Double Robustness of Local Projections and Some Unpleasant VARithmetic

José Luis Montiel Olea
Cornell

Mikkel Plagborg-Møller
Princeton

Eric Qian
Princeton

Christian K. Wolf
MIT

May 11, 2024

Inference on impulse responses

- Impulse response:

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

- **Vector autoregression (VAR)** Sims (1980, 20k GS cites): extrapolate from dynamic model

$$y_t = \hat{A} y_{t-1} + \hat{H} \hat{\varepsilon}_t, \quad \hat{\delta}_h \propto e_{i^*}' \hat{A}^h \hat{H}_{\bullet,1}.$$

- **Local projection (LP)** Jordà (2005, 3.5k GS cites): direct OLS regression

$$y_{i^*,t+h} = \hat{\beta}_h y_{1,t} + \text{controls} + \hat{\xi}_{h,t}.$$

- Perennial issues in applied work: LP or VAR? How to select controls var's and #lags?

Inference on impulse responses: Misspecification

- Jordà (2005) on LP vs. VAR: "[T]hese projections are local to each forecast horizon and therefore **more robust to misspecification** of the unknown DGP."
  - Echoed in influential reviews by Ramey (2016) and Nakamura & Steinsson (2018).
  - Essentially no general theoretical results to support this yet.
  - Not strictly true: LP $\approx$ VAR with many lags p. P-M & Wolf (2021); Xu (2023)
- **Bias-variance trade-off** in simulations: Li, P-M & Wolf (2024)
  - VAR (with moderate lag length) extrapolates: low variance, potentially high bias.
  - LP does not extrapolate: low bias, high variance.
- Open questions: How bad can the biases of VAR & LP get relative to their variances? How do biases distort (frequentist) **inference**?

Our paper: Model

- SVAR(p) model with small MA remainder: Schorfheide (2005); Müller & Stock (2011)

$$y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + H\left(\varepsilon_t + T^{-\zeta} \sum_{\ell=1}^{\infty} \alpha_\ell \varepsilon_{t-\ell}\right), \quad \varepsilon_t \overset{i.i.d.}{\sim} (0, D).$$

- Empirically plausible: dynamics well-approximated by finite-order VAR, but not exact fit.
    - Local-to-0 device: generates tractable asy. bias-variance trade-off, mimicking finite sample. Neyman (1937); Pitman (1948); Rothenberg (1984); Armstrong & Kolesár (2021)
- Parameter of interest: impulse response of $y_{i^*,t+h}$ wrt. first shock $\varepsilon_{1,t}$.
    - Shock directly observed or identified as residual.
- Assume stationarity, fixed horizon h.

Our paper: Main results

$$y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + H\left(\varepsilon_t + T^{-\zeta}\sum_{\ell=1}^{\infty}\alpha_\ell \varepsilon_{t-\ell}\right), \quad \varepsilon_t \overset{i.i.d.}{\sim} (0, D)$$

1. LP CI is robust: correct asy. coverage when $\zeta > 1/4$ due to *double robustness*.
2. Some unpleasant VARithmetic:
   i. VAR CI generically under-covers when $\zeta \leq 1/2$.
   ii. No free lunch: Worst-case bias – given bound on noise-to-signal ratio – is small iff. aVar(VAR) $\approx$ aVar(LP).
   iii. Low VAR coverage for "reasonable" MA coef's that are difficult to detect statistically.
   iv. Fixing VAR coverage w/ large lag length or bias-aware critical value yields wide CI – might as well have done LP.

Lit

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Local-to-AR(1) model

$$y_t = \rho y_{t-1} + [1 + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$$

- Parameter of interest (h fixed):

$$\theta_{h,T} \equiv \frac{\partial y_{t+h}}{\partial \varepsilon_t} = \rho^h + T^{-\zeta} \sum_{\ell=1}^{h} \rho^{h-\ell}\alpha_\ell.$$

- Assumptions (ignoring regularity cond'ns):

  i. $\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$.

  ii. **Stationarity:** $\rho \in (-1, 1)$.

  iii. **Local misspecification:** $\zeta > 1/4$.

Types of misspecification

- Why might small MA terms arise?
  - Discrete-time DSGE models generally have VARMA representations, not finite-order VAR.
  - Dynamic misspecification of true finite-order VAR:
    - Under-specified lag length.
    - Failure to control for relevant variables (special case: non-invertibility).
    - Aggregation (cross-sectional or temporal), measurement error. Granger & Morris (1976)
- Our framework encompasses general additive misspec'n: $y_t = \rho y_{t-1} + \varepsilon_t + T^{-\zeta} \upsilon_t$, with param. of interest $\theta_h \equiv \text{proj}[y_{t+h} \mid \varepsilon_t = 1] - \text{proj}[y_{t+h} \mid \varepsilon_t = 0]$.
  - Omitted nonlinearities, stationary time-varying parameters.

Estimators

- **LP:** Coefficient $\hat{\beta}_h$ in OLS regression

$$y_{t+h} = \hat{\beta}_h y_t + \hat{\gamma}_h y_{t-1} + \hat{\xi}_{h,t}.$$

- **AR:**

$$\hat{\delta}_h \equiv \hat{\rho}^h, \quad \text{where} \quad \hat{\rho} \equiv \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}.$$

- The two estimators coincide on impact: $\hat{\beta}_0 = \hat{\delta}_0 = 1$.

Robustness of LP to local misspecification

Proposition: LP representation

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\xi_{h,t}\varepsilon_t + o_p(T^{-1/2}),$$

where

$$\xi_{h,t} \equiv \sum_{\ell=1}^{h}\rho^{h-\ell}\varepsilon_{t+\ell}.$$

- LP limit does not depend on misspecification parameters ζ or $\alpha(L)$ (as long as $\zeta > 1/4$).
- Note: MA terms of order $T^{-\zeta}$ with $\zeta < 1/2$ can be detected with prob. $\to 1$.

