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


Dake Li — Princeton University

Mikkel Plagborg-Møller — Princeton University

Christian K. Wolf — MIT


Slides: `https://scholar.princeton.edu/mikkelpm`

November 12, 2021

# Estimation of IRFs

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

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

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)

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

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, shrinkage variants, . . .
  - Several identification schemes: observed shock, recursive, proxy/instrument.
  - Pay attention to researcher's loss function and role of horizon.
- Question: Which estimators perform well on average across many DGPs? Answers:
  1. Must care a lot about bias to prefer least-squares LP over VAR.
  2. Shrinkage estimation attractive unless concern for bias is overwhelming.

# 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 (2019)

- Inference about IRFs:
  Inoue & Kilian (2020); Montiel Olea & Plagborg-Møller (2020)

# Outline


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

# Simple analytical example

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

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

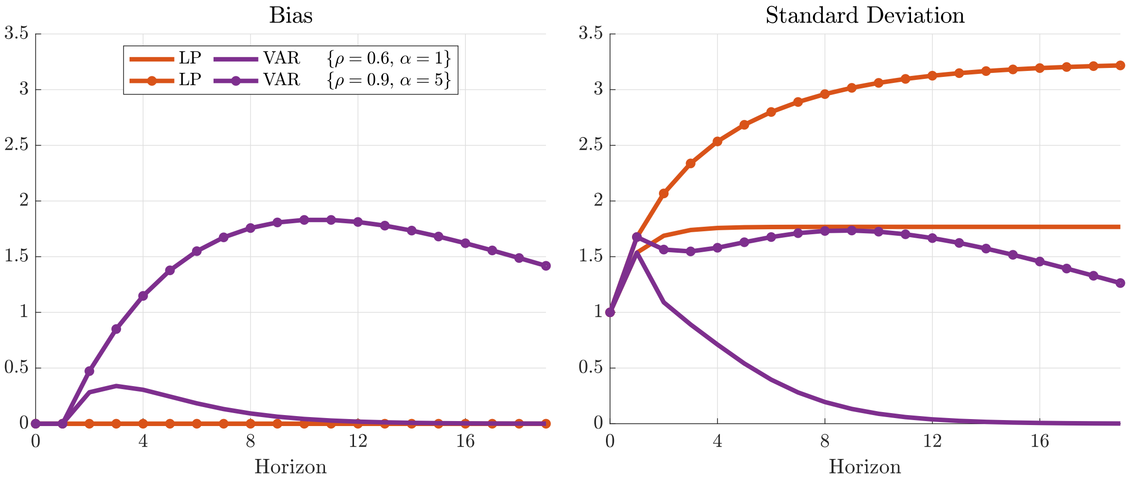
- Parameter of interest: $\theta_h \equiv \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}} = \rho^h$.
- Two estimators (later consider other ones):

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 $\hat{C} = \text{Cholesky}$. 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 ▸ t

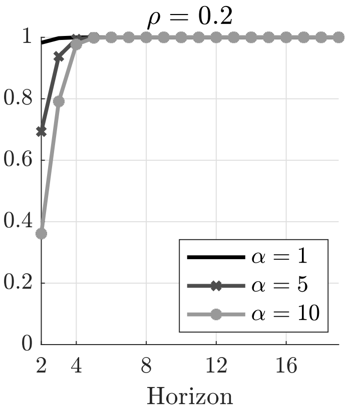
$$\sqrt{T}(\hat{\beta}_h - \theta_h) \xrightarrow{d} N(0, \text{aVar}_{\text{LP}}), \quad \sqrt{T}(\hat{\delta}_h - \theta_h) \xrightarrow{d} N(\text{aBias}_{\text{VAR}}, \text{aVar}_{\text{VAR}}).$$

# Asymptotic bias and standard deviation




- LP has zero asymptotic bias because it projects $y_{t+h}$ directly on shock $\varepsilon_{1,t}$.
- VAR extrapolates, which lowers variance at the cost of bias when $\alpha \neq 0$.

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




- Given loss function

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

how much weight $\omega = \omega_h^*$ should we attach to bias$^2$ to be indifferent btw. LP and VAR?

