Exploring Data Dynamics with Local Projections

Òscar Jordà
Department of Economics
U.C. Davis

Web, e-mail, address:

Google Oscar Jordà

November 08
What is a Local Projection?

- It refers to regressions of the form:

\[ y_{t+h} = \beta_0 y_t + \ldots + \beta_p y_{t-p} + u_{t+h} \]

for \( h = 1, \ldots, H \)

- In forecasting, these regressions are sometimes called direct forecasts.

- What are they useful for? Essentially, they provide a semi-parametric method to estimate the coefficients of the Wold decomposition.
Some Applications of Local Projections

- As a flexible way to calculate the impulse response function (looks like a “treatment effect”):

\[ IR(y_{t+h}; \delta) = E(y_{t+h} | \epsilon_t = \delta; y_{t-1}, ...) - E(y_{t+h} | \epsilon_t = 0; y_{t-1}, ...) \]

- As a flexible way to estimate parameters from dynamic moment conditions, e.g., the parameters of a Phillips curve

\[ \pi_t = \beta E_t(\pi_{t+1}) + \gamma y_t + \epsilon_t \]
... and

- As a flexible way to compute path forecasts

\[
\begin{bmatrix}
E(y_{t+1}|y_t, \ldots) \\
\vdots \\
E(y_{t+H}|y_t, \ldots)
\end{bmatrix}
\]

and hence derive the joint predictive density and appropriate statistics.
What is the motivation for using local projections?

- Models for vector time series are well-known (e.g. VARs, etc), and their likelihoods and their statistical properties are well-understood.
- However, the model’s parameters are rarely of interest themselves. Usually it is a nonlinear function of these parameters that interest us (such as an impulse response or a forecast).
- Therefore, restrictions that are sensible from the perspective of a model, may not be useful for the objects that we want to estimate.
Example: Properties of IRs from a VAR

- **Symmetry:** the response of a variable to a positive treatment has the same shape if the shock is negative instead.
- **Shape-invariance:** the size/sign of the treatment does not affect the shape of the impulse response (it scales it).
- **History independence:** the impulse response is independent of the value of recent observations.
Basic Intuition

Exploring Data Dynamics with Local Projections

November 08
I hope to review 3 useful applications of local projections:

1. Estimation and Inference of Impulse Responses (AER, 2005 + ReStat forthcoming)
2. Projection Minimum Distance (under review)
3. Path Forecasting (JAE, forthcoming)
1. Impulse Responses

- Suppose the data have a Wold (impulse response) representation

\[ y_t = \epsilon_t + B_1 \epsilon_{t-1} + B_2 \epsilon_{t-2} + \ldots \]

- and that it is invertible

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

- Notice this includes all VARs, VARMAss, etc. but excludes some others
Local Projections

• Then, iterating the VAR(∞) representation forward

\[ y_{t+h} = A^h_1 y_t + A^h_2 y_{t-1} + \ldots + \epsilon_{t+h} + B_1 \epsilon_{t+h-1} + \ldots + B_{h-1} \epsilon_{t+1} \]

• with the convenient result

\[ A^h_1 = B_h \text{ for } h \geq 1 \]
Estimation

- Truncate the iterated VAR(∞) representation at some lag $k$ (not terribly important how chosen).
- Let $Y$ collect $T$ obs. of $\{y_{t+1}', \ldots, y_{t+H}'\}$
- Let $Z$ collect $T$ obs. of $\{1, y_{t-1}', \ldots, y_{t-k+1}'\}$
- Let $X$ collect $T$ obs. of $y_t$
- Let $M = I - Z(Z'Z)^{-1}Z'$
- Let $\mathbf{B}$ stack the impulse response matrices $B_h$
- Let $\hat{V} = MY - MX\hat{\mathbf{B}}_T$
Then, the least-squares estimate of the system’s impulse responses is

\[ \hat{B}_T = (X'MX)^{-1} (X'MY) \]

\[ \sqrt{T - k - H} vec(\hat{B}_T - B_0) \overset{d}{\rightarrow} N(0, \Omega_b) \]

\[ \Omega_b = [(X'MX)^{-1} \otimes \Sigma_v] \]

\[ \hat{\Sigma}_v = \frac{\hat{\nu} \hat{\nu}'}{T - k - H} \]
Properties

- When the model is correctly specified, slightly less efficient than VAR-based IRs
- Consistent (and robust to misspecification)
- Can be estimated equation-by-equation (convenient for panels, non-linearities and nonparametrics)
- Hence they can be generalized easily with univariate nonlinear models: stress testing (Drehmann, Patton and Sorensen, 2006); thresholds (Jordà, 2005); STAR (Jordà and Taylor, 2008); spatial correlation in housing markets (Brady, 2007)
Local Projections for Cointegrated Systems

