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

October 16, 2024

Inference on impulse responses: LP or VAR?

• Impulse response:

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

• 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

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

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).
- LP \approx VAR with many lags \implies Only interesting comparison is with small/moderate lag length. P-M & Wolf (2021); Xu (2023)
- Bias-variance trade-off in simulations: Li, P-M & Wolf (2024)
 - VAR (moderate lag length) extrapolates: low variance, potentially high bias.
 - LP does not extrapolate: low bias, high variance.
- Dearth of analytical results on this central trade-off or its consequences for inference.

Our paper

- Study general locally misspecified VAR: VARMA with small (asy. vanishing) MA terms.
 - Idea: VAR is good but imperfect approximation to reality. E.g., slightly wrong specification of lags, controls, aggregation, etc.
- **1** LP CI (and VAR w/ many lags) is surprisingly **robust** to large misspecification.
 - Related to recent literature on double robustness in machine learning.
- **2 Unpleasant VARithmetic**: With moderate lag length, severe VAR coverage distortions for MA coef's that are small, economically plausible, and difficult to detect statistically.
- **3** No free lunch: VAR CI robust *only* if we use so many lags that VAR \approx LP.

Literature

- Misspecified VARs: Braun & Mittnik (1993); Schorfheide (2005); Müller & Stock (2011); González-Casasús & Schorfheide (2024)
 - Our contributions: structural analysis, general $T^{-\zeta}$ 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)
- $O(T^{-1})$ bias under correct spec'n: 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); Ai & Chen (2007); Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey & Robins (2018); Chernozhukov, Escanciano, Ichimura, Newey & Robins (2022)

Outline

1 Robustness of LP, fragility of VAR

2 Some unpleasant VARithmetic

- 3 Simulations
- 4 Conclusion

General local-to-SVAR(p) model



General local-to-SVAR(p) model

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

• Idea: VAR is good but imperfect approximation. Schorfheide (2005)

General local-to-SVAR(p) model

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

00

- Idea: VAR is good but imperfect approximation. Schorfheide (2005)
- 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:

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

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)
- Conjecture: Can allow general additive misspec'n: y_t = Ay_{t-1} + ε_t + T^{-ζ}υ_t, with param. of interest θ_h ≡ proj[y_{t+h} | ε_t = 1] proj[y_{t+h} | ε_t = 0].
 - Omitted nonlinearities, stationary time-varying parameters.

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 $\varepsilon_t \stackrel{i.i.d.}{\sim} (0, D), D = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$, shocks mutually indep'nt.

(ii) Stationarity: All absolute eigenvalues of A < 1.

(f) Approximately correct identification: $H_{1,1} = 1$, $H_{1,j} = 0$ for j = 2, ..., m.

- In paper: general recursive identification. IV/proxy identif'n is minor extension.
- **(v)** Local misspecification: $\zeta > 1/4$.
 - When $\zeta < 1/2$, misspecification detected w/ prob. \rightarrow 1 by conventional tests.
- Regularity conditions on shocks and $\alpha(L)$.

Estimators

(1) LP: Coefficient $\hat{\beta}_h$ in OLS regression

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

2 VAR: Run reduced-form OLS regression

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

and report impulse response estimate

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

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

$$\hat{\beta}_{h} - \theta_{h,T} = T^{-1} \sum_{t=1}^{T} \Upsilon_{\text{LP},h,t} + o_{p}(T^{-1/2}) \hat{\delta}_{h} - \theta_{h,T} = T^{-1} \sum_{t=1}^{T} \Upsilon_{\text{VAR},h,t} + T^{-\zeta} \operatorname{aBias}(\hat{\delta}_{h}) + o_{p}(T^{-\zeta} + T^{-1/2}),$$

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

- LP is robust to misspecification.
 - Even when $\zeta \in (1/4, 1/2)$, which is detectable w/ prob $\rightarrow 1$.
- VAR has bias of order $T^{-\zeta}$. Dominates std. dev. when $\zeta < 1/2$.
- If $h \leq p p_0$, then $\operatorname{aBias}(\hat{\delta}_h) = 0$ and LP & VAR are asy. equivalent.