# Analytical illustration: take-aways

- Even in simple DGP, bias-variance trade-off is non-trivial. Depends on...
  - ... persistence $\rho$ 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 "$\rho$" 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 example
2. Data generating processes
3. Estimators
4. Results
5. Conclusion

# Encompassing model

- Empirical Dynamic Factor Model (DFM):

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

- $X_t$: 207 quarterly macro time series, spanning various categories.
- $f_t$: six latent driving factors, evolve as VAR(2), driven by six aggregate shocks $\varepsilon_t$.
- $v_{i,t}$: idiosyncratic noise, evolves as AR(2), independent across $i$.
- Parameters fixed at estimates from U.S. data in Stock & Watson (2016). ($H$: next slide.)
- Gaussian shocks + noise.

# Lower-dimensional DGPs and estimands

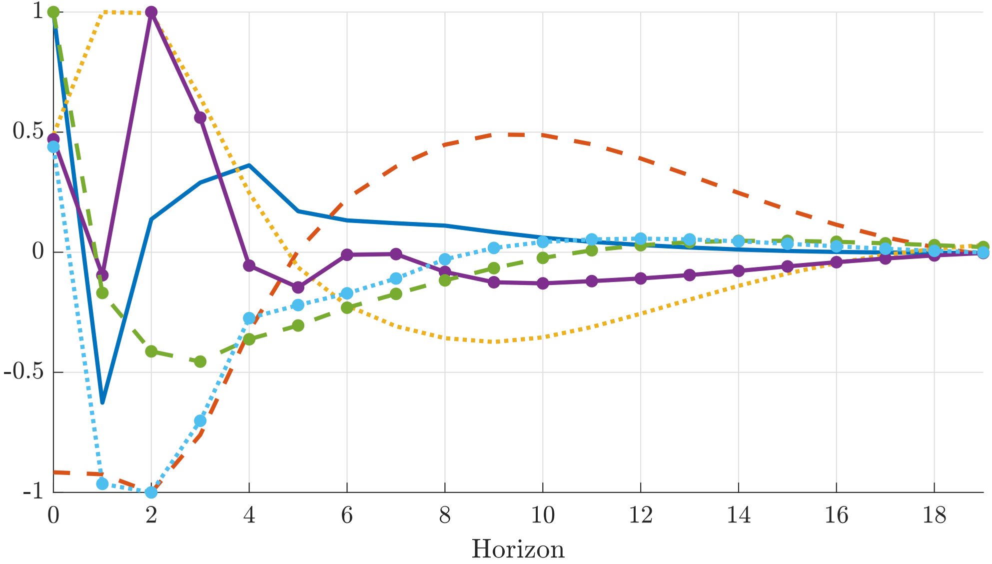
- Draw 6,000 subsets of 5 variables $\bar{w}_t \subset X_t$. DFM implies that $\bar{w}_t$ follows VAR($\infty$).
- $\bar{w}_t$ contains at least one output and one price series, and – depending on type of DGP...
  1. Monetary shock: $i_t =$ federal funds rate.
  2. Fiscal shock: $i_t =$ federal government spending.
- Select response variable $y_t \in \bar{w}_t$ at random (not $i_t$).
- For today, two identification schemes (recursive ID in paper):
  1. Observed shock $\varepsilon_{1,t}$. Impulse response estimand: $\theta_h = \frac{\partial y_{t+h}}{\partial \varepsilon_{1,t}}$, $h = 0, 1, 2, \dots, 19$.
  2. IV/proxy $z_t = \rho z_{t-1} + \varepsilon_{1,t} + \nu_t$ observed. Calibrate signal-to-noise empirically. $\theta_h = \frac{\partial y_{t+h}/\partial \varepsilon_{1,t}}{\partial i_t/\partial \varepsilon_{1,t}}$.
- $H = \frac{\partial f_t}{\partial \varepsilon_t'}$ 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) | 0.42 | 0.93 | 0.98 | 1.14 | 2.29 | 4.78 | 18.09 |
| Largest VAR eigenvalue | 0.82 | 0.84 | 0.84 | 0.84 | 0.84 | 0.86 | 0.91 |
| Fraction of VAR coef's $\ell \geq 5$ | 0.02 | 0.10 | 0.15 | 0.23 | 0.34 | 0.44 | 0.84 |
| Degree of shock invertibility | 0.14 | 0.16 | 0.19 | 0.28 | 0.41 | 0.47 | 0.65 |
| IV first stage F-statistic | 9.49 | 9.61 | 9.70 | 17.29 | 27.58 | 28.20 | 29.10 |
| *Impulse responses up to $h = 19$* | | | | | | | |
| 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) | -0.44 | -0.16 | -0.09 | -0.02 | 0.06 | 0.11 | 0.46 |
| $R^2$ in regression on quadratic | 0.01 | 0.11 | 0.23 | 0.49 | 0.71 | 0.84 | 0.98 |

Combining 6,000 monetary and fiscal DGPs. Observed shock or IV identification.

