Online Appendix for “Local Projections vs. VARs: Lessons From Thousands of DGPs”

Dake Li Mikkel Plagborg-Møller Christian K. Wolf

March 30, 2021

Contents

Appendix D Examples of estimated IRFs

Figures D.1 to D.6 provide a visual illustration of estimated impulse response functions (IRFs) from the six estimation procedures defined in Section 4. We fix a single (randomly chosen) DGP with an observed fiscal shock and simulate ten data sets with sample size $T = 200$. We then apply the six estimation methods to these ten data sets.

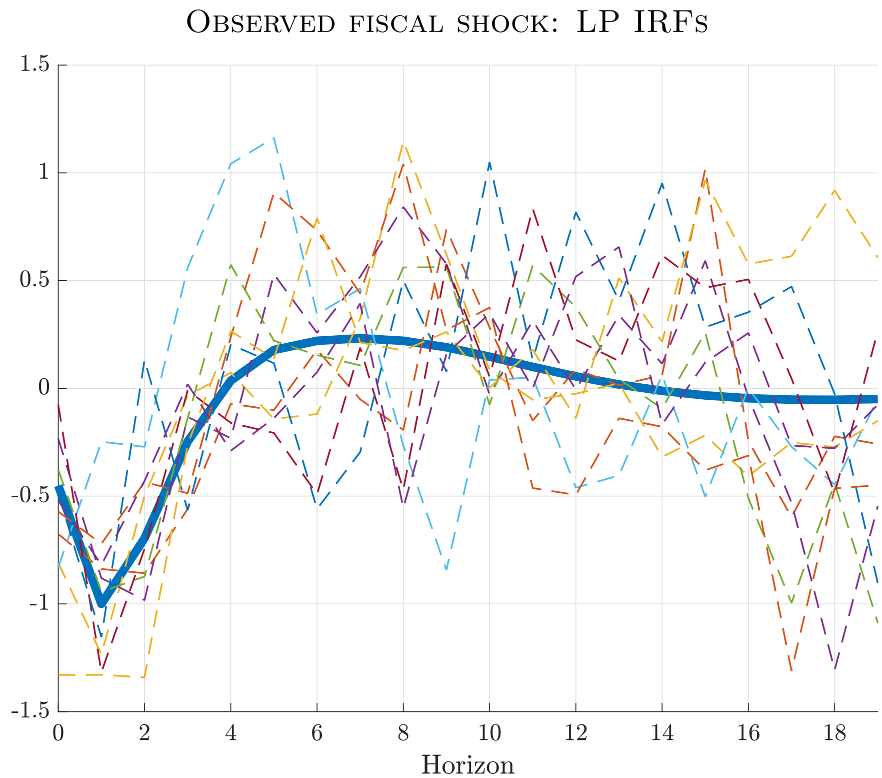

Figure D.1: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares LP impulse response estimates.

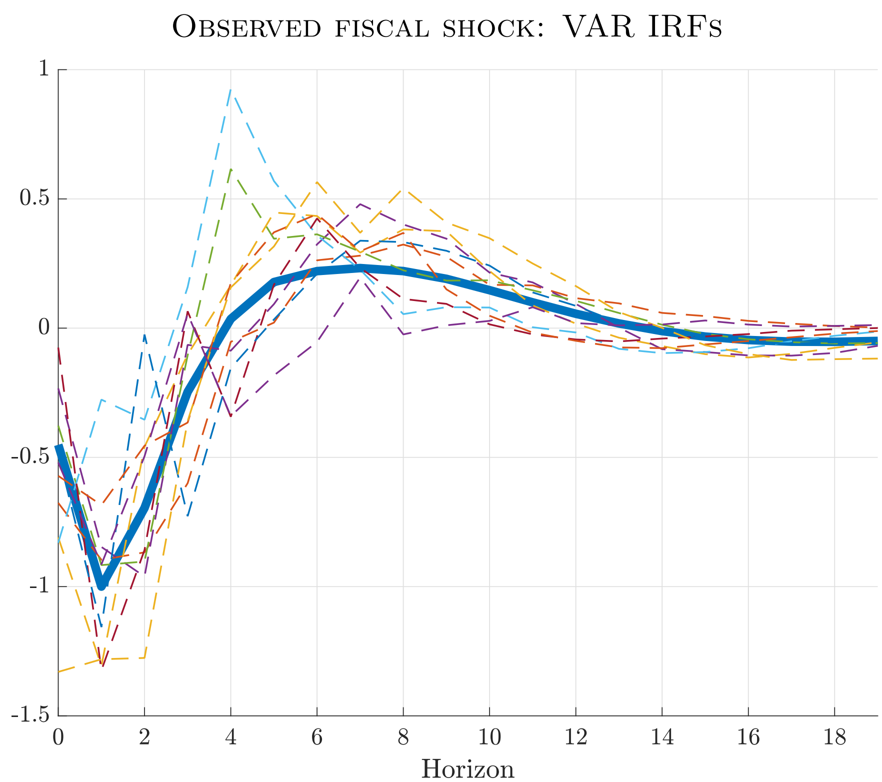

Figure D.2: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares VAR impulse response estimates.

Observed fiscal shock: BC-VAR IRFs

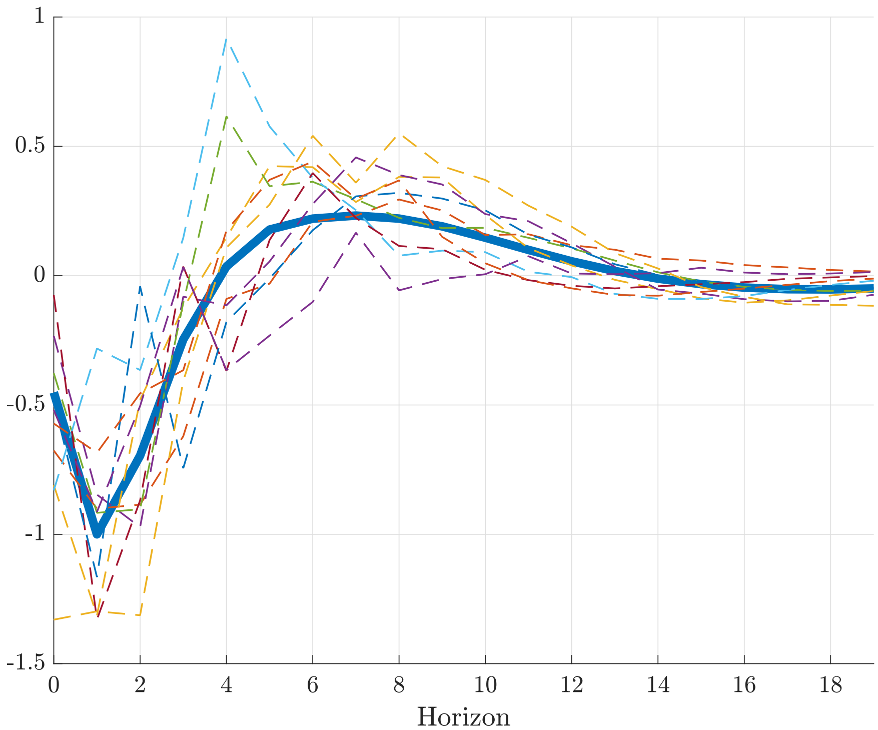

Figure D.3: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten bias-corrected VAR impulse response estimates.

Observed fiscal shock: BVAR IRFs

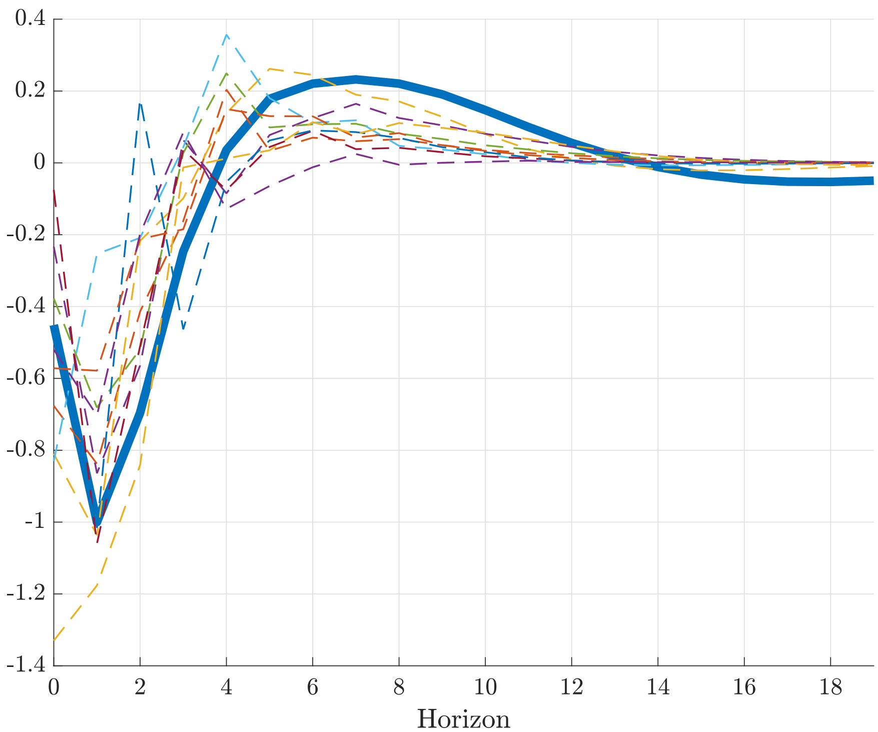

Figure D.4: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten Bayesian VAR impulse response estimates.

