Local Projections or VARs?
A Primer for Macroeconomists

José Luis Montiel Olea
Cornell

Mikkel Plagborg-Møller
Princeton

Eric Qian
Princeton

Christian K. Wolf
MIT

August 7, 2025

Dynamic causal effects in macro

- Key objects in applied macro: structural impulse responses (dynamic causal effects).

$$\theta_h = E[y_{t+h} \mid \varepsilon_{1t} = 1] - E[y_{t+h} \mid \varepsilon_{1t} = 0], \quad h = 0, 1, 2, \ldots$$

  - Not a forecast. Shock ε_{1t} may only explain small fraction of variation.

- Estimation methods: vector autoregressions (VARs) and local projections (LPs).

  1. VAR: iterate on dynamic multivariate model. Sims (1980, 21.5k cites)
  2. LP: direct regression of future outcome y_{t+h} on current covariates. Jordà (2005, 4.5k cites)

This talk: LPs or VARs?

- Literature synthesis of core principles guiding the choice between LP and VAR:
  1. LP & VAR are two *estimation* methods, $\perp$ to questions of identification.
  2. Must navigate a stark bias-variance trade-off:
     - LP: low bias, high variance.
     - VAR (few lags): potentially high bias, low variance. More lags $\Rightarrow$ closer to LP.
  3. For reliable uncertainty assessments, choose (a) LP or (b) VAR with very many lags.
- Provide recommendations for practical implementation of LP.

Outline

Local projection

- LP: linear regression, separately for each horizon $h = 0, 1, 2, \ldots$:

$$y_{t+h} = \mu_h + \theta_h^{\text{LP}} x_t + \gamma_h' r_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' w_{t-\ell} + \xi_{h,t}.$$

  - y_t: outcome, x_t: "impulse", r_t: contemporaneous controls, $w_t = (r_t', x_t, y_t, q_t')'$: all data.
  - This is a projection, not a generative model.

- **Shock:** by FWL theorem, LP estimates impulse response of y_{t+h} with respect to

$$\tilde{x}_t = x_t - \text{proj}(x_t \mid r_t, w_{t-1}, \ldots, w_{t-p}).$$

  Economically interesting? Requires identifying assumptions.

  - E.g., $\tilde{x}_t =$ narrative shock (Romer x 2) or Taylor rule residual (Christiano, Eichenbaum & Evans).

- **Projection:** LP uses autocorrelations in the data out to the horizon h of interest.

Vector autoregression

- VAR: estimate reduced-form multivariate dynamic model

$$w_t = c + A_1 w_{t-1} + A_2 w_{t-2} + \cdots + A_p w_{t-p} + u_t.$$

- Orthogonalize $u_t = H\varepsilon_t$. For now, assume H lower triangular (recursive/Cholesky id'n).

- Structural impulse responses $\Psi_h = \partial w_{t+h}/\partial \varepsilon_t'$ from iterative propagation:

$$\Psi_0 = H, \quad \Psi_1 = A_1\Psi_0, \quad \Psi_2 = A_1\Psi_1 + A_2\Psi_0, \quad \ldots \quad \Psi_h = \sum_{\ell=1}^{\min\{p,h\}} A_\ell \Psi_{h-\ell},$$

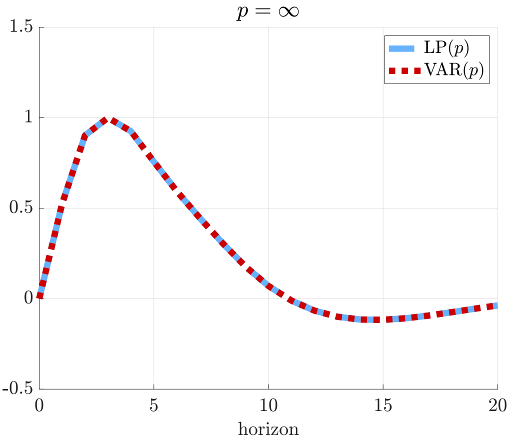
$$\theta_h^{\text{VAR}} = \partial y_{t+h}/\partial \varepsilon_{x,t} = e_y' \Psi_h e_x.$$

- **Shock:** residual in projection of $u_{x,t} = e_x' u_t$ on $u_{r,t} = e_r' u_t$. Same as LP shock $\tilde{x}_t$!

- **Projection:** VAR matches first p autocovariances of the data, but extrapolates to longer horizons $h > p$.