# Impulse response estimands are also heterogeneous



# Outline

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

# Impulse response estimators

- Local projection methods:
  1. **Least squares**. Jordà (2005)
  2. **Penalized**: shrinks towards quadratic polynomial in $h$. Barnichon & Brownlees (2019)
- VAR methods:
  3. **Least squares**.
  4. **Bias-corrected**: corrects small-sample bias due to persistence. Pope (1990)
  5. **Bayesian**: Minnesota-type prior, shrinks towards white noise. Canova (2007)
  6. **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 (2019)
- IV estimation: "external" (SVAR-IV) vs. "internal" (LP-IV and VAR version).
  Ramey (2011); Stock & Watson (2012, 2018); Mertens & Ravn (2013); Plagborg-Møller & Wolf (2021)

# Outline

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

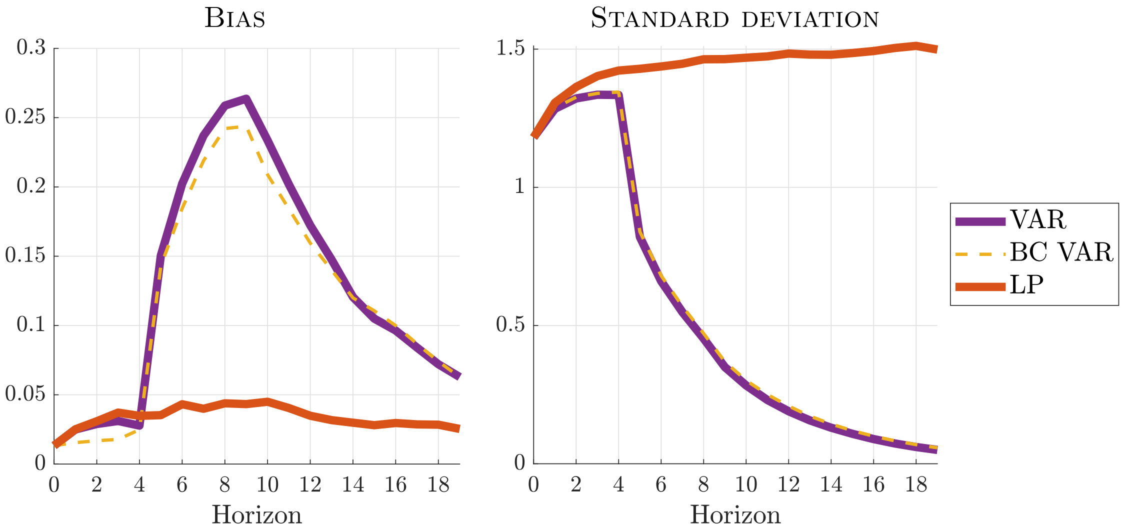
# Specification and simulation settings

- Include $p = 4$ lags in LP and VAR, except VAR model averaging ($p = 8$ in paper). AIC almost always selects fewer than 4 lags.
- Show results for 6,000 monetary and fiscal shock DGPs jointly (separate results in paper).
- Loss function:

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

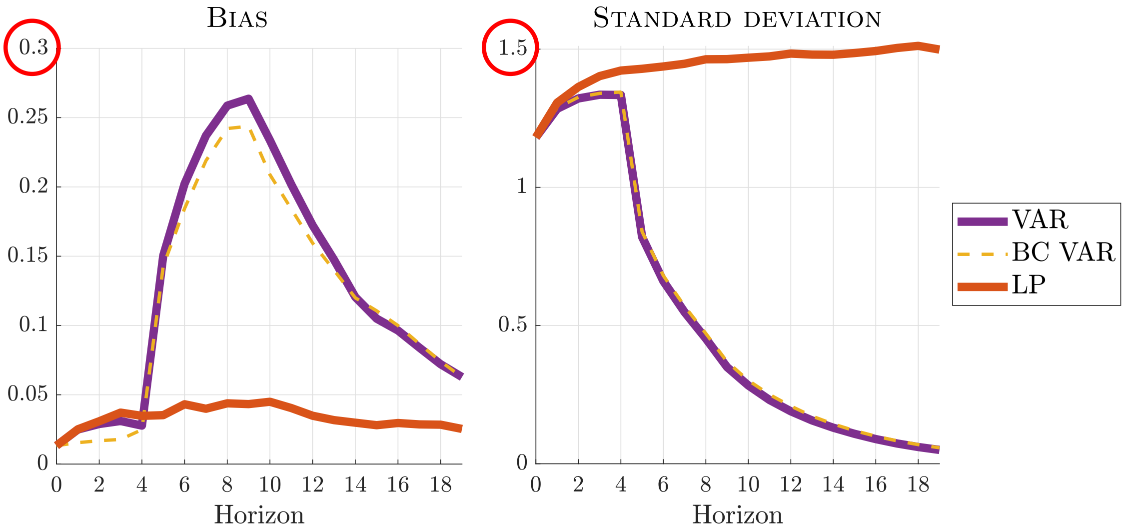
  Divide estimator bias/std by $\sqrt{\frac{1}{20}\sum_{h=0}^{19} \theta_h^2}$ to remove units.
- $T = 200$. 5,000 Monte Carlo repetitions per DGP.
  - Simulations take two weeks on the cluster in Matlab with 25 parallel cores.