Robustness of LP to local misspecification: Why?

$$y_{t+h} = \hat{\beta}_h y_t + \hat{\gamma}_h y_{t-1} + \hat{\xi}_{h,t}$$

- Intuition: omitted variable bias formula for LP coefficient $\hat{\beta}_h$.

$$\text{OVB} \propto \underbrace{\frac{\partial y_{t+h}}{\partial(\text{omitted lags})}}_{O(T^{-\zeta})} \times \underbrace{\text{Cov}(\underbrace{y_t - E[y_t \mid y_{t-1}]}_{\varepsilon_t + T^{-\zeta} \times \text{lags}}, \text{omitted lags})}_{= \text{Cov}(\varepsilon_t, \text{omitted lags}) + O(T^{-\zeta})} = O(T^{-2\zeta}) = o(T^{-1/2}),$$

since $\text{Cov}(\varepsilon_t, \text{omitted lags}) = 0$.

- Equivalent with **double robustness** in partially linear regression. Chernozhukov et al. (2018)
  - LP consistent if we correctly specify *either* lagged controls *or* shock.

Asymptotic bias of AR estimator

Proposition: AR representation

$$\hat{\delta}_h - \theta_{h,T} = T^{-\zeta}\,\text{aBias}(\hat{\delta}_h) + \frac{h\rho^{h-1}(1-\rho^2)}{\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \tilde{y}_{t-1} + o_p(T^{-\zeta} + T^{-1/2}),$$

where $\tilde{y}_t$ satisfies correctly specified AR(1) model with $\alpha(L) = 0$, and

$$\text{aBias}(\hat{\delta}_h) \equiv \underbrace{h\rho^{h-1}}_{\frac{\partial(\rho^h)}{\partial\rho}} \underbrace{(1-\rho^2)\sum_{\ell=1}^{\infty}\rho^{\ell-1}\alpha_\ell}_{\text{aBias}(\hat{\rho})=\frac{\text{Cov}(\tilde{y}_{t-1},\alpha(L)\varepsilon_t)}{\text{Var}(\tilde{y}_{t-1})}} - \underbrace{\sum_{\ell=1}^{h}\rho^{h-\ell}\alpha_\ell}_{\theta_{h,T}-\rho^h}.$$

- Bias dominates when $\zeta \in (1/4, 1/2)$.
- When $\zeta = 1/2$ (detectable with prob. $\to (0,1)$): nontrivial asy. bias; asy. variance same as in correctly specified case ($\alpha(L) = 0$).

Conventional confidence intervals

$$\text{CI}(\hat{\beta}_h) \equiv \left[\hat{\beta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\beta}_h)/T}\right], \quad \text{CI}(\hat{\delta}_h) \equiv \left[\hat{\delta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\delta}_h)/T}\right]$$

Proposition: Coverage of LP and AR

Robust coverage for LP:

$$\lim_{T\to\infty} P(\theta_{h,T} \in \text{CI}(\hat{\beta}_h)) = 1 - a.$$

Fragile coverage for AR: If $\rho \neq 0$ and $\text{aBias}(\hat{\delta}_h) \neq 0$,

$$\lim_{T\to\infty} P(\theta_{h,T} \in \text{CI}(\hat{\delta}_h)) = \begin{cases} 0 & \text{for } \zeta \in (1/4, 1/2), \\ < 1 - a & \text{for } \zeta = 1/2. \end{cases}$$

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

General local-to-SVAR(p) model

$$y_t = Ay_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$$

- y_t is n-dimensional, ε_t is m-dimensional.
- Encompasses general local-to-SVAR(p) models via companion form.
  - Allows estimation lag length $p >$ true lag length p_0 (VAR coef's $= 0$ at lags $> p_0$).
- Parameter of interest:

$$\theta_{h,T} \equiv \frac{\partial y_{i^*,t+h}}{\partial \varepsilon_{1,t}} = e'_{i^*,n} \left(A^h H + T^{-\zeta} \sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell \right) e_{1,m}.$$

General local-to-SVAR(p) model: Assumptions

$$y_t = Ay_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \theta_{h,T} = \partial y_{i^*,t+h}/\partial\varepsilon_{1,t}$$

1. $\varepsilon_t \overset{i.i.d.}{\sim} (0, D)$, $D = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$.
2. **Stationarity:** All absolute eigenvalues of $A < 1$.
3. **Approximately correct identification:** $H_{1,1} = 1$, $H_{1,j} = 0$ for $j = 2, \ldots, m$.
   - In paper: general recursive identification. IV/proxy identif'n is minor extension.
4. **Local misspecification:** $\zeta > 1/4$.
5. Regularity conditions on shocks and $\alpha(L)$.

Estimators

- **LP:** Coefficient $\hat{\beta}_h$ in OLS regression

$$y_{i^*,t+h} = \hat{\beta}_h y_{1,t} + \hat{\gamma}_h' y_{t-1} + \hat{\xi}_{h,t}.$$

- **VAR:** Run reduced-form OLS regression

$$y_t = \hat{A} y_{t-1} + \hat{u}_t,$$

  and report impulse response estimate

$$\hat{\delta}_h \equiv e_{i^*,n}' \hat{A}^h \hat{\nu},$$

  where $\hat{\nu}_i$ is OLS coef. in regr. of $\hat{u}_{i,t}$ on $\hat{u}_{1,t}$ (normalized Cholesky decomp'n).

- The two estimators coincide on impact: $\hat{\beta}_0 = \hat{\delta}_0$.

Asymptotic representations of LP and VAR

Proposition: Representations of LP and VAR

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{\text{LP},h,t} + o_p(T^{-1/2})$$

$$\hat{\delta}_h - \theta_{h,T} = T^{-\zeta}\,\text{aBias}(\hat{\delta}_h) + \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{\text{VAR},h,t} + o_p(T^{-\zeta} + T^{-1/2}),$$

where $\Upsilon_{\text{LP},h,t}$ and $\Upsilon_{\text{VAR},h,t}$ are the same as in the correctly specified case ($\alpha(L) = 0$).