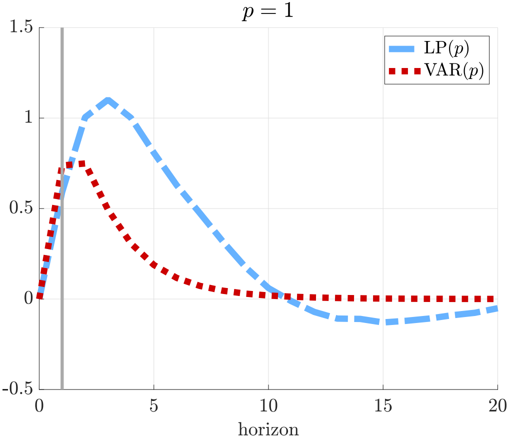
LP = VAR with very long lag length

$p = \infty$: same shock, same projection, so same impulse responses



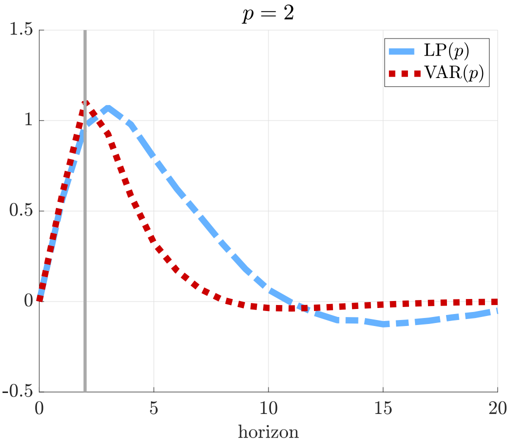
LP $\approx$ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



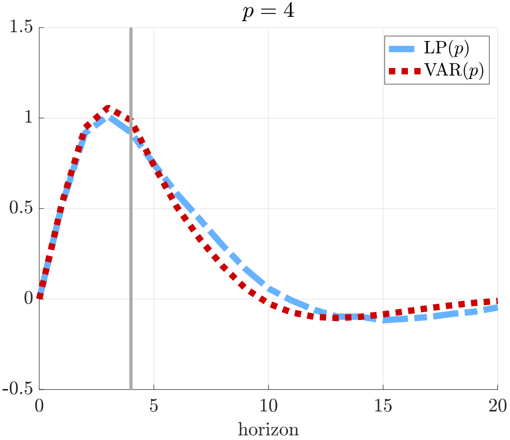
LP $\approx$ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



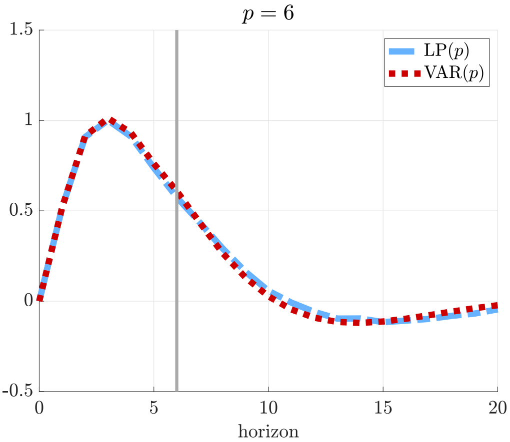
LP $\approx$ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



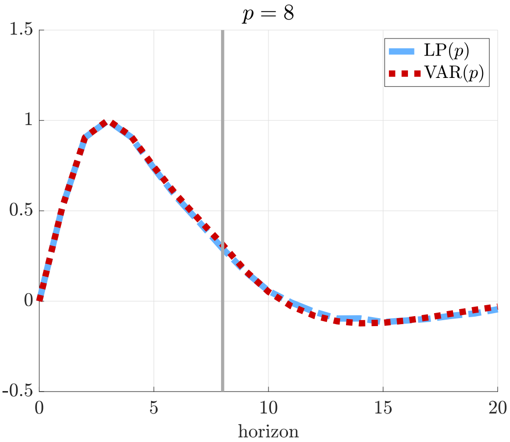
LP $\approx$ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



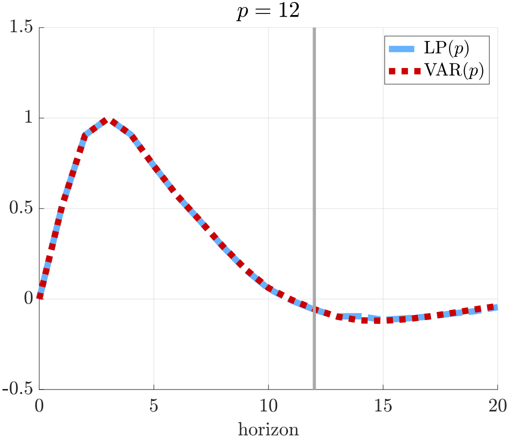
LP $\approx$ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



LP ≈ VAR up to horizon p

$p < \infty$: same shock so same responses at $h = 0$,
approx'ly same for $0 < h \leq p$, but then VAR extrapolates



