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

Dake Li Mikkel Plagborg-Møller Christian K. Wolf

September 26, 2023

Contents

Appendix C Estimation of non-stationary DFM

We here elaborate on our estimation of the non-stationary DFM, complementing the brief discussion in Section 3.3.

Data transformation. We use the same dataset as Stock & Watson (2016) but, unlike those authors, do not transform the series to stationarity prior to estimating the DFM. This means that we change Stock & Watson's transformation codes 2 (first differences), 3 (second differences), 5 (first log differences), and 6 (second log differences) to 1 (levels), 2, 4 (log levels), and 5, respectively. We then correct outliers using the same procedure as Stock & Watson, meaning that when the outlier adjustment procedure is applied to a series in levels, we difference the series, then adjust outliers, and then finally cumulate the series again. There are 8 series who have outliers adjusted, all of which are in levels. Finally, and unlike Stock & Watson, we refrain from subtracting a nonparametric trend estimate from the series. Instead, we control for a linear time trend in the estimation, as discussed below.

Estimation. We set the number of factors n_f equal to 6 as in Stock & Watson (2016), since these authors selected the number of factors based on information criteria applied to ΔX_t (essentially), and this also remains a valid way to select n_f in our non-stationary model. We then apply Stock & Watson's PCA procedure for unbalanced panels to the differenced data ΔX_t.[C.1] This gives us estimates $\Delta\hat{f}_t$ of the differenced factors (up to rotation), and we then cumulate $\hat{f}_t = \sum_{s=1}^{t} \Delta\hat{f}_s$. Asymptotically, $\hat{f}_t$ equals f_t up to rotation and a linear time trend (Barigozzi et al., 2021). We then estimate the loadings $\Lambda = (\lambda_1, \ldots, \lambda_{n_X})'$ through series-by-series OLS regressions of $X_{i,t}$ onto $\hat{f}_t$, controlling for a linear time trend. The OLS residuals $\hat{v}_{i,t}$ are estimates of the idiosyncratic components $v_{i,t}$.

As in Stock & Watson (2016), we next fit AR(p_v) processes to each idiosyncratic residual $\hat{v}_{i,t}$ by OLS, separately for each i. Unlike Stock & Watson, we apply the Pope (1990) bias correction to the AR coefficient estimates to avoid understating persistence.

Differently from the stationary specification in Stock & Watson (2016), we fit a VECM to the 6 estimated factors $\hat{f}_t$, explicitly allowing for unit roots and cointegration. We specify that the VECM in error correction form contains $p_f - 1$ lagged difference terms, implying a VAR(p_f) model for the factors, and control for a time trend. We estimate the cointe-

[C.1] Like Stock & Watson, we estimate the factors using only a subset of the variables in X_t, since some variables are essentially aggregates of other variables.

gration rank using the Johansen (1995) maximum eigenvalue test, again controlling for an unrestricted time trend.[C.2] We apply a sequential testing procedure using a 5% significance level. Given the final non-rejected cointegration rank, the corresponding MLE of the VECM parameters is then transformed into estimates of the VAR parameters. Note that, while the empirical estimation has controlled for deterministic time trends, we omit any such deterministic terms from the final calibrated DFM (5)–(7) used in our simulation study.

To ensure comparability with Stock & Watson (2016), the principal components routine uses data starting in 1959Q1, while the loadings, idiosyncratic components, and factor VECM are estimated on data starting in 1959Q3.

FURTHER ESTIMATION DETAILS AND RESULTS. We select the factor and idiosyncratic error lag lengths using the Akaike information criterion (AIC). For the factors, AIC is minimized at $p_f = 3$, but the criterion is essentially flat for $p_f \in [2, 4]$; thus, to err on the side of allowing for richer long-run dynamics, we set $p_f = 4$. For the idiosyncratic errors, we apply the AIC to each individual series. The 90th percentile of selected lags equals 4; we thus set $p_v = 4$ to be consistent with the clear majority of the series. All estimated idiosyncratic AR(4) processes are technically stationary, but half the processes have largest AR root exceeding 0.86, with 25% exceeding 0.93.

The Johansen test selects a cointegration rank of 2 for the factor VECM, corresponding to $6 - 2 = 4$ common stochastic trends.

[C.2] We use the command `jcitest` in Matlab's Econometrics Toolbox.

Appendix D Definition of recursive shock estimand

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

Recall that the encompassing DFM takes the form (5)–(7). We map this DFM into the "ABCD" form of Fernández-Villaverde et al. (2007) as follows. The general ABCD representation takes the form

$$s_t = As_{t-1} + Be_t, \tag{D.1}$$
$$y_t = Cs_{t-1} + De_t, \tag{D.2}$$

where $e_t \sim \mathcal{N}(0, I)$. Define the notation $\Phi_{1:p} = (\Phi_1, \ldots, \Phi_p)$, $\Delta_\ell = \text{diag}(\delta_{1,\ell}, \ldots, \delta_{n_X,\ell})$, $\Delta_{1:p} = (\Delta_1, \ldots, \Delta_p)$, and $\Xi = \text{diag}(\Xi_1, \ldots, \Xi_{n_X})$. Then the mapping from (5)–(7) to (D.1)–(D.2) given a selected set of observables $\bar{w}_t = \bar{S}X_t$ is

$$\underbrace{\begin{pmatrix} f_t \\ \vdots \\ f_{t-p_f+1} \\ v_t \\ \vdots \\ v_{t-p_v+1} \end{pmatrix}}_{s_t} = \underbrace{\begin{pmatrix} \Phi_{1:p_f-1} & \Phi_{p_f} & 0 & 0 \\ I & 0 & 0 & 0 \\ 0 & 0 & \Delta_{1:p_v-1} & \Delta_{p_v} \\ 0 & 0 & I & 0 \end{pmatrix}}_{A} \begin{pmatrix} f_{t-1} \\ \vdots \\ f_{t-p_f} \\ v_{t-1} \\ \vdots \\ v_{t-p_v} \end{pmatrix} + \underbrace{\begin{pmatrix} H & 0 \\ 0 & 0 \\ 0 & \Xi \\ 0 & 0 \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} \varepsilon_t \\ \xi_t \end{pmatrix}}_{e_t}, \tag{D.3}$$

$$\bar{w}_t = \underbrace{\bar{S}\begin{pmatrix} \Lambda\Phi_{1:p_f} & \Delta_{1:p_v} \end{pmatrix}}_{C} \begin{pmatrix} f_{t-1} \\ \vdots \\ f_{t-p_f} \\ v_{t-1} \\ \vdots \\ v_{t-p_v} \end{pmatrix} + \underbrace{\bar{S}\begin{pmatrix} \Lambda H & \Xi \end{pmatrix}}_{D} \begin{pmatrix} \varepsilon_t \\ \xi_t \end{pmatrix}. \tag{D.4}$$

Next we derive the "innovations representation" of the ABCD model, proceeding exactly as in Fernández-Villaverde et al. (2007):[D.1]

$$\hat{x}_t = A\hat{x}_{t-1} + K\bar{u}_t,$$

[D.1] To ensure that the matrix $\Sigma = \text{Var}(s_t \mid \bar{w}_t, \bar{w}_{t-1}, \ldots)$ is positive semidefinite despite numerical rounding

$$\bar{w}_t \quad = \quad C\hat{x}_{t-1} + \bar{u}_t,$$

where $\hat{x}_t = E[s_t \mid \bar{w}_t, \bar{w}_{t-1}, \dots]$ and $\bar{u}_t = \bar{w}_t - E[\bar{w}_t \mid \bar{w}_{t-1}, \bar{w}_{t-2}, \dots]$. The innovations representation immediately yields the impulse responses of the observables $\bar{w}_t$ with respect to the Wold innovations $\bar{u}_t$. We orthogonalize the Wold innovations using a Cholesky decomposition, given the chosen ordering of the variables. In particular, letting $\text{Var}(\bar{u}_t) = \Sigma_{\bar{u}} = \bar{B}\bar{B}'$, with $\bar{B}$ lower triangular, and denoting the Wold innovation impulse responses by $\bar{C}(L)$, we define the recursive impulse response estimands as

$$\theta_h \equiv \frac{\bar{C}_{\iota_y,\bullet,h}\bar{B}_{\bullet,\iota_i}}{\bar{C}_{\iota_i,\bullet,0}\bar{B}_{\bullet,\iota_i}}, \quad h = 0, 1, 2, \dots, \tag{D.5}$$

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 (D.5) 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.[D.2] In other words, the expression (D.5) is the impulse response with respect to a potentially non-structural innovation.

To interpret the estimand (D.5), 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 (D.5) gives impulse responses to those innovations, normalized by the impact response of government spending.

---

errors, we write Equation 9 in Fernández-Villaverde et al. (2007) in the following equivalent way:

$$\Sigma = L\tilde{\Sigma}L', \quad \text{where } L \equiv \tilde{A} - \tilde{A}\tilde{\Sigma}\tilde{C}'(\tilde{C}\tilde{\Sigma}\tilde{C}')^{-1}\tilde{C},\ \tilde{\Sigma} \equiv \begin{pmatrix} \Sigma & 0 \\ 0 & I \end{pmatrix},\ \tilde{A} \equiv (A, B),\ \tilde{C} \equiv (C, D).$$

[D.2] A necessary condition for the impulse responses (D.5) 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 $\bar{w}_t$.

Appendix E Examples of estimated IRFs

Figures E.1 to E.7 provide a visual illustration of estimated impulse response functions (IRFs) from the seven 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 seven estimation methods to these ten data sets.

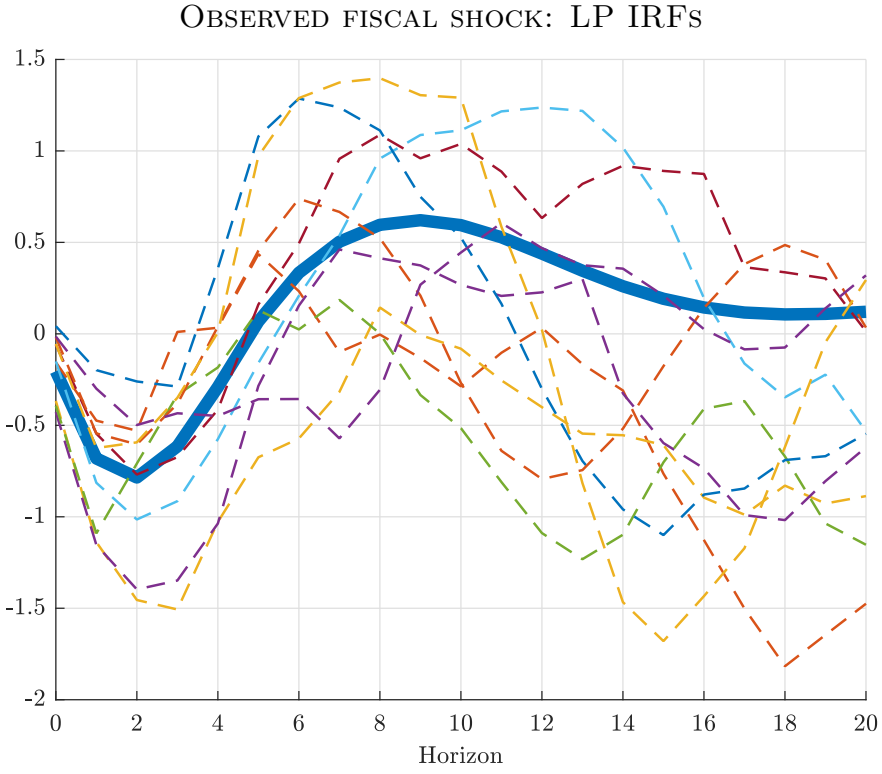

Figure E.1: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares LP impulse response estimates.

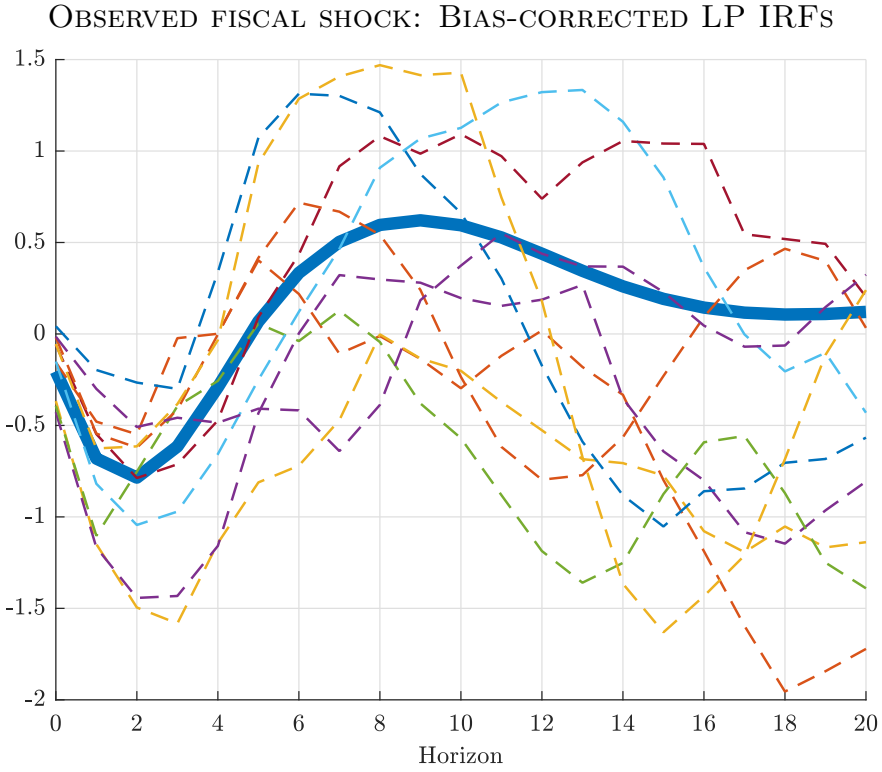

Figure E.2: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten bias-corrected LP impulse response estimates.

