Introduction to Local Projections
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
1. BASIC IDEAS
2. VAR-LP NEXUS
3. MULTIPLIERS AND COUNTERFACTUALS
4. PANEL DATA BASICS
5. COINTEGRATION
6. VARIANCE DECOMPOSITIONS
7. SMOOTHING
8. EXTENSIONS
BASIC IDEAS
Borrowing from applied micro to draw a parallel
Impulse responses: a comparison of two averages

\[ R(h) = E( E[y_{t+h}|s_t = s + \delta, x_t] - E[y_{t+h}|s_t = s, x_t] ) \]

\( y_{t+h} \): outcome
\( s_t \): intervention
\( s \): baseline, e.g., \( s = 0 \)
\( \delta \): dose, e.g., \( \delta = 1; \delta = \text{var}(\epsilon)^{1/2}; \ldots \)
\( x_t \): vector of exogenous and predetermined variables
Main issues to be solved

- Identification: next section
- Estimation of $E[y_{t+h}|s_t; x_t]$
- Interpretation: multipliers
- Inference: discussed later
A trivial example

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

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

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

Remarks:

- inefficient (not using $x_t$), but *consistent*
- could control for $x_t$ 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 $x_t$ easily)
Estimation by Local projections

Linear case:

\[ y_{t+h} = \alpha_h + \beta_h s_t + \gamma_h x_t + v_{t+h}; \quad v_{t+h} = u_{t+h} + \psi_1 u_{t+h-1} + \ldots + \psi_h u_t \]

As long as \( s_t, x_t \) exogenous w.r.t. \( v_t \), then \( \hat{\beta}_h \to \beta_h \) (identification) and then:

\[ R_{sy}(h) = E[y_{t+h}|s_t = s_1; x_t] - E[y_{t+h}|s_t = s_0; x_t] = \beta_h(s_1 - s_0) \]

General case:

\[ y_{t+h} = m(s_t, x_t; \theta_h) + v_{t+h} \to R_{sy}(h) = m(s_1, x_t; \theta_h) - m(s_0, x_t; \theta_h) \]

i.e. \( m(s_t, x_t; \theta_h) \) can be a nonlinear function
Remarks

- **single equation estimation**: easily scales to panel, easy to extend to nonlinear specifications
- **effects ‘local’ to each $h$**: no cross-period restrictions
- **errors serially correlated**: needs fixing
- **from binary to continuous treatment (dose)**

Many assumptions implicit in **linear** formulation:

- **symmetry**: increase in dose same as decrease
- **scale independence**: double dose, double the effect
- **state independence**: the $x_t$ don’t affect $\mathcal{R}(h)$
- **treatment does not affect covariate effects**: $\gamma^0_h = \gamma^1_h$
- **$\delta|x$ randomly assigned**

We will analyze/generalize each of these assumptions
simple illustration of different variable transformations:
- *levels vs. differences* (e.g. price index vs inflation)
- *levels = long-differences = cumulative of differences*

\[
\Delta y_{t+h} + \ldots + \Delta y_t = y_{t+h} - y_{t+h-1} + y_{t+h-1} - y_{t+h-2} + \ldots y_t - y_{t-1} = y_{t+h} - y_{t-1}
\]

- shows a simple way to construct the loop and plot LPs
- maybe useful to build upon. Much left undone. Will come back to it
RELATION TO VARs REMINDER
Set aside identification discussion for now
Propagation in an AR(1)
suppose:

\[(y_t - \mu) = \psi(y_{t-1} - \mu) + u_t\]

by recursive substitution:

\[(y_{t+h} - \mu) = \psi^{h+1}(y_{t-1} - \mu) + u_{t+h} + \psi u_{t+h-1} + \ldots + \psi^h u_t\]

\{intrinsic MA residuals\}

suppose the intervention is \(u_t = \delta; (u_{t+1} = \ldots = u_{t+h} = 0); y_{t-1} = y^*\)

\[
\mathcal{R}(h) = E\left( E[y_{t+h} | u_t = \delta; y_{t-1} = y^*] - E[y_{t+h} | u_t = 0; y_{t-1} = y^*] \right)