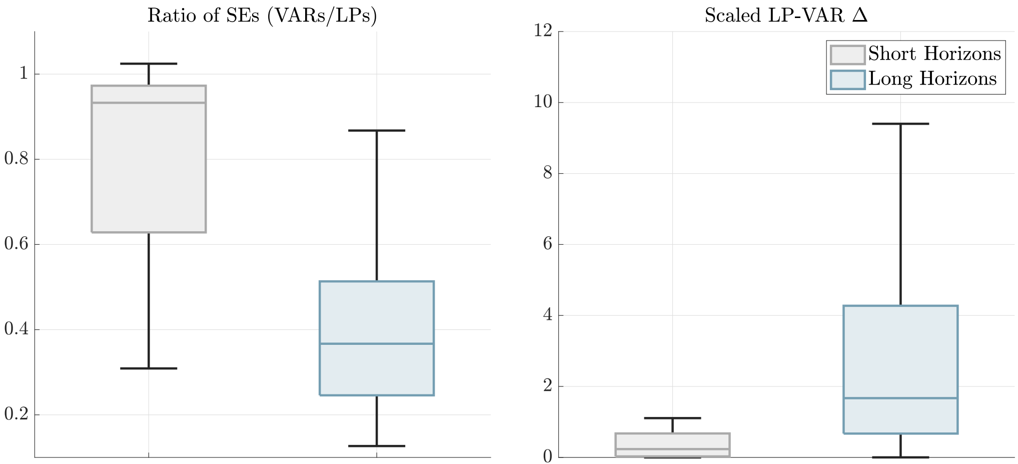
LPs and VARs share the same estimand

- Have only considered recursive identif'n so far.
- But equivalence extends to more complicated identification schemes.
  - External instruments/proxies, long-run restrictions, sign restrictions, ...
  - Intuition: "shock" is still just some (potentially complicated) f'n of autocovariances of the data. With many lags, both LP and VAR approximate these well in large samples.
- Take-away: LP vs. VAR debate $\perp$ questions of identification.
  - Only difference is how a finite data set is exploited to estimate the common estimand.

Outline

1. Identification
2. Bias-variance trade-off
3. Uncertainty assessments
4. Conclusion

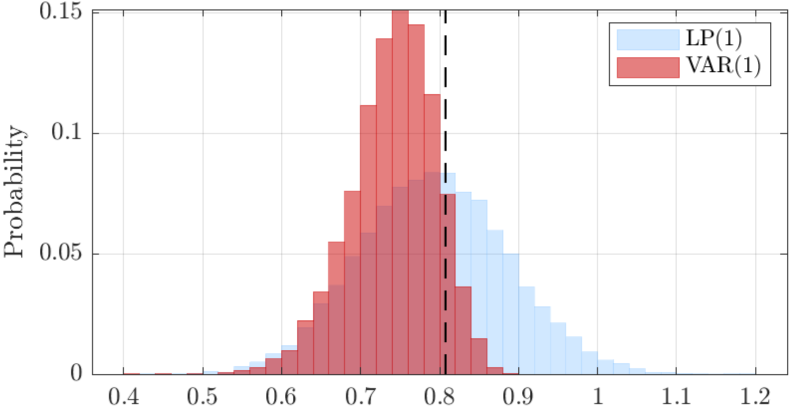
VAR vs. LP in finite samples



Replication of 4 empirical applications in Ramey (2016), total of 385 impulse responses

Illustrative simulation

$$y_t = \rho y_{t-1} + \varepsilon_t + \alpha \varepsilon_{t-1}, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$$



$h = 2,\ \rho = 0.85,\ \alpha = 0.1,\ T = 240$

Analytics of the bias-variance trade-off

- Consider a structural VAR model contaminated by small MA terms:

$$w_t = A_1 w_{t-1} + \cdots + A_{p_0} w_{t-p_0} + H\left(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \ldots\right).$$

  - Why? Low-order VARs are known to deliver good forecasts, but not literal truth.
  - Technically, assume $\alpha_\ell \propto$ std. dev. of VAR estimator.
- In this environment, estimators should control for infinitely many lags. Infeasible.
- Suppose both LP & VAR use $p \geq p_0$ estimation lags. Then in large samples,

$$\hat{\theta}_h^{\text{VAR}} \overset{\cdot}{\sim} N\left(\theta_h + b_h(p), \tau^2_{h,\text{VAR}}(p)\right), \quad \hat{\theta}_h^{\text{LP}} \overset{\cdot}{\sim} N\left(\theta_h, \tau^2_{h,\text{LP}}\right).$$

  - Benefit and cost of extrapolation: VAR more efficient ($\tau^2_{h,\text{VAR}}(p) \leq \tau^2_{h,\text{LP}}$) but biased.
  - $h \leq p - p_0$: VAR bias $b_h(p) = 0$ and variance coincide with LP.