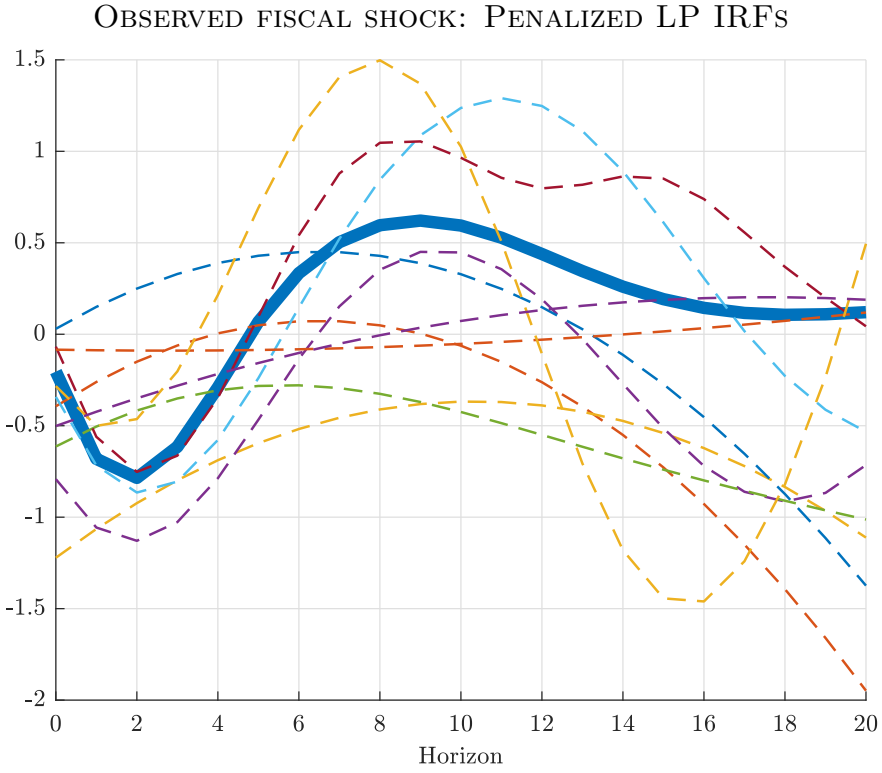

Figure E.3: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten penalized LP impulse response estimates.

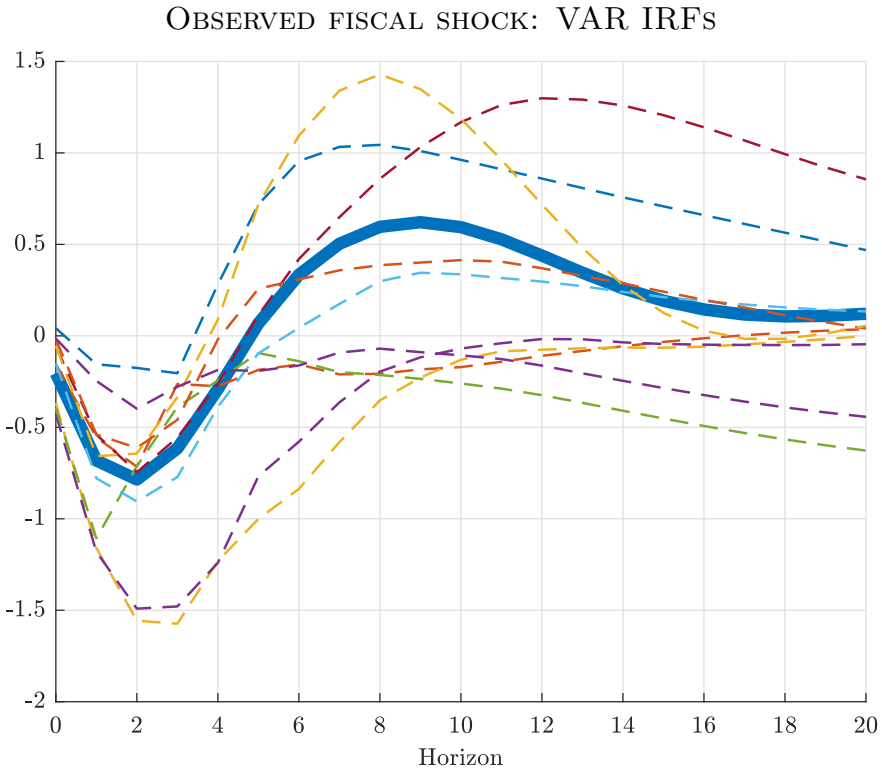

Figure E.4: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten least-squares VAR impulse response estimates.

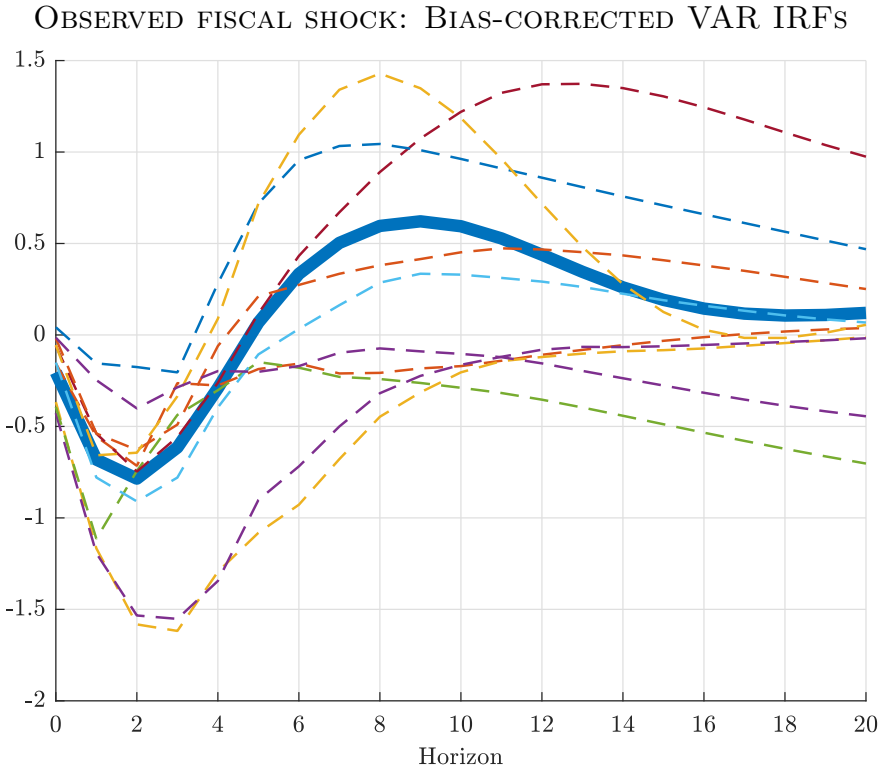

Figure E.5: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten bias-corrected VAR impulse response estimates.

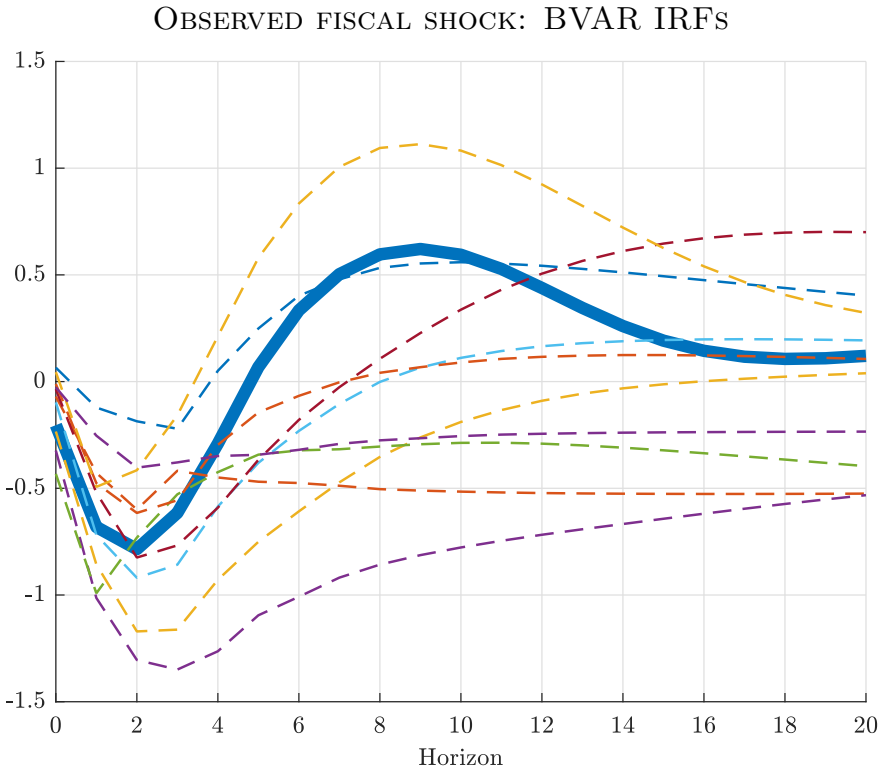

Figure E.6: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten Bayesian VAR impulse response estimates.

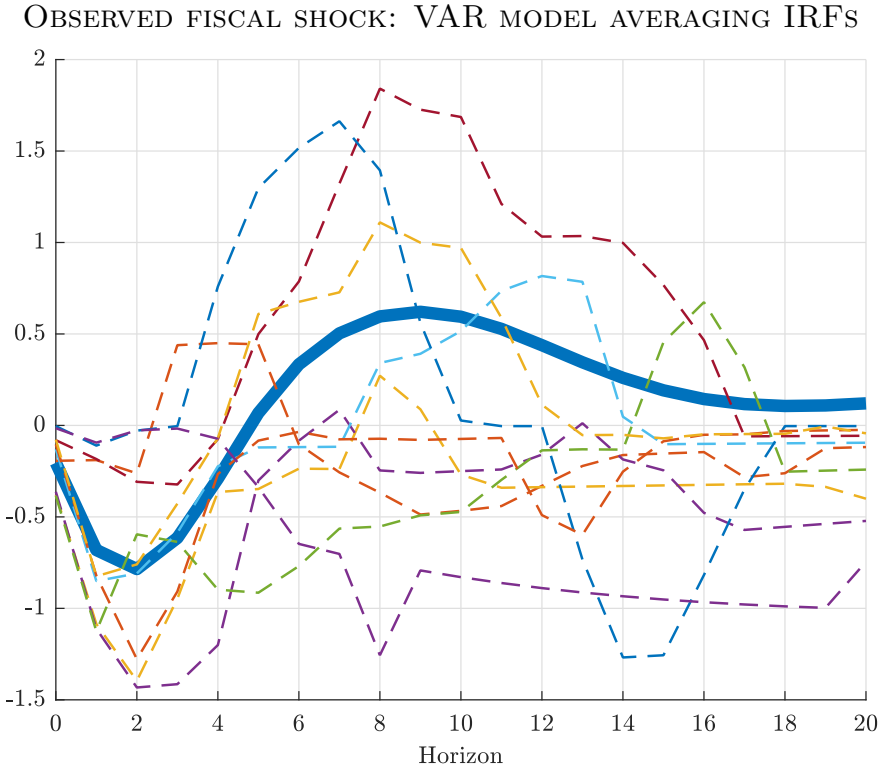

Figure E.7: Structural impulse response estimand (thick blue) for one specification with an observed fiscal spending shock vs. ten VAR model averaging impulse response estimates.

Appendix F Further simulation results and robustness

F.1 IV estimators

Figures F.1 and F.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 our main analysis in Section 5.4.

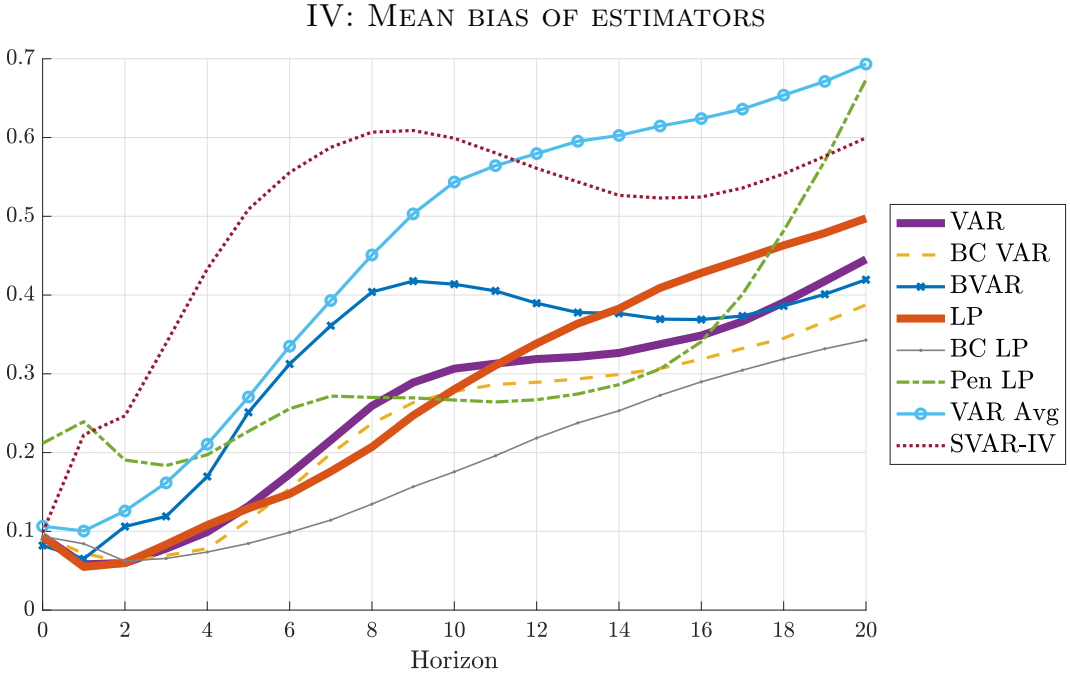

Figure F.1: Median (across DGPs) of absolute mean bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