= E \left( \{\psi^{h+1}(y^* - \mu) + \psi^h \delta\} - \psi^{h+1}(y^* - \mu) \right)

= E(\psi^h \delta) = \psi^h \delta\]
Remarks

- **iterative approach** with AR(1): from $\hat{\psi}$ obtain $\hat{\psi}_h$
- inference based on **delta method**:
  \[ H_0 : \psi = 0 \implies H_0 : ATE(h) = \mathcal{R}(h) = \psi_h = 0 \]
- **direct approach** with local projections:
  \[
y_{t+h} = \alpha_{h+1} + \psi_{h+1}y_{t-1} + v_{t+h}; \quad h = 0, 1, \ldots
\]
- note: $v_{t+h} = u_{t+h} + \psi u_{t+h-1} + \ldots + \psi^h u_t$
- hence $E[y_{t-1}, v_{t+h}] = 0 \implies \hat{\psi}_{h+1} \xrightarrow{p} \psi^{h+1}$
- inference: correct error serial correlation (we will see how)
- $H_0 : ATE(h) = \mathcal{R}(h) = \psi_h = 0$
propagation in a VAR(2) just to see the details

\[ y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t \]

by recursive substitution:

\[ y_{t+1} = (A_1^2 + A_2) y_{t-1} + A_1 A_2 y_{t-2} + u_{t+1} + A_1 u_t \]

one more time:

\[ y_{t+2} = (A_1^3 + A_2 A_1 + A_1 A_2) y_{t-1} + (A_1^2 A_2 + A_2^2) y_{t-2} + u_{t+2} + A_1 u_{t+1} + (A_1^2 + A_2) u_t \]

takeaway: \( R(h) \) a complicated function of \( A_1, A_2 \) (more on this later, an issue also raised in recent Plagborg-Møller papers)
FURTHER EXPLORATION OF THE VAR–LP NEXUS
A note on lag lengths

- iterated VAR-based forecasts need *correct specification*
- if not, responses will be biased
- consistency of $\mathcal{R}(h)$ only if in VAR($p$) s. t. $p \rightarrow h$ as $h \rightarrow \infty$
- local projections are approximations
- no correct specification assumed
- smaller lag lengths ok for consistency under mild assumptions
- however, lag-augmentation can be very helpful for inference (later)

Some results derived more formally later
Using a VAR to construct $E[y_{t+h}|s_t, x_t]$
Reduced-form only to explain VAR(p) vs. VAR(∞) issues

consider a VAR(p): (assume $s_t$ and $x_t$ in $y_t$)

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t; \quad E(u_t u_t') = \Sigma_u$$

by recursive substitution, $VMA(\infty)$:

$$y_t = u_t + B_1 u_{t-1} \ldots + B(\infty) y_0;$$

$B(\infty) y_0 \to 0$ if $|A(z)| \neq 0$ for $|z| \leq 1$  MA invertibility

$B(\infty) = B(A_1, \ldots, A_p)$, e.g., see Slide 13

$y_0$ is distant initial condition. MA invertibility $\implies B(\infty) \to 0$
Relation between $VAR(p)$ and $VMA(\infty)$

Recall the impulse response representation

\[ B_1 = A_1 \]
\[ B_2 = A_1 B_1 + A_2 \]
\[ \vdots = \vdots \]
\[ B_i = A_1 B_{i-1} + A_2 A_{i-2} + \ldots + A_p B_{i-p}; \quad i \geq p \]

or compactly

\[ B_i = \sum_{j=1}^{i} B_{i-j} A_j; \quad i = 1, 2, \ldots; \quad B_0 = I_k \]
Constructing $E[y_{t+h}|s_t, x_t]$ using $VMA(\infty)$ from:

$$y_{t+h} = u_{t+h} + \ldots + B_{h-1}u_{t+1} + B_h u_t + B_{h+1}u_{t-1} + \ldots$$

then:

$$E[y_{i,t+h}|u_{j,t} = 1, u_{t-1}, \ldots] = B_h(i, j)$$

where, $s_t = u_{j,t}$ and $x_t = u_{t-1}, u_{t-2}, \ldots$ hence

$$R(h) = B_h(i, j); \quad \hat{B}_h = \sum_{j=1}^{h} \hat{B}_{h-j} \hat{A}_j; \quad \hat{A}_j \text{ from VAR(p)}$$