How bad can the VAR bias be in theory?

- Both LP & VAR require controlling for the most important predictors/lags. But LP is robust to omitting moderately important ones, while VAR is not.
- Theoretical bound on bias: letting $\mathcal{M}$ denote the fraction of the variance of the MA residual that's due to lagged terms,

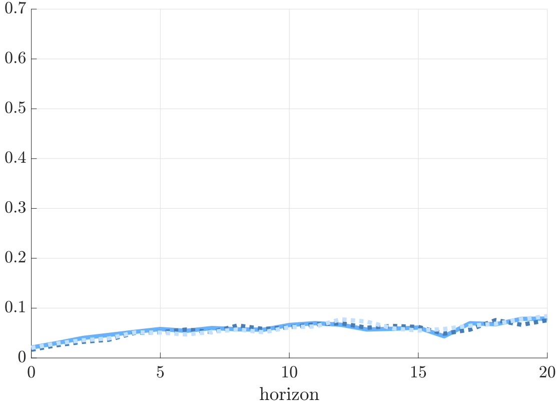
$$|b_h(p)| \leq \sqrt{T \times \mathcal{M} \times \left\{\tau^2_{h,\text{LP}} - \tau^2_{h,\text{VAR}}(p)\right\}},$$

  and there exist MA coefficients that attain the bound.
    - Example: if $T = 100$, $\mathcal{M} = 1\%$, $\tau_{h,\text{VAR}}(p)/\tau_{h,\text{LP}} = 0.5$, then bias can be $1.73 \times$ SE.
- No free lunch for VARs: if precision gain is large, then so is the potential bias.
    - VAR only robust if we use so many lags that VAR = LP.

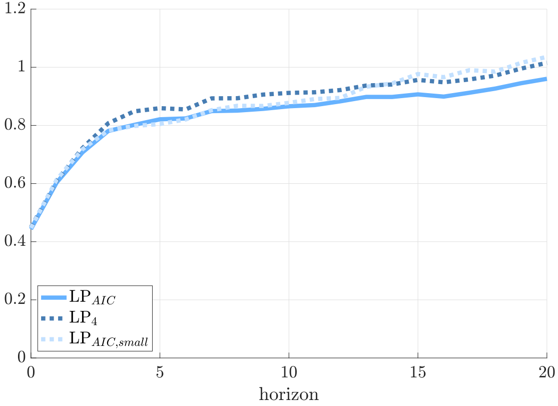
The bias-variance trade-off in practice

- Conduct large-scale simulation study. Extends Li, Plagborg-Møller & Wolf (2024)
- DGP: extension of Stock-Watson dynamic factor model fitted to 207 macro series.
  - Both stationary and non-stationary versions.
  - To be useful for applied work, an econometric procedure should at least work well here.
- Construct 100s of specifications:
  - Randomly draw subsets of 5 salient macro series from the DFM. Outcome y_t chosen at random from this list.
  - Additionally, econometrician observes a monetary/fiscal shock (in paper: recursive identif'n).
- Simulate data with $T = 240$, then estimate LPs, VARs, and several variants.