- Qualitatively same bias-variance trade-off and coverage as in local-to-AR(1) model.
- If $h \leq p - p_0$, then $\text{aBias}(\hat{\delta}_h) = 0$ and LP & VAR are asy. equivalent.

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Outline

Restricting the amount of misspecification

- In the following, set $\zeta = 1/2$ to have nontrivial bias/variance trade-off:

$$y_t = Ay_{t-1} + H[I + T^{-1/2}\alpha(L)]\varepsilon_t.$$

- Noise-to-signal ratio in VAR error term:

$$\text{trace}\left\{\text{Var}(T^{-1/2}\alpha(L)\varepsilon_t)\,\text{Var}(\varepsilon_t)^{-1}\right\} = \text{trace}\left\{\left(T^{-1}\sum_{\ell=1}^{\infty}\alpha_\ell D\alpha_\ell'\right)D^{-1}\right\} = T^{-1}\|\alpha(L)\|^2,$$

where

$$\|\alpha(L)\| \equiv \sqrt{\sum_{\ell=1}^{\infty}\text{trace}\{D\alpha_\ell' D^{-1}\alpha_\ell\}}.$$

- Suppose we are willing to impose *a priori* bound on misspecification: $\|\alpha(L)\| \leq M$.

- Next: worst-case analysis over local parameter space $\{\|\alpha(L)\| \leq M\}$, treating the easier-to-estimate VAR parameters (A, H, D) as fixed.

Worst-case VAR bias: No free lunch

Proposition: Worst-case VAR bias

$$\max_{\|\alpha(L)\| \leq M} \left| \frac{\text{aBias}(\hat{\delta}_h)}{\sqrt{\text{aVar}(\hat{\delta}_h)}} \right| = M \sqrt{\frac{\text{aVar}(\hat{\beta}_h)}{\text{aVar}(\hat{\delta}_h)} - 1}.$$

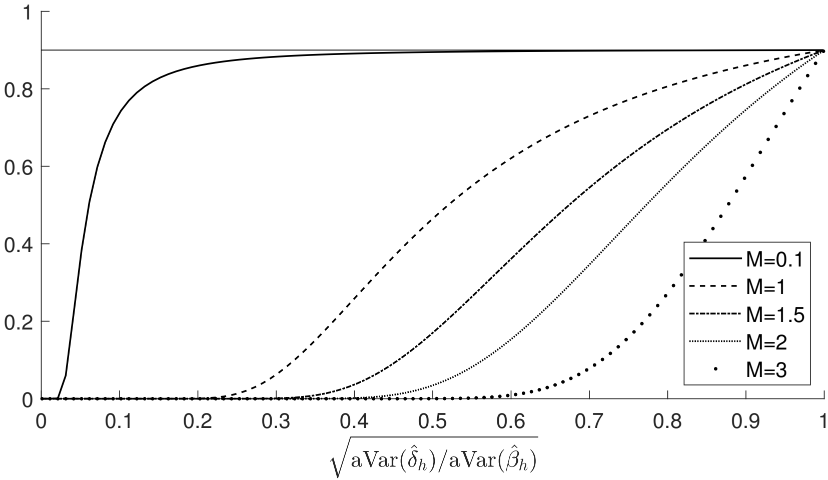
- Worst-case analysis in very large class of DGPs characterized by only 2 parameters!
  - Regardless of #variables n, lag length p, specific VAR parameters (A, H, D), and horizon h, worst-case scaled bias depends only on M and relative precision $\text{aVar}(\hat{\beta}_h)/\,\text{aVar}(\hat{\delta}_h)$.
- **No free lunch:** Worst-case (scaled) bias is large iff. relative precision of VAR is high.
  - Increasing VAR estimation lag length reduces worst-case bias, but *only* at expense of variance. If p is chosen so large that max bias $= 0$, then necessarily $\text{aVar}(\hat{\delta}_h) = \text{aVar}(\hat{\beta}_h)$.

▸ MSE

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Worst-case coverage of conventional 90% VAR CI



- For $M = 1$, worst-case coverage $< 48\%$ when $\sqrt{\mathrm{aVar}(\hat{\delta}_h)/\,\mathrm{aVar}(\hat{\beta}_h)} \leq 0.5$.

Not so easy to rule out the least favorable MA misspecification

- Difficult to rule out worst-case $\alpha^{\dagger}(L; h, M)$ based on *ex ante* theory:
  - Small (by definition).
  - Scales proportionally with M, decays exponentially as $\ell \to \infty$.
  - Numerically, tends to have Λ or $\surd$ shape, with largest value at $\ell = h$. Consistent with gradual/lumpy adjustment, time to build, info frictions, overshooting...
- Difficult to detect *ex post* with Hausman test of correct VAR specification:

$$\lim_{T\to\infty} P_{\alpha^{\dagger}(L;h,M)}\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)}} > z_{1-a/2}\right) = \begin{cases} 26\% & \text{for } M = 1,\ a = 10\%, \\ 17\% & \text{for } M = 1,\ a = 5\%. \end{cases}$$