- A state-space representation of a VECM

\[
\begin{bmatrix}
    z_{t+1} \\
    \Delta y_{t+1} \\
    \Delta y_t \\
    \vdots \\
    \Delta y_{t-p+1}
\end{bmatrix}
= 
\begin{bmatrix}
    (I_k - A' B) & A' \Psi_1 & \ldots & A' \Psi_{p-2} & A' \Psi_{p-1} \\
    -B & \Psi_1 & \ldots & \Psi_{p-2} & \Psi_{p-1} \\
    0_{k,k} & I_n & \ldots & 0_{n,n} & 0_{n,n} \\
    \vdots & \vdots & \ldots & \vdots & \vdots \\
    0_{k,k} & 0_{n,n} & \ldots & I_n & 0_{n,n}
\end{bmatrix}
\begin{bmatrix}
    z_t \\
    \Delta y_t \\
    \Delta y_{t-1} \\
    \vdots \\
    \Delta y_{t-p+2}
\end{bmatrix}
+ 
\begin{bmatrix}
    A' \varepsilon_{t+1} \\
    \varepsilon_{t+1} \\
    0 \\
    \vdots \\
    0
\end{bmatrix}
\]

- ...or compactly

\[
Y_{t+1} = GY_t + \varepsilon_{t+1}
\]
Estimator and Properties

- Let $Z_H$ collect $\{z'_{t+1}, \ldots, z'_{t+H}\}_{t=p+1}^{T-H}$
- Let $Y_H$ collect $\{\Delta y'_{t+1}, \ldots, \Delta y'_{t+H}\}_{t=p+1}^{T-H}$
- Let $X$ collect $\{z'_t, \Delta y'_t\}_{t=p+1}^{T-H}$
- Let $W$ collect $\{1, \Delta y'_{t-1}, \ldots, \Delta y'_{t-p+2}\}_{t=p+1}^{T-H}$
- Let $M_W = I - W(W'W)^{-1}W'$
- Then

\[
\hat{IR}(z_{t+h}; \delta_j) = \hat{G}_{1,1}^{h} \hat{A}' \delta_j + \hat{G}_{1,2}^{h} \delta_j
\]

\[
\hat{IR}(\Delta y_{t}; \delta_j) = \hat{G}_{2,1}^{h} \hat{A}' \delta_j + \hat{G}_{2,2}^{h} \delta_j
\]
... estimator continued

\[
\hat{G}_{z,H} = \begin{bmatrix}
\hat{G}^1_{1,1} & \hat{G}^1_{1,2} \\
\vdots & \vdots \\
\hat{G}^H_{1,1} & \hat{G}^H_{1,2}
\end{bmatrix} = Z'_H M_W X (X' M_W X)^{-1}
\]

\[
\hat{G}_{y,H} = \begin{bmatrix}
\hat{G}^1_{2,1} & \hat{G}^1_{2,2} \\
\vdots & \vdots \\
\hat{G}^H_{2,1} & \hat{G}^H_{2,2}
\end{bmatrix} = Y'_H M_W X (X' M_W X)^{-1}
\]

\[
\sqrt{T - p - H} \left( \text{vec}(\hat{G}_{i,H}) - \text{vec}(G_{i,H}) \right) \xrightarrow{d} N (0, \Omega_i) \quad i = z, y
\]
Advantages

- Given the cointegrating vector (estimated or imposed), recall that:

\[
\begin{align*}
\hat{IR}(z_{t+h}; \delta_j) &= \hat{G}_{1,1}^h \hat{A}' \delta_j + \hat{G}_{1,2}^h \delta_j \\
\hat{IR}(\Delta y_t; \delta_j) &= \hat{G}_{2,1}^h \hat{A}' \delta_j + \hat{G}_{2,2}^h \delta_j
\end{align*}
\]