Simulation evidence: bias and standard deviation



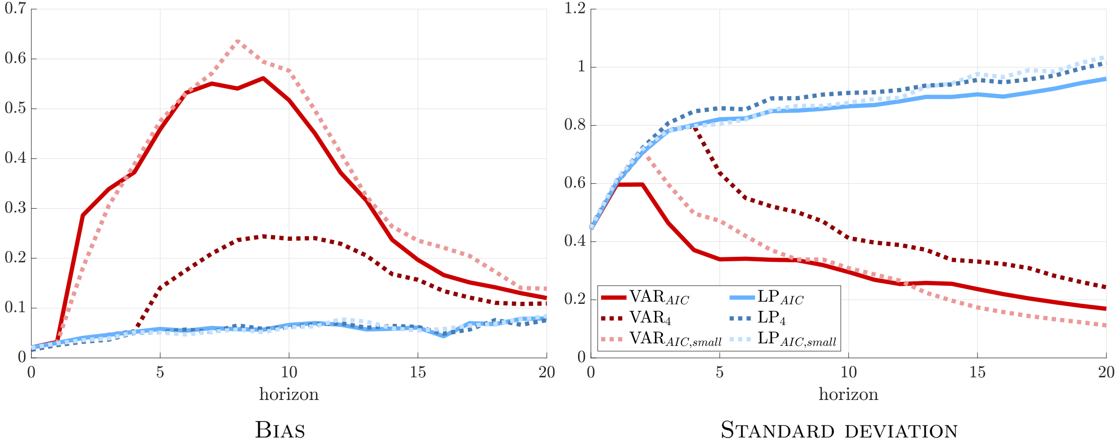
BIAS



STANDARD DEVIATION

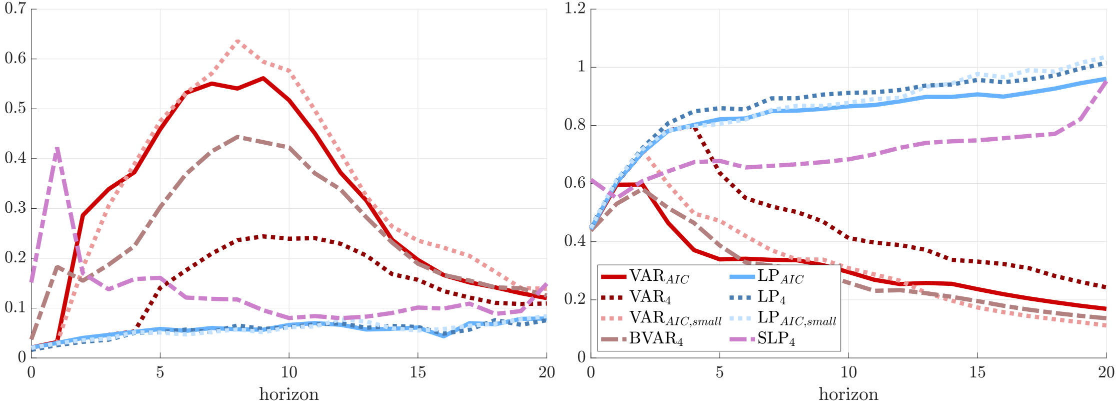
average across 200 stationary and 200 non-stationary DGPs

Simulation evidence: bias and standard deviation



average across 200 stationary and 200 non-stationary DGPs

Simulation evidence: bias and standard deviation

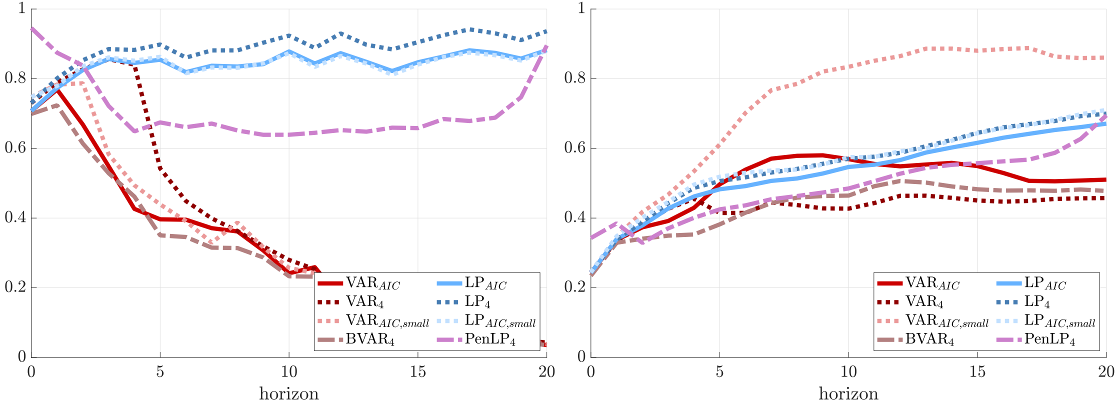


Bias Standard deviation

average across 200 stationary and 200 non-stationary DGPs

MSE loss: (B)VAR preferred over LP on average

Conventional way to trade off bias and variance: $\text{MSE} = \text{bias}^2 + \text{variance}$



MSE FOR STATIONARY DGPs

MSE FOR NON-STATIONARY DGPs

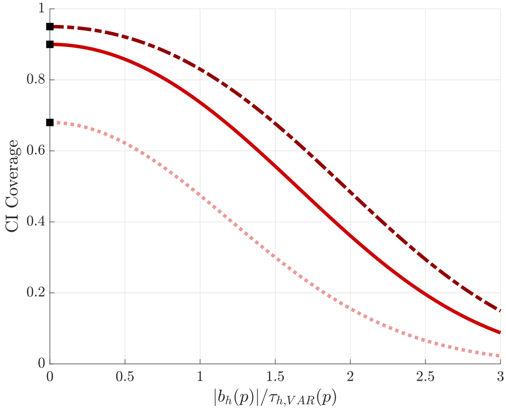
Bias-variance trade-off: recap

- Take-away: bias-variance trade-off is stark in practice.
- Robustness of LP to dynamic misspecification comes at significant variance cost.
- Under MSE loss, VAR is preferred over LP in the *average* simulation DGP.
  - Shrinkage (penalized LP or BVAR) often preferred over OLS.
- But MSE only evaluates the accuracy of the point estimate. This is not worth much without an accompanying uncertainty assessment.

