Micro responses to macro shocks

Martín Almuzara[1] Víctor Sancibrián[2]

[1]Federal Reserve Bank of New York

[2]CEMFI

NBER Summer Institute
Forecasting and Empirical Methods

July 10, 2024

Disclaimer: The views below are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System

Motivation

- Estimates of transmission of aggregate shocks to individual outcomes are key objects.
- Panel local projection (LP) at horizon h:

$$Y_{i,t+h} = \beta(h) s_i X_t + \text{controls} + \xi_{it}(h)$$

with micro outcome Y_{it}, macro shock X_t and micro covariate $s_i \implies$ least squares $\hat{\beta}(h)$.

➡ Ottonello, Winberry (2020)
$Y =$ firm-level investment
$X =$ monetary policy shock,
$s =$ leverage/distance to default

➡ Holm, Paul, Tibshirek (2021)
$Y =$ household income/spending,
$X =$ monetary policy shock,
$s =$ liquid assets indicator.

- Despite a lot of progress in time series, little is known about the panel data case.

This paper

1. What is $\hat{\beta}(h)$ estimating?

2. How to compute standard errors/confidence intervals?

- We study these questions in a general setup:
  - Observed and unobserved, macro and micro shocks.
  - Heterogeneous, dynamic transmission.
    - ➡ $Y_{it} = \mu_i + \sum_{\ell=0}^{\infty} \beta_{i\ell} X_{t-\ell} + v_{it}$
- Macro shock of interest X is observed.
  - Can be relaxed under finite-order VAR or LP-IV assumptions.
- We allow for DGPs with potentially low macro-micro signal-noise.

Contributions

1. Estimand of $\hat{\beta}(h)$.
   - Population projection of impulse response β_{ih} on s_i:

$$\beta(h) = \frac{\text{Cov}(s_i, \beta_{ih})}{\text{Var}(s_i)}.$$

   Nonparametric in the sense of permitting unrestricted unobserved heterogeneity.

2. Panel LP inference.
   - Clustering on t, not on i. It's not necessary.
   - Lags + heteroskedasticity-robust or HAR inference.
     - ➡ Connection with a synthetic time series of sample regression coefficients.

   Uniform validity over DGPs with different macro-micro signal-noise.

Empirical relevance

- Disagreement in the choice of standard errors in applied work.
  In our review of almost 50 recent empirical papers:
  - (i, t)-clustering (two-way) $\approx$ 50%
  - i-clustering (within units) $\approx$ 33%
  - Driscoll, Kraay (1998) $\approx$ 15%

- Instead, our recommendation:
  - Time clustering + lags + heteroskedasticity robust standard errors
  - Small-sample refinements if T is small; Imbens, Kolesár (2016)
  - Inference is simple and robust to the pervasiveness of micro variation.

Literature

- Time series literature on local projections.
  Jordà (2005), Stock, Watson (2018), Montiel Olea, Plagborg-Møller (2021), Xu (2023) ...
- Estimation and inference with aggregate shocks.
  Hahn, Kuersteiner, Mazzocco (2020), Arkhangelsky, Korovkin (2023), Majerovitz, Sastry (2023)
- Models with cross-sectional dependence.
  Driscoll, Kraay (1998), Andrews (2005), Pesaran (2006), Gonçalves (2011)

➡ This paper: panel data + aggregate shocks + robustness to macro signal strength

Outline

1. Introduction
2. Panel local projections
3. Empirical illustration
4. Conclusion

Panel local projections

Model: setup

- General DGP:

$$Y_{it} = \mu_i + \sum_{\ell=0}^{\infty} \beta_{i\ell} X_{t-\ell} + v_{it}, \qquad t = 1, \ldots, T, \quad i = 1, \ldots, N,$$

$$v_{it} = \sum_{\ell=0}^{\infty} \gamma_{i\ell} Z_{t-\ell} + \kappa \sum_{\ell=0}^{\infty} \delta_{i\ell} u_{i,t-\ell}.$$

- Macro errors Z and micro errors u (serially uncorrelated).
- Unobserved heterogeneity $\theta_i = \{\mu_i, \{\beta_{i\ell}\}_\ell, \{\gamma_{i\ell}\}_\ell, \{\delta_{i\ell}\}_\ell\}$.
  - Micro-macro Wold representation, more flexible than VAR.
- Macro-micro signal noise κ.
  - We consider a range of DGPs P_κ where κ might grow as $N \to \infty$.

