Local Projection Inference is Simpler and More Robust Than You Think

José Luis Montiel Olea — Columbia University

Mikkel Plagborg-Møller — Princeton University

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

November 24

Inference on impulse responses

- **Impulse responses** are central objects for causal/counterfactual analysis in macro.
  - Response of $y_{i,t}$ due to exogenous policy change, at horizons $h = 1, 2, \ldots$
- How to do frequentist inference?

1. **Vector Autoregression (VAR):** Iterate on

$$y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t.$$

   Standard errors: delta method or bootstrap.

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

$$y_{i,t+h} = \beta_i(h)' y_t + \text{controls} + \xi_{i,t}(h), \quad h = 1, 2, \ldots$$

   Standard errors: HAC/HAR, since $\xi_{i,t}(h)$ is serially correlated.

VAR vs. LP

- Common issues in applied work:

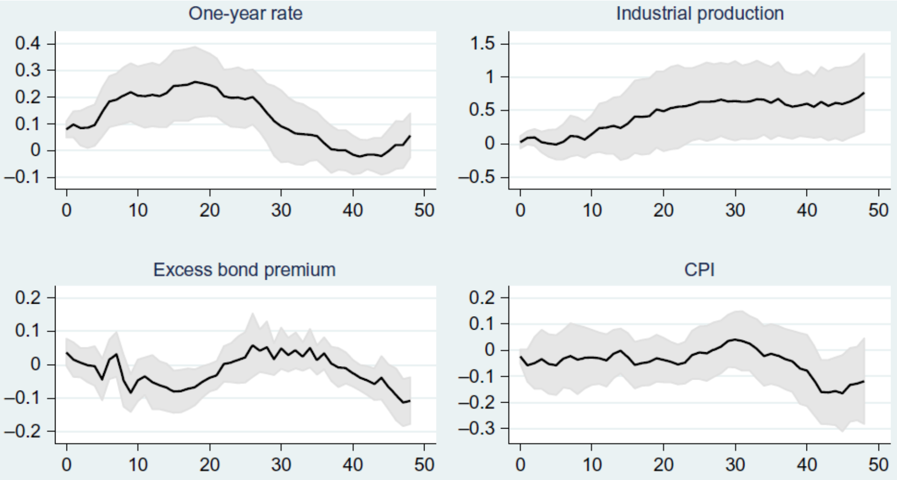
  1. Persistent data.
  2. Interest in long impulse response horizons h.

- VAR inference: Standard procedures break down in parts of parameter space.

  - (Near-)unit roots. Phillips (1998); Inoue & Kilian (2020)
  - Delta method singularity. Benkwitz, Neumann & Lütkepohl (2000)
  - Asymptotics: fixed h. Non-normal limit when $h = h_T \propto \sqrt{T}$ or $\propto T$.
    Wright (2000); Pesavento & Rossi (2006, 2007); Mikusheva (2012)

- LP inference: Despite widespread use, no theoretical comparisons with VARs.
  Jordà (2005); Kilian & Kim (2011); Brugnolini (2018)

Example: Ramey (2016) handbook chapter



Gertler-Karadi monetary shock, 90% CI. Sample: 1990m1–2012m6.

Largest horizon h = 18% of sample size T.

Our contributions

- Assume VAR(p) model.

1. Lag-augmented LP inference:

$$y_{i,t+h} = \hat{\beta}_i(h)' y_t + \sum_{\ell=1}^{p} \hat{\gamma}_{i,\ell}(h)' y_{t-\ell} + \hat{\xi}_{i,t}(h), \quad h = 1, 2, \ldots$$

2. Lag-augmented LP inference is uniformly valid over both...
    1. DGP. Includes unit root.
    2. Horizon h. Includes $h = h_T \propto T^{\eta}$ for $\eta \in [0, 1)$ (and $\propto T$ if no unit root).

3. Lag augmentation obviates need for HAC/HAR s.e.
    - Heteroskedasticity-robust (Eicker-Huber-White) s.e. suffice.
    - Simple. No need to choose HAR procedure, tuning parameters.

