# Local Projections vs. VARs: Lessons From Thousands of DGPs


| Dake Li | Mikkel Plagborg-Møller | Christian K. Wolf |
| --- | --- | --- |
| Two Sigma | Princeton University | MIT |


September 12, 2023

# Estimation of IRFs

- How to estimate **impulse response functions (IRFs)** in finite samples?

$$\theta_h \equiv E(y_{t+h} \mid \varepsilon_{j,t} = 1) - E(y_{t+h} \mid \varepsilon_{j,t} = 0), \quad h = 0, 1, 2, \dots$$

1. **Structural Vector Autoregression (VAR)**: Sims (1980)

$$w_t = \sum_{\ell=1}^{p} A_\ell w_{t-\ell} + B\varepsilon_t, \quad \varepsilon_t \sim WN(0, I_n).$$

Extrapolates $\theta_h$ from first $p$ autocovariances. Low variance, potentially high bias.

2. **Local Projections (LP)**: Jordà (2005)

$$y_{t+h} = \beta_h \varepsilon_{j,t} + \text{controls} + \text{residual}_{h,t}, \quad h = 0, 1, 2, \dots$$

Estimates $\theta_h$ from sample autocovariances out to lag $h$. Low bias, high variance.

# LP or VAR?

- Choice of LP or VAR seems to matter for important applied questions. Ramey (2016)
- LP and VAR share same population IRF estimand at horizons $h \leq p$ (lag length). Plagborg-Møller & Wolf (2021)
  - No meaningful trade-off if interest centers on short horizons ...
  - ... or if we choose very large lag length (high variance).
- Applied interest in LP suggests concerns about substantial VAR misspecification at intermediate/long horizons. Justified? Nakamura & Steinsson (2018)
- Analytical guidance is murky: Under local misspecification of VAR($p$) model, bias-variance trade-off depends on numerous aspects of DGP. Schorfheide (2005)

# This paper

- Our approach: Large-scale simulation study of impulse response estimators.
  - Draw 1,000s of DGPs from empirical Dynamic Factor Model. Stock & Watson (2016)
  - Several estimation methods: LP, VAR, bias correction, shrinkage, . . .
  - Several identification schemes: observed shock, recursive, proxy/instrument.
  - Pay attention to researcher's loss function and role of horizon.
- Which estimators perform well on average across many DGPs?

# This paper

- Our approach: Large-scale simulation study of impulse response estimators.
  - Draw 1,000s of DGPs from empirical Dynamic Factor Model. Stock & Watson (2016)
  - Several estimation methods: LP, VAR, bias correction, shrinkage, ...
  - Several identification schemes: observed shock, recursive, proxy/instrument.
  - Pay attention to researcher's loss function and role of horizon.
- Which estimators perform well on average across many DGPs?
  1. Bias-corrected LP preferred method if *and only if* researcher overwhelmingly prioritizes bias.
  2. For researchers who also care about precision, BVAR attractive at short horizons, least-squares VAR at intermediate horizons, similar at long horizons.

# Literature

- Direct vs. iterated forecasts:
  Marcellino, Stock & Watson (2006)
- LP vs. VAR simulation studies:
  Jordà (2005); Meier (2005); Kilian & Kim (2011); Brugnolini (2018); Choi & Chudik (2019); Austin (2020); Bruns & Lütkepohl (2021)
- Analytical comparisons:
  Schorfheide (2005); Kilian & Lütkepohl (2017); Plagborg-Møller & Wolf (2021)
- Shrinkage estimation:
  Hansen (2016); Barnichon & Brownlees (2019); Miranda-Agrippino & Ricco (2021)
- Inference about IRFs:
  Inoue & Kilian (2020); Montiel Olea & Plagborg-Møller (2021); Xu (2023)

# Outline

1. Analytical illustration
2. Data generating processes
3. Estimators
4. Results
5. Conclusion

# Simple analytical example

- Locally misspecified VAR(1) in the data $w_t \equiv (\varepsilon_{1,t}, y_t)'$:

$$y_t = y_{t-1} + \varepsilon_{1,t} + \tau\varepsilon_{1,t-1} + \frac{\alpha}{\sqrt{T}}\varepsilon_{1,t-2} + \varepsilon_{2,t}, \quad (\varepsilon_{1,t}, \varepsilon_{2,t})' \overset{i.i.d.}{\sim} N(0, \text{diag}(1, \sigma_2^2)).$$

- Parameter of interest: $\theta_{h,T} \equiv \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}} = 1 + \tau\mathbb{1}(h \geq 1) + \frac{\alpha}{\sqrt{T}}\mathbb{1}(h \geq 2)$.

- Two estimators (later consider other ones):

  1. **LP**: $y_{t+h} = \hat{\beta}_h \varepsilon_{1,t} + \hat{\zeta}_h' w_{t-1} + \text{residual}_{h,t}$.

  2. **VAR**: $w_t = \hat{A}w_{t-1} + \hat{C}\hat{\eta}_t$, where $\hat{C} =$ Cholesky. Impulse response estimate $\hat{\delta}_h \propto e_2'\hat{A}^h\hat{C}e_1$ normalized so first variable $w_{1,t}$ responds by 1 unit on impact.

- Proposition (building on Schorfheide, 2005): ▸ Prop

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

# How much should we care about bias to pick LP over VAR?

- At horizons $h \in \{0, 1\}$, LP and VAR asy. equivalent. Plagborg-Møller & Wolf (2021)
- At horizons $h \geq 2$:
  - VAR extrapolation causes bias: $|\,\text{aBias}_{\text{VAR},h}\,| = |\alpha| > 0 = |\,\text{aBias}_{\text{LP},h}\,|$.
  - LP less precise: $\text{aVar}_{\text{LP},h} - \text{aVar}_{\text{VAR},h} = 1 + \sigma_2^2 + (h-2)\big[(1+\tau)^2 + \sigma_2^2\big] > 0$.
  - Given "loss function"

$$\mathcal{L}_\omega(\theta_{h,T}, \hat{\theta}_h) = \omega \times \left(\mathbb{E}[\hat{\theta}_h - \theta_{h,T}]\right)^2 + (1-\omega) \times \text{Var}(\hat{\theta}_h),$$