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Bias-aware VAR CI

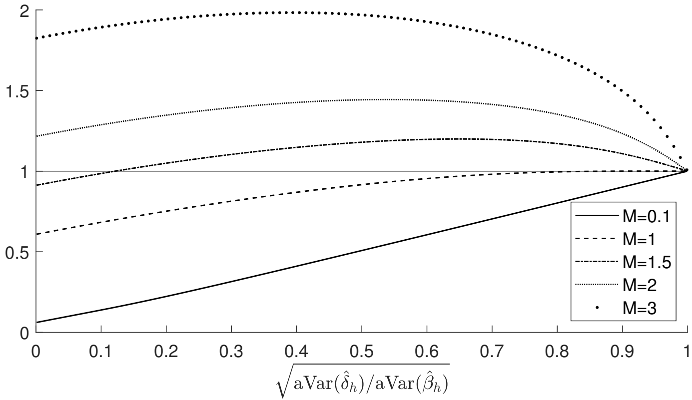
- Bias-aware CI: enlarge critical value to reflect worst-case bias. Armstrong & Kolesár (2021)

$$\mathrm{CI}_B(\hat{\delta}_h; M) \equiv \left[\hat{\delta}_h \pm \mathrm{cv}_{1-a}\left(M\sqrt{\frac{\mathrm{aVar}(\hat{\beta}_h)}{\mathrm{aVar}(\hat{\delta}_h)} - 1}\right)\sqrt{\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

where $P_{Z\sim N(0,1)}(|Z+b| > \mathrm{cv}_{1-a}(b)) = a$.

- Controls coverage by construction, as long as $\|\alpha(L)\| \leq M$.

Bias-aware 90% VAR CI: Length relative to LP CI

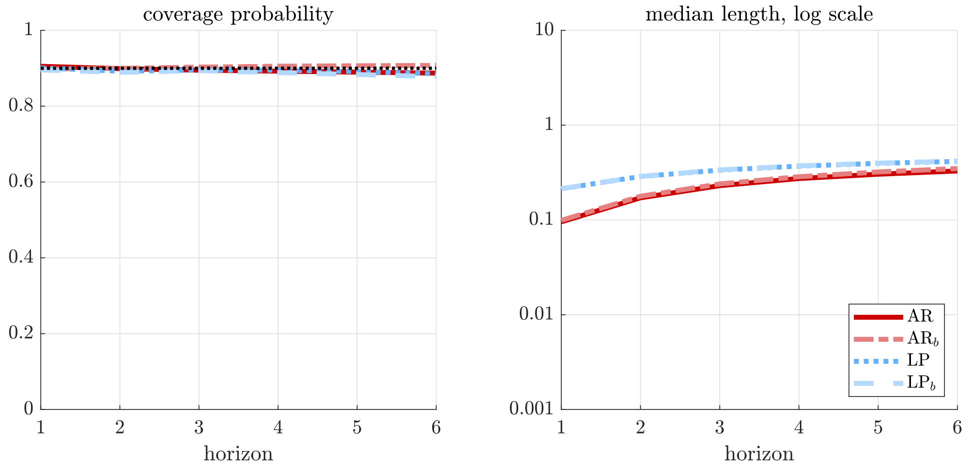


- For $M \geq 2$ (noise-to-signal ratio $\geq 4/T$), LP CI dominates bias-aware VAR CI.
- Also consider bias-aware CI centered at model avg. estimator $\omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$. Length-optimal ω yields only small gains over LP CI when $M \geq 2$.

Outline

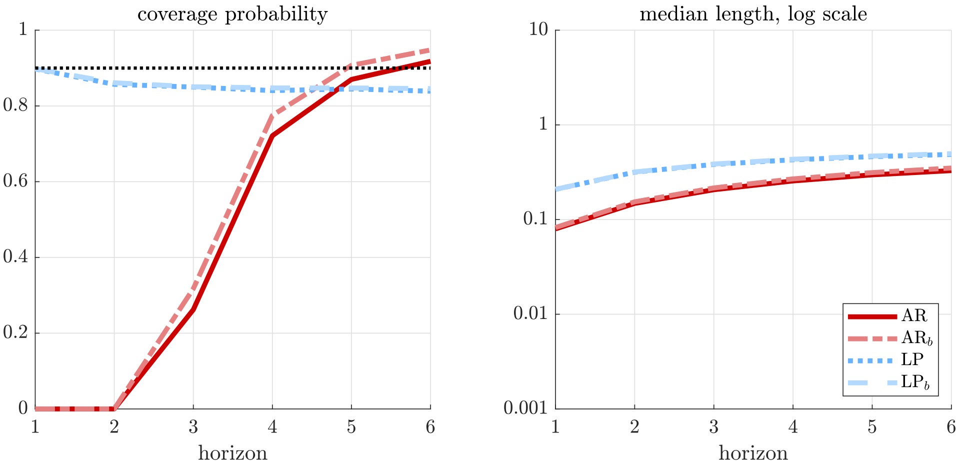
1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

AR(1) — correct specification



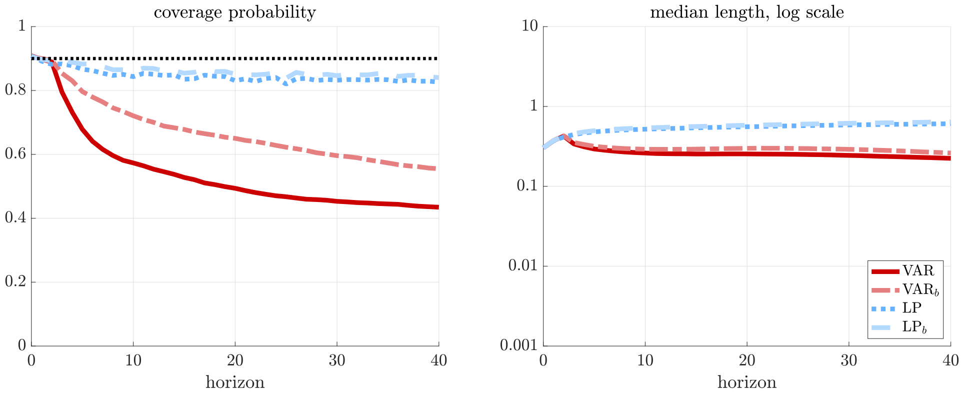
$$y_t = 0.9 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1), \quad p = 1, \quad T = 240$$

ARMA(1,1)



$$y_t = 0.9 y_{t-1} + \varepsilon_t + 0.25 \varepsilon_{t-1}, \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,1), \quad p = 1, \quad T = 240$$

Smets-Wouters DGP



- Smets & Wouters (2007) model (VARMA), posterior mode estimate.
  - $y_t =$ (cost-push shock, inflation, wage, hours). IRF: inflation wrt. cost-push shock.
  - $T = 240$. p selected by AIC (median $= 2$).

