Local Projections: Inference

Òscar Jordà and Karel Mertens

AEA Continuing Education 2023

Last updated: December 27, 2022

The views expressed herein do not necessarily represent those of any of the institutions in the Federal Reserve System

See also:
https://sites.google.com/site/oscarjorda/home/local-projections

Review of identification with Local Projections

Most of this already discussed in the previous lecture

The issue

(Some) threats to identification

Recall, we need $s_t|\boldsymbol{x}_t$ *randomly assigned*

Some examples when identification fails:

- excluded observables: correlated with s_t and y_t
- unobservables: correlated with s_t and y_t
- simultaneity: s_t and y_t jointly determined

(Some) **solutions** (well known from VARs):

- parametric zero restrictions
- *internal* instruments
- *external* instruments
- identification through heteroscedasticity
- ... and others

Recall: zero short-run restrictions

Cholesky decomposition – Wold causal ordering

$\Sigma = PP'$ with P lower triangular:
always exists and is unique, but ...

- different ordering of the variables, different P
- implied 0 restrictions may be incorrect
- just-identification $\implies$ ordering cannot be tested
- however, trivial to implement

Interpretation:

- $y_{(1),t}$ does not contemporaneously depend on others
- $y_{(2),t}$ only depends on $y_{(1),t}$ contemporaneously
- $y_{(3),t}$ only depends on $y_{(1),t}, y_{(2),t}$ contemporaneously
- and so on...

Recursive identification in LPs

Suppose $n \times 1$ vector y_t

Decide the *causal* ordering.
Include the contemporaneous values of variables causally ordered first:

$$y_{j,t+h} = \mu_j^h + \beta_{j,1}^h y_{1,t} + \ldots + \beta_{j,i-1}^h y_{i-1,t} + \beta_{j,i}^h y_{i,t} + \sum_{k=1}^{p} \mathbf{c}_{j,k}^h \mathbf{y}_{t-k} + v_{j,t+h}$$

Structural LP Estimate

$$\hat{\mathcal{R}}_{ij}(h) = \hat{\beta}_{j,i}^h; \quad h = 0, 1 \ldots, H; \quad i, j \in \{1, \ldots, n\}$$

Remark: good idea to order treatment variable ($y_{i,t}$) last $\rightarrow$ variation cannot be explained by observables

Long-run zero restrictions with LPs

Two step procedure

Blanchard and Quah (1989) example:
$\boldsymbol{y}_t = (x_t, u_t)$, x_t log real GDP; u_t unemployment rate

Step 1: long-run LP

$$x_{t+H} - x_{t-1} = \alpha_H + \boldsymbol{\delta}_{x,H} \boldsymbol{y}_t + \sum_{k=1}^{p} \boldsymbol{c}_{x,k}^{H} \boldsymbol{y}_{t-k} + v_{x,t+H}$$

$\boldsymbol{\delta}_{x,H}$: linear combination that best explains long-run GDP (i.e. supply shock)

Remark: choose *H large*

Long-run identification

Step 2

$$y_{j,t+h} = \mu_h + \beta_{j,h}(\hat{\boldsymbol{\delta}}_{x,H}\boldsymbol{y}_t) + \sum_{k=1}^{p} \boldsymbol{c}_{j,k}^{h}\boldsymbol{y}_{t-k} + v_{j,t+h}; \quad j = x, u; \quad h = 0, 1, \ldots, H$$

Remarks:

- $\beta_{j,h}$ is the response of the j^{th} variable to supply shock, in period h
- $\hat{\boldsymbol{\delta}}_{x,H}\boldsymbol{y}_t$ comes from first step
- little guidance on how to choose H. Try different values
- Idea can be generalized in a number of ways:
  medium-run identification?

Sign restrictions

Example: monetary shock $\rightarrow$ positive response of r_{t+h} for $h = 0, 1, \ldots, H$ with $\mathcal{R}_r(0) = 1$ normalization

Idea: find all linear combinations $\boldsymbol{\delta}$ such that $\mathcal{R}_r(h) > 0$ and $\mathcal{R}_r(0) = 1$

Step 1: $r_{t+h} = \mu_{r,h} + \boldsymbol{g}_{r,h}\boldsymbol{y}_t + \sum_{k=1}^{p} \boldsymbol{c}_{r,k}^{h}\boldsymbol{y}_{t-k} + \boldsymbol{v}_{r,t+h} \rightarrow \hat{\boldsymbol{g}}_{r,h}$

Step 2: $y_{j,t+h} = \mu_{j,h} + \boldsymbol{\gamma}_{j,h}\boldsymbol{y}_t + \sum_{k=1}^{p} \boldsymbol{c}_{j,k}^{h}\boldsymbol{y}_{t-k} + \boldsymbol{v}_{j,t+h} \rightarrow \hat{\boldsymbol{\gamma}}_{j,h}$

