Online Appendix for "Local Projections or VARs? A Primer for Macroeconomists"

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

January 23, 2025

This online appendix contains supplemental material on: the comparison of LP and VAR estimation results in prior empirical work (Supplemental Appendix C); implementation details for the LP and VAR estimators that we consider in our simulations (Supplemental Appendix D); a detailed description of our simulation study set-up (Supplemental Appendix E); and further simulation results (Supplemental Appendix F).

Appendix C VARs vs. LPs in empirical work

We describe how we construct the point estimate and standard error comparison of LPs and VARs in existing applied work in Figure 3.1. Our implementation closely follows our earlier work in Montiel Olea, Plagborg-Møller, Qian, and Wolf (2024, Online Appendix C), which in turn is based on the literature summary in Ramey (2016).

We consider four applications in which the researcher has access to a direct measure of a structural shock: to monetary policy, taxes, government spending, and technology. She then estimates the dynamic causal effects of these macroeconomic shocks using either LPs or the equivalent (internal-instrument) recursive VAR. The choice of shocks, outcomes, controls, and lags is exactly the same as in our earlier work, as is the computation of standard errors. Overall we obtain LP and VAR impulse response point estimates and standard errors for 385 impulse responses, across all shocks, outcome variables, and horizons. For each we compute standard error ratios and point estimate differences, scaled by the VAR standard error. We finally split the impulse responses into short horizons ($\leq$ one year, 84 observations overall) and long horizons (> one year, 301 observations overall), and report our results for standard error ratios and scaled point estimate differences as boxplots.

Appendix D Estimation method details

We provide implementation details for the LP and VAR estimator variants that we consider in our simulations in Sections 5.2 and 6.4.

D.1 LPs

Recall from Section 2 that the LP estimator of the horizon-h impulse response is the coefficient $\hat{\beta}_h$ in the h-step-ahead regression

$$y_{t+h} = \hat{\mu}_h + \hat{\theta}_h^{\text{LP}} x_t + \hat{\gamma}_h' r_t + \sum_{\ell=1}^{p} \hat{\delta}_{h,\ell}' w_{t-\ell} + \hat{\xi}_{h,t}. \tag{D.1}$$

For our different structural identification schemes the outcome, impulse, and control variables are selected as follows:

1. **Observed shock.** x_t is the observed shock $\varepsilon_{1,t}$, there are no contemporaneous controls r_t, and the outcome variable y_t is selected at random from the observed data series. The lagged controls are either the shock plus the full five-dimensional vector of observables, or only the shock plus the outcome of interest.

2. **Recursive identification.** x_t is the policy variable (i.e., either the federal funds rate or government purchases), r_t contains all the variables ordered before x_t in the structural shock identification scheme, and w_t always contains the full vector of observables, consistent with the invertibility assumption.

We consider both least-squares as well as bias-corrected versions of the LP estimators. For bias correction we follow Herbst and Johannsen (2024), using their approximate analytical bias formula for LPs with controls.[D.1] The lag length p is either set exogenously (in our simulations typically to $p = 4$) or selected the Akaike Information Criterion (AIC) for the equivalent reduced-form VAR. Standard errors and confidence intervals are computed using

[D.1] We substitute population autocovariances with sample analogues. We implement an iterative bias correction, with the impulse response estimate at horizon h bias-corrected using the previously corrected impulse response estimates at horizons $1, 2, \ldots, h-1$.

either the delta method (assuming homoskedasticity) or equal-tailed percentile-t bootstrap (with 500 draws).

D.2 VARs

Recall from Section 2 that the VAR impulse response estimator is based on the reduced-form VAR

$$w_t = \hat{c} + \sum_{\ell=1}^{p} \hat{A}_\ell w_{t-\ell} + \hat{u}_t. \tag{D.2}$$

with $\text{Var}(u_t) = \hat{B}\hat{B}'$ where $\hat{B}$ is the Cholesky decomposition of the estimated forecast error variance-covariance matrix. Our different identification schemes are implemented as follows:

1. **Observed shock.** w_t contains the observed shock as well as either all other five observed series, or only the outcome variable of interest. The observed shock is ordered first in the recursive orthogonalization of the reduced-form innovations.

2. **Recursive identification.** w_t consists of the five observed series, ordered as indicated in our discussion of the structural monetary and fiscal shock identification schemes (see Supplemental Appendix E.2). We do not consider a small version of this system, consistent with the invertibility assumption.

