Tradeoffs Between Impulse Response Inference Methods

Mikkel Plagborg-Møller

Princeton University

January 7, 2023

This talk

- Selective summary (manifesto?) of recent research about estimation/inference for

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

- Comparison of semi-structural estimation procedures:
  - Local projections (LP), vector autoregressions (VAR), shrinkage variants (Bayesian, smoothing).
- Large-sample equivalence, but finite-sample trade-offs:
  - Point estimation: bias vs. variance.
  - Inference: different notions of confidence interval coverage.
- Point out areas for future research.

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

Assumptions

- For simplicity, focus on simple case where x_t (shock of interest) is...
  1. observed.
  2. independent of all past data.
  3. mean $= 0$, variance $= 1$.
- Denote the full data vector by $Y_t = (x_t, y_t, \dots)'$. Assume stationary.
- Abstract from any deterministic terms.
- These assumptions can be relaxed.

Impulse responses estimators

- VAR estimator: extrapolates $\hat{\theta}_h$ from first p sample autocov's. Sims (1980)

  1. Estimate model
  $$Y_t = \sum_{\ell=1}^{p} \hat{A}_\ell Y_{t-\ell} + \hat{u}_t, \quad \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'.$$
  2. $\hat{\theta}_h = e_2' \hat{\Psi}_h \hat{\nu}$, where $\hat{\nu} = \text{chol}(\hat{\Sigma}) e_1$, and we iterate
  $$\hat{\Psi}_h = \sum_{\ell=1}^{\min\{p,h\}} \hat{A}_\ell \hat{\Psi}_{h-\ell}, \; h = 1, 2, \dots, \quad \hat{\Psi}_0 = I.$$

- LP estimator: direct projection, separately for each h. Jordà (2005)
$$y_{t+h} = \hat{\theta}_h x_t + \sum_{\ell=1}^{p} \hat{\delta}_{\ell,h}' Y_{t-\ell} + \hat{e}_{t,h}.$$

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

LP and VAR estimate the same impulse responses

- In population, LP and VAR estimate the same impulse responses:

$$\operatorname*{plim}_{T\to\infty} \hat{\theta}_h^{\text{LP}} = \operatorname*{plim}_{T\to\infty} \hat{\theta}_h^{\text{VAR}} = \text{Proj}[y_{t+h} \mid x_t = 1] - \text{Proj}[y_{t+h} \mid x_t = 0] \equiv \theta_h$$

  for all horizons $h \leq p$. P-M & Wolf (2021)

- Intuition:
  1. $\hat{\theta}_h^{\text{VAR}}$ = MSE-optimal forecast given model-implied autocov's out to horizon h.
  2. VAR(p) estimator consistently estimates true autocov's $\text{Cov}(Y_t, Y_{t-\ell})$ for $\ell \leq p$.
  3. LP estimates the MSE-optimal forecast by definition.

- Note: Result does not assume any particular parametric model (e.g., linear or VAR(p)).

Implications of large-sample equivalence

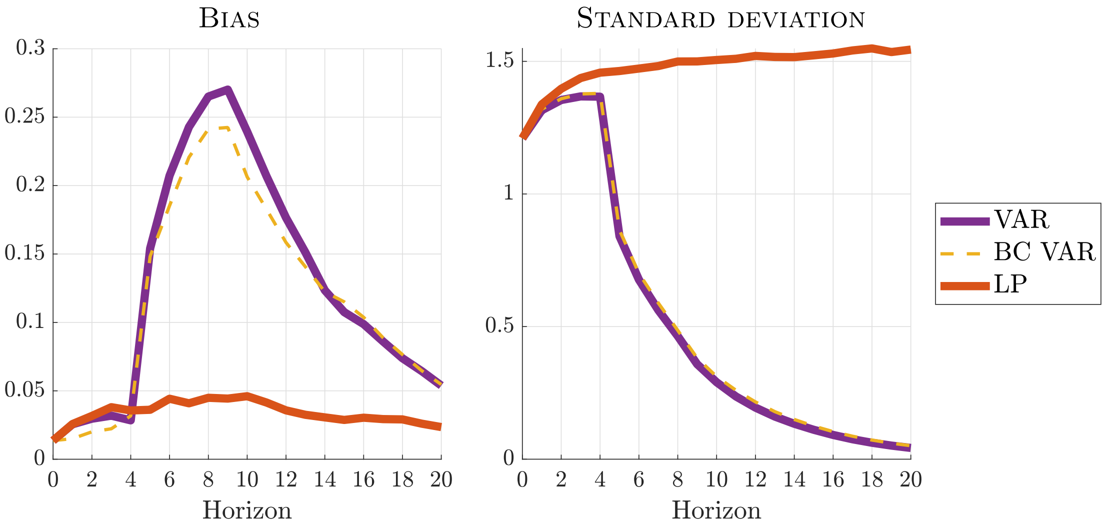
1. LP $\approx$ VAR at short horizons h, but not necessarily for $h > p$.
2. LP IRF $\approx$ VAR IRF with sufficiently large p.
   - In fact, estimators are asymptotically equivalent as $p, T \to \infty$. Xu (2022)
3. LP and VAR are not conceptually distinct paradigms. Just two techniques for estimating projections with shared large-sample estimand.
4. Identification $\perp$ estimation: Any SVAR identification procedure can be implemented using LP and *vice versa* (also when x_t is not observed).
   - Short-run/long-run zero restrictions, sign restrictions, narrative restrictions, higher moments.
5. LP is not more "robust to misspecification" than VAR (unless this means sensitivity to p).