Why is LP robust? (univariate case)

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

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

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

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

• Equivalent w/ double robustness in partially linear regression, recently exploited in machine learning literature. Newey (1990); Chernozhukov et al. (2018)

Asymptotic bias of VAR estimator: AR(1) case

• Consider univariate local-to-AR(1) special case:

$$y_t = \rho y_{t-1} + [1 + T^{-\zeta} \alpha(L)] \varepsilon_t.$$

• Then

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

- First term due to endogeneity bias in $\hat{\rho}$.
- Second term due to error in extrapolating from parametric formula ρ^h .
- Both terms can be reduced by increasing lag length.

Coverage of conventional confidence intervals

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

Proposition: Coverage of LP and VAR

Robust coverage for LP:

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

Fragile coverage for VAR: If $aVar(\hat{\delta}_h) > 0$ and $aBias(\hat{\delta}_h) \neq 0$,

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

 Intuition: VAR CI has correct length but wrongly centered due to bias. Concern for coverage ⇒ high concern for bias ⇒ LP preferred. Li, P-M & Wolf (2024)

Outline

1 Robustness of LP, fragility of VAR

- **2** Some unpleasant VARithmetic
- 3 Simulations
- 4 Conclusion

Restricting the amount of misspecification

- What if we're willing to assume that the misspecification is very small?
- 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:

$$\operatorname{trace}\left\{\operatorname{Var}(T^{-1/2}\alpha(L)\varepsilon_{t})\operatorname{Var}(\varepsilon_{t})^{-1}\right\} = T^{-1} \underbrace{\|\alpha(L)\|^{2}}_{\equiv \sum_{\ell=1}^{\infty}\operatorname{trace}\{D\alpha_{\ell}^{\prime}D^{-1}\alpha_{\ell}\}}$$

- Impose a priori bound on misspecification: $\|\alpha(L)\| \leq M$.
 - With T = 100, misspec'n contributes at most $\approx M^2\%$ of variance of error term.
- Next: worst-case analysis over local parameter space { ||α(L)|| ≤ 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
$$\max_{\|\alpha(L)\| \le M} \left| \frac{a \text{Bias}(\hat{\delta}_h)}{\sqrt{a \text{Var}(\hat{\delta}_h)}} \right| = M \sqrt{\frac{a \text{Var}(\hat{\beta}_h)}{a \text{Var}(\hat{\delta}_h)} - 1}.$$

- 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*.
- No free lunch: Worst-case (scaled) bias is small iff. VAR is almost as variable as LP.
 - 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 $aVar(\hat{\delta}_h) = aVar(\hat{\beta}_h)$.

Worst-case coverage of conventional 90% VAR CI



• For M = 1, worst-case coverage < 48% when relative SE ≤ 0.5 .

Worst-case coverage of conventional 90% VAR CI



 Shaded area: 10th−90th percentile range of SE ratios for impulse responses at horizons ≥ 1 year in 4 applications from Ramey (2016) handbook chapter.

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 γ 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:

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

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



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

Bias-aware VAR CI

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

$$\mathsf{Cl}_{B}(\hat{\delta}_{h};M) \equiv \left[\hat{\delta}_{h} \pm \mathsf{cv}_{1-a}\left(M\sqrt{rac{\mathsf{a}\mathsf{Var}(\hat{eta}_{h})}{\mathsf{a}\mathsf{Var}(\hat{\delta}_{h})}-1}
ight)\sqrt{\mathsf{a}\mathsf{Var}(\hat{\delta}_{h})/\mathcal{T}}
ight],$$

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

Controls coverage by construction, as long as ||α(L)|| ≤ M.

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



- For $M \ge 2$ (noise-to-signal ratio $\ge 4/T$), 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 $M \ge 2$.

Outline

1 Robustness of LP, fragility of VAR

2 Some unpleasant VARithmetic

- **3** Simulations
- 4 Conclusion

AR(1) — correct specification



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

ARMA(1,1)



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

ARMA(1,1): Close to worst case



• Given T = 240, represent

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



- Smets & Wouters (2007) model (VARMA), posterior mode estimate.
 - $y_t = (\text{cost-push shock, inflation, wage, hours})$. IRF: inflation wrt. cost-push shock.

