A nonparametric dynamic causal model for macroeconometrics

Ashesh Rambachan[1] Neil Shephard[2]

March 27, 2019

[1]Department of Economics, Harvard University

[2]Department of Economics and Department of Statistics, Harvard University

Overview

- **Dynamic causal effects** – pervasive throughout macroeconomics.
  - Example: FOMC raises the target FFR by "surprise" 25 bps. How will this affect unemployment over the next several quarters?

- **Goal:** Common set of foundational assumptions for causal macroeconometrics methods.

- **How:** Take potential outcome time series framework of Bojinov & Shephard (2018) built for experiments on time series
  - Extend to observational data and specialize to where treatments are "shocks:" **shocked potential outcome time series.**

- **Results:** Provide *sufficient conditions* for causal inference in time series, introduce *non-parametric estimator* of dynamic causal effects & provide *causal interpretations* to common strategies in macroeconomics: SVAR, LP, LP-IV.

Potential outcome time series

- **Single unit** for $t = 1, \dots, T$. Treatment: $W_t \in \mathcal{W}$; outcome: Y_t. Both observed.
- k-th potential outcome at time t is $Y_{k,t}(w_{1:T})$
  - Treatment path $w_{1:T} = (w_1, \dots, w_T)$ & vector of outcomes $Y_t(w_{1:T})$.
- Assume
  1. **Non-anticipating POs**: $Y_t(w_{1:t}, w_{t+1:T}) = Y_t(w_{1:t}, w'_{t+1:T})$ for all $w_{1:T}, w'_{t+1:T}$. Write:

$$Y_t(\bullet) = \{Y_t(w_{1:t}) : w_{1:t} \in \mathcal{W}^t\}, \quad \text{and} \quad Y_{1:t}(\bullet) = \{Y_1(\bullet), ..., Y_t(\bullet)\}.$$

  2. **Non-anticipating treatments**: For each t, w_t:

$$P(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(W_{1:t-1}), Y_{t:T}(W_{1:t-1}, \bullet)) = P(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(W_{1:t-1})).$$

  3. **Data**: Observe $w_{1:T}^{obs}$ & $y_{1:T}^{obs} = y_{1:T}(w_{1:T}^{obs})$.
- Time series of treatments and potential outcomes that satisfy Assumptions 1, 2 and 3 are potential outcome time series (POTS).

Time-t causal effects

- **Causal effect**: comparisons of potential outcomes at a fixed time along different treatment paths.
- Time-t causal effect on Y_t for treatment path $w_{1:t}$ and counterfactual path $w'_{1:t}$ is

$$\tau_t(w, w') = Y_t(w_{1:t}) - Y_t(w'_{1:t}).$$

- Fix treatment path up to period $t-p$ and vary the j-th treatment in period $t-p$

$$w_{1:t} = (w_{1:t-p-1}, w_{1,t-p}, \dots, w_{j-1,t-p}, w, w_{j+1,t-p}, \dots, w_{n_w,t-p}, w_{t-p+1:t}),$$
$$w'_{1:t} = (w_{1:t-p-1}, w_{1,t-p}, \dots, w_{j-1,t-p}, w', w_{j+1,t-p}, \dots, w_{n_w,t-p}, w^*_{t-p+1:t}).$$

  k, j-th lag-p time-t causal effect is

$$\tau_{k,j,t}(w, w')(p) = Y_{k,t}(w_{1:t}) - Y_{k,t}(w'_{1:t}). \tag{1}$$

- If $w_{t-p+1:t} = w^*_{t-p+1:t}$, call this k, j-th lag-p time-t impulse causal effect, denoted $\tau^*_{k,j,t}(w, w')(p)$.