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

Finite-sample bias-variance trade-off

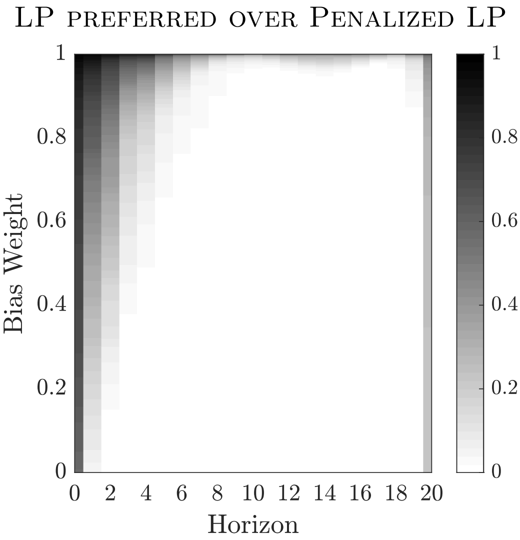
- At horizons $h > p$, we face non-trivial bias-variance trade-off:
  - VAR extrapolates longer-run responses from first p autocov's. Low variance, potentially high bias if DGP $\neq$ VAR(p).
  - LP does not extrapolate. High variance, low bias.
- In ongoing simulation project, my coauthors and I explore this trade-off across 1,000s of empirically calibrated DGPs (caveat: stationarity). Li, P-M & Wolf (2022)

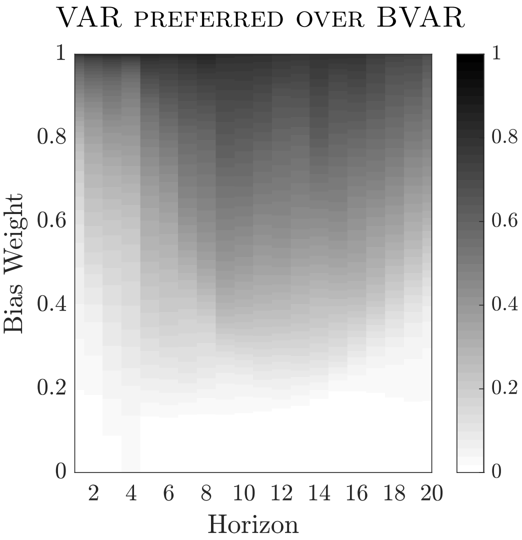


Medians across 6,000 DGPs. Source: Li, P-M & Wolf (2022).

Improving the trade-off through shrinkage/penalization

- Shrinkage/penalization: nudge least-squares LP/VAR estimates towards *a priori* more reasonable values.
  - Bayesian VAR: shrink towards impulse responses for independent random walks or white noise. Doan, Litterman & Sims (1984)
  - Smoothness: shrink LP towards impulse responses that are smoother functions of h. Shiller (1973); P-M (2016); Barnichon & Brownlees (2019)
  - Functional form: shrink LP towards VAR-implied impulse responses or towards exponential shapes. Barnichon & Matthes (2018); Miranda-Aggripino & Ricco (2021)
- Introduces some bias (unless prior is exactly right) in order to lower variance.
- Degree of shrinkage chosen based on Bayesian considerations or frequentist objective fct.





Fraction of DGPs where estimators are preferred according to loss function = $\omega \times \text{bias}^2 + (1-\omega) \times \text{variance}$. Source: Li, P-M & Wolf (2022).