Response to
Long-Run Equilibrium
Response to
Short-run Dynamics

- Example: effect of PPP on carry trade (with Alan Taylor)
PPP Adjustment Speeds

Exploring Data Dynamics with Local Projections

November 08
2. Projection Minimum Distance

The idea

- The first order Euler conditions of many dynamic macroeconomic models provide moment conditions, often estimated by GMM.
- However, models often simplify reality considerably.
- Hence we want a method to cast these moment conditions against a statistical model of the data that imposes the least constraints possible (certainly avoid the restrictive economic model).
Illegitimate Instruments

- Suppose we want to estimate
  \[ y = Y\beta + u \]
  with a set of instruments \( Z \) for \( Y \)
- Suppose the DGP is instead
  \[ y = Y\beta + x\gamma + \varepsilon \]
  where \( Z \) are valid instruments for \( Y \)
- *Are the Z valid instruments for the equation we want to estimate? Usually, no.*
Here is Why

- If $E(Z'x) \neq 0$ and $\gamma \neq 0$ then the moment condition required to ensure that instruments are valid is

  $$E(Z'u) = E(Z'x)\gamma + E(Z'\varepsilon) = E(Z'x)\gamma \neq 0$$

  which is violated...

- What can one do? Even though the $x$ where not included in the economic model, they can be used to restore the legitimacy of the instruments.
Two ways to restore legitimacy

1. Orthogonalize the explanatory variable and the regressors with respect to the omitted $x$

2. Orthogonalize the instruments with respect to the omitted $x$, e.g. regress

$$Z = x\delta + \nu$$

and do the usual IV with the $\hat{\nu}$. This is different than the usual TSLS – here it is the residuals (not the predicted values) that are the valid instruments.
What do Local Projections Have to Do with All This?

- Plenty: local projections provide a semi-parametric method to consistently estimate the first $H$ coefficients of the Wold representation.
- Moment conditions expressed in terms of their Wold representation are simply a collection of restrictions between the parameters of interest and the impulse response coefficient matrices.
- Hence, consistent and asymptotically normal estimates can be obtained by minimum distance methods.
The Mechanics

• Suppose the Euler conditions from a model can be summarized in the system:

\[ y_t = \Phi_F E_t y_{t+1} + \Phi_B y_{t-1} + u_t \]

• Stability of the system means that it has a reduced-form Wold representation

\[ y_t = \epsilon_t + B_1 \epsilon_{t-1} + B_2 \epsilon_{t-2} + \ldots \]

• Plugging back to the Euler equations…
The Mechanics (cont.)

• ... we get the set of conditions

\[ B_h = \Phi_F B_{h+1} + \Phi_B B_{h-1} \] for \( h \geq 1 \)

• These conditions are linear in the parameters \( \Phi_F, \Phi_B \) and hence unique.

• Do not require structural identification since they are based on serial correlation properties of the data.

• Can be estimated by GLS-type step
The Estimator in a nutshell

- Let \( f(\hat{b}_T; \phi) = vec(\hat{B} - \Phi_F \hat{B}_F - \Phi_B \hat{B}_B) \)

- Then a consistent and asymptotically normal estimate of the \( \phi = vec(\Phi_F, \Phi_B) \) is found from

\[
\min_{\phi} f(\hat{b}_T; \phi)' \hat{W} f(\hat{b}_T; \phi)
\]

- Let \( F_b = \frac{\partial f(\hat{b}_T; \phi)}{\partial b} \); \( F_\phi = \frac{\partial f(\hat{b}_T; \phi)}{\partial \phi} \) and

\[
\hat{W} = (F_b' \Omega_b^{-1} F_b)^{-1}
\]
Then...

$$\hat{\phi} = (\hat{F}'\hat{W} \hat{F}_\phi)^{-1}(\hat{F}'\hat{W} \hat{b}_T)$$

$$\sqrt{T - H - k} \left( \hat{\phi}_T - \phi_0 \right) \overset{d}{\rightarrow} N(0, \Omega_{\phi})$$

$$\hat{\Omega}_\phi = (F'_\phi \hat{W} F_\phi)^{-1}$$

- with overidentifying restrictions test

$$Q(\hat{b}_T; \hat{\phi}_T) \overset{d}{\rightarrow} \chi^2_{\dim(f(\hat{b}_T;\phi)) - \dim(\phi)}$$
So What is the Connection between PMD and Illegitimate Instruments?