LP preferred over VAR (asymptotically) if and only if

$$\omega \geq \omega_h^* \equiv 1 - \frac{\alpha^2}{\alpha^2 + \text{aVar}_{\text{LP},h} - \text{aVar}_{\text{VAR},h}} \in (0, 1).$$

# Analytical illustration: take-aways

- Even in simple DGP, bias-variance trade-off is non-trivial. Depends on...
  - ... IRF shape $\tau$, importance $\sigma_2^2$ of nuisance shocks, and degree $\alpha$ of misspecification.
  - ... bias weight $\omega$ in loss function.
  - ... impulse response horizon $h$.
- Our approach going forward:
  - Study trade-off through simulations in thousands of empirically calibrated DGPs. Will inform us about empirically relevant "$\tau$", "$\sigma_2^2$", and "$\alpha$".
  - Enrich menu of estimation procedures to trace out bias-variance possibility frontier.
  - Also consider identification schemes that don't require observed shocks.

# Outline

1. Analytical illustration
2. Data generating processes
3. Estimators
4. Results
5. Conclusion

# Encompassing model

- Dynamic Factor Model (DFM):

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

- $X_t$: 207 quarterly macro time series in levels, spanning various categories.
  - Stock & Watson (2016) argue that DFM captures 2nd moments of U.S. data well.
- $f_t$: six latent driving factors, evolve as VAR(4), driven by six aggregate shocks $\varepsilon_t$.
- $v_{i,t}$: idiosyncratic noise, evolves as AR(4), independent across $i$.
- Estimation: PCA on $\Delta X_t$, cumulate factors, VECM for $\hat{f}_t$ with 4 common stoch'c trends. ($H$: next slide.) Gaussian shocks. Bai & Ng (2004); Barigozzi, Lippi & Luciani (2021)

# Lower-dimensional DGPs and estimands

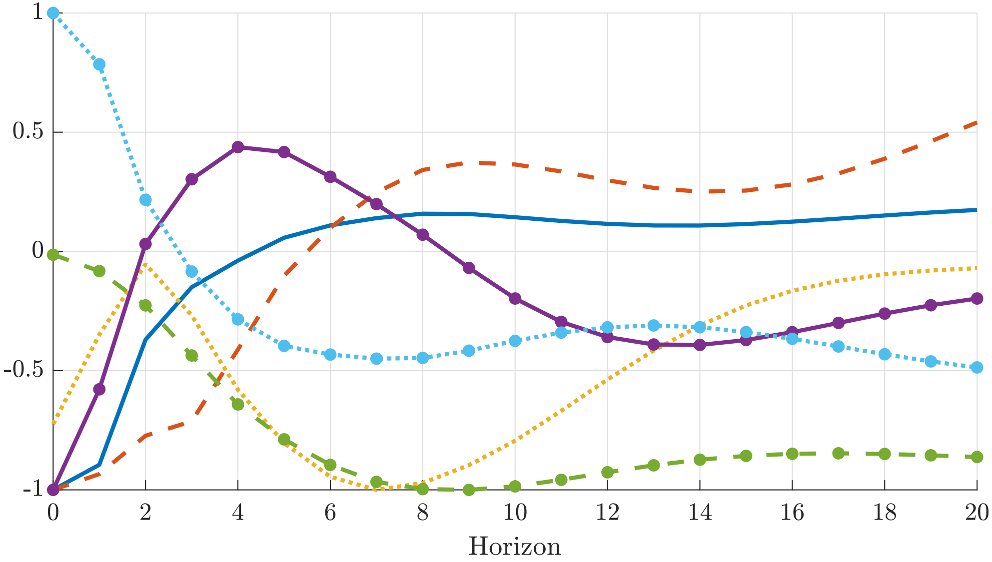
- Draw 6,000 subsets of 5 variables $\bar{w}_t \subset X_t$. DFM implies that $\bar{w}_t$ follows VAR($\infty$).
- $\bar{w}_t$ contains at least one activity and one price series, and – depending on type of DGP...
  1. Monetary shock: $i_t =$ federal funds rate.
  2. Fiscal shock: $i_t =$ federal government spending.
- Select response variable $y_t \in \bar{w}_t$ at random (not $i_t$).
- For today, assume shock $\varepsilon_{1,t}$ is observed. Estimand: $\theta_h = \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}}$, $h = 0, 1, 2, \dots, 20$.
  - In paper: proxy/IV, recursive identification.
- $H = \frac{\partial f_t}{\partial \varepsilon_t'}$ chosen to maximize impact response of $i_t$ wrt. $\varepsilon_{1,t}$.

# DGPs are heterogeneous along various dimensions

| Percentile | min | 10 | 25 | 50 | 75 | 90 | max |
|---|---|---|---|---|---|---|---|
| *Data and shocks* | | | | | | | |
| trace(long-run var)/trace(var)* | 0.03 | 0.27 | 0.54 | 1.02 | 1.98 | 3.54 | 23.73 |
| Fraction of VAR coef's $\ell \geq 5$ | 0.07 | 0.14 | 0.17 | 0.23 | 0.29 | 0.37 | 0.81 |
| Degree of shock invertibility | 0.24 | 0.30 | 0.34 | 0.39 | 0.44 | 0.49 | 0.65 |
| IV first stage F-statistic | 7.18 | 7.91 | 10.55 | 21.13 | 24.20 | 33.29 | 33.97 |
| *Impulse responses up to* $h = 20$ | | | | | | | |
| No. of interior local extrema | 0 | 1 | 2 | 2 | 3 | 3 | 5 |
| Horizon of max abs. value | 0 | 0 | 1 | 4 | 8 | 19 | 20 |
| Average/(max abs. value) | -0.87 | -0.67 | -0.48 | 0.01 | 0.33 | 0.64 | 0.89 |
| $R^2$ in regression on quadratic | 0.04 | 0.46 | 0.70 | 0.85 | 0.95 | 0.98 | 1.00 |

Combining 6,000 monetary and fiscal DGPs. Observed shock or IV identification. *: first diff.