Causal equivalence

- $\{Y_{1:T}(\bullet), X_{1:T}(\bullet)\}$ are POTS. If for every $w_{1:t}$, $w'_{1:t}$,

$$Y_t(w_{1:t}) - Y_t(w'_{1:t}) \stackrel{as}{=} X_t(w_{1:t}) - X_t(w'_{1:t}),$$

  then $Y_{1:t}(\bullet)$, $X_{1:t}(\bullet)$ are time-t causally equivalent.

- If $Y_{1:T}(\bullet), X_{1:T}(\bullet)$ are causally equivalent,

$$Y_t(w_{1:t}) = X_t(w_{1:t}) + U_t$$

  where U_t is a *causal nuisance*.
  - U_t does not vary with $w_{1:t}$ – it produces different time series properties but not different time-t causal effects.
  - Nonparametric version of "counterfactual equivalence" in Beraja (2018).

Additive time-t causal effects

- New assumptions to connect potential outcome time series to existing literature.
- For a potential outcome time series, if

$$\tau_t(w, w') = \sum_{s=0}^{t-1} \beta_{s,t}(w_{t-s} - w'_{t-s}), \tag{2}$$

where $\beta_{s,t}$ non-random, $n_y \times n_w$ matrices then time-t causal effects are additive. $\beta_{s,t}$ called "*causal coefficients*."
  - If $\beta_{s,t} = \beta_s$, then causal effects are l-additive.
- **Proposition**:
  1. Under additive causal effects, the impulse causal effect simplifies to $\tau^*_{k,j,t}(w, w')(p) = \beta_{p,k,j,t}(w - w')$ – similar result for the k, j-th lag-p causal effect.
  2. If $Y_{1:T}(\bullet), X_{1:T}(\bullet)$ have l-additive causal effects, they are causally equivalent if and only if the causal coefficients match with $\beta^Y_{0:T-1} = \beta^X_{0:T-1}$.

Time-t weighted causal effects

- **Now**: Define additional causal estimands by building upon the k, j-th lag-p time-t causal effects.
  - Fix observed path in the treatment and counterfactual:

$$\tau_{k,j,t}(w, w')(p) = Y_{k,t}(w^{obs}_{1:t-p-1}, w_{t-p:t}) - Y_{k,t}(w^{obs}_{1:t-p-1}, w'_{t-p:t})$$

    and use weights to average over possible future paths $w_{t-p:t}$.

- The weighted causal effect $\bar{\tau}_{k,j,t}(w, w')(p)$ is

$$\mathbb{E}^{W|Y(\bullet)}[Y_{k,t} \mid w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p} = w] - \mathbb{E}^{W|Y(\bullet)}[Y_{k,t} \mid w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p} = w'],$$

  $\mathbb{E}^{W|Y(\bullet)}[\cdot]$ denotes expectation w.r.t. $W_{-j,t-p:t} \mid \{w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p}, Y_{t-p:t}(\bullet)\}$.

  - Averages over the random treatment path conditional on the past history of treatments, fixing the potential outcomes.

Time-t Causal response function

- **Now**: Additionally view $Y_{t-p:t}(\bullet)$ as random.
  - Average weighted causal effect over $Y_{t-p:t}(\bullet)$ by applying iterated expectations.
- The causal response function is

$$CRF_{k,j,t}(w, w')(p) = \mathbb{E}^{W,Y(\bullet)}[\bar{\tau}_{k,j,t}(w, w')(p) \mid w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}].$$

  - *Different from the IRF* – it conditions on the observed outcome and treatment path.
    CRF vs. IRF
  - Think of this as the time-series analogue of a *"super-population" causal effect*.