The reduced-form VAR coefficient matrices are estimated either using least-squares or with the analytical bias correction of Pope (1990), following the recommendations in Kilian (1998). The lag length p is either set exogenously (in our simulations typically to $p = 4$) or selected using the AIC for (D.2). Standard errors and confidence intervals are computed using either the delta method (assuming homoskedasticity) or Efron's bootstrap (500 draws), following recommendations of Kilian and Lütkepohl (2017).

Appendix E Simulation study details

E.1 DFM estimation

We estimate the encompassing stationary and non-stationary DFMs on the data set of Stock and Watson (2016), proceeding exactly as in Li, Plagborg-Møller, and Wolf (2024). For the stationary DFM, we follow the same steps as in Online Appendix F.2 (p.16) of Li, Plagborg-Møller, and Wolf, which in turn replicates the original analysis by Stock and Watson as well as in Lazarus, Lewis, Stock, and Watson (2018). For the non-stationary DFM, we transform variables, select the number of factors and lags, and estimate the factor VECM as in Online Appendix C (pp.2–4) of Li, Plagborg-Møller, and Wolf.

E.2 DGP selection and impulse response estimands

We draw our individual DGPs from the two encompassing DFMs by proceeding as follows. For all DGPs, we restrict attention to the following 17 oft-used series (with Stock and 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).* We then draw several random combinations of five series from this overall set of salient time series, subject to the constraint that each DGP contains either the federal funds rate or government spending (for monetary or fiscal shock estimands, respectively, as discussed further below) as well as at least one real activity series (categories 1–3) and one price series (category 6). For both the stationary and non-stationary DFM we draw 100 monetary and 100 fiscal DGPs in this way.

Observed shock identification. We define the monetary policy shock as the (unique) linear combination of the innovations in the factor equation that maximizes the impact impulse response of the federal funds rate, and analogously for the fiscal policy shock, with government spending as the maximized response variable. We then assume that the econometrician directly observes this shock of interest, together with the five observables drawn from the list of salient time series, as we discussed above. She estimates the propagation of the shock using either LPs with the shock as the impulse variable (and no contemporaneous controls) or a recursive VAR with the observed shock ordered first. The outcome of interest is randomly selected among the observable series, not including the fiscal or monetary policy instruments (i.e., government purchases or the federal funds rate).

We also estimate LP and VAR variants with a smaller set of observables. Here the system only contains the observed shock as well as the outcome variable of interest.

Recursive identification. For recursive identification, the researcher only observes the five time series drawn from the encompassing DFM. We then define as the object of interest impulse responses with respect to a recursive orthogonalization of the reduced-form (Wold) forecast errors in the VAR(∞) representation of the observables. For monetary policy, we order the federal funds rate last and then call the orthogonalized innovation to that variable a monetary policy shock, restricting all other variables to not respond contemporaneously to monetary policy, as in Christiano, Eichenbaum, and Evans (1999). For fiscal policy, we order government spending first and then call the innovation to that variable a fiscal policy shock, thus restricting the fiscal authority to respond to all of the other innovations with a lag, following Blanchard and Perotti (2002). The overall identifying assumptions here are thus invertibility together with particular temporal orderings. See Li, Plagborg-Møller, and Wolf (2024, Online Appendix D) for further details.

E.3 Population estimands

For our visual illustration of LP-VAR equivalence in Figure 2.1 we consider one particular monetary policy DGP from the stationary DFM, with recursive shock identification. The observables in our system are (Stock and Watson Data Appendix series # in brackets): the unemployment rate (56); real GDP (1); the core consumer price index (121); the BAA 10-year spread (151); and the federal funds rate (142). The outcome of interest is unemployment. We then simulate a large sample, and use recursive LPs and VARs with $p \in \{2, 6, 12\}$ lags to estimate the propagation of the recursively identified monetary shock, as discussed above. We compare these finite-lag LP and VAR estimands with the true population projection on the recursively identified monetary policy innovation, which we estimate using a numerical approximation to an infinite-lag VAR.