Important: in reduced form, $E(u_{i,t}u_{l,t}) \neq 0$ for $i \neq l$, usually hence, this is not yet a well defined experiment
Fitting a finite $\text{VAR}(p)$ to a $\text{VAR}(\infty)$ (1 of 2)

A good assumption if true DGP is VARMA (e.g. many DSGE models)

Suppose the DGP is:

$$y_t = \sum_{i=1}^{\infty} A_i y_{t-i} + u_t \quad \text{with} \quad \sum_{i=1}^{\infty} ||A_i|| < \infty$$

hence:

$$y_t = \sum_{i=0}^{\infty} B_i u_{t-i}; \quad B_0 = I; \quad \det \left( \sum_{i=0}^{\infty} B_i z_i \right) \neq 0$$

for $|z_i| \leq 1$ and $\sum_{i=0}^{\infty} i^{1/2} ||B_i|| < \infty$
Fitting a finite $VAR(p)$ to a $VAR(\infty)$ (2 of 2)

Results from Lewis and Reinsel (1985), a key paper in this literature

Let $p_T$ denote the order of the $VAR(p_T)$. If:

$$p_T \rightarrow \infty; \quad \frac{p_T^3}{T} \rightarrow 0; \quad \sqrt{T} \sum_{i=p_T+1}^{\infty} ||A_i|| \rightarrow 0 \quad \text{as} \quad T \rightarrow \infty$$

then:

$$\sqrt{T} \left[ \text{vec}(\hat{A}_1' \ldots \hat{A}_{p_T}') - \text{vec}(A_1 \ldots A_{p_T}) \right] \xrightarrow{d} N(0, \Sigma^*); \quad \Sigma^* \neq \Sigma_a$$

where $\Sigma_a$ refers to finite $VAR(p)$, and

$$\sqrt{T} \left[ \text{vec}(\hat{B}_h') - \text{vec}(B_h) \right] \xrightarrow{P} N \left( 0, \Sigma_u \otimes \sum_{j=0}^{h-1} B_j \Sigma_u B_j' \right); \quad h \leq p_T$$

Note: consistency not guaranteed for $h > p_T$
Takeaways and references

- $\text{VAR}(\infty)$ results in, e.g., Lütkepohl (2005, Chapter 15)
- many DSGE have VARMA reduced form or $\text{VAR}(\infty)$
- note $p_T$ grows with $T$ but at a slower rate
- consistency of $B_h$ only guaranteed up to $h = p_T$
- unlike $\text{VAR}(p)$, response S.E.s $\to 0$ as $h \to \infty$
- Plagborg-Møller and Wolf (2021): for $h \leq p_T$ VARs and LPs estimate the same response
- Jordà, Singh, and Taylor (2020): for $h > p_T$ VAR responses are biased, but LPs are not (under certain conditions)
VAR vs. LP Bias in infinite lag processes
Or why LPs can be more reliable for long-horizon responses

Intuition:

■ Suppose D.G.P. is:

\[ y_t = \sum_{j=0}^{\infty} A_j y_{t-j} + u_t; \quad \sum_{j=1}^{\infty} ||A_j|| < \infty \]