- The constraints implied by the Euler equations are not used to restrict the data generating process used as in, e.g., MLE, Bayesian estimation, or GMM.
- The impulse responses are estimated semi-parametrically – no restrictions on the dynamics of the data but also, one is free to include other variables not originally in the analysis.
- Instruments are sequentially orthogonalized for omitted dynamics and/or omitted variables – this becomes useful for model checking.
Example

• You consider estimating the forward-looking Phillips curve: $\pi_t = \beta E_t \pi_{t+1} + u_t$ using $\pi_{t-2}$ as an instrument, hoping to avoid biases with $\pi_{t-1}$

• The true model is: $\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \varepsilon_t$

• GMM:
  \[
  \hat{\beta}_{GMM} = \frac{\sum_1^T \pi_{t-2} \pi_t}{\sum_h \pi_{t-2} \pi_{t+1}} \xrightarrow{p} \gamma_f + \gamma_b \frac{\phi_1}{\phi_3}
  \]

• PMD:
  \[
  \hat{\beta}_{PMD} = \frac{\sum_1^T \pi_{t-2} M_{t-1} \pi_t}{\sum_h \pi_{t-2} M_{t-1} \pi_{t+1}} \xrightarrow{p} \gamma_f + \gamma_b \frac{0}{\delta_3} = \gamma_f
  \]
Other Advantages: An ARMA(1,1) Monte Carlo

\[ y_t = \rho y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \]

<table>
<thead>
<tr>
<th></th>
<th>( h = 2 )</th>
<th>( h = 5 )</th>
<th>( h = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \rho )</td>
<td>( \theta )</td>
<td>( \rho )</td>
</tr>
<tr>
<td>PMD</td>
<td>Est.</td>
<td>0.24</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>SE (MC)</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>MLE</td>
<td>Est.</td>
<td>0.25</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>SE (MC)</td>
<td>0.15</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Table 2.1.2 Monte Carlo Comparison: GMM vs. PMD.

Case 1: \( w_f = \beta_f = 0.7; w_b = \beta_b = 0.3; \gamma = 0.13; \beta_r = 0.09; \rho = 0.5; \gamma_{\pi} = 1.50; \gamma_y = 0.5 \)

D.G.P.

\[
\begin{align*}
\pi_t &= w_f \pi_{t+1} + w_b^1 \pi_{t-1} + w_b^2 \pi_{t-2} + \gamma y_t + \epsilon_{x,t} \\
y_t &= \beta_f E_t y_{t+1} + \beta_b^1 y_{t-1} + \beta_b^2 y_{t-2} - \beta_r (R_t - E_t \pi_{t+1}) + \epsilon_{y,t} \\
R_t &= (1 - \rho) (\gamma_{\pi} \pi_t + \gamma_y y_t) + \rho R_{t-1} + \epsilon_{R,t}
\end{align*}
\]

\[
\begin{align*}
\epsilon_{x,t} &= u_{x,t} \\
\epsilon_{y,t} &= \rho_y \epsilon_{y,t-1} + u_{y,t} \\
\epsilon_{R,t} &= \rho_R \epsilon_{R,t-1} + u_{R,t}
\end{align*}
\]

\[
\begin{align*}
u_{x,t} &\sim N(0, 0.5^2) \\
u_{y,t} &\sim N(0, 0.288^2) \\
u_{R,t} &\sim N(0, 0.252^2)
\end{align*}
\]

Instrument list includes \( R_t \)

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>( \rho_g = \rho_R = 0 )</th>
<th>( \rho_g = 0.5; \rho_R = 0.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMD</td>
<td>GMM</td>
</tr>
<tr>
<td>( w_f = 0.7 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.710</td>
<td>0.684</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>( w_b^1 = 0.3; w_b^2 = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.290</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>( \gamma = 0.13 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.130</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.070)</td>
</tr>
</tbody>
</table>
3. Path Forecasting