Related literature

- LP vs. VAR: identification, estimation. Plagborg-Møller & Wolf (2020)
- VAR inference.
  - Lag augmentation: Toda & Yamamoto (1995); Dolado & Lütkepohl (1996); Inoue & Kilian (2020)
  - Long horizons: Phillips (1998); Wright (2000); Gospodinov (2004); Pesavento & Rossi (2006, 2007); Mikusheva (2012); I&K (2020)
  - Uniformity: Mikusheva (2007, 2012); I&K (2020)
- LP inference.
  - Pointwise asymptotics: Jordà (2005); Kilian & Lütkepohl (2017); Stock & Watson (2018)
  - Lag augmentation: Dufour, Pelletier & Renault (2006); Breitung & Brüggemann (2020)

This paper: uniform LP inference, long+short horizons, simple s.e.

Outline

1. Overview of results: AR(1) case
2. Comparison with alternative methods: theory and simulations
3. General VAR(p) case
4. Empirical illustration
5. Conclusion

Model

- Start with univariate AR(1) model for clarity:

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, T, \quad y_0 = 0.$$

- Parameter of interest: impulse response at horizon h.

$$\beta(\rho, h) \equiv \rho^h, \quad \rho \in [-1, 1],\ h \in \mathbb{N}.$$

Assumption 1: Mean independence

$\{u_t\}$ is strictly stationary, and $E(u_t \mid \{u_s\}_{s \neq t}) = 0$.

- Stronger than MDS: $E(u_t \mid \{u_s\}_{s<t}) = 0$.
- Satisfied for i.i.d. u_t. Also allows many types of heteroskedasticity/SV.

Non-augmented local projection: fragile, HAR s.e.

- AR(1) model implies

$$y_{t+h} = \underbrace{\beta(\rho, h)}_{\equiv \rho^h} y_t + \underbrace{\xi_t(\rho, h)}_{\equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}} .$$

- Non-augmented LP estimator: regress y_{t+h} on y_t (no controls).
  - Consistent and asy. normal when $|\rho| \ll 1$.
  - Non-normal limit when $\rho \approx 1$ since y_t non-stationary.
  - Requires HAR s.e. even when $|\rho| \ll 1$, since $\xi_t(\rho, h)$ serially correlated. HAR inference challenging in small samples. Involves tuning parameters.
    Müller (2007, 2014); Lazarus, Lewis, Stock & Watson (2018)

Lag-augmented local projection: robust inference

- Lag-augmented LP:

$$\begin{pmatrix} \hat{\beta}(h) \\ \hat{\gamma}(h) \end{pmatrix} \equiv \left(\sum_{t=1}^{T-h} x_t x_t' \right)^{-1} \sum_{t=1}^{T-h} x_t y_{t+h}, \quad x_t \equiv (y_t, y_{t-1})'.$$

- Would get same $\hat{\beta}(h)$ if we regressed on (u_t, y_{t-1}), since $u_t = y_t - \rho y_{t-1}$.
- $\hat{\beta}(h)$ has uniform normal limit, since

$$y_{t+h} = \beta(\rho, h) \underbrace{u_t}_{\text{stationary}} + \beta(\rho, h+1) \underbrace{y_{t-1}}_{\text{control}} + \xi_t(\rho, h).$$

- $\hat{\gamma}(h)$ non-normal when $\rho \approx 1$, but we don't care.

Lag-augmented local projection: simple standard errors

$$y_{t+h} = \beta(\rho, h)u_t + \beta(\rho, h+1)y_{t-1} + \underbrace{\xi_t(\rho, h)}_{\equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}}$$

- Bonus: lag augmentation simplifies standard errors.
- Leading term in asymptotic expansion:

$$\hat{\beta}(h) \approx \beta(\rho, h) + \frac{\sum_{t=1}^{T-h} \xi_t(\rho, h) u_t}{\sum_{t=1}^{T-h} u_t^2}.$$

- $\xi_t(\rho, h)$ is serially correlated, but scores $\xi_t(\rho, h)u_t$ are not: For $s < t$,

$$E[\xi_t(\rho, h)u_t \xi_s(\rho, h)u_s] = E\big[\xi_t(\rho, h)u_t \xi_s(\rho, h) \underbrace{E(u_s \mid u_{s+1}, u_{s+2}, \dots)}_{=0}\big].$$

- Requires $E(u_t \mid \{u_s\}_{s>t}) = 0$. MDS is not enough.

