Lecture 5: Time Series Estimation Techniques

Christian Wolf

MIT

14.461, Fall 2022

Overview

- So far: population arguments for **identification** of SVMA coefficients
  - Basic idea: use addt'l identifying information (e.g., invertibility + X, IVs, ...) to assign a causal interpretation to certain estimable moments of aggregate time series data
  - Open **Q**: what econometric techniques should we use to estimate those moments?
- Today's lecture: **VARs** vs. **local projections**
  - Will show that they are nothing more than two different techniques for estimating second moments. In population, under the same structural ID asn's, they yield the same objects
  - Finite-sample recommendations: should we use LPs? VARs? something else?

Outline

1. VARs vs. Local Projections
   - Example I: recursive identification
   - Example II: IV identification

2. Finite-Sample Recommendations
   - A Menu of Estimation Strategies
   - Insights from Simulation Studies

Outline

VARs vs. Local Projections

- Dominant methods in semi-structural time series: **VARs** and **Local Projections**
  Going back to Sims (1980) and Jorda (2005)
- Can find a lot of claims in the literature that these are somehow **fundamentally different methods**. We will show that 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**
  - Data $\{w_t\}$ are covariance stationary and nondeterministic, with absolutely summable Wold representation coefficients
  - Split 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
- **LP IRF estimator**
  - A linear projection is a regression of an outcome y_t on a "shock" x_t plus controls
  - The LP(∞) regression equation for horizon $h = 0, 1, 2, \ldots$ is

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

  - The LP IRFs are the $\{\beta_h\}$

Setting and estimators

- **VAR IRF estimator**
  - Consider the reduced-form VAR(∞)

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

  - Under a recursive ordering, we arrive at the following "structural" VAR:

$$A(L) w_t = c + B\eta_t$$

    where $B = \text{chol}(\text{Var}(u_t))$ and $\eta_t \equiv B^{-1}u_t$, with $\eta_{1,t} \propto u_{1,t}$, $\eta_{2,t} \propto u_{2,t} - \mathbb{E}^*(u_{2,t} \mid u_{1,t})$, ...

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

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

    Then the VAR IRF of y_{t+h} to an innovation to x_t is

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

Equivalence result

Proposition

Let $\tilde{x}_t \equiv x_t - \mathbb{E}^* \left[x_t \mid r_t, \{w_\tau\}_{-\infty<\tau<t} \right]$. *Then*

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

- (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

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

- The VAR impulse responses equal

$$\theta_h = C_{n_r+2,\bullet,h} B_{\bullet,n_r+1} = \text{Cov}(y_{t+h}, \eta_{x,t})$$

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

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

where $u_t = (u_{r,t}', u_{x,t}, u_{y,t}, u_{q,t}')'$ and $\tilde{u}_{x,t} \equiv u_{x,t} - \mathbb{E}^*(u_{x,t} \mid u_{r,t}) = \tilde{x}_t$

- Comparing the above relations, the result follows

Equivalence result: discussion

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

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

  - 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 id schemes**

Outline

VAR vs. LP: recursive identification

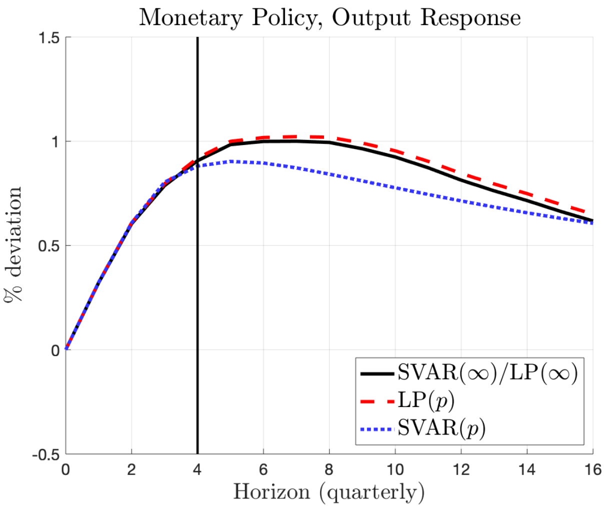
- Recall the identification scheme of Christiano-Eichenbaum-Evans (1999)
    - Assume SVMA model + invertibility + recursive ordering of macro var's, consistent with slow-moving real effects of monetary policy
- How could we implement this as a **local projection**?
    - Note that the model fits immediately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

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

Q: what happens if $h = 0$ and y_t is in r_t, e.g. GDP?

