Lecture 4: Estimation with time-series data

Diego Känzig
Northwestern University

416-2 Empirical Macro, Winter 2025

Overview

- So far: population arguments for **identification** of SVMA coefficients
  - Basic idea: use identifying information (e.g. invertibility $+$X, IVs, ...) to assign a **causal interpretation** to certain estimable moments of aggregate time series data
  - Question: how should we best *estimate* those second-moment properties in finite samples?
- Today's lecture: **VARs** vs. **local projections**
  - We'll see that they are just two different techniques for estimating second moments. In population, under same identifying assumptions, they yield the same objects
  - Finite-sample recommendations: should we use LPs? VARs? something else?

Outline

1. VARs vs. local projections

   Recursive case

   Identification using IVs

2. Finite-sample recommendations

   A menu of estimation strategies

   Insights from simulation studies

VARs vs. local projections

- Dominant methods in semi-structural time series: **VARs** and **local projections**
  Going back to Sims (1980) and Jordà (2005)
- Recall the key challenges with the VAR approach:
  - Results may be biased if VAR does not span all relevant information, i.e. is non-invertible
  - VAR may not be an adequate representation of the dynamics of the variables in the system because we have to truncate the VAR(∞) to a VAR with p lags

Are local projections a panacea?

- Local projections very influential approach to estimate dynamic causal effects (Jordà, 2005; Jordà, Schularick, and Taylor, 2015; Ramey and Zubairy, 2018)
  - motivated by the SVMA representation
- **Idea**: estimate responses directly by regressing future outcomes on shock measure and controls
- Key advantage: very simple to implement, straightforward to allow for non-linearities

VARs vs. local projections

- Can find a lot of claims in the literature that these are somehow **fundamentally different** methods
  - LP rely on less restrictive identifying assumptions, more robust to misspecification
  - VARs more efficient
- We will see that this is **not** the case
- The argument will proceed in two steps:
  1. Define VAR and LP estimators
  2. Show that their estimands are the same, using linear projection arguments

  None of this will rely on an underlying SVMA model
- Rather, the point of the SVMA model + invertibility/proxies/... is to establish that this common LP/VAR estimand is actually **structurally interesting**

Setting and estimators

- **Data-generating process**
  - Denote our time-series data $\{w_t\}$. For simplicity, we assume that w_t are covariance stationary and nondeterministic, with absolutely summable Wold representation coefficients
  - Partition the data as $w_t = (r_t', x_t, y_t, q_t')$ where r_t and q_t are "controls", and we are interested in the response of y_t to an impulse to x_t

VAR estimator

- Consider the reduced-form VAR(p):

$$w_t = c + \sum_{\ell=1}^{p} A_\ell w_{t-\ell} + u_t$$

  - This can easily be estimated by OLS, see Kilian and Lütkepohl (2017)

- Under invertibility, we arrive at the following structural VAR:

$$A(L) w_t = c + S\varepsilon_t$$

  where $\varepsilon_t \equiv S^{-1} u_t$

- Let $C(L) = A(L)^{-1}$. We can then write

$$w_t = \chi + C(L) u_t = \chi + \sum_{\ell=0}^{\infty} C_\ell S \varepsilon_t, \quad \chi \equiv C(1)c$$

VAR estimator

- The VAR IRF of y_{t+h} to an innovation to x_t is

$$\theta_h \equiv C_{n_r+2,\bullet,h} S_{\bullet,n_r+1}$$

- How to estimate this in practice?
    - Write the VAR in companion form: $\mathbf{w}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{w}_{t-1} + \mathbf{u}_t$, where $\mathbf{w}_t$ is an $np \times 1$ vector including $p-1$ lags of the variables
    - The reduced-form IRF is given by $C_{\bullet,\bullet,h} = [(\mathbf{A}_1)^h]_{1:n,1:n}$
    - Our identifying assumption will pin down S, e.g. in recursive case $S = \operatorname{chol}(\operatorname{Var}(u_t))$