Non-parametric estimation of dynamic causal effects

- Large literature attempts to construct measures of shock of interest.
  - Narrative approach: Romer & Romer (1989), Romer & Romer (2010); High frequency movements around FOMC announcments: Rudebusch (1998) & Kuttner (2001); Oil shocks: Hamilton (2003) & Killian (2008).
- Given W_t, common approach: directly regress $Y_{k,t}$ on $W_{j,t-p}$ at a variety of lags – "Local projections" (LP) (Jordá 2005).
  - FOMC announcements: Cochrane & Piazessi (2002), Faust et al. (2003), Bernanke & Kuttner (2005), Gurkaynak et al. (2005); Oil shocks: Hamilton (2003) & Killian (2008).
- **Now**: Develop non-parametric estimators of weighted causal effect $\bar{\tau}_{k,j,t}(w, w')(p)$, its average $\bar{\tau}_{k,j}(w, w')(p)$ and provide tools for inference.
  - As notation, $\mathcal{F}_{T,t}$ denotes history $\{w_{1:t}^{obs}, y_{1:t}^{obs}, Y_{t+1:T}(\bullet)\}$ and $p_{j,t}(w) \equiv \mathbb{P}(W_{j,t} = w | \mathcal{F}_{T,t-1})$.
  - First assume support $\mathcal{W}$ discrete and then deal with continuous treatment case.

Non-parametric estimator of weighted causal effect

- Probabilistic treatment: For all $t \geq 1$, $\mathcal{F}_{T,t-1}$ & $w \in \mathcal{W}$, $p_{j,t}(w) > 0$.
  - Analogue of "overlap" assumption made in cross-sectional causal inference.

- Define Horvitz-Thompson style estimator

$$\hat{\hat{\tau}}_{k,j}(w, w')(p) \equiv \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\hat{\tau}}_{k,j,t}(w, w')(p),$$

$$\hat{\hat{\tau}}_{k,j,t}(w, w')(p) \equiv \frac{y_{k,t}^{obs} \left\{ \mathbb{1}(w_{j,t}^{obs} = w) - \mathbb{1}(w_{j,t-p}^{obs} = w') \right\}}{p_{j,t-p}(w_{j,t-p}^{obs})}.$$

  - Similar estimator appears in Angrist et al. (2018) – introduced for a super-population estimand that is most similar to our CRF.

Properties of Horvitz-Thompson estimator

- **Theorem**: Assume POTS, probabilistic treatment & define error $u_{k,j,t-p} := \hat{\bar{\tau}}_{k,j,t}(w, w')(p) - \bar{\tau}_{k,j,t}(w, w')(p)$. Over the law of the non-anticipating treatment path,

$$\mathbb{E}^{W \mid Y(\bullet)}_{T,t-p-1}[u_{k,j,t-p} \mid \mathcal{F}_{T,t-p-1}] = 0, \quad \text{and} \quad \mathbb{E}^{W \mid Y(\bullet)}[\hat{\bar{\tau}}_{k,j}(w, w')(p)] = \bar{\tau}_{k,j}(w, w')(p).$$

Furthermore, $\eta^2_{k,j,t-p} \equiv Var^{W|Y(\bullet)}[u_{k,j,t-p}|\mathcal{F}_{T,t-p-1}]$ exists.

- **Theorem (cont.)** If the non-anticipating treatment paths satisfy $\lim_{T\to\infty} \frac{1}{T-p}\sum_{t=p+1}^{T} \mathbb{E}^{W \mid Y(\bullet)}[\eta^2_{k,j,t-p}] < \infty$, then $\hat{\bar{\tau}}_{k,j}(p) - \bar{\tau}_{k,j}(p) \xrightarrow{p} 0$ as $T \to \infty$. Finally, if $\frac{1}{T-p}\sum_{t=p+1}^{T} \eta^2_{k,j,t-p} \xrightarrow{p} \eta^2_{k,j} > 0$, then

$$\sqrt{T}\frac{\{\hat{\bar{\tau}}_{k,j}(p) - \bar{\tau}_{k,j}(p)\}}{\eta_{k,j}} \xrightarrow{d} N(0,1).$$