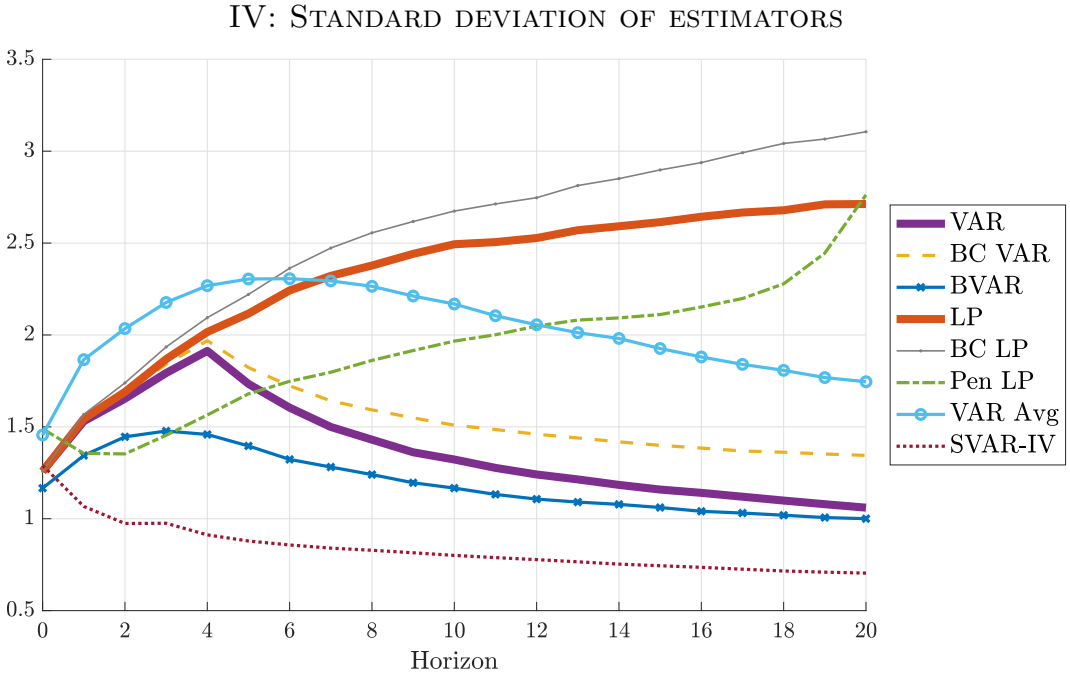

Figure F.2: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

F.2 Stationary DGPs

For our baseline analysis we used a non-stationary DFM with variables in levels as our encompassing model. We did this to follow the predominant practice in applied work of using variables in levels in LP and VAR regressions. Nevertheless, some researchers work with data that has been transformed to stationarity prior to the analysis. We here discuss results from an analogous simulation study conducted with a stationary encompassing DFM.

Encompassing model. We consider a stationary version of the encompassing DFM (5)–(7). In particular, we parametrize the model based on the empirical reduced-form parameter estimates from Stock & Watson (2016), using the same specification as in Lazarus et al. (2018). We provide a brief summary here and refer to Stock & Watson for further details. Differently from our main analysis, each series in the vector of observables X_t is now transformed to ensure approximate stationarity. We follow Stock & Watson in selecting $n_f = 6$ factors, two lags in the factor equation (5), and two lags in the idiosyncratic component equation (7). The reduced-form parameters are estimated by principal components and least-squares procedures; in particular, the factor VAR equation (5) is estimated by OLS instead of imposing a VECM model. This pins down all parameters of the DFM except for the structural impact response matrix H, which we construct exactly as in our main analysis.

DGP summary statistics. Table F.1 shows summary statistics of the 6,000 stationary DGPs (3,000 monetary DGPs and 3,000 fiscal ones, as in our baseline analysis). The persistence of the DGPs varies widely, as measured by the ratio $\text{trace}(LRV(\bar{w}_t))/\text{trace}(\text{Var}(\bar{w}_t))$, but note that none of the DGPs feature near-unit roots, with the largest absolute eigenvalue of the VAR companion matrix being below 0.91 in all cases. As in our baseline analysis, the DGPs feature marked heterogeneity in the characteristics of impulse response functions and in the degree to which the DGPs are well-approximated by a VAR(4) model.

Estimators. The estimation procedures are the same as in the baseline analysis, except for the Bayesian VAR. Rather than centering the Minnesota prior at independent random walks, we center at independent white noise, i.e., all autoregressive coefficients are shrunk towards 0. We also remove the "sum-of-coefficients" and "dummy-initial-observation" priors.

Results. Figures F.3 to F.5 show bias, standard deviation, as well as the best estimation method choice for our experiments based on the stationary DFM, focusing here on the case

STATIONARY DGPS: SUMMARY STATISTICS

Percentile	min	10	25	50	75	90	max
Data and shocks							
trace(long-run var)/trace(var)	0.36	0.76	1.00	1.45	2.18	3.71	18.09
Largest VAR eigenvalue	0.84	0.84	0.84	0.84	0.85	0.86	0.91
Fraction of VAR coef's $\ell \geq 5$	0.02	0.11	0.15	0.21	0.27	0.34	0.77
Impulse responses up to $h = 20$							
No. of interior local extrema	1	2	2	2	3	4	6
Horizon of max abs. value	0	0	0	0	1	2	8
Average/(max abs. value)	-0.42	-0.16	-0.08	-0.02	0.06	0.11	0.43
R^2 in regression on quadratic	0.01	0.10	0.20	0.46	0.69	0.83	0.97

Table F.1: Quantiles of various population parameters across the stationary DGPs for observed shock identification. "long-run var": long-run variance of series. "var": variance of series. "Largest VAR eigenvalue": largest absolute eigenvalue of reduced-form VAR companion matrix. "Fraction of VAR coef's $\ell \geq 5$": $\sum_{\ell=5}^{1000} \|A_\ell^w\| / \sum_{\ell=1}^{1000} \|A_\ell^w\|$, where A_ℓ^w are the population VAR(∞) coefficient matrices and $\|\cdot\|$ is the Frobenius norm. "Average/(max abs. value)": $(\frac{1}{21}\sum_{h=0}^{20}\theta_h)/\max_h\{|\theta_h|\}$. "$R^2$ in regression on quadratic": R^2 from a regression of the impulse response function $\{\theta_h\}_{h=0}^{20}$ on a quadratic polynomial in h.

of an observed shock. We summarize the findings in three lessons that are analogous to those in Sections 5.1 to 5.3.

1. Least-squares LP and VAR estimators again lie on opposite ends of the bias-variance spectrum. Furthermore, as in our main analysis, the slope of this trade-off is stark, with indifference between the two methods requiring the researcher to almost fully prioritize bias. Since the stationary DGPs are not highly persistent, bias correction has negligible impact, and indeed the uncorrected LP estimator has near-zero bias at all horizons (as predicted by asymptotic theory). Note also that the bias and standard deviation of the VAR estimators converge to zero as the horizon increases, a mechanical feature of VAR estimators in stationary DGPs that is not shared by LP methods.

2. LP (bias-corrected or uncorrected) remains the preferred method when the weight ω on bias in the loss function is very high, but this requires ω to be even closer to 1 than in our baseline analysis. The region (gray with vertical lines) where bias-corrected LP is preferred at the top of Figure F.5 is very thin. Whenever the loss function puts non-trivial

weight on precision, penalized LP is usually preferred to least-squares or bias-corrected LP due to the substantial variance reduction.

3. VAR methods tend to be preferred to least-squares or bias-corrected LP whenever the weight on variance in the loss function is non-trivial. However, for MSE loss, BVAR is now outperformed (on average across DGPs) at very short horizons $h \leq 2$ by penalized LP, another shrinkage method. BVAR tends to outperform least-squares VAR at shorter horizons, but this ranking is reversed at intermediate horizons; at long horizons, the methods perform similarly since the estimated impulse responses are close to 0 anyway. Though at first sight bias-corrected VAR (yellow with horizontal lines) looks attractive in Figure F.5 at intermediate horizons when the weight on bias is moderately high, we remind the reader that least-squares VAR (purple) performs very similarly. VAR model averaging continues to perform poorly regardless of horizon and bias-variance preferences.

Observed shock, stationary DGPs: Bias of estimators

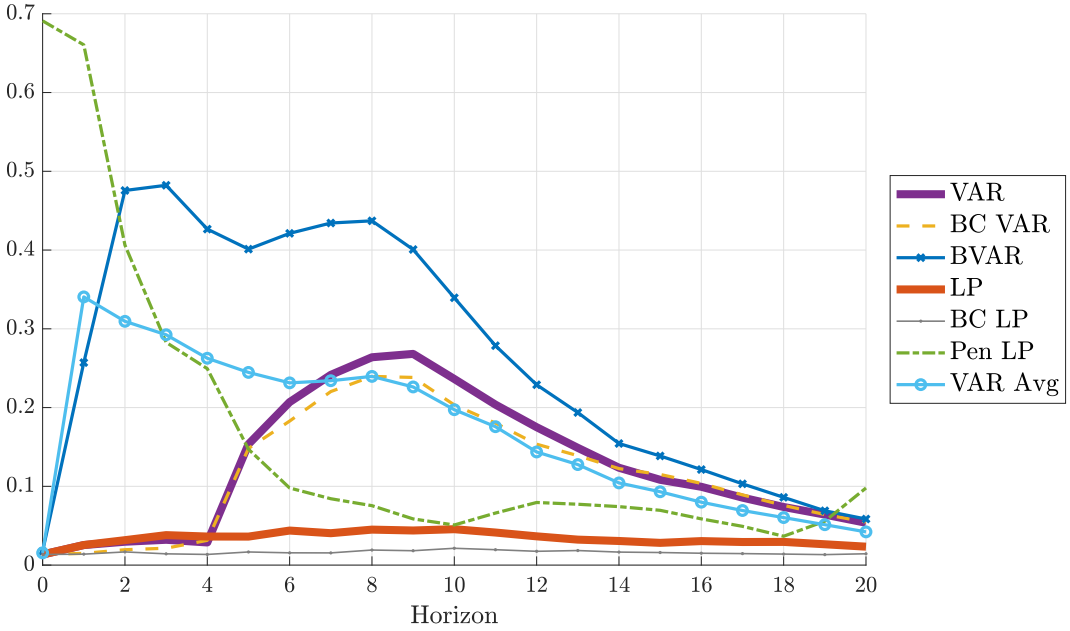

Figure F.3: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, stationary DGPs: Standard deviation of estimators