Step 3: find $\boldsymbol{\delta}$ such that

$$\sup_{\delta} \boldsymbol{\delta}'\hat{\boldsymbol{\gamma}}_{j,h} \quad s.t. \quad \boldsymbol{\delta}'\hat{\boldsymbol{g}}_{r,0} = 1$$

$$\boldsymbol{\delta}'\hat{\boldsymbol{g}}_{r,h} \geq 0 \quad for \quad h = 1, \ldots, H$$

same for $\inf$ to obtain upper and lower bounds for $\mathcal{R}_{ry}(h)$

Remarks

- note this is *set* identification not *point* identification
- hence inference is much more complicated
- Plagborg-Møller and Wolf (2021, ECTA) provide solution algorithm
- choice of H matters, could be relatively short
- simulation methods (bayesian) another way to go?
- may combine with other constraints

LP-IV

Stock and Watson (2018, Economic Journal) Assumptions

Suppose $\boldsymbol{z}_t$ is a vector of instruments for the structural shock $\epsilon_{1,t}$ and denote $\boldsymbol{z}_t^P = \boldsymbol{z}_t - \mathcal{P}(\boldsymbol{z}_t | \boldsymbol{w}_t)$ where $\boldsymbol{w}_t$ collects all controls in the LP (e.g. $\boldsymbol{y}_{t-j}$)

1. Relevance: $E\left(\epsilon_{1,t}^P \boldsymbol{z}_t^{P\prime}\right) = \boldsymbol{\alpha}' \neq 0$
2. Basic exogeneity: $E\left(\epsilon_{j,t}^P \boldsymbol{z}_t^{P\prime}\right) = 0, \quad j \neq 1$
3. Lead-Lag exogeneity: $E\left(\epsilon_{j,t+h}^P \boldsymbol{z}_t^{P\prime}\right) = 0, \forall j, h \neq 0$

Remarks:

- usual IV conditions except lead-lag exogeneity because dynamics

LP-IV: Assumptions 1

Plagborg-Møller and Wolf (2021 ECTA)

Assumption 1: $\underset{n_y \times 1}{\boldsymbol{y}_t} = \boldsymbol{\mu} + \Theta(L) \underset{n_\epsilon \times 1}{\boldsymbol{\epsilon}_t}$; where:

$$\underset{n_y \times n_\epsilon}{\Theta(L)} \equiv \sum_{h=o}^{\infty} \Theta_h L^h \text{ s.t. } \sum_{h=0}^{\infty} ||\Theta_h|| < \infty \text{ with } ||\Theta_h||^2 = tr(\Theta_h' \Theta_h)$$

and $\Theta(x)$ has full column rank for all complex scalars x on the unit circle.

Remarks:

- $\boldsymbol{\epsilon}_t$ are structural, hence possibly $\Theta_0 \neq I$
- we can have $n_\epsilon > n_y$ (non-invertibility)
- $\boldsymbol{y}_t$ is strictly stationary
- Θ_h is the structural impulse response coefficient matrix

LP-IV: Assumptions 2

Assumption 2: $z_t = c_z + \sum_{h=1}^{\infty}(G_h z_{t-h} + \Lambda_h \boldsymbol{y}_{t-h}) + \alpha \epsilon_{1,t} + \nu_t$ with:

- $\alpha \neq 0$ relevance condition
- $1 - \sum_{h=1}^{\infty} G_h L^h$ has all roots outside unit circle
- $\sum_{h=1}^{\infty} ||\Lambda_h|| < \infty$
- $\nu_t \perp \boldsymbol{\epsilon}_{t-j}$ for any j, ν_t is measurement error

Remarks:

- Assumptions 1 and 2 $\rightarrow$ validity of LP-IV and SVAR-IV
- but LP-IV does not require invertibility

See Plagborg-Møller and Wolf (2021) for more details
Example code: LPIV_example.do

Recall: Impulse responses as a comparison of two averages

$$\mathcal{R}(h) = E(E[y_{t+h}|s_t = s + \delta, \mathbf{x}_t] - E[y_{t+h}|s_t = s, \mathbf{x}_t]))$$

- y_{t+h}: outcome
- s_t: intervention
- s: baseline, e.g., $s = 0$
- δ: dose, e.g., $\delta = 1$; $\delta = var(\epsilon)^{1/2}; \ldots$
- $\mathbf{x}_t$: exogenous and predetermined variables