Outline

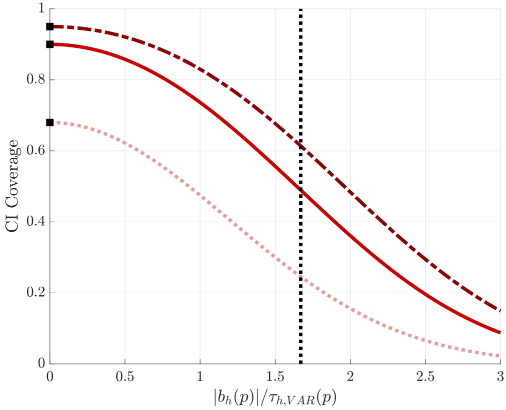
Uncertainty assessments: bias is costly

- Conventional to summarize uncertainty using confidence interval.
- Want coverage probability close to (say) 90% *regardless* of true DGP (not just for avg DGP!).
- Challenge for VARs: bias is really costly for coverage. CI has correct width, but off-center.
- Remember: easy to get worst-case bias $\approx 1.73 \times$ SE.

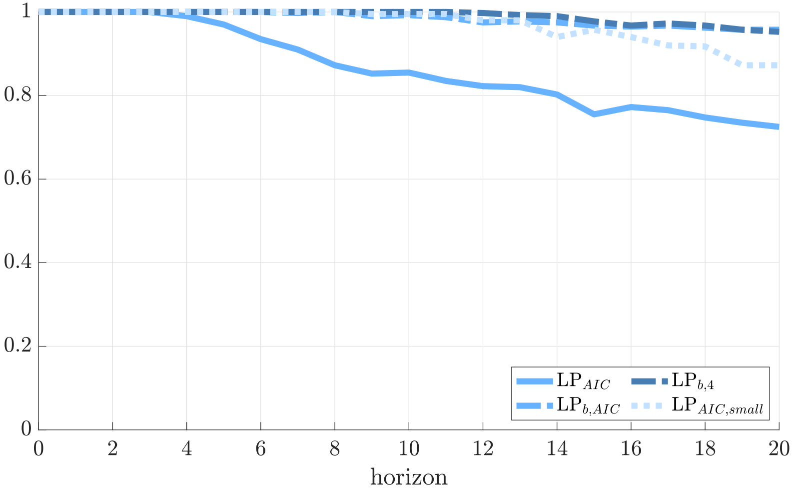


Uncertainty assessments: bias is costly

- Conventional to summarize uncertainty using confidence interval.
- Want coverage probability close to (say) 90% *regardless* of true DGP (not just for avg DGP!).
- Challenge for VARs: bias is really costly for coverage. CI has correct width, but off-center.
- Remember: easy to get worst-case bias $\approx 1.73 \times$ SE.