Model: setup

- R^2's of aggregate shocks:

$$\bar{R}^2 = 1 - \frac{\text{Var}\left(\bar{Y}_t \middle| \{X_\tau, Z_\tau\}, \{\theta_i\}\right)}{\text{Var}\left(\bar{Y}_t \middle| \{\theta_i\}\right)} = 1 - O\left(\frac{\kappa^2}{N}\right),$$

with $\bar{Y}_t = N^{-1} \sum_{i=1}^{N} Y_{it}$.

  - High-signal case $\implies$ κ fixed, $\bar{R}^2 \approx 1$.
  - Moderate-signal case $\implies$ $\kappa \propto \sqrt{N}$, $\bar{R}^2 \in (0, 1)$.
  - Low-signal case $\implies$ $\kappa \gg \sqrt{N}$, $\bar{R}^2 \approx 0$.

- But... κ is not estimable.

- Object of interest. Features of the distribution of $\{\beta_{ih}\}$.

Model: assumptions

Assumption: stationarity and iidness

$\{X_t, Z_t, \{u_{it}\}_i\}$ stationary given $\{\theta_i, s_i\}_i$.

$\{\theta_i, s_i, \{u_{it}\}\}_i$ i.i.d. over i given $\{X_t, Z_t\}$.

Assumption: shocks and mean independence

$$E\left[X_t \middle| \{X_\tau\}_{\tau\neq t}, \{Z_\tau, \{u_{i\tau}\}_i\}, \{\theta_i, s_i\}_i\right] = 0.$$
$$E\left[Z_t \middle| \{Z_\tau\}_{\tau\neq t}, \{X_\tau, \{u_{i\tau}\}_i\}, \{\theta_i, s_i\}_i\right] = 0.$$
$$E\left[u_{it} \middle| \{u_{i\tau}\}_{\tau\neq t}, \{X_\tau, Z_\tau\}, \theta_i, s_i\right] = 0.$$

Regularity cond's: decay of β, γ, δ + moments of X, Z, u + summability of squares

Panel local projections: estimator

- Panel LP at horizon h with p lags + unit and time FEs:

$$
\begin{aligned}
Y_{i,t+h} &= \hat{\beta}(h) s_i X_t + \sum_{\ell=1}^{p} \hat{\varphi}_j(h) s_i X_{t-\ell} + \hat{\mu}_i(h) + \hat{\nu}_t(h) + \hat{\xi}_{it}(h) \\
&= \hat{\beta}(h) \hat{s}_i \hat{X}_t + \sum_{\ell=1}^{p} \tilde{\varphi}_j(h) s_i X_{t-\ell} + \tilde{\mu}_i(h) + \tilde{\nu}_t(h) + \hat{\xi}_{it}(h)
\end{aligned}
$$

  with $\hat{X}_t =$ residual from regressing X_t on $1, X_{t-1}, \ldots, X_{t-p}$ and $\hat{s}_i = s_i - N^{-1} \sum_{j=1}^{N} s_j$.

- Can include additional macro and micro controls.

- Easy to extend to unbalanced panels and time-varying s.

Panel local projections: inference

- Confidence interval based on sandwich formula for standard errors:

$$\hat{C}_\alpha(h) = \left[\hat{\beta}(h) \pm z_{1-\alpha/2}\hat{\sigma}(h)\right], \qquad \hat{\sigma}(h) = \sqrt{\frac{\hat{V}(h)}{(T-h-p)\hat{G}^2}}$$

where $\hat{G} = N^{-1}(T-h-p)^{-1}\sum_{i=1}^{N}\sum_{t=p+1}^{T-h}\hat{s}_i^2\hat{X}_t^2$ is the OLS denominator.

- Score variance term $\hat{V}$... should we cluster on i, t? should we HAC?
- Right choice relies on time clustering: $\hat{V}(h) = \hat{V}_0(h) + 2\sum_{\ell=p+1}^{h}\hat{V}_\ell(h)$,

$$\hat{V}_\ell(h) = \frac{1}{N^2(T-h-p)}\sum_{t=\ell+p+1}^{T-h}\left(\sum_{i=1}^{N}\hat{s}_i\hat{X}_t\hat{\xi}_{it}(h)\right)\left(\sum_{i=1}^{N}\hat{s}_i\hat{X}_{t-\ell}\hat{\xi}_{i,t-\ell}(h)\right).$$

Main result

- Asymptotics. $T, N_T \to \infty$ with $T/N_T \to 0$ holding h, p fixed.
- Population regression coefficient $\beta(h) = \text{Cov}(s_i, \beta_{ih}) / \text{Var}(s_i)$.