E.4 Degree of mis-specification

Given a DGP—i.e., a five-variable (for recursive identification) or six-variable (for observed shock identification) system randomly drawn from the encompassing DFM—and a lag length p, we can represent that DGP as a VARMA(p,∞), following the same steps as those outlined in Montiel Olea, Plagborg-Møller, Qian, and Wolf (2024, Footnote 8). For any given sample size T and degree of local mis-specification ζ, we can then map that VARMA(p,∞) into the locally mis-specified VAR(p) framework of Montiel Olea, Plagborg-Møller, Qian, and Wolf, obtaining in particular the misspecification lag polynomial $\alpha(L)$. We then obtain the total degree of misspecification as

$$\sqrt{T \times \mathcal{M}} = \|\alpha(L)\| \equiv \sqrt{\sum_{\ell=1}^{\infty} \text{trace}\{D\alpha_\ell' D^{-1}\alpha_\ell\}},$$

where D is the variance-covariance matrix of the contemporaneous innovations.

Appendix F Supplementary simulation results

We provide several supplementary simulation results to complement our main findings reported in Sections 5.2 and 6.4. We report detailed results for: mean squared error; recursive identification; monetary and fiscal shocks considered separately; and finally coverage together with confidence interval width.

F.1 MSE

To complement the bias and variance plots in Figure 5.1, we in Figure F.1 show mean squared error results, again in the observed shock case, averaging over fiscal and monetary shocks, and separately for stationary and non-stationary DGPs. Consistent with the discussion in Section 5.2 and in particular Lesson 7 we see that, quantitatively, the variance cost necessary for LP to attain its low bias is substantial, and so—in terms of mean squared error—VAR estimation methods tend to dominate. Looking across stationary and non-stationary DGPs, VARs with somewhat extended lag length (here $p = 4$) tend to attain a reasonable balance between bias and variance, attaining substantially (for stationary DGPs) or moderately (for non-stationary DGPs) lower MSEs than LPs at medium and long horizons.

F.2 Recursive identification

While our headline results in Sections 5.2 and 6.4 are reported for observed shock identification, broadly similar lessons for both the bias-variance trade-off and for inference emerge under recursive shock identification. Visual summaries are provided in Figures F.2 and F.3.

Figure F.2 shows the bias-variance trade-off, now averaging across both monetary and fiscal shocks and across stationary and non-stationary DGPs. We see the same patterns as for observed shock identification: the bias-variance trade-off is stark, long lag lengths for the VAR align it with LPs, and the variance cost of LPs is large. Differently from the observed shock case we do not report any results for "small" specifications, simply because the control

MSE, OBSERVED SHOCK, STATIONARY & NON-STATIONARY

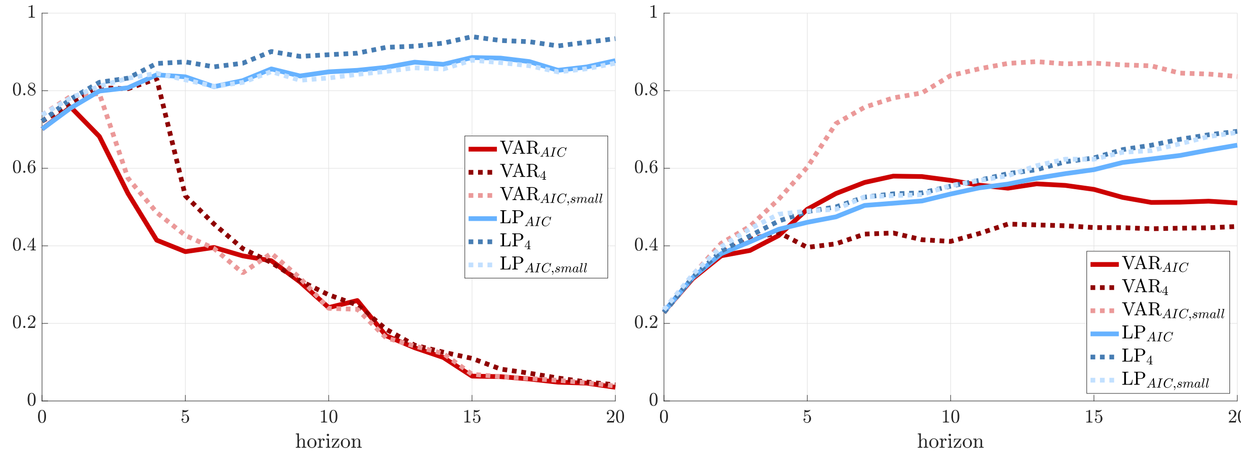
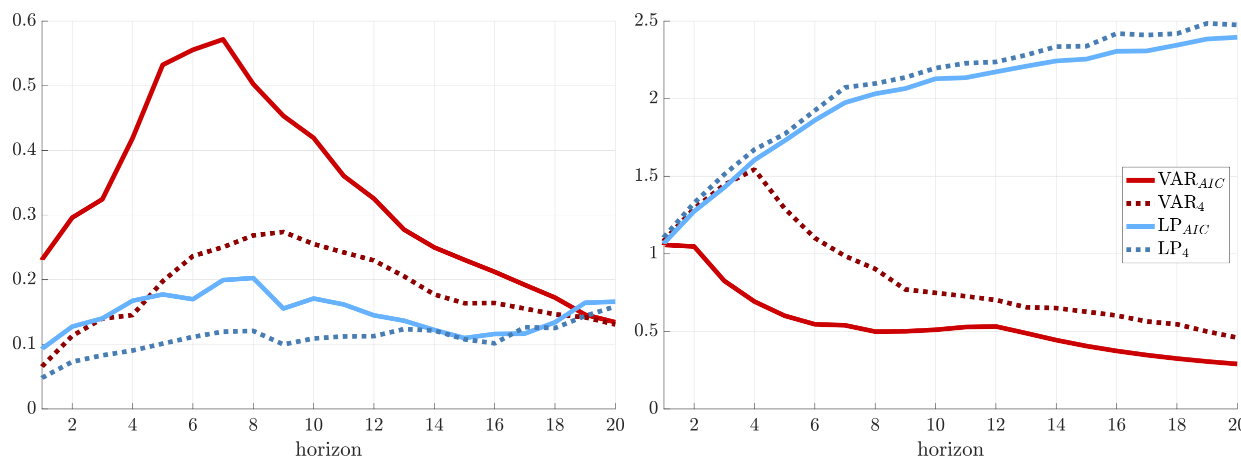


