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

Dake Li Mikkel Plagborg-Møller Christian K. Wolf

June 24, 2022

Contents

Appendix C Definition of recursive shock estimand

Here we formally define the impulse response estimand for the recursive identification scheme discussed in Section 3.2.

It is straightforward to show from (5)–(7) that w_t follows an *infinite-order* VAR process,

$$w_t = \sum_{\ell=1}^{\infty} A_\ell^w w_{t-\ell} + u_t^w, \tag{C.1}$$

where u_t^w is a white noise process. We provide formulas for $\{A_\ell^w, \Sigma_u^w\}$ as a function of DFM primitives below. Define the Cholesky decomposition $\text{Var}(u_t^w) = \Sigma_u^w = B^w(B^w)'$, with B^w lower triangular, and define $C^w(L) \equiv A^w(L)^{-1}$, where $A^w(L) = I_{n_{\bar{w}}} - \sum_{\ell=1}^{\infty} A_\ell^w L^\ell$. Then we define the recursive impulse response estimands as

$$\theta_h \equiv \frac{C^w_{\iota_y,\bullet,h} B^w_{\bullet,\iota_i}}{C^w_{\iota_i,\bullet,0} B^w_{\bullet,\iota_i}}, \quad h = 0, 1, 2, \ldots, \tag{C.2}$$

where ι_y and ι_i are the indices corresponding to y_t and i_t in the vector $\bar{w}_t$, respectively. Unlike the observed shock and IV estimands considered in Section 3.2, the estimand (C.2) might not equal the model-implied *structural* impulse response of the variable y_t with respect to any aggregate shock $\varepsilon_{j,t}$ in the DFM.[C.1] In other words, the expression (C.2) is the impulse response with respect to a potentially non-structural innovation.

To interpret the estimand (C.2), consider two popular applied identification schemes. First, in the monetary policy shock identification scheme of Christiano et al. (1999), y_t may be aggregate output, i_t is the nominal interest rate, and the nominal rate is typically ordered *after* all other observables. The recursive estimand is then the impulse response of output to a residualized interest rate innovation, normalized by the impact response of interest rates. Second, for the fiscal policy shock identification procedure in Blanchard & Perotti (2002), we may again take y_t to be aggregate output and let i_t be aggregate government spending. In this case, reduced-form innovations in the government spending equation are treated as structural shocks, and so (C.2) gives impulse responses to those innovations, normalized by the impact response of government spending.

We now provide an explicit formula for the recursive shock estimand in terms of the

[C.1] A necessary condition for the impulse responses (C.2) to equal structural impulse responses from (5)–(7) is that $\varepsilon_{j,t} \in \text{span}(\{\bar{w}_{t-\ell}\}_{\ell=0}^{\infty})$ for at least one shock j. A sufficient condition for $\varepsilon_t \in \text{span}(\{\bar{w}_{t-\ell}\}_{\ell=0}^{\infty})$ is that $n_{\bar{w}} = n_f$, $\bar{\Lambda}$ is non-singular, and $\Xi_i = 0$ for all i in $\bar{w}_t$.

primitive parameters of the DFM. We first express the reduced-form VAR representation (C.1) in terms of factor model primitives. We can reduce the encompassing factor model to only include the selected observables $w_t = \bar{w}_t$ as

$$\begin{aligned} f_t &= \Phi(L) f_{t-1} + H\varepsilon_t, && \text{(C.3)} \\ w_t &= \Lambda^w f_t + v_t^w, && \text{(C.4)} \\ v_t^w &= \Delta^w(L) v_{t-1} + \Xi^w \xi_t^w. && \text{(C.5)} \end{aligned}$$

Combining (C.4) and (C.5), we get

$$[I - \Delta^w(L)L] w_t = [I - \Delta^w(L)L]\Lambda^w f_t + \Xi^w \xi_t^w. \tag{C.6}$$

We can then write the system in state-space form (Fernández-Villaverde et al., 2007),

$$\begin{aligned} s_{t+1} &= A s_t + B\zeta_t, && \text{(C.7)} \\ w_t^* &= C s_t + D\zeta_t, && \text{(C.8)} \end{aligned}$$

where $s_t \equiv [f_t', f_{t-1}', \ldots, f_{t-q}']'$, $w_t^* = [I - \Delta^w(L)L] w_t$, $\zeta_t \equiv (\varepsilon_{t+1}', \xi_t')'$, A is the companion matrix for the lag polynomial $\Phi(L)$, $B \equiv \left(\begin{smallmatrix} H' & 0 & \cdots \\ 0 & 0 & \cdots \end{smallmatrix}\right)'$, $C \equiv [I, -\Delta_1^w, \ldots, -\Delta_q^w] \left(\begin{smallmatrix} \Lambda^w & 0 & \cdots \\ 0 & \Lambda^w & \cdots \\ \vdots & \vdots & \ddots \end{smallmatrix}\right)$, and $D \equiv (0, \Xi^w)$.[C.2]

Given the state-space system (C.7)–(C.8), straightforward manipulations give the standard reduced-form VAR representation for w_t^*:

$$w_t^* = C[I - (A - KC)L]^{-1} K w_{t-1}^* + u_t^{w^*}, \tag{C.9}$$

where $u_t^{w^*}$ is the Wold innovation for w_t^*, $\mathrm{Var}(u_t^{w^*}) = C\Sigma C' + DD'$, and (K, Σ) satisfy

$$\begin{aligned} \Sigma &= A\Sigma A' + BB' - (A\Sigma C' + BD')(C\Sigma C' + DD')^{-1}(A\Sigma C' + BD')', \\ K &= (A\Sigma C' + BD')(C\Sigma C' + DD')^{-1}. \end{aligned}$$

Given (C.9) as the VAR(∞) representation for $w_t^* = [I - \Delta^w(L)L] w_t$, we obtain the VAR($\infty$)

[C.2] This state-space form is valid if $p \leq q+1$, which is true in our encompassing factor model. When $p > q+1$, we need to accordingly change $s_t = [f_t', f_{t-1}', \ldots, f_{t-p+1}']'$ and $C = [I, -\Delta_1^w, \ldots, -\Delta_q^w] \left(\begin{smallmatrix} \Lambda^w & & & \\ & \Lambda^w & & \\ & & \ddots & \\ & & & 0 \end{smallmatrix}\right)$.

representation for w_t itself as[C.3]

$$\{I - C[I - (A - KC)L]^{-1}KL\}[I - \Delta^w(L)L]w_t = u_t^w.$$

The representation (C.1) then gives the recursive shock estimand (C.2), as discussed above.

[C.3]Note that these two Wold innovations are equivalent, i.e., $u_t^{w^*} \equiv u_t^w$.

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.

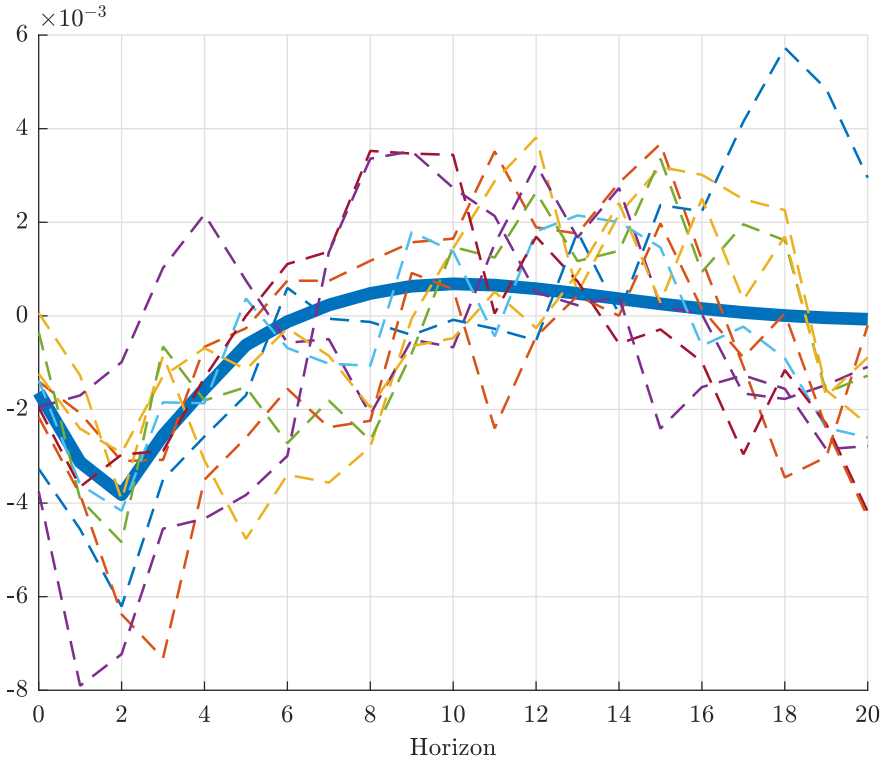

Figure D.1: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares LP impulse response estimates.

Observed fiscal shock: VAR IRFs

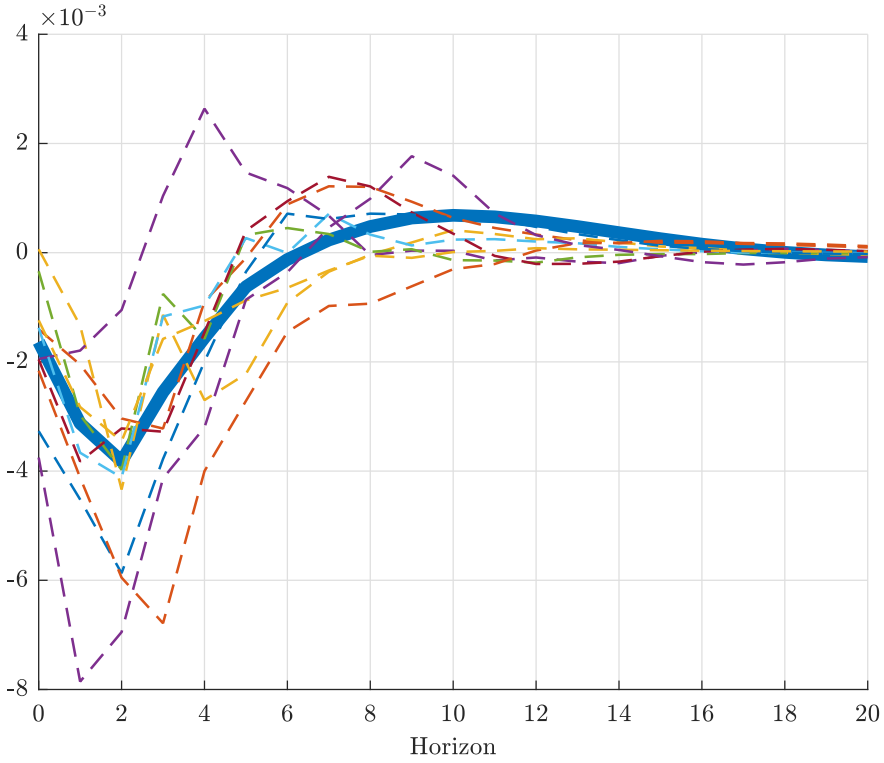

Figure D.2: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares VAR impulse response estimates.

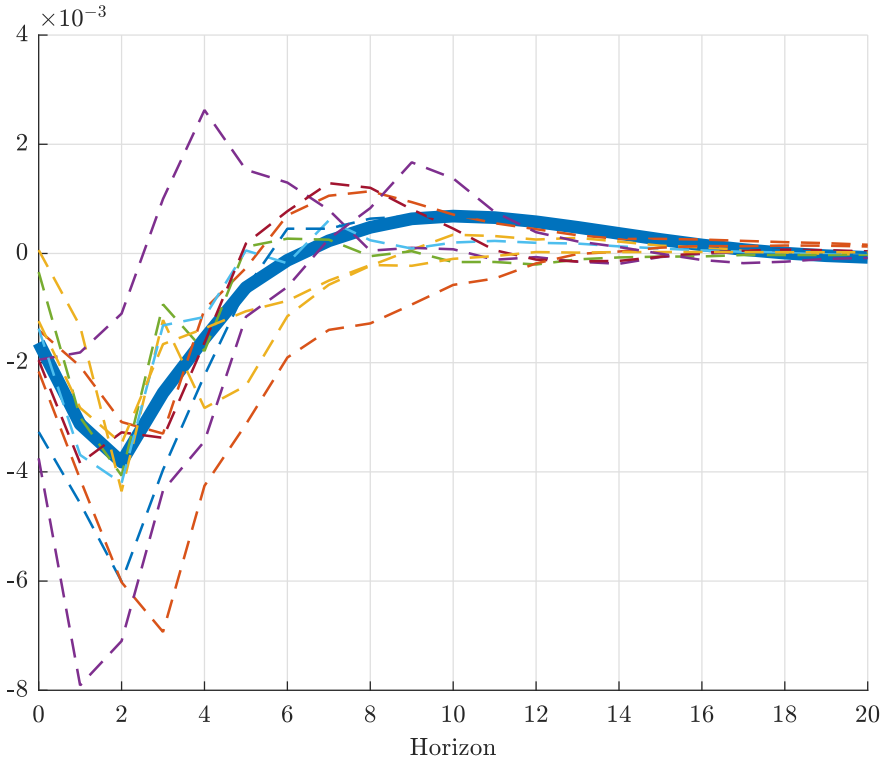

Figure D.3: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten bias-corrected VAR impulse response estimates.