LP estimator

- A linear projection is a regression of an outcome y_t on a "shock" x_t plus controls
- The $\mathrm{LP}(p)$ regression equation for horizon h is

$$y_{t+h} = \mu_h + \beta_h x_t + \gamma_h' r_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' w_{t-\ell} + \xi_{h,t}$$

- The LP IRFs are the $\{\beta_h\}$
- Controlling for r_t is like imposing a zero restriction on these variables
  $\Rightarrow$ we can impose the recursiveness assumption by choosing the controls correspondingly

Equivalence result

Proposition (VAR/LP equivalence)

Assume that we are in population and set $p = \infty$. Let $\tilde{x}_t \equiv x_t - \textit{Proj}(x_t \mid r_t, \{w_\tau\}_{-\infty<\tau<t})$. Then

$$\theta_h = \sqrt{\mathbb{E}\left(\tilde{x}_t^2\right)} \times \beta_h, \quad h = 0, 1, \ldots \tag{1}$$

- (1) states that LPs and VARs estimate the same impulse responses
  - Note that so far we haven't assumed an underlying SVMA model. That will only be needed to argue that the object in (1) is structurally interesting
  - The factor of proportionality in (1) just reflects a different scaling normalization, with VARs normalizing impulses to have unit variance. Will see that in the proof
- Proof intuition: both techniques simply estimate certain linear projections (= functions of 2nd moments) and are willing to call them "structural"

Equivalence result: proof sketch

- By the Frisch-Waugh theorem, the LP estimand is

$$\beta_h = \frac{\text{Cov}\left(y_{t+h}, \tilde{x}_t\right)}{\mathbb{E}\left(\tilde{x}_t^2\right)}$$

- The VAR impulse responses equal [why? use the Wold representation]

$$\theta_h = C_{n_r+2,\bullet,h} S_{\bullet,n_r+1} = \text{Cov}\left(y_{t+h}, \varepsilon_{x,t}\right)$$

  where $\varepsilon_t = \left(\varepsilon_{r,t}', \varepsilon_{x,t}, \varepsilon_{y,t}, \varepsilon_{q,t}'\right)'$. By the properties of Cholesky decompositions we also have

$$\varepsilon_{x,t} = \frac{1}{\sqrt{\mathbb{E}\left(\tilde{u}_{x,t}^2\right)}} \times \tilde{u}_{x,t}$$

  where $u_t = \left(u_{r,t}', u_{x,t}, u_{y,t}, u_{q,t}'\right)'$ and $\tilde{u}_{x,t} \equiv u_{x,t} - \text{Proj}\left(u_{x,t} \mid u_{r,t}\right) = \tilde{x}_t$

- Comparing the above relations the result follows, see Plagborg-Møller and Wolf (2021) for full proof

Equivalence result: discussion

- Plagborg-Møller and Wolf (2021) also discuss the **scope** of this equivalence result
  - Previous argument was for recursive SVAR(∞) estimands. Can show that the same logic works for non-recursive identification: VAR shock $= s'u_t$ so we could run the LP

$$y_{t+h} = \mu_h + \beta_h \left(s'u_t\right) + \sum_{\ell=1}^{\infty} \delta'_{h,\ell} w_{t-\ell} + \xi_{h,t}$$

  - Can also show: with p lags (rather than ∞), equivalence $\approx$ up to horizon p
- The generality of the result reflects its **very simple intuition**:
  - VAR(p) IRF $=$ mean-square optimal forecast given the second moments implied by the VAR(p) model. But a VAR(∞) matches all second-moment properties of the data
  - Thus VAR IRF $=$ optimal forecast given second moments of the data $=$ LP
- Next: walk through LP/VAR implementations of some canonical identification schemes