Lag-augmented local projection: robust inference

- Heteroskedasticity-robust (Eicker-Huber-White) s.e. $\hat{s}(h)$ suffice. No tuning param's.
- Define usual confidence interval:

$$\hat{C}(h,\alpha) \equiv \left[\hat{\beta}(h) - z_{1-\alpha/2}\,\hat{s}(h)\,,\ \hat{\beta}(h) + z_{1-\alpha/2}\,\hat{s}(h)\right].$$

- Proposition: This CI is uniformly valid.

$$\inf_{\rho\in[-1,1]} \inf_{1\leq h\leq \bar{h}_T} P_\rho\left(\beta(\rho,h) \in \hat{C}(h,\alpha)\right) \to 1-\alpha,$$

  for any seq $\{\bar{h}_T\} \in \mathbb{N}$ such that $\bar{h}_T/T \to 0$.

- Further result: If we restrict $|\rho| \leq 1-a$ for $a > 0$, then even $\bar{h}_T \propto T$ is OK.
- Non-normal limit for $\rho = 1$, $h = h_T \propto T$.

Outline

1. Overview of results: AR(1) case
2. Comparison with alternative methods: theory and simulations
3. General VAR(p) case
4. Empirical illustration
5. Conclusion

Fragility of AR inference

- Simple lag-aug LP inference robust to persistence ρ and horizon h. Not true for textbook AR delta method inference.
  Phillips (1998); Benkwitz et al. (2000); Inoue & Kilian (2002, 2020); Pesavento & Rossi (2007)
- Lag-augmented AR: $\hat{\beta}_{\text{ARLA}}(h) \equiv \hat{\rho}_1^h$, where $y_t = \hat{\rho}_1 y_{t-1} + \hat{\rho}_2 y_{t-2} + \hat{u}_t$.
  - Uniformly $\sqrt{T}$-normal limit for fixed h.
  - Efron bootstrap CI valid at long horizons. Inoue & Kilian (2020)
  - But estimator is inconsistent at horzs $h = h_T \geq \kappa\sqrt{T}$ when $\rho \approx 1$. Confidence interval does not shrink with T (length can even explode), unlike LP.
- AR grid bootstrap valid at short+long horizons, but *not* intermediate. Computationally intensive. Hansen (1999); Mikusheva (2012)

Simulation study

- Confidence interval procedures:
  1. Non-augmented AR, delta method s.e. (straw man).
  2. Lag-augmented AR, Efron bootstrap CI. Inoue & Kilian (2020)
  3. Non-augmented LP, percentile-t bootstrap CI, HAR s.e.
  4. Lag-augmented LP, percentile-t bootstrap CI, EHW s.e.
- Bootstrap: wild recursive AR design.
- AR(1) model. $T = 240$. Nominal confidence level: 90%.
- $u_t \sim N(0, 1)$ i.i.d. (ARCH innovations qualitatively similar.)

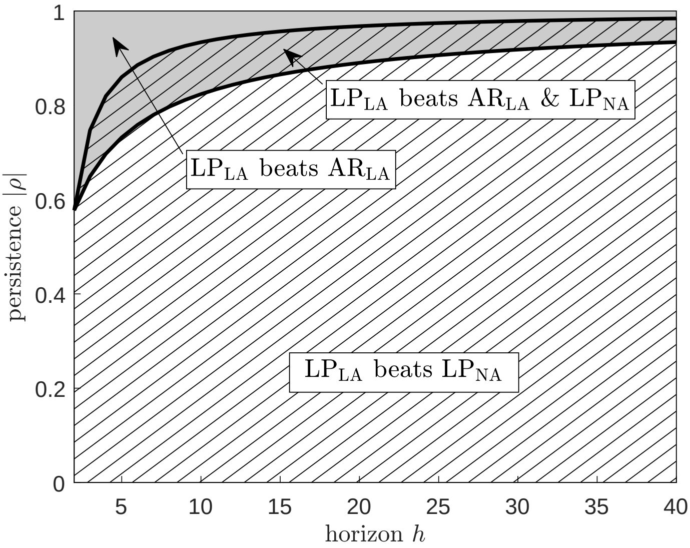
h	Coverage AR_d	AR_b^*	LP_b	LP_b^*	Median length AR_d	AR_b^*	LP_b	LP_b^*
$\rho = 0.50$								
1	0.897	0.897	0.912	0.906	0.184	0.211	0.205	0.219
6	0.832	0.897	0.906	0.895	0.032	0.046	0.293	0.252
12	0.766	0.897	0.903	0.906	0.001	0.002	0.293	0.255
36	0.643	0.897	0.901	0.900	0.000	0.000	0.309	0.271
60	0.595	0.897	0.903	0.905	0.000	0.000	0.333	0.291
$\rho = 0.95$								
1	0.850	0.882	0.842	0.892	0.075	0.212	0.076	0.220
6	0.810	0.882	0.851	0.903	0.318	1.011	0.395	0.523
12	0.769	0.882	0.853	0.889	0.430	1.744	0.644	0.678
36	0.656	0.882	0.865	0.885	0.272	6.567	0.859	0.728
60	0.595	0.882	0.892	0.892	0.095	23.050	0.942	0.731
$\rho = 1.00$								
1	0.532	0.877	0.820	0.895	0.039	0.210	0.040	0.219
6	0.494	0.877	0.836	0.875	0.214	1.206	0.243	0.564
12	0.459	0.877	0.827	0.843	0.379	2.553	0.477	0.821
36	0.348	0.877	0.755	0.741	0.670	21.107	1.200	1.338
60	0.295	0.877	0.712	0.642	0.731	161.250	1.667	1.434