■ Fit VAR(1)
■ True vs. VAR(1) IRFs

\[
\begin{align*}
VAR(\infty) & \quad VAR(1) \\
B_1 &= A_1 & B_1^* &= A_1 \\
B_2 &= A_1^2 + A_2 & B_2^* &= A_1^2 \\
B_3 &= A_1^3 + 2A_1A_2 + A_3 & B_3^* &= A_1^3 \\
B_4 &= A_1^4 + 3A_1^2A_2 + 2A_1A_3 + A_4 & B_4^* &= A_1^4 \\
\end{align*}
\]
VAR bias

Consistency guaranteed up to $p$ only for $\text{VAR}(\infty)$

**objective:** truncate $\text{VAR}(\infty)$ so that remaining lags are "small"

$$\frac{1}{T^{1/2}} \sum_{j=p+1}^{\infty} ||A_j|| \to 0; \quad p, T \to \infty$$

however, from the usual $\text{VAR} \to \text{VMA}$ recursion, these terms are missing for $h > p$:

$$\text{BIAS} : A_{p+1}B_{h-(p+1)} + \ldots + A_{h-1}B_1 + A_h; \quad h > p$$

**problem:** in practice VARs are truncated too early
LP bias
or lack thereof

when is the LP consistent? i.e., when is this condition met:

$$||\hat{A}_{h,1} - B_h|| \xrightarrow{p} 0; \quad p, T \to \infty$$

in the LP:

$$y_{t+h} = A_{h,1}y_{t-1} + \ldots + A_{h,p}y_{t-p} + u_{t+h}$$

turns out same as consistency of VAR($p$), i.e.

$$p^{1/2} \sum_{j=0}^{\infty} ||A_{k+j}|| \to 0$$

see proof in Jordà, Singh, Taylor (2020)
Illustration of VAR vs. LP bias

Based on MA(24) model
Another example
Figure 2 in Palgborg-Møller and Wolf (2021, ECTA)
Multipliers and Counterfactuals
Two models, same response, different conclusions
Alloza, Gonzalo, Sanz (2020)

\[
\begin{align*}
(a) \quad \begin{cases}
\Delta y_t &= \beta \Delta s_t + u_t^y \\
\Delta s_t &= \rho \Delta s_{t-1} + u_t^s
\end{cases} \quad \text{; (b) } \begin{cases}
\Delta y_t &= \beta \Delta s_t + \rho \Delta y_{t-1} + u_t^y \\
\Delta s_t &= u_t^s
\end{cases} \quad \text{; } u_t \sim D(0, I)
\end{align*}
\]

Note: $R_{sy}^a(h) = \beta \rho^h = R_{sy}^b(h)$. Both can be estimated with the LP:

\[
\Delta y_{t+h} = \gamma_h \Delta s_t + \psi_h \Delta y_{t-1} + v_{t+h}
\]

Propagation in (a), due to correlated treatment, in (b) correlated outcome. Consider augmenting LP with treatment leads:

\[
\Delta y_{t+h} = \gamma_h \Delta s_t + \psi_h \Delta y_{t-1} + \sum_{i=1}^{h} \phi_i \Delta s_{t+i} + v_{t+h};
\]

$\tilde{R}_{sy}^a(h) = \beta$; $\tilde{R}_{sy}^b(h) = \beta \rho^h$
What is going on?

- in both cases, $\Delta s_t$ is strictly exogenous. Leads are allowed in the LP
- in model (a), including leads removes the effect from future potential treatments (due to treatment serial correlation)
- in model (b), on average, there is no expectation of additional treatment. The leads do not matter
- what is the effect of a single treatment? In (a) $\beta$, in (b) $\beta \rho^h$
- think of the LP MA(h) residual structure. In general, the MA would have terms in $u^Y_{t+i}$ and $u^S_{t+i}$. But in model (b) coeffs on $u^S_{t+i}$ are all zero
- another way to think about these effects is using multipliers
From previous example

Consider the following model (model (a) earlier):

\[
\begin{align*}
\Delta y_t &= \beta \Delta s_t + u_t^y, \quad u_t \sim D \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_y & 0 \\ 0 & \sigma_s \end{pmatrix} \right) \\
\Delta s_t &= \rho \Delta s_{t-1} + u_t^s
\end{align*}
\]