Outline

1. VARs vs. local projections

   Recursive case

   Identification using IVs

2. Finite-sample recommendations

   A menu of estimation strategies

   Insights from simulation studies

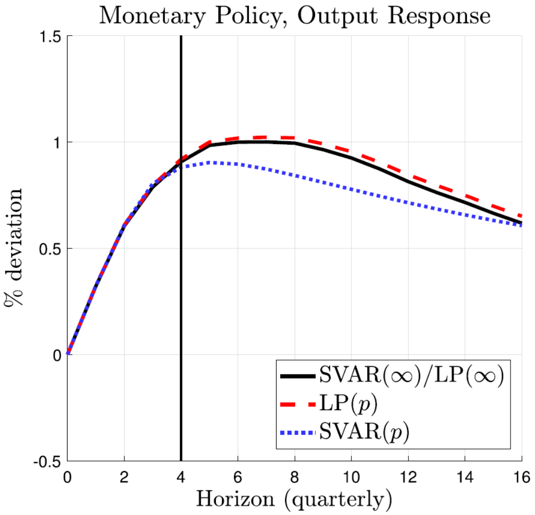
VAR vs. LP: recursive identification

- Recall the identification scheme of Christiano, Eichenbaum, and Evans (1999)
  - Assume SVMA model + invertibility + recursive ordering of macro var's, consistent with slow-moving real effects of monetary policy
- How could we implement this as a **local projection**?
  - Note that the model fits immediately into our structure from before: r_t = GDP, prices, ..., x_t = federal funds rate, q_t = reserves, M2, y_t = any variable of interest in VAR
  - We could thus consider running the LP

$$y_{t+h} = \mu_h + \beta_h x_t + \gamma_h' r_t + \sum_{\ell=1}^{p} \delta_{h,\ell}' w_{t-\ell} + \xi_{h,t}$$

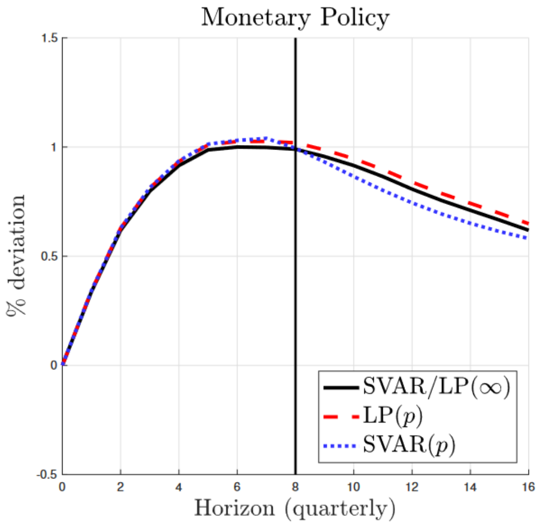
- Let's see the equivalence result in action by computing the common LP/VAR estimand in the Smets and Wouters (2007) model as a simple example DGP ...

VAR vs. LP: recursive identification



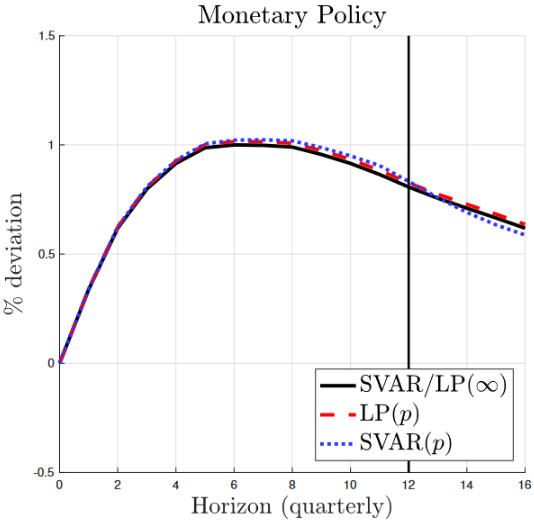
$p = 4$

VAR vs. LP: recursive identification



$p = 8$

VAR vs. LP: recursive identification



$p = 12$

Plagborg-Møller and Wolf (2021) summary

- We can (should?) separate between identification and dimension reduction technique in empirical work
    - VAR/LP can be thought of as particular dimension reduction technique
    - Any identification strategy can be implemented in VARs and LPs
- No estimation method dominates in terms of mean squared error across every possible DGP
- In finite samples, choice between VAR/LP depends on bias-variance trade-off
    - Especially in short samples, VARs may be preferable, more on this later ...