# Lesson 1: There is a clear bias-variance trade-off between LP and VAR



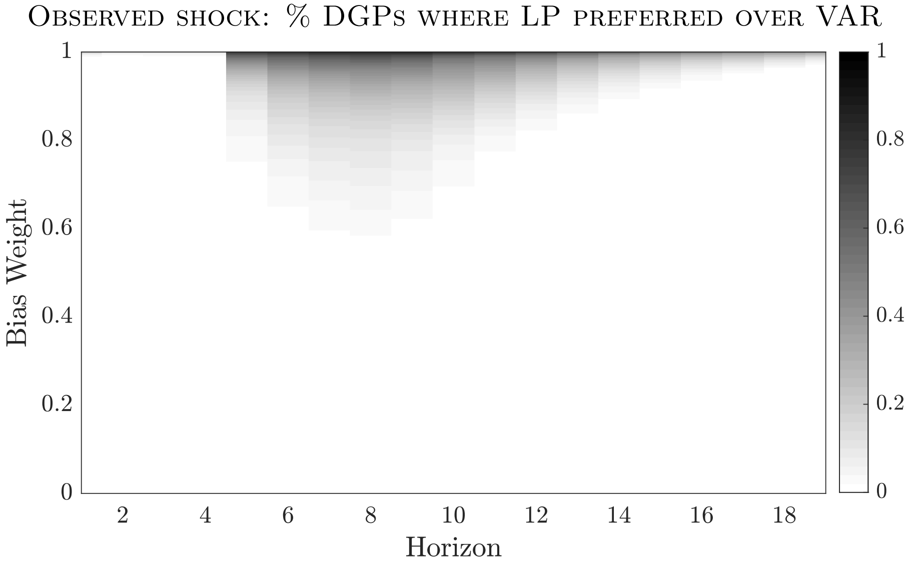

Observed shock identification, medians across 6,000 DGPs

# Lesson 1: There is a clear bias-variance trade-off between LP and VAR

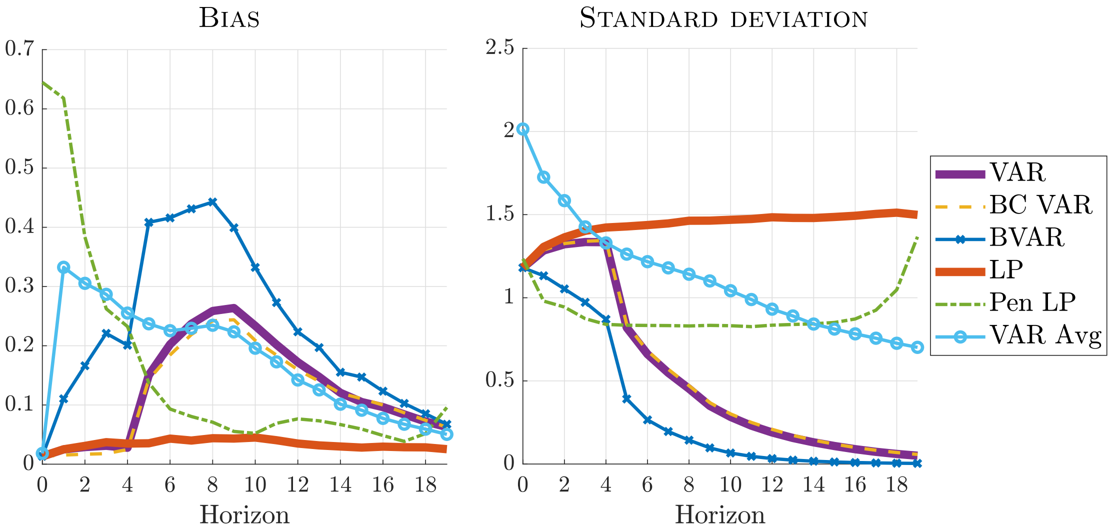



Observed shock identification, medians across 6,000 DGPs

# Lesson 1: There is a clear bias-variance trade-off between LP and VAR

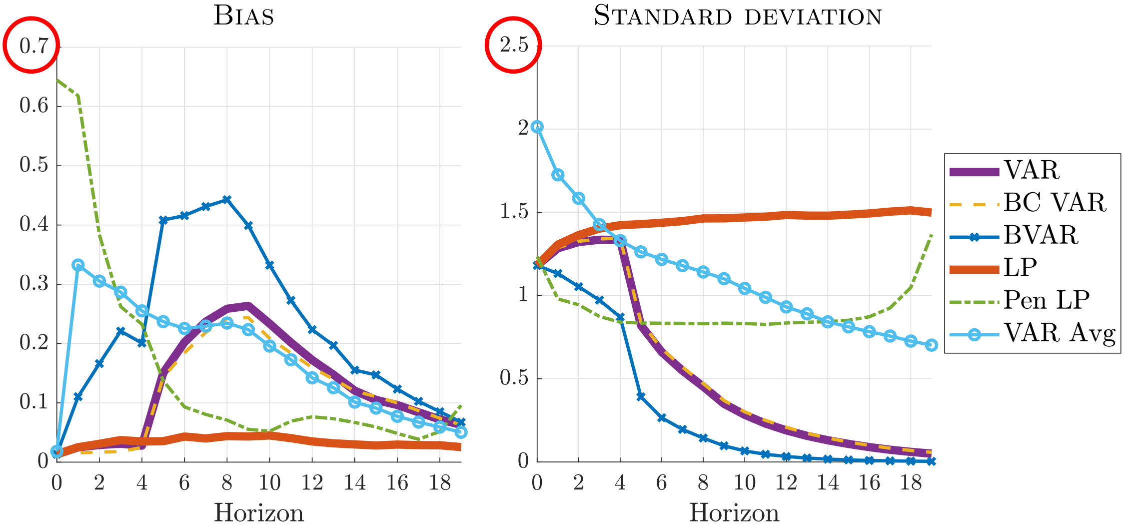


# Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias



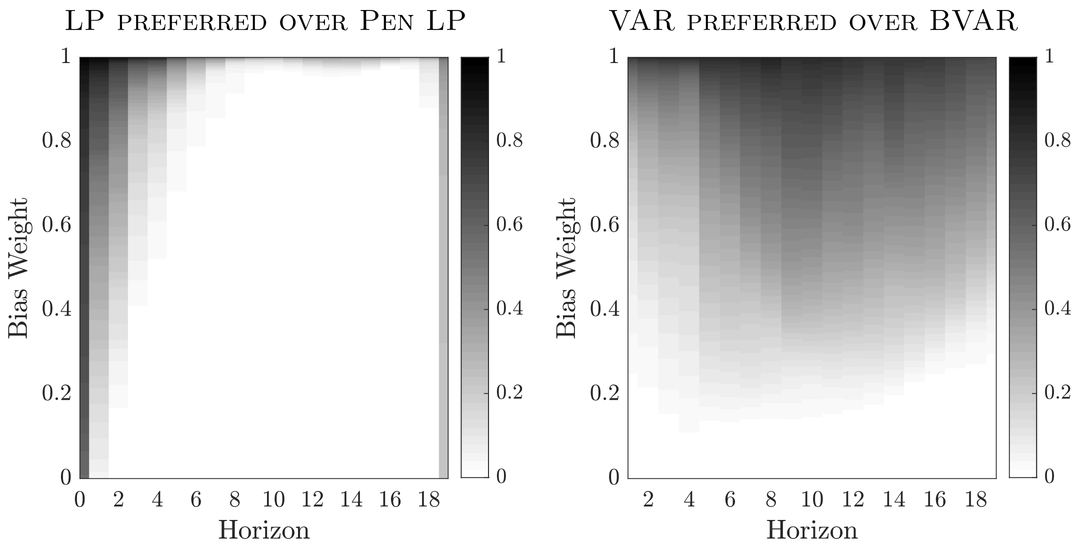

Observed shock identification, medians across 6,000 DGPs

# Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias



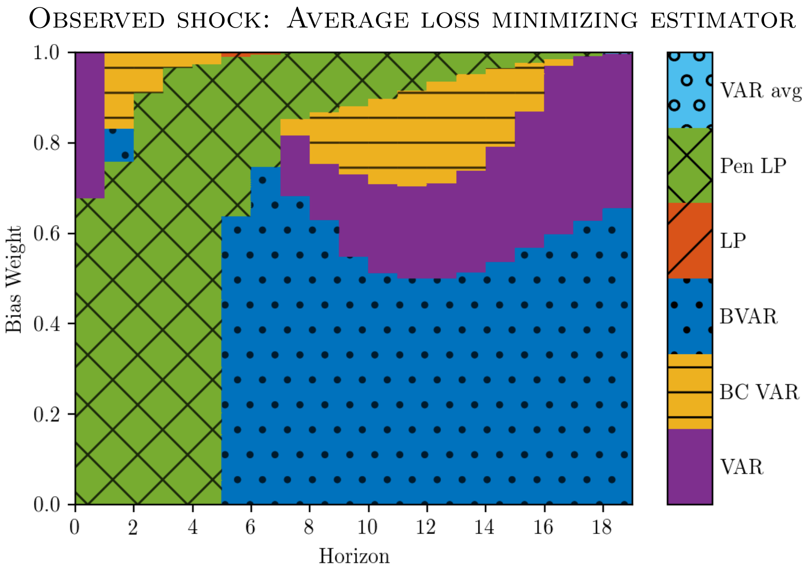

Observed shock identification, medians across 6,000 DGPs

# Lesson 2: Shrinkage dramatically lowers variance, at some cost of bias

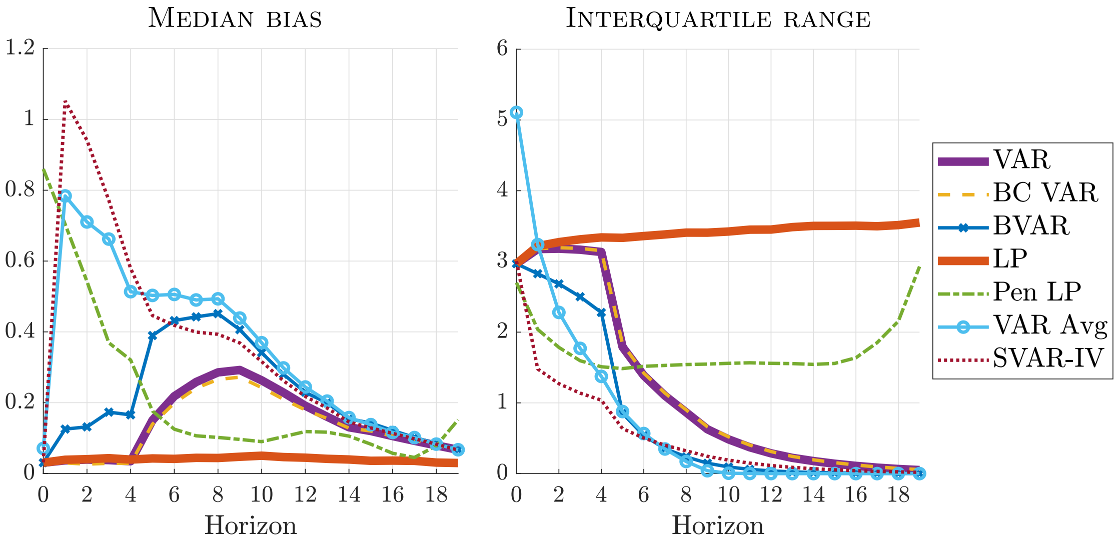



Observed shock identification

# Lesson 3: No method dominates, but shrinkage is generally welcome



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




IV identification, medians across 6,000 DGPs

# Can we select the estimator based on the data?

- In-sample, data-driven estimator choice $\Longrightarrow$ best of both worlds?
- Disappointing performance of VAR model averaging estimator suggests caution.
- In our DGPs, conventional model selection/evaluation criteria are unable 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: rejection probability below 25% in 99.9% of DGPs (signif. level = 10%).

# Outline

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

# 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 least-squares LP and VAR. Loss fct weight on bias must be high to prefer LP over VAR. Caveat: moderately persistent DGPs.
  2. Shrinkage dramatically lowers variance, at some cost of bias.
  3. No method dominates at all horizons, but shrinkage is generally welcome. Penalized LP good at short horizons, BVAR good at intermediate+long.
  4. SVAR-IV is heavily biased, but has relatively low dispersion.