Outline

1. Robustness of LP, fragility of VAR
   - AR(1)
   - VAR(p)
2. Some unpleasant VARithmetic
   - Worst-case bias
   - Worst-case coverage
   - Bias-aware CI
3. Simulations
4. Conclusion

Conclusion

- LP robust to MA misspecification of order $T^{-1/4-\epsilon}$. Consequence of double robustness.
- Some unpleasant VARithmetic:
  - No free lunch: If we only constrain noise-to-signal ratio, then worst-case VAR bias is small precisely when aVar(VAR) $\approx$ aVar(LP).
  - Severe coverage distortion of VAR CI for difficult-to-detect MA terms $\propto T^{-1/2}$.
  - If we fix coverage with bias-aware critical value or $p \to \infty$, might as well do LP.
- How to rescue VARs?
  - Impose more elaborate restrictions on misspecification. (VARMA with prior on MA?)
  - Relax coverage criterion: average (over h) coverage, cover smooth projection of IRF, ...

Appendix

Literature

- Local misspecification in VAR forecasting: Schorfheide (2005); Müller & Stock (2011)
  - Our contributions: structural analysis, not just $T^{-1/2}$ MA misspec'n (double robustness of LP), worst-case bias, consequences for inference.
- LP vs. VAR simulations: Kilian & Kim (2011); Li, P-M & Wolf (2024)
- Order-T^{-1} bias of VAR and LP under correct specification: Pope (1990); Kilian (1998); Herbst & Johanssen (2023)
- Robustness of LP to long horizons and persistence: Montiel Olea & P-M (2021)
  - This paper: lag augmentation of LP also key to robustness to misspecification.
- Doubly robust: Newey (1990); Robins, Mark & Newey (1992); Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey & Robins (2018); Chernozhukov, Escanciano, Ichimura, Newey & Robins (2022)

Companion form

$$\check{y}_t = \sum_{\ell=1}^{p} \check{A}_\ell \check{y}_{t-\ell} + \check{H}[I + T^{-\zeta}\alpha(L)]\varepsilon_t$$

$$\Updownarrow$$

$$y_t = Ay_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \text{where}$$

$$y_t = \begin{pmatrix} \check{y}_t \\ \check{y}_{t-1} \\ \check{y}_{t-2} \\ \vdots \\ \check{y}_{t-p+1} \end{pmatrix}, \quad A = \begin{pmatrix} \check{A}_1 & \check{A}_2 & \dots & \check{A}_{p-1} & \check{A}_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}, \quad H = \begin{pmatrix} \check{H} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Robustness of LP to local misspecification: Why? (cont.)

- Consider any model (e.g., ARMA(∞, ∞)) that implies LP representation

$$y_{t+h} = \theta_{0,h} y_t + \gamma_0(y^{t-1}) + \xi_{h,t}, \quad \text{where} \quad \xi_{h,t} \perp\!\!\!\perp y^t \equiv (y_t, y_{t-1}, \dots).$$

- Define $\nu_0(y^{t-1}) \equiv E[y_t \mid y^{t-1}]$.
- By Frisch-Waugh, LP estimator $\hat{\beta}_h$ of $\theta_{0,h}$ solves sample analogue of moment cond'n

$$\begin{aligned} 0 &= E[\{y_{t+h} - \theta_{0,h} y_t - \gamma(y^{t-1})\}\{y_t - \nu(y^{t-1})\}] \\ &= E[\{\gamma(y^{t-1}) - \gamma_0(y^{t-1})\}\{\nu(y^{t-1}) - \nu_0(y^{t-1})\}]. \end{aligned}$$

- LP is **doubly robust** (like partially linear regression): Chernozhukov et al. (2018)
  - Consistent if *either* γ or ν is well-specified.
  - Estimated $\hat{\gamma}$ and $\hat{\nu}$ influence asy. distr'n of $\hat{\beta}_h$ only through product $\|\hat{\gamma} - \gamma_0\| \times \|\hat{\nu} - \nu_0\|$.
  - In local-to-AR(1) model, $\|\hat{\gamma} - \gamma_0\| \times \|\hat{\nu} - \nu_0\| = O_p(T^{-\zeta}) \times O_p(T^{-\zeta}) = o_p(T^{-1/2})$.

Double robustness

$$y_{t+h} = \theta_{0,h} y_t + \gamma_0(y^{t-1}) + \xi_{h,t}, \quad \text{where } \xi_{h,t} \perp\!\!\!\perp y^t \equiv (y_t, y_{t-1}, \dots)$$

$$\nu_0(y^{t-1}) \equiv E[y_t \mid y^{t-1}]$$

$$\begin{aligned}
& E[\{y_{t+h} - \theta_{0,h} y_t - \gamma(y^{t-1})\}\{y_t - \nu(y^{t-1})\}] \\
&= E[\{\underbrace{y_{t+h} - \theta_{0,h} y_t - \gamma_0(y^{t-1})}_{=\xi_{h,t} \perp\!\!\!\perp y^t} + \gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{y_t - \nu(y^{t-1})\}] \\
&= E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\underbrace{y_t - \nu_0(y^{t-1})}_{\perp y^{t-1}} + \nu_0(y^{t-1}) - \nu(y^{t-1})\}] \\
&= E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]
\end{aligned}$$