Observed fiscal shock: Penalized LP IRFs

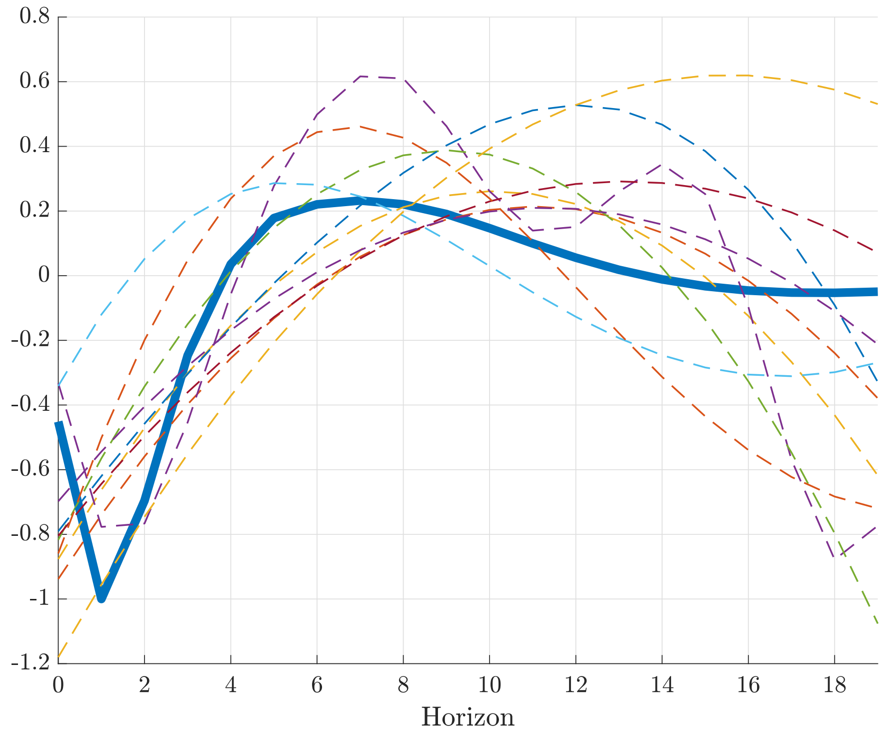

Figure D.5: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten penalized LP impulse response estimates.

Observed fiscal shock: VAR Averaging IRFs

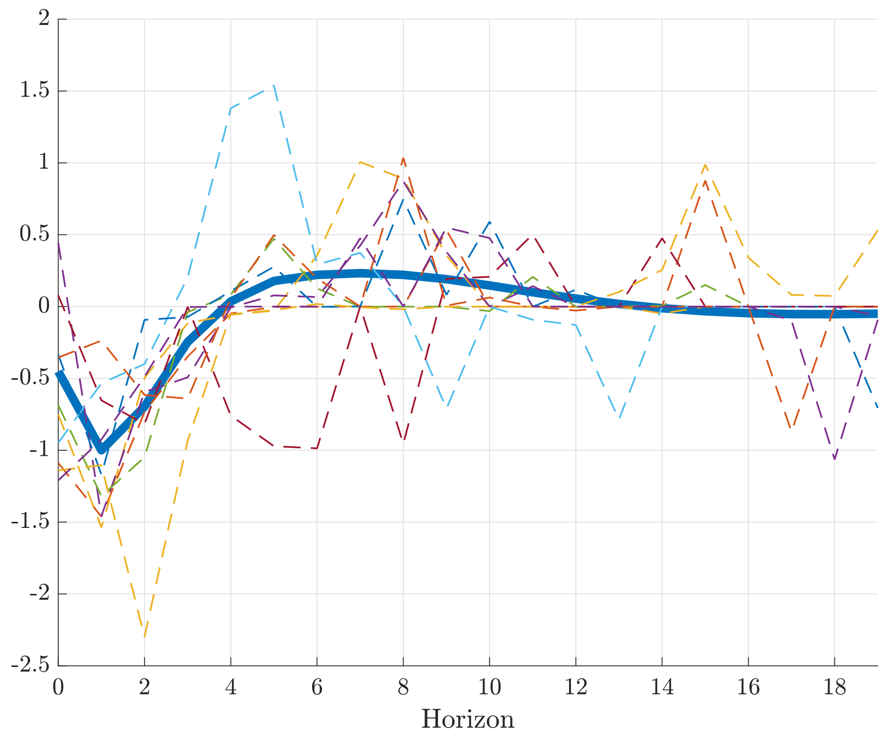

Figure D.6: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten VAR Averaging impulse response estimates.

DGP SUMMARY STATISTICS: RECURSIVE IDENTIFICATION

Percentile	min	10	25	50	75	90	max
Data and shocks							
trace(long-run var)/trace(var)	0.42	0.93	0.98	1.14	2.29	4.78	18.09
Largest VAR eigenvalue	0.82	0.84	0.84	0.84	0.84	0.86	0.91
Fraction of VAR coef's $\ell \geq 5$	0.02	0.10	0.15	0.23	0.34	0.44	0.84
Degree of shock invertibility	0.14	0.16	0.19	0.28	0.41	0.47	0.65
Impulse responses up to $h = 19$							
No. of interior local extrema	0	2	2	2	3	4	7
Horizon of max abs. value	0	0	0	1	2	4	11
Average/(max abs. value)	-0.50	-0.29	-0.16	-0.04	0.10	0.23	0.54
R^2 in regression on quadratic	0.00	0.21	0.39	0.57	0.67	0.81	0.97

Table E.1: Quantiles of various population parameters across the 6,000 DGPs for recursive identification. "long-run var": long-run variance. "Fraction of VAR coef's $\ell \geq 5$": $\sum_{\ell=5}^{50} \|A_\ell^w\| / \sum_{\ell=1}^{50} \|A_\ell^w\|$, where $\|\cdot\|$ is the Frobenius norm. "Average/(max abs. value)": $(\frac{1}{20}\sum_{h=0}^{19} \theta_h)/\max_h\{|\theta_h|\}$. "$R^2$ in regression on quadratic": R-squared from a regression of the impulse response function $\{\theta_h\}_{h=0}^{19}$ on a quadratic polynomial in h.

Appendix E Results: recursive identification

Here we provide results for the recursive impulse response estimand defined in Section 3.2.

Table E.1 shows summary statistics for the recursive identification setting, analogous to the summary statistics for the "observed shock" case in Section 3.4.

Figures E.1 and E.2 show the median (across DGPs) absolute bias and standard deviation of the various estimators. Figure E.3 depicts the best estimation method as a function of the horizon and the bias weight ω in the loss function (which is averaged across DGPs). These three figures are qualitatively and quantitatively similar to the corresponding figures for the "observed shock" estimands in Section 5.