Figure F.1: Median (across DGPs) of mean squared error $\text{MSE}(\hat{\theta}_h)$ of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$, for the stationary DGPs (left panel) and the non-stationary DGPs (right panel).

vector is now integral to the economic identifying assumptions.

RECURSIVE IDENTIFICATION: BIAS-VARIANCE TRADE-OFF



BIAS

STANDARD DEVIATION

Figure F.2: Median (across DGPs) of absolute bias $\left|\mathbb{E}\left[\hat{\theta}_h - \theta_h\right]\right|$ (left panel) and standard deviation $\sqrt{\text{Var}(\hat{\theta}_h)}$ (right panel) of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Figure F.3 displays coverage results, again across all DGPs. As in the observed shock case, VARs with lag length selected by standard information criteria significantly undercover, Recursive identification: CI coverage

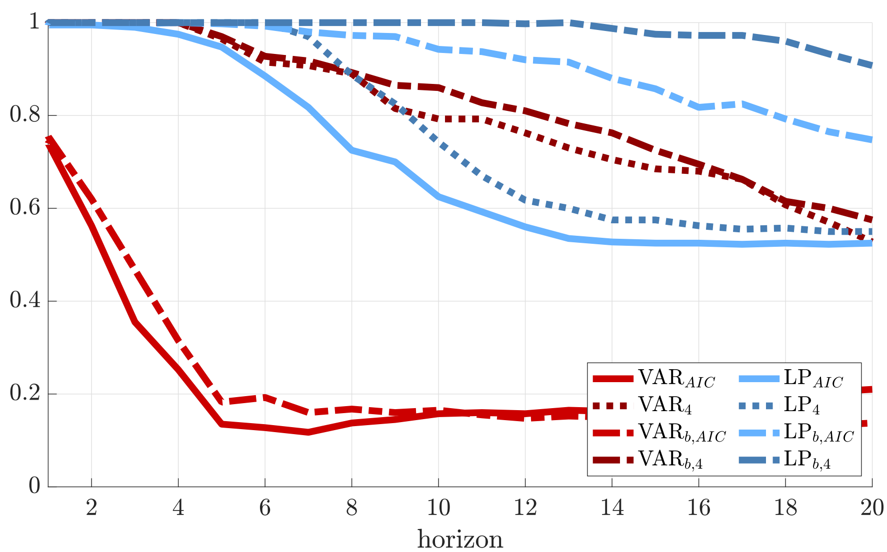


Figure F.3: Fraction of DGPs (both stationary and non-stationary) for which the confidence interval coverage probability is above 80 per cent, by estimation procedure; the b subscript in the figure legend indicates bootstrap standard errors.

while LPs—in particular with bootstrapped confidence intervals—cover well. We now also see that inclusion of longer lag lengths is more important for LP, simply because those lags are important now for the identification scheme, ensuring that the shock of interest is indeed spanned by the regression residuals.