# Impulse response estimands are also heterogeneous



# Outline

1. Analytical illustration
2. Data generating processes
3. Estimators
4. Results
5. Conclusion

# Impulse response estimators

- Local projection methods:
  1. **Least squares**. Jordà (2005)
  2. **Bias-corrected**: corrects $O(T^{-1})$ bias due to persistence. Herbst & Johanssen (2021)
  3. **Penalized**: shrinks towards quadratic polynomial in $h$. Barnichon & Brownlees (2019)
- VAR methods:
  4. **Least squares**.
  5. **Bias-corrected**: corrects $O(T^{-1})$ bias due to persistence. Pope (1990)
  6. **Bayesian**: Minnesota prior favoring cointegration, automatic hyper-parameter selection via marginal likelihood. Giannone, Lenza & Primiceri (2015)
  7. **Model averaging**: Data-dependent weighted average of estimates from 40 models, AR(1) to AR(20) and VAR(1) to VAR(20). Hansen (2016); Miranda-Agrippino & Ricco (2021)

# Outline

1. Analytical illustration
2. Data generating processes
3. Estimators
4. Results
5. Conclusion

# Specification and simulation settings

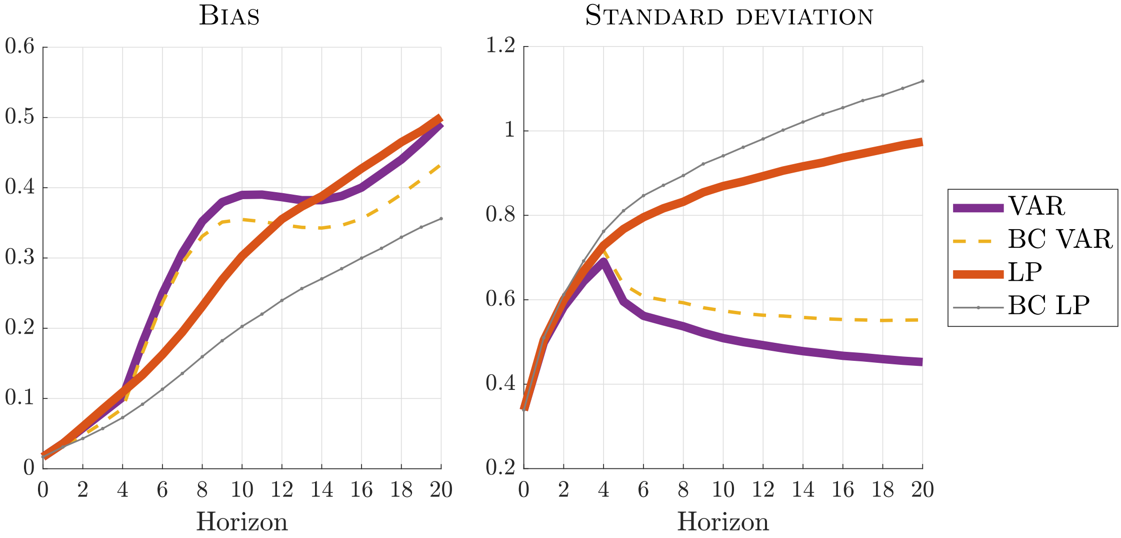
- $p = 4$ lags in LP and VAR, except VAR model averaging.
  - AIC almost always selects fewer than 4 lags.
- Show results for 6,000 monetary and fiscal shock DGPs jointly.
- Loss function:

$$\mathcal{L}_\omega(\theta_h, \hat{\theta}_h) = \omega \times \left(\mathbb{E}[\hat{\theta}_h - \theta_h]\right)^2 + (1 - \omega) \times \text{Var}(\hat{\theta}_h).$$

Divide estimator bias/std by $\sqrt{\frac{1}{21}\sum_{h=0}^{20}\theta_h^2}$ to remove units.

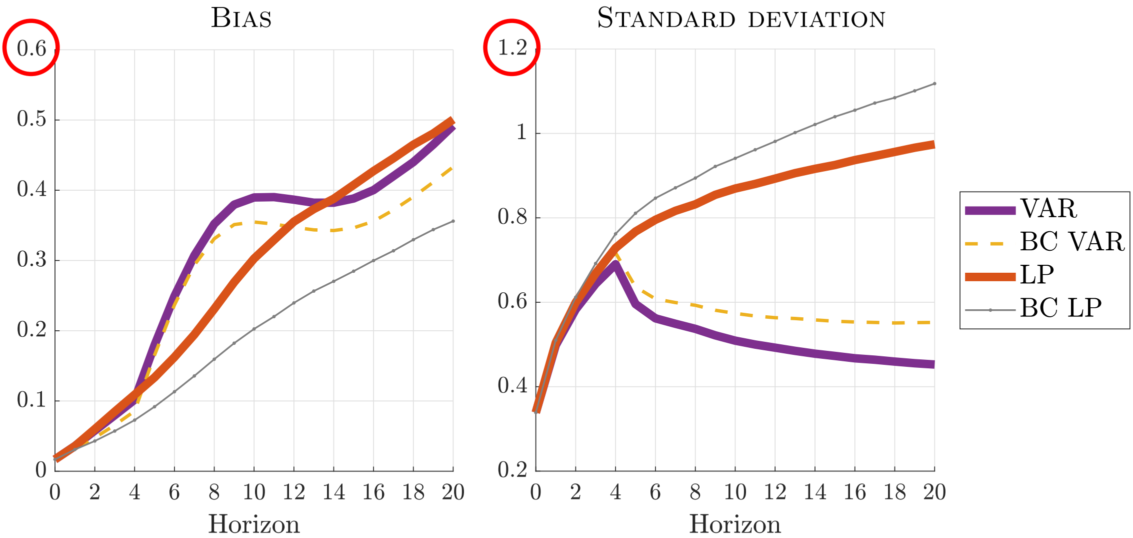
- $T = 200$. 5,000 Monte Carlo repetitions per DGP.
  - Simulations take about a week in Matlab on research cluster with 300 parallel cores.