Observed fiscal shock: BVAR IRFs

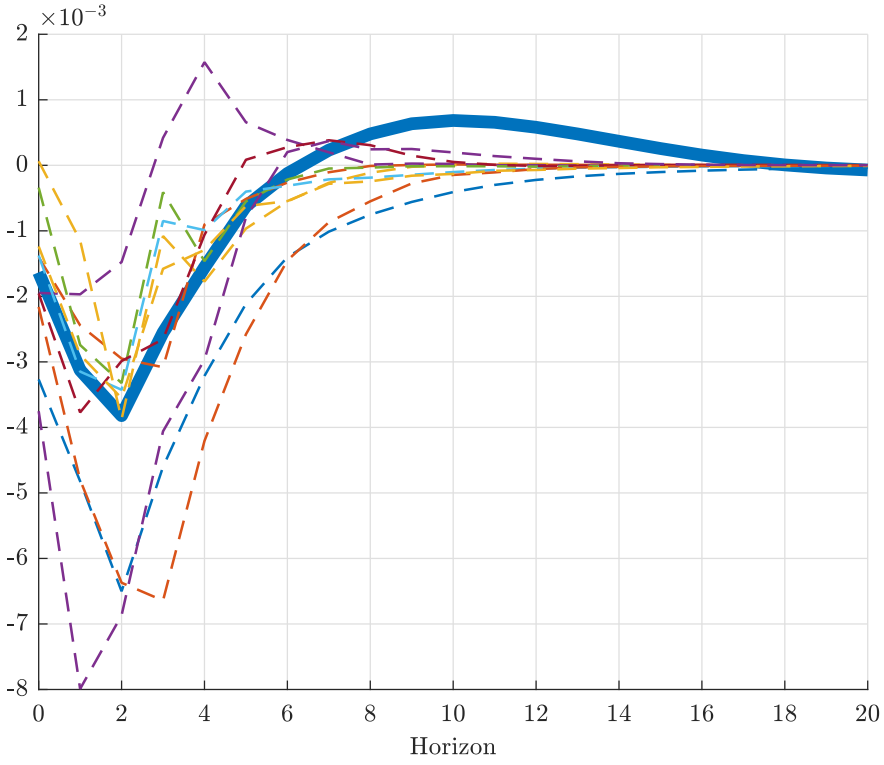

Figure D.4: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten Bayesian VAR impulse response estimates.

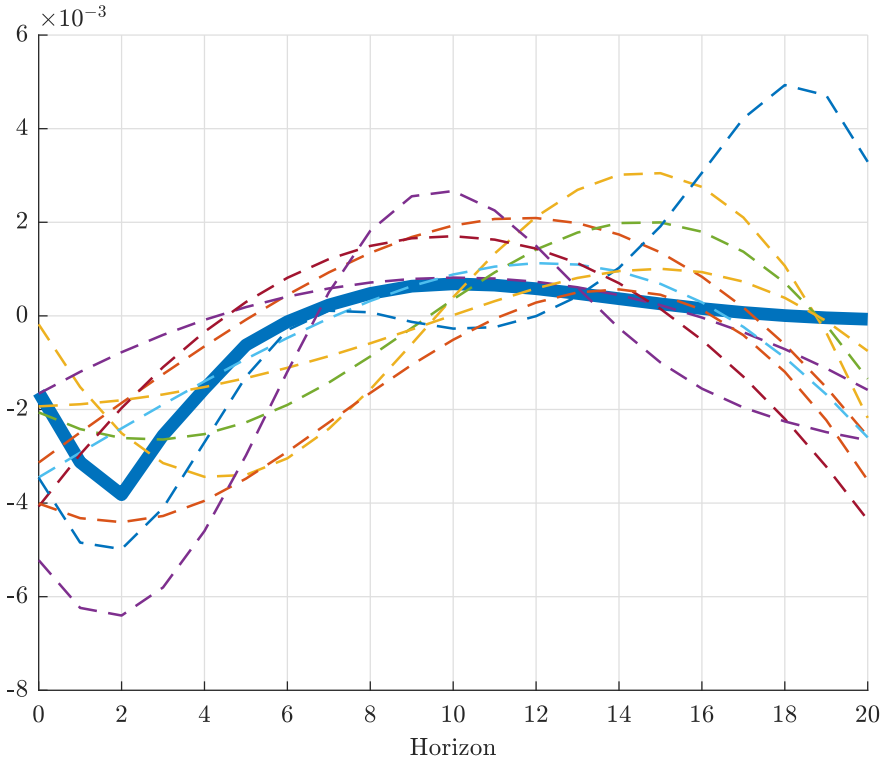

Figure D.5: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten penalized LP impulse response estimates.

Observed fiscal shock: VAR Averaging IRFs

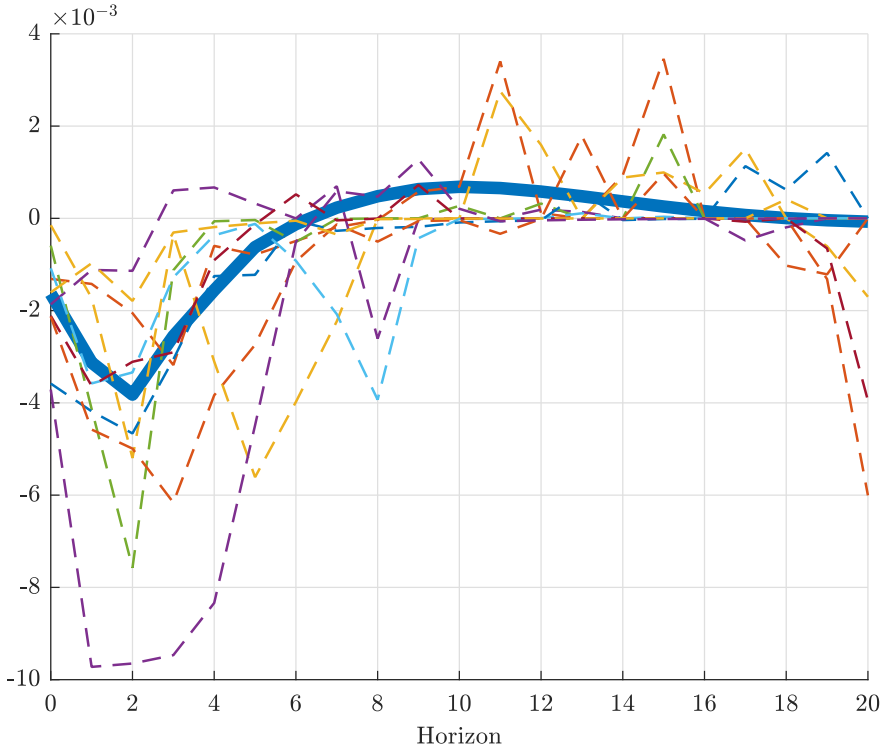

Figure D.6: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten VAR Averaging impulse response estimates.

Appendix E Further simulation results and robustness checks

E.1 IV estimators

Figures E.1 and E.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. The only significant differences concern the VAR averaging estimator, which has a particularly fat-tailed sampling distribution compared to the other estimators.

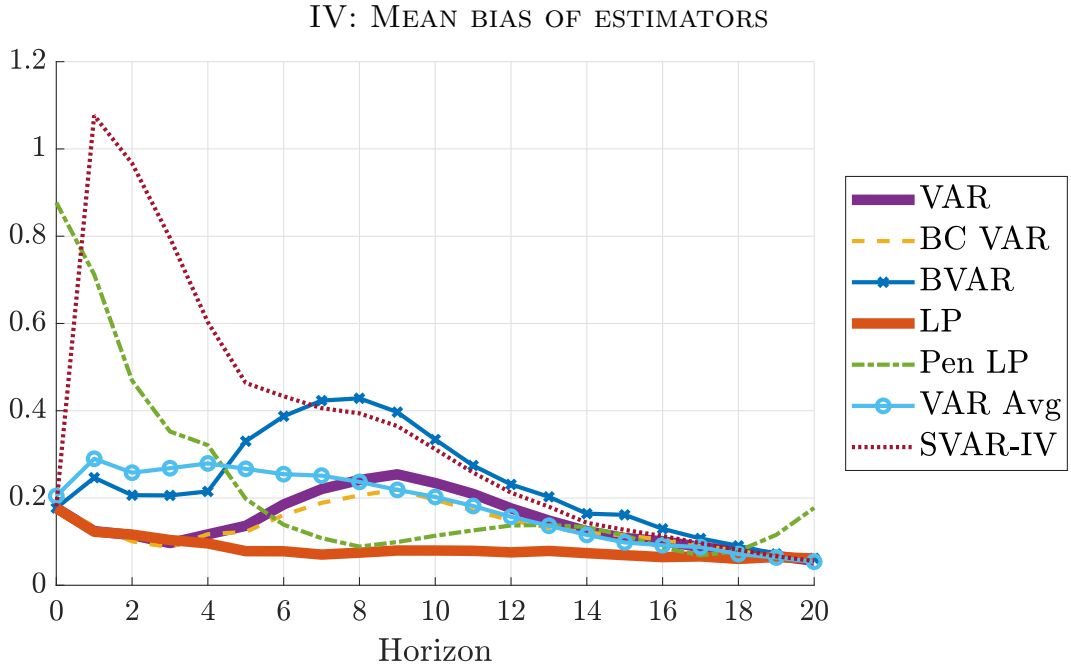

Figure E.1: Median (across DGPs) of absolute mean bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

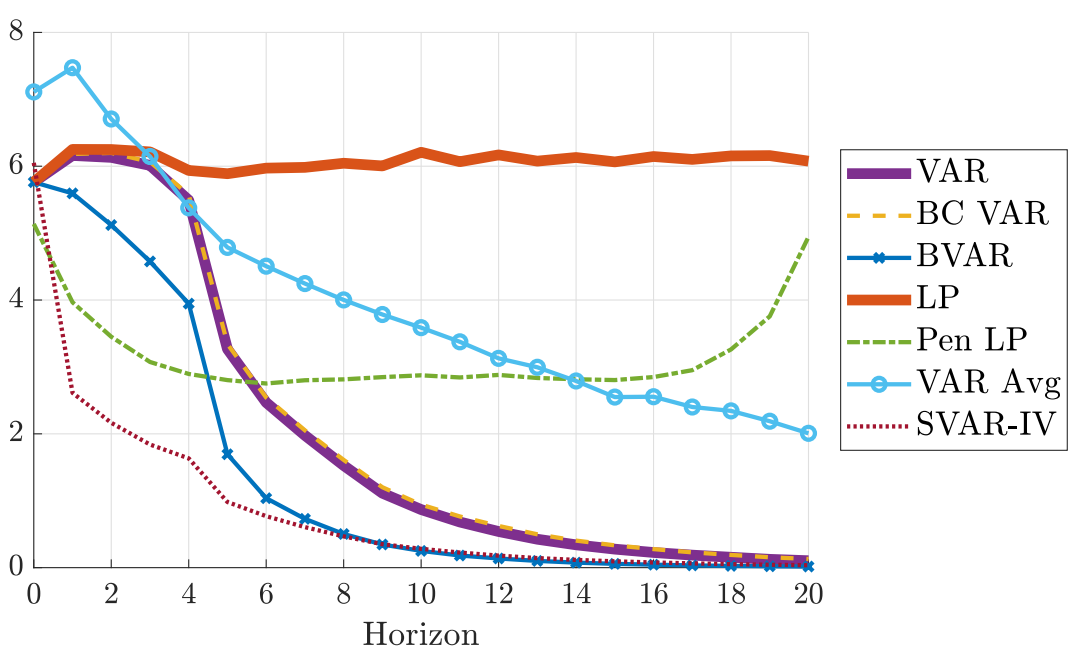

Figure E.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

E.2 More persistent DGPs

We here provide detailed results for "observed shock" identification for DGPs drawn from the alternative, more persistent factor model defined in Appendix A.3. All other settings are kept exactly as in our baseline experiments.

Figures E.3 and E.4 show the median (across DGPs) absolute bias and standard deviation of the estimation methods; the corresponding optimal method choice plot was already provided in Section 5.5. For most estimation methods, results are qualitatively similar to the baseline analysis: at intermediate and long horizons, there continues to be a non-trivial bias-variance trade-off; least-squares LP is only optimal for researchers almost exclusively concerned with bias; and the variance cost of bias reduction is steep, so shrinkage methods generally remain attractive. The two most notable differences are, first, that VAR bias reduction now makes a larger (but still moderate) difference, and second, that the bias cost of shrinking towards white noise—as done by our BVAR implementation—increases.