•
$$T = 240$$
. p selected by AIC (median = 2)

No free lunch in action: p = 1



No free lunch in action: p = 4



No free lunch in action: p = 8



Outline

1 Robustness of LP, fragility of VAR

2 Some unpleasant VARithmetic

Simulations

4 Conclusion

Conclusion

- LP CI surprisingly robust to misspecification supports Jordà's (2005) claim.
- **Unpleasant VARithmetic**: VAR CI severely undercovers for misspec'n that is small, economically plausible, and difficult to detect.
- No free lunch: VARs only robust if we control for so many lags that VAR \approx LP.
- How to rescue VARs?
 - Smaller #lags with bias-aware crit. val. doesn't help.
 - Impose more elaborate, application-specific restrictions on misspecification.
 - Relax coverage criterion: average (over h) coverage, cover smooth projection of IRF, ...

Appendix

Companion form

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

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

Why is LP robust? (cont.)

• Consider any model (e.g., $\mathsf{ARMA}(\infty,\infty))$ that implies LP representation

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

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

$$0 = E[\{y_{t+h} - \theta_{0,h}y_t - \gamma(y^{t-1})\}\{y_t - \nu(y^{t-1})\}]$$

= E[{\gamma(y^{t-1}) - \gamma_0(y^{t-1})}\{\nu(y^{t-1}) - \nu_0(y^{t-1})\}].

- 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

$$y_{t+h} = \theta_{0,h} y_t + \gamma_0(y^{t-1}) + \xi_{h,t}, \quad \text{where } \xi_{h,t} \perp y^t \equiv (y_t, y_{t-1}, \dots)$$
$$\nu_0(y^{t-1}) \equiv E[y_t \mid y^{t-1}]$$

$$E[\{y_{t+h} - \theta_{0,h}y_t - \gamma(y^{t-1})\}\{y_t - \nu(y^{t-1})\}]$$

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

$$= E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\underbrace{y_t - \nu_0(y^{t-1})}_{\perp y^{t-1}} + \nu_0(y^{t-1}) - \nu(y^{t-1})\}]$$

$$= E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]$$

Asymptotic representations: Details

$$\begin{split} \Upsilon_{\mathsf{LP},h,t} &\equiv \frac{1}{\sigma_1^2} \xi_{h,i^*,t} \varepsilon_{1,t} \\ \Upsilon_{\mathsf{VAR},h,t} &\equiv \mathsf{trace} \left\{ S^{-1} \Psi_h H \varepsilon_t \tilde{y}_{t-1}' \right\} + \frac{1}{\sigma_1^2} e_{i^*,n}' A^h \xi_{0,t} \varepsilon_{1,t}, \\ \mathsf{aBias}(\hat{\delta}_h) &\equiv \mathsf{trace} \left\{ S^{-1} \Psi_h H \sum_{\ell=1}^{\infty} \alpha_\ell D H'(A')^{\ell-1} \right\} - e_{i^*,n}' \sum_{\ell=1}^h A^{h-\ell} H \alpha_\ell e_{1,m}, \end{split}$$

where

$$\begin{split} \xi_{h,t} &\equiv A^{h} \overline{H}_{1} \overline{\varepsilon}_{1,t} + \sum_{\ell=1}^{h} A^{h-\ell} H \varepsilon_{t+\ell}, \quad \overline{H}_{1} = (H_{\bullet,2}, \dots, H_{\bullet,m}), \quad \overline{\varepsilon}_{1,t} = (\varepsilon_{2,t}, \dots, \varepsilon_{m,t})', \\ \Psi_{h} &\equiv \sum_{\ell=1}^{h} A^{h-\ell} H_{\bullet,1} e_{i^{*},n}' A^{\ell-1}. \end{split}$$