# #1: Clear bias-variance trade-off between LP & VAR after bias-corr'n



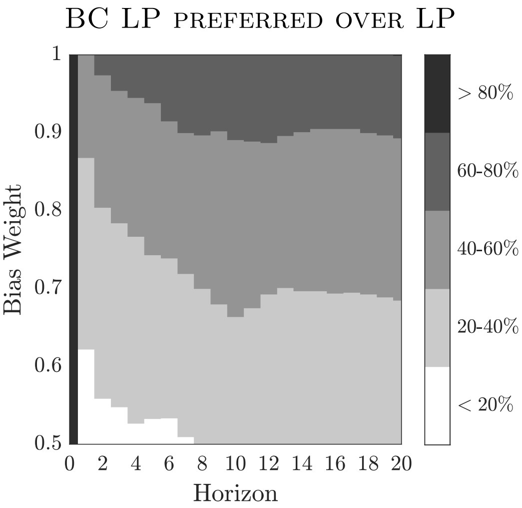

Observed shock identification, medians across 6,000 DGPs

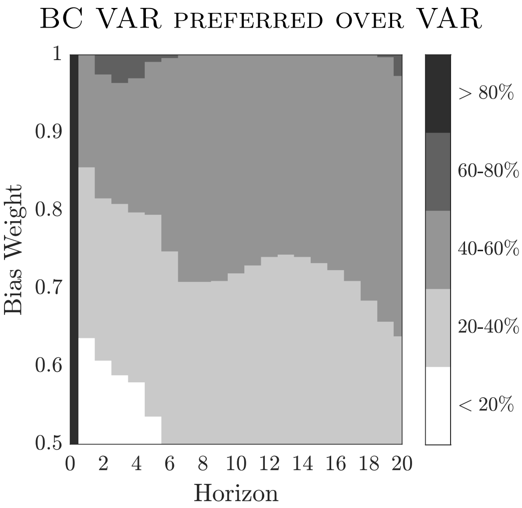
# #1: Clear bias-variance trade-off between LP & VAR after bias-corr'n




Observed shock identification, medians across 6,000 DGPs

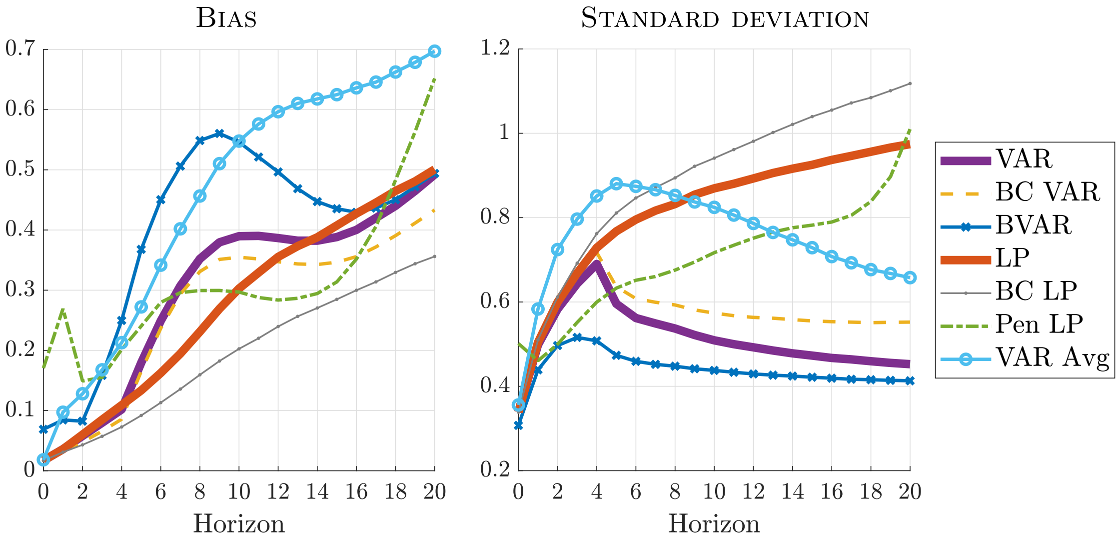
# . . . but bias-correction is not a free lunch