Can the performance of the BVAR estimator on persistent DGPs be improved by changing its prior so that it shrinks towards independent random walks rather than towards independent white noise processes? Figures E.5 and E.6 show the bias and standard deviation after this change (the other procedures are unchanged but have also been plotted for comparison). Comparing with Figures E.3 and E.4, we see that changing the BVAR prior in this simple way leads to only a very small decrease in bias at intermediate horizons. Hence, more careful and potentially case-specific prior elicitation is needed to reduce the bias of the BVAR procedure substantially.

Observed shock, persistent DGPs: Bias of estimators

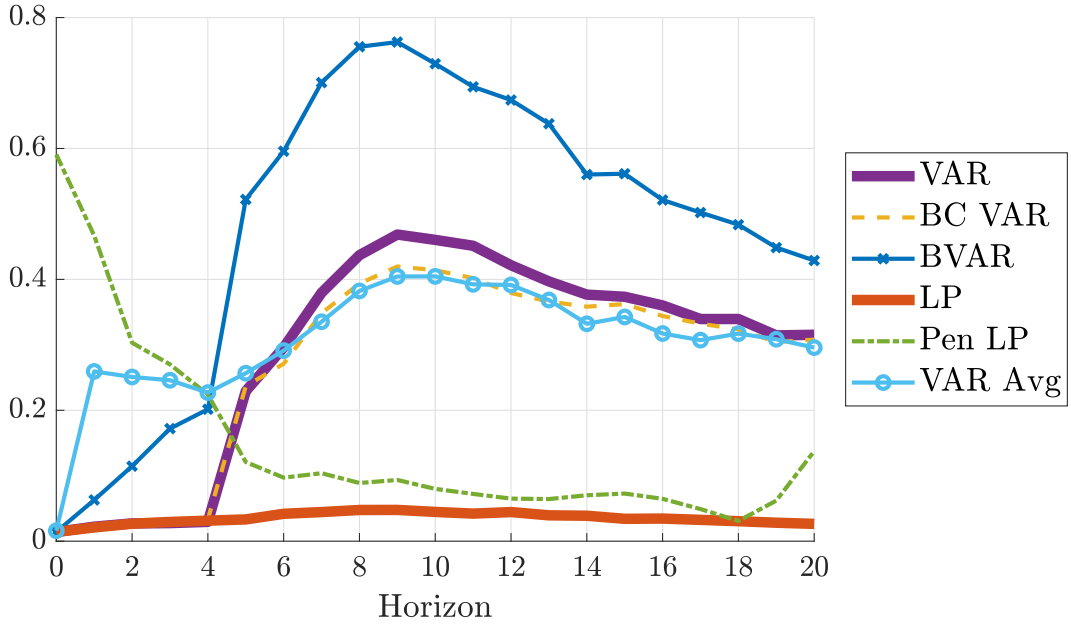

Figure E.3: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, persistent DGPs: Standard deviation of estimators

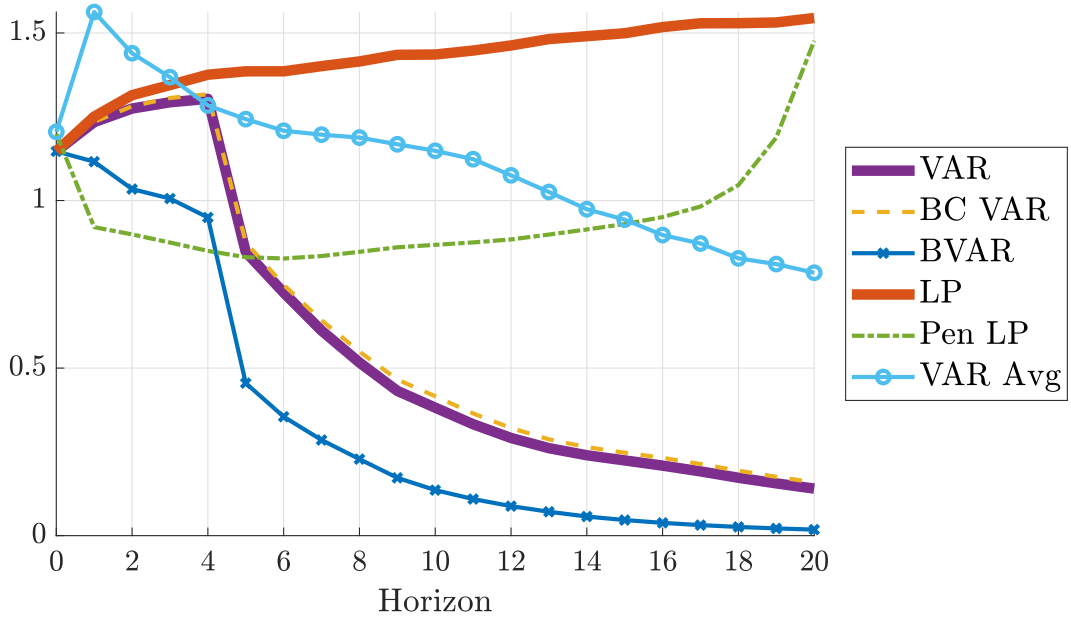

Figure E.4: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, persistent DGPs, alternative BVAR prior: Bias of estimators

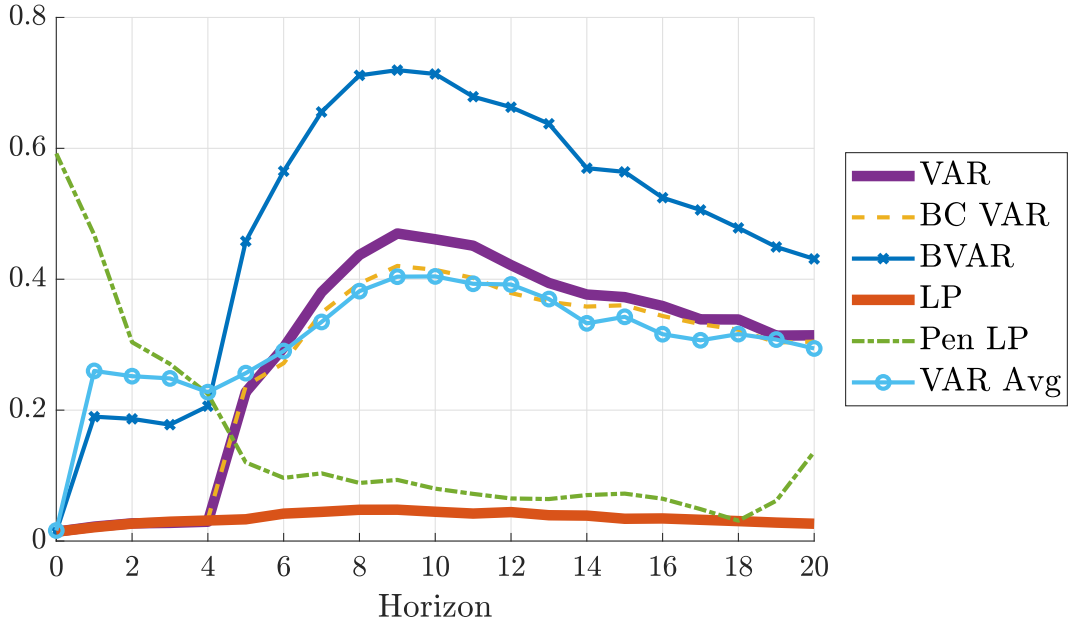

Figure E.5: Same as Figure E.3, except that the prior used for the BVAR procedure is centered at independent random walks rather than at white noise.

Observed shock, persistent DGPs, alternative BVAR prior: Standard deviation of estimators

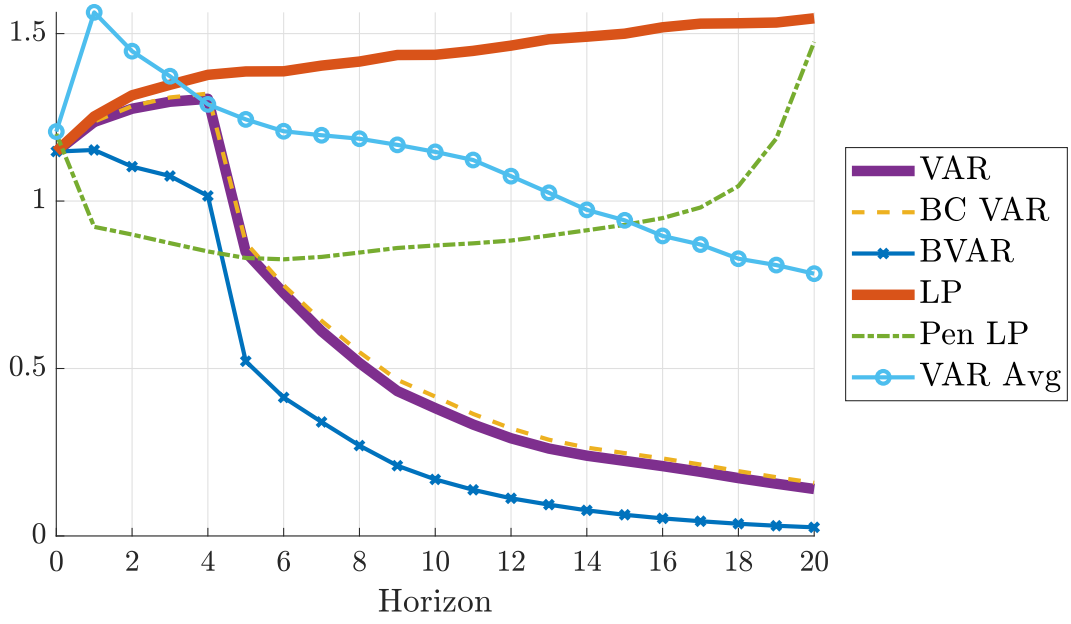

Figure E.6: Same as Figure E.4, except that the prior used for the BVAR procedure is centered at independent random walks rather than at white noise.

E.3 Cumulated impulse responses

Following Stock & Watson (2016), all variables entering the encompassing factor model have been transformed to ensure (approximate) stationarity. For example, activity series are typically in first differences (in logs), while price series are typically in second differences (in logs). However, since applied interest often centers on *level* impulse responses, we here repeat our analysis for cumulated impulse responses.[E.1] In the interest of space, we only report results for "observed shock" identification. All other settings are kept exactly as in our baseline experiments.

Figures E.7 to E.9 show the bias, standard deviation, and method choice plots for cumulated impulse responses. For LP methods, bias and standard deviation now increase with the horizon, while for VAR methods they level off with the horizon (rather than tending to zero). Importantly, however, the *relative magnitudes* of bias and standard deviation are similar to our baseline analysis, so there continues to be a bias-variance trade-off, and bias reduction through direct projection continues to incur a steep variance cost. Thus, unless concern for bias is overwhelming, researchers will still prefer to use shrinkage methods (penalized LP, BVAR) or least-squares VAR. Indeed, Figure E.9 is similar to Figure 9 (provided one groups together the areas for the quantitatively nearly indistinguishable least-squares VAR and bias-corrected VAR methods).

Note that the penalized LP and VAR model averaging procedures have been tuned to target the original un-cumulated impulse responses. That is, the cross-validation criterion or MSE criterion are formulated in terms of the differenced data. We therefore conjecture that the performance of these two estimators on cumulated impulse responses could be improved by modifying the way that tuning parameters are selected.

[E.1] Specifically, for all variables that do not enter in levels (i.e., all transformation codes in Stock & Watson (2016) except for 1 and 4), we cumulate. For twice-differenced series we only cumulate once.