Properties of Horvitz-Thompson estimator

- Series of important remarks:
  1. Cannot compute $\eta^2_{k,j,t-p}$ in general as some potential outcomes are unobserved. We provide *upper bound* for $\eta^2_{k,j,t-p}$ in proof of Theorem. Then construct an unbiased estimator for the upper bound

$$\hat{\hat{\eta}}^2_{k,j,t-p}(w,w')(p) = \frac{(y^{obs}_{k,t})^2 \left\{\mathbb{1}(w^{obs}_{j,t} = w) + \mathbb{1}(w^{obs}_{j,t-p} = w')\right\}}{p_{j,t-p}(w^{obs}_{j,t-p})^2}.$$

  2. Proxy outcomes are useful to reduce the variance of $\hat{\hat{\tau}}_{k,j,t}(w,w')(p)$ – produces *doubly robust estimator*. E.g.

$$\hat{\hat{\tau}}^{dr}_{k,j,t}(w,w')(p) = \frac{(y^{obs}_{k,t} - y^{obs}_{k,t-p-1})\left\{\mathbb{1}(w^{obs}_{j,t} = w) - \mathbb{1}(w^{obs}_{j,t-p} = w')\right\}}{p_{j,t-p}(w^{obs}_{j,t-p})}$$

  3. Can extend results to provide an unbiased, non-parametric estimator of the CRF.

Hypothesis tests: Sharp and Neyman nulls

- Sharp null: $H_0 :: \; Y_{k,t}(w_{1:t}) = Y_{k,t}(w'_{1:t})$ for all $w_{1:t}, w'_{1:t}$.
  - Under sharp null, all potential outcome paths are known and can simulate randomization distribution of estimator if $p_{j,t}(w)$ is known.
  - Enables researcher to compute exact p-value of observed test statistic and construct exact confidence intervals through test inversion.

- Neyman null: $H_0 \; : \; \bar{\tau}_{k,j}(w, w')(p) = 0$.
  - Restricts the average temporal causal effect to be zero.
  - Con construct tests using the CLT in Theorem. Under H_0,

$$\frac{\sqrt{T}\hat{\bar{\tau}}_{k,j}(w, w')(p)}{\bar{\eta}_{k,j}(w, w')(p)} \xrightarrow{d} N(0, 1)$$

    as $T \to \infty$. Using the estimator of upper-bound of the variance, use this to construct a conservative test.

Extension to continuous treatments

- Now suppose treatments are continuously valued. For bandwidth $h > 0$, consider estimator

$$\hat{\hat{\tau}}_{k,j,t}(w, w') \equiv \frac{y_{k,t}^{obs}\left(\mathbb{1}(w_{j,t-p}^{obs} \in [w-h, w+h]) - \mathbb{1}(w_{j,t-p}^{obs} \in [w'-h, w'+h])\right)}{F_{j,t-p}(w_{j,t-p}^{obs} + h) - F_{j,t-p}(w_{j,t-p}^{obs} - h)}.$$

  - Difference of two kernel estimators. Non-parametrics only in $W_{j,t-p} \implies$ no curse of dimensionality.
  - Show that rate of convergence falls from $T^{-}1$ in discrete case to $T^{-4/5}$ in continuous. Also derive bias, variance & optimal selection of h to minimize mean square error under additional smoothness restrictions on potential outcomes.

Frisch-Slutzky paradigm and treatments as shocks

- In time-series macroeconometrics, treatments are assumed to be *unpredictable* given past – referred to as **shocks**.
  - This is a stronger assumption than the non-anticipation assumption introduced earlier.
  - We now nest this work into the general POTS framework.

- For a POTS, the time-t treatment is a shock if it is non-anticipating and satisfies

$$\mathbb{E}[W_t \mid W_{1:t-1}, Y_{1:t-1}] = 0, \text{ and } Var(W_t | W_{1:t-1}, Y_{1:t-1}) = diag\{\sigma^2_{1,t}, \ldots, \sigma^2_{n_w,t}\}.$$

  Formalizes interpretation provided in Ramey (2016), Stock & Watson (2016):
  - Shocks are unpredictable – mds w.r.t. filtration produced by POTS.
  - Shocks are mutually uncorrelated – not necessary but simplifies many results.