A trivial example

Suppose $s_t \in \{0, 1\}$ is *randomly assigned*, then:

$$\mathcal{R}(h) = \frac{1}{N_1} \sum_{t=1}^{T-h} y_{t+h} s_t - \frac{1}{N_0} \sum_{t=1}^{T-h} y_{t+h}(1 - s_t)$$

$$N_1 = \sum_{t=1}^{T-h} s_t; \quad T - h = N_1 + N_0$$

Remarks:

- inefficient (not using $\boldsymbol{x}_t$), but *consistent*
- could control for $\boldsymbol{x}_t$ with Inverse Propensity score Weighting (IPW)
- feels like the potential outcomes paradigm used in micro
- could have regressed y_{t+h} on s_t, same thing (could add $\boldsymbol{x}_t$ easily)

Inverse propensity score weighting

The basics: an alternative/complement to regression control

let $s_t \in \{0,1\}$ be policy treatment;

$\boldsymbol{y}_{t,H} = (y_t, y_{t+1}, \ldots, y_{t+H})$

Selection on observables or conditional ignorability:

$$\boldsymbol{y}(s) \perp s | \boldsymbol{x} \quad s \in \{0,1\}$$

suppose s randomly assigned, then no need for $\boldsymbol{x}$:

$$\hat{\mathcal{R}}(h) = \underbrace{\frac{1}{T_1}\sum_{t=1}^{T} s_t \, y_{t+h}}_{\mu_1^h} - \underbrace{\frac{1}{T_0}\sum_{t=1}^{T}(1-s_t)\, y_{t+h}}_{\mu_0^h}$$

$$\boldsymbol{y}_{t+h} = \mu_0^h + s_t\gamma_h + v_{t+h} \rightarrow \mathcal{R} = \boldsymbol{\gamma}$$

Rosenbaum and Rubin 1983

the propensity score as a sufficient statistic

before: $\boldsymbol{y}(s) \perp s|\boldsymbol{x}$; now: $\boldsymbol{y}(s) \perp s|p(s=1|\boldsymbol{x}) \quad s \in \{0,1\}$

hence, if $\hat{p}_t = p(s_t = 1|\boldsymbol{x}_t; \hat{\boldsymbol{\theta}})$ then:

$$\hat{\mathcal{R}}(h) = \frac{1}{T_1^*}\sum_{t=1}^{T}\left(\frac{s_t\, y_{t+h}}{\hat{p}_t}\right) - \frac{1}{T_0^*}\sum_{t=1}^{T}\left(\frac{(1-s_t)\, y_{t+h}}{(1-\hat{p}_t)}\right)$$

with

$$T_1^* = \sum_{t=1}^{T}\frac{s_t}{\hat{p}_t}; \quad T_0^* = \sum_{t=1}^{T}\frac{1-s_t}{1-\hat{p}_t}$$

Doubly robust IPW estimators

regression augmented IPW:

$$y_{t+h} = \frac{s_t}{\hat{p}_t}\left(\mu_0^h + (\boldsymbol{x}_t - \boldsymbol{\mu}_x)\boldsymbol{\gamma}_0^h\right) + \frac{1-s_t}{1-\hat{p}_t}\left(\mu_1^h + (\boldsymbol{x}_t - \boldsymbol{\mu}_x)\boldsymbol{\gamma}_1^h\right) + v_{t+h}$$

see also *augmented IPW* by Lunceford and Davidian (2004)

Remarks:

- $\hat{p}_t$ usually a first-stage logit/probit $\rightarrow$ affects inference
- IPW literature provides SE formulas, but not for time series settings
- one solution is to use the bootstrap

IPW code available here

Inference

Why is inference different with local projections?

It is the MA structure of the residuals

recall the AR(1) example, $y_t = \rho y_{t-1} + u_t$. By recursive substitution:

$$y_{t+h} = \rho^{h+1} y_{t-1} + u_{t+h} + \rho u_{t+h-1} + \ldots + \rho^h u_t$$

so in a local projection:

$$y_{t+h} = \beta_{h+1} y_{t-1} + v_{t+h}; \quad v_{t+h} = u_{t+h} + \rho u_{t+h-1} + \ldots + \rho^h u_t$$

In general, we don't know the MA structure
Jordà (2005) recommended HAC standard errors, e.g. Newey-West

Lag augmentation a simpler, more elegant solution

Montiel-Olea and Plagborg-Møller. 2021. Econometrica

The logic of lag augmentation

A simple example

DGP:: $y_t = \rho y_{t-1} + u_t$; u_t strictly stationary, $E(u_t | \{u_s\}_{s \neq t}) = 0$