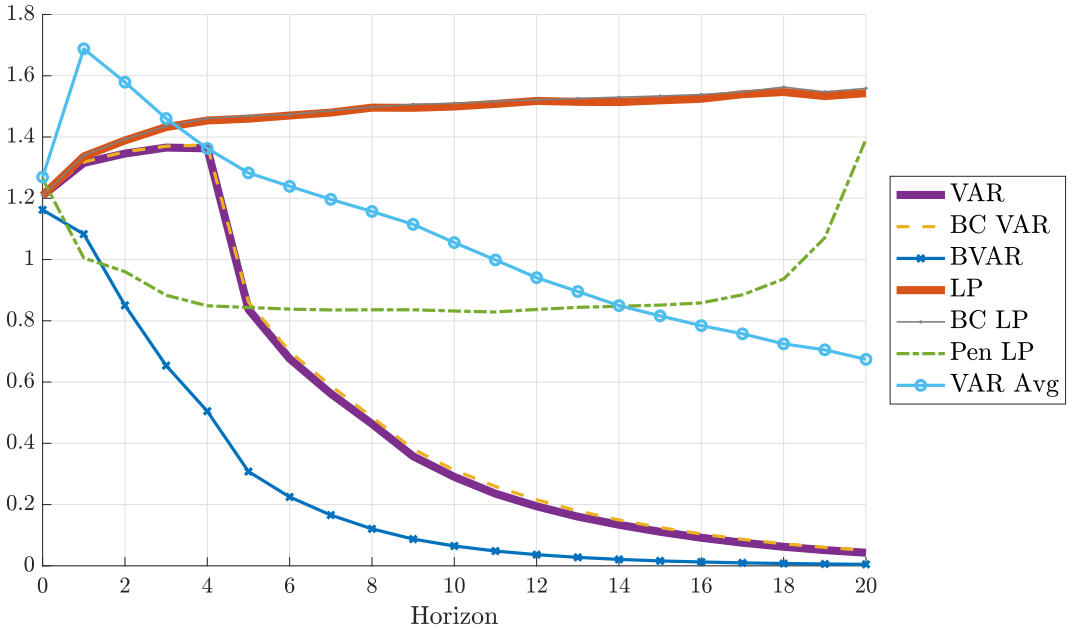

Figure F.4: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

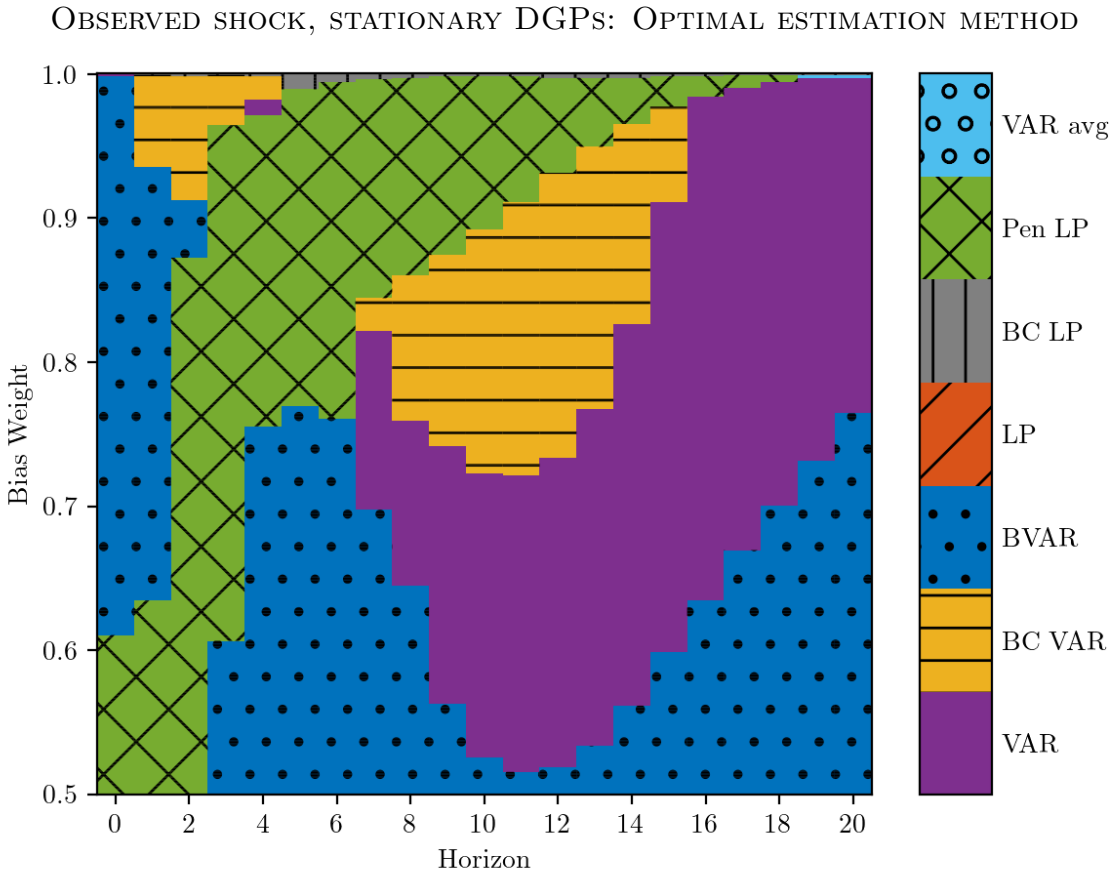

Figure F.5: Method that minimizes the average (across DGPs) loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures F.3 and F.4.

RECURSIVE IDENTIFICATION: DGP SUMMARY STATISTICS

Percentile	min	10	25	50	75	90	max
Impulse responses up to $h = 20$							
No. of interior local extrema	0	1	2	2	3	3	7
Horizon of max abs. value	0	1	1	6	10	20	20
Average/(max abs. value)	-0.83	-0.70	-0.58	-0.07	0.38	0.65	0.94
R^2 in regression on quadratic	0.01	0.45	0.65	0.83	0.93	0.98	1.00

Table F.2: 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.

F.3 Recursive identification

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

Table F.2 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 first three entries of the top half of that table apply without change here.[F.1]

Figures F.6 and F.7 show the median (across DGPs) absolute bias and standard deviation of the various estimators. Figure F.8 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 reasonably similar—both qualitatively as well as quantitatively—to the corresponding figures for the "observed shock" estimands in Section 5, with one notable difference: the relative bias increase for BVAR is smaller, and so now BVAR looks even more attractive, leading to the large solid-dotted blue area in Figure F.8. Intuitively, since recursive (i.e., Cholesky) shock identification depends heavily on estimating the reduced-form innovation variance-covariance matrix of the multivariate system, the stylized prior information about this matrix imposed by the BVAR provides helpful additional shrinkage that the other methods do not exploit.

[F.1] The last two entries of ourse do not apply here: the degree of invertibility is trivially equal to 1, and we are not studying IV identification.

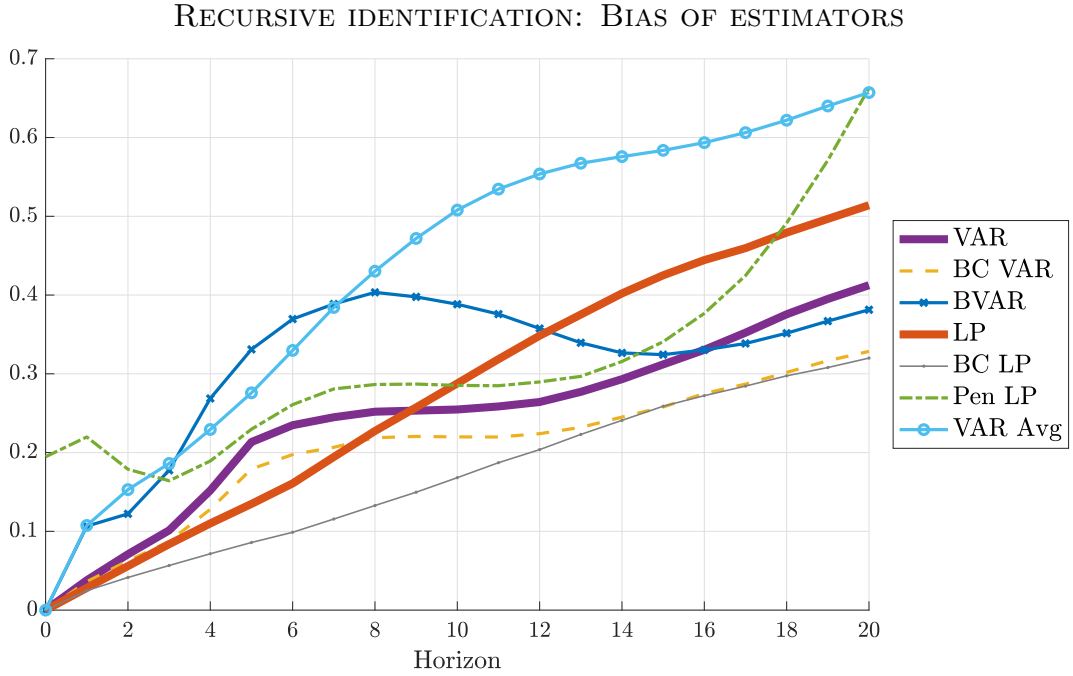

Figure F.6: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

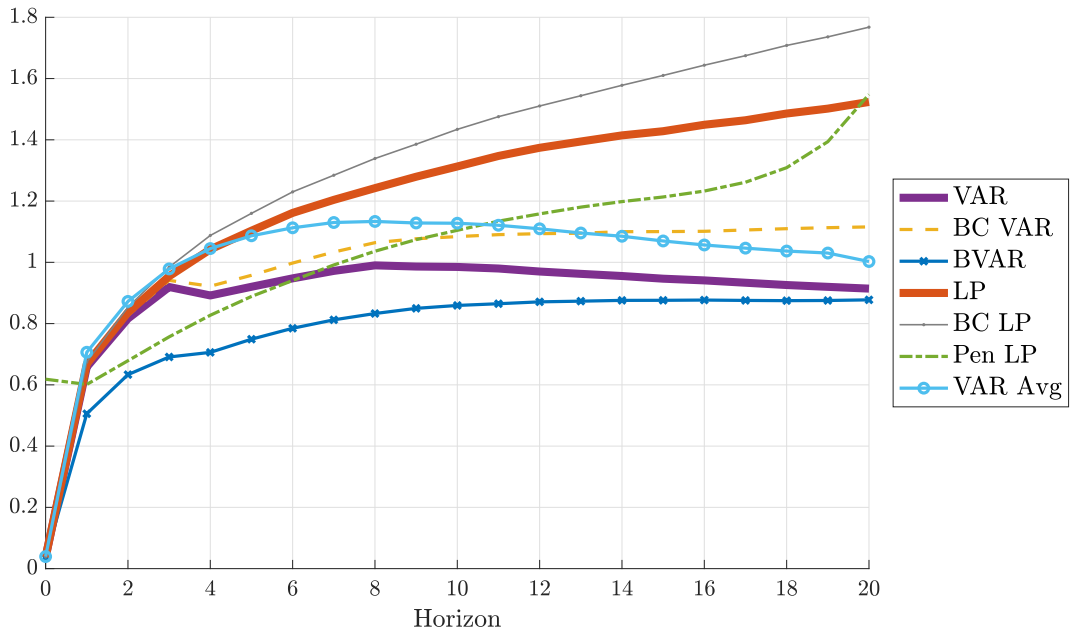

Figure F.7: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

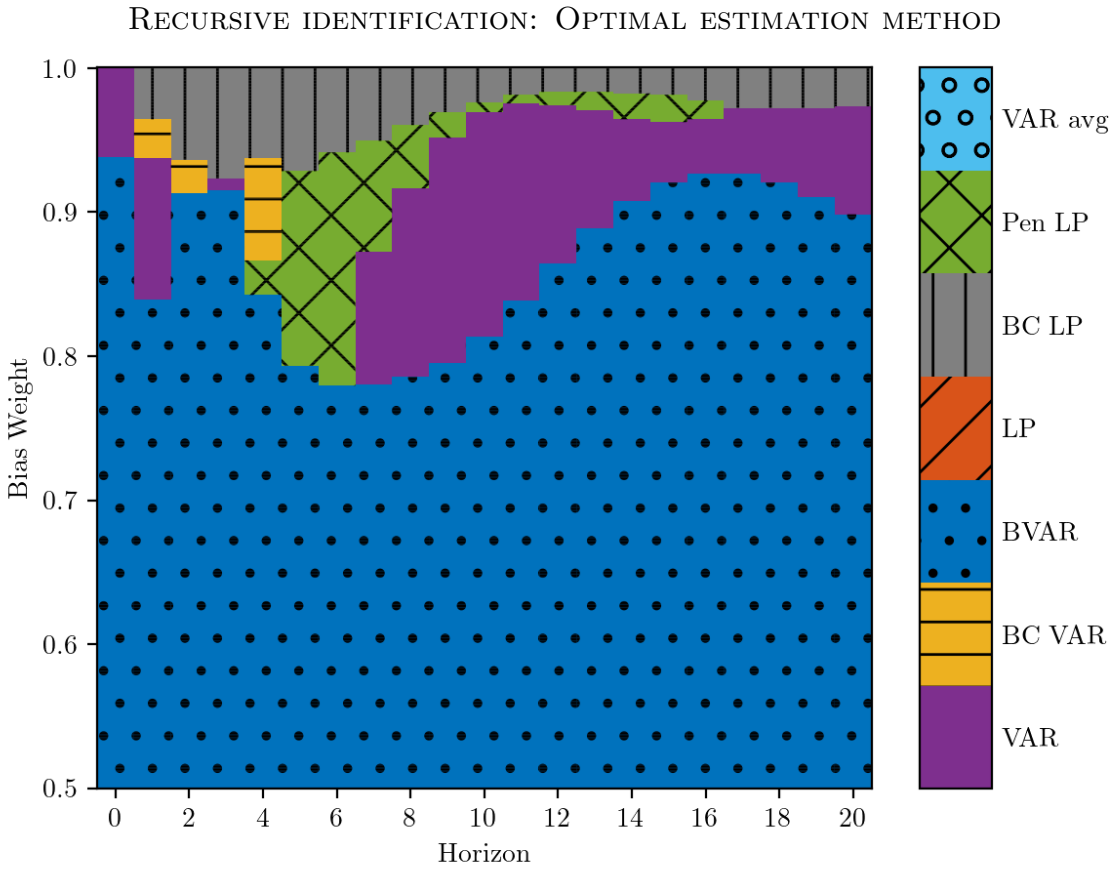

Figure F.8: Method that minimizes the average (across DGPs) loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2 and 3. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

F.4 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 policy shock estimand.

Figures F.9 and F.10 show the bias and standard deviation plots for the 3,000 fiscal shock DGPs, while Figures F.11 and F.12 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 standard deviations is somewhat higher in the fiscal shock case for all estimation methods.