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

Confidence intervals for impulse responses

- VAR(p)-based CIs are non-robust at longer horizons (lack uniform validity).
  - Consider AR(1): $\hat{\theta}_h = \hat{\rho}^h$.
  - Sampling distr'n highly sensitive to estimation error $\hat{\rho} - \rho$ when $h \to \infty$. Methods that are tailored to the case $h \to \infty$ don't work at shorter horizons.
    Wright (2000); Gospodinov (2004); Pesavento and Rossi (2007), Mikusheva (2012)
  - Unit roots can cause non-normal limiting distr'n (not always). Want to avoid pre-testing for unit roots. Inoue & Kilian (2002, 2020)
- In contrast, linearity of LP means normal approx'n works regardless of h and unit roots.
  Dufour, Pelletier & Renault (2006); Breitung & Brüggemann (2019); Montiel Olea & P-M (2021)
  - Important to control for lagged data so variance of OLS residual doesn't blow up.
  - Bonus: No need for Newey-West standard errors if we're regressing on a shock x_t.

Can we shorten the CIs?

- LP (or VAR with large p) often yields wide CIs in practice. Can shrinkage help here?
  - No, not if we insist on usual coverage notion: $1-\alpha$ coverage prob. separately at each horizon, regardless of true impulse responses. Pratt (1961); Amstrong & Kolesár (2018)
  - Intuition: Must guard against worst-case bias over parameter space.
- If we want narrower CIs, we have to relax the coverage requirement. Ideas:
  - Control coverage prob. only *on average* across horizons. Armstrong, Kolesár & P-M (2022)
  - Restrict parameter space (e.g., *a priori* upper bound on jaggedness of IRF).
  - Coverage for simpler "surrogate IRF". Genovese & Wasserman (2008)
  - Don't report confidence band. Think harder about what specific hypothesis we want to test.

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

Impulse responses in nonlinear models

- What do linear LP/VAR estimate in nonlinear DGPs?
- Assume general nonlinear causal representation, given shock vector $\varepsilon_t = (x_t, \varepsilon_{2:n,t})$:

$$y_{t+h} = g_h(\varepsilon_{t+h}, \dots, \varepsilon_{t+1}, \varepsilon_t, Y_{t-1}).$$

  - E.g., binary outcomes, ZLB, discrete/smooth regime-switching, higher-order lag dynamics.
- Many possible definitions of nonlinear impulse responses. Linear LP/VAR do not consistently estimate all of these. Gonçalves, Herrera, Kilian & Pesavento (2021, 2022)
- Consider manipulating x_t from δ_1 to δ_2, while averaging over other shocks and the history:

$$\theta_h(\delta_1, \delta_2) \equiv E\Big[g_h\big(\varepsilon_{t+h}, \dots, \varepsilon_{t+1}, (\delta_2, \varepsilon_{2:n,t}), Y_{t-1}\big) - g_h\big(\varepsilon_{t+h}, \dots, \varepsilon_{t+1}, (\delta_1, \varepsilon_{2:n,t}), Y_{t-1}\big)\Big].$$

Interpretation of linear LP/VAR estimand in nonlinear models

- If shocks are mutually and serially independent, and $h \leq p$,

$$\operatorname*{plim}_{T\to\infty} \hat{\theta}_h^{\text{LP}} = \operatorname*{plim}_{T\to\infty} \hat{\theta}_h^{\text{VAR}} = \lim_{\delta\to 0} \int_{-\infty}^{\infty} w(x) \frac{\theta_h(x, x+\delta)}{\delta}\, dx,$$

  where $w(\cdot) \geq 0$ and $\int_{-\infty}^{\infty} w(x)\, dx = 1$.
  Yitzhaki (1996); Angrist & Pischke (2009); Rambachan & Shephard (2021); Kolesár & P-M (2022)

- Linear LP/VAR estimate a meaningful summary: weighted average of scaled causal effects $\theta_h(x, x+\delta)/\delta$ for infinitesimal shock size $\delta \approx 0$.

- Weight function $w(\cdot)$ does not depend on h. Estimable. Angrist & Krueger (1999)

- Ongoing work: implications for state/sign-dependence, compare with nonlinear estim'rs.

Outline

1. Econometric framework
2. Large-sample estimand
3. Finite-sample estimation
4. Confidence intervals
5. Nonlinearities
6. Topics for future research

Conclusions and areas for future research

- LP and VAR belong to a menu of procedures with shared large-sample estimand.
- Shrinkage/penalization usually improves bias-variance trade-off.
  - Can new procedures further optimize trade-off? Does trade-off differ when we have panel data?
- However, shrinkage does not help confidence interval construction in the usual sense.
  - Should we relax the coverage requirement, and how? How to deal with very long horizons? Müller & Watson (2017, 2018)
- Linear LP/VAR useful also in nonlinear DGPs, but much work to be done.
  - Do existing nonlinear estimators have interpretable estimands under misspecification? Can we hope to accurately estimate specific nonlinearities in macro data, and how?