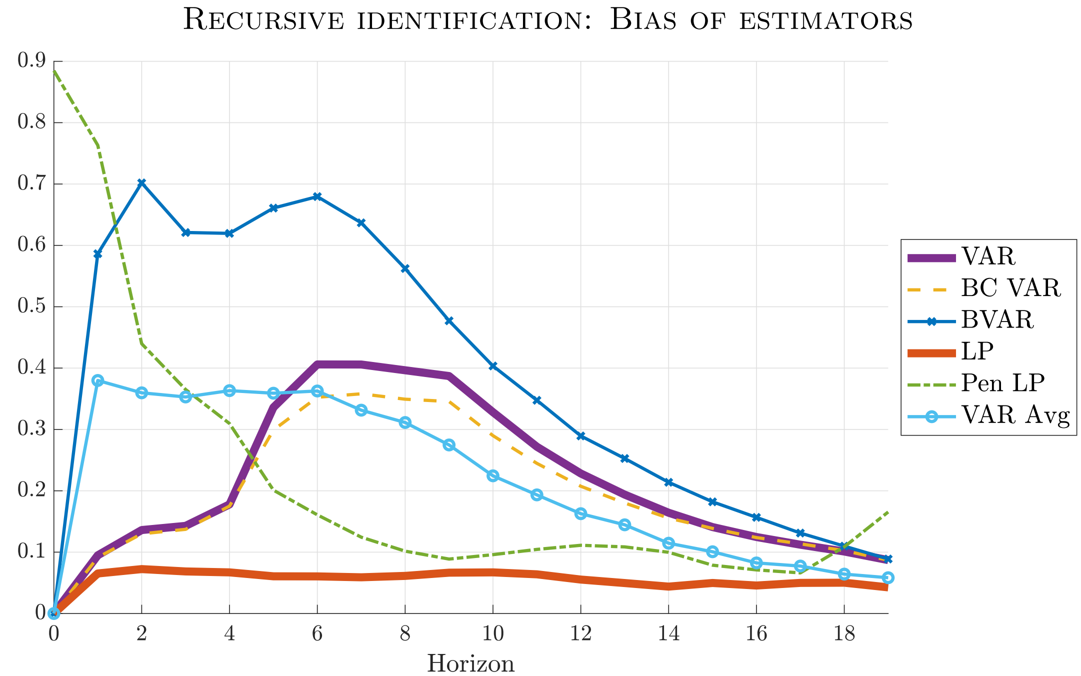

Figure E.1: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

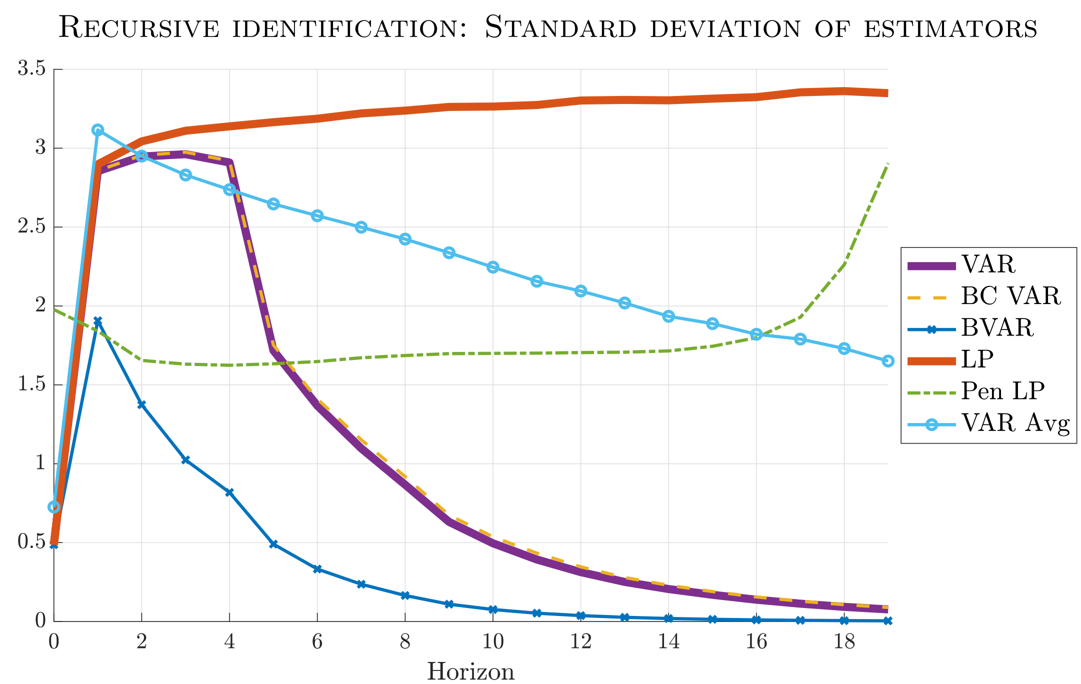

Figure E.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

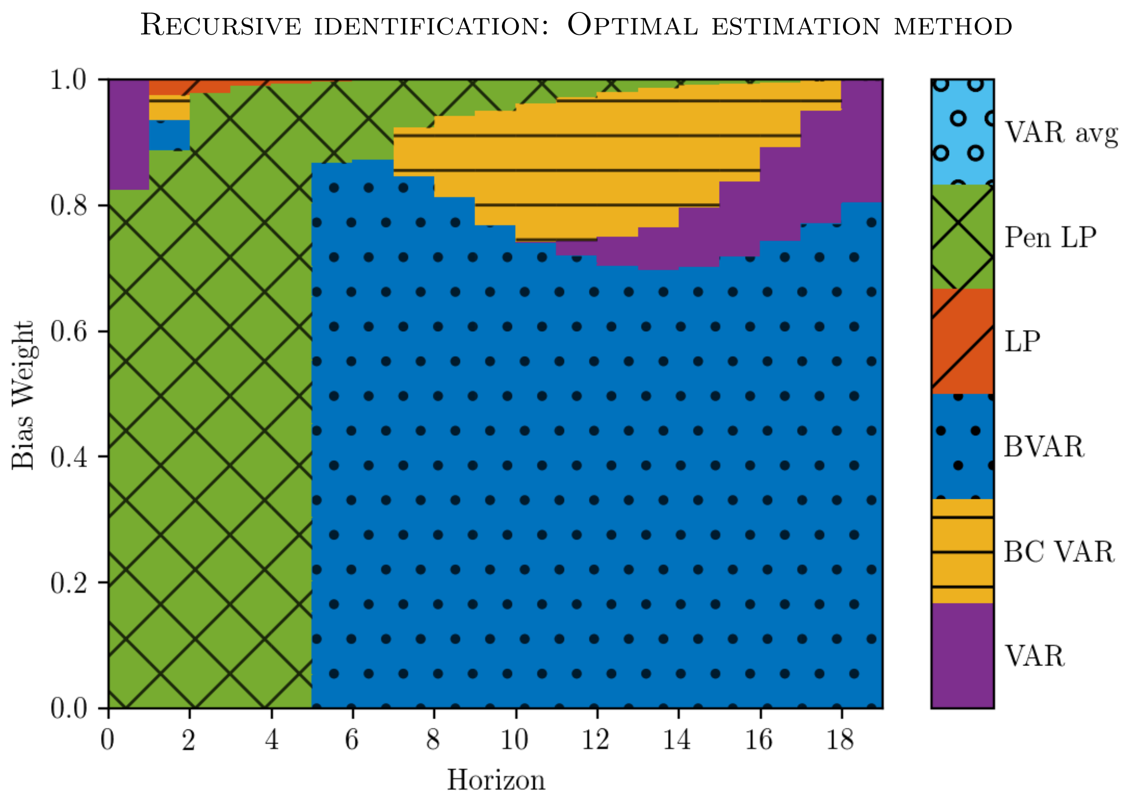

Figure E.3: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

Appendix F Results: longer estimation lag length

Here we provide results for "observed shock" identification when the estimation lag length is increased to $p = 8$ (recall that we set $p = 4$ in Section 5).

Figures F.1 and F.2 show the median (across DGPs) absolute bias and standard deviation of the estimation methods, while Figure F.3 shows the optimal method choice according to the loss function (which has been averaged across DGPs). The qualitative conclusions from Section 5 are unchanged, as long as we redefine "intermediate horizons" to mean horizons that are moderately longer than $h = p = 8$ (instead of 4). In particular, and consistent with the theoretical results of Plagborg-Møller & Wolf (2021), least-squares LP performs similarly to least-squares VAR at all horizons $h \leq p = 8$. Moreover, unless the weight on bias in the loss function is very close to 1, penalized LP remains attractive at short horizons $h \leq p$, while BVAR remains attractive at intermediate and long horizons.