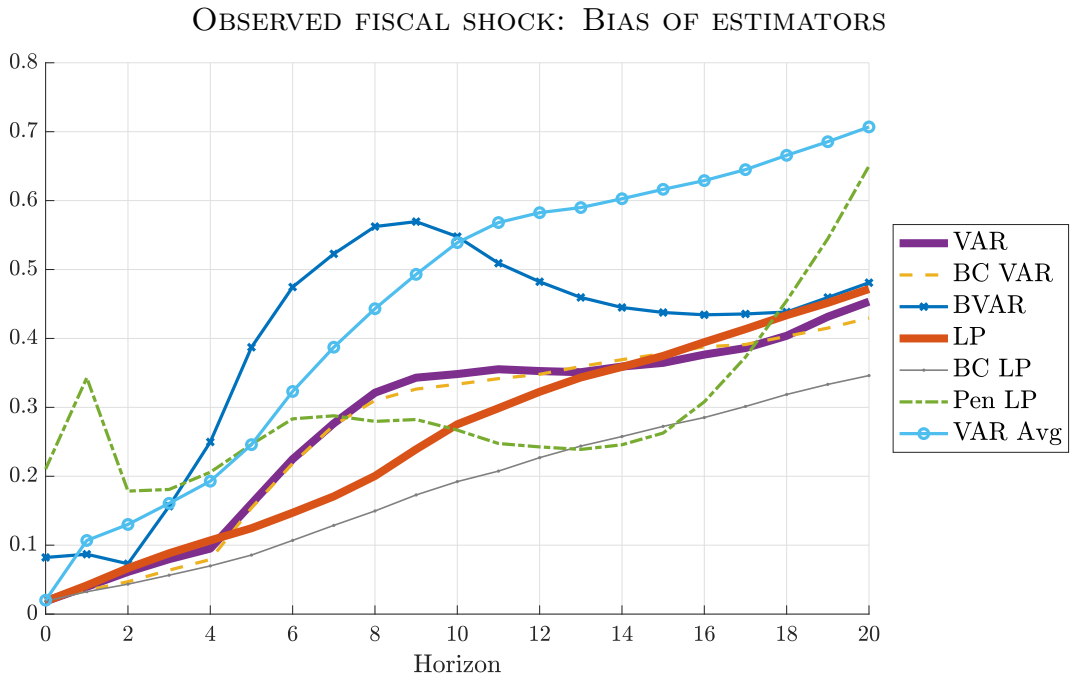

Figure F.9: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

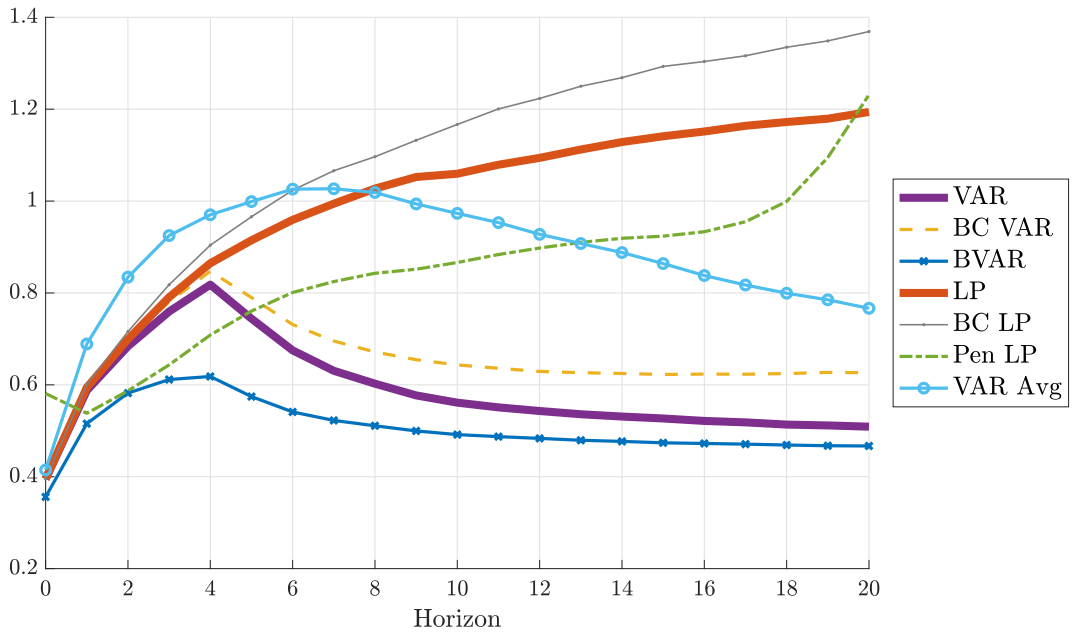

Figure F.10: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

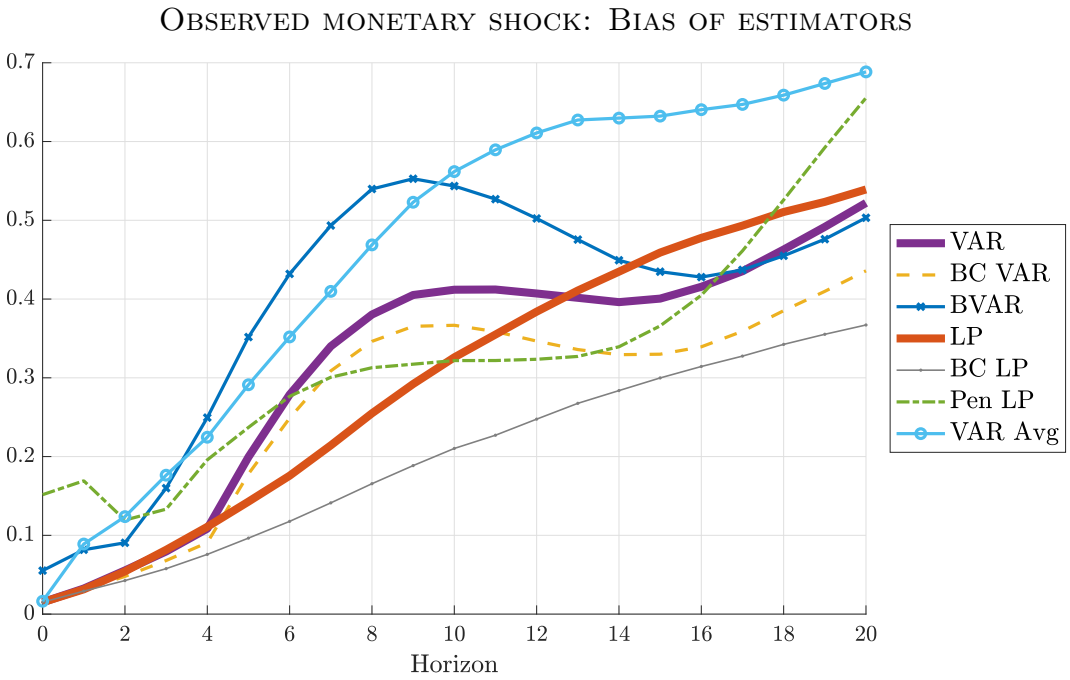

Figure F.11: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

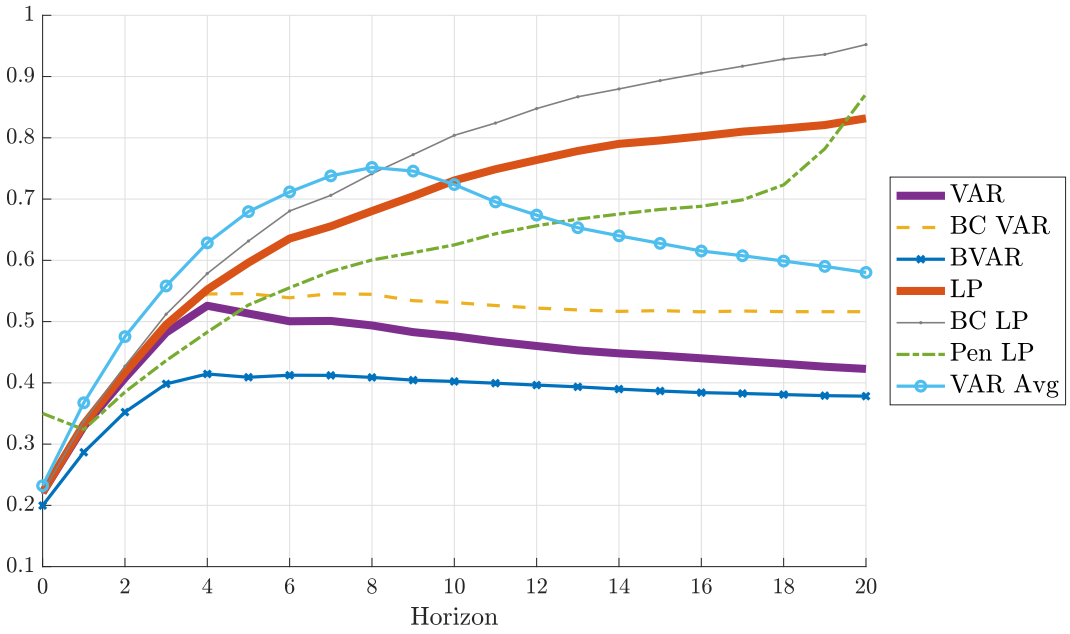

Figure F.12: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

F.5 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.13 and F.14 show the median (across DGPs) absolute bias and standard deviation of the estimation methods, while Figure F.15 shows the optimal method choice according to the loss function (which has been averaged across DGPs). Consistent with the theoretical results in Plagborg-Møller & Wolf (2021), least-squares LPs and VARs now perform similarly at all horizons $h \leq p = 8$. BVAR shrinkage now looks even more appealing than in our main analysis for loss functions with bias weight $\omega \leq 0.7$, due to the much lower standard deviation of this procedure. VAR bias correction has more bite in terms of reducing bias than in the baseline results. Otherwise the qualitative conclusions from Section 5 are unchanged.

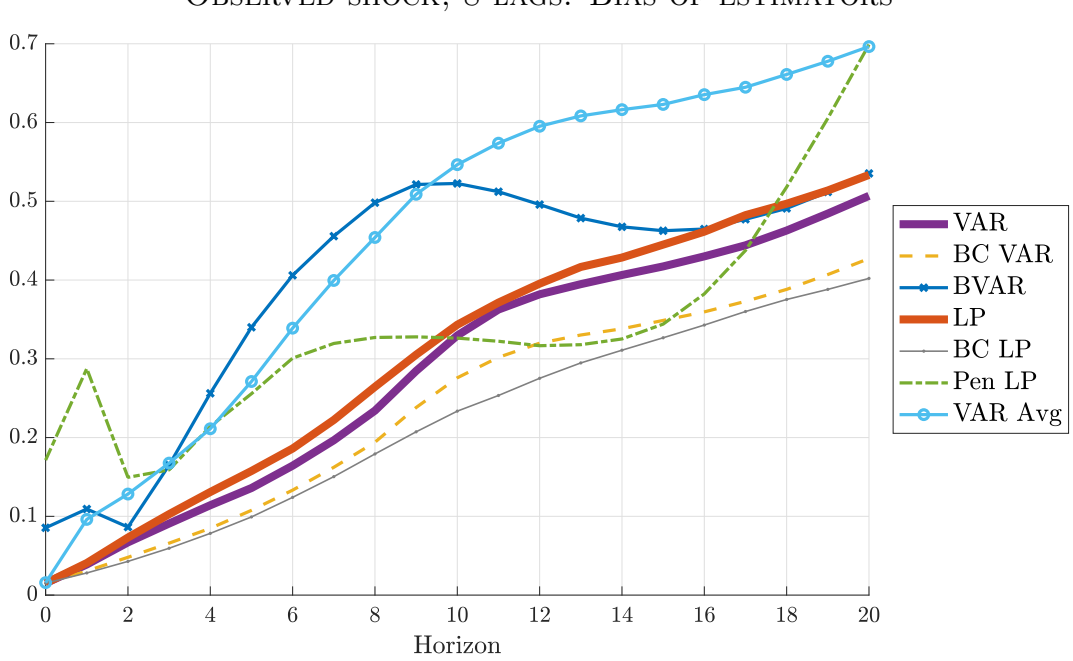

Figure F.13: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

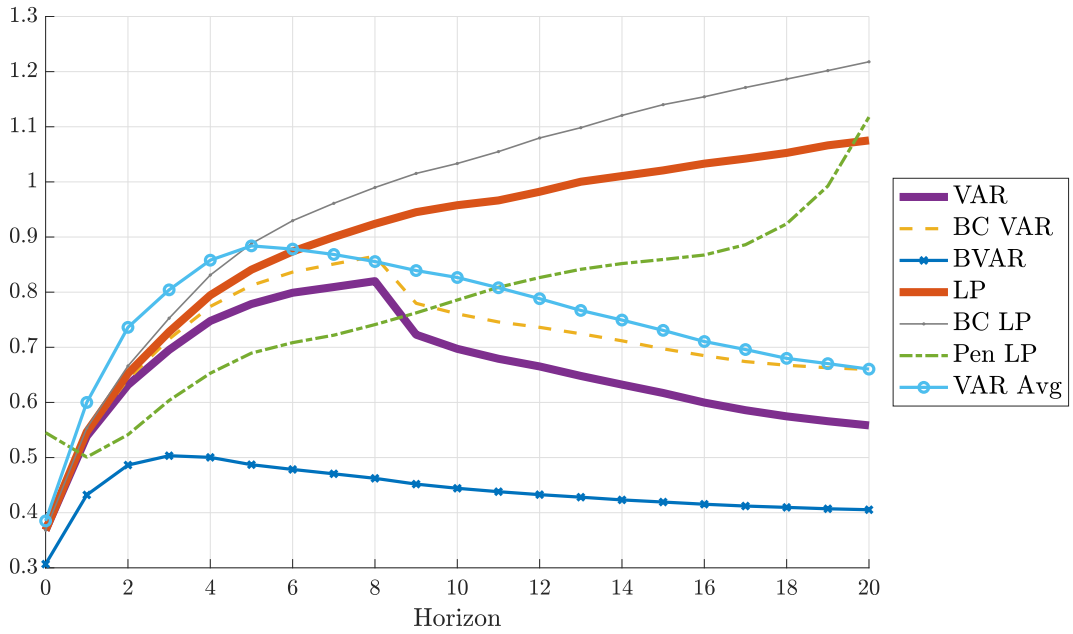

Figure F.14: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

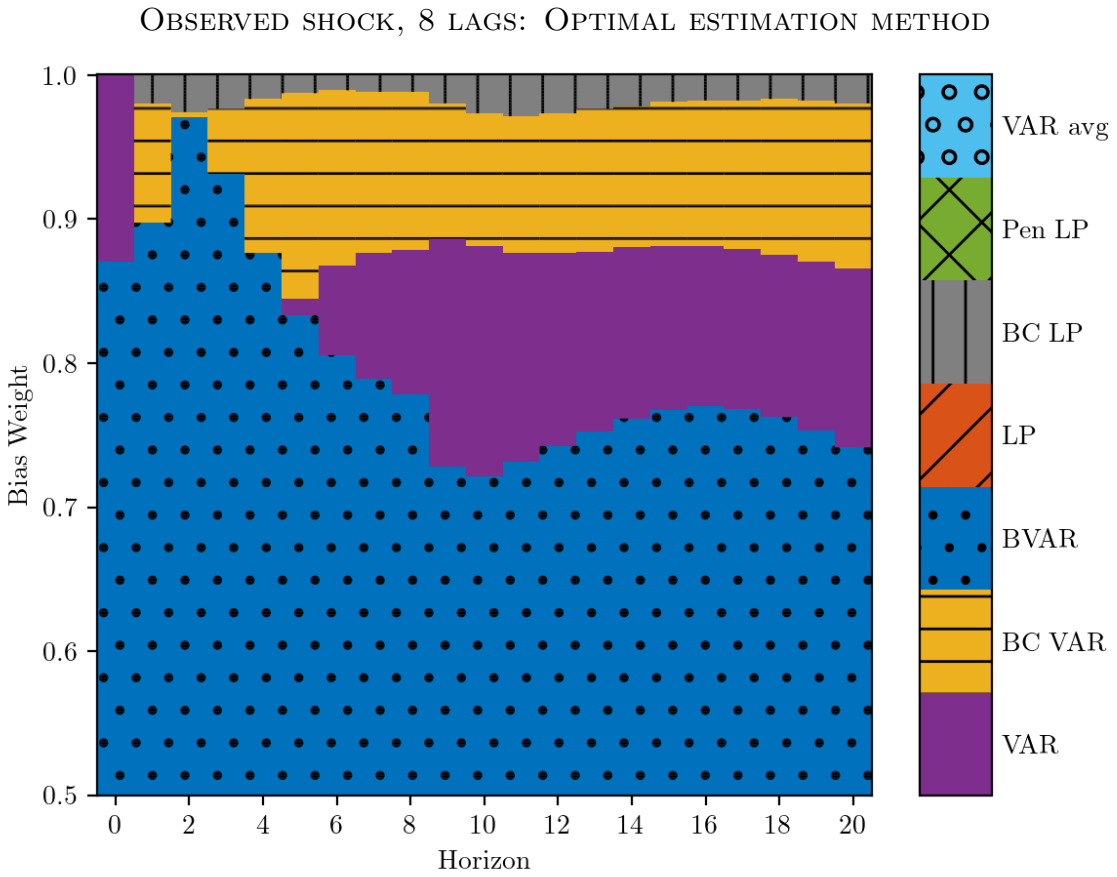

Figure F.15: Method that minimizes the average (across DGPs) loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2 and 3. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