Proposition: estimand and consistency

$\lim_{T\to\infty} \sup_{\kappa/\sqrt{NT}=o(1)} P_\kappa(|\hat{\beta}(h) - \beta(h)| > M) = 0.$

Proposition: valid inference

$\lim_{T\to\infty} \sup_\kappa \left| P_\kappa(\beta(h) \in \hat{C}_\alpha(h)) - (1-\alpha) \right| = 0.$

- Proofs use drifting parameter sequences (Andrews, Cheng, Guggenberger (2020)).

Synthetic time series representation

- FWL + orthogonality of $\hat{s}_i \hat{X}_t$ wrt all other controls:

$$\hat{\beta}(h) = \frac{\sum_{t=p+1}^{T-h} \sum_{i=1}^{N} \hat{s}_i \hat{X}_t Y_{i,t+h}}{\sum_{t=1}^{T-h} \sum_{i=1}^{N} \hat{s}_i^2 \hat{X}_t^2} = \frac{\sum_{t=p+1}^{T-h} \hat{X}_t \hat{Y}_t(h)}{\sum_{t=p+1}^{T} \hat{X}_t^2},$$

where $\hat{Y}_t(h) = \left(\sum_{i=1}^{N} \hat{s}_i Y_{i,t+h}\right) / \left(\sum_{i=1}^{N} \hat{s}_i^2\right)$.

- Synthetic residual:

$$\hat{\xi}_t(h) = \frac{\sum_{i=1}^{N} \hat{s}_i \hat{\xi}_{it}(h)}{\sum_{i=1}^{N} \hat{s}_i^2} = \hat{Y}_t(h) - \left(\hat{\beta}(h)\hat{X}_t + \sum_{\ell=1}^{p} \tilde{\varphi}_\ell(h) X_{t-\ell} + \tilde{\mu}(h)\right).$$

$\hat{C}_\alpha(h), \hat{\sigma}(h)$ numerically the same as synthetic time series-based CI/SE.

Macro-micro decomposition

- Representation of estimation error:

$$\hat{\beta}(h) = \underbrace{\frac{\sum_{i=1}^{N} \hat{s}_i \beta_{ih}}{\sum_{i=1}^{N} \hat{s}_i^2}}_{\beta(h)+o_p\left(N^{-1/2}\right)} + \frac{\sum_{t=p+1}^{T-h} X_t \xi_t(h)}{E\left[X_t^2\right]} + o_p\left(T^{-1/2}\right)$$

where

$$\xi_t(h) = \left(\sum_{\ell \notin [h,h+p]} \tilde{\beta}_\ell X_{t+h-\ell} + \sum_{\ell=0}^{\infty} \tilde{\gamma}_\ell Z_{t+h-\ell} \right) + \frac{\kappa}{\sqrt{N}} \left(\frac{\sum_{i=1}^{N} \hat{s}_i \sum_{\ell=0}^{\infty} \delta_{i\ell} u_{i,t+h-\ell}}{\sqrt{N} \mathrm{Var}(s_i)} \right)$$

- Nature of estimation error depends on κ. Micro noise non-negligible if $\kappa \propto \sqrt{N}$.

Which inference procedures work?

- Regression score is MA(h) with $\text{Cov}(X_t\xi_t(h), X_{t-\ell}\xi_{t-\ell}(h)) = 0$ if $1 \leq |\ell| \leq p$.
- Moreover, only $\text{Var}(X_t\xi_t(h))$ depends on $\kappa/\sqrt{N}$.
- Whether a CI works hinges on whether it captures score's sum of autocovariances.
  - Unit-level clustering neglects cross-sectional dependence induced by macro shocks.
  - Driscoll-Kraay is OK in theory. Tricky in practice (kernel + difficulties with HAC).
    - Also estimates a lot of unnecessary autocovariances.
  - Two-way clustering will have some distortion, but not too bad if κ is large.
    - Also unit-level clustering part is redundant.
- One issue with $\hat{\sigma}(h)$ is that it runs into problems if h or p are large.

