac{1}{T-p}\sum_{t=p+1}^{T} w_{t-\ell}\hat{u}_{t-m}(p)'.
\end{aligned}$$

By Steps 1 and 6, the second term above is $O_p(\frac{1}{T-p}\sup_t \|w_t\|^2)$. By Steps 1 and 5, the first term above equals

$$\sum_{\ell=1}^{m} \hat{A}_\ell(p) \frac{1}{T-p} \sum_{t=p+1}^{T} w_{t-\ell} \hat{u}_{t-m}(p)' = \sum_{\ell=1}^{m} \hat{A}_\ell(p) \frac{1}{T-p} \sum_{t=p+1}^{T} w_t \hat{u}_{t-(m-\ell)}(p)' + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right)$$
$$= \sum_{\ell=1}^{m} \hat{A}_\ell(p) \hat{H}_{m-\ell}(p) + O_p\left(\frac{\sup_t \|w_t\|^2}{T-p}\right).$$

Step 8. We will show that $\hat{H}_m(p) = \hat{C}_m(p)\hat{\Sigma}(p) + O_p(\hat{R}(p))$ for all $m = 0, \ldots, h$. We proceed by induction on m. The claim is true by definition for $m = 0$. Assume the claim is true for all $m \leq \tilde{m} - 1$. Then Step 7 implies

$$\begin{aligned}
\hat{H}_{\tilde{m}} &= \sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p) \hat{H}_{\tilde{m}-\ell}(p) + O_p(\hat{R}(p)) \\
&= \sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p) \{\hat{C}_{\tilde{m}-\ell}(p)\hat{\Sigma}(p) + O_p(\hat{R}(p))\} + O_p(\hat{R}(p)) \\
&= \left(\sum_{\ell=1}^{\tilde{m}} \hat{A}_\ell(p) \hat{C}_{\tilde{m}-\ell}(p)\right) \hat{\Sigma}(p) + O_p(\hat{R}(p)) \\
&= \hat{C}_{\tilde{m}}(p)\hat{\Sigma}(p) + O_p(\hat{R}(p)).
\end{aligned}$$

Here the penultimate equality uses Step 1, and the last equality uses the recursive definition of $\hat{C}_{\tilde{m}}(p)$.

Step 9. We will show that $\|\hat{B}(p)^{-1}\| = O_p(1)$. This follows from assumption (ii), the continuity of the Cholesky decomposition at any positive definite matrix, and the assumption (i) that Σ is positive definite.

Step 10. Let e_x be the $(n_r + 2)$-th n_w-dimensional unit vector, i.e., $x_t = e_x' w_t$. Then

$$e_x' \hat{B}(p)^{-1} \hat{u}_t(p) = \frac{1}{(\frac{1}{T-p}\sum_{t=p+1}^{T} \hat{x}_t(p)^2)^{1/2}} \hat{x}_t(p)$$

for all $t = p+1, p+2, \ldots, T$. This is just the sample analogue of the population result (6)–(7), so we refrain from giving the details of the proof.

Step 11. We will show that

$$\hat{C}_h(p)\hat{B}(p)e_x = \frac{1}{(\frac{1}{T-p}\sum_{t=p+1}^{T}\hat{x}_t(p)^2)^{1/2}} \times \frac{1}{T-p}\sum_{t=p+1}^{T} w_t\hat{x}_{t-h}(p)' + O_p(\hat{R}(p)).$$

By Steps 8 and 9,

$$\hat{C}_h(p)\hat{B}(p) = \hat{C}_h(p)\hat{\Sigma}(p)\hat{B}(p)^{-1\prime} = \hat{H}_m\hat{B}(p)^{-1\prime} + O_p(\hat{R}_p).$$

Hence,

$$\hat{C}_h(p)\hat{B}(p)e_x = \frac{1}{T-p}\sum_{t=p+1}^{T} w_t\left(e_x'\hat{B}(p)^{-1}\hat{u}_{t-h}(p)\right)' + O_p(\hat{R}_p),$$

so the claim follows from Step 10.

Step 12. The statement of the proposition follows from Step 11 and the fact that $\hat{\theta}_h(p)$ by definition equals the (n_r+2)-th element of $\hat{C}_h(p)\hat{B}(p)e_x$. □

B.2 General system identification

We here show how arbitrary sign and zero identification restrictions on entire multivariate systems can be implemented in a local projection framework. The logic of our procedure is analogous to the examples in Section 3.2 – we estimate reduced-form impulse responses using LPs rather than a VAR, and then rotate the reduced-form impulse responses to conform with the chosen identifying restrictions.