LP: $y_{t+h} = \beta_h y_t + v_{t+h}; \; v_{t+h} \sim MA(h)$

Plug DGP into LP: $y_{t+h} = \beta_h u_t + \gamma_h y_{t-1} + v_{t+h}$

FWL logic: obtain β_h by regressing $y_{t+h} - \gamma_h y_{t-1}$ on $y_t - \rho y_{t-1}$

$$\hat{\beta}_h = \frac{\sum_{t=1}^{T-h} (y_{t+h} - \gamma_h y_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=1}^{T-h} (y_t - \rho y_{t-1})^2} = \frac{\sum_{t=1}^{T-h} (\beta_h u_t + v_{t+h}) u_t}{\sum_{t=1}^{T-h} u_t^2}$$

$$= \beta_h + \frac{\sum_{t=1}^{T-h} v_{t+h} u_t}{\sum_{t=1}^{T-h} u_t^2}$$

Key insight

Same logic if DGP is VAR(p)

Recall:

$$\hat{\beta}_h = \beta_h + \frac{\sum_{t=1}^{T-h} v_{t+h} u_t}{\sum_{t=1}^{T-h} u_t^2} \quad \rightarrow \quad \hat{\sigma}^2(\hat{\beta}_h) = \frac{\sum_{t=1}^{T-h} \hat{v}_{t+h}^2 \hat{u}_t^2}{\left(\sum_{t=1}^{T-h} \hat{u}_t^2\right)^2}$$

although $v_{t+h} \sim MA(h)$, note that $v_{t+h}u_t \sim MA(0)$ since for any $s < t$:

$$\begin{aligned} E[v_{t+h}u_t v_{s+h}u_s] &= E[E[v_{t+h}u_t v_{s+h}u_s | u_{s+1}, u_{s+2}, \ldots]] \\ &= E[v_{t+h}u_t v_{s+h} \underbrace{E[u_s | u_{s+1}, u_{s+2}, \ldots]}_{=0}] \end{aligned}$$

Takeaway: do lag-augmented LP with White corrected errors.
No need for Newey-West

Wild boostrap with lag augmentation

Response of j^{th} variable to a shock

1. Lag-augmented LP $\rightarrow$ collect $\hat{\beta}_{j,h}, \hat{\sigma}_{j,h} = \hat{\sigma}(\hat{\beta}_{j,h})$
2. VAR(p) $\rightarrow \hat{\boldsymbol{u}}_t$ (option: bias-adjust VAR coeffs Pope, 1990 procedure)
3. VAR(p) $\rightarrow \hat{\beta}^{VAR}_{j,h}$
4. For each boostrap iteration $b = 1, \ldots, B$:
    1. Generate bootstrap residuals $\hat{\boldsymbol{u}}^*_t \equiv Z_t \hat{\boldsymbol{u}}_t; \quad Z_t \sim N(0,1)$ (wild bootstrap)
    2. draw a block of p initial observations $(\boldsymbol{y}^*_1, \ldots, \boldsymbol{y}^*_p)$ at random from $T - p + 1$ blocks of p observations from the data
    3. Generate $\boldsymbol{y}^*_t$ with $(\boldsymbol{y}^*_1, \ldots, \boldsymbol{y}^*_p)$ initial observations, the bias-corrected $VAR(p)$ coeffs, and $\hat{\boldsymbol{u}}^*_t$
    4. Apply augmented LP to $\{\boldsymbol{y}^*_t\} \rightarrow \hat{\beta}^*_{j,h},\ \hat{\sigma}^*_{j,h}$
    5. Store $\hat{T}^*_b = (\hat{\beta}^*_{j,h} - \hat{\beta}^{VAR}_{j,h})/\hat{\sigma}^*_{j,h}$
5. Compute $\alpha/2$ and $1 - \alpha/2$ quantiles of $\{\hat{T}^*_b\}^B_{b=1}$, say $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ respectively
6. the percentile confidence interval is:

$$[\hat{\beta}_{j,h} - \hat{\sigma}_{j,h}\hat{q}_{1-\alpha/2},\ \hat{\beta}_{j,h} - \hat{\sigma}_{j,h}\hat{q}_{\alpha/2}]$$

See https://github.com/jm4474/Lag-augmented_LocalProjections

Parametrically adjusted standard errors

General LP:

$$y_{t+h} = \beta_h s_t + \boldsymbol{\gamma}_h \boldsymbol{x}_t + v_{t+h}; \quad v_{t+h} = u_{t+h} + \phi_1 u_{t+h-1} + \ldots + \phi_h u_t$$