Heterogeneous VAR model

- Heterogeneous VAR DGP:

$$Y_{it} = m_i + \sum_{\ell=1}^{p} A_{i\ell} Y_{i,t-\ell} + \sum_{\ell=0}^{p} B_{i\ell} X_{t-\ell} + C_{i0} Z_t + \kappa D_{i0} u_{it}.$$

- Local projection augmented with p lags of Y_{it} and $s_i X_t$.

$$Y_{i,t+h} = \hat{\beta}(h) s_i X_t + \sum_{\ell=1}^{p} \left(\hat{\psi}_{i\ell}(h) Y_{i,t-\ell} + \hat{\varphi}_\ell(h) s_i X_{t-\ell} \right) + \hat{\mu}_i(h) + \hat{\nu}_t(h) + \hat{\xi}_{it}(h).$$

- Time-level aggregation of $\hat{\xi}_{it}(h)$ + Eicker–Huber–White works.
  - Dimension reduction when a low-order VAR ($p \ll h$) offers a good approximation.
  - As in Montiel-Olea, Plagborg-Møller (2021), but lags serve another purpose.

Simulation evidence: $T = 100$, $N = 1000$

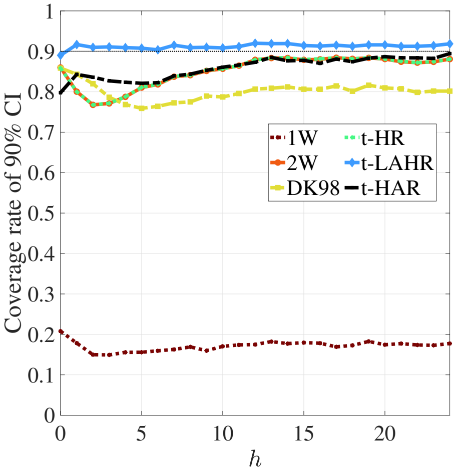


FIGURE. $\bar{R}^2 = 0.99$

Simulation evidence: $T = 100$, $N = 1000$

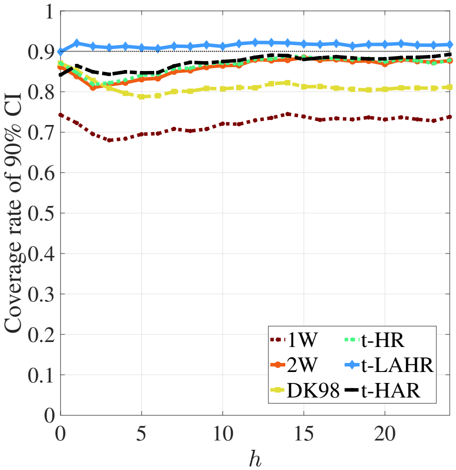


FIGURE. $\bar{R}^2 = 0.66$

Simulation evidence: $T = 100$, $N = 1000$

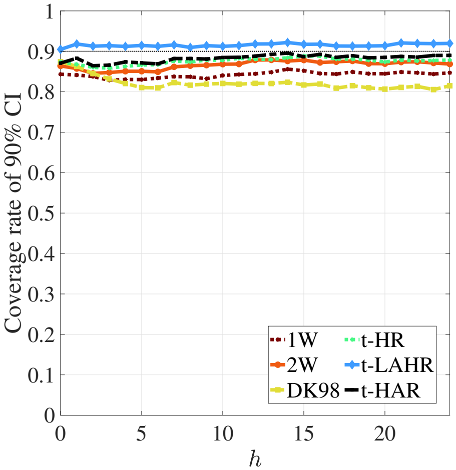


FIGURE. $\bar{R}^2 = 0.33$

Simulation evidence: $T = 30$, $N = 1000$

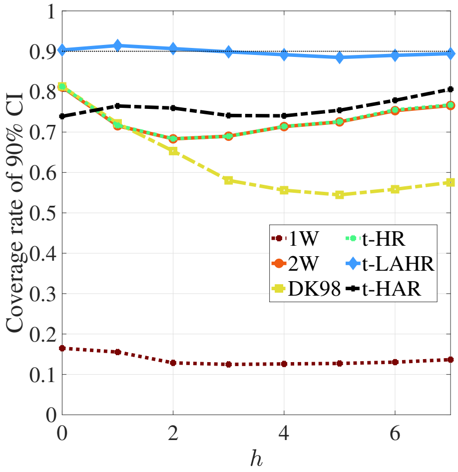