Asymptotic representations: Details

$$\Upsilon_{\text{LP},h,t} \equiv \frac{1}{\sigma_1^2}\xi_{h,i^*,t}\varepsilon_{1,t}$$

$$\Upsilon_{\text{VAR},h,t} \equiv \text{trace}\left\{S^{-1}\Psi_h H\varepsilon_t\tilde{y}_{t-1}'\right\} + \frac{1}{\sigma_1^2}e_{i^*,n}'A^h\xi_{0,t}\varepsilon_{1,t},$$

$$\text{aBias}(\hat{\delta}_h) \equiv \text{trace}\left\{S^{-1}\Psi_h H\sum_{\ell=1}^{\infty}\alpha_\ell DH'(A')^{\ell-1}\right\} - e_{i^*,n}'\sum_{\ell=1}^{h}A^{h-\ell}H\alpha_\ell e_{1,m},$$

where

$$\xi_{h,t} \equiv A^h\overline{H}_1\overline{\varepsilon}_{1,t} + \sum_{\ell=1}^{h}A^{h-\ell}H\varepsilon_{t+\ell}, \quad \overline{H}_1 = (H_{\bullet,2},\dots,H_{\bullet,m}), \quad \overline{\varepsilon}_{1,t} = (\varepsilon_{2,t},\dots,\varepsilon_{m,t})',$$

$$\Psi_h \equiv \sum_{\ell=1}^{h}A^{h-\ell}H_{\bullet,1}e_{i^*,n}'A^{\ell-1}.$$

aVar

Asymptotic variances

$$\text{aVar}(\hat{\beta}_h) = \frac{1}{\sigma_1^2}\left(e'_{i^*,n}A^h\overline{H}_1\overline{D}_1\overline{H}'_1(A')^h e_{i^*,n} + \sum_{\ell=1}^{h} e'_{i^*,n}A^{h-\ell}\Sigma(A')^{h-\ell}e_{i^*,n}\right),$$

$$\text{aVar}(\hat{\delta}_h) = \frac{1}{\sigma_1^2}e'_{i^*,n}A^h\overline{H}_1\overline{D}_1\overline{H}'_1(A')^h e_{i^*,n} + \text{trace}(\Psi_h\Sigma\Psi'_h S^{-1}),$$

where

$$\overline{D}_1 \equiv \text{diag}(\sigma_2^2, \ldots, \sigma_m^2).$$

The role of the lag length

- Local-to-SVAR(p_0) model:

$$\check{y}_t = \sum_{\ell=1}^{p_0} \check{A}_\ell \check{y}_{t-\ell} + \check{H}[I + T^{-\zeta}\alpha(L)]\varepsilon_t.$$

- Suppose we use $p \geq p_0$ lags for estimation.

Proposition: Lag length

Assume $\zeta = 1/2$. Then $T^{1/2}(\hat{\beta}_h - \hat{\delta}_h) = o_p(1)$ if either of the following two sufficient conditions hold:

i. $h \leq p - p_0$.

ii. Shock of interest is directly observed (i.e., $\check{A}_{1,j,\ell} = 0$ for all j, ℓ), and $h \leq p$.

Worst-case MSE comparison

Proposition: Worst-case MSE

Assume $\zeta = 1/2$ and $\text{aVar}(\hat{\beta}_h) > \text{aVar}(\hat{\delta}_h)$.

(i) Worst-case regret of VAR vs. LP:

$$\sup_{\|\alpha(L)\| \leq M} \{\text{aMSE}(\hat{\delta}_h) - \text{aMSE}(\hat{\beta}_h)\} = (M^2 - 1)\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

(ii) Minimax optimal *ex ante* model averaging weights:

$$\operatorname*{argmin}_{\omega \in \mathbb{R}} \sup_{\|\alpha(L)\| \leq M} \text{aMSE}\left(\omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h\right) = \frac{M^2}{1+M^2}.$$

- When $M > 1$ (noise-to-signal $> T^{-1}$), worst-case VAR regret is positive, and optimal model averaging weight on LP exceeds 50%.

Coverage of confidence intervals

Proposition: Coverage

$$\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a,$$

$$\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = \lim_{T\to\infty} \{1 - r\left(T^{1/2-\zeta} b_h; z_{1-a/2}\right)\} \quad (\text{if aVar}(\hat{\delta}_h) > 0),$$

where $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$ and $r(b; c) \equiv P_{Z\sim N(0,1)}(|Z + b| > c)$.

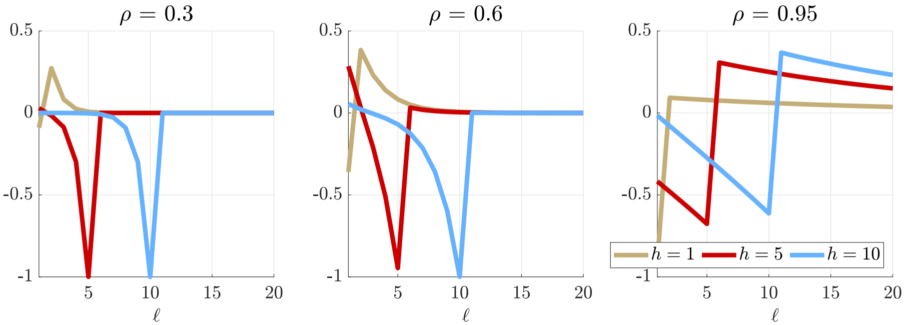
Least favorable MA misspecification

- General case:

$$\alpha^{\dagger}_{h,M,\ell} \propto D^{1/2} H' \Psi_h' S^{-1} A^{\ell-1} H D^{1/2} - \mathbb{1}(\ell \leq h) D^{1/2} H' (A')^{h-\ell} e_{i^*,n} e_{1,m}' D^{-1/2}.$$

- Local-to-AR(1) special case:

$$\alpha^{\dagger}_{h,M,\ell} \propto \underbrace{h\rho^{h-1}(1-\rho^2)\rho^{\ell-1}}_{\text{decreasing in } \ell} - \underbrace{\mathbb{1}(\ell \leq h)\rho^{h-\ell}}_{\text{increasing in } \ell \text{ (until } \ell = h)}.$$



Hausman test of correct VAR specification

- Hausman (1978) test comparing LP estimator (always consistent but inefficient) to VAR estimator (consistent and efficient under correct specif'n).