D

A

Relative efficiency in stationary, fixed-horizon case

- Consider a stationary, homoskedastic VAR model. Fix the horizon h.
- Then textbook non-augmented AR estimator achieves semiparametric efficiency bound.
- Ambiguous ranking of asymptotic variances of inefficient procedures:
  1. Lag-augmented AR.
  2. Lag-augmented LP.
  3. Non-augmented LP (requires HAC standard errors).
- Next slide: Ranking in homoskedastic AR(1) model, as function of (ρ, h).



Trade-off $\mathrm{LP}_{\mathrm{LA}}$ vs. $\mathrm{AR}_{\mathrm{LA}}$:

- Non-linear transformation ρ^h.

Trade-off $\mathrm{LP}_{\mathrm{LA}}$ vs. $\mathrm{LP}_{\mathrm{NA}}$:

- Effective regressor u_t vs. y_t.
- Serial correl'n of $\xi_t(\rho, h)y_t$.

Outline

1. Overview of results: AR(1) case
2. Comparison with alternative methods: theory and simulations
3. General VAR(p) case
4. Empirical illustration
5. Conclusion

VAR(p) model

- VAR(p) model for n-dimensional data vector:

$$y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t, \quad t = 1, \dots, T, \quad y_0 = \cdots = y_{1-p} = 0.$$

- Still impose conditional mean independence: $E(u_t \mid \{u_s\}_{s \neq t}) = 0$.
- Known lag length p.
- Reduced-form impulse responses of $y_{1,t}$ at horizon h: $\beta_1(A, h) \in \mathbb{R}^n$.
- Parameter of interest: $\nu' \beta_1(A, h)$, where $\nu \in \mathbb{R}^n$, $\nu \neq 0$.
  - Simple extension: joint inference on vector $\beta_1(A, h)$.
  - Extensions for future work: structural impulse responses, deterministic dynamics.

Multivariate lag-augmented local projection

- VAR model implies: Jordà (2005); Kilian & Lütkepohl (2017)

$$y_{1,t+h} = \beta_1(A,h)' y_t + \sum_{\ell=1}^{p-1} \delta_{1,\ell}(A,h)' y_{t-j} + \xi_{1,t}(A,h), \quad \xi_{1,t}(A,h) \equiv \sum_{\ell=1}^{h} \beta_1(A,h-\ell)' u_{t+\ell}.$$

- Lag-augmented LP estimator: regression with p lags as controls.

$$y_{1,t+h} = \hat{\beta}_1(h)' y_t + \sum_{\ell=1}^{p} \hat{\delta}_{1,\ell}(h)' y_{t-j} + \hat{\xi}_{1,t}(A,h).$$

- Confidence interval for $\nu'\beta_1(A,h)$:

$$\hat{C}_1(h,\nu,\alpha) \equiv \nu'\hat{\beta}_1(h) \pm z_{1-\alpha/2} \underbrace{\hat{s}_1(h,\nu)}_{\text{EHW}}.$$

Uniformly valid multivariate LP inference

- Proposition: Impose As'n "Mean independence". Then the CI $\hat{C}_1(h, \nu, \alpha)$ is asymptotically uniformly valid over the DGP and horizon h, provided...