Outline

1. VARs vs. local projections

   Recursive case

   Identification using IVs

2. Finite-sample recommendations

   A menu of estimation strategies

   Insights from simulation studies

Case with (external) instruments

- Now return again to the **SVMA-IV model**

$$w_t = \Theta(L)\varepsilon_t$$
$$z_t = \alpha\varepsilon_{1,t} + \sigma_\nu \nu_t$$

- We have seen that **relative IRFs** are identified. How could we estimate them?
- Here the original approach is **external instruments VAR** estimator (Stock and Watson, 2012; Mertens and Ravn, 2013)
    - Estimate VAR(p): $w_t = \sum_{\ell=1}^{p} A_\ell w_{t-\ell} + u_t$ to get u_t
    - Apply the following 2SLS estimator to obtain s
        1. Regress $u_{1,t}$ on z_t to get $\hat{u}_{1,t}$. This is the first stage
        2. Regress u_t on $\hat{u}_{1,t}$
        3. Impose normalization restriction to get s

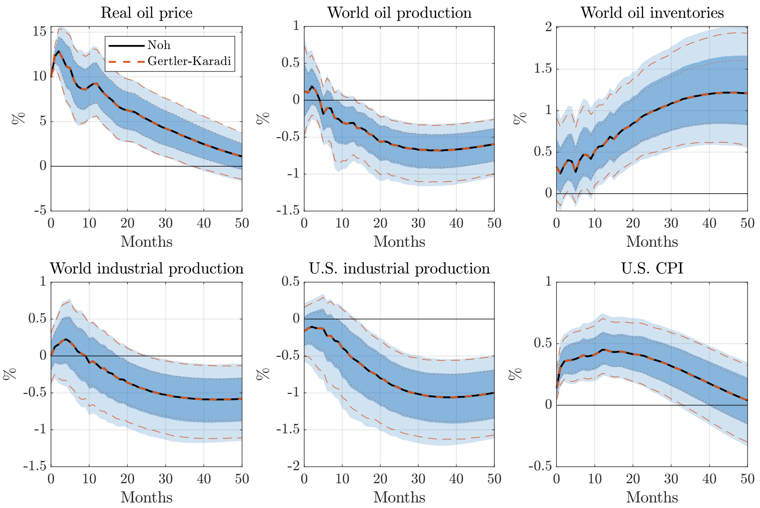
Practical considerations

- In principle, we can instrument **any reduced-from innovation** $\mathrm{u}_{i,t}$ as they all are a linear combination of the structural shocks
- In finite sample, instrument strength may vary depending on i
- Can pick different **normalizations**
    - one standard deviation shock versus shock with unit impact on variable of interest
- Which normalization to pick depends on application
- Often unit normalization more intuitive
    - as we have seen impact responses depend on (slightly) less restrictive assumptions

Practical considerations

- Often instrument is only available for a sub-sample. How to deal with that?
  - One approach is to censor the values in the missing sample to zero
  - Noh (2019) shows that this is fine provided that the timing when the instrument is available is independent to the structural shocks and the measurement errors
  - Gertler and Karadi (2015) instead estimate the VAR coefficients to get u_t on a longer sample than the 2SLS estimator to get s
  - In practice, the two yield very similar results

Practical considerations



First stage regression: F: 20.26, robust F: 10.55, R^2: 4.65%, Adjusted R^2: 4.42%

Testing instrument strength

- As in any IV analysis, crucial to test **instrument strength**
- Different tests available in the literature
  - Reliability measure (Mertens and Ravn, 2013)
  - First-stage F-statistic (Stock and Watson, 2018; Montiel Olea and Pflueger, 2013)
  - Alternative F-test (Lunsford, 2016)

How to deal with weak instruments?