Note: make no assumptions on how y, s, and $\boldsymbol{x}$ are dynamically related

hence no assumption on $\phi_1, \ldots, \phi_h$

Can view the LP as the DGP and estimate the ϕ_j directly as XMA(h) model

Lusompa (2019) FGLS

Lusompa's (2019) FGLS procedure

See his paper for a bootstrap and Bayesian approaches

Step 1 (usual LP for $h = 0$):

$$y_t = \alpha_0 + \boldsymbol{x}_t \boldsymbol{\beta}_0 + s_t \gamma_0 + u_t \quad \rightarrow \quad \{\hat{u}_t\}; \hat{\gamma}_0$$

Step 2 (use step 1 to fix LHS variable):

$$\tilde{y}_{t+1} = \alpha_1 + \boldsymbol{x}_t \boldsymbol{\beta}_1 + s_t \gamma_1 + v_{t+1}; \quad \tilde{y}_{t+1} = y_{t+1} - \hat{u}_t \hat{\gamma}_0 \quad \rightarrow \hat{\gamma}_1$$

Step 3 (use estimates from Step 1 and 2):

$$\begin{aligned} \tilde{y}_{t+2} &= \alpha_2 + \boldsymbol{x}_t \boldsymbol{\beta}_2 + s_t \gamma_2 + v_{t+2} \\ \tilde{y}_{t+2} &= y_{t+2} - (\hat{u}_t \hat{\gamma}_1 + \hat{u}_{t+1} \hat{\gamma}_0) \quad \rightarrow \hat{\gamma}_2 \end{aligned}$$

rinse and repeat for steps 4 ... H

Note: always use Step 1 residuals $\hat{u}_t$ in all steps

Further comments and remarks

Many interesting results from Lusompa (2019)

- VAR need not be DGP for FGLS to work
- in small samples with high persistence,
  NW has small sample bias
- similar result in Herbst and Johannsen (2020)
- shows two bootstrap algorithms
- shows bayesian approach with time-varying example
- focus is on pointwise uncertainty, however

Joint inference

LPs as a GMM problem

A simplification first

The Frisch-Waugh-Lovell theorem

Elements of the problem:

- y_t: outcome variable (response)
- $\boldsymbol{x}_t$: control variables (constant, predetermined endogenous and exogenous variables)
- s_t: treatment variable (impulse)
- $\boldsymbol{z}_t$: instrumental variables (possibly none in which case, $s_t = \boldsymbol{z}_t$)

Let $\mathcal{P}_L(w_t|\boldsymbol{v}_t)$ denote the linear regression of w_t on $\boldsymbol{v}_t$

From now on, assume:

- $y^e_{t+h} \stackrel{def}{=} y_{t+h} - \mathcal{P}_L(y_{t+h}|\boldsymbol{x}_t)$
- $s^e_t \stackrel{def}{=} s_t - \mathcal{P}_L(s_t|\boldsymbol{x}_t)$
- $\boldsymbol{z}^e_t \stackrel{def}{=} \boldsymbol{z}_t - \mathcal{P}_L(\boldsymbol{z}_t|\boldsymbol{x}_t)$

Basic univariate LP results

$$y^e_{t+h} = s^e_t \gamma_h + v_{t+h}; \quad h = 0, 1, \ldots, H$$

$$\sqrt{T}(\hat{\gamma}_h - \gamma_h) = \frac{\frac{1}{T^{1/2}} \sum_{t=1}^{T-h} v^e_{t+h} s^e_t}{\frac{1}{T} \sum_{t=1}^{T} s^{e2}_t}; \quad \frac{1}{T} \sum_{t=1}^{T} s^{e2}_t \xrightarrow{p} E(s^{e2}_t) = Q_s$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T-h} v^e_{t+h} s^e_t \xrightarrow{d} N(0, \Omega); \quad \Omega = V\left(\frac{1}{T^{1/2}} \sum_{t=1}^{T-h} v^e_{t+h} s^e_t\right)$$

$$\Omega \approx \sum_{j=-\infty}^{\infty} E(s^e_t v_{t+h} v_{t+h-j} s^e_{t-j}) \approx$$

$$\frac{1}{T} \sum_{t=1}^{T-h} v^2_{t+h} s^{e2}_t + \frac{1}{T} \sum_{l=1}^{L} \sum_{t=l+1}^{T-h} \omega_l v_{t+h} v_{t+h-l} s^e_t s^e_{t-l}; \ \ \omega_l = 1 - \frac{l}{L+1}$$

Remarks

- I am using T instead of $T - h$ to keep it simple
  asymptotically, it makes no difference
- Newey-West or any other HAC estimator ok
- In principle, $L = h; \; h = 1, \ldots, H$
  can truncate at L_{max} for efficiency
- Lusompa (2020) GLS directly tackles MA errors