Proposition: Power of Hausman test

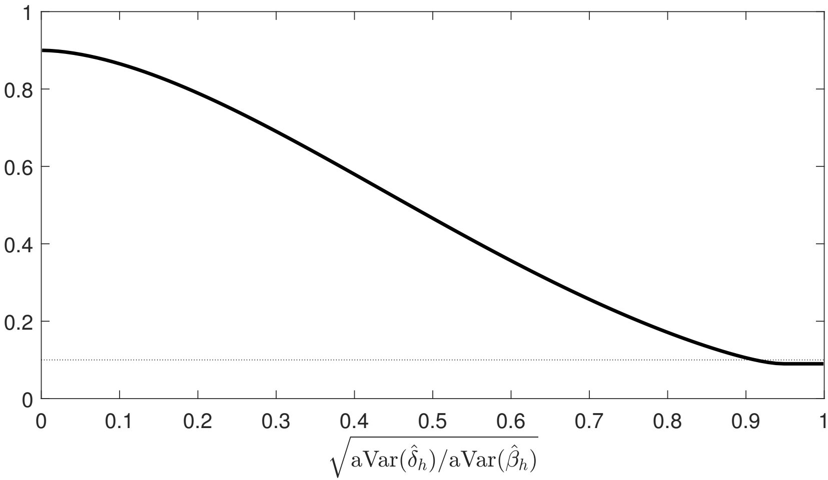
Assume $\zeta = 1/2$. Then $\{\hat{\beta}_h - \hat{\delta}_h\}$ is asymptotically independent of $\hat{\delta}_h$.

Moreover, if $\text{aVar}(\hat{\beta}_h) > \text{aVar}(\hat{\delta}_h) > 0$, then

$$\lim_{T\to\infty} P\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)}} > z_{1-a/2}\right) = r\left(\frac{b_h}{\sqrt{\text{aVar}(\hat{\beta}_h)/\,\text{aVar}(\hat{\delta}_h) - 1}}; z_{1-a/2}\right),$$

where $b_h \equiv \text{aBias}(\hat{\delta}_h)/\sqrt{\text{aVar}(\hat{\delta}_h)}$.

$\sup_{\alpha(L)} \lim_{T\to\infty} P(\text{Hausman test fails to reject} \cap \text{VAR CI doesn't cover})$



Supremum taken over <u>all</u> absolutely summable $\alpha(L)$. Dotted line: nominal signif. level 10%.

Optimal bias-aware CI

- Bias-aware CI centered at model averaging estimator $\hat{\theta}_h(\omega) = \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$:

$$\mathrm{CI}_B(\hat{\theta}_h(\omega); M) \equiv \left[\hat{\theta}_h(\omega) \pm \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\lambda}{\sqrt{1+\omega^2\lambda^2}}\right)\sqrt{(1+\omega^2\lambda^2)\,\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

  where $\lambda \equiv \sqrt{\mathrm{aVar}(\hat{\beta}_h)/\,\mathrm{aVar}(\hat{\delta}_h) - 1}$.

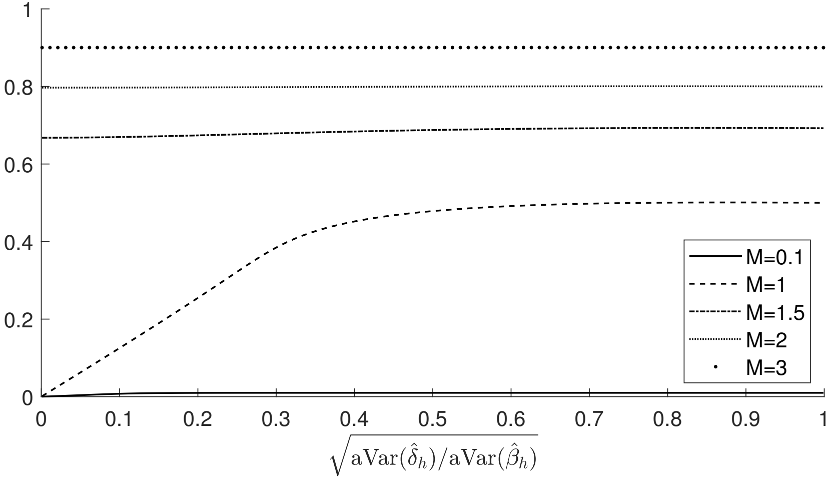
- $\omega = 1$: conventional LP CI. $\omega = 0$: bias-aware VAR CI.

- **Proposition** (by construction): controls asy. coverage regardless of ω.

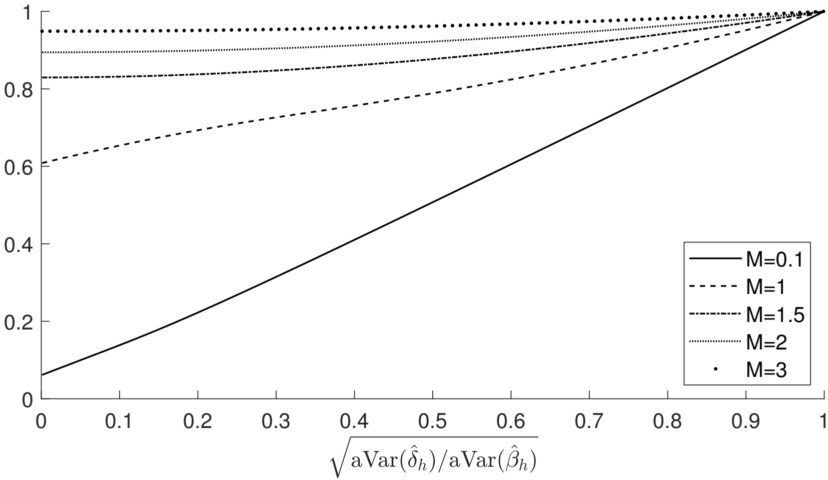
- Consider length-optimal choice of ω:

$$\omega^* \equiv \operatorname*{argmin}_{\omega\in[0,1]}\ \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\lambda}{\sqrt{1+\omega^2\lambda^2}}\right)\sqrt{1+\omega^2\lambda^2}.$$