# 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 least-squares LP and VAR. Loss fct weight on bias must be high to prefer LP over VAR. Caveat: moderately persistent DGPs.
  2. Shrinkage dramatically lowers variance, at some cost of bias.
  3. No method dominates at all horizons, but shrinkage is generally welcome. Penalized LP good at short horizons, BVAR good at intermediate+long.
  4. SVAR-IV is heavily biased, but has relatively low dispersion.

**Thank you!**

# Appendix

# Bias-variance trade-off in simple DGP

**Proposition 1**

Fix $h \geq 0$, $\rho \in (-1, 1)$, $\sigma_2 > 0$, and $\alpha \in \mathbb{R}$. Assume $E(\varepsilon_{j,t}^4) < \infty$ for $j = 1, 2$. Define $\sigma_{0,y}^2 \equiv \frac{1+\sigma_2^2}{1-\rho^2}$. Then, as $T \to \infty$,

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

where for all $h \geq 0$,

$$\text{aBias}_{\text{LP}} \equiv 0, \;\; \text{aVar}_{\text{LP}} \equiv \sigma_{0,y}^2(1 - \rho^{2(h+1)}) - \rho^{2h},$$

and for $h \geq 1$,

$$\text{aBias}_{\text{VAR}} \equiv \rho^{h-1}(h-1)\frac{\alpha\sigma_2^2}{\sigma_{0,y}^2 - 1}, \;\; \text{aVar}_{\text{VAR}} \equiv \rho^{2(h-1)}(1-\rho^2)\sigma_{0,y}^2\left(1 + \frac{(h-1)^2}{\sigma_{0,y}^2 - 1}\right) + \rho^{2h}\sigma_2^2.$$

# Interpretation of degree $\alpha$ of misspecification

**Proposition 2**

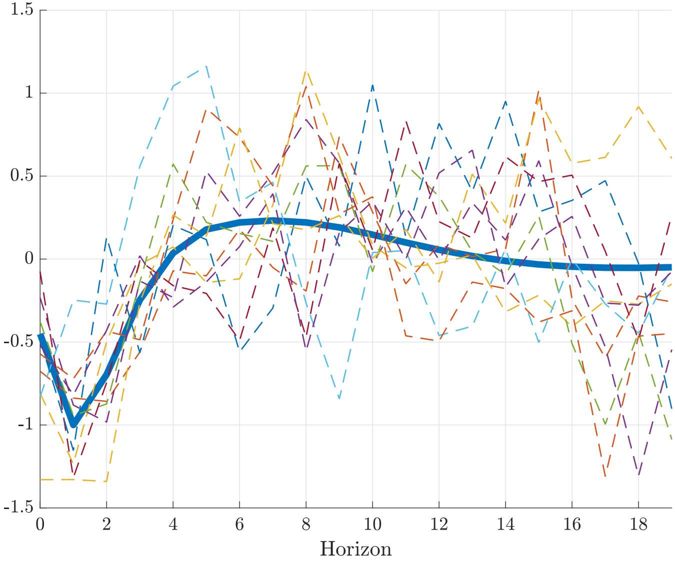
Impose same assumptions as in Proposition 1.

Let $\hat{\tau}$ denote the t-statistic for testing the significance of the second lag in a univariate AR(2) regression for $\{y_t\}$.

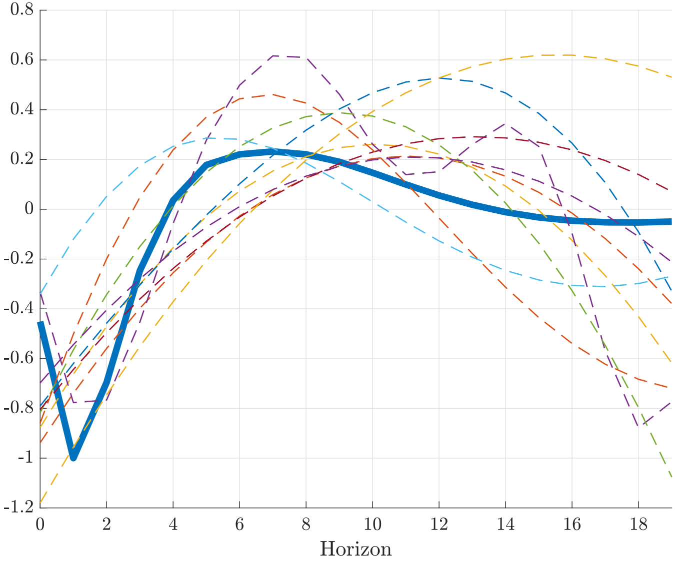
Then, as $T \to \infty$,

$$\hat{\tau} \xrightarrow{d} N\left(-\rho \frac{\sigma_2^2}{1+\sigma_2^2}\alpha,\ 1\right).$$