- A POTS whose treatments are shocks is a shocked potential outcome time series.

Shocks and causal effects

- Return to causal estimands introduced earlier. Enormous simplification occurs under additive causal effects + shocks.

- **Theorem**: Consider shocked POTS with additive causal effects. Then, for $t \geq p+1$,

$$CRF_{k,j,t}(w, w')(p) = \bar{\tau}_{k,j,t}(w, w') = \tau^{*}_{k,j,t}(w, w)(p) = \beta_{p,k,j,t}(w - w').$$

  Additionally, if $\{Y_{k,t}(W_{1:t}), W_{j,t}\}$ is a jointly stationary stochastic process and the causal effects are I-additive, then

$$IRF_{k,j}(w, w')(p) = \beta_{p,k,j}(w - w').$$

  Proof

  - The choice of causal estimand *no longer matters* as they identify the same objects – the causal coefficients.

Interpreting the SVMA

- How do we get to linear versions of shocked POTS?

$$Y_t(w_{1:t}) = U_t^Y + \sum_{s=0}^{t-1} \Theta_{s,t} w_s.$$

(1) $Y_t(w_{1:t})$ is differentiable w.r.t. $w_{1:t}$ & taylor expand around $\bar{w}_{1:t}$; (2) Directly assume time-t additive causal effects

- Assume shocked POTS, $\Theta_0, \dots, \Theta_{t-1}$ non-stoch. matrices, $dim(Y_t) = dim(W_t)$, and

$$Y_t(w_{1:t}) = \mu + \sum_{s=0}^{t-1} \Theta_s w_{t-s} \quad \forall w_{1:t} \in \mathbb{W}^t.$$

Then, $Y_{1:t}(\bullet)$ is structural vector moving average.

  - SVMA has I-additive causal effects.
  - SVMA causally equivalent to POTS with I-additive causal effects.

Interpreting local projection

- Given W_t, then just directly regress $Y_{k,t}$ on $W_{j,t-p}$ at a variety of lags – "local projections" (LP) (Jordá 2005).
  - Under what conditions does LP identify a dynamic causal effect?

- **Theorem**: Consider a shocked POTS and assume that for $p \geq 0$:
  A. The causal effects are additive.
  B. $\mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t}] = \alpha$, where $\bar{w}_{1:t} = 0$ is fixed.
  C. $Var^W(W_t) < \infty$, $Var^Y(Y_{k,t}(\bar{w}_{1:t}) < \infty$ for all t.

  Then, for $p \geq 0$, the causal regression is

$$Y_{k,t} = \alpha + \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}, \quad t = p+1, 2, \ldots, T,$$

  where $\beta_{p,k,j,t}$ is the causal coefficient and $\mathbb{E}^{W,Y}[\eta_{k,j,t}] = 0$, $Cov^{W,Y}(W_{j,t-p}, \eta_{k,j,t}) = 0$ if they exist. Proof
  - Error $\eta_{k,j,t}$ may depend on all of $W_{1:t}$ except $W_{j,t-p}$. It is not a mds or white noise in general.
  - Time series generalization of result for when linear regression identifies the ATE.

LP under I-additivity

- If additionally the causal effects are I-additive, causal regression becomes

$$Y_{k,t}(W_{1:t}) = \alpha + \beta_{p,k,j} W_{j,t-p} + \eta_{k,j,t} \quad t = p+1, 2, \ldots, T. \tag{3}$$

- **Remarks**:
  i. Can *estimate* causal coefficient using *OLS*. Since, $\eta_{k,j,t} \perp W_{j,t-p}$, this is consistent and inference can be based on standard CLT arguments.
     - May require HAR techniques as $\eta_{k,j,t}$ may be serially correlated.

  ii. If $X_{k,j,t}$ is a control correlated $\eta_{k,j,t}$ but unrelated to $W_{j,t-p}$,

$$Y_{k,t}(W_{1:t}) = \alpha + \beta_{p,k,j} W_{j,t-p} + \gamma'_{p,k,j} X_{k,j,t} + \eta^*_{k,j,t},$$

  may produce a more efficient estimator, where $\gamma_{p,k,j}$ is a nuisance parameter.

  iii. If causal effects are *additive*, OLS identifies *time average* of the causal coefficients. If causal effects are *not* additive, OLS identifies *best linear approx.* to time-t causal effect.