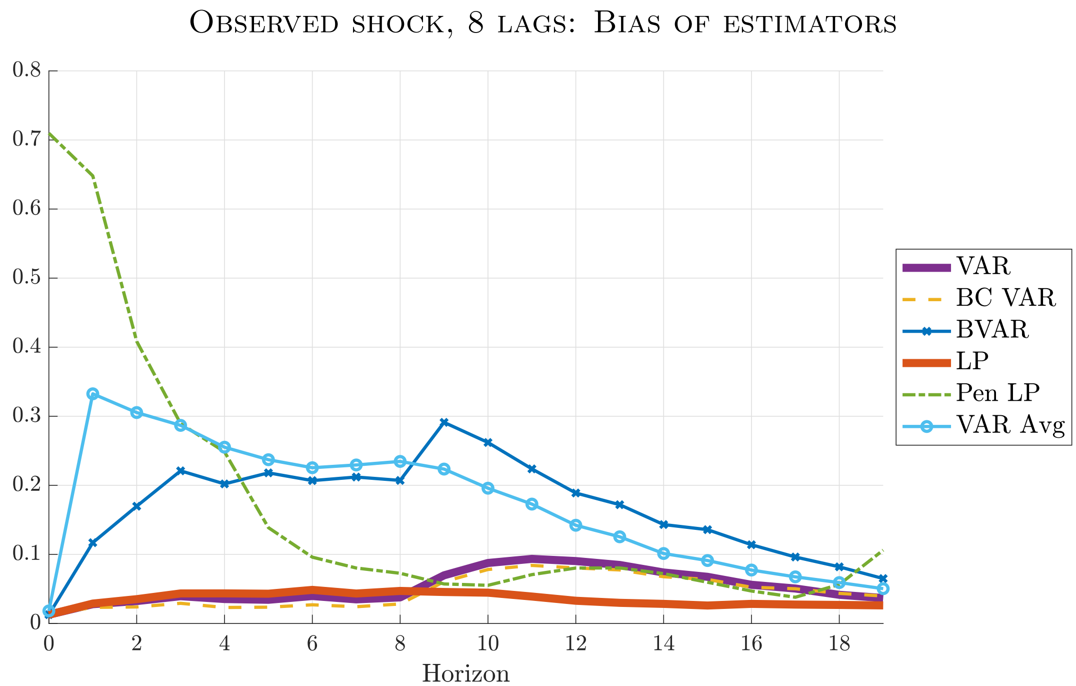

Figure F.1: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

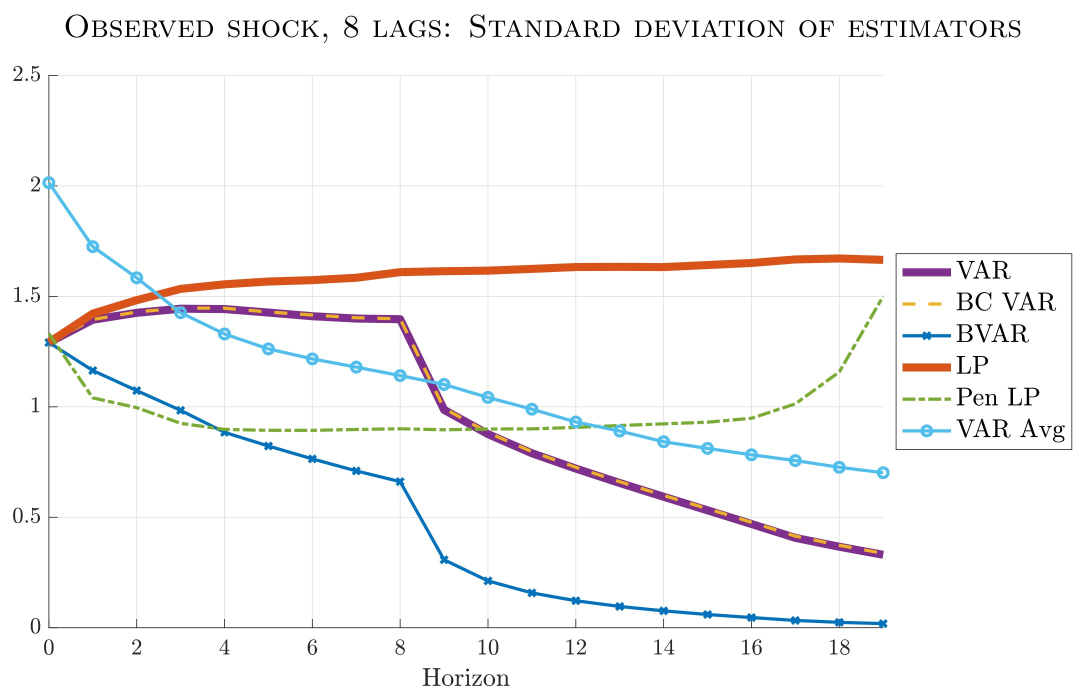

Figure F.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

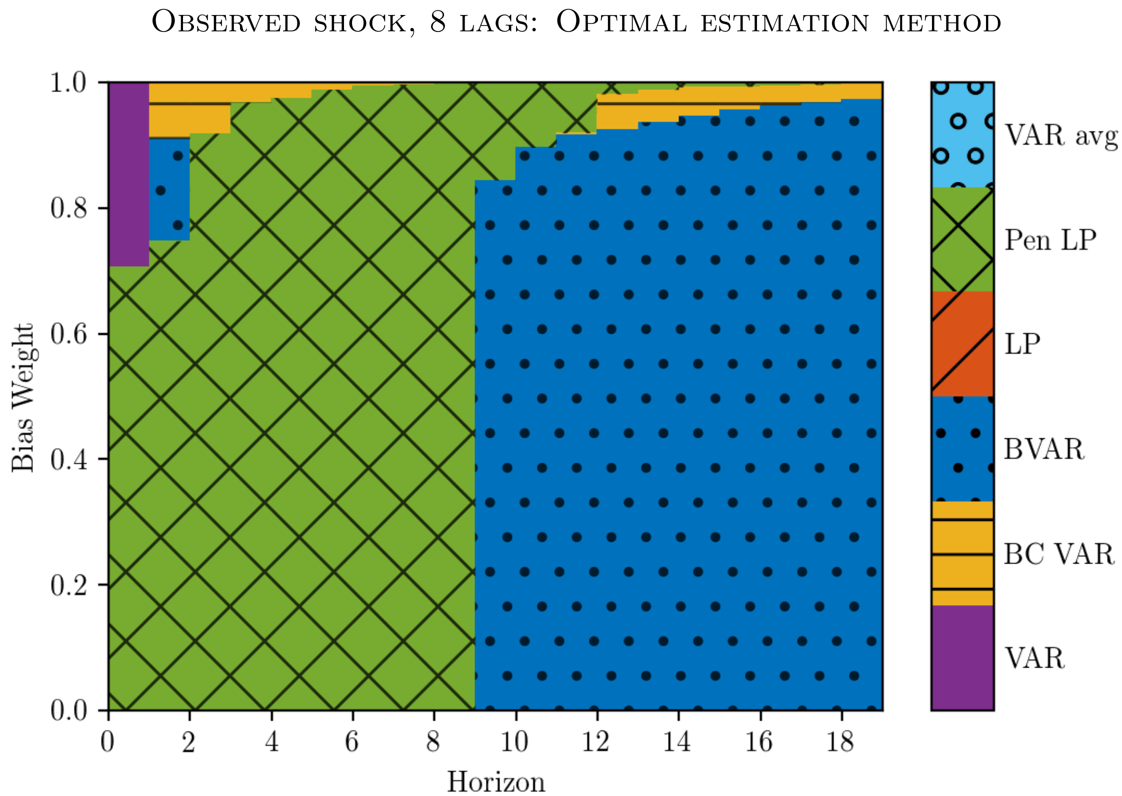

Figure F.3: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

Appendix G Results: fiscal and monetary shocks

Here we break down the results from Section 5 into separate results for fiscal shock DGPs and monetary shock DGPs.

Figures G.1 and G.2 show the bias and standard deviation plots for the 3,000 fiscal shock DGPs, while Figures G.3 and G.4 show the analogous figures for the 3,000 monetary shock DGPs. The results are qualitatively similar across the two kinds of DGPs, including the relative rankings of the various estimation procedures. However, the overall level of the absolute biases and standard deviations is somewhat higher in the fiscal shock case for all estimation methods.