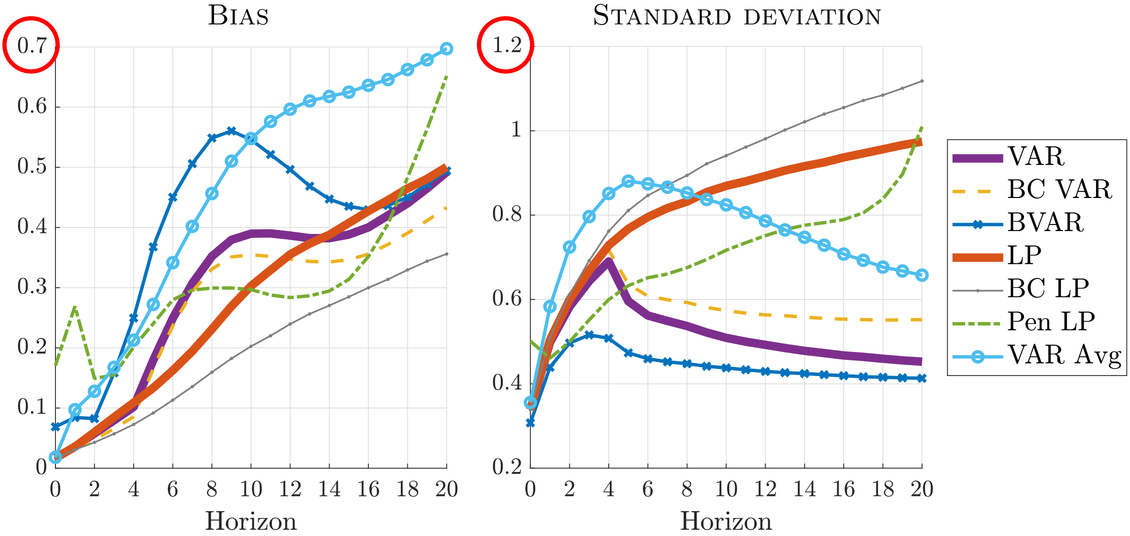

Observed shock identification

# #2: Bias-corr'd LP best iff. researcher heavily prioritizes bias



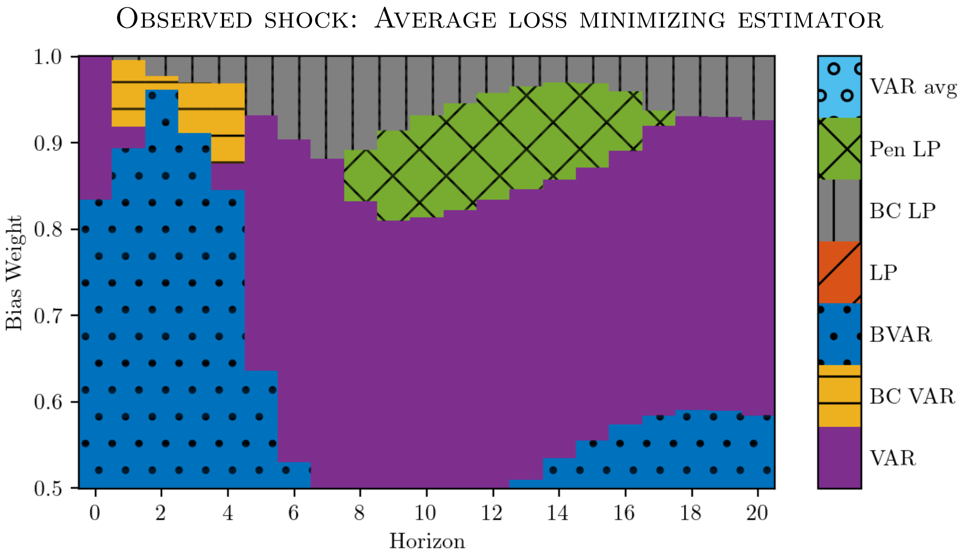

Observed shock identification, medians across 6,000 DGPs

# #2: Bias-corr'd LP best iff. researcher heavily prioritizes bias

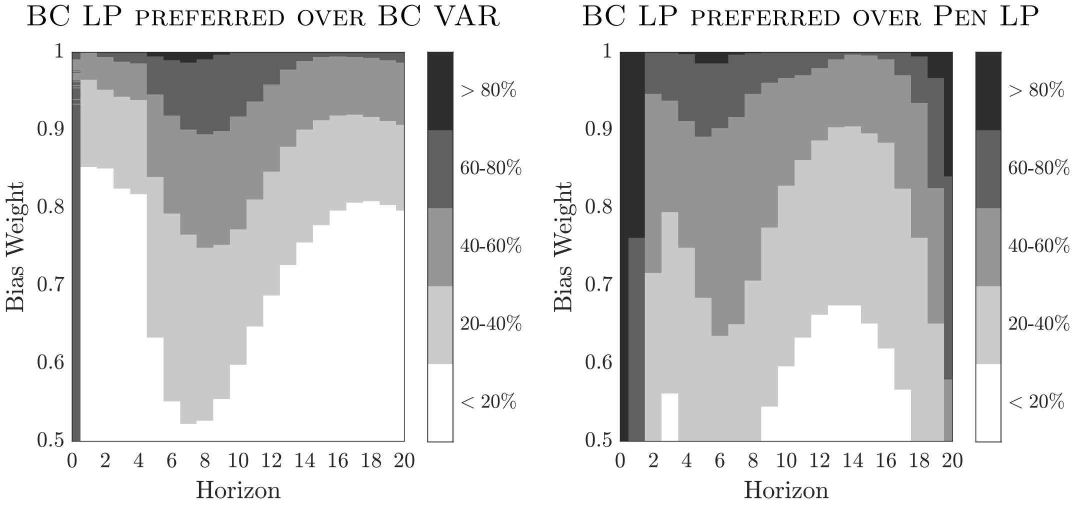



Observed shock identification, medians across 6,000 DGPs

# #2: Bias-corr'd LP best iff. researcher heavily prioritizes bias

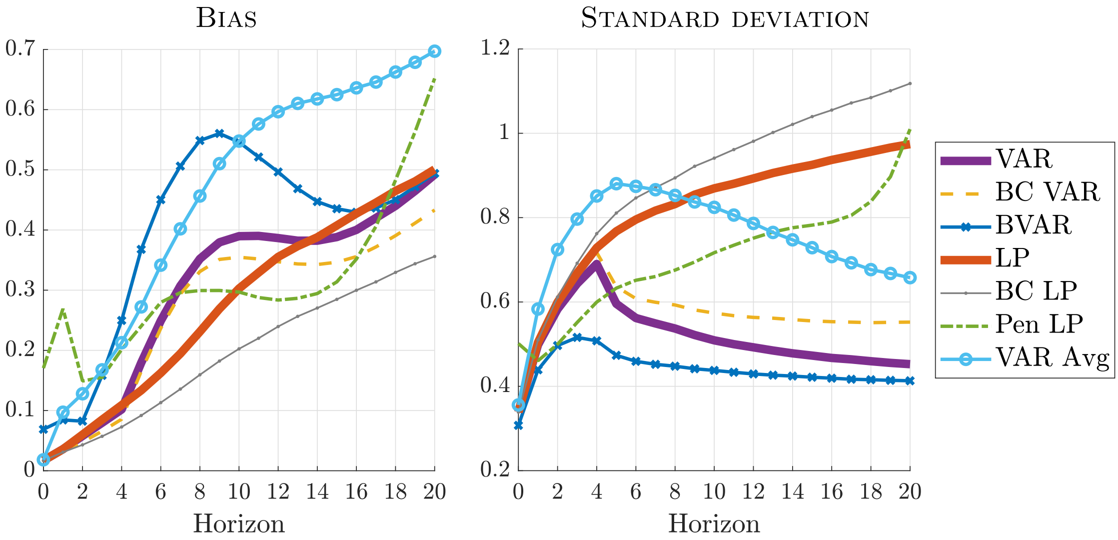


# #2: Bias-corr'd LP best iff. researcher heavily prioritizes bias



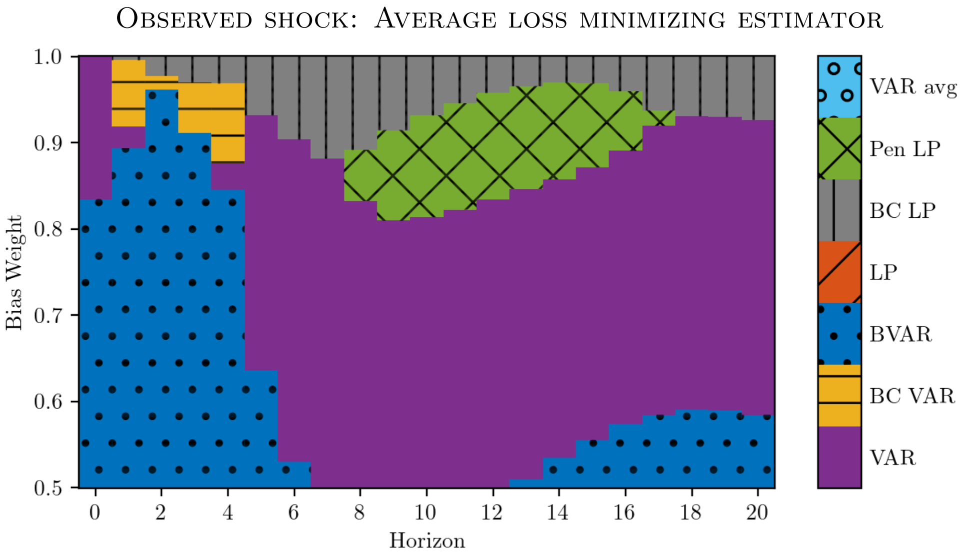

Observed shock identification

# #3: For researchers who also care about precision, (B)VAR is attractive

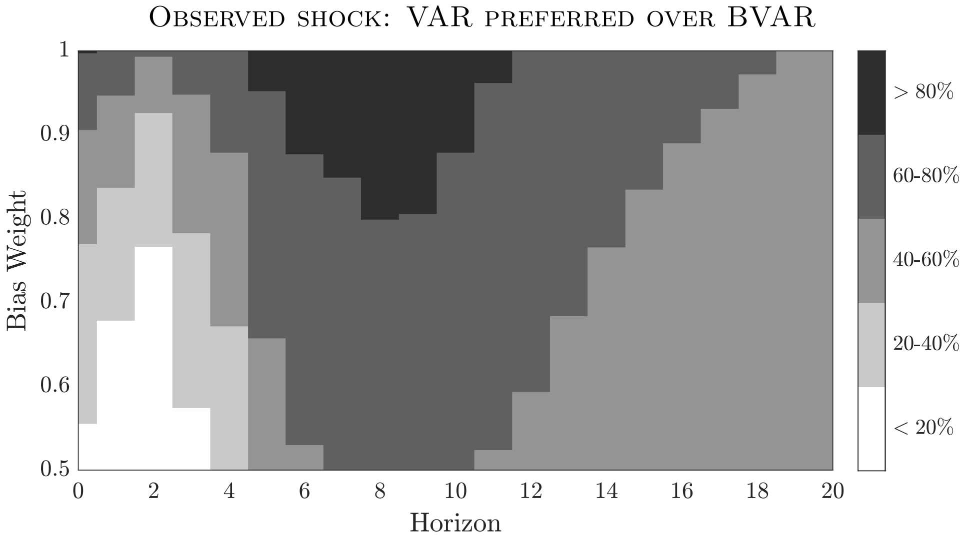



Observed shock identification, medians across 6,000 DGPs

# #3: For researchers who also care about precision, (B)VAR is attractive



# #3: For researchers who also care about precision, (B)VAR is attractive



# Robustness checks in paper

- Stationary DGPs (first diff): bias-correction doesn't matter, shrinkage attractive. ▸
- Other identification schemes: IV, recursive. ▸ IV
- Monetary and fiscal shocks considered separately.
- Longer estimation lag length $p = 8$.
- Smaller sample size $T = 100$.
- Break down results by variable categories.
- Smaller, salient set of observables.
- Near-worst-case performance: 90th percentile loss across DGPs instead of median.

# Can we select the estimator based on the data?

- In-sample, data-driven estimator choice $\Longrightarrow$ best of all worlds?