Local projection with an instrument – review

- Some econometricians worry constructed measures are noisy proxies of shock.
  - E.g. changes in FFR future rates around FOMC announcements only describes part of the monetary policy shock.
- Recent literature develops inference strategy that treats constructed measure $\hat{W}_{j,t-p}$ as instrument for shock – referred to as LP-IV.
  - Proposed in Jordà et al. (2015) – use two noisy measures of shock $(Y_{j,t-p}, \hat{W}_{j,t-p})$ in an IV estimator.
  - See Stock & Watson (2018), Plagborg-Moller & Wolf (2018), Fieldhouse et al. (2018) for discussion of LP-IV.
- Place LP-IV in POTS and provide conditions under which it identifies a causal effect.

Potential outcome time series with added proxies

- Let $\hat{W}_{1:T}$ be a proxy for the treatment $W_{1:T}$.
- **Proxy potential outcome time series**:

A. *Exclusion restriction*: For all $w_{1:T}, \hat{w}_{1:T}, \hat{w}'_{1:T}$,

$$Y_{1:T}(w_{1:T}, \hat{w}_{1:T}) = Y_{1:T}(w_{1:T}, \hat{w}'_{1:T}).$$

B. *Non-anticipating treatment*:

$$W_t | W_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}(W_{1:t-1}), Y_{t:T}(W_{1:t-1}, \bullet) \sim W_t | W_{1:t-1}, Y_{1:t-1}(W_{1:t-1})$$

C. *Non-anticipating proxy*:

$$\hat{W}_t | W_{1:t}, \hat{W}_{1:t-1}, Y_{1:t-1}(W_{1:t-1}), Y_{1:T}(W_{1:t-1}, \bullet) \sim \hat{W}_t | W_{1:t}, \hat{W}_{1:t-1}, Y_{1:t-1}(W_{1:t-1}).$$

LP-IV under additive causality

- **Theorem**: Consider a proxy potential outcome time series and assume that
  A. The causal effects are additive.
  B. $(W_t, \hat{W}_t)$ is a joint mds with respect to $W_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}$.
  C . $\mathbb{E}^{W,\hat{W}}[W_t \hat{W}_s]$ and $\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}]$ exists for any fixed counterfactual path $w_{1:t}$.
  D. $\mathbb{E}^{W,\hat{W}}[W_{i,t}\hat{W}_{j,t}] = 1_{i=j}\gamma_j$ with $\gamma_j \neq 0$.

  Then, for $p \geq 0$, $\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}] = \beta_{p,k,j,t}\gamma_j$. Proof

- Therefore, the *Wald estimand* is

$$\frac{\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}}{\mathbb{E}^{Y,W,\hat{W}}[Y_{j,t-p}(W_{1:t-p})\hat{W}_{j,t-p}]} = \frac{\beta_{p,k,j,t}}{\beta_{0,j,j,t}}.$$

  Imposing the unit effect normalization $\beta_{0,j,j,t} = 1$ then delivers the causal coefficient (e.g. Stock & Watson 2018).