• What is a path forecast? A collection of 1 to $H$ step-ahead forecasts

$$\hat{Y}_\tau(H) = \begin{bmatrix} \hat{y}_\tau(1) \\ \vdots \\ \hat{y}_\tau(H) \end{bmatrix} ; Y_{\tau,H} = \begin{bmatrix} y_{\tau+1} \\ \vdots \\ y_{\tau+H} \end{bmatrix}$$

• Suppose for convenience

$$\sqrt{T} \left( \hat{Y}_\tau(H) - Y_{\tau,H} | y_\tau, y_{\tau-1}, \ldots \right) \xrightarrow{d} N \left(0, \Xi_H \right)$$
What is the Uncertainty Associated with the Path Forecast?

- Formally, invert the statistic for the null

\[ H_0 : E \left( \hat{Y}_\tau(H) - Y_{\tau,H} | y_\tau, y_{\tau-1}, \ldots \right) = 0 \]

- ...but this is a multidimensional ellipse since

\[ W_H = T \left( \hat{Y}_\tau(H) - Y_{\tau,H} \right)' \Xi_H^{-1} \left( \hat{Y}_\tau(H) - Y_{\tau,H} \right) \xrightarrow{d} \chi^2_H \]

- and the region we want to plot is that which satisfies

\[ \Pr \left[ W_H \leq c^2_{\alpha}(H) \right] = 1 - \alpha \]
Scheffe Bands

- Scheffe’s (1953) S-Method: provides a method to obtain statistics for nulls that involve linear combinations of the original joint null
- A particularly appealing linear combination allows one to maximize the joint variation of the path
- Let $P$ be the Cholesky factor for $\Xi_H = PP'$ then bands can be constructed as...
Scheffe Bands (cont.)

\[ \hat{Y}_\tau(H) = P \left[ \sqrt{\frac{c^2_\alpha(h)}{h}} \right]_h^{H} \]

- instead of the usual

\[ \hat{Y}_\tau(H) = z_{\alpha/2} \text{diag}(\Xi_H)^{1/2} \]

- or the Bonferroni version

\[ \hat{Y}_\tau(H) = z_{\alpha/2H} \text{diag}(\Xi_H)^{1/2} \]
Intuition

95% Scheffe Upper Bound (1.73, 3.03)

Traditional 2 S.E. Box

95% Scheffe Lower Bound (-1.73, -3.03)

95% Confidence Ellipse

Estimated Values
Some Pictures...
... and some Monte Carlos

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Nominal Coverage: 68%</th>
<th>Nominal Coverage: 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FWER</td>
<td>WALD</td>
</tr>
<tr>
<td>T=100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>19.1</td>
<td>85.7</td>
</tr>
<tr>
<td>UN</td>
<td>15.7</td>
<td>85.0</td>
</tr>
<tr>
<td>FF</td>
<td>9.1</td>
<td>83.3</td>
</tr>
<tr>
<td>T=400</td>
<td>18.2</td>
<td>86.4</td>
</tr>
<tr>
<td>UN</td>
<td>14.8</td>
<td>85.9</td>
</tr>
</tbody>
</table>
Summary

• **Some General Principles:**

1. Tailor the statistics for the question you want to answer
2. When you are unsure about the “truth,” choose methods that are robust to misspecification
3. Simple is often times better: if you hear hooves, think horses, not zebras...
4. Models can sometimes be too restrictive
Where next?

- **IRs with Local Projections:**
  - In a panel setting with cointegration and nonlinearities, investigate the returns to carry trade and long-run PPP adjustment
  - Harrod-Balassa-Samuelson

- **PMD**
  - Estimation of GARCH models, specifically multivariate GARCH

- **Path Forecasting:**
  - Mahalanobis vs. RMSE and other alternatives
  - Value at Risk and Path-dependent options
  - Path Predictive Ability Testing
Pie in the Sky

- Using IV to estimate structural IRs with local projections
- Semiparametric Dynamic Treatment Effects: use the Mahalanobis distance instead of propensity scores for dynamic treatments, e.g. the effects of monetary policy