We consider the most general case, in which the researcher wishes to impose sign and/or zero restrictions on impulse responses for all variables w_t and all horizons h up to some maximum H. Consider the coefficient vectors $\{\check{\beta}_{i,h}\}$ obtained from the $n_w \times (H+1)$ projections

$$w_{i,t+h} = \check{\mu}_{i,h} + \check{\beta}'_{i,h} w_t + \sum_{\ell=1}^{\infty} \check{\delta}'_{i,h,\ell} w_{t-\ell} + \check{\xi}_{i,h,t}, \quad i = 1, 2, \ldots, n_w, \quad h = 0, 1, 2, \ldots, H. \tag{B.4}$$

Let $C_h \equiv (\check{\beta}_{1,h}, \check{\beta}_{2,h}, \ldots, \check{\beta}_{n_w,h})'$ denote the $n_w \times n_w$ matrix of horizon-h stacked projection coefficients. The logic of Section 2 shows that C_h are the reduced-form impulse responses of w_t with respect to the Wold innovation $u_t \equiv w_t - E(w_t \mid \{w_\tau\}_{\tau<t})$ at horizon h. Similarly, the projection residuals $\check{\xi}_{\bullet,1,t} \equiv (\check{\xi}_{1,1,t}, \ldots, \check{\xi}_{n_w,1,t})'$ for the horizon-1 projections equal these Wold innovations u_t. Let $\text{Var}(\check{\xi}_{\bullet,1,t}) = BB'$ denote the Cholesky decomposition of the variance-covariance matrix of u_t, where B is lower triangular with strictly positive diagonal elements.

As in Rubio-Ramírez et al. (2010), for a given $n_w \times n_w$ orthogonal matrix Q (i.e., $Q'Q = QQ' = I_{n_w}$), the "structural" horizon-h impulse response matrix is

$$\Theta_h(Q, C, B) \equiv C_h B Q,$$

where entry (i, j) gives the response of variable i to shock j. Write the stacked $n_w(H+1) \times n_w$ matrix of impulse responses as $\Theta(Q, C, B) \equiv (\Theta_0(\bullet)', \Theta_1(\bullet)', \ldots, \Theta_H(\bullet)')'$.

Without loss of generality, normalize the first shock to be the shock of interest. Suppose also we normalize the scale of the shock so that it has variance 1. We can then compute the supremum of the identified set for the impulse response of the i-th variable at horizon h as the solution to the quadratic program[B.1]

$$\sup_{Q \in \mathbb{R}^{n_w \times n_w}} e_i' \Theta_h(Q, C, B) e_1 \tag{B.5}$$

[B.1] It is straight-forward to see that the solution to this program is independent of the choice of base rotation matrix (here the matrix B of the Cholesky decomposition of $\text{Var}(\check{\xi}_{\bullet,1,t})$).

subject to

$$S_j \Theta(Q, C, B) e_j \geq 0, \quad j = 1, 2, \ldots, n_w, \tag{B.6}$$

$$Z_j \Theta(Q, C, B) e_j = 0, \quad j = 1, 2, \ldots, n_w, \tag{B.7}$$

$$Q'Q = I_{n_w}, \tag{B.8}$$

where e_j is the j-th column of the n_w-dimensional identity matrix. The $n_w(H+1) \times n_w(H+1)$-dimensional matrices $\{S_j, Z_j\}$ are set in line with the chosen identification scheme as in Rubio-Ramírez et al. (2010) (see the example below). Vector inequalities are to be understood elementwise.

Note that the only difference between the above formulation and the SVAR approach in Rubio-Ramírez et al. (2010) is that here we obtain the reduced-form impulse responses C_h from the direct projections (B.4).

Remarks.

1. The infimum of the identified set is computed analogously.

2. Our set-up is purposefully general. If only the responses of a subset of variables at a subset of horizons are restricted, then the projections (B.4) need to only be estimated for that subset (and of course for the response variable of interest, at all relevant horizons). Our analysis in Section 3.2 is a special case with only one restricted variable and restrictions imposed only on responses to a single shock. In that special case we were also able to drop the quadratic constraint (B.8) and turn the program into a computationally convenient linear program, provided we changed the normalization of the scale of the shock to the unit-effect normalization (Stock & Watson, 2016).

3. When all sign restrictions are placed on a single shock, and if this shock is the shock of interest, then the boundaries of the identified set can be conveniently computed using the algorithm of Gafarov et al. (2018). Otherwise, the boundaries of the identified set can be approximated numerically through random draws of orthogonal matrices Q, either drawn from Haar measure on the space of orthogonal rotation matrices (for pure sign restrictions) or from subspaces consistent with any imposed zero restrictions, as in Arias et al. (2018).