- Sometimes difficult to find strong instruments, see e.g. Kilian (2008)
- Two ways of dealing with weak instruments
    - Adjust frequentist inference (Montiel Olea, Stock, and Watson, 2016)
    - Use Bayesian approach, which provides correct inference even when instrument is weak (Caldara and Herbst, 2019; Arias, Rubio-Ramírez, and Waggoner, 2021)

More shocks, more instruments

- What if we want to identify multiple shocks using multiple instruments?
- We can identify all shocks individually, as in Stock and Watson (2012)
- Alternatively, we can identify them jointly, as in Mertens and Ravn (2013)

More shocks, more instruments

- Assume we have k instruments, z_t
- Partition structural shocks into $\boldsymbol{\varepsilon}_t = [\boldsymbol{\varepsilon}'_{1,t},\ \boldsymbol{\varepsilon}'_{2,t}]'$, where $\boldsymbol{\varepsilon}_{1,t}$ is the $k \times 1$ vector of structural shocks to be identified and $\boldsymbol{\varepsilon}_{2,t}$ is a $(n-k) \times 1$
- Identifying restrictions:

$$\mathbb{E}[z_t \boldsymbol{\varepsilon}'_{1,t}] = \boldsymbol{\alpha}$$
$$\mathbb{E}[z_t \boldsymbol{\varepsilon}'_{2,t}] = 0_{k \times (n-k)},$$

where $\boldsymbol{\alpha}$ is a $k \times k$ matrix (of full rank)

More shocks, more instruments

- Together with the covariance restrictions

$$SS' = \Sigma,$$

  this identifies S up to a rotation:

$$SQ = (S_1,\ S_2)Q = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{pmatrix}$$

- With $k > 1$, to identify S_1, an additional $k(k-1)/2$ restrictions are needed
- Pick Q_k accordingly to implement additional zero/long-run/sign restrictions
- Exercise: Try to show this!

Overidentification

- If we have more instruments than shocks available, we can in principle test for exogeneity of the instruments
- e.g. using Sargan–Hansen test

How to conduct inference?

- Typically done in frequentist setting, using bootstrapping (provided that instrument is strong)
  - Caution: the Wild bootstrap used in Mertens and Ravn (2013); Gertler and Karadi (2015) is not valid: it underestimates uncertainty, especially at short horizons
  - Jentsch and Lunsford (2019) propose a simple moving-block bootstrap
  - Mertens and Ravn (2019) consider a wide range of valid alternatives, such as the Delta method
- Pro: point-identifies the responses of interest
- Con: inference under weak instruments not straightforward, see Montiel Olea, Stock, and Watson (2016)

Bayesian inference

- Easy way: only consider sampling uncertainty treating instrument as given
- Draw from reduced-form VAR posterior and compute external instruments estimator for each draw to map out posterior distribution
- More involved: model instrument and model variables jointly (Caldara and Herbst, 2019; Arias, Rubio-Ramírez, and Waggoner, 2021)
- Pro: valid inference also under weak instruments
- Con: responses only set-identified

LP-IV

- Return again to the **SVMA-IV model**

$$w_t = \Theta(L)\varepsilon_t$$

$$z_t = \alpha\varepsilon_{1,t} + \sigma_\nu \nu_t$$

- Alternatively, we can also directly estimate the causal effects of interests. With instruments, this is referred to as **LP-IV**:
    - Basic idea: use z_t to instrument for x_t in a projection of y_{t+h} on x_t

$$y_{t+h} = \mu_h + \beta_h x_t + \text{ controls } + \xi_{h,t}$$

    - Will decompose this into reduced-form and first-stage to arrive at equivalent VAR representation, using our previous results

LP-IV

- Let $W_t = (z_t, w_t')'$. The reduced-form and first-stage projections are

$$y_{t+h} = \mu_{RF,h} + \beta_{RF,h} z_t + \sum_{\ell=1}^{\infty} \delta'_{RF,h,\ell} W_{t-\ell} + \xi_{RF,h,t}$$

$$x_t = \mu_{FS} + \beta_{FS} z_t + \sum_{\ell=1}^{\infty} \delta'_{FS,\ell} W_{t-\ell} + \xi_{FS,t}$$