- Disappointing performance of VAR model averaging estimator suggests caution.
- In our DGPs, conventional model selection/evaluation criteria fail to detect even substantial misspecification of VAR(4) model.
  - AIC: 90th percentile of $\hat{p}_{AIC}$ does not exceed 2 in any DGP.
  - LM test of residual serial correlation (signif. level $=$ 10%): rejection probability below 25% in 92% of DGPs; below 50% in all DGPs.

# Outline

# Conclusion

- Large-scale simulation study of LP, VAR, and related impulse response estimators.
- Thousands of DGPs drawn from encompassing empirical DFM.
- Lessons:
  1. Clear bias-variance trade-off between LP and VAR after persistence-bias-correction.
  2. Bias-corrected LP preferred method if *and only if* researcher overwhelmingly prioritizes bias.
  3. For researchers who also care about precision, VAR is attractive: BVAR at short horizons, OLS at intermediate horizons, similar at long horizons.
- Future research: Panel data, model selection, shrinkage for medium-run responses.

# Conclusion

- Large-scale simulation study of LP, VAR, and related impulse response estimators.
- Thousands of DGPs drawn from encompassing empirical DFM.
- Lessons:
    1. Clear bias-variance trade-off between LP and VAR after persistence-bias-correction.
    2. Bias-corrected LP preferred method if *and only if* researcher overwhelmingly prioritizes bias.
    3. For researchers who also care about precision, VAR is attractive: BVAR at short horizons, OLS at intermediate horizons, similar at long horizons.
- Future research: Panel data, model selection, shrinkage for medium-run responses.