4. It is straightforward to extend the above program to allow for magnitude restrictions

as in Kilian & Murphy (2012). For those, the zeros on the right-hand side of (B.6) are replaced by a vector of restriction-specific constants.

Example. We illustrate the construction of the restriction matrices $\{S_j, Z_j\}$ through an example. The researcher observes output and inflation, $w_t = (y_t, \pi_t)'$, and wishes to disentangle demand and supply shocks through the following identifying restrictions. First, expansionary demand shocks increase output for three periods, and increase inflation on impact. Second, expansionary supply shocks increase output on impact, and lower inflation for two periods. Let the first shock be the demand shock and the second the supply shock.

We collect impulse responses up to horizon $H = 2$ in the matrix Θ, and impose the sign restrictions through the two matrices S_1 (for demand shocks) and S_2 (for supply shocks). These matrices are given by

$$S_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad S_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The first two rows of S_1 and S_2 select the impact responses of $(y_t, \pi_t)'$, the second two rows select the $h = 1$ responses, and the final two rows the $h = 2$ responses.

B.3 Narrative identification

Researchers often observe only sparse, qualitative indicators of structural macro shocks (e.g., Romer & Romer, 1989; Ramey, 2011). For concreteness, consider the following model of a sparse narrative indicator z_t:

$$z_t = \begin{cases} -1 & \text{if } g(\varepsilon_{1,t} + \omega_t) \leq \underline{a}, \\ 0 & \text{if } \underline{a} \leq g(\varepsilon_{1,t} + \omega_t) \leq \bar{a}, \\ 1 & \text{if } g(\varepsilon_{1,t} + \omega_t) \geq \bar{a}, \end{cases} \tag{B.9}$$

for some function $g(\bullet)$ and some measurement error process ω_t that is independent of the structural shocks ε_t at all leads and lags. A particularly simple case is $g(\bullet) = \text{sign}(\bullet)$ – the researcher observes the sign of the contaminated shock $\varepsilon_{1,t} + \omega_t$ if it is sufficiently large.

The macroeconometric literature has developed two main ways of using sparse narrative instruments in semi-structural analysis. First, researchers restrict VAR-identified shocks $\tilde{\varepsilon}_{1,t}$ to have the same *sign* as z_t whenever $z_t \neq 0$ (Antolín-Díaz & Rubio-Ramírez, 2018; Ludvigson et al., 2020). In finite samples, this approach necessarily delivers a *set* of possible structural shocks. This finite-sample focus on one-off dates, however, does not lend itself naturally to the population focus in this paper. Second, narrative instruments can be cast in a proxy framework (Budnik & Rünstler, 2020), and so Section 3.3 of our analysis applies: Under the model (B.9), the indicator z_t is an IV that satisfies the exclusion and relevance restrictions (16). Hence, z_t can be used as an instrument in linear estimation procedures such as LP-IV, despite the presence of the measurement error ω_t. Furthermore, by Corollary 1, a recursive VAR with z_t ordered first estimates the same impulse responses as LP-IV. Such "dummy variable" instruments have been used in VAR analysis at least since Hamilton (2003).

B.4 Estimands in non-linear models

Our main result in Section 2.1 implies that linear local projections are exactly as "robust to non-linearities" as linear VAR methods, in population. We now show that the common LP/VAR estimand can be given a mathematically well-defined "best linear approximation" interpretation when the true underlying structural DGP is in fact non-linear. We also discuss to what extent this best linear approximation is structurally interesting.

Assume that the underlying structural DGP has the nonparametric causal structure

$$w_t = g(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \tag{B.10}$$

where $g(\cdot)$ is any non-linear function that yields a well-defined covariance stationary process $\{w_t\}$, and $\{\varepsilon_t\}$ is an n_ε-dimensional i.i.d. process with $\text{Cov}(\varepsilon_t) = I_{n_\varepsilon}$. The number of structural shocks ε_t may exceed the number of variables in w_t.

We show formally in Section B.4.1 below that we can represent the process (B.10) as the *linear* Structural Vector Moving Average model

$$w_t = \mu^* + \sum_{\ell=0}^{\infty} \Theta_\ell^* \varepsilon_{t-\ell} + \sum_{\ell=0}^{\infty} \Psi_\ell^* \zeta_{t-\ell},$$

where ζ_t is an n_w-dimensional white noise process that is uncorrelated at all leads and lags with the structural shocks ε_t. The argument exploits the Wold decomposition of the residual of w_t after projecting on the structural shocks. Hence, the linear SVMA model (10) in Assumption 3 should not be thought of as restrictive, provided we do not restrict the number of "shocks" relative to the number of variables.