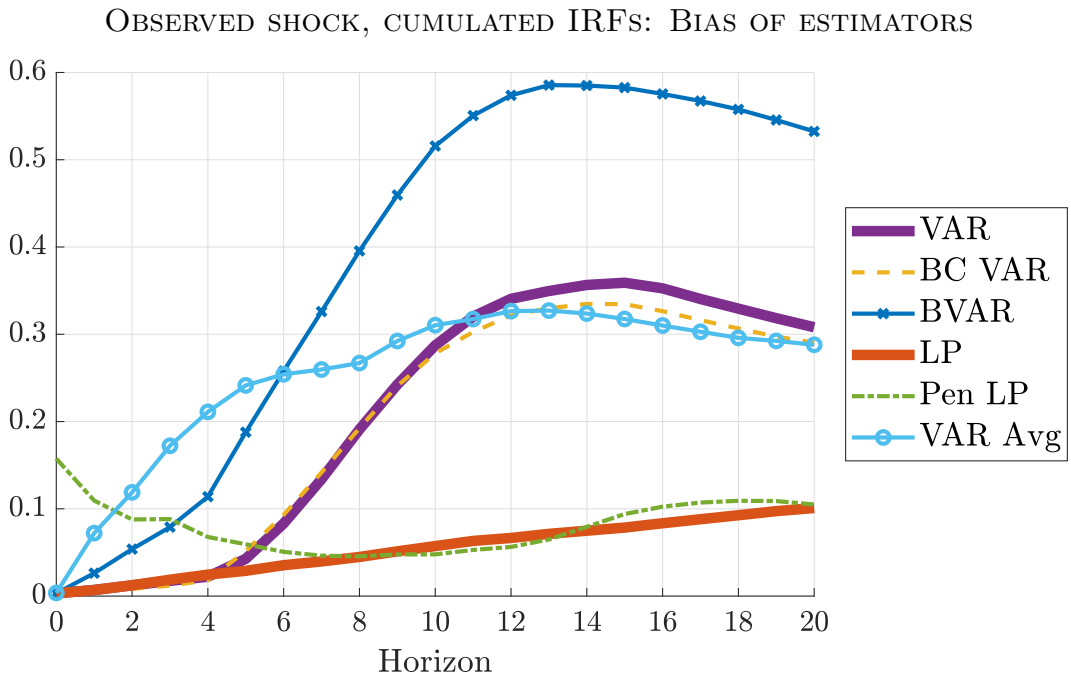

Figure E.7: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

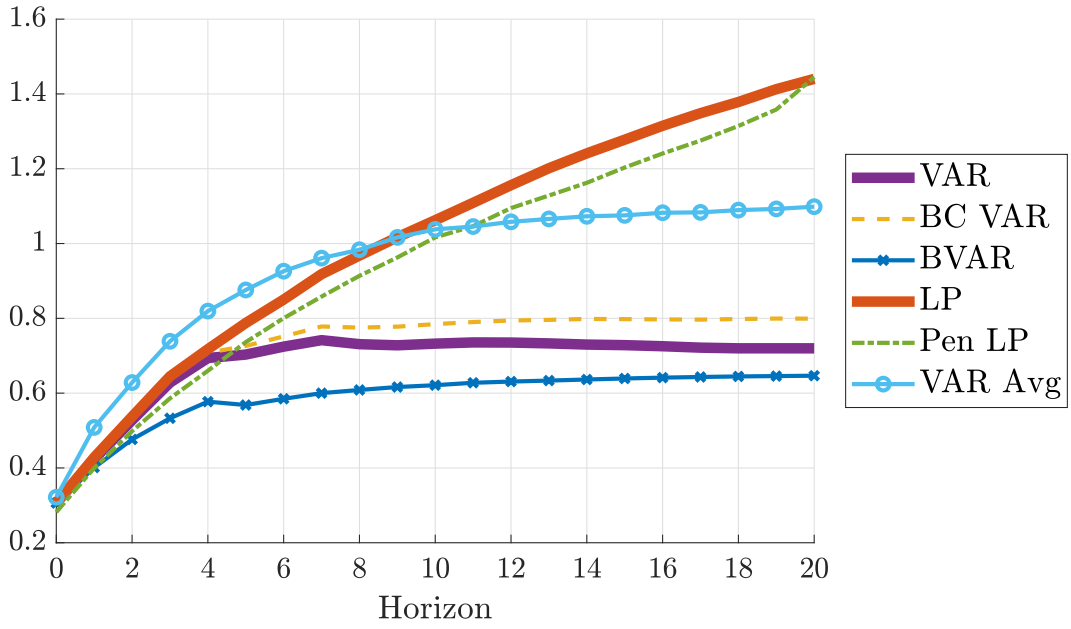

Figure E.8: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

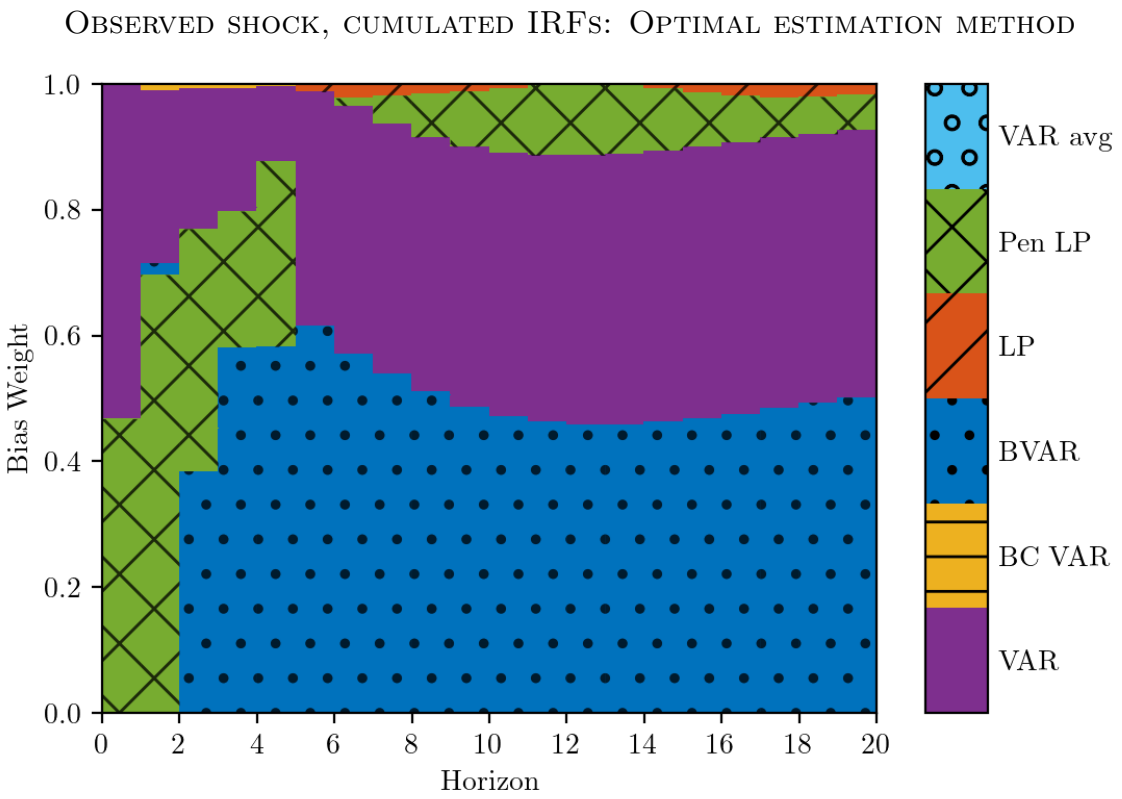

Figure E.9: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

DGP SUMMARY STATISTICS: RECURSIVE IDENTIFICATION

Percentile	min	10	25	50	75	90	max
Impulse responses up to $h = 20$							
No. of interior local extrema	0	2	2	3	3	5	8
Horizon of max abs. value	0	0	0	1	2	4	11
Average/(max abs. value)	-0.48	-0.27	-0.15	-0.04	0.09	0.22	0.52
R^2 in regression on quadratic	0.00	0.19	0.38	0.56	0.64	0.78	0.97

Table E.1: Quantiles of various population parameters across the 6,000 DGPs for recursive identification. "Average/(max abs. value)": $(\frac{1}{21}\sum_{h=0}^{20}\theta_h)/\max_h\{|\theta_h|\}$. "$R^2$ in regression on quadratic": R-squared from a regression of the impulse response function $\{\theta_h\}_{h=0}^{20}$ on a quadratic polynomial in h.

E.4 Recursive identification

Here we provide results for the recursive impulse response estimand defined in Section 3.2 and Supplemental Appendix C.

Table E.1 shows summary statistics for the impulse response functions in the recursive identification setting, analogous to the summary statistics for the "observed shock" case in the bottom half of Table 1 in Section 3.4. Note that the top half of that table is unchanged in the recursive identification setting.

Figures E.10 and E.11 show the median (across DGPs) absolute bias and standard deviation of the various estimators. Figure E.12 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.