Trivially: \( R_{sy}(h) = \beta \rho^h; \ R_{ss}(h) = \rho^h \)

The cumulative impact, \( C_{ij}(h) = \sum_{k=0}^{h} R_{ij}(k) \) can be directly estimated from:

\[
\begin{align*}
y_{t+h} - y_{t-1} &= \Delta_h y_{t+h} = \theta_h \Delta s_t + v^y_{t+h}; \quad v^y_{t+h} \sim MA(h) \\
s_{t+h} - s_{t-1} &= \Delta_h s_{t+h} = \psi_h \Delta s_t + v^s_{t+h}; \quad v^s_{t+h} \sim MA(h)
\end{align*}
\]

with \( C_{sy}(h) = \theta_h = \beta \sum_{k=0}^{h} \rho^k; \ C_{ss}(h) = \psi_h = \sum_{k=0}^{h} \rho^k \)
Calculating the multiplier

Define:

\[ m_h = \frac{c_{sy}(h)}{c_{ss}(h)} = \frac{\beta \sum_{k=0}^{h} \rho^k}{\sum_{k=0}^{h} \rho^k} = \beta; \text{cum. change in } y \text{ due to cum. change in } s \]

Suppose \( \Delta z_t \) is a valid instrument for \( \Delta s_t \) then:

\[
E(\Delta h y_{t+h}, \Delta z_t) = \theta_h E(\Delta s_t \Delta z_t)
\]
\[
E(\Delta h s_{t+h}, \Delta z_t) = \psi_h E(\Delta s_t \Delta z_t)
\]

hence \( m_h \) can be directly estimated from the IV projection:

\[
\Delta h y_{t+h} = m_h \Delta h s_{t+h} + \eta_{t+h}; \quad \text{instrumented with } \Delta z_t
\]


LPs in panels

The set-up

\[ y_{i,t+h} = \alpha_i + \delta_t + s_{i,t}\beta_h + x_{i,t}\gamma_h + v_{i,t+h}; \quad i = 1, \ldots, n; \quad t = 1, \ldots, T \]

- \( \alpha_i \) unit-fixed effects
- \( \delta_t \) time-fixed effects
- \( x_{i,t} \) exogenous and pre-determined variables
- \( s_{i,t} \) treatment variable
- \( \beta_h \) response coefficient of interest

Sample code: LP_example_panel.do
Panel-LPs
Remarks: usual panel data issues appear here too

- LP is costly in short-panels (lost time dimension cross-sections)
- but cross-section brings more power
- incidental parameter issues (fixed effects):
  - beware of high autocorr and low T (Alvarez and Arellano, 2003 ECTA)
  - will need Arellano-Bond or similar estimator
- inference
  - $n, T$ large → two-way clustering helps MA(h) and heteroscedasticity
  - $n$ large, $T$ small → cluster by unit helps with MA(h)
  - $T$ large, $n$ small → cluster by time helps heteroscedasticity
  - else, Driscoll-Kraay is like Newey-West for panel data
  - when clustering with small $n, T$, may need bootstrap.
    See papers here and here.
  See also summclust and boottest STATA ado files
COINTEGRATION

A brief detour
What is cointegration?

Idea: two variables can be I(1) but their linear combination is I(0). Example:

\[
\begin{align*}
  y_{1,t} &= \gamma y_{2,t} + u_{1,t} \\
  y_{2,t} &= y_{2,t} + u_{2,t}
\end{align*}
\]

\[ y_1, t, y_2, t \sim I(1) \quad \text{but} \quad z_t = y_{1,t} - \gamma y_{2,t} \sim I(0) \]