Set-up

$$\underset{(H+1)\times 1}{\boldsymbol{y}^e_{t,H}} \equiv (y^e_t \;\ldots\; y^e_{t+H})'; \qquad \underset{(H+1)\times(H+1)}{S^e_t} \equiv I_{(H+1)} \otimes \underset{1\times 1}{s^e_t}$$

$$\underset{(H+1)\times 1}{\boldsymbol{v}_{t,H}} \equiv (v_t \;\ldots\; v_{t+H})'; \qquad \underset{(H+1)\times(H+1)(k+l)}{Z^e_t} \equiv (\underset{(H+1)\times(H+1)k}{X^e_t} \;\; (I_{(H+1)} \otimes \underset{1\times l}{\boldsymbol{z}^e_t}));$$

moment condition:

$$E[Z_t'(\boldsymbol{y}^e_{t,H} - S^e_t\boldsymbol{\beta})] = E[Z^{e\prime}_t \boldsymbol{v}_{t,H}] = 0$$

with

$$\mathcal{R} = \underset{H+1\times 1}{\boldsymbol{\beta}} = (\beta_0 \;\ldots\; \beta_H)'$$

Objective function

recall the moment condition:

$$E[Z_t'(\mathbf{y}_{t,H}^e - S_t^e\boldsymbol{\beta})] = E[Z_t^{e\prime}\mathbf{v}_{t,H}] = 0$$

objective function:

$$\min_{\boldsymbol{\beta}} \left[\sum_{t=1}^{T-H} Z_t^{e\prime}(\mathbf{y}_{t,H}^e - S_t^e\boldsymbol{\beta})\right]' \hat{W} \left[\sum_{t=1}^{T-H} Z_t^{e\prime}(\mathbf{y}_{t,H}^e - S_t^e\boldsymbol{\beta})\right]$$

$$\hat{W} = \left(\frac{1}{T}\sum_{t=1}^{T-H} Z_t^{e\prime}\mathbf{v}_{t,H}\mathbf{v}_{t,H}' Z_t^e\right)^{-1}$$

Estimator

In the simple case

$$\hat{\gamma} = \left(\frac{1}{T} \sum_{t=1}^{T-H} Z_t^{e\prime} S_t^e \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T-H} Z_t^{e\prime} \mathbf{y}_{t,H}^e \right)$$

more generally:

$$\hat{\gamma} = \left(\frac{1}{T} \sum_{t=1}^{T-H} S_t^{e\prime} Z_t^e \hat{W} Z_t^{e\prime} S_t^e \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T-H} S_t^{e\prime} Z_t^e \hat{W} Z_t^{e\prime} \mathbf{y}_{t,H}^e \right)$$

The residual structure

Useful later when we construct GLS

$$\mathbf{v}_{t,H} = \begin{pmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_{t+H} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 \\ \phi_1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \phi_H & \phi_{H-1} & \dots & 1 \end{pmatrix}}_{\Phi} \underbrace{\begin{pmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+H} \end{pmatrix}}_{\mathbf{u}_{t,H}}$$

in the AR(1) example, $\phi_h = \phi^h$ and $\beta_h = \phi_h$

Note $\hat{\phi}_h = \hat{\beta}_h \implies$ exploit for GLS

Estimating LP covariance matrix Σ

Using optimal $\hat{W}$ defined earlier, usual GMM result is:

$$\Sigma = \left(\frac{1}{T} \sum_{t=1}^{T} Z_t' S_t \left(\frac{1}{T} \sum_{t=1}^{T} Z_t' \Phi \boldsymbol{u}_{t,H} \boldsymbol{u}_{t,H}' \Phi' Z_t \right)^{-1} S_t' Z_t \right)^{-1}$$

but Φ unknown. solutions:

- Newey-West (as we saw earlier)
- recursive estimates of Φ (GLS)
- block bootstrap
- Bayesian methods

Comments on GMM

- nothing unusual in using GMM to estimate LPs
- LPs induce MA structure on residuals
- optimal weighting matrix should reflect this
- GMM results on LM test useful later
- also useful later for Gaussian Basis Functions

Error bands

Inference on the trajectory of the response

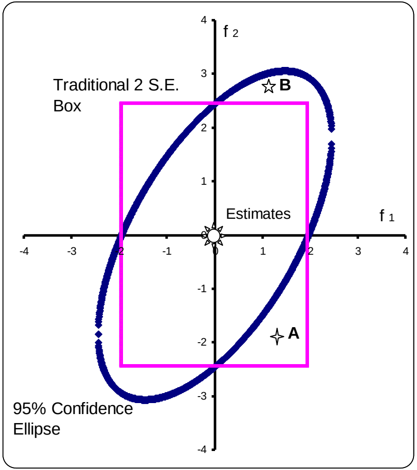
key reference

"Simultaneous confidence bands: theory, implementation, and an application to SVARs" by José Luis Montiel Olea and Mikkel Plagborg-Møller