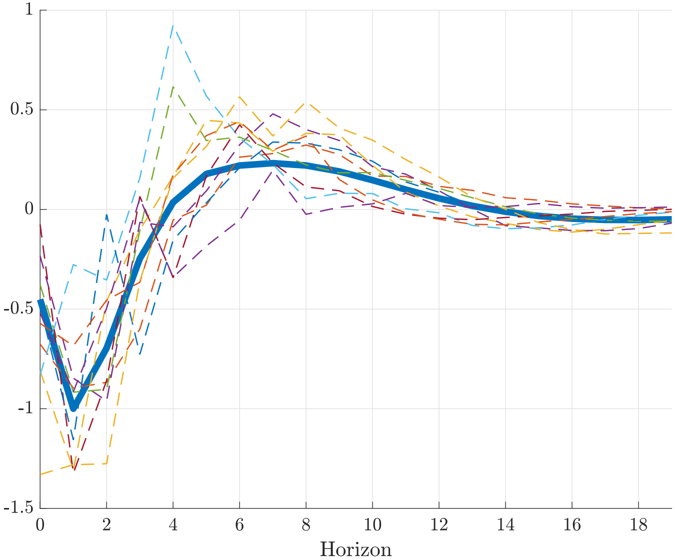
# Example IRF estimates: Least-squares LP



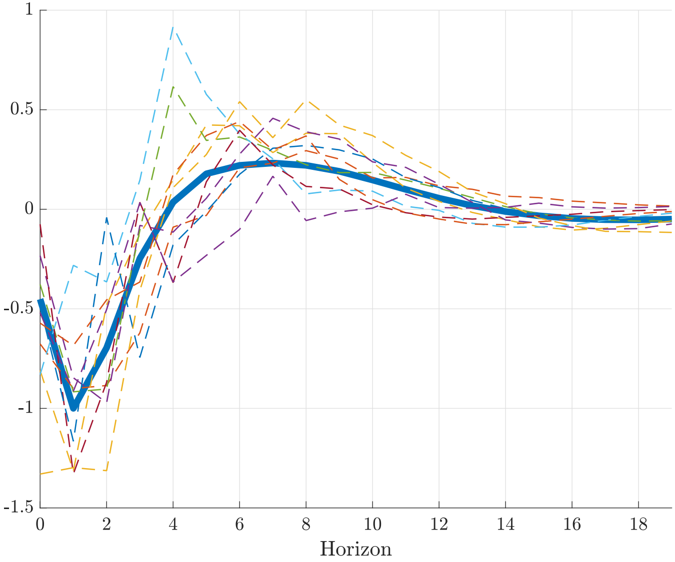
# Example IRF estimates: Penalized LP



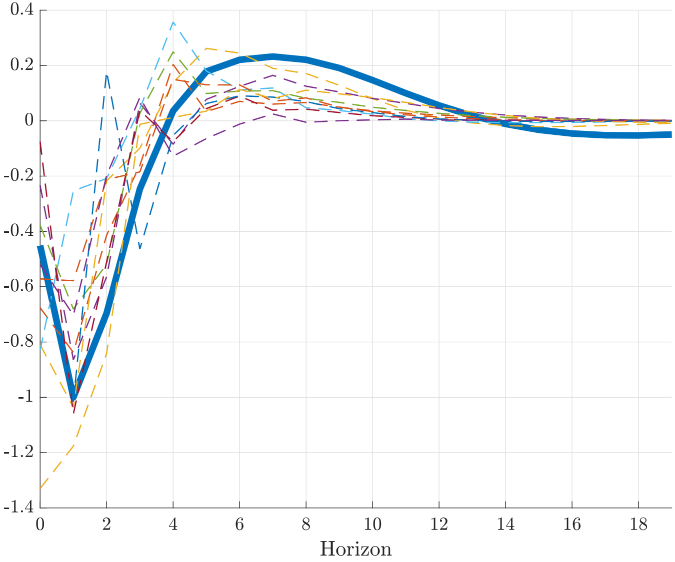
# Example IRF estimates: Least-squares VAR



# Example IRF estimates: Bias-corrected VAR



# Example IRF estimates: Bayesian VAR



# Example IRF estimates: VAR model averaging