LP-IV under additive causality

- **Remarks**:
  1. Strengthen to I-additivity and Theorem 3 implies
  $$\frac{\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}]}{\mathbb{E}^{Y,W,\hat{W}}[Y_{j,t-p}(W_{1:t-p})\hat{W}_{j,t-p}]} = \frac{\beta_{p,k,j}}{\beta_{0,j,j}}.$$
  Use IV estimator and analyze its properties with conventional asymptotic estimators. May require HAR and/or weak IV robust techniques.
  2. Lit. assumes proxy satisfies *lead-lag exogeneity*. This is not sufficient – does not imply $\mathbb{E}^{Y,\hat{W}}[Y_t(\bar{w}_{1:t})\hat{W}_{t-p}] = 0$.
     - Restricts joint distribution of unobserved counter-factual and proxy.
     - Implied by proxy POTS assumptions.
  3. If causal effects are *additive*, IV estimator identifies time average of causal coefficients. Not a LATE result – still restricting relationship between proxy and treatment to be time invariant.

Empirical application: Romer & Romer shocks

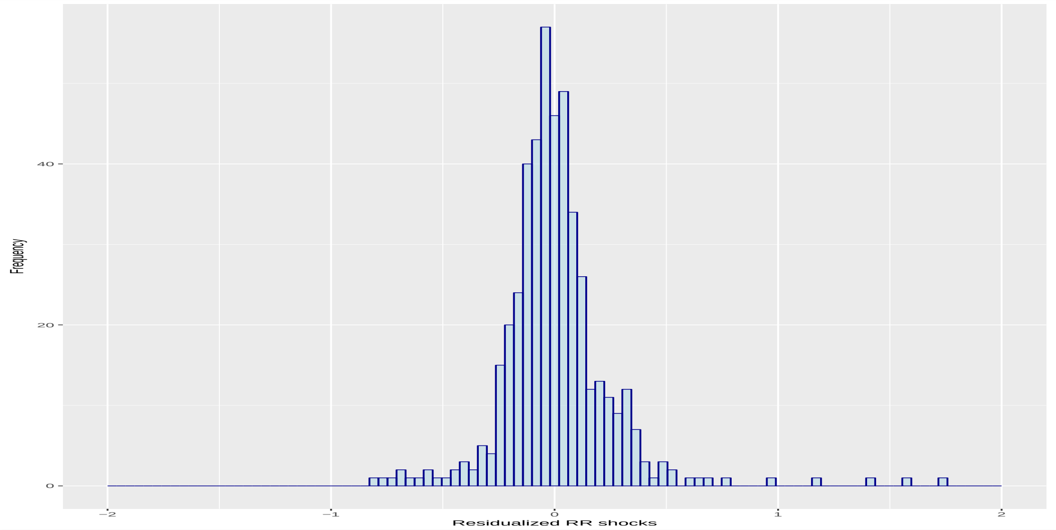
- **Now**: Compare nonparametric estimator with local projection approach in simple, well-known example.
  - Replicate analysis in Ramey (2016) on real effects of monetary policy.
  - Use shock series from Romer & Romer (2004) as the observed treatments W_t and maintain identifying assumptions as in Ramey (2016).

- **Local projection**: For $p = 1, \ldots, 24$, estimate

$$Y_{k,t} = \alpha_{p,k} + \beta_{p,k} W_{t-p} + X'_{t-p} \lambda_{p,k} + \epsilon_{p,k,t}$$

  - $X_{k,t}$ is a vector of additional controls: two lags of RR shocks, log of IP, log of CPI, UR, FFR and Commodity Price Index.
  - Estimate over monthly data from January 1969 – December 2007.