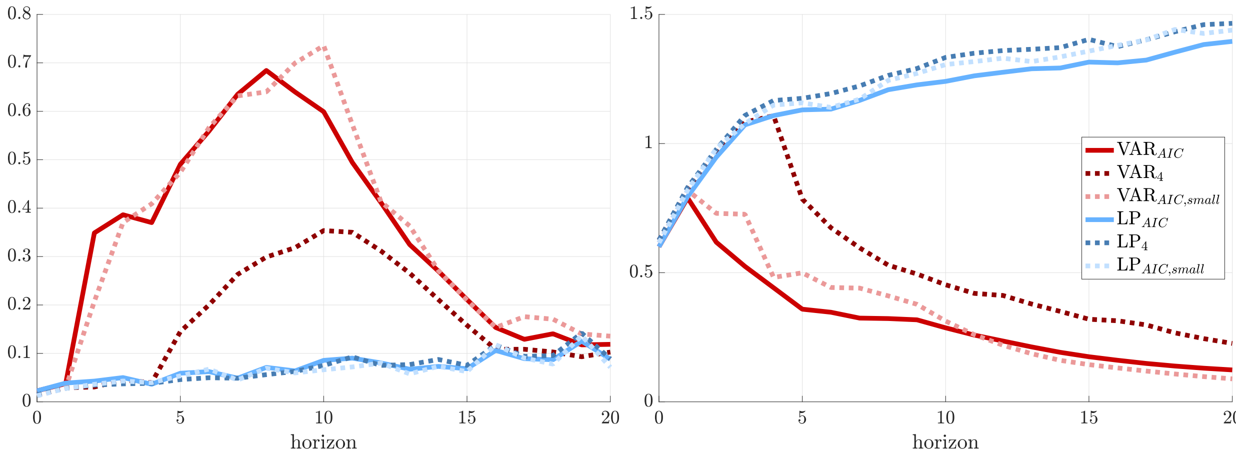
F.3 Fiscal and monetary shocks

Our reported conclusions are not sensitive to the type of (policy) shock that we consider. To establish this, Figures F.4 and F.5 show bias, variance, and coverage results for fiscal shocks (across stationary and non-stationary DGPs), while Figures F.4 and F.5 do the same for monetary shocks. The figures by shock echo the messages of our main figures, which average across shocks: there is a meaningful bias-variance trade-off; the variance cost of LPs is high; and only LP methods robustly attain high coverage.

F.4 Coverage probability and CI width

While in Section 6.4 we report the fraction of LP and VAR confidence intervals with coverage above 80%, Figure F.8 here instead shows the *average* coverage probability (left panel) and median confidence interval length (right panel) of our different estimators across DGPs. The

Observed fiscal shock: bias-variance trade-off



Bias

Standard deviation

Figure F.4: Median (across DGPs) of absolute bias $\left|\mathbb{E}\left[\hat{\theta}_h - \theta_h\right]\right|$ (left panel) and standard deviation $\sqrt{\text{Var}(\hat{\theta}_h)}$ (right panel) of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

picture that emerges is again consistent with our theoretical and practical messages: VAR confidence intervals are quite a bit shorter, but this comes at the cost of (sometimes material) under-coverage, in particular at longer horizons. Long-lag VARs attain correct coverage only at the cost of confidence intervals essentially as wide as those of LP.

References

Blanchard, O., and R. Perotti (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. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?," in *Handbook of Macroeconomics*, ed. by J. Taylor, and M. Woodford, vol. 1, chap. 2, pp. 65–148. Elsevier.

Herbst, E. P., and B. K. Johannsen (2024): "Bias in local projections," *Journal of Econometrics*, 240(1), 105655.

Kilian, L. (1998): "Small-sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80(2), 218–230.