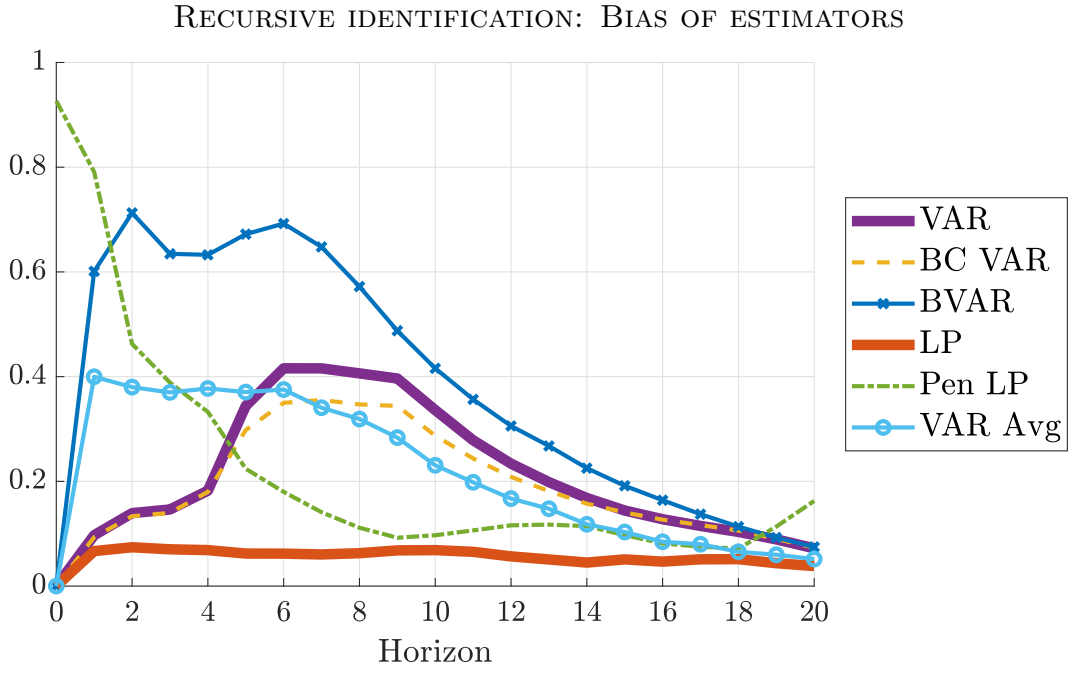

Figure E.10: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

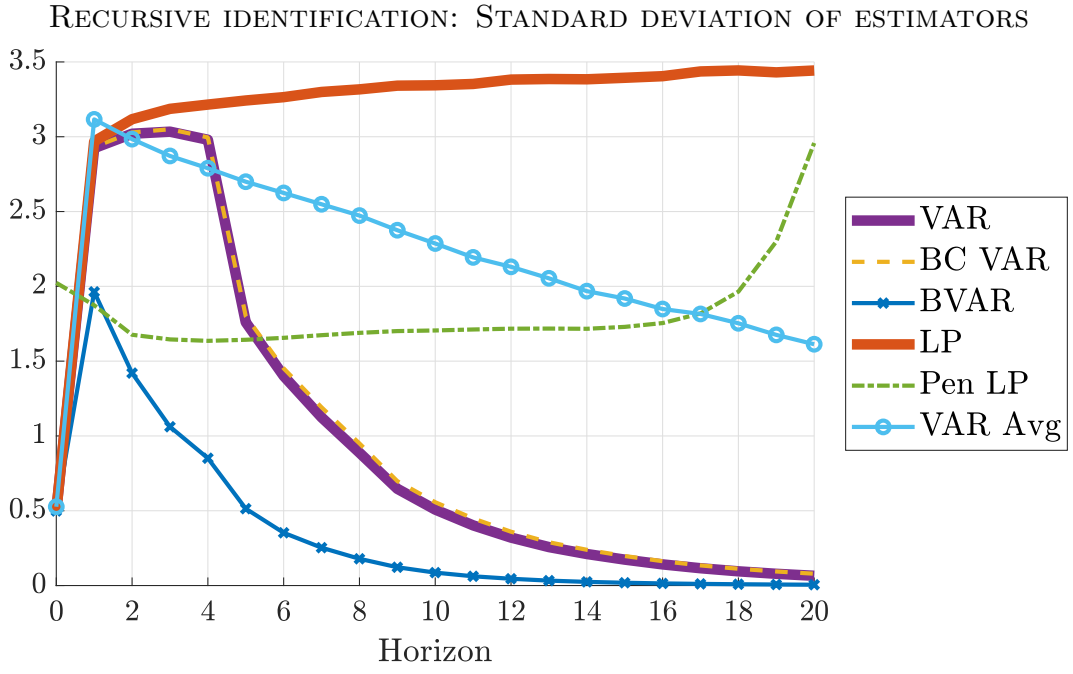

Figure E.11: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

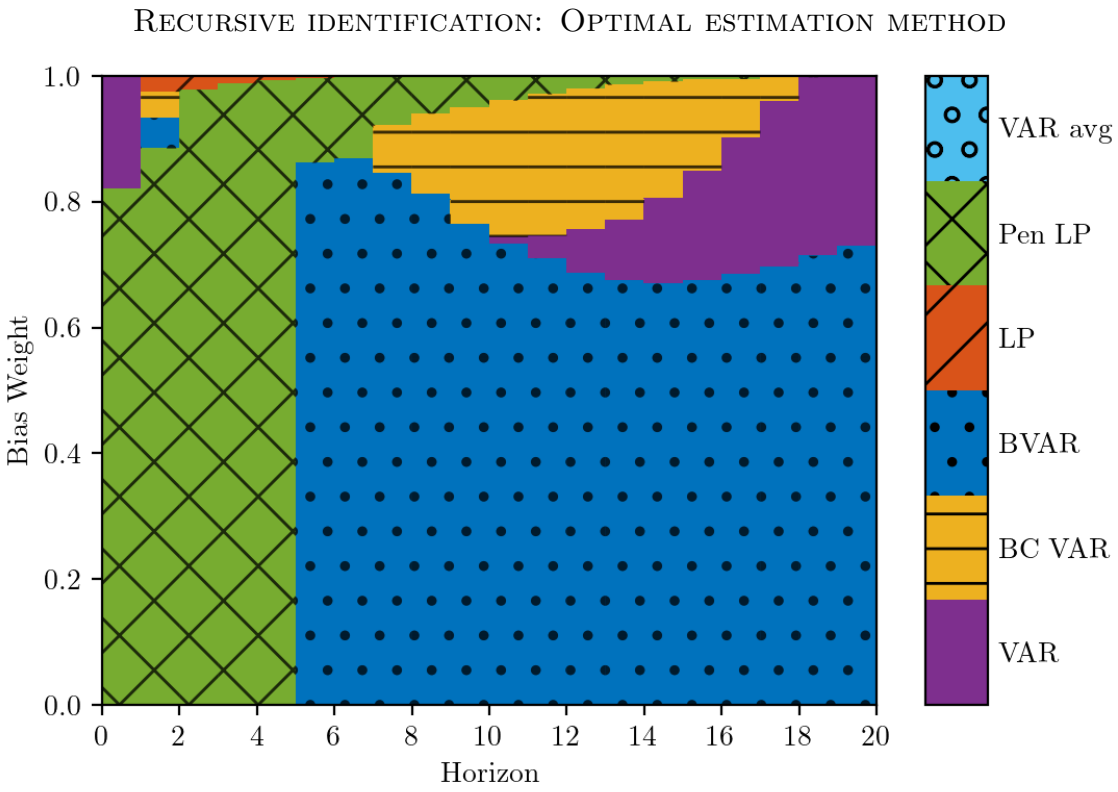

Figure E.12: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

E.5 Fiscal and monetary shocks

Recall that the results from Section 5 combine fiscal and monetary policy shock estimands. We here break the results down by shock estimand.

Figures E.13 and E.14 show the bias and standard deviation plots for the 3,000 fiscal shock DGPs, while Figures E.15 and E.16 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.

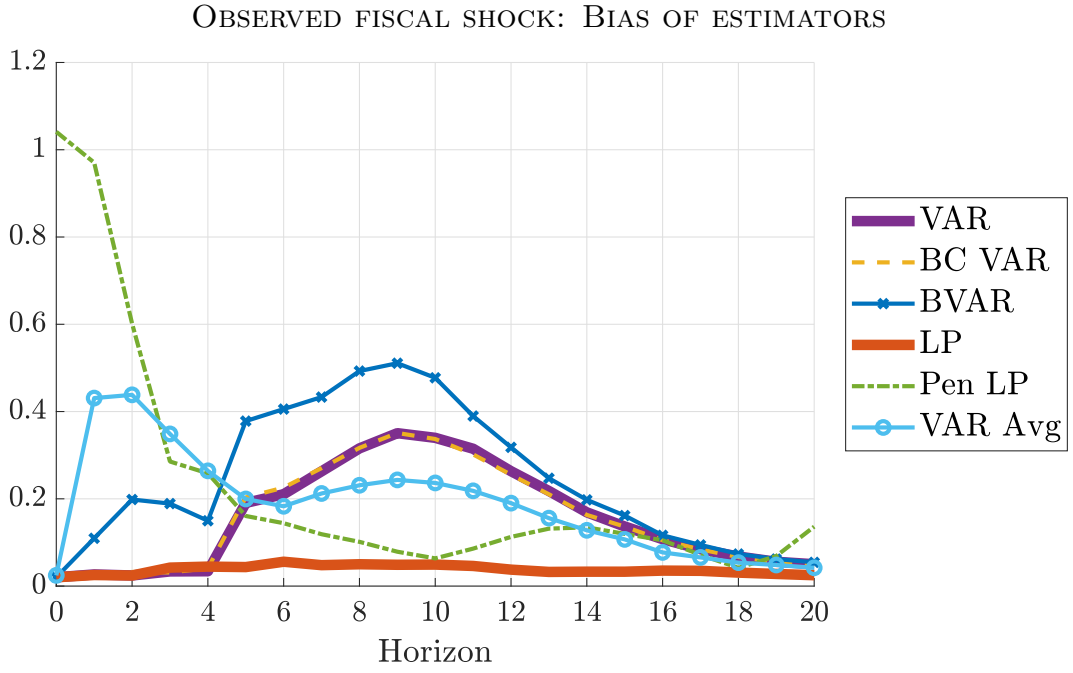

Figure E.13: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

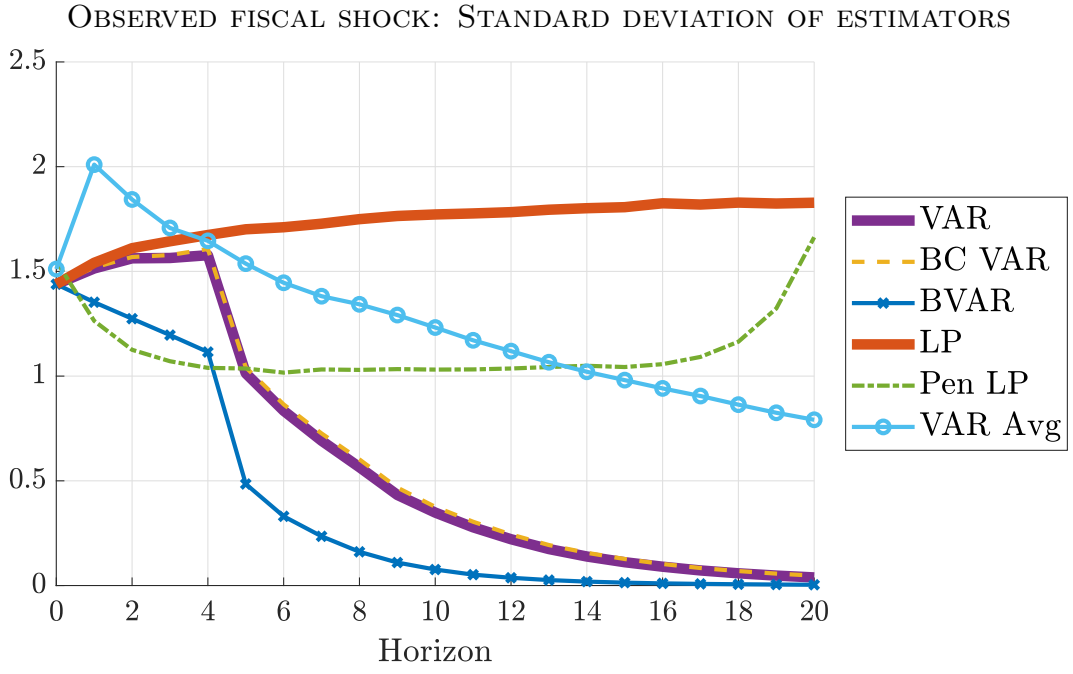

Figure E.14: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

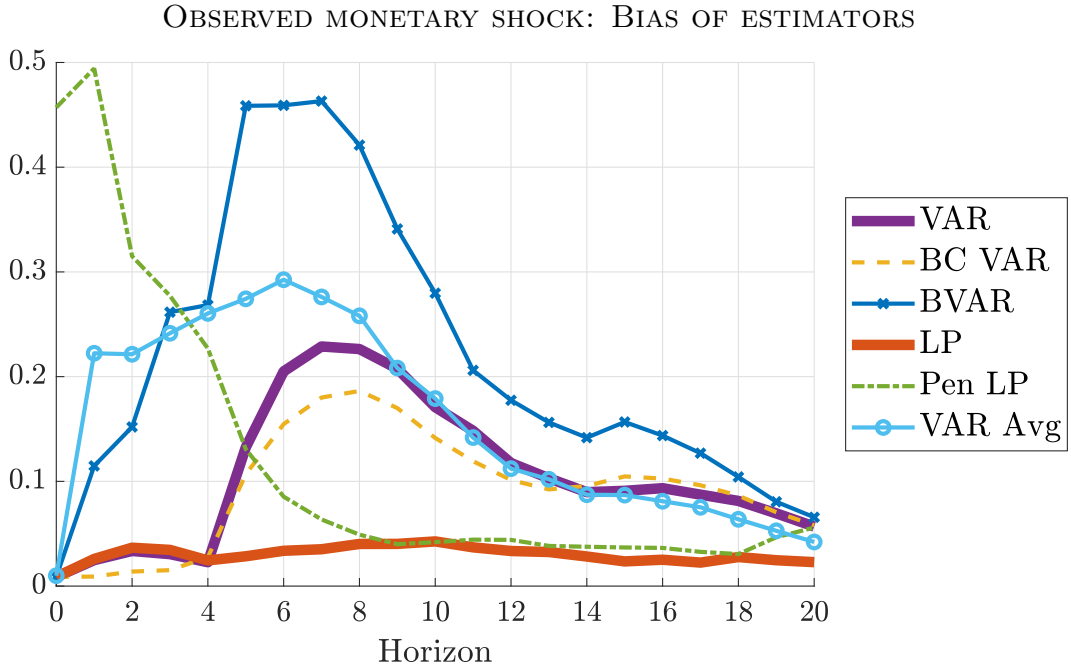

Figure E.15: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

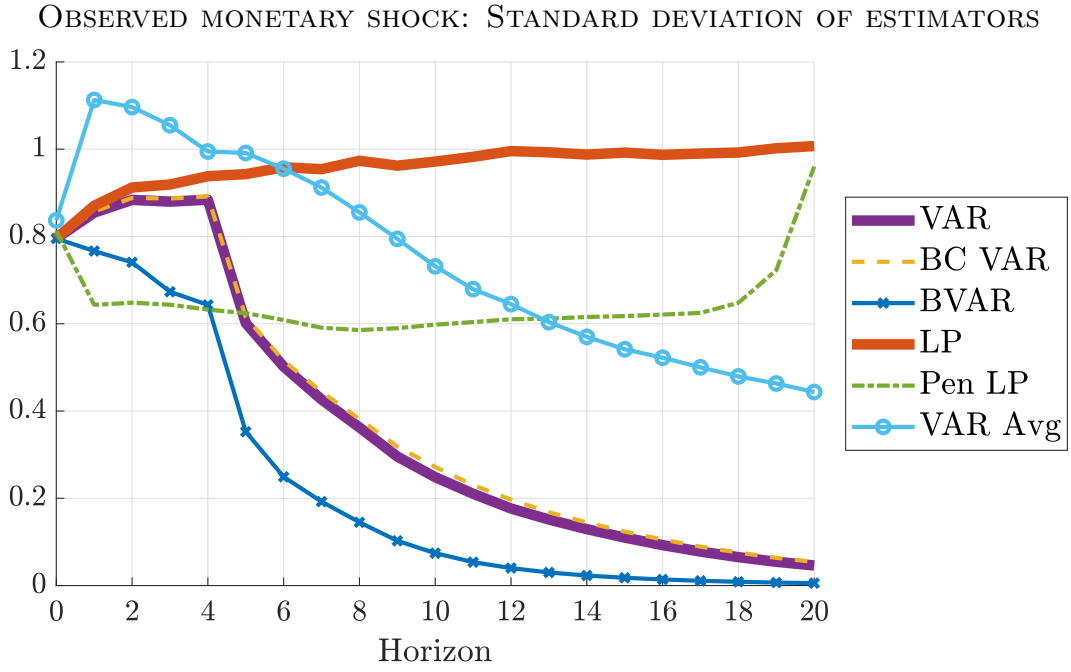

Figure E.16: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

E.6 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 E.17 and E.18 show the median (across DGPs) absolute bias and standard deviation of the estimation methods, while Figure E.19 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.