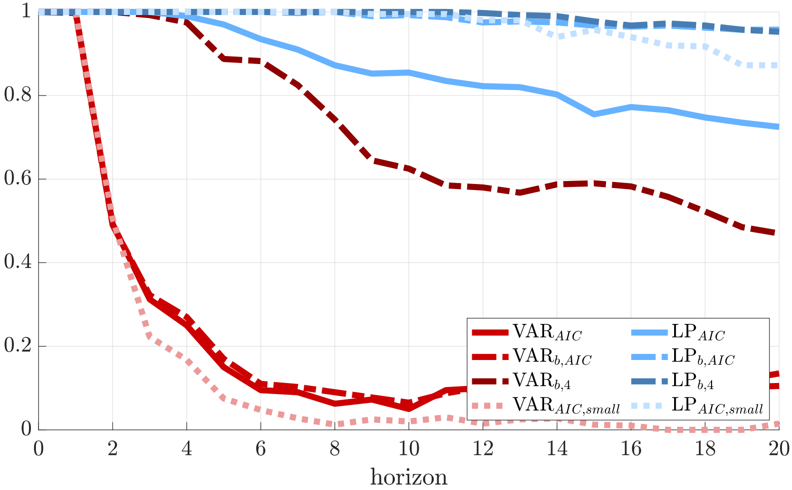


Simulation evidence: confidence interval coverage



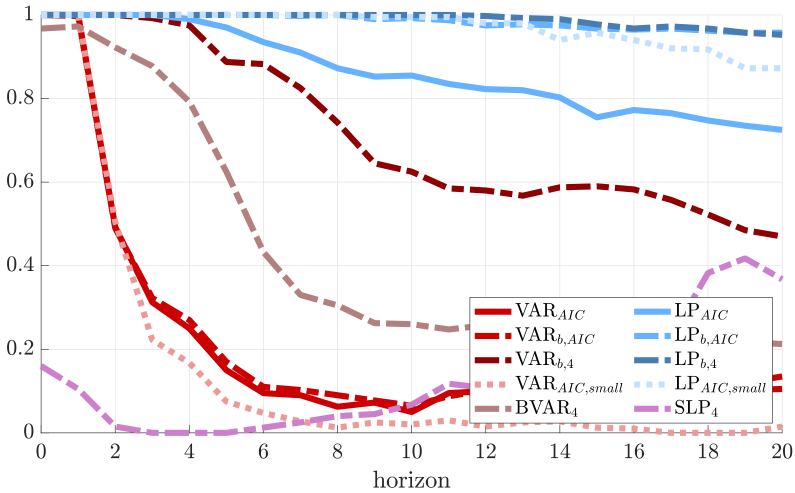
FRACTION OF DGPS WITH COVERAGE $\geq 80\%$ (TARGET COVERAGE 90%)

Simulation evidence: confidence interval coverage



FRACTION OF DGPS WITH COVERAGE $\geq 80\%$ (TARGET COVERAGE 90%)

Simulation evidence: confidence interval coverage



Fraction of DGPs with coverage $\geq 80\%$ (target coverage 90%)

Outline

Summary of take-aways

1. Choice of VAR vs. LP $\perp$ identification.
2. Stark bias-variance trade-off.
    - LP robust to dynamic misspecification (low bias), but comes at significant variance cost.
    - MSE loss: VAR (or BVAR) preferred for the avg DGP.
    - Here "VAR" $=$ conventional number of lags (e.g., AIC/BIC).
3. Only LP (or VAR with very many lags) yield uncertainty assessments that are reliable across a wide range of DGPs.

- Comparison extends beyond VARs: *no* procedure can be more efficient than LP without sacrificing robustness. P-M & Wolf (2021); Xu (2023)

Practical recommendations

- To analyze what—and how much—the data can say about causal effects, use *either* (a) LPs *or* (b) VARs with very many lags ($\approx$ LP).
  - VARs with conventional lag lengths remain useful for forecasting.
- Guidelines for implementing LP (details in paper):

1. Control for all var's and lags that are strong predictors of either outcome or impulse. OK to omit weak predictors. Can use information criteria as guide.
2. Analytical bias correction. Herbst & Johannsen (2024)
3. Heteroskedasticity-robust SE (no need for Newey-West).
4. For persistent data, report bootstrap CI.

Appendix

Encompassing model

- Dynamic Factor Model (DFM): Stock & Watson (2016)

$$
\begin{aligned}
f_t &= \Phi(L) f_{t-1} + H \varepsilon_t \\
X_t &= \Lambda f_t + v_t \\
v_{i,t} &= \Gamma_i(L) v_{i,t-1} + \Xi_i \xi_{i,t}
\end{aligned}
$$

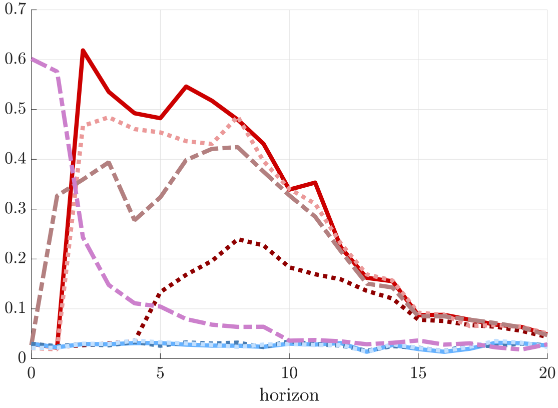
  - f_t: six latent factors, evolve as VECM or VAR, driven by six aggregate shocks ε_t.
  - X_t: 207 quarterly macro time series, spanning various categories.
  - $v_{i,t}$: idiosyncratic noise, evolves as AR(4), independent across i.
- Parameters estimated from quarterly U.S. data. Li, Plagborg-Møller & Wolf (2024)
- New: ARCH processes for the innovations $\{\varepsilon_t, \xi_{i,t}\}$.