FIGURE. $\bar{R}^2 = 0.99$

Simulation evidence: $T = 30$, $N = 1000$

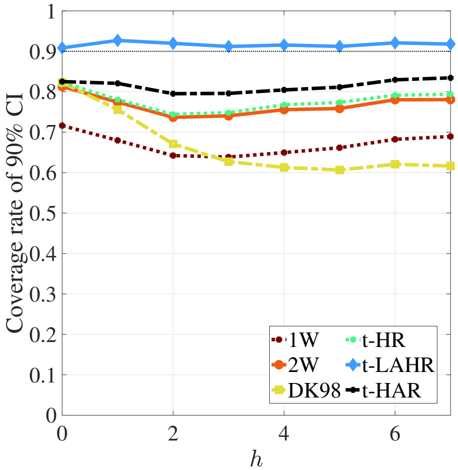


FIGURE. $\bar{R}^2 = 0.66$

Simulation evidence: $T = 30$, $N = 1000$

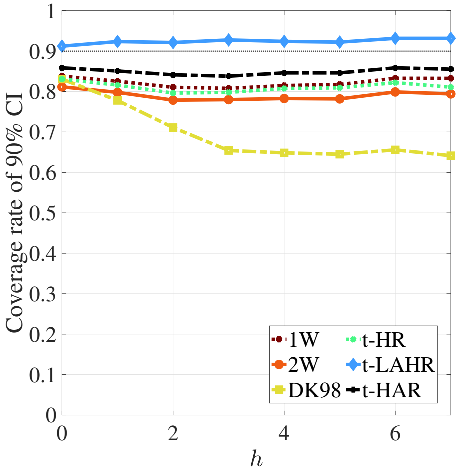
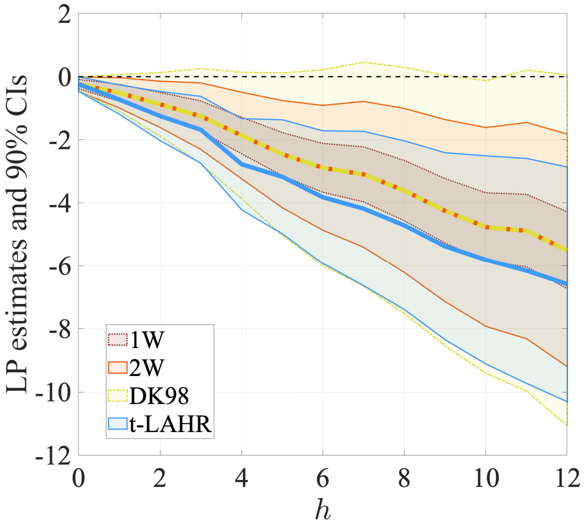


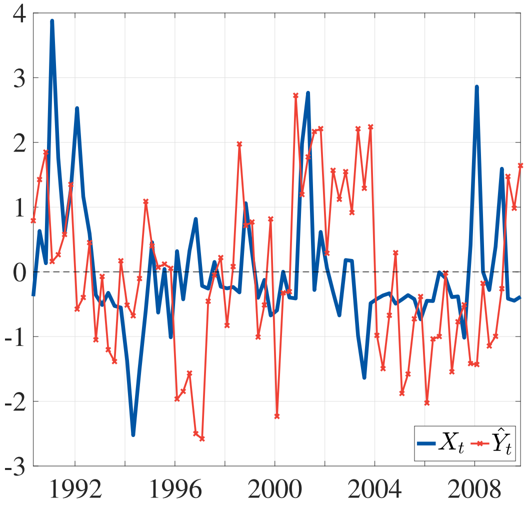
FIGURE. $\bar{R}^2 = 0.33$

Empirical illustration

Empirical illustration: confidence intervals



Empirical illustration: synthetic time series



Conclusion

Conclusion and practical recommendations

- Explosion of empirical work using panel local projections with aggregate shocks.
- Estimand under unrestricted heterogeneity $=$ population regression.
- Simple inference:
  - Time-level aggregation of residuals + lags + heteroskedasticity-robust SE.
  - Easier to refine in small samples.
  - Remains tractable over moderate horizons if a low-order VAR is reasonable.
  - Seems to perform better in low-signal environments.
- We also study the validity of popular inferential choices.
  - Unit-level clustering is either wrong (one-way) or unnecessary (two-way).
- Extensions.

Thank you!