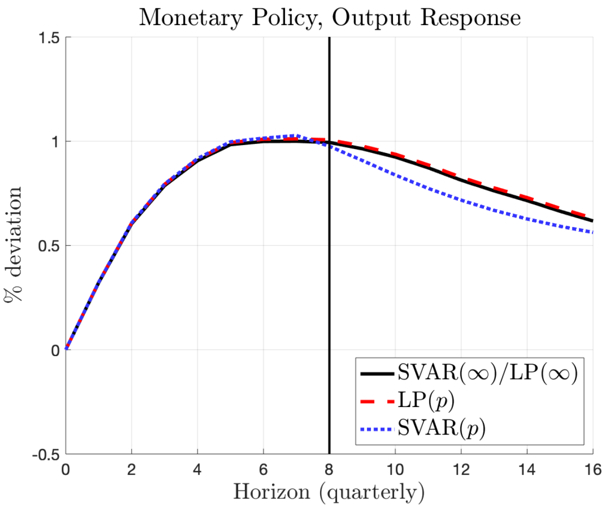
- Let's see the equivalence result in action by computing the common LP/VAR estimand in the Smets & Wouters (2007) model as a simple example DGP ...

VAR vs. LP: recursive identification



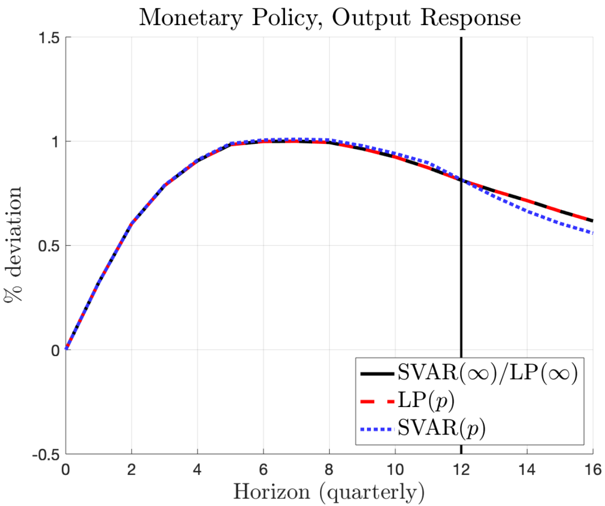
$p = 4$

VAR vs. LP: recursive identification



$p = 8$

VAR vs. LP: recursive identification



$p = 12$

Outline

1. VARs vs. Local Projections
    - Example I: recursive identification
    - Example II: IV identification

2. Finite-Sample Recommendations
    - A Menu of Estimation Strategies
    - Insights from Simulation Studies

VAR vs. LP: IV identification

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

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

- We have seen that **relative IRFs** are identified. How could we estimate them?
- Here the typical approach is an LP implementation, known 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

VAR vs. LP: IV identification

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

$$
\begin{aligned}
y_{t+h} &= \mu_{RF,h} + \beta_{RF,h} z_t + \sum_{\ell=1}^{\infty} \delta_{RF,h,\ell}' W_{t-\ell} + \xi_{RF,h,t} \\
x_t &= \mu_{FS} + \beta_{FS} z_t + \sum_{\ell=1}^{\infty} \delta_{FS,\ell}' W_{t-\ell} + \xi_{RF,t}
\end{aligned}
$$

  - We thus get relative LP-IV IRFs as $\beta_h \equiv \beta_{RF,h} / \beta_{FS}$
- **Recursive VAR**
  - From previous results: can recover $\beta_{RF,h}$ and β_{FS} from a recursive VAR in $W_t = (z_t, w_t')'$
  - Aside: 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

Outline

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
    - VAR: extrapolate longer-run impulse responses from first few autocorrelations. Low variance, possibly high bias.
    - LP: no extrapolation. High variance, low bias.
- Natural **Q**: which to pick in finite samples? Only very brief discussion here, since this not a class on finite-sample estimation. Main points:
    1. 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"
    2. Presents results from a comprehensive **simulation study** 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