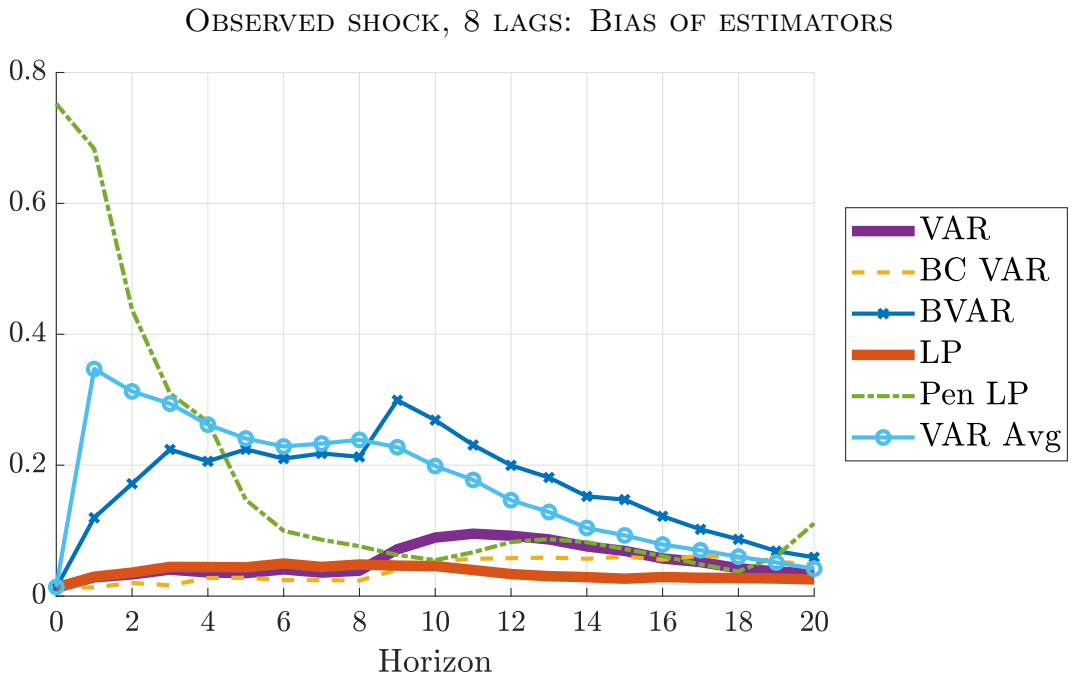

Figure E.17: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

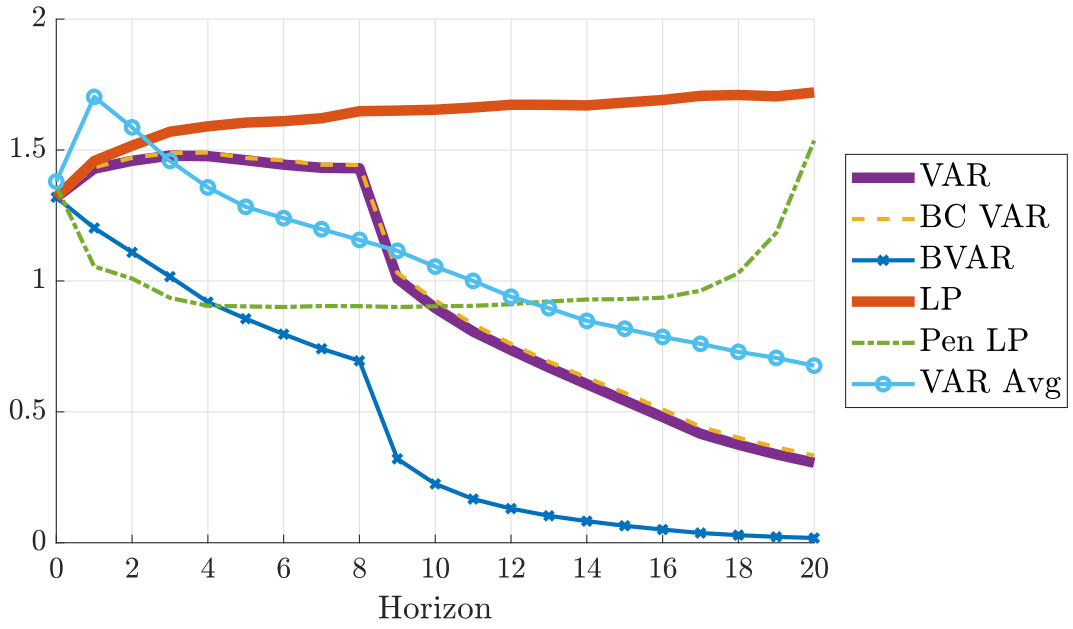

Figure E.18: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

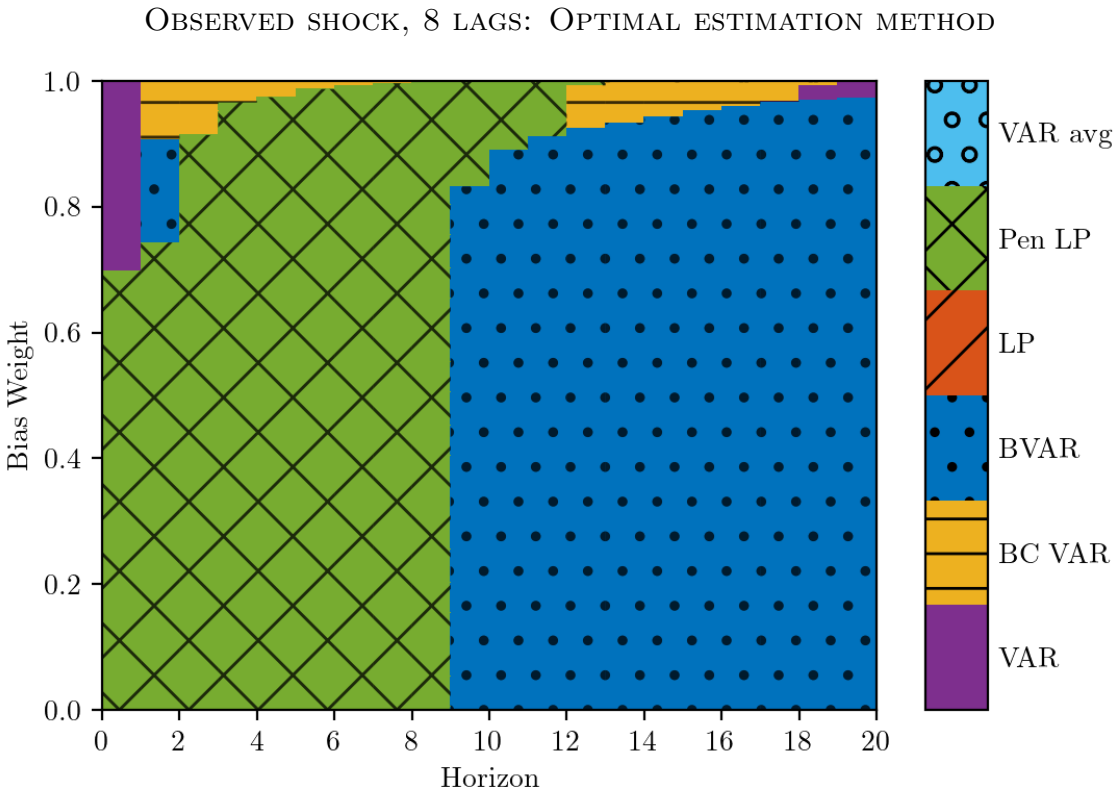

Figure E.19: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

E.7 Smaller sample size

Recall that our baseline experiments consider the sample size $T = 200$ quarters. We here instead present results for $T = 100$. In the interest of space, we focus on results for "observed shock" identification.

Results for bias, standard deviation, and optimal method choice are displayed in Figures E.20 to E.22, respectively. The figures are qualitatively similar to those for our baseline analysis, though there is a quantitative difference: reducing the sample size increases standard deviation by more (in relative terms) than bias. As a result, the method choice plot indicates an even more pronounced preference for shrinkage, with now almost the entire figure either cross-hatched green (penalized LP) or solid-dotted blue (BVAR).

Observed shock, small sample: Bias of estimators

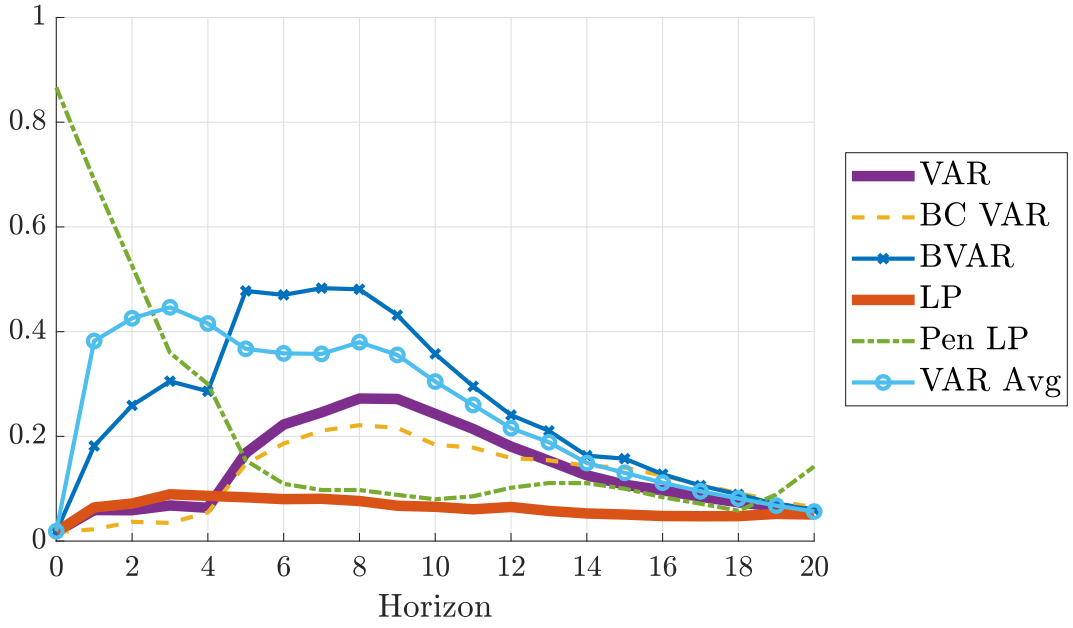

Figure E.20: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, small sample: Standard deviation of estimators