- We thus get relative LP-IV IRFs as $\beta_h \equiv \beta_{RF,h}/\beta_{FS}$

Inference in LP

- Inference is straightforward in LPs
- Because residual can be autocorrelated, use HAC standard errors
- Alternative: lag-augmentation approach (Montiel Olea and Plagborg-Møller, 2021)
  - show that when augmenting the lag order by one, White standard errors are valid
  - true for both stationary and non-stationary data and a wide range of response horizons

Internal instruments VAR

- Recall from last time that external instruments VAR and LP-IV do not rely on the same identifying assumptions
  - Partial invertibility versus limited lag exogeneity
- But from Plagborg-Møller and Wolf (2021), we know that there must exist a VAR implementation of LP-IV
- This is the **internal instruments** approach
  - Can show that $\beta_{RF,h}$ and β_{FS} can be recovered from recursive VAR in $W_t = (z_t, w_t')'$
  - Note that this is a VAR that works without invertibility. Why? non-invertibility is related to measurement error $\sigma_\nu \nu_t$, which merely induces a constant attenuation bias
  - Cancels out when computing the **relative** impulse responses

Outline

1. VARs vs. local projections

   Recursive case

   Identification using IVs

2. Finite-sample recommendations

   A menu of estimation strategies

   Insights from simulation studies

LP vs. VAR in finite samples

- We have seen that LPs and VARs estimate the **same IRFs in population**
- In finite samples there's a standard *bias-variance* trade-off
  Note that this trade-off is analogous to direct vs. iterated forecasting
    - VAR: extrapolate longer-run impulse responses from first few autocorrelations. Low variance, possibly high bias
    - LP: no extrapolation. High variance, low bias
- Natural question: which to pick in finite samples?
    - Just looking at VARs vs. LPs is a **false dichotomy**. We should look for methods that estimate autocovariance functions = second-moment properties "as well as possible"
    - Li, Plagborg-Møller, and Wolf (2021) provide comprehensive **simulation study** results for a variety of estimation techniques Note: this will be informative for typical time-series context. Trade-off with panel data may look quite different

Outline

1. VARs vs. local projections

   Recursive case

   Identification using IVs

2. Finite-sample recommendations

   A menu of estimation strategies

   Insights from simulation studies

A menu of estimation strategies

- I want to emphasize that plain-vanilla **VARs & LPs** are not the only game in town
  - LP & adjacent methods: OLS LP, penalized LP, Bayesian LP

    References: Jordà (2005), Barnichon and Brownlees (2019), Miranda-Agrippino and Ricco (2021)
  - VAR & adjacent methods: OLS VAR, bias-corrected VAR, Bayesian VAR, VAR averaging

    Hansen (2016), Kilian (1998), Kilian and Lütkepohl (2017)
- Active research area in applied macroeconometrics. We'll summarize main simulation results, see references for more info

Outline

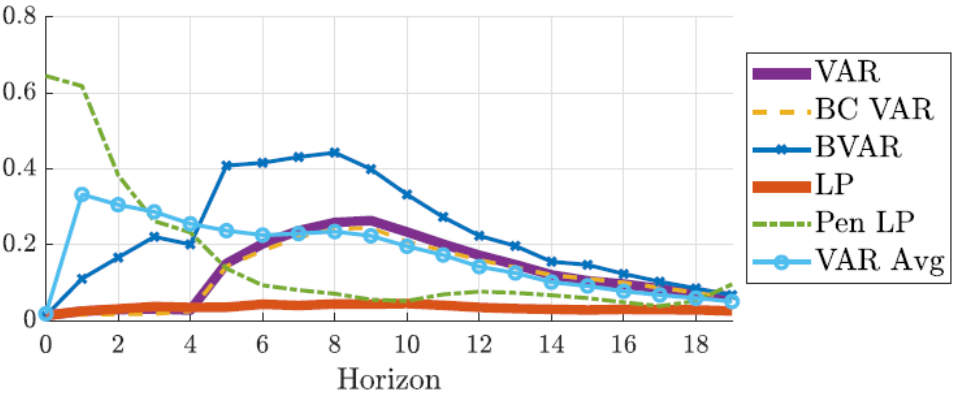
Li, Plagborg-Møller, and Wolf (2021)