Observed fiscal shock: Bias of estimators

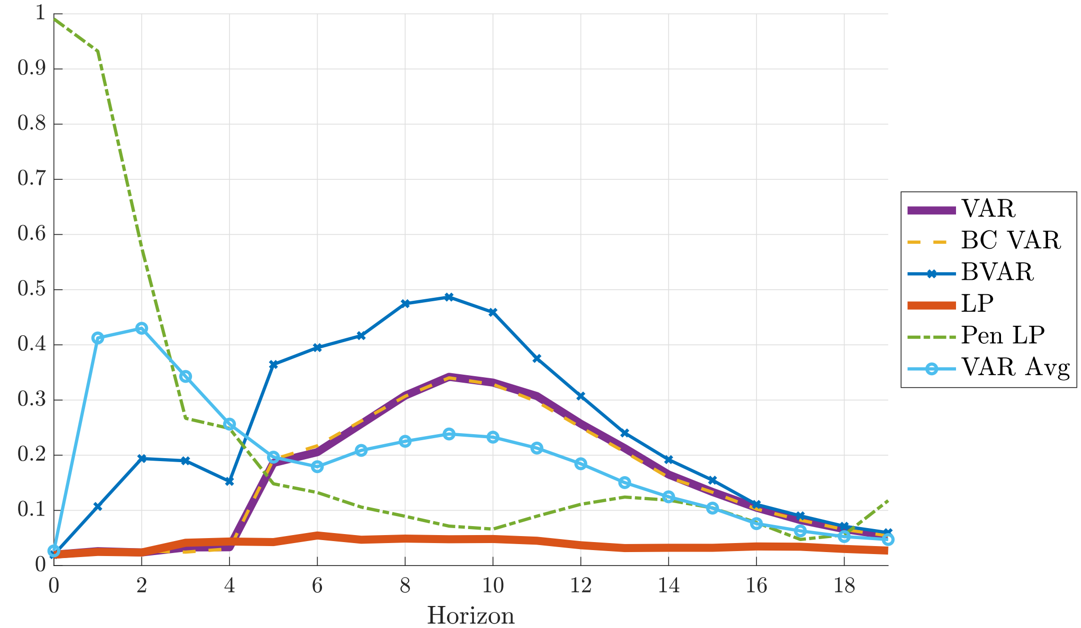

Figure G.1: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

Observed fiscal shock: Standard deviation of estimators

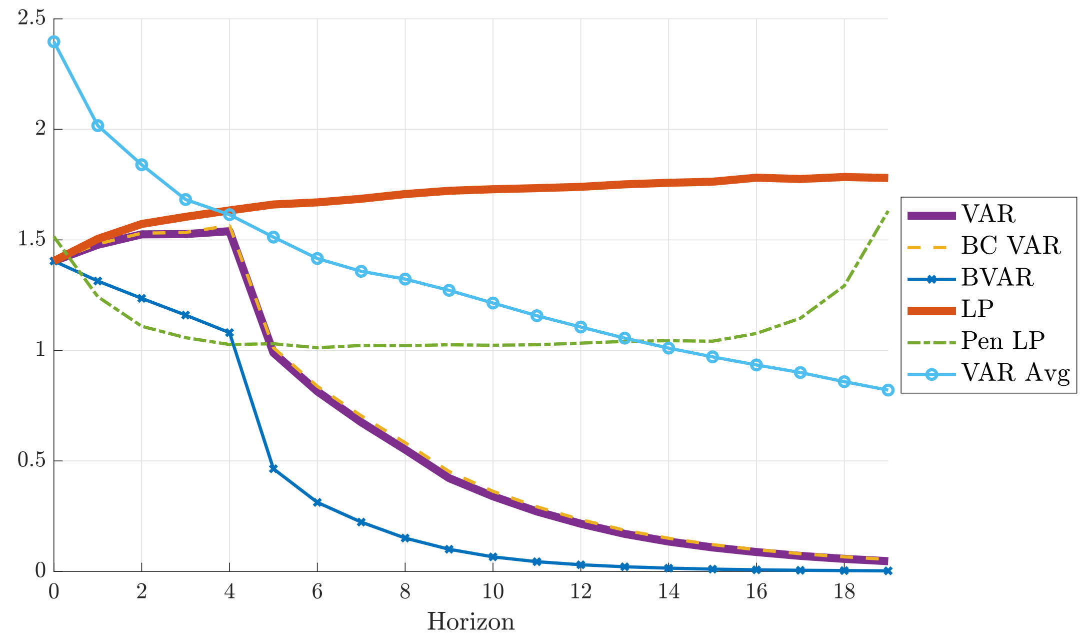

Figure G.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

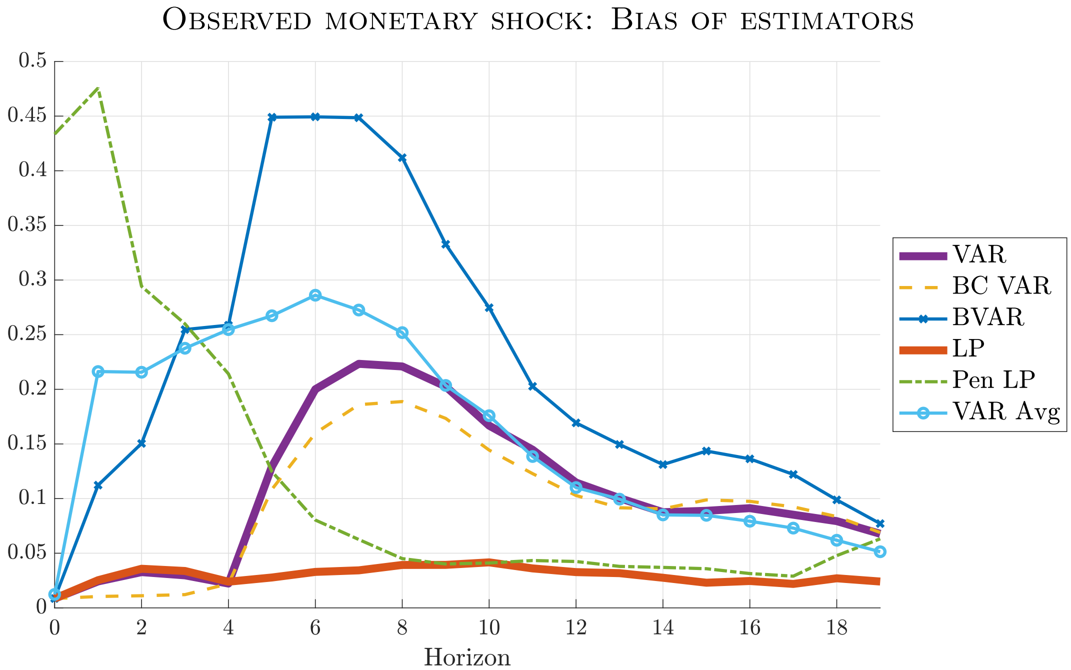

Figure G.3: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

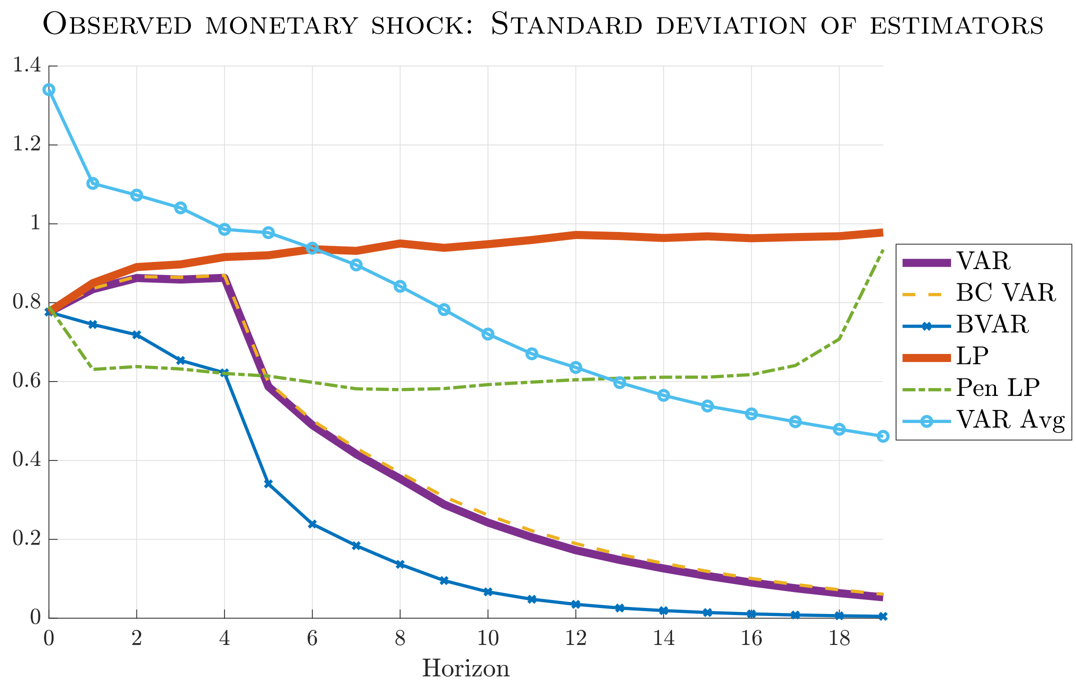

Figure G.4: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

Appendix H Results: IV estimators

Figures H.1 and H.2 plot the mean bias and standard deviation of the estimation procedures in the case of IV identification. The relative ranking of the various estimation procedures is essentially the same as in the median bias and interquartile range plots presented in Section 5.4. However, the standard deviation plot shows that the VAR averaging estimator has a particularly fat-tailed sampling distribution compared to the other estimators.