1. The VAR(p) model can be written in the form $y_t = \Lambda y_{t-1} + \tilde{y}_t$, where: Phillips (1988)
   - $\tilde{y}_t$ is uniformly stationary VAR($p-1$) (geometrically decaying IRFs).
   - $\Lambda = \text{diag}(\rho_1, \dots, \rho_n)$ with $\rho_i \in [-1, 1]$. Mikusheva (2012)
2. Either:
   1. $h \leq (1-a)T$ and $|\rho_i| \leq 1-a$, $i = 1, \dots, n$, where $a > 0$. OR:
   2. $h \leq \bar{h}_T$, where $\bar{h}_T / T \to 0$.
3. Further regularity conditions: moments, uniform non-singularity of OLS denominator.

Uniformly valid LP inference: key proof challenges, AR(1) case

- Study all drifting sequences $\{\rho_T, h_T\}$. Andrews, Cheng & Guggenberger (2019)

$$\hat{\beta}(h_T) - \beta(\rho_T, h_T) \propto \sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T)u_t + \sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) \underbrace{[\hat{u}_t(h_T) - u_t]}_{=[\rho_T - \hat{\rho}(h_T)]y_{t-1}}$$

- 1st term: $E[\xi_t(\rho_T, h_T)^2 u_t^2] \to \infty$ if $\rho_T \to 1$, $h_T \to \infty$.
  - CLT for MDS. Must "reverse time" b/c $E[\xi_t(\rho, h)u_t \mid \{u_s\}_{s<t+h}] \neq 0$.
  - Explicitly calculate uniform moment bounds.
- 2nd term: Convergence rate of $\hat{\rho}(h_T)$ depends on whether $\rho_T \approx 1$. Mikusheva (2007)
  - Explicitly calculate uniform moment bounds.

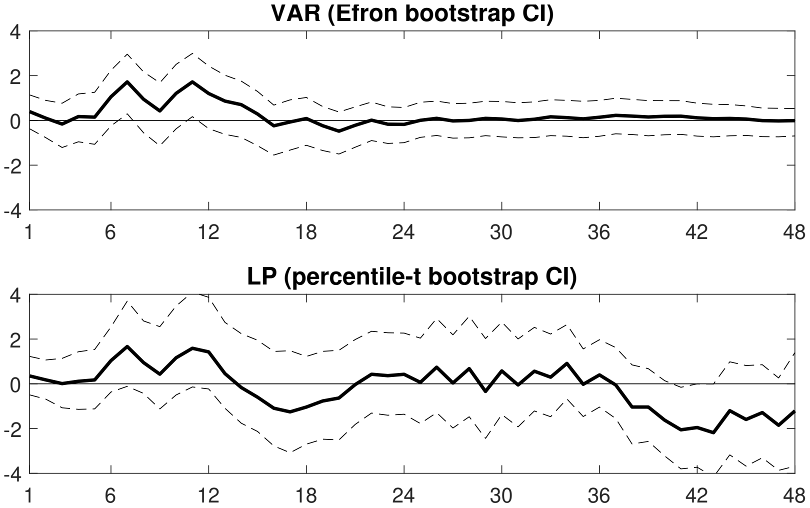
Uniformly valid multivariate LP inference: discussion

- VAR(p) proof follows AR(1) intuition. Main challenge: uniform bounds on IRFs.
- We are not aware of other uniformity results that allow multiple (near-)unit roots.
- Corollary: Can allow for cointegration among control variables $y_{2,t}, \dots, y_{n,t}$.
- Possibly non-normal limit when $\rho_i \approx 1$ for some i and $h \propto T$.

Outline

1. Overview of results: AR(1) case
2. Comparison with alternative methods: theory and simulations
3. General VAR(p) case
4. Empirical illustration
5. Conclusion

Response of Excess Bond Premium to 100 bp monetary shock



Data: Gertler & Karadi (2015). Monetary shock: high-freq. change in 3-mth Fed Funds futures. $p = 12$. $1 - \alpha = 90\%$. Horiz. axis: months. Vert. axis: percentage points.