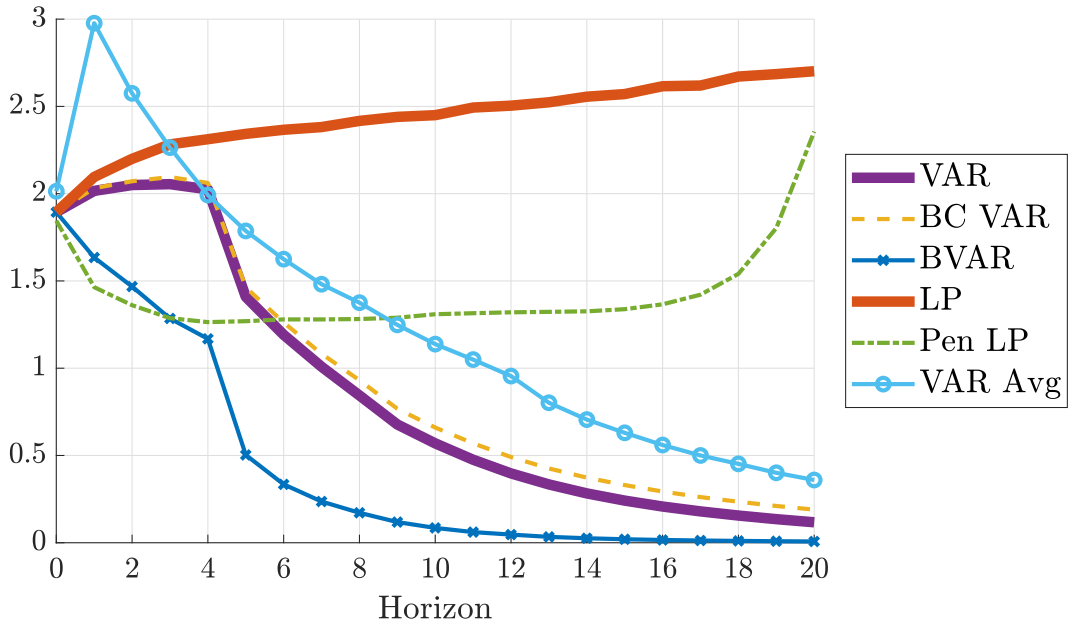

Figure E.21: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

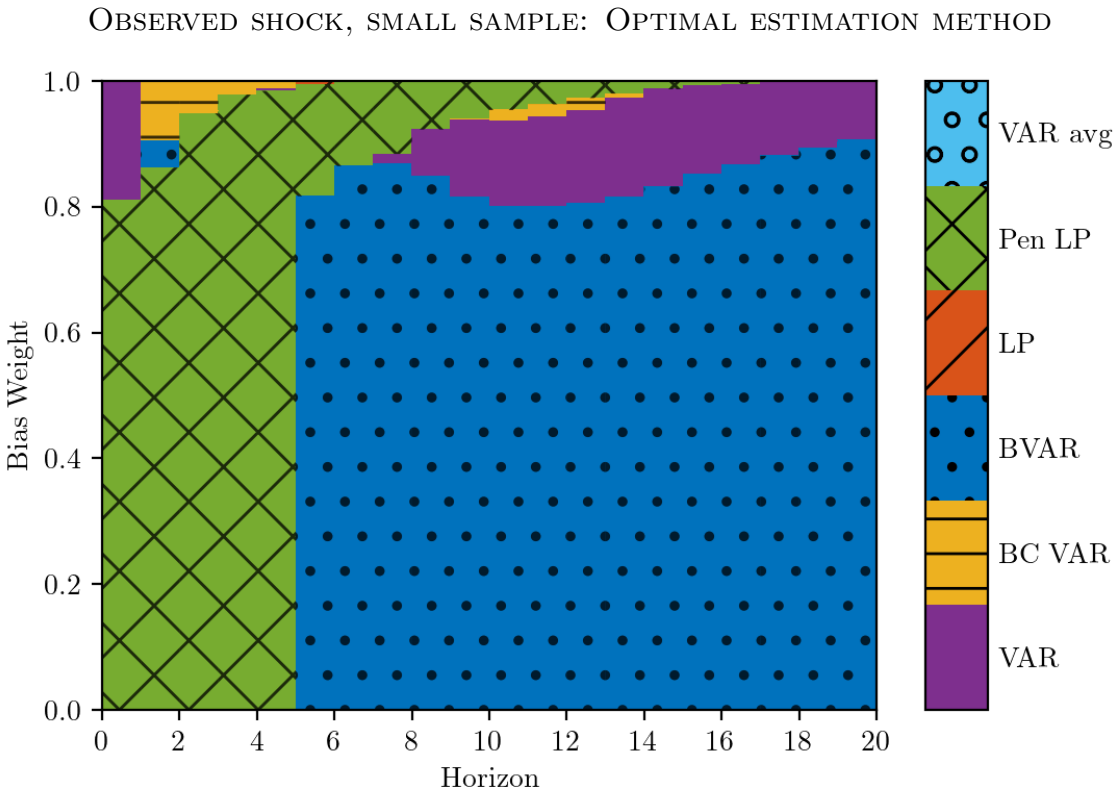

Figure E.22: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

E.8 Splitting by variable categories

Here we show that the categories of variables included in the DGP are not highly predictive of the bias or standard deviation of any estimator. We group the 207 variables in the Stock & Watson (2016) data set into 11 categories, proceeding exactly as in the Data Appendix to that paper, except that we lump together their final three categories in a single "Asset Price & Sentiment" category. Let $\Delta_{j,d}$ denote the total number of variables in category j that are included in DGP d. We are interested in relating the absolute bias and standard deviation of the various estimation methods with $\Delta_{j,d}$, across DGPs and horizons.

Using the baseline simulation results from Sections 5.1 to 5.3, we run the following OLS regression separately for each estimation method m, pooling observations across all DGPs and all horizons:

$$\log \text{bias}_{m,d,h} = \sum_{j=1}^{10} \hat{\gamma}_j \Delta_{j,d} + \sum_{i=0}^{20} \hat{\iota}_i \mathbb{1}(i = h) + \hat{e}_{m,d,h}, \quad d = 1, \ldots, 6000, \; h = 0, 1, \ldots, 20.$$

Here $\text{bias}_{m,d,h}$ is the absolute bias of estimator m in DGP d at horizon h, $\hat{\gamma}_j$ are coefficients on the category counts, $\hat{\iota}_i$ are coefficients on indicator variables for each horizon, and $\hat{e}_{m,d,h}$ is the OLS residual. Notice that we omit category 11 from the regression for reasons of multicollinearity (there are 5 variables in each DGP). We also run the above regression with the log standard deviation on the left-hand side in place of the log absolute bias.

Table E.2 shows the coefficients $\hat{\gamma}_j$ on the category counts. A coefficient of 0.1, say, means that adding one variable of that category to the DGP (while removing a variable from the omitted "Asset Price & Sentiment" category) leads to a 10% higher absolute bias (resp., standard deviation). Few coefficients in the table exceed 0.2 in absolute value, indicating that none of the categories are highly predictive of the bias or standard deviation of the resulting impulse response estimates. The only exception is that adding variables from the "Interest Rates" category does seem to moderately lower both the bias and the standard deviation of most estimation procedures, which is consistent with the results reported in Supplemental Appendix E.5. Though the regression results reported in Table E.2 pool across all horizons $h \in [0, 20]$, we obtain similar coefficients if we instead restrict the regressions only to the intermediate horizons $h \in [5, 12]$.

We conclude that our baseline results do not conceal significant across-category heterogeneity. In particular, and differently from Marcellino et al. (2006), we do not find significant differences when comparing price and real activity series.

Bias regression

	VAR	VAR BC	BVAR	LP	Pen LP	VAR Avg
NIPA	0.05	0.10	0.08	-0.02	0.12	0.08
Industrial Production	0.00	0.02	0.10	-0.13	0.02	0.09
Employment & Unemployment	0.04	0.07	0.13	-0.03	0.03	0.08
Orders, Inventories & Sales	0.02	0.01	0.16	-0.14	-0.01	0.11
Housing Starts & Permits	-0.02	-0.04	0.09	-0.07	0.02	0.01
Prices	0.00	0.05	-0.05	-0.08	-0.03	-0.03
Productivity & Earnings	-0.04	-0.04	-0.10	-0.04	-0.06	-0.03
Interest Rates	-0.18	**-0.25**	0.03	-0.17	**-0.30**	-0.06
Money & Credit	0.09	0.06	0.08	0.04	0.06	0.12
International	-0.03	-0.02	-0.02	-0.08	-0.07	-0.05

Standard deviation regression

	VAR	VAR BC	BVAR	LP	Pen LP	VAR Avg
NIPA	0.05	0.03	-0.10	0.10	0.10	0.09
Industrial Production	0.02	-0.01	-0.11	-0.07	-0.03	-0.06
Employment & Unemployment	0.04	0.03	0.03	-0.04	0.00	-0.02
Orders, Inventories & Sales	0.00	-0.03	-0.13	-0.14	-0.08	-0.12
Housing Starts & Permits	-0.03	-0.06	**-0.31**	-0.08	-0.09	-0.08
Prices	-0.04	-0.06	-0.15	0.12	0.04	0.11
Productivity & Earnings	-0.04	-0.05	-0.09	0.00	0.02	-0.01
Interest Rates	-0.14	-0.16	**-0.21**	**-0.34**	**-0.29**	**-0.31**
Money & Credit	0.07	0.08	**0.24**	-0.04	0.02	-0.03
International	-0.10	-0.10	**-0.25**	-0.02	-0.01	-0.02

Table E.2: Coefficients from OLS regressions of log bias (top table) or log standard deviation (bottom table) on variable category counts (along rows), controlling for horizon fixed effects. Regressions are run separately by estimation method (along columns), and observations are pooled across DGPs and horizons. Coefficients greater than 0.2 in absolute value highlighted in bold. Bias and standard deviation normalized as in Figures 4 and 5. Observed shock identification.

E.9 Salient observables

We here re-run our analysis on a restricted set of DGPs that use a smaller subset of particularly salient time series. Whereas our baseline analysis randomly draws series from the large set of 207 variables included in the empirical DFM of Stock & Watson (2016), we now consider the exhaustive list of all possible combinations of 18 oft-used series.

Our subset of salient series includes (Stock & Watson Data Appendix series # in brackets): *real GDP (1); real consumption (2); real investment (6); real government expenditure (12); the unemployment rate (56); personal consumption expenditure prices (95); the GDP deflator (97); the consumer price index (121); average hourly earnings (132); the federal funds rate (142); the 10-year Treasury rate (147); the BAA 10-year spread (151); an index of the U.S. dollar exchange rate relative to other major currencies (172); the S&P 500 (181); a real house price index (193); consumer expectations (196); and real oil prices (202).* As in our baseline analysis, we force each GDP to include either the federal funds rate or government spending (for monetary or fiscal shock estimands, respectively) as well as at least one real activity series (categories 1–3) and one price series (category 6). Subject to these constraints, we then generate the exhaustive list of all five-variable combinations of the salient series. This yields a total of 1581 DGPs.

Results for bias, standard deviation, and optimal method choice are reported in Figures E.23 to E.25. The figures look extremely similar to those from our baseline analysis. We conclude that there are no systematic differences between the larger set of variables included in the full DFM and our smaller subset of particularly salient macroeconomic time series.

Observed shock, salient observables: Bias of estimators

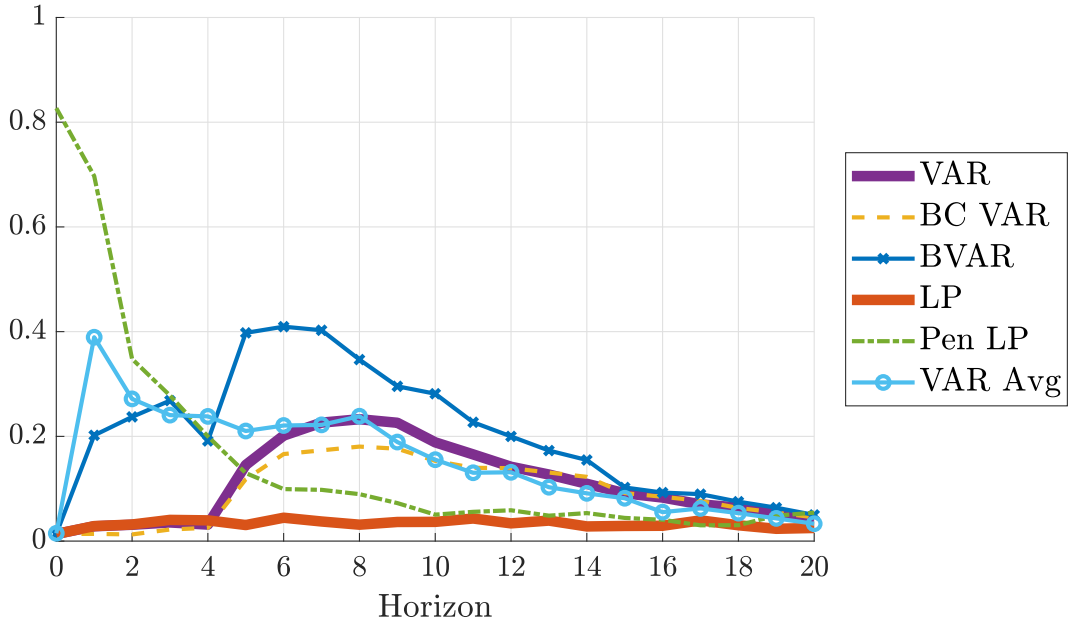

Figure E.23: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, salient observables: Standard deviation of estimators