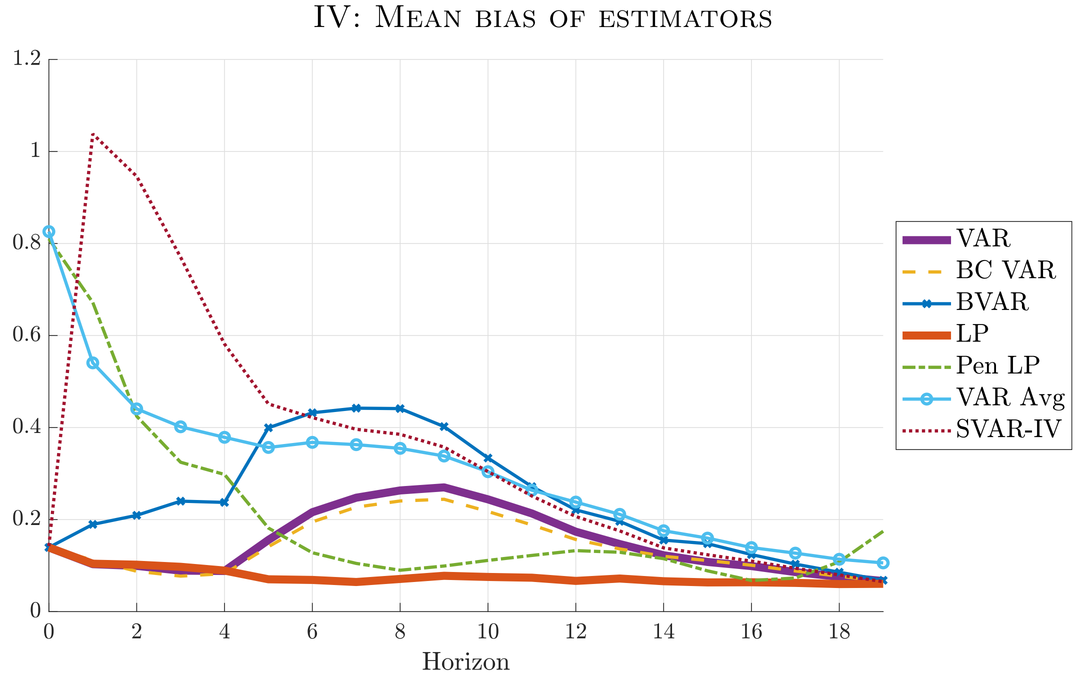

Figure H.1: Median (across DGPs) of absolute mean bias of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

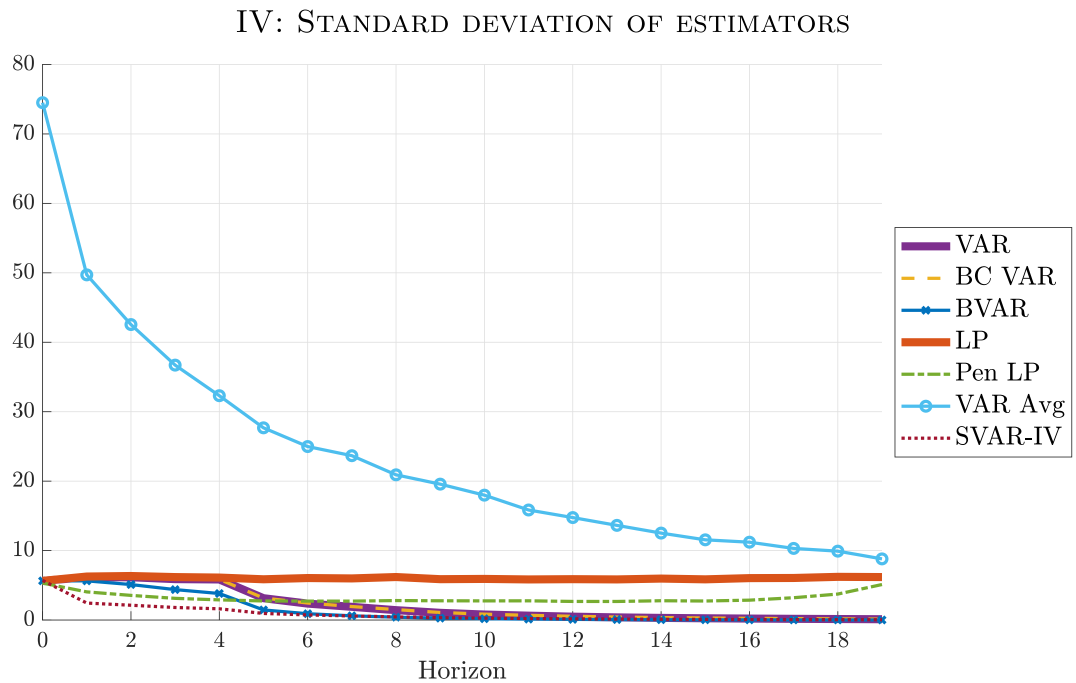

Figure H.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{20}\sum_{h=0}^{19}\theta_h^2}$.

Appendix I Proofs

I.1 Auxiliary lemmas

Before proving Proposition 1, we state and prove some auxiliary lemmas. All lemmas below impose the assumptions of Proposition 1.

Lemma I.1. *Define the process $\tilde{y}_t \equiv (1-\rho L)^{-1}(\varepsilon_{1,t}+\varepsilon_{2,t})$ for all t. Then for all $j=1,2$ and $\ell \geq 0$,*

$$\frac{1}{T}\sum_{t=1}^{T}(y_t-\tilde{y}_t)^2 = O_p(T^{-1}), \quad \frac{1}{T}\sum_{t=1}^{T}(y_t-\tilde{y}_t)\varepsilon_{j,t+\ell} = O_p(T^{-1}).$$

Proof. From the DGP (1) we have

$$y_t-\tilde{y}_t = \frac{\alpha}{\sqrt{T}}B(L)\varepsilon_{2,t}, \quad B(L) \equiv (1-\rho L)^{-1}L.$$

Since the moving average coefficients of $B(L)$ are geometrically decaying, the first statement of the lemma follows from Phillips & Solo (1992, Theorem 3.7) and the assumption of finite variances of the shocks. The second statement of the lemma follows from Chebyshev's inequality and the fact that the process $\varepsilon_{j,t+\ell} \times B(L)\varepsilon_{2,t}$ is serially uncorrelated under our assumptions on the shocks. □

In the following, we define $\tilde{w}_t \equiv (\varepsilon_{1,t}, \tilde{y}_t)'$, where $\tilde{y}_t$ was defined in Lemma I.1. Recall also the definitions of θ_h, $\hat{\beta}_h$, $\hat{\delta}_h$, $\hat{A}$, $\hat{\kappa}$, and the unit vector e_j from Section 2.

Lemma I.2. *We have*

$$\hat{\beta}_h-\theta_h = \frac{1}{T}\sum_{t=2}^{T-h}\left\{\sum_{\ell=1}^{h}\rho^{h-\ell}\varepsilon_{1,t+\ell}+\sum_{\ell=0}^{h}\rho^{h-\ell}\varepsilon_{2,t+\ell}\right\}\varepsilon_{1,t}+o_p(T^{-1/2}). \tag{I.1}$$

Proof. Let $\hat{\varepsilon}_{1,t} \equiv \varepsilon_{1,t}-\hat{b}'w_{t-1}$ be the residual from an auxiliary regression of $\varepsilon_{1,t}$ on w_{t-1}. Using Lemma I.1, it is straightforward to show that

$$\hat{b} = \left\{E(\tilde{w}_t\tilde{w}_t')^{-1}+o_p(1)\right\}\left\{\frac{1}{T}\sum_{t=2}^{T-h}\tilde{w}_{t-1}\varepsilon_{1,t}+O_p(T^{-1})\right\} = O_p(T^{-1/2}),$$

where the last step applies Chebyshev's inequality to the sample average of the serially uncorrelated process $\tilde{w}_{t-1}\varepsilon_{1,t}$, using the assumption $E(\varepsilon_{j,t}^4)<\infty$.