A menu of estimation strategies

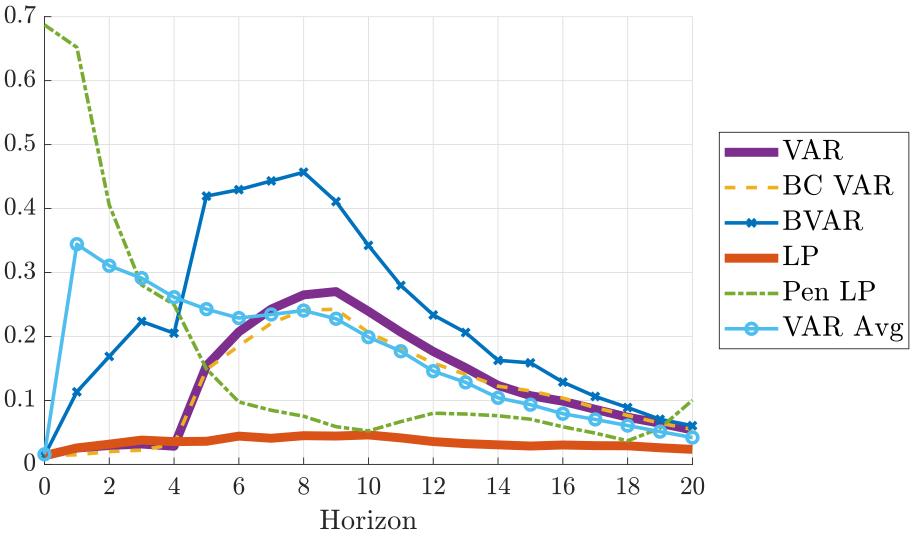
- I just want to emphasize that **vanilla VARs & LPs** are not the only game in town:
  - LP & adjacent methods: OLS LP, penalized LP, Bayesian LP
    References: Jorda (2005), Barnichon & Brownlees (2019), Miranda-Agrippino & Ricco (2021)
  - VAR & adjacent methods: OLS VAR, bias-corrected VAR, Bayesian VAR, VAR averaging
    Hansen (2016), Kilian (1998), Kilian & Lütkepohl (2017)
- Active research area in applied macroeconometrics. I will only provide references and summarize **main simulation study results**.

Outline

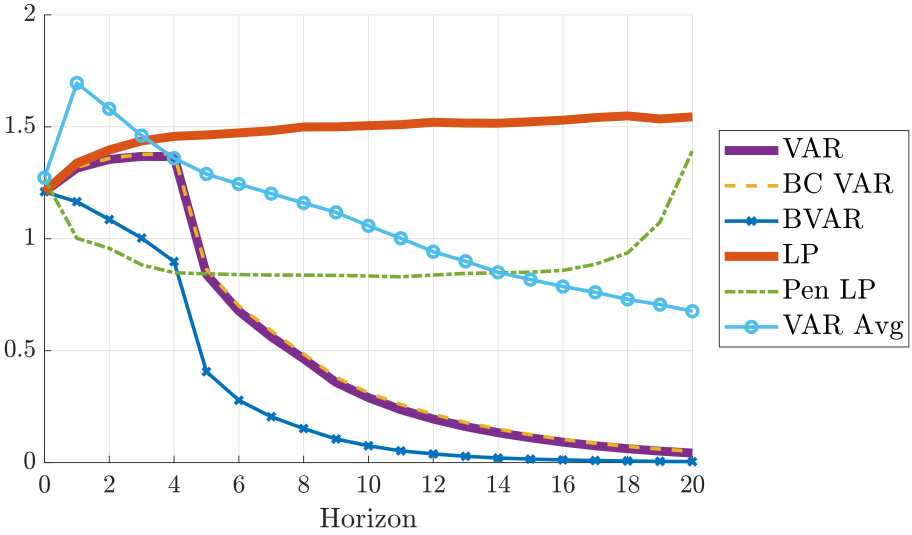
Overview

- Will briefly the main results from Li, Plagborg-Møller and Wolf (2022)
- 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 are 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
- **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.
  - **Q**: 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": $\mathcal{L}_\omega(\theta_h, \hat{\theta}_h) = \omega \times \left(\mathbb{E}\left[\hat{\theta}_h - \theta_h\right]\right)^2 + (1-\omega) \times \mathrm{Var}(\hat{\theta}_h)$