Outline

1. Overview of results: AR(1) case
2. Comparison with alternative methods: theory and simulations
3. General VAR(p) case
4. Empirical illustration
5. Conclusion

Conclusion

- Local projections already popular in applied work. We provide statistical rationale.
- Lag-augmented LP inference on impulse responses is robust to:
  1. Persistence of data.
  2. Length of impulse response horizon.
- Efficiency loss for stationary DGPs at short horizons, but modest in absolute terms.
- Lag augmentation obviates need for HAR s.e. Simple!
- Only known VAR-based methods with comparable robustness are either computationally demanding or can yield very long CIs. Mikusheva (2012); Inoue & Kilian (2020)

Conclusion

- Local projections already popular in applied work. We provide statistical rationale.
- Lag-augmented LP inference on impulse responses is robust to:
  1. Persistence of data.
  2. Length of impulse response horizon.
- Efficiency loss for stationary DGPs at short horizons, but modest in absolute terms.
- Lag augmentation obviates need for HAR s.e. Simple!
- Only known VAR-based methods with comparable robustness are either computationally demanding or can yield very long CIs. Mikusheva (2012); Inoue & Kilian (2020)

Thank you!

Appendix

Heteroskedastic innovations

- Assume $u_t = \tau_t \epsilon_t$, where $\tau_t \geq 0$.
- Assume ϵ_t is i.i.d., $E(\epsilon_t) = 0$.
- Then $E(u_t \mid \{u_s\}_{s \neq t}) = 0$ if either...
  - i) $\{\tau_t\} \perp\!\!\!\perp \{\epsilon_t\}$ (SV).
  - ii) $\tau_t = f(\epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots)$ and distribution of ϵ_t is symmetric (GARCH).

Lag-augmented local projection: simple standard errors

- Heteroskedasticity-robust (Eicker-Huber-White) s.e. suffice:

$$\hat{s}(h) \equiv \frac{\left(\sum_{t=1}^{T-h} \hat{\xi}_t(h)^2 \hat{u}_t(h)^2\right)^{1/2}}{\sum_{t=1}^{T-h} \hat{u}_t(h)^2},$$

  where

$$\hat{\xi}_t(h) \equiv y_{t+h} - \hat{\beta}(h) y_t - \hat{\gamma}(h) y_{t-1},$$
$$\hat{u}_t(h) \equiv y_t - \hat{\rho}(h) y_{t-1},$$
$$\hat{\rho}(h) \equiv \left(\sum_{t=1}^{T-h} y_t y_{t-1}\right) / \left(\sum_{t=1}^{T-h} y_{t-1}^2\right).$$

- Readily computed by standard statistical software.

- No tuning parameters.

AR inference: medium-long horizons

- Suppose $\rho > 0$ and we use some asy. normal estimator $\hat{\rho}$:

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \tau^2).$$

- Delta method s.e.:

$$\text{se}(\hat{\rho}) \equiv h|\hat{\rho}|^{h-1}\hat{\tau}/\sqrt{T}, \quad \hat{\tau} \xrightarrow{p} \tau.$$

- Then at horizon $h = h_T = \sqrt{T}$,

$$\frac{\hat{\rho}^{h_T}}{\rho^{h_T}} = e^{\sqrt{T}(\log \hat{\rho} - \log \rho)} \xrightarrow{d} e^{N(0,\tau^2/\rho^2)},$$

$$\left| \frac{\sqrt{T}(\hat{\rho}^{h_T} - \rho^{h_T})}{h_T \hat{\rho}^{h_T - 1}\hat{\tau}} \right| = \left| \frac{\hat{\rho}}{\hat{\tau}} \left(1 - \frac{\hat{\rho}^{h_T}}{\rho^{h_T}}\right) \right| \xrightarrow{d} \left| \frac{\rho}{\tau} \left(1 - N(0, \tau^2/\rho^2)\right) \right|.$$

Simulation results: delta method procedures

h	Coverage AR_d	AR_d^*	LP_d	LP_d^*	Median length AR_d	AR_d^*	LP_d	LP_d^*
				$\rho = 0.50$				
1	0.897	0.898	0.885	0.896	0.184	0.212	0.187	0.212
6	0.832	0.858	0.875	0.886	0.032	0.040	0.266	0.245
12	0.766	0.802	0.887	0.894	0.001	0.001	0.280	0.248
36	0.643	0.678	0.884	0.889	0.000	0.000	0.296	0.262
60	0.595	0.631	0.880	0.891	0.000	0.000	0.316	0.279
				$\rho = 0.95$				
1	0.850	0.881	0.827	0.878	0.075	0.212	0.072	0.212
6	0.810	0.873	0.789	0.838	0.318	0.975	0.345	0.452
12	0.769	0.845	0.752	0.806	0.430	1.436	0.518	0.550
36	0.656	0.736	0.674	0.814	0.272	1.249	0.612	0.625
60	0.595	0.682	0.693	0.833	0.095	0.603	0.641	0.651
				$\rho = 1.00$				
1	0.532	0.883	0.554	0.874	0.039	0.211	0.040	0.211
6	0.494	0.851	0.503	0.777	0.214	1.205	0.222	0.498
12	0.459	0.801	0.429	0.676	0.379	2.253	0.385	0.671
36	0.348	0.683	0.200	0.428	0.670	5.239	0.592	0.950
60	0.295	0.632	0.156	0.276	0.731	6.764	0.637	0.978