F.6 Smaller sample size

Recall that our baseline experiments consider a sample size of $T = 200$ 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 F.16 to F.18, 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 estimation method choice plot indicates a more pronounced preference for shrinkage, with a larger area now solid-dotted blue (BVAR). Use of bias-corrected LP (grey with vertical lines) requires an even larger preference for low bias over high precision than in our baseline analysis.

Observed shock, small sample: Bias of estimators

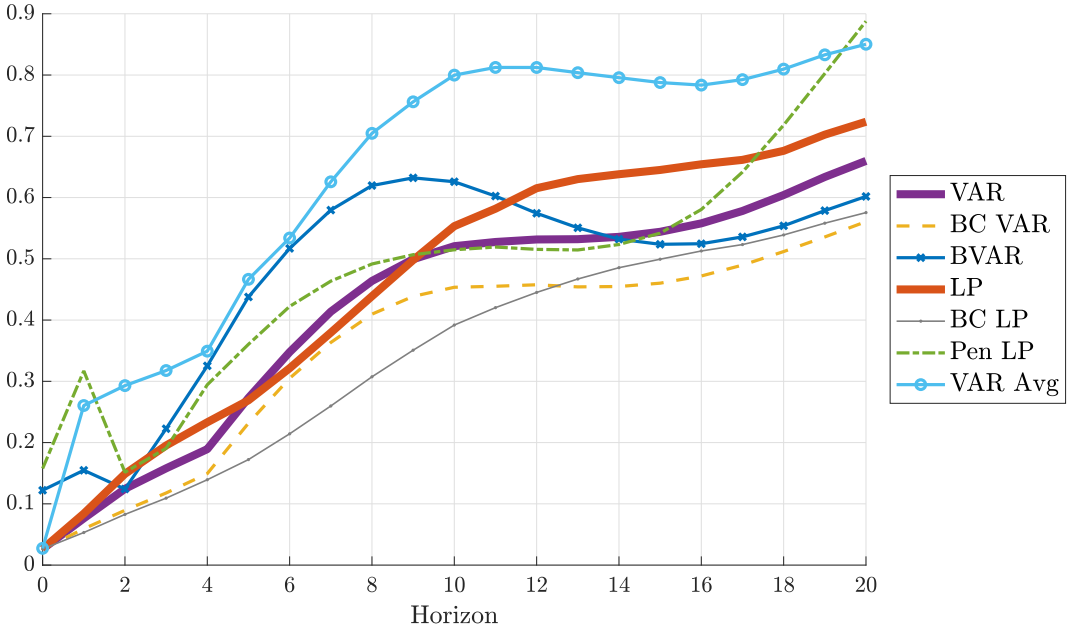

Figure F.16: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, small sample: Standard deviation of estimators

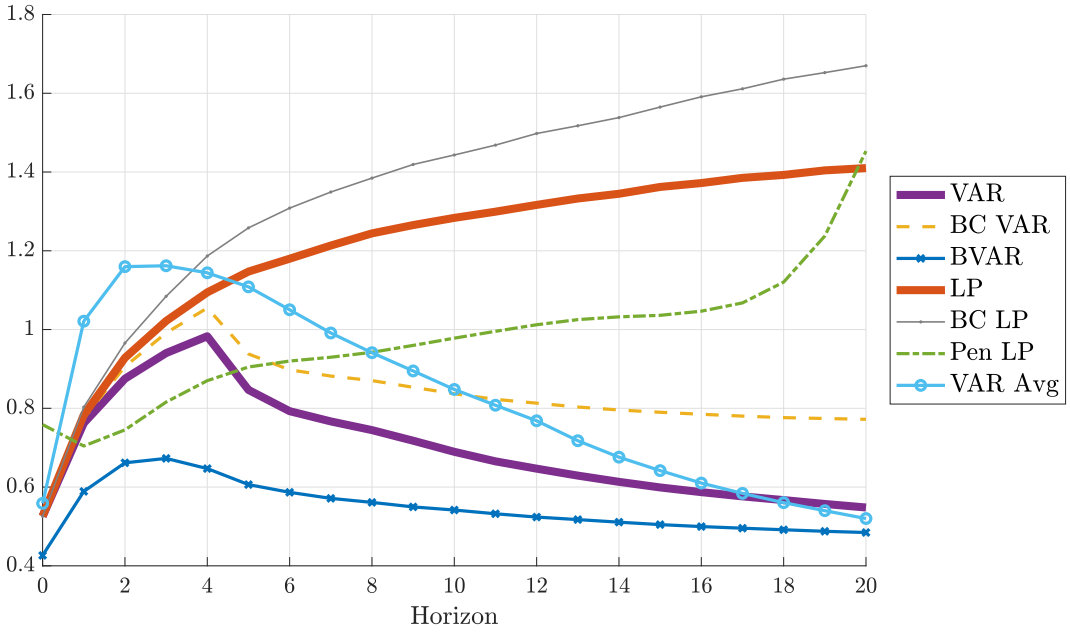

Figure F.17: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

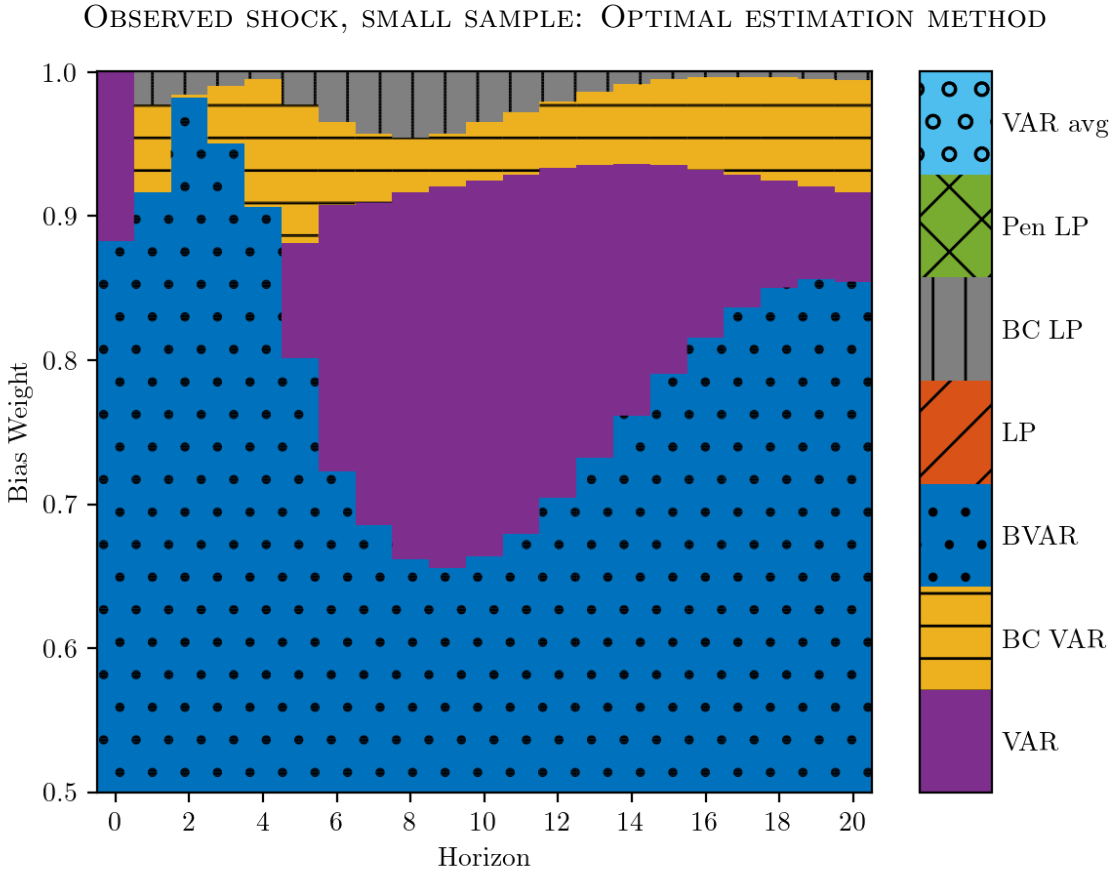

Figure F.18: Method that minimizes the average (across DGPs) loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2 and 3. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

F.7 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 F.3 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). No coefficients in the table exceed 0.2 in absolute value, indicating that none of the categories are particularly 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 the standard deviation of most of our estimation procedures, which is consistent with the results reported in Supplemental Appendix F.4. Though the regression results reported in Table F.3 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 major differences when comparing price and real activity series.

Bias regression

	VAR	VAR BC	BVAR	LP	LP BC	Pen LP	VAR Avg
NIPA	-0.12	-0.10	-0.03	-0.05	-0.06	0.01	-0.04
Industrial Production	-0.06	-0.06	0.66	-0.01	-0.01	0.06	0.02
Employment & Unemployment	0.02	0.07	0.08	-0.02	0.05	0.04	-0.03
Orders, Inventories & Sales	-0.05	-0.09	0.01	-0.03	0.00	0.02	0.00
Housing Starts & Permits	-0.01	0.01	0.04	0.02	0.08	0.08	0.06
Prices	0.05	0.15	0.12	-0.18	-0.03	-0.17	-0.05
Productivity & Earnings	0.02	-0.01	0.03	-0.02	-0.05	-0.02	-0.01
Interest Rates	-0.01	-0.01	0.03	-0.03	-0.03	-0.04	-0.03
Money & Credit	0.04	0.07	0.08	-0.03	0.01	0.00	-0.03
International	0.07	0.12	0.15	-0.11	-0.05	-0.09	-0.05

Standard deviation regression

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

Table F.3: 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. Bias and standard deviation normalized as in Figures 2 and 3. Observed shock identification.

F.8 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 five-variable combinations of 17 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 core 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 DGP 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 1,581 DGPs (845 monetary shock DGPs and 736 fiscal shock DGPs).

Results for bias, standard deviation, and optimal method choice are reported in Figures F.19 to F.21. The figures look very similar to those from our baseline analysis. We therefore conclude that there is little in the way of systematic differences between the larger set of variables included in the full DFM and this smaller subset of particularly salient macroeconomic time series.