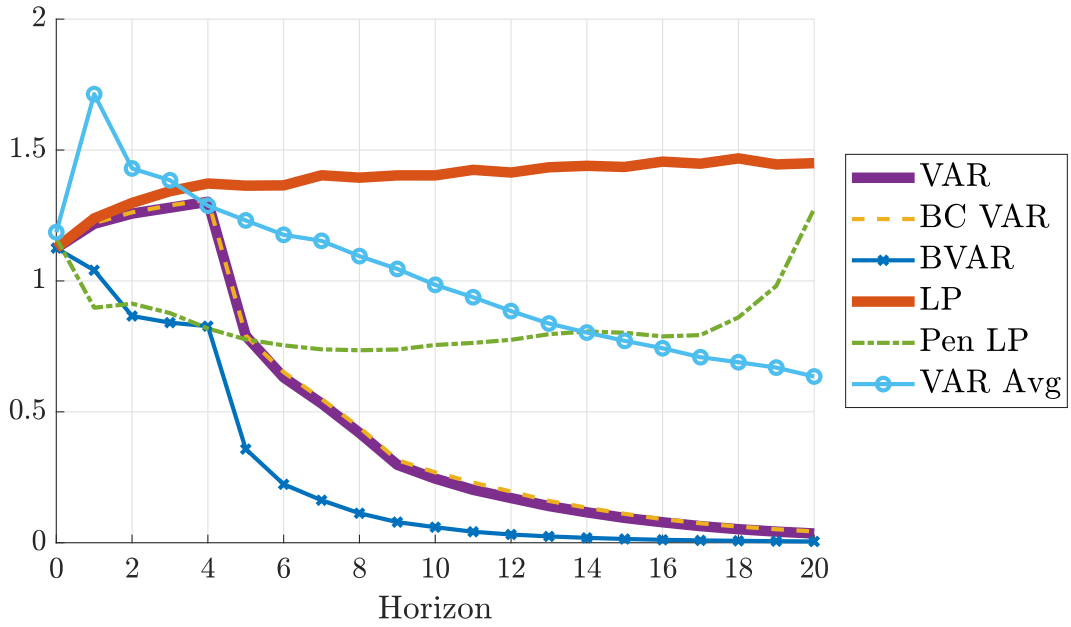

Figure E.24: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

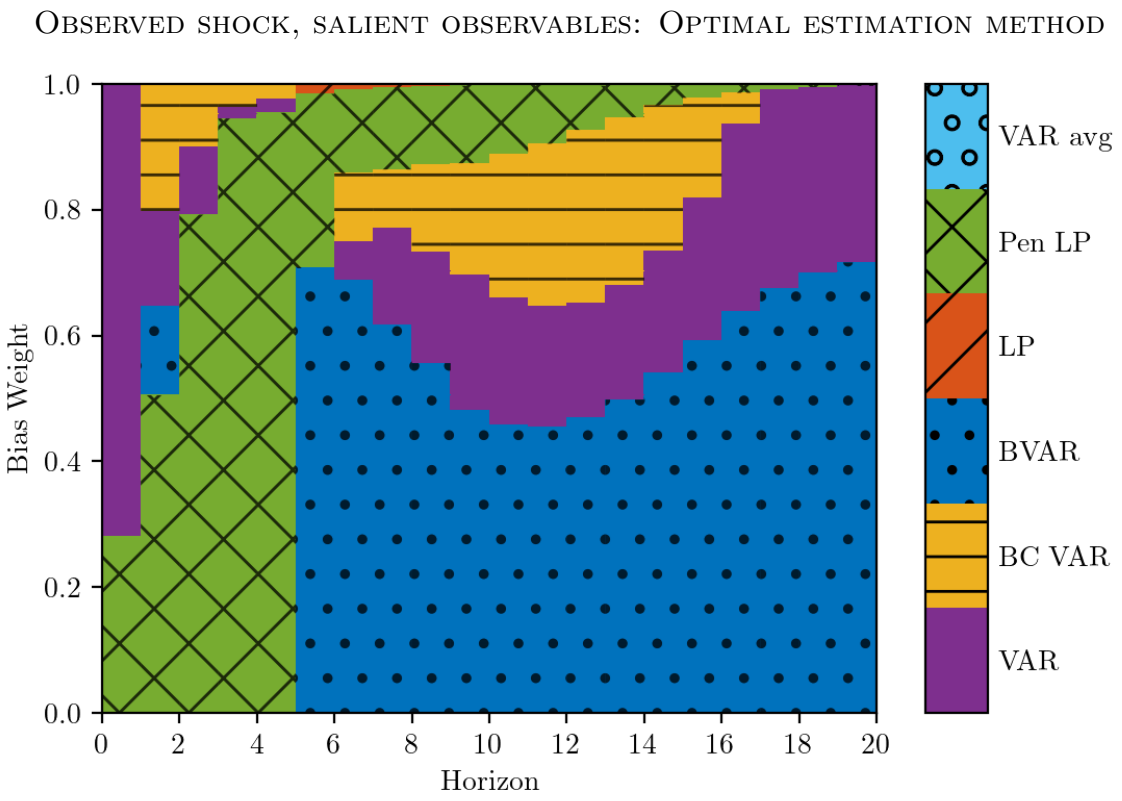

Figure E.25: Method that minimizes the average (across DGPs) loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

E.10 90th percentile loss

Whereas our baseline results report medians (for bias and standard deviation) across DGPs, we now report the 90th percentiles across DGPs. This places the spotlight on those DGPs that are particularly challenging for impulse response estimation.

Figures for bias, standard deviation, and optimal method choice are displayed in Figures E.26 to E.28. By construction, bias and standard deviation are now higher for all methods. Importantly, however, *relative* magnitudes do not change by much; that is, for DGPs in which VARs or shrinkage techniques do poorly, least-squares LP tends to do just as poorly (relative to their respective median performance). As a result, optimal method choice for a researcher that evaluates loss at the 90th percentile looks similar to our baseline.

Observed shock, 90th percentile of loss: Bias of estimators

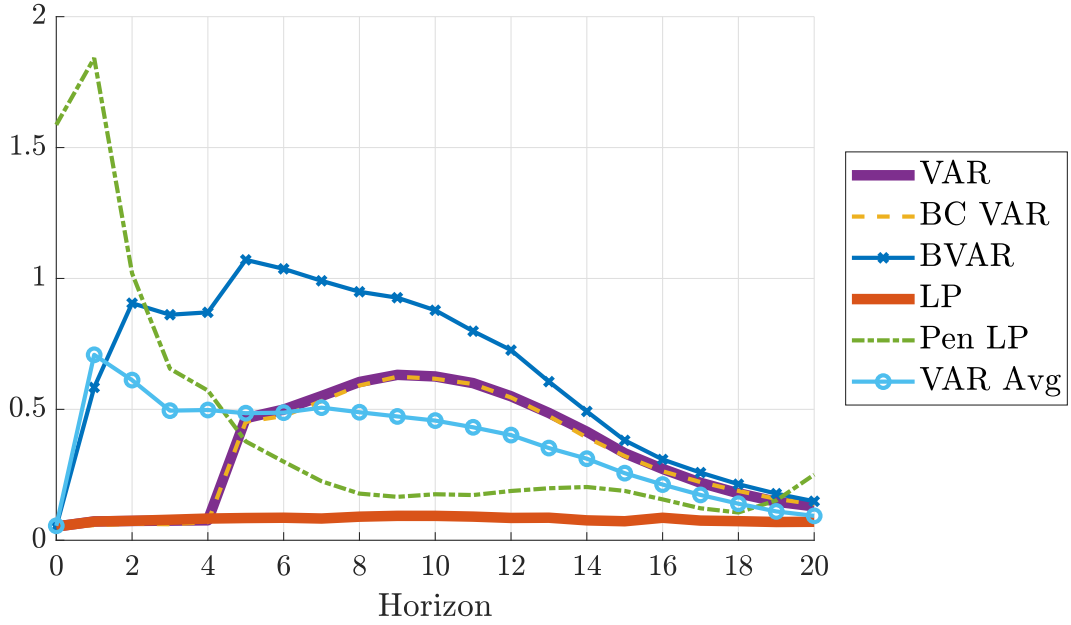

Figure E.26: 90th percentile (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, 90th percentile of loss: Standard deviation of estimators

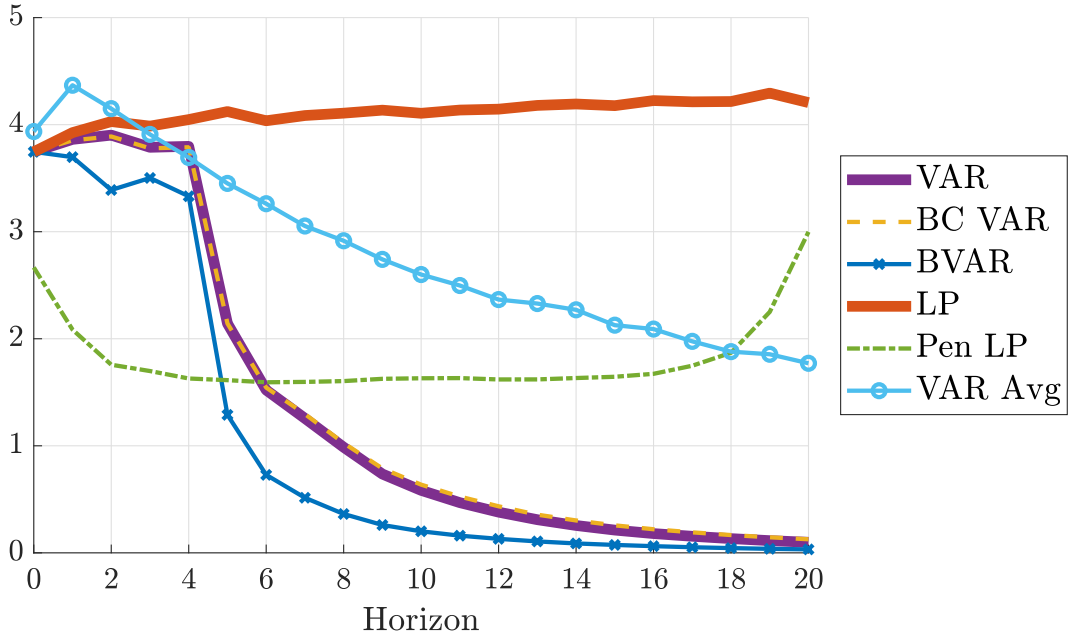

Figure E.27: 90th percentile (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

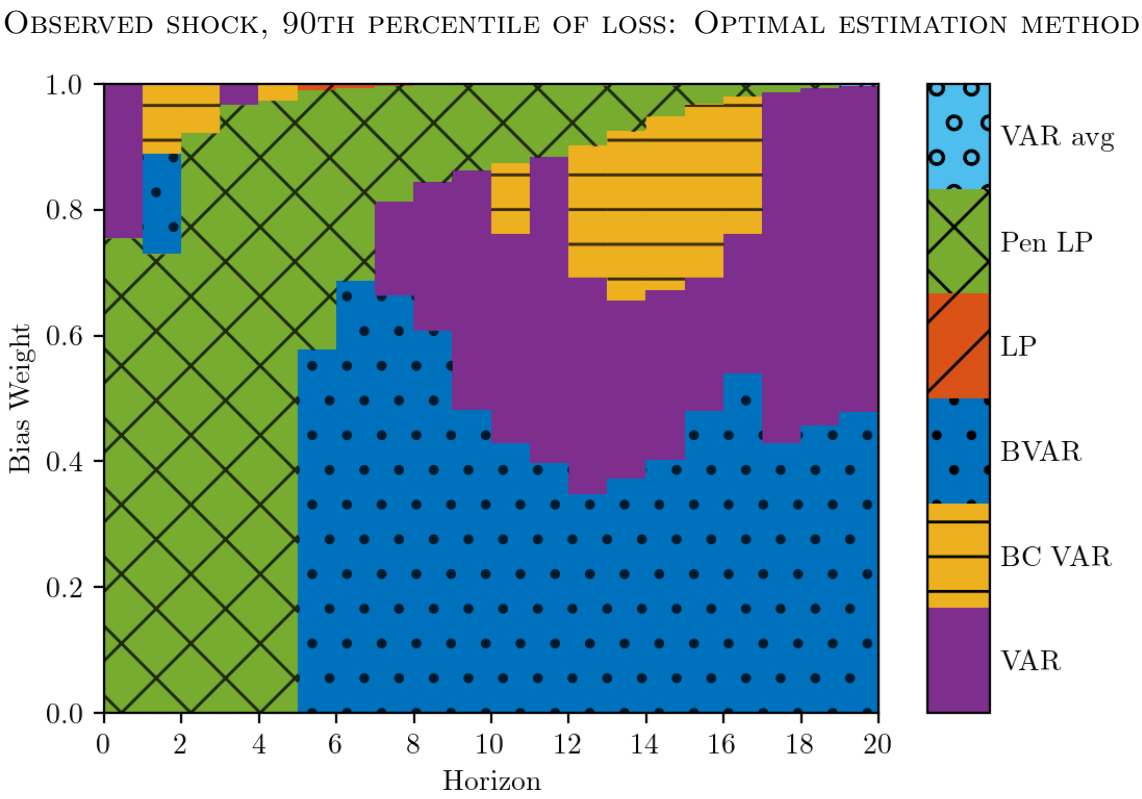

Figure E.28: Method that minimizes the 90th percentile (across DGPs) of the loss function (2). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 4 and 5. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

Appendix F Proofs

F.1 Auxiliary lemmas

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

Lemma F.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 F.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 F.2. *We have*

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

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 F.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

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

Lemma F.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 (F.2), which we decompose as

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

We first show that the first term above equals the sum on the right-hand side of (F.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 (F.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 F.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. □

Lemma F.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 F.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 F.2 applied to $h = 0$. □

F.2 Proof of Proposition 1

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

LP. It follows from Lemma F.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

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

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 $\text{vec}(A)$ and κ. According to Lemma F.3, the Jacobians should be evaluated at $\text{plim}\,\hat{A}=A_0\equiv\left(\begin{smallmatrix}0&0\\0&\rho\end{smallmatrix}\right)$ and $\text{plim}\,\hat{\kappa}=1$. Thus, γ should be evaluated at $\gamma_0\equiv(1,1)'$.

First, for $h\geq 2$,

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

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 $h \geq 1$,

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

Next, Lemma F.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 F.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

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

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

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

F.3 Asymptotic power of t-statistic

Here we prove the following result, which was discussed in Footnote 7.

Proposition F.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 F.1. Straight-forward calculations employing Lemma F.1 imply that

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

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 F.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,

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

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

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

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

Blanchard, O. & Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics*, *117*(4), 1329–1368.

Christiano, L., Eichenbaum, M., & Evans, C. (1999). Monetary Policy Shocks: What Have We Learned and to What End? In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics*, volume 1A chapter 2, (pp. 65–148). Elsevier.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., & Watson, M. W. (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review*, *97*(3), 1021–1026.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.

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.

Stock, J. H. & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, volume 2 chapter 8, (pp. 415–525). Elsevier.