Empirical application: Distribution of residualized RR shocks



Empirical application: Romer & Romer shocks

- **Non-parametric estimator**: $\hat{\hat{\tau}}_k(p)$ is

$$\frac{1}{(T-p)}\sum_{t=p+1}^{T}\frac{(Y_{k,t}-Y_{k,t-p-1})\left(\mathbb{1}(\tilde{W}_{t-p}\in[w-h,w+h])-\mathbb{1}(\tilde{W}_{t-p}\in[w'-h,w'+h])\right)}{F_{t-p}(\tilde{W}_{t-p}+h)-F_{t-p}(\tilde{W}_{t-p}-h)}.$$

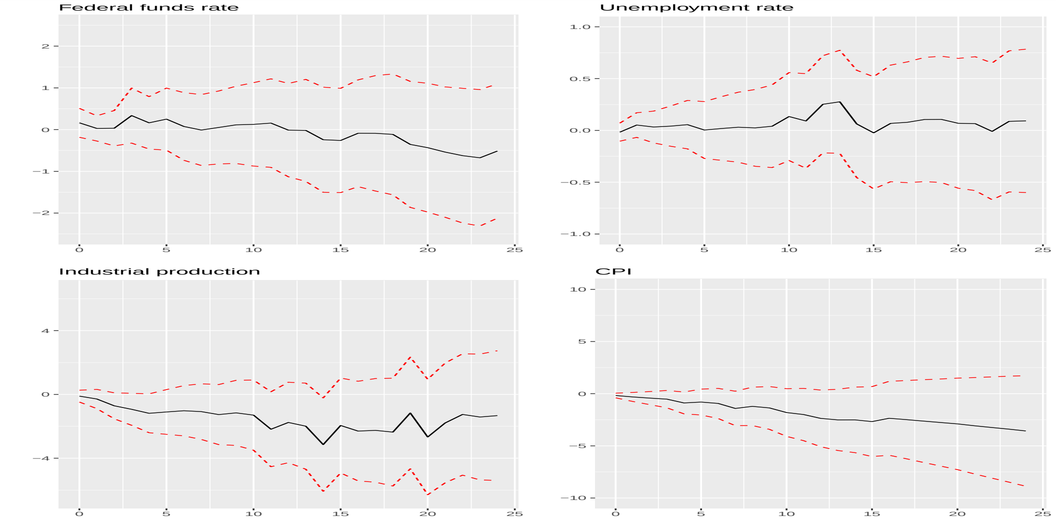
  - Select $w=0.1, w'=-0.1$. If causal effects are l-additive, expect estimates to be 20% of LP estimates.
  - Follow Hirano & Imbens (2004) – estimate F_{t-p} using MLE where we assume $W_{t-p}\sim N(\gamma,\sigma^2)$.
  - Choose bandwidth h to minimize MSE.

Use conservative estimator of variance $h^{-1}v_k^2(p)/(T-p)$, where $v_k^2(p)$ is

$$\frac{1}{(T-p)}\sum_{t=p+1}^{T}\frac{(Y_{k,t}-Y_{k,t-p-1})^2\left(\mathbb{1}(\tilde{W}_{t-p}\in[w-h,w+h])+\mathbb{1}(\tilde{W}_{t-p}\in[w'-h,w'+h])\right)}{f_{t-p}(\tilde{W}_{t-p})^2}.$$

Empirical application: Local projection estimates

Empirical application: Non-parametric estimates



Structural vector autoregression

- Alternative approach – use a model that a prior links outcomes to unobserved treatments.
- Assume $Y_{1:t}(\bullet)$ is a POTS. If

$$\Gamma(L)_t(Y_t(w_{1:t}) - \mu) = w_t, \text{ where } \Gamma(L)_t = \Gamma_0 - \Gamma_1 L - \ldots - \Gamma_{t-1} L^{t-1}, \tag{4}$$

  we call $Y_{1:t}(\bullet)$ a structural vector autoregression.

- Assume $Y_{1:t}(\bullet)$ is a POTS. If

$$A(L)_t\{Y_t(w_{1:t}) - \mu\} = \Theta_0 w_t, \quad \text{where} \quad A(L)_t = I - A_1 L - \ldots - A_{t-1} L^{t-1}, \tag{5}$$

  we call $Y_{1:t}(\bullet)$ a vector autoregression.

- **Note**: We do not assume W_t is a shock but we restrict $n_w = n_y$.

SVAR