Observed fiscal shock: CI coverage

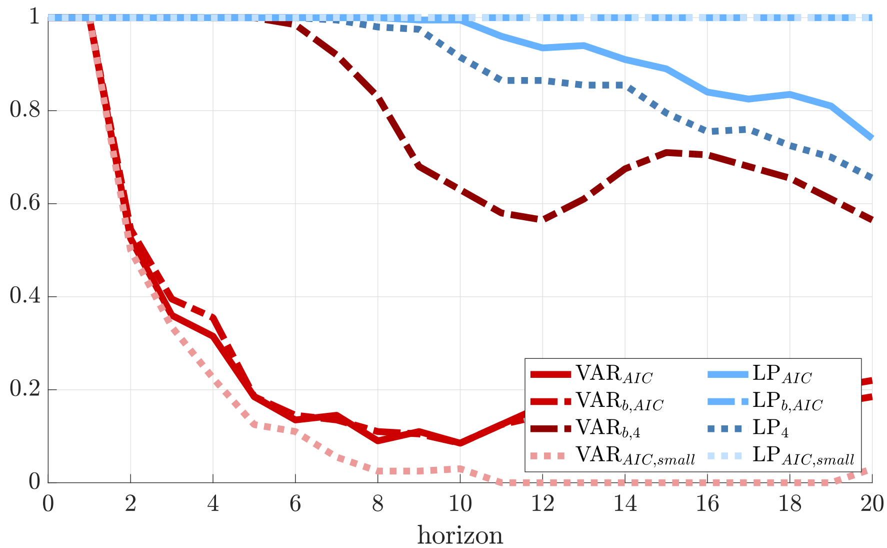
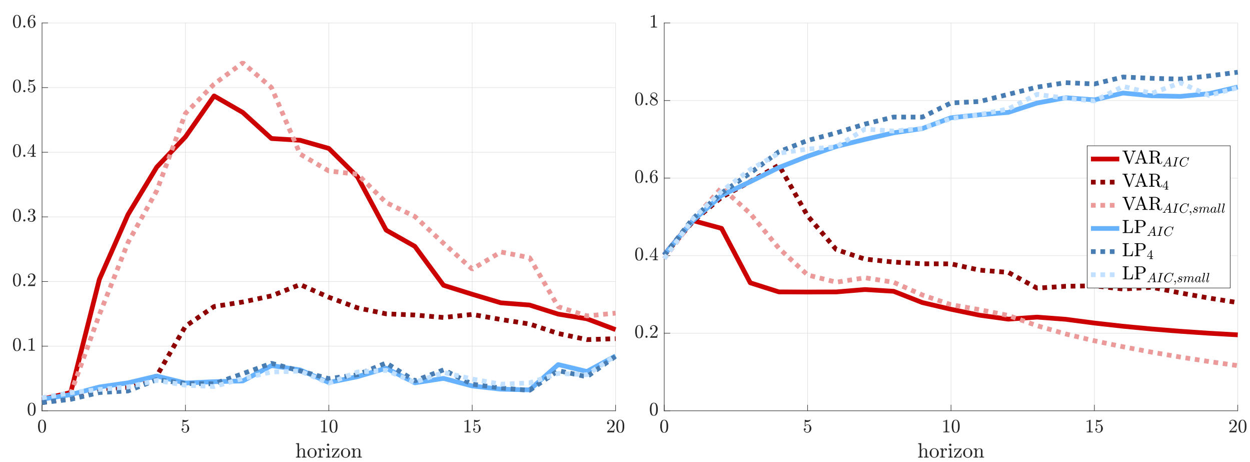


Figure F.5: Fraction of DGPs (both stationary and non-stationary) for which the confidence interval coverage probability is above 80 per cent, by estimation procedure; the b subscript in the figure legend indicates bootstrap standard errors.

Observed monetary shock: bias-variance trade-off



Bias

Standard deviation

Figure F.6: Median (across DGPs) of absolute bias $\left|\mathbb{E}\left[\hat{\theta}_h - \theta_h\right]\right|$ (left panel) and standard deviation $\sqrt{\text{Var}(\hat{\theta}_h)}$ (right panel) of the different estimation procedures, relative to $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$.