**Thank you!**

# Appendix

# Bias-variance trade-off in simple DGP

**Proposition 1**

Fix $h \geq 0$, $\tau \in \mathbb{R}$, $\sigma_2 > 0$, and $\alpha \in \mathbb{R}$. Assume $E(\varepsilon_{j,t}^4) < \infty$ for $j = 1, 2$.

Then, as $T \to \infty$,

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

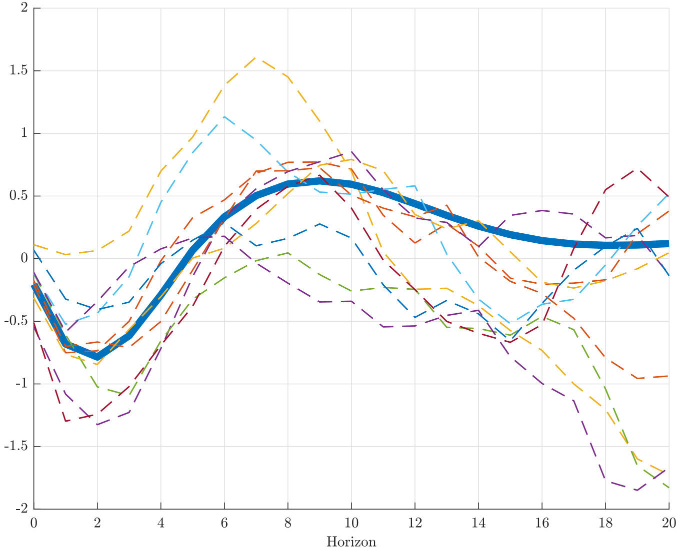
where for all $h \geq 0$,

$$\text{aBias}_{\text{LP},h} \equiv 0, \quad \text{aVar}_{\text{LP},h} \equiv \{1 + (h-1)(1+\tau)^2\}\mathbb{1}(h \geq 1) + (h+1)\sigma_2^2,$$

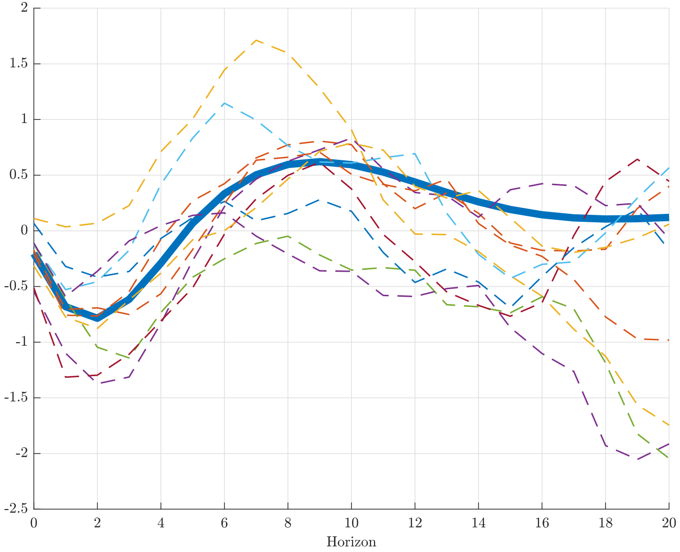
and for $h \geq 1$,

$$\text{aBias}_{\text{VAR},h} \equiv -\alpha\mathbb{1}(h \geq 2), \quad \text{aVar}_{\text{VAR},h} \equiv \{1 + \tau\mathbb{1}(h \geq 2)\}^2 + 2\sigma_2^2.$$

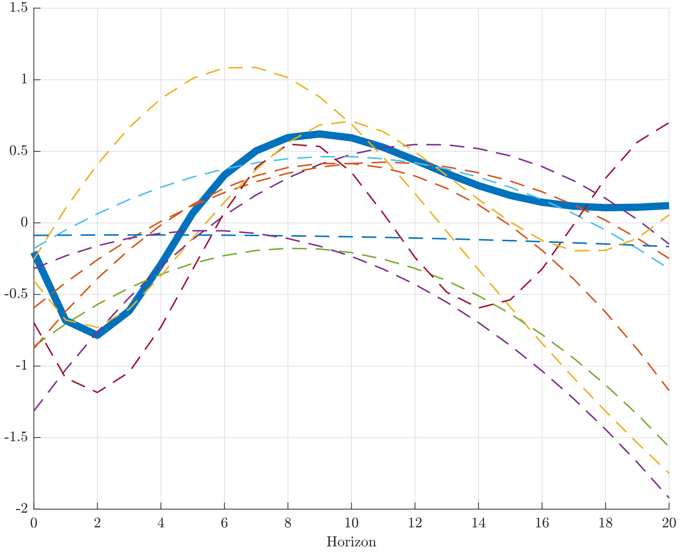
# Example IRF estimates: Least-squares LP



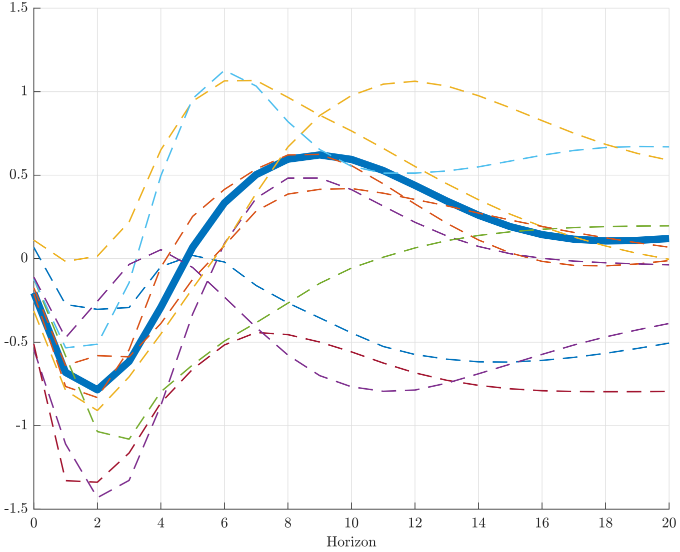
# Example IRF estimates: Bias-corrected LP