The linear SVMA impulse responses Θ_ℓ^* corresponding to the structural shocks ε_t have a "best linear approximation" interpretation. Specifically,

$$(\Theta_0^*, \Theta_1^*, \dots) \in \underset{(\tilde{\Theta}_0, \tilde{\Theta}_1, \dots)}{\operatorname{argmin}} \; E\left[\left(g(\varepsilon_t, \varepsilon_{t-1}, \dots) - \textstyle\sum_{\ell=0}^{\infty} \tilde{\Theta}_\ell \varepsilon_{t-\ell}\right)^2\right]. \tag{B.11}$$

Thus, if a second-moment LP/VAR identification scheme is known to correctly identify the impulse responses in a linear SVMA model (10), and there is doubt about whether the true underlying DGP is in fact linear, the population estimand of the identification procedure can be given a formal "best linear approximation" interpretation. This is analogous to the "best linear predictor" property of Ordinary Least Squares in cross-sectional regression. In contrast, identification approaches that depart from standard linear projections – such as identification through higher moments or through heteroskedasticity – may not have a clear interpretation under functional form misspecification.

Of course, this best linear approximation may bear little resemblance to the impulse responses in the underlying nonlinear model, which will generally depend on the history and magnitudes of current and past shocks, unlike the linear impulse responses. In applications where the non-linearities of the true underlying DGP are of interest *per se*, non-linear versions of VAR or LP estimators can be applied, for example by adding interaction or polynomial terms, regime switching, stochastic volatility, etc. Such issues are outside the scope of this paper, which deals exclusively with linear estimators.

B.4.1 Technical details

We now give the technical details behind the "best linear approximation" interpretation of a non-linear model. Assume the nonparametric model (B.10), and that $\{w_t\}$ is covariance stationary and purely nondeterministic. Let the linear projection of w_t on the orthonormal basis $(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ be denoted $\sum_{\ell=0}^{\infty} \Theta_\ell^* \varepsilon_{t-\ell}$, with projection residual v_t. Assume v_t is either identically zero or purely non-deterministic. Then it has a Wold decomposition

$$v_t = \mu^* + \sum_{\ell=0}^{\infty} \Psi_\ell^* \zeta_{t-\ell},$$

where $\{\zeta_t\}$ is n_w-dimensional white noise with $\text{Cov}(\zeta_t) = I_{n_w}$. Since v_t is a function of $\{\varepsilon_\tau\}_{\tau \leq t}$, and $\{\varepsilon_t\}$ is i.i.d., we have $\text{Cov}(\varepsilon_{t+\ell}, v_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \geq 1$. Moreover, since v_t is a residual from a projection onto $\{\varepsilon_\tau\}_{\tau \leq t}$, we also have $\text{Cov}(\varepsilon_{t+\ell}, v_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \leq 0$. By the Wold decomposition theorem, ζ_t lies in the closed linear span of $\{v_\tau\}_{\tau \leq t}$, so we must have $\text{Cov}(\varepsilon_{t+\ell}, \zeta_t) = 0_{n_\varepsilon \times n_w}$ for all $\ell \in \mathbb{Z}$. Finally, the best linear approximation property (B.11) is a standard consequence of linear projection. We have thus verified all claims made in Section B.4. □

References

Antolín-Díaz, J. & Rubio-Ramírez, J. F. (2018). Narrative Sign Restrictions for SVARs. *American Economic Review*, *108*(10), 2802–29.

Arias, J. E., Rubio-Ramírez, J. F., & Waggoner, D. F. (2018). Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications. *Econometrica*, *86*(2), 685–720.

Budnik, K. & Rünstler, G. (2020). Identifying SVARs from Sparse Narrative Instruments. European Central Bank Working Paper No. 2353.

Gafarov, B., Meier, M., & Montiel Olea, J. L. (2018). Delta-Method Inference for a Class of Set-Identified SVARs. *Journal of Econometrics*, *203*(2), 316–327.

Gonçalves, S. & Kilian, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews*, *26*(6), 609–641.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*(2), 363–398.

Kilian, L. & Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models. *Journal of the European Economic Association*, *10*(5), 1166–1188.

Lewis, R. & Reinsel, G. C. (1985). Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis*, *16*(3), 393–411.

Ludvigson, S. C., Ma, S., & Ng, S. (2020). Shock Restricted Structural Vector-Autoregressions. Manuscript, New York University.

Ramey, V. A. (2011). Identifying Government Spending Shocks: It's All in the Timing. *Quarterly Journal of Economics*, *126*(1), 1–50.

Romer, C. D. & Romer, D. H. (1989). Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz. *NBER Macroeconomics Annual*, *4*, 121–170.

Rubio-Ramírez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, *77*(2), 665–696.

Stock, J. H. & Watson, M. W. (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2 chapter 8, (pp. 415–525). Elsevier.