Observed shock, salient observables: Bias of estimators

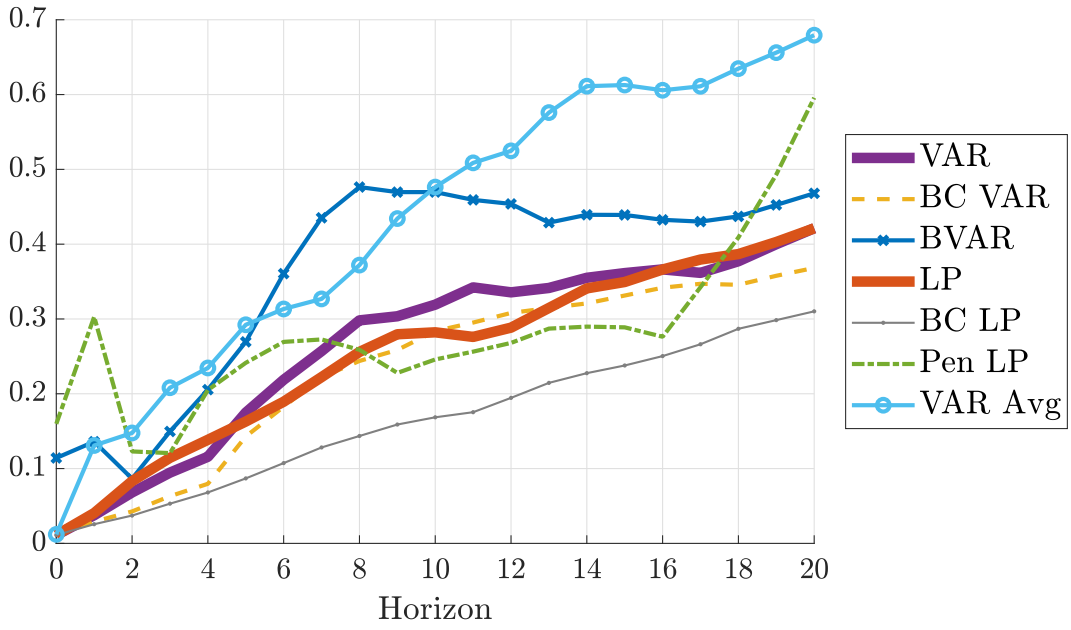

Figure F.19: Median (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Observed shock, salient observables: Standard deviation of estimators

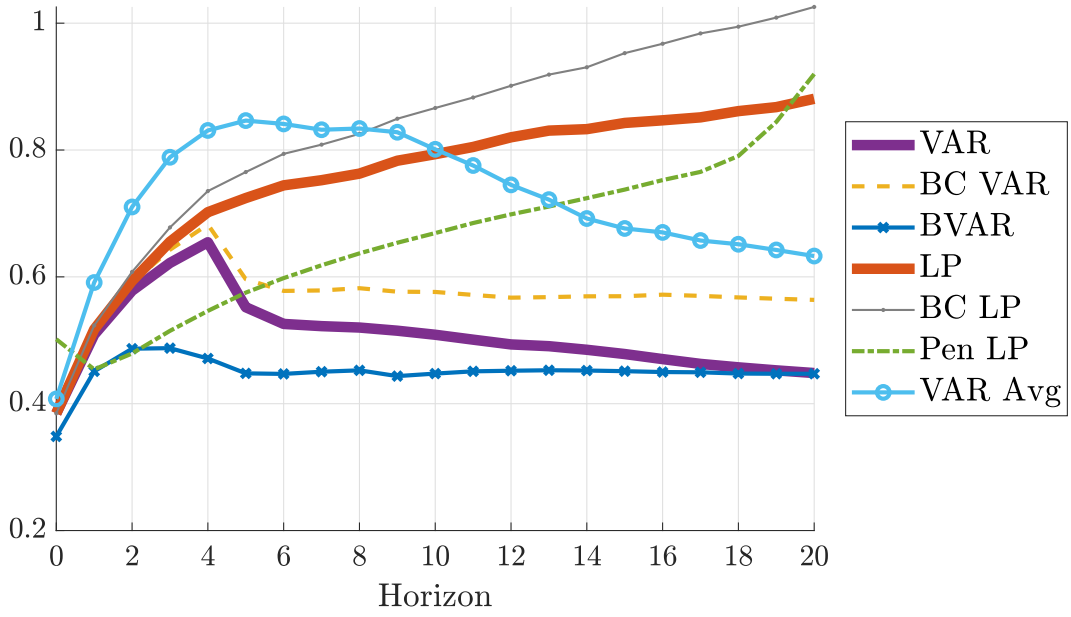

Figure F.20: Median (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

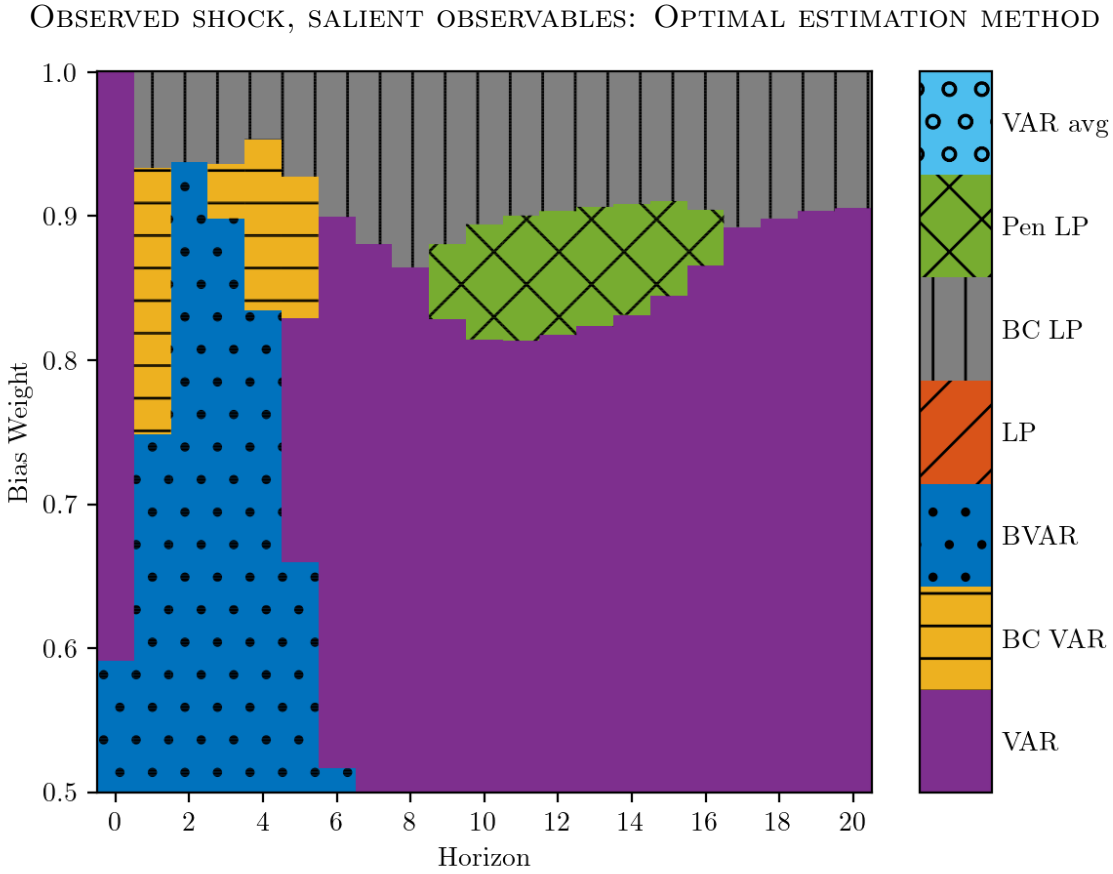

Figure F.21: Method that minimizes the average (across DGPs) loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2 and 3. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

F.9 90th percentile loss

Whereas our baseline results report medians (of 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 F.22 to F.24. By construction, the bias and standard deviation numbers are now higher for all estimators. Importantly, however, *relative* magnitudes do not change by much; that is, for DGPs in which VARs or shrinkage techniques do poorly, least-squares and bias-corrected LP tend 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

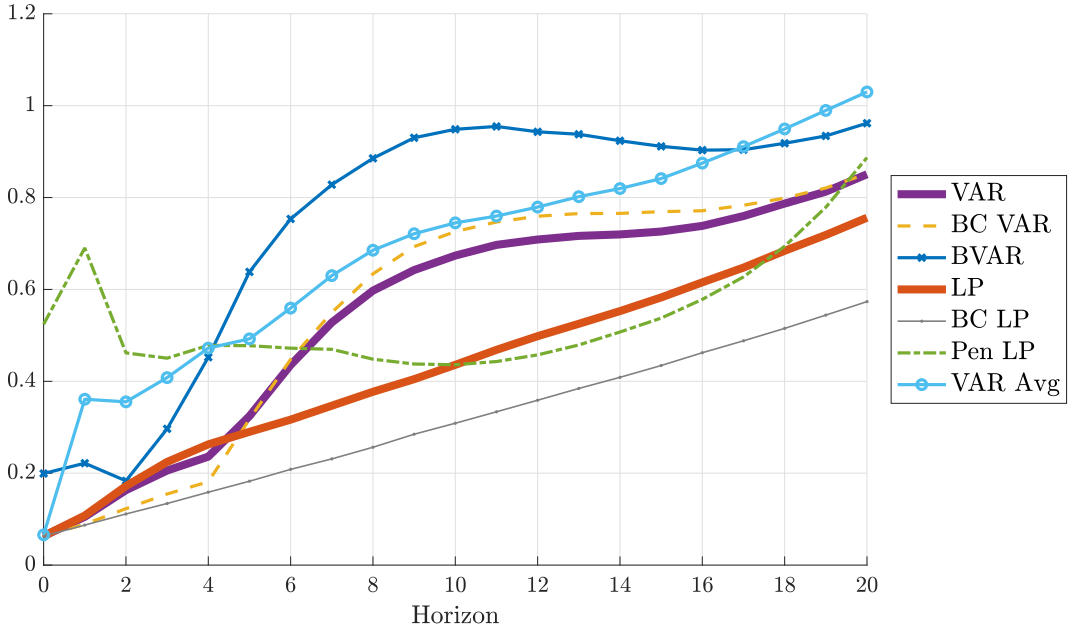

Figure F.22: 90th percentile (across DGPs) of absolute bias of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

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

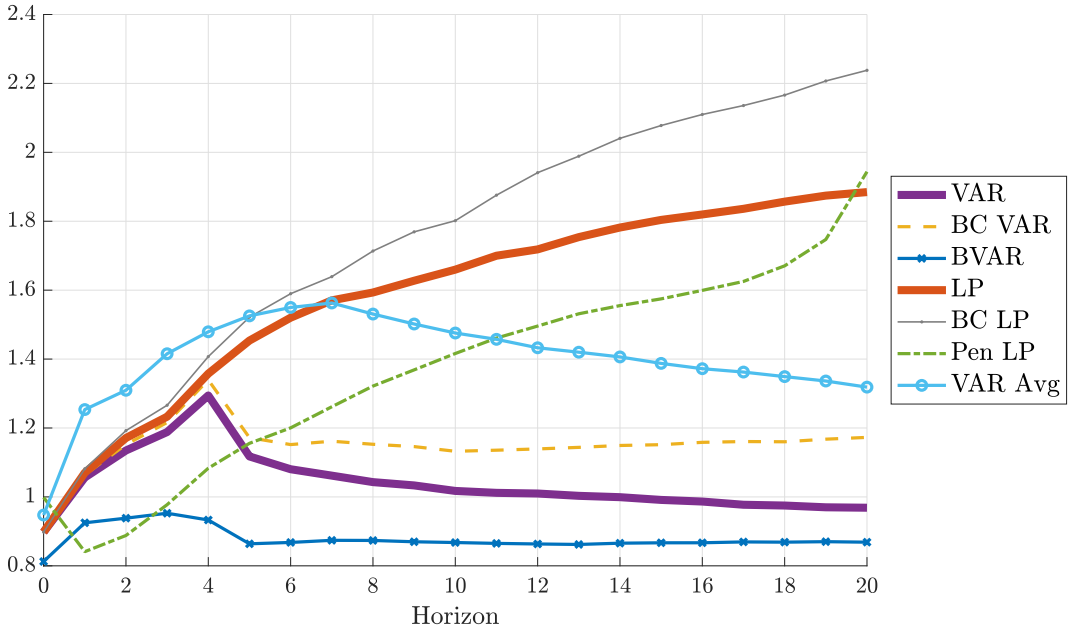

Figure F.23: 90th percentile (across DGPs) of standard deviation of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

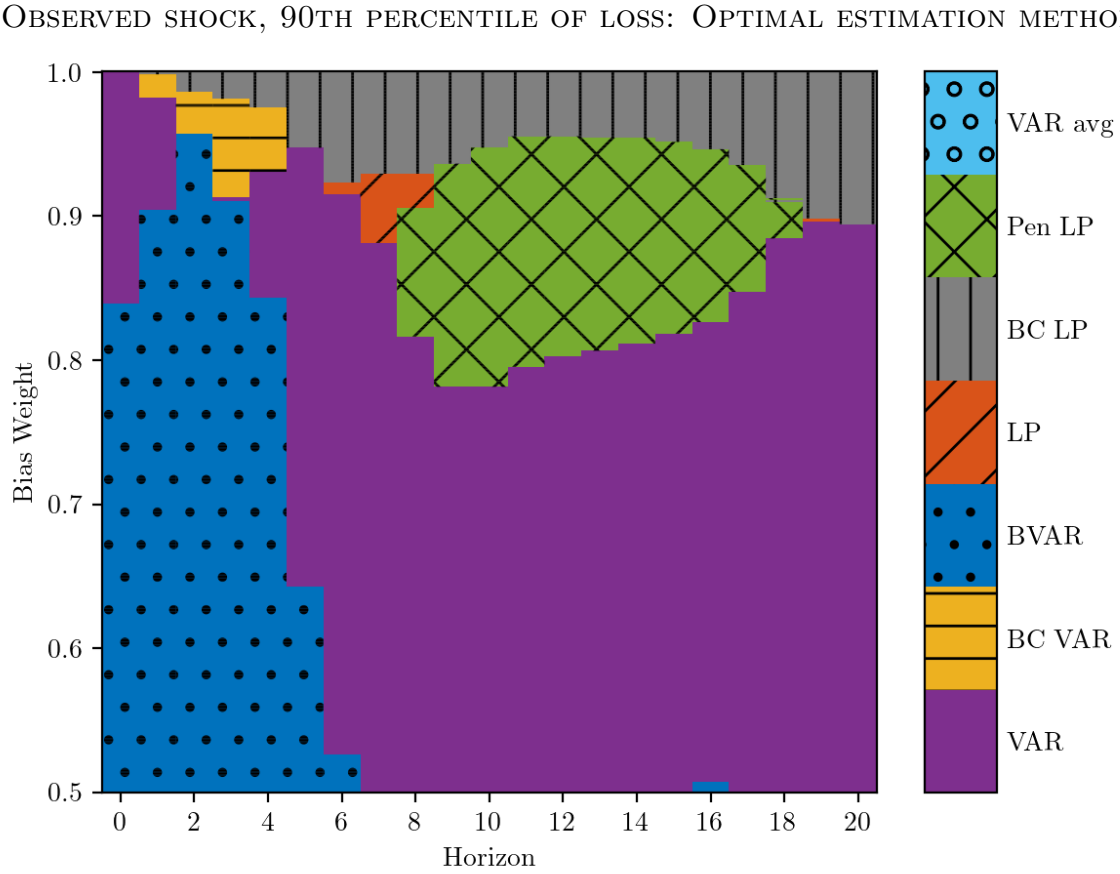

Figure F.24: Method that minimizes the 90th percentile (across DGPs) of the loss function (4). Horizontal axis: impulse response horizon. Vertical axis: weight on squared bias in loss function. The loss function is normalized by the scale of the impulse response function, as in Figures 2 and 3. At $h = 0$, VAR and LP are numerically identical; we break the tie in favor of VAR.