In general:

\[ y_t = \alpha + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t \]

cointegration means:

\[ \Phi(1) \equiv I - \Phi_1 - \ldots - \Phi_p \quad \text{then} \quad \text{rank}(\Phi(1)) = g < n \]

that is, the system has \(n - g\) unit roots and \(g\) cointegrating vectors, s.t. \(\Phi(1) = BA'\) with \(A, B n \times g\) matrices, and \(A'y_t = z_t\) cointegrating vectors
The VECM representation
Using general representation of a VAR(p)

\[ y_{t+1} = \Phi_1 y_t + \ldots + \Phi_{p+1} y_{t-p} + \alpha + u_{t+1} \]
\[ y_{t+1} = \Psi_1 \Delta y_t + \ldots + \Psi_p \Delta y_{t-p+1} + \Pi y_t + \alpha + u_{t+1} \]

with \( \Psi_j = -[\Phi_{j+1} + \ldots + \Phi_{p+1}] \); for \( j = 1, \ldots, p \) and \( \Pi = \sum_{j=1}^{p+1} \Phi_j \)

subtracting \( y_t \) on both sides:

\[ \Delta y_{t+1} = \Psi_1 \Delta y_t + \ldots + \Psi_p \Delta y_{t-p+1} + \Psi_0 y_t + \alpha + u_{t+1} \]

Note: \( \Psi_0 = -\Phi(1) = BA' \) when there is cointegration, and \( z_t = A'y_t \)
How does cointegration affect impulse responses?

Remarks

- responses from levels VAR **always** correct
- responses from differenced VAR **only** correct if no cointegration
- cointegration improves efficiency ...
- ... but estimation and inference more difficult
- responses often not used to investigate LR equilibrium relationships but should
- useful to impose LR exclusion identification restrictions
Cointegrated systems in state-space form

notice:

\[
\Psi_0 = \Pi - I = -\Phi(1);
\]

if \( \text{rank}(\Psi_0) < n \rightarrow \Phi(1) = BA' \); cointegrating vector: \( z_t = A'y_t \)

\[
\begin{bmatrix}
  z_{t+1} \\
  \Delta y_{t+1} \\
  \Delta y_t \\
  \vdots \\
  \Delta y_{t-p+1}
\end{bmatrix} =
\begin{bmatrix}
  A'\Pi & A'\Psi_1 & \ldots & A'\Psi_{p-1} & A'\Psi_p \\
  -B & \Psi_1 & \ldots & \Psi_{p-1} & \Psi_p \\
  0 & I & \ldots & 0 & 0 \\
  \vdots & \vdots & \ldots & \vdots & \vdots \\
  0 & 0 & \ldots & I & 0
\end{bmatrix}
\begin{bmatrix}
  z_t \\
  \Delta y_t \\
  \Delta y_{t-1} \\
  \vdots \\
  \Delta y_{t-p}
\end{bmatrix}
+ \begin{bmatrix}
  A'u_{t+1} \\
  u_{t+1} \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\]

\( Z_{t+1} = \Psi Z_t + V_{t+1} \)
Usefulness of state-space representation

Calculating impulse responses through recursive substitution

long-run dynamics:

\[ z_{t+h} = \Psi^h_{[1,1]} z_t + \Psi^h_{[1,2]} \Delta y_t + \sum_{j=3}^{p-2} \Psi^h_{[1,j]} \Delta y_{t-j+2} + \nu_{t+h} \]

\[ \nu_{t+h} = A' u_{t+h} + A'(I + \Gamma_1) U_{t+h-1} + ... + A'(I + \Gamma_1 + ... + \Gamma_{h-1}) U_{t+1} \]

short-run dynamics:

\[ \Delta y_{t+h} = \Psi^h_{[2,1]} z_t + \Psi^h_{[2,2]} \Delta y_t + \sum_{j=3}^{p-2} \Psi^h_{[2,j]} \Delta y_{t-j+2} + \nu_{t+h} \]

\[ \nu_{t+h} = u_{t+h} + \Gamma_1 u_{t+h-1} + ... + \Gamma_{h-1} u_{t+1} \]

where

\[ \Delta y_t = \sum_{j=0}^{\infty} \Gamma_j u_{t-j} \]
Responses to equilibrium shocks

equilibrium dynamics, short- vs. long-run effects:

\[ R_z(h; A'u_{t+1} = 1) = (I + \Gamma_1 + \ldots + \Gamma_h)A = \Psi^h_{[1,1]} + \Psi^h_{[1,2]}A \]

short-run dynamics, short- vs long-run effects:

\[ R_{\Delta y}(h; A'u_{t+1} = 1) = \Gamma_hA = \Psi^h_{[2,1]} + \Psi^h_{[2,2]}A \]

remarks:

- note shock cointegrating vector, \( z \), not a variable
- each response, 2 parts:
  1. return to equilibrium  (LR)
  2. short-run frictions   (SR)
VARIANCE DECOMPOSITIONS
two important recent references:


can always write $y_{t+h} = \hat{E}_t(y_{t+h}) + \hat{v}_{t+h}$

then $R^2$ of regression of $\hat{v}_{t+h}$ on $\epsilon_{j,t+h}, ..., \epsilon_{j,t}$ measures percent of FEV explained by j-shock

assumes structural shock $\epsilon_{j,t}$ available
SMOOTH LOCAL PROJECTIONS
Smoothing

relevant references:

- Barnichon, Regis and Christian Brownlees. 2018. Impulse response estimation by smooth local projections. Available at: https://sites.google.com/site/regisbarnichon/research

Many solutions. A simple one: Gaussian Basis Functions

**Intuition:** impose some cross-horizon discipline to smooth LP wiggles. Can improve efficiency

**Other options:** bayesian shrinkage
see, e.g. Miranda-Agrippino and Rico. 2018. *Bayesian Local Projections*
A general approach to smoothing

GMM provides local projection estimates of the response $\hat{R}$ given by $\hat{\gamma}$ and $\hat{\Sigma}_{\gamma}$

a natural solution is minimum distance

let $\psi(\hat{\gamma}, \theta)$ be a function that returns a smoothed estimate of $\hat{\gamma}$ based on auxiliary parameters $\theta$, then:

$$\min_{\theta} [\hat{\gamma} - \psi(\theta)]' \hat{\Sigma}_{\gamma} [\hat{\gamma} - \psi(\theta)]$$

delivering, $\hat{\theta}$, $\hat{\Sigma}_{\hat{\theta}}$ and if $dim(\gamma) > dim(\theta)$, a test of overidentifying restrictions for $\psi(\theta)$
Smoothing with Gaussian Basis Functions

Suppose no controls to simplify

\[ R(h; a, b, c) = \psi(h) = ae^{-(\frac{h-b}{c})^2} \]

Using GMM set-up, two estimators: direct v. 2-step

Direct estimator:

\[
\min_{a,b,c} \left[ \sum_{t=1}^{T} Z_t'(y_{t,H} - S_t \psi(h)) \right] \' \hat{W} \left[ \sum_{t=1}^{T} Z_t'(y_{t,H} - S_t \psi(h)) \right]
\]

2-step: Step-1 is usual LP, get \( \hat{\gamma}, \hat{\Sigma}_\gamma \), then min. distance

\[
\min_{a,b,c} [\hat{\gamma} - \phi(h)]' \hat{\Sigma}_\gamma [\hat{\gamma} - \phi(h)]
\]
GBF–GMM
Remarks

- direct method requires NL estimation techniques
- however, problem is reasonably well behaved
- 2-step method provides useful intuition
- note H-period LP, but 3 parameters so \((H + 1) - 3\) overidentifying restrictions
- regardless of method, J-test natural specification test
- considerable gain in parsimony \(\Rightarrow\) efficiency
- GBF approximation works well with "single humps"
- multiple "humps" require more basis functions \(\Rightarrow\) GBF approach no longer as practical
approximation using gaussian basis functions

recall:

\[ R(h) = \psi(h) = ae^{-\left(\frac{h-b}{c}\right)^2} \]

what does each parameter do?