By the Frisch-Waugh theorem and sample orthogonality of $\hat{\varepsilon}_{1,t}$ and w_{t-1}, we may write

$$\hat{\beta}_h = \theta_h + \frac{\frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_h \hat{\varepsilon}_{1,t})\hat{\varepsilon}_{1,t}}{\frac{1}{T}\sum_{t=2}^{T-h}\hat{\varepsilon}_{1,t}^2} = \theta_h + \frac{\frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_h \varepsilon_{1,t} - \theta_h y_{t-1})\hat{\varepsilon}_{1,t}}{\frac{1}{T}\sum_{t=2}^{T-h}\hat{\varepsilon}_{1,t}^2}. \tag{I.2}$$

Lemma I.1 and $\hat{b} = o_p(1)$ yield $\frac{1}{T}\sum_{t=2}^{T-h}\hat{\varepsilon}_{1,t}^2 \xrightarrow{p} E(\varepsilon_{1,t}^2) = 1$. We can therefore focus on the numerator in the fraction in (I.2), which we decompose as

$$\frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_h \varepsilon_{1,t} - \theta_h y_{t-1})\varepsilon_{1,t} + \frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_h \varepsilon_{1,t} - \theta_h y_{t-1})(\hat{\varepsilon}_{1,t} - \varepsilon_{1,t}). \tag{I.3}$$

We first show that the first term above equals the sum on the right-hand side of (I.1). Iteration on the DGP (1) implies

$$y_{t+h} - \theta_h \varepsilon_{1,t} - \theta_h y_{t-1} = \sum_{\ell=1}^{h} \rho^{h-\ell}\varepsilon_{1,t+\ell} + \sum_{\ell=0}^{h}\rho^{h-\ell}\left(\varepsilon_{2,t+\ell} + \frac{\alpha}{\sqrt{T}}\varepsilon_{2,t+\ell-1}\right).$$

The desired conclusion then follows from

$$\frac{1}{T}\sum_{t=2}^{T-h}\sum_{\ell=0}^{h}\rho^{h-\ell}\frac{\alpha}{\sqrt{T}}\varepsilon_{2,t+\ell-1}\varepsilon_{1,t} = \frac{\alpha}{T^{3/2}}\sum_{t=2}^{T-h}\sum_{\ell=0}^{h}\rho^{h-\ell}\varepsilon_{2,t+\ell-1}\varepsilon_{1,t} = O_p(T^{-1}),$$

which can be verified using Chebyshev's inequality and the fact that the summand in the sum over t is a serially uncorrelated process.

We finish the proof by showing that the second term in (I.3) is $O_p(T^{-1})$. This term equals

$$-\frac{1}{T}\sum_{t=2}^{T-h}\left\{\sum_{\ell=1}^{h}\rho^{h-\ell}\varepsilon_{1,t+\ell} + \sum_{\ell=0}^{h}\rho^{h-\ell}\left(\varepsilon_{2,t+\ell} + \frac{\alpha}{\sqrt{T}}\varepsilon_{2,t+\ell-1}\right)\right\} w_{t-1}'\hat{b}.$$

Using $\hat{b} = O_p(T^{-1/2})$ and Lemma I.1, it suffices to show that

$$\frac{1}{T}\sum_{t=2}^{T-h}\varepsilon_{j,t+\ell}\tilde{w}_{t-1} = O_p(T^{-1/2})$$

for all $j = 1, 2$ and $\ell \geq 0$, and

$$\frac{1}{T^{3/2}}\sum_{t=2}^{T-h}\varepsilon_{2,t-1}\tilde{w}_{t-1} = O_p(T^{-1/2}).$$

Both of these statements follow easily from Chebyshev's inequality. $\square$

Lemma I.3. *Define* $A_0 \equiv \left(\begin{smallmatrix} 0 & 0 \\ 0 & \rho \end{smallmatrix}\right)$. *We have*

$$\hat{A} - A_0 = \left(\frac{1}{T} \sum_{t=2}^{T} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix} \tilde{w}'_{t-1} + \frac{\alpha \sigma_2^2}{\sqrt{T}} e_2 e'_2 \right) E(\tilde{w}_t \tilde{w}'_t)^{-1} + o_p(T^{-1/2}),$$

$$\hat{\kappa} - 1 = \frac{1}{T} \sum_{t=2}^{T} \varepsilon_{1,t} \varepsilon_{2,t} + o_p(T^{-1/2}).$$

Proof. By appealing repeatedly to Lemma I.1, and applying standard laws of large numbers and mean-square moment bounds, we get

$$\begin{aligned} \hat{A} - A_0 &= \left(\frac{1}{T} \sum_{t=2}^{T} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} + \frac{\alpha}{\sqrt{T}} \varepsilon_{2,t-1} \end{pmatrix} w'_{t-1} \right) \left(\frac{1}{T} \sum_{t=2}^{T} w_{t-1} w'_{t-1} \right)^{-1} \\ &= \left(\frac{1}{T} \sum_{t=2}^{T} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix} \tilde{w}'_{t-1} + \frac{\alpha}{T^{3/2}} e_2 \sum_{t=2}^{T} \varepsilon_{2,t-1} \tilde{w}'_{t-1} + O_p(T^{-1}) \right) \\ &\quad \times \left(\frac{1}{T} \sum_{t=2}^{T} \tilde{w}_{t-1} \tilde{w}'_{t-1} + o_p(1) \right)^{-1} \\ &= \left(\frac{1}{T} \sum_{t=2}^{T} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix} \tilde{w}'_{t-1} + \frac{\alpha}{\sqrt{T}} e_2 E(\varepsilon_{2,t-1} \tilde{w}'_{t-1}) + O_p(T^{-1}) \right) \\ &\quad \times \left(E(\tilde{w}_t \tilde{w}'_t) + o_p(1) \right)^{-1}. \end{aligned}$$

Note that $E(\tilde{w}_{t-1} \varepsilon_{2,t-1}) = \sigma_2^2 e_2$. This proves the first statement of the lemma.

Next, by the Frisch-Waugh Theorem, $\hat{\kappa} \equiv \hat{\Sigma}_{21}/\hat{\Sigma}_{11}$ equals the coefficient on $\varepsilon_{1,t}$ in an OLS regression of y_t on $\varepsilon_{1,t}$ and w_{t-1}. In other words, $\hat{\kappa}$ equals the impact LP estimate $\hat{\beta}_0$. The second statement of the lemma then follows from Lemma I.2 applied to $h = 0$. □

I.2 Proof of Proposition 1

We derive the asymptotic distributions of the LP and VAR estimators in that order.

LP. It follows from Lemma I.2 and a standard martingale central limit theorem that

$$\sqrt{T}(\hat{\beta}_h - \theta_h) \xrightarrow{d} N(0, \text{aVar}_{\text{LP}}),$$

where

$$
\begin{aligned}
\mathrm{aVar}_{\mathrm{LP}} &= E(\varepsilon_{1,t}^2) E\left(\left\{\sum_{\ell=1}^{h} \rho^{h-\ell}\varepsilon_{1,t+\ell} + \sum_{\ell=0}^{h} \rho^{h-\ell}\varepsilon_{2,t+\ell}\right\}^2\right) \\
&= E\left(\left\{\sum_{\ell=1}^{h} \rho^{h-\ell}\varepsilon_{1,t+\ell} + \sum_{\ell=0}^{h} \rho^{h-\ell}\varepsilon_{2,t+\ell}\right\}^2\right) \\
&= \sum_{\ell=1}^{h} \rho^{2(h-\ell)} E(\varepsilon_{1,t}^2) + \sum_{\ell=0}^{h} \rho^{2(h-\ell)} E(\varepsilon_{2,t}^2) \\
&= \sum_{\ell=0}^{h} \rho^{2(h-\ell)}(1+\sigma_2^2) - \rho^{2h} \\
&= (1+\sigma_2^2)\frac{1-\rho^{2(h+1)}}{1-\rho^2} - \rho^{2h}.
\end{aligned}
$$

VAR. We derive the asymptotic distribution of $\hat{\delta}_h$ by appealing to the delta method. Let $f_h(A,\kappa) \equiv e_2' A^h \gamma$, where $\gamma = (1,\kappa)'$, so that $\hat{\delta}_h = f_h(\hat{A},\hat{\kappa})$. We need the Jacobians of this transformation with respect to $\mathrm{vec}(A)$ and κ. According to Lemma I.3, the Jacobians should be evaluated at $\operatorname{plim}\hat{A} = A_0 \equiv \left(\begin{smallmatrix} 0 & 0 \\ 0 & \rho \end{smallmatrix}\right)$ and $\operatorname{plim}\hat{\kappa} = 1$. Thus, γ should be evaluated at $\gamma_0 \equiv (1,1)'$.

First, for $h \geq 2$,

$$
\begin{aligned}
\left.\frac{\partial e_2' A^h \gamma}{\partial \operatorname{vec}(A)}\right|_{A=A_0,\gamma=\gamma_0} &= (\gamma' \otimes e_2') \sum_{j=1}^{h} (A_0')^{h-j} \otimes A_0^{j-1} \\
&= (\gamma' \otimes e_2') \left((A_0')^{h-1} \otimes I + I \otimes A_0^{h-1} + \sum_{j=2}^{h-1} (A_0')^{h-j} \otimes A_0^{j-1} \right) \\
&= (\gamma_0' \otimes e_2') \rho^{h-1} \left(\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \otimes I + I \otimes \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) + \sum_{j=2}^{h-1} \left(\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \otimes \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \right) \right) \\
&= (\gamma_0' \otimes e_2') \rho^{h-1} \Big(\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \otimes I + I \otimes \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) + (h-2) \left(\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \otimes \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \right) \Big) \\
&= \rho^{h-1} \Big(e_2' \otimes e_2' + \gamma_0' \otimes e_2' + (h-2)(e_2' \otimes e_2') \Big) \\
&= \rho^{h-1} \Big((\gamma_0 + (h-1)e_2)' \otimes e_2' \Big) \\
&= \rho^{h-1} \Big((1,h) \otimes e_2' \Big),
\end{aligned}
$$

where the third-last equality uses $\gamma_0' \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) = e_2' \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) = e_2'$, and the last equality uses $\gamma_0 = (1,1)'$. It's clear that the final formula is also the correct Jacobian for $h = 1$. Note that the form of the above Jacobian implies that the only part of $\hat{A}$ that will contribute to the asymptotic distribution of $\hat{\delta}_h$ is the second row $e_2'\hat{A}$.