Appendix G Proofs

G.1 Auxiliary lemmas

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

Define the process $\tilde{y}_t \equiv \sum_{s=1}^{t}(\varepsilon_{1,s} + \tau\varepsilon_{1,s-1} + \varepsilon_{2,s})$ for all $t \geq 1$, with $\tilde{y}_0 = 0$, and let $\tilde{w}_t \equiv (\varepsilon_{1,t}, \tilde{y}_t)'$. Define also the 2×2 matrix $D_T \equiv \operatorname{diag}(T^{1/2}, T)$.

Lemma G.1. *For all $j = 1, 2$ and $\ell \geq 0$,*

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

Proof. From the DGP (1) we have

$$y_t - \tilde{y}_t = \frac{\alpha}{\sqrt{T}}\sum_{s=1}^{t}\varepsilon_{1,s-2}.$$

The first statement of the lemma follows from Markov's inequality and a simple moment calculation. Then second statement of the lemma follows from Chebyshev's inequality, since the process $\varepsilon_{j,t+\ell}\sum_{s=1}^{t}\varepsilon_{1,s-2}$ is serially uncorrelated and has a variance of order $O(T)$. □

Lemma G.2. *For any $\ell \geq 0$ and $j = 1, 2$,*

$$\sum_{t=1}^{T}\varepsilon_{j,t+\ell}\tilde{w}_{t-1}'D_T^{-1} = O_p(1).$$

Proof. This follows from Chebyshev's inequality and standard variance calculations, using that $\varepsilon_{j,t+\ell}\tilde{w}_{t-1}'$ is a serially uncorrelated process. □

Lemma G.3. *For any $\ell \geq 1$ and $j = 1, 2$,*

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{j,t-\ell}\tilde{y}_{t-1} = O_p(1).$$

Proof. Write $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{j,t-\ell}\tilde{y}_{t-1} = \sum_{b=0}^{\ell-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{j,t-\ell}\Delta\tilde{y}_{t-b-1} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{j,t-\ell}\tilde{y}_{t-\ell-1}$, where we define $\tilde{y}_t = 0$ for $t \leq 0$. The first term is clearly $O_p(1)$, being composed of sample averages of stationary variables. The second term is $O_p(1)$ by Lemma G.2. □

Lemma G.4. *$D_T^{-1}\sum_{t=1}^T w_t w_t' D_T^{-1} \xrightarrow{d} \operatorname{diag}\left(1, \{(1+\tau)^2+\sigma_2^2\}\int_0^1 W(r)^2\,dr\right)$, where $W(\cdot)$ is a standard Brownian motion.*

Proof. By Lemma G.1, it suffices to show that $D_T^{-1}\sum_{t=1}^T \tilde{w}_t\tilde{w}_t' D_T^{-1}$ converges to the stated limiting distribution. The $(1,1)$ element converges in probability to 1 by the law of large numbers. The $(1,2)$ element is $O_p(T^{-1/2})$ by Lemma G.3. Finally, the convergence in distribution of the $(2,2)$ element follows from Phillips & Solo (1992, Thm. 3.4), the continuous mapping theorem, and the long-run variance of the process $(\varepsilon_{1,t}+\tau\varepsilon_{1,t-1}+\varepsilon_{2,t})$ being equal to $(1+\tau)^2+\sigma_2^2$. □

Recall the definitions of $\theta_{h,T}$, $\hat{\beta}_h$, $\hat{\delta}_h$, $\hat{A}$, $\hat{\kappa}$, and the unit vector e_j from Section 2.

Lemma G.5. *For any $0 \le \ell \le h$, define $b_{\ell,h} \equiv \mathbb{1}(\ell > 0) + \tau\mathbb{1}(0 < \ell < h)$. Then*

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{T}\sum_{t=2}^{T-h}\sum_{\ell=0}^{h}(b_{\ell,h}\varepsilon_{1,t+\ell} + \varepsilon_{2,t+\ell})\varepsilon_{1,t} + o_p(T^{-1/2}). \tag{G.1}$$

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 Lemmas G.1, G.2 and G.4, it is straightforward to show that

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

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

$$\begin{aligned}\hat{\beta}_h &= \theta_{h,T} + \frac{\frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_{h,T}\hat{\varepsilon}_{1,t})\hat{\varepsilon}_{1,t}}{\frac{1}{T}\sum_{t=2}^{T-h}\hat{\varepsilon}_{1,t}^2} \\ &= \theta_{h,T} + \frac{\frac{1}{T}\sum_{t=2}^{T-h}(y_{t+h} - \theta_{h,T}\varepsilon_{1,t} - \tau\varepsilon_{1,t-1} - y_{t-1})\hat{\varepsilon}_{1,t}}{\frac{1}{T}\sum_{t=2}^{T-h}\hat{\varepsilon}_{1,t}^2}.\end{aligned} \tag{G.2}$$

Lemmas G.1 and G.4 and $D_T\hat{b} = O_p(1)$ 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 (G.2), which we decompose as

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

We first show that the first term above equals the right-hand side of (G.1). Iteration on the

DGP (1) implies

$$y_{t+h} - \theta_{h,T}\varepsilon_{1,t} - \tau\varepsilon_{1,t-1} - y_{t-1} = \sum_{\ell=0}^{h}(b_{\ell,h}\varepsilon_{1,t+\ell} + \varepsilon_{2,t+\ell}) + \frac{\alpha}{\sqrt{T}}\sum_{\ell=0}^{h}\varepsilon_{1,t+\ell-2}\mathbb{1}(\ell \neq 2).$$

The desired conclusion then follows from

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

which can be easily verified with Chebyshev's inequality, using that $\varepsilon_{1,t+\ell}\varepsilon_{1,t}$ is a serially uncorrelated process for $\ell \neq 0$.

We finish the proof by showing that the second term in (G.3) is $O_p(T^{-1})$, i.e.,

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

This follows from $D_T\hat{b} = O_p(1)$ and Lemmas G.1 to G.3. □

Lemma G.6. *Define* $A_0 \equiv \left(\begin{smallmatrix} 0 & 0 \\ \tau & 1 \end{smallmatrix}\right)$. *We have*

$$\hat{A} - A_0 = \frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix}\varepsilon_{1,t-1}e_1' + o_p(T^{-1/2}),$$
$$\hat{\kappa} - 1 = \frac{1}{T}\sum_{t=2}^{T}\varepsilon_{1,t}\varepsilon_{2,t} + o_p(T^{-1/2}).$$

Proof. By appealing repeatedly to Lemmas G.1 to G.4, we get

$$\begin{aligned}
\hat{A} - A_0 &= \left(\frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} + \frac{\alpha}{\sqrt{T}}\varepsilon_{1,t-2} \end{pmatrix} w_{t-1}'D_T^{-1}T^{1/2}\right)\left(D_T^{-1}\sum_{t=2}^{T}w_{t-1}w_{t-1}'D_T^{-1}\right)^{-1}D_T^{-1}T^{1/2} \\
&= \left(\frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix}\tilde{w}_{t-1}'D_T^{-1}T^{1/2} + \frac{\alpha}{T}e_2\sum_{t=2}^{T}\varepsilon_{1,t-2}\tilde{w}_{t-1}'D_T^{-1} + O_p(T^{-1})\right)\{e_1e_1' + o_p(1)\} \\
&= \left(\frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix}\tilde{w}_{t-1}'D_T^{-1}T^{1/2} + O_p(T^{-1})\right)\{e_1e_1' + o_p(1)\} \\
&= \left(\frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix}(\varepsilon_{1,t-1}e_1' + T^{-1/2}\tilde{y}_{t-1}e_2') + O_p(T^{-1})\right)\{e_1e_1' + o_p(1)\}
\end{aligned}$$

$$= \frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix}\varepsilon_{1,t}\\ \varepsilon_{1,t}+\varepsilon_{2,t}\end{pmatrix}\varepsilon_{1,t-1}e_1' + \frac{1}{T^{3/2}}\sum_{t=2}^{T}\begin{pmatrix}\varepsilon_{1,t}\\ \varepsilon_{1,t}+\varepsilon_{2,t}\end{pmatrix}\tilde{y}_{t-1}e_2' \times o_p(1) + o_p(T^{-1/2})$$
$$= \frac{1}{T}\sum_{t=2}^{T}\begin{pmatrix}\varepsilon_{1,t}\\ \varepsilon_{1,t}+\varepsilon_{2,t}\end{pmatrix}\varepsilon_{1,t-1}e_1' + O_p(T^{-1/2}) \times o_p(1) + o_p(T^{-1/2}).$$

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 G.5 applied to $h = 0$. □

G.2 Proof of Proposition 1

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

LP. It follows from Lemma G.5 and a standard martingale central limit theorem that

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

where

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

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. In fact, we only require the Jacobians with respect to $\text{vec}(Ae_1)$ and κ, since Lemma G.6 implies that the second column of $\hat{A}$ is super-consistent. The Jacobians should be evaluated at $\text{plim}\,\hat{A} = A_0 \equiv \left(\begin{smallmatrix}0 & 0\\ \tau & 1\end{smallmatrix}\right)$ and $\text{plim}\,\hat{\kappa} = 1$. Thus, γ should be evaluated at $\gamma_0 \equiv (1, 1)'$.

First, for $h \geq 2$, the Jacobian with respect to Ae_1 equals (Magnus & Neudecker, 2007, p.

208)

$$\begin{aligned}\left.\frac{\partial e_2' A^h \gamma}{\partial \operatorname{vec}(Ae_1)'}\right|_{A=A_0,\gamma=\gamma_0} &= (\gamma' \otimes e_2') \sum_{j=1}^{h} \left((A_0')^{h-j} \otimes A_0^{j-1} \right) (e_1 \otimes I) \\ &= (\gamma' \otimes e_2') \sum_{j=1}^{h} \{ (A_0')^{h-j} e_1 \} \otimes A_0^{j-1} \\ &= (\gamma' \otimes e_2')(e_1 \otimes A_0 + 0) \\ &= (\gamma' e_1 \otimes e_2' A_0) \\ &= (\tau, 1),\end{aligned}$$

where the third equality uses that A_0 is an idempotent matrix, and $A_0' e_1 = 0$. For $h = 1$, the Jacobian with respect to $\operatorname{vec}(Ae_1)$ obviously equals $(0, 1)$. So we can write the Jacobian for all $h \geq 1$ as $(\tau \mathbb{1}(h \geq 2), 1)$.

Second, for any $h \geq 1$, the Jacobian with respect to κ equals

$$\left.\frac{\partial e_2' A^h \gamma}{\partial \kappa}\right|_{A=A_0,\gamma=\gamma_0} = e_2' A_0^h e_2 = e_2' A_0 e_2 = 1.$$

Next, Lemma G.6 and a standard martingale central limit theorem imply

$$\sqrt{T}(\hat{A} - A_0)e_1 \xrightarrow{d} N(0, \operatorname{aVar}(\hat{A}e_1)), \quad \sqrt{T}(\hat{\kappa} - 1) \xrightarrow{d} N(0, \operatorname{aVar}(\hat{\kappa})),$$

where

$$\begin{aligned}\operatorname{aVar}(\hat{A}e_1) &= \operatorname{Var}\left(\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix} \varepsilon_{1,t-1} \right) = \begin{pmatrix} 1 & 1 \\ 1 & 1 + \sigma_2^2 \end{pmatrix}, \\ \operatorname{aVar}(\hat{\kappa}) &= \operatorname{Var}(\varepsilon_{1,t}\varepsilon_{2,t}) = \sigma_2^2.\end{aligned}$$

Moreover, $\hat{A}e_1$ and $\hat{\kappa}$ are asymptotically independent by Lemma G.6, since

$$\operatorname{Cov}\left(\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} + \varepsilon_{2,t} \end{pmatrix} \varepsilon_{1,t-1}, \varepsilon_{1,t}\varepsilon_{2,t} \right) = 0.$$

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

$$\sqrt{T}(\hat{\delta}_h - \theta_{h,T}) = \sqrt{T}(\hat{\delta}_h - e_2' A^h \gamma) + \sqrt{T}(e_2' A^h \gamma - \theta_{h,T}) \xrightarrow{d} N(\operatorname{aBias}_{\mathrm{VAR}}, \operatorname{aVar}_{\mathrm{VAR}}),$$

where, for $h \geq 1$,

$$\begin{aligned}
\text{aBias}_{\text{VAR}} &= \sqrt{T}(e_2' A^h \gamma - \theta_{h,T}) = -\alpha \mathbb{1}(h \geq 2), \\
\text{aVar}_{\text{VAR}} &= (\tau \mathbb{1}(h \geq 2), 1) \begin{pmatrix} 1 & 1 \\ 1 & 1+\sigma_2^2 \end{pmatrix} \begin{pmatrix} \tau \mathbb{1}(h \geq 2) \\ 1 \end{pmatrix} + \sigma_2^2 \\
&= (1 + \tau \mathbb{1}(h \geq 2))^2 + 2\sigma_2^2.
\end{aligned}$$

□

References

Barigozzi, M., Lippi, M., & Luciani, M. (2021). Large-dimensional Dynamic Factor Models: Estimation of Impulse–Response Functions with I(1) cointegrated factors. *Journal of Econometrics*, *221*(2), 455–482.

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.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR Inference: Recommendations for Practice. *Journal of Business & Economic Statistics*, *36*(4), 541–559.

Magnus, J. & Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (3rd ed.). Wiley Series in Probability and Statistics. John Wiley & Sons.

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.

Pope, A. L. (1990). Biases of Estimators in Multivariate Non-Gaussian Autoregressions. *Journal of Time Series Analysis*, *11*(3), 249–258.

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.