Specifications and estimands

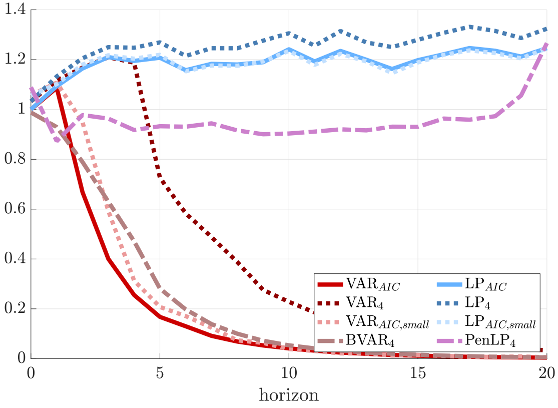
- Draw subsets of 5 variables. DFM implies these follow VAR(∞).
  - Restrict attention to 17 salient series.
  - Spec'n always contains at least one real activity and one price series, + policy instrument (either fed funds rate or gov't spending).
  - Select response variable y_t at random (not policy instrument).
- Estimands for two structural identification schemes:
  1. Observed shock $\varepsilon_{1,t}$: estimand $\theta_h = \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}}$, $h = 0, 1, 2, \ldots, 20$. $H = \frac{\partial f_t}{\partial \varepsilon_t'}$ chosen to maximize impact response of policy instrument wrt. $\varepsilon_{1,t}$.
  2. Recursive: fiscal shock ordered first, monetary shock ordered last.

Additional simulation results: bias and standard deviation

Stationary DGPs:



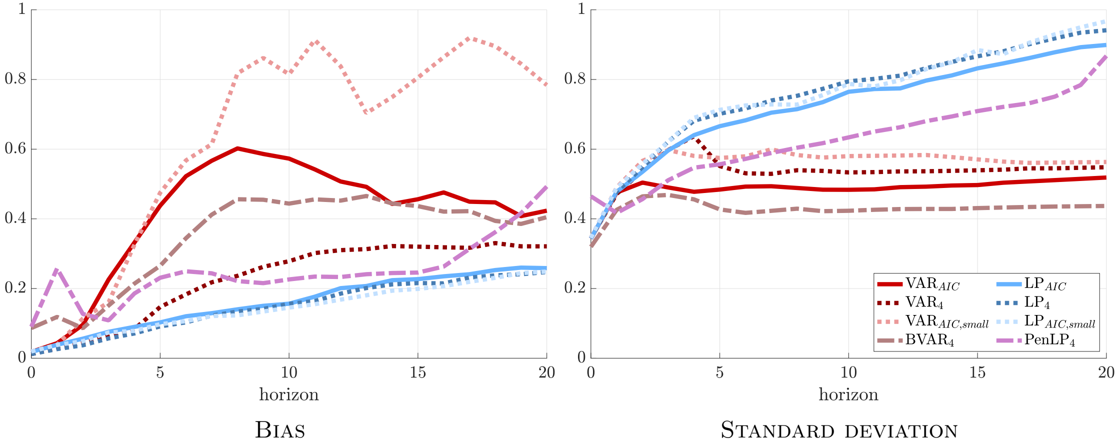
Bias



Standard deviation

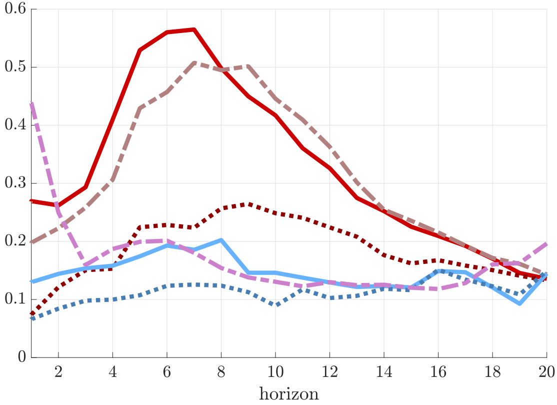
Additional simulation results: bias and standard deviation

Non-stationary DGPs:

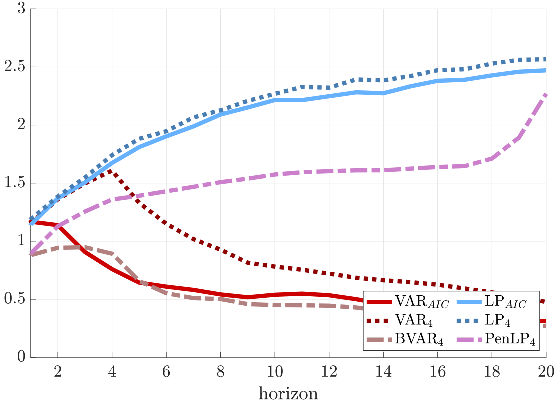


Additional simulation results: bias and standard deviation

Recursive identification:



Bias



Standard deviation

Additional simulation results: confidence interval coverage