Asymptotic variances

$$\begin{aligned} \mathsf{aVar}(\hat{\beta}_{h}) &= \frac{1}{\sigma_{1}^{2}} \left(e_{i^{*},n}^{\prime} A^{h} \overline{H}_{1} \overline{D}_{1} \overline{H}_{1}^{\prime} (A^{\prime})^{h} e_{i^{*},n} + \sum_{\ell=1}^{h} e_{i^{*},n}^{\prime} A^{h-\ell} \Sigma(A^{\prime})^{h-\ell} e_{i^{*},n} \right), \\ \mathsf{aVar}(\hat{\delta}_{h}) &= \frac{1}{\sigma_{1}^{2}} e_{i^{*},n}^{\prime} A^{h} \overline{H}_{1} \overline{D}_{1} \overline{H}_{1}^{\prime} (A^{\prime})^{h} e_{i^{*},n} + \mathsf{trace}(\Psi_{h} \Sigma \Psi_{h}^{\prime} S^{-1}), \end{aligned}$$

where

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

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 $p \ge p_0$ 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:

• $h \le p - p_0$.

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

Coverage of confidence intervals

Proposition: Coverage

$$\begin{split} &\lim_{T \to \infty} P(\theta_{h,T} \in \mathsf{CI}(\hat{\beta}_{h})) = 1 - a, \\ &\lim_{T \to \infty} P(\theta_{h,T} \in \mathsf{CI}(\hat{\delta}_{h})) = \lim_{T \to \infty} \left\{ 1 - r \left(T^{1/2 - \zeta} b_{h}; z_{1 - a/2} \right) \right\} \quad (\text{if } \mathsf{aVar}(\hat{\delta}_{h}) > 0), \\ &\text{where } b_{h} \equiv \mathsf{aBias}(\hat{\delta}_{h}) / \sqrt{\mathsf{aVar}(\hat{\delta}_{h})} \text{ and } r(b; c) \equiv P_{Z \sim N(0,1)}(|Z + b| > c). \end{split}$$

Worst-case MSE comparison

Proposition: Worst-case MSE

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

1 Worst-case regret of VAR vs. LP:

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

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

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

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

Worst-case bias: Multi-dimensional case

• Let $\hat{\beta}$ and $\hat{\delta}$ be k-dimensional vectors of LP and VAR impulse response estimates for different response variables i^* , shocks j^* , and/or horizons h.

Proposition: Multi-dimensional worst-case VAR bias

Let R be a constant matrix with k columns. Then

$$\max_{\alpha(L) \colon \|\alpha(L)\| \leq M} \|R \operatorname{aBias}(\hat{\delta})\|^2 = M^2 \lambda_{\mathsf{max}} \Big(R[\operatorname{aVar}(\hat{\beta}) - \operatorname{aVar}(\hat{\delta})] R' \Big)$$

- $\lambda_{max} = largest$ efficiency gain for VAR over LP across all linear combinations (with norm 1) of the estimated parameters.
- VAR CI has fragile coverage even if target parameter is the integral (sum) of impulse responses across multiple horizons.

Worst-case coverage of Wald ellipsoid



37

Least favorable MA misspecification

• General case:

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

• Local-to-AR(1) special case:



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
$$\operatorname{aVar}(\hat{eta}_h) > \operatorname{aVar}(\hat{eta}_h) > 0$$
, then

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

Optimal bias-aware CI

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

$$\mathsf{CI}_B(\hat{ heta}_h(\omega);M) \equiv \left[\hat{ heta}_h(\omega) \pm \mathsf{cv}_{1-s}\left(rac{(1-\omega)M\lambda}{\sqrt{1+\omega^2\lambda^2}}
ight)\sqrt{(1+\omega^2\lambda^2)}\,\mathsf{aVar}(\hat{\delta}_h)/T
ight],$$

where $\lambda \equiv \sqrt{\mathsf{aVar}(\hat{eta}_h)}/\mathsf{aVar}(\hat{eta}_h) - 1.$

- $\omega = 1$: conventional LP CI. $\omega = 0$: bias-aware VAR CI.
- **Proposition** (by construction): controls asy. coverage regardless of ω .
- Consider length-optimal choice of ω :

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

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



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