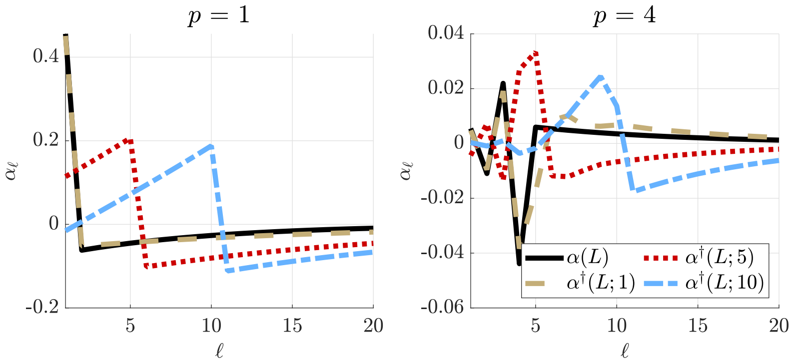
Optimal bias-aware 90% CI: Weight ω^* on LP



Optimal bias-aware 90% CI: Length relative to LP CI



ARMA(1,1): Close to worst case



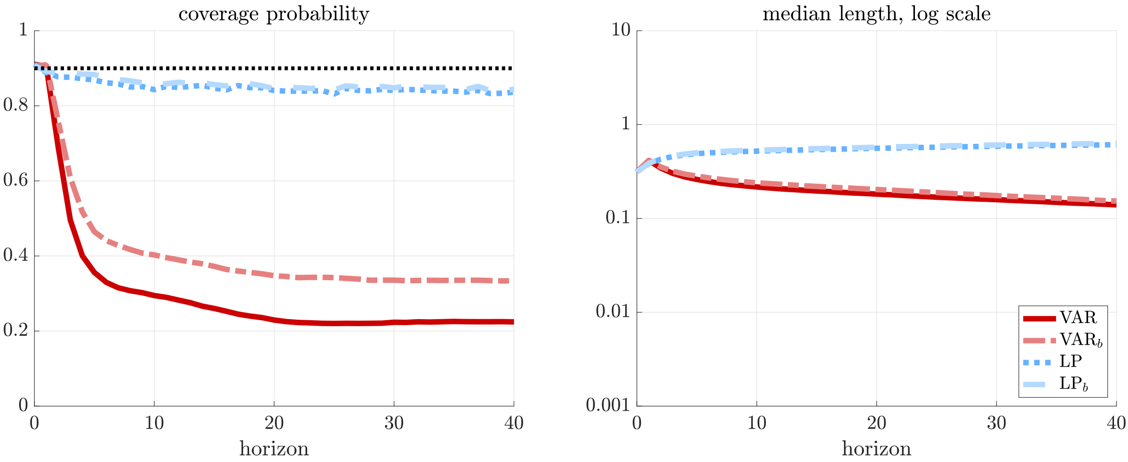
- Given $T = 240$, represent

$$y_t = 0.9y_{t-1} + \varepsilon_t + 0.25\varepsilon_{t-1} \quad \Longrightarrow \quad y_t = \sum_{\ell=1}^{p} A_\ell^* y_{t-\ell} + [1 + T^{-1/2}\alpha(L)]\varepsilon_t,$$

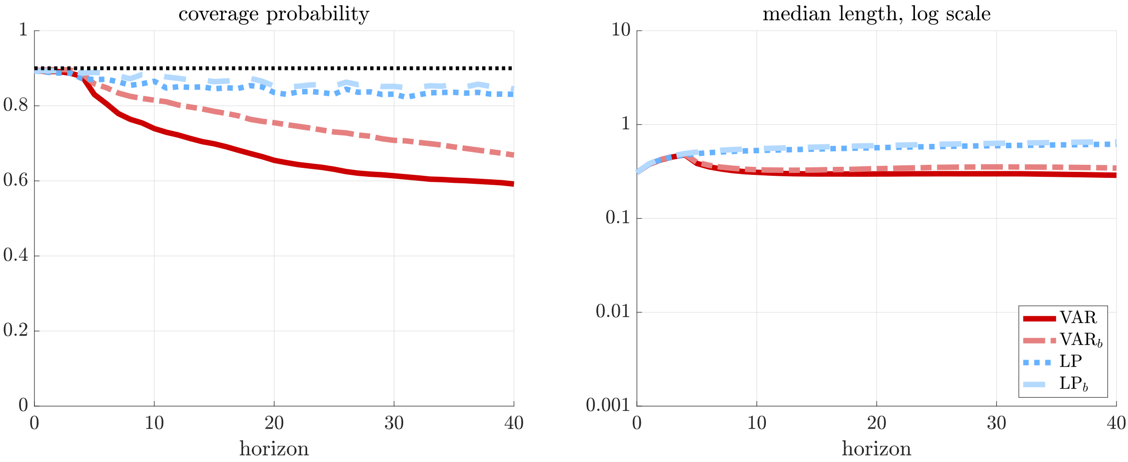
where A_ℓ^* are population regression coef's.

- For $p \in \{1, 4\}$, $\alpha(L)$ is close to worst case $\alpha^\dagger(L)$ at $h = 1$!

Smets-Wouters DGP: $p = 1$



Smets-Wouters DGP: $p = 4$



Smets-Wouters DGP: $p = 8$