- Li, Plagborg-Møller, and Wolf (2021) perform a simulation study of LP and VAR estimators of structural impulse responses
- Design of the **simulation study**
  - Take large-scale factor model as encompassing DGP, pick random subsets of observables w_t, try to estimate standard recursive/IV estimands

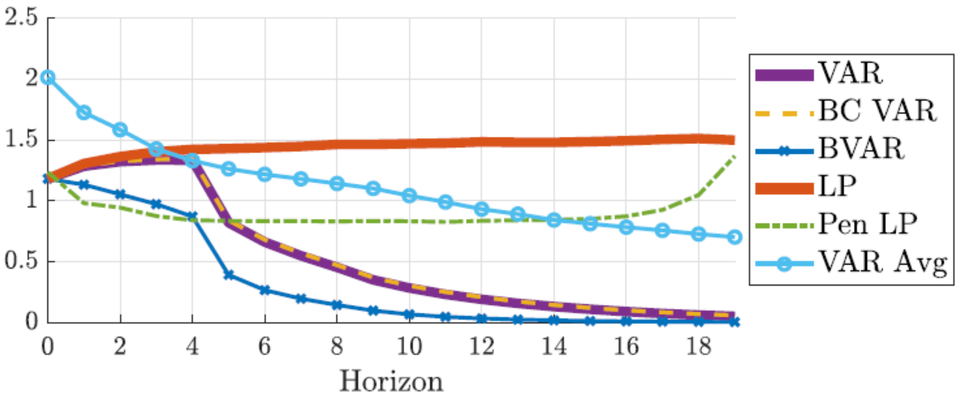
    Why? such models known to fit the properties of the "universe" of U.S. time series quite well
  - Estimation methods: VAR, BC-VAR, BVAR, VAR averaging, LP, pen. LP and various identification schemes
- **Results reporting**
  - Show bias and standard deviation by IRF horizon

    Note that this averages over the random subsets of observables. Results do not differ much by structural identification scheme
  - Question: given an IRF horizon and a relative weight on bias vs. variance, which method performs the best on average?

Results: bias

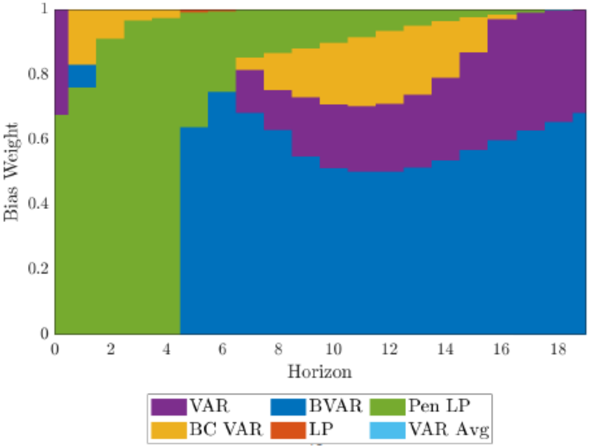


Results: standard deviation



Results: preferred method

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



Results: summary

- The results point to a clear **bias-variance trade-off**
    - LP estimators have lower bias than VAR estimators but substantially higher variance at intermediate and long horizon
- Thus, unless researchers are overwhelmingly concerned with bias, shrinkage via Bayesian VARs or penalized LPs is attractive

Summary

- Identification and estimation can be thought of separately in empirical work
    - VAR/LP can be thought of as particular dimension reduction technique
    - Any identification strategy can be implemented in VARs and LPs
    - The term VAR identification is misleading
- No estimation method dominates in terms of mean squared error across every possible DGP, in finite samples choice between VAR/LP depends on bias-variance trade-off
- Next time: how to come up with **instruments** in macro & applications