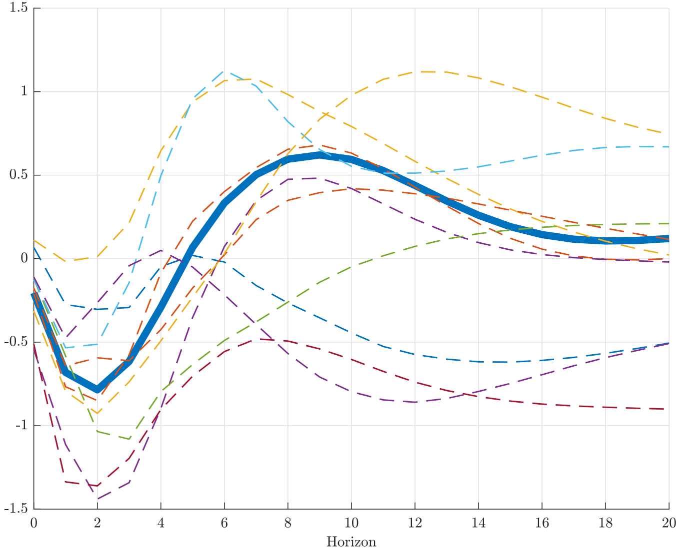
# Example IRF estimates: Penalized LP



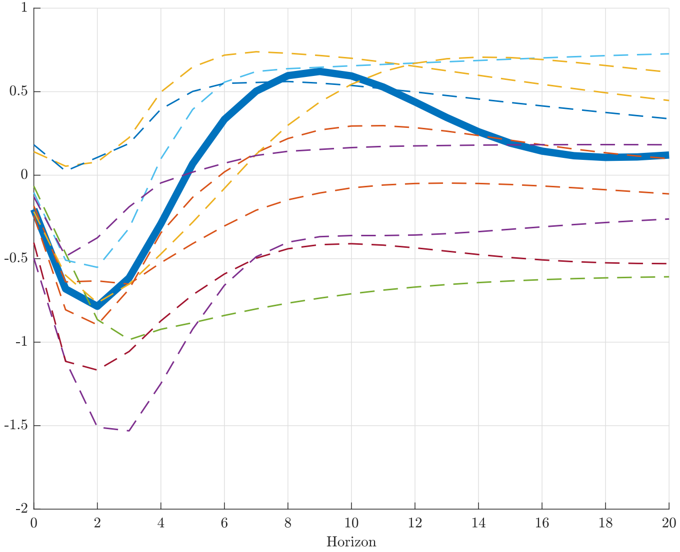
# Example IRF estimates: Least-squares VAR



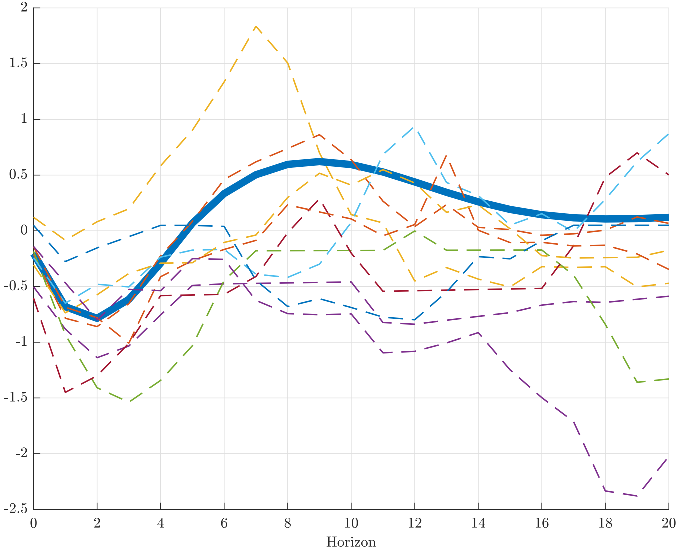
# Example IRF estimates: Bias-corrected VAR



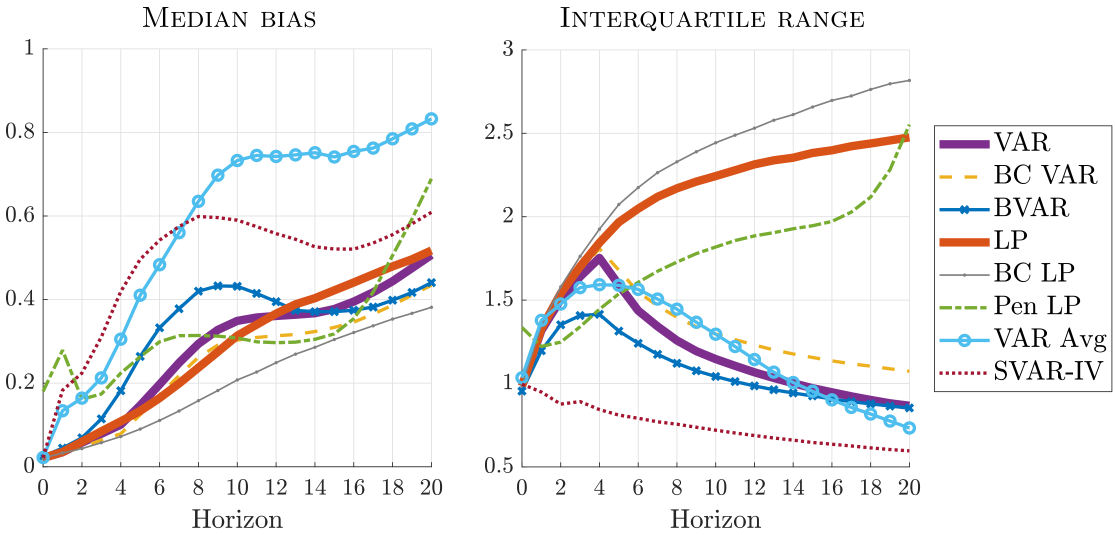
# Example IRF estimates: Bayesian VAR



# Example IRF estimates: VAR model averaging



# #4: SVAR-IV is heavily biased, but has relatively low dispersion




IV identification, medians across 6,000 DGPs

# Stationary DGPs: bias-correction doesn't matter, shrinkage attractive