Kilian, L., and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Observed monetary shock: CI coverage

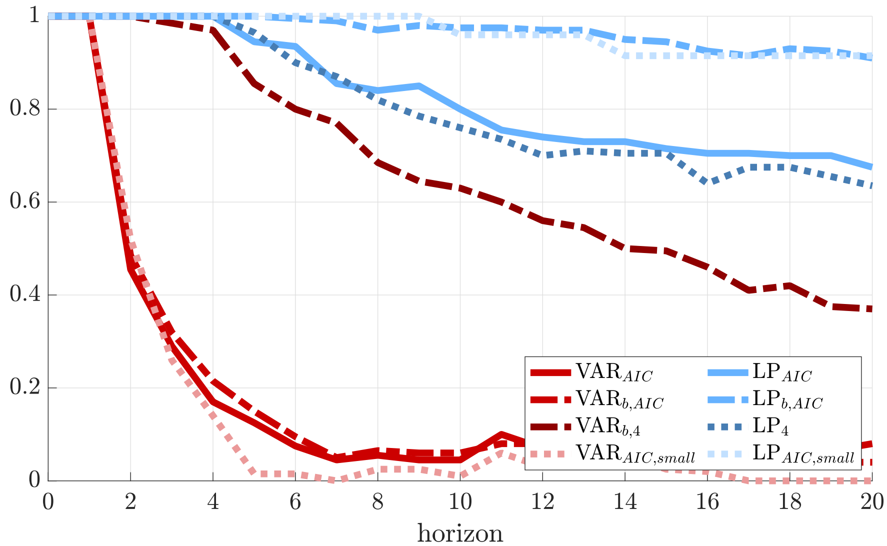


Figure F.7: Fraction of DGPs (both stationary and non-stationary) for which the confidence interval coverage probability is above 80 per cent, by estimation procedure; the b subscript in the figure legend indicates bootstrap standard errors.

Observed shock: CI coverage probability & width

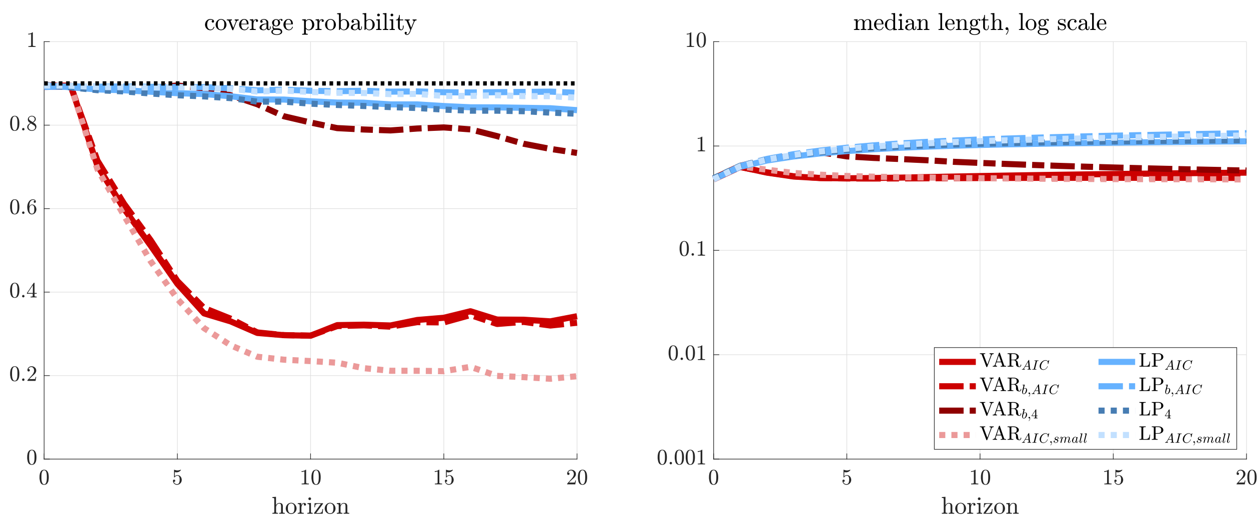


Figure F.8: Coverage probabilities (left panel) and median confidence interval length (right panel) for VAR (red) and LP (blue) confidence intervals, by estimation procedure, across all DGPs (both stationary and non-stationary); the b subscript in the figure legend indicates bootstrap standard errors.

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

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024): "Local projections vs. VARs: Lessons from thousands of DGPs," *Journal of Econometrics*, 244(2), 105722, Themed Issue: Macroeconometrics.

Montiel Olea, J. L., M. Plagborg-Møller, E. Qian, and C. K. Wolf (2024): "Double Robustness of Local Projections and Some Unpleasant VARithmetic," Working Paper 32495, National Bureau of Economic Research.

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

Ramey, V. A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 2, pp. 71–162. Elsevier.

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