Simulation results: ARCH innovations ($\alpha_1 = 0.7$)

h	Coverage AR_d	AR_b^*	LP_b	LP_b^*	Median length AR_d	AR_b^*	LP_b	LP_b^*
$\rho = 0.50$								
1	0.874	0.836	0.908	0.891	0.294	0.336	0.330	0.387
6	0.776	0.837	0.908	0.900	0.048	0.090	0.272	0.246
12	0.689	0.837	0.896	0.904	0.001	0.008	0.265	0.240
36	0.579	0.837	0.894	0.897	0.000	0.000	0.277	0.254
60	0.540	0.837	0.902	0.901	0.000	0.000	0.300	0.273
$\rho = 0.95$								
1	0.856	0.838	0.823	0.897	0.086	0.335	0.084	0.392
6	0.806	0.838	0.854	0.896	0.355	1.746	0.381	0.621
12	0.758	0.838	0.850	0.880	0.467	3.942	0.604	0.724
36	0.643	0.838	0.859	0.869	0.291	64.319	0.816	0.717
60	0.579	0.838	0.885	0.881	0.095	1032.604	0.900	0.711
$\rho = 1.00$								
1	0.560	0.839	0.841	0.896	0.041	0.330	0.040	0.386
6	0.513	0.839	0.859	0.879	0.223	2.035	0.240	0.686
12	0.468	0.839	0.845	0.854	0.391	5.510	0.473	0.902
36	0.352	0.839	0.752	0.731	0.669	177.260	1.170	1.384
60	0.294	0.839	0.697	0.640	0.729	5593.663	1.647	1.475

VAR parameter space

- Let there be given constants $a \in [0,1)$, $C > 0$, and $\epsilon \in (0,1)$.
- $\mathcal{A}(a, C, \epsilon) \equiv$ space of autoregressive coefficients $A = (A_1, \ldots, A_p)$ such that the associated lag polynomial $A(L) = I_n - \sum_{\ell=1}^{p} A_\ell L^\ell$ admits the factorization

$$A(L) = B(L)(I_n - \text{diag}(\rho_1, \ldots, \rho_n)L).$$

- $\rho_i \in [a-1, 1-a]$ for all $i = 1, \ldots, n$.
- $B(L)$ is a lag polynomial of order $p-1$ with companion matrix $\mathbf{B}$ satisfying $\|\mathbf{B}^\ell\| \leq C(1-\epsilon)^\ell$ for all $\ell = 1, 2, \ldots$.

Proof sketch: reversing time

$$\xi_t(\rho, h) \equiv \sum_{\ell=1}^{h} \rho^{h-\ell} u_{t+\ell}$$

- Run sum "backwards in time":
$$\sum_{t=1}^{T-h_T} \xi_t(\rho_T, h_T) u_t = \sum_{t=1}^{T-h_T} \chi_{T,t}, \quad \chi_{T,t} \equiv \xi_{T-h-t+1}(\rho_T, h_T) u_{T-h-t+1}.$$
- Define filtration
$$\mathcal{F}_{T,t} \equiv \sigma(u_{T-h_T-t+1}, u_{T-h_T-t+2}, \dots).$$
Then $\chi_{T,t} \in \mathcal{F}_{T,t}$ and $\mathcal{F}_{T,t} \subset \mathcal{F}_{T,t+1}$ for all t.
- $\{\chi_{T,t}, \mathcal{F}_{T,t}\}$ is a martingale difference array:

$$E(\chi_{T,t} \mid \mathcal{F}_{T,t-1}) = \xi_{T-h_T-t+1}(\rho_T, h_T) \underbrace{E(u_{T-h_T-t+1} \mid \{u_{T-h_T-t+s}\}_{s>1})}_{=0 \text{ by As'n "Mean independence"}}$$