- \( a \) scales the entire response
- \( b \) dates the peak effect
- \( c \) measures the half-life
gaussian basis functions

the picture

\[ \psi(h) = ae^{-\left(\frac{h-b}{c}\right)^2} \]

Sample code: LP_GBF.do
GBF–GMM example
unemployment v. inflation response to monetary policy shock

![Graph showing the response of unemployment and inflation to a monetary policy shock.](image-url)
NONLINEARITIES AND OTHER POTENTIAL EXTENSIONS
The principle

What we are after:

\[ R_{sy}(h) = E[y_{t+h} | s_t = s_0 + \delta; x_t] - E[y_{t+h} | s_t = s_0; x_t] \]

No reason to assume the conditional expectation is linear

Example:

\[ y_{t+h} = \gamma_1 s_t + \gamma_2 s_t^2 + \gamma x_t + v_{t+h} \quad \rightarrow \quad R_{sy}(h) = \gamma_1 (s_0 + \delta) + \gamma_2 (s_0 + \delta)^2 + \gamma x_t - (\gamma_1 s_0 + \gamma_2 s_0^2 + \gamma x_t) \]

\[ = \gamma_1 + \gamma_2 (\delta^2 + 2s_0 \delta) \]

Hence, \( R_{sy}(h) \) depends on \( \delta \) and \( s_0 \), just like NL regression
Binary dependent variable

Example: response probability of financial crisis to today’s credit shock

\[ R_{sy}(h) = P(y_{t+h} = 1|s_t = s_0 + \delta; x_t) - P(y_{t+h} = 1|s_t = s_0; x_t) \]

Remarks:
- logit/probit → \( R_{sy}(h) \) depends on \( s_0, \delta \) and \( x_t \)
- can estimate a linear probability model. But crises are tail events

Another example: Text-based recession probabilities
Response of recession probability
Marginal effect of 1% increase in newspaper-based index

Fig. 2 Marginal effects from Eq. (4.1). Notes: Marginal effects \( \frac{\partial P(Recession_{t+1}=1|x)}{\partial \text{Index}_t} \) from the probit regression for a 1% increase in the newspaper-based index (i.e., a 1% increase in the share of newspaper articles discussing a recession in the USA). Grey shaded areas report 95% confidence intervals.
Quantile LPs

**Example**: does high corporate debt increase risk of left tail GDP draws? Does it depend on legal bankruptcy framework?

\[
\hat{\gamma}_{h,\tau} = \arg\min_{\gamma_{h,\tau}} \sum_{t(p)} \left( \tau \mathbf{1}(\Delta_h y_{it(p)} + h \geq s_{it(p)} \gamma_{h,\tau}) |\Delta_h y_{it(p)} + h - s_{it(p)} \gamma_{h,\tau}| 
+ (1 - \tau) \mathbf{1}(\Delta_h y_{it(p)} + h < s_{it(p)} \gamma_{h,\tau}) |\Delta_h y_{it(p)} + h - s_{it(p)} \gamma_{h,\tau}| \right)
\]

Figure A.4: Business and household debt, responses at 20th percentile of real GDP per capita growth

Notes: Figures show the predictive effects on growth of a two-SD business/household debt buildup in the five years preceding the recession based on a LP series of quantile regressions. Business credit booms shown in the left-hand side panel and household debt booms shown in the right-hand side panel. Shaded areas denote the 95% confidence interval based on bootstrap replications. See text.
Factor models

Idea: control for many covariates using factor model. Suppose:

\[
\begin{aligned}
\begin{cases}
  x_t^{k \times 1} &= \lambda(L)f_t^{q \times 1} + e_t \\
  f_t &= \pi(L)f_{t-1} + \eta_t
\end{cases}
\end{aligned}
\]

\( k >> q \); \( E(e_t) = E(\eta_t) = 0 \); \( E(e_t\eta'_{t-j}) = 0 \ \forall j \)

Then LP can be specified as:

\[
y_{t+h} = \beta_h S_t + \sum_{j=0}^{p} \gamma_j f_{t-j} + v_{t+h}
\]