Second, for any $h \geq 1$,

$$\left.\frac{\partial e_2' A^h \gamma}{\partial \kappa}\right|_{A=A_0,\gamma=\gamma_0} = e_2' A_0^h e_2 = \rho^h.$$

Next, Lemma I.3 and a standard martingale central limit theorem imply

$$\sqrt{T}(\hat{A} - A_0)' e_2 \xrightarrow{d} N(\text{aBias}(\hat{A}' e_2), \text{aVar}(\hat{A}' e_2)), \quad \sqrt{T}(\hat{\kappa} - 1) \xrightarrow{d} N(0, \text{aVar}(\hat{\kappa})),$$

where

$$\begin{aligned}
\text{aBias}(\hat{A}' e_2) &= \alpha \sigma_2^2 E(\tilde{w}_t \tilde{w}_t')^{-1} e_2, \\
\text{aVar}(\hat{A}' e_2) &= E(\tilde{w}_t \tilde{w}_t')^{-1}(1 + \sigma_2^2), \\
\text{aVar}(\hat{\kappa}) &= \text{Var}(\varepsilon_{1,t}\varepsilon_{2,t}) = \sigma_2^2.
\end{aligned}$$

Moreover, $\hat{A}' e_2$ and $\hat{\kappa}$ are asymptotically independent by Lemma I.3, since

$$\text{Cov}(\tilde{w}_{t-1}(\varepsilon_{1,t} + \varepsilon_{2,t}), \varepsilon_{1,t}\varepsilon_{2,t}) = E(\tilde{w}_{t-1}) E[(\varepsilon_{1,t} + \varepsilon_{2,t})\varepsilon_{1,t}\varepsilon_{2,t}] = 0.$$

Finally, note that

$$E(\tilde{w}_t \tilde{w}_t')^{-1} = \frac{1}{\sigma_{0,y}^2 - 1} \begin{pmatrix} \sigma_{0,y}^2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Given all the preceding ingredients, we can apply the delta method to conclude that

$$\sqrt{T}(\hat{\delta}_h - \theta_h) \xrightarrow{d} N(\text{aBias}_{\text{VAR}}, \text{aVar}_{\text{VAR}}),$$

where

$$\begin{aligned}
\text{aBias}_{\text{VAR}} &= \rho^{h-1}(1, h)\, \text{aBias}(\hat{A}' e_2) + \rho^h \times 0 \\
&= \rho^{h-1}(h - 1) \frac{\alpha \sigma_2^2}{\sigma_{0,y}^2 - 1},
\end{aligned}$$

$$\begin{aligned}
\text{aVar}_{\text{VAR}} &= \rho^{2(h-1)}(1, h)\, \text{aVar}(\hat{A}' e_2)(1, h)' + \rho^{2h}\, \text{aVar}(\hat{\kappa}) \\
&= \rho^{2(h-1)} \frac{1 + \sigma_2^2}{\sigma_{0,y}^2 - 1}(\sigma_{0,y}^2 - 2h + h^2) + \rho^{2h} \sigma_2^2
\end{aligned}$$

$$= \rho^{2(h-1)}(1+\sigma_2^2)\left(1+\frac{(h-1)^2}{\sigma_{0,y}^2-1}\right)+\rho^{2h}\sigma_2^2. \qquad \square$$

I.3 Asymptotic power of t-statistic

Here we prove the following result, which was discussed in Footnote 9.

Proposition I.1. *Assume the conditions of Proposition 1. Let $\hat{\tau}$ be the conventional t-statistic for testing the significance of the second lag in a univariate AR(2) regression for $\{y_t\}$. Then $\hat{\tau} \xrightarrow{d} N\left(-\rho\frac{\sigma_2^2}{1+\sigma_2^2}\alpha, 1\right)$.*

Proof. Denote $Y_t \equiv (y_t, y_{t-1})'$ and $\hat{S} \equiv \frac{1}{T-2}\sum_{t=3}^T Y_{t-1}Y_{t-1}'$. Let $\hat{\rho}_2$ be the coefficient on the second lag in the AR(2) regression, and let $\hat{s}^2$ be the sample variance of the residuals from this regression. Then standard OLS algebra gives

$$\hat{\tau} \equiv \frac{\hat{\rho}_2}{\sqrt{\hat{s}^2(e_2'\hat{S}^{-1}e_2)/T}} = \frac{e_2'\hat{S}^{-1}\frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t-1}(\varepsilon_{1,t}+\varepsilon_{2,t}+\frac{\alpha}{\sqrt{T}}\varepsilon_{2,t-1})}{\sqrt{\hat{s}^2(e_2'\hat{S}^{-1}e_2)}}.$$

Denote $\tilde{Y}_t \equiv (\tilde{y}_t, \tilde{y}_{t-1})'$ and $S_0 \equiv E(\tilde{Y}_t\tilde{Y}_t')$, using the notation from Supplemental Appendix I.1. Straight-forward calculations employing Lemma I.1 imply that

$$\hat{S} \xrightarrow{p} S_0, \quad \hat{s}^2 \xrightarrow{p} \operatorname{Var}(\varepsilon_{1,t}+\varepsilon_{2,t}) = 1+\sigma_2^2.$$

Since

$$S_0^{-1} = \begin{pmatrix} \sigma_{0,y}^2 & \rho\sigma_{0,y}^2 \\ \rho\sigma_{0,y}^2 & \sigma_{0,y}^2 \end{pmatrix}^{-1} = \frac{1}{1+\sigma_2^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

we have $\hat{s}^2(e_2'\hat{S}^{-1}e_2) \xrightarrow{p} (1+\sigma_2^2)(e_2'S_0^{-1}e_2) = 1$. Lemma I.1 also implies

$$\frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t-1}\left(\varepsilon_{1,t}+\varepsilon_{2,t}+\frac{\alpha}{\sqrt{T}}\varepsilon_{2,t-1}\right) = \frac{1}{\sqrt{T}}\sum_{t=3}^T \tilde{Y}_{t-1}\left(\varepsilon_{1,t}+\varepsilon_{2,t}+\frac{\alpha}{\sqrt{T}}\varepsilon_{2,t-1}\right) + O_p(T^{-1/2}).$$

Thus, by a standard martingale law of large numbers and central limit theorem,

$$\begin{aligned}\hat{\tau} &= \frac{1}{\sqrt{T}}\sum_{t=3}^T (e_2'S_0^{-1}\tilde{Y}_{t-1})(\varepsilon_{1,t}+\varepsilon_{2,t}) + \frac{\alpha}{T}\sum_{t=3}^T (e_2'S_0^{-1}\tilde{Y}_{t-1})\varepsilon_{2,t-1} + o_p(1) \\ &\xrightarrow{d} N\left(0, \operatorname{Var}\left[(e_2'S_0^{-1}\tilde{Y}_{t-1})(\varepsilon_{1,t}+\varepsilon_{2,t})\right]\right) + \alpha E\left[(e_2'S_0^{-1}\tilde{Y}_{t-1})\varepsilon_{2,t-1}\right].\end{aligned}$$

Since $e_2' S_0^{-1} \tilde{Y}_{t-1} = (\tilde{y}_{t-2} - \rho \tilde{y}_{t-1})/(1+\sigma_2^2)$, it is easily verified that

$$\operatorname{Var}\left[(e_2' S_0^{-1} \tilde{Y}_{t-1})(\varepsilon_{1,t} + \varepsilon_{2,t})\right] = \operatorname{Var}\left(\frac{\tilde{y}_{t-2} - \rho \tilde{y}_{t-1}}{1+\sigma_2^2}\right) \operatorname{Var}(\varepsilon_{1,t} + \varepsilon_{2,t}) = 1,$$

and

$$E\left[(e_2' S_0^{-1} \tilde{Y}_{t-1})\varepsilon_{2,t-1}\right] = \frac{0 - \rho E(\tilde{y}_{t-1}\varepsilon_{2,t-1})}{1+\sigma_2^2} = -\rho \frac{\sigma_2^2}{1+\sigma_2^2}.$$

This concludes the proof. □

References

Phillips, P. C. B. & Solo, V. (1992). Asymptotics for Linear Processes. *Annals of Statistics*, *20*(2), 971–1001.

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