idea

$\mathcal{R}_h$ is correlated with $\mathcal{R}_{h-1}$

In AR(1) example $CORR(\hat{\mathcal{R}}_h, \hat{\mathcal{R}}_{h-1}) = \phi$



The *sup-t* procedure for joint inference

let the $H \times 1$ vector $\hat{\mathcal{R}}$ collect impulse response coeffs
assume

$$\hat{\mathcal{R}} \xrightarrow{d} \mathcal{N}(\mathcal{R}, \Sigma)$$

can show error bands for response are such that:

$$P\left(\bigcap_{h=1}^{H}\left[\mathcal{R}_h \in \hat{\mathcal{R}}_h \pm c\,\hat{\sigma}_h\right]\right) \to P\left(\max_h |\sigma_h\, v_h| \leq c\right)$$

choose c as smallest c.v. with simultaneous coverage

$$c = q_{1-\alpha}(\Sigma) \equiv q_{1-\alpha}\left(\max_h |\sigma_h^{-1}\, v_h|\right)$$

where $\boldsymbol{v} = (v_1, \ldots, v_H)' \sim \mathcal{N}(\mathbf{0}_H, \Sigma)$ and $\sigma_h = \Sigma_{[h,h]}$

A simple algorithm to implement sup-t procedure

based on asymptotic normality

start with estimates of the response: $\hat{\mathcal{R}}, \hat{\Sigma}$

1. draw i.i.d. vectors $\hat{\boldsymbol{v}}^{(s)} \sim \mathcal{N}(\mathbf{0}_H, \hat{\Sigma})$, for $s = 1, ..., S$
2. define $\hat{q}_{1-\alpha}$ as the empirical $1-\alpha$ quantile of $\max_h |\hat{\sigma}_h^{-1}\, \hat{v}_h^{(s)}|$ across $s = 1, \ldots, S$ with $\hat{\sigma}_h = \Sigma_{[h,h]}$
3. construct bands as $\bigcap_{h=1}^{H} [\hat{\mathcal{R}}_h - \hat{\sigma}_h \hat{q}_{1-\alpha}, \hat{\mathcal{R}}_h + \hat{\sigma}_h \hat{q}_{1-\alpha}]$

Bootstrap/Bayesian version of sup-t algorithm

denote $\hat{P}$ as either the bootstrap or posterior $\hat{\phi}$

1. $\hat{\phi}$ can be VAR parameters so that $\hat{\mathcal{R}} = \mathcal{R}(\hat{\phi})$
2. $\hat{\phi}$ can be local projection estimates so that $\hat{\mathcal{R}} = \hat{\phi}$

and generate $s = 1, \ldots, S$ draws $\hat{\mathcal{R}}^{(s)}$
Hence:

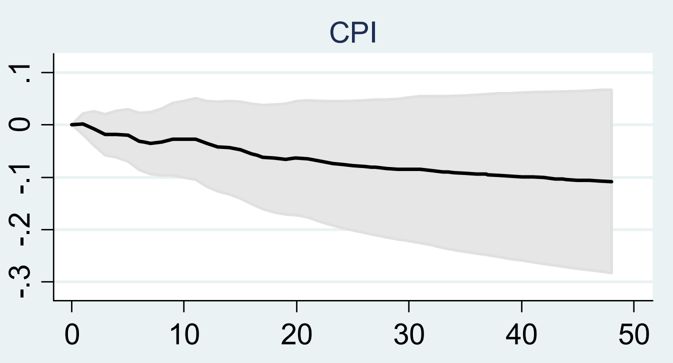
1. let $\hat{q}_{h,\delta}$ denote the empirical δ quantile of $\hat{\mathcal{R}}_h^{(s)}$
2. $$\hat{\delta} = \sup \left\{ \delta \in \left[\frac{\alpha}{(2H)}, \frac{\alpha}{2} \right] \middle| \frac{\sum_{s=1}^{S} \mathbb{I}\left(\hat{\mathcal{R}}^{(s)} \in \bigcap_{h=1}^{H} [\hat{q}_{h,\delta}, \hat{q}_{h,1-\delta}] \right)}{S} \geq 1 - \alpha \right\}$$
3. construct bands as $\bigcap_{h=1}^{H} [\hat{q}_{h,\hat{\delta}}, \hat{q}_{h,1-\hat{\delta}}]$

Significance bands

Motivation

a common situation with VARs

Response of log CPI to a monetary shock



Basic idea
some observations

- **temptation:** the response of CPI is basically zero
- **observation 1:** all (48) coefficients negative rather than randomly alternating between +/-
- **observation 2:** response coefficients (highly) correlated
- **observation 3:** collinearity $\rightarrow$ low individual t-stats (wide bands), sometimes high F-stat

proposition: often the key question is significance of the overall response rather than estimation uncertainty

is the average treatment effect (ATE) different from zero?

A simple example

let $\{y_t\}_{t=1}^{T}$ be mean zero, stationary and homoscedastic AR(1). Using local projections (LPs):

$$y_{t+h} = \beta_h y_t + u_{t+h} \qquad l = 1, ..., H$$

so that

$$\hat{\beta}_h = \frac{\frac{1}{n}\sum_{t=1}^{n} y_{t+h} y_t}{\frac{1}{n}\sum_{t=1}^{n} y_t^2}$$

with n subset of T observations available for estimation

under the null

$$H_0 : \beta_h = 0, \; \forall \; h \qquad \rightarrow \qquad \tilde{u}_{t+h} = y_{t+h}$$

here $\tilde{u}$ denotes the residuals under the null

A simple example
continued

using usual OLS formula for variance of $\hat{\beta}_h$, under the null,

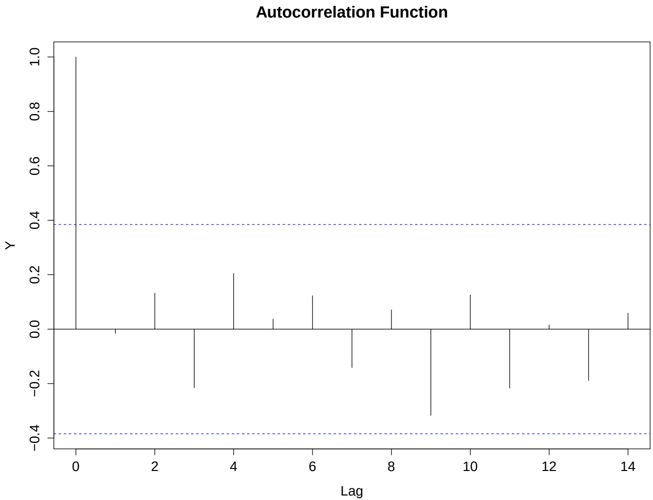
$$\tilde{\sigma}^2_{\hat{\beta}} = \frac{\frac{1}{n}\sum_{t=1}^{n} y_{t+h}^2}{\sum_{t=1}^{n} y_t^2} \xrightarrow{p} \frac{1}{n}$$

since y_t is stationary and under H_0, no serial correlation

- hence, asymptotic confidence interval is $\pm c_{(1-\alpha/2)}/\sqrt{n}$
- $c_{(1-\alpha/2)}$ standard Gaussian critical value
- same as autocorrelogram error bands

Significance bands in a local projection

the autocorelogram is the LP in an AR(1)



Significance bands

LPIV set up and using x_t^e notation for $x_t - \mathcal{P}_L(x_t|I_t)$

LPIV: $y^e_{t+h} = s^e_t \gamma_h + u_{t+h}$. Instrument: z^e_t. Null: $H_0 : \gamma_h = 0$

$$\sqrt{T}(\hat{\gamma}_h - 0) = \frac{\frac{1}{T^{1/2}} \sum_{t=1}^{T-h} z^e_t y^e_{t+h}}{\frac{1}{T} \sum_{t=1}^{T-h} z^e_t s^e_t}; \quad \frac{1}{T^{1/2}} \sum_{t=1}^{T-h} z^e_t y^e_{t+h} \xrightarrow{d} N(0, V);$$

$$\frac{1}{T} \sum_{t=1}^{T-h} z^e_t s^e_t \xrightarrow{p} q_{zs}$$

What is V under the null hypothesis?

The variance under the null

Key: the variance is not a function of h!

$$V = V\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t^e y_{t+h}^e\right) \approx \sum_{j=-\infty}^{\infty} E(z_t^e y_{t+h}^e z_{t-j}^e y_{t_h-j}^e)$$

$$= \sum_{j=-\infty}^{\infty} E(z_t^e z_{t-j}^e) E(y_{t+h}^e y_{t+h-j}^e) \quad \text{under } H_0 + \text{lead-lag exogeneity}$$

$$= \sum_{j=-\infty}^{\infty} \varphi_{z,j}\varphi_{y,j} = \varphi_{z,0}\varphi_{y,0} \quad \text{if } \boldsymbol{z} \text{ serially uncorrelated}$$

hence

$$\tilde{\sigma}_h^2 = \hat{q}_{zs}^{-1}\hat{V}\hat{q}_{zs}^{-1}$$

Note: use Barlett-type correction for $\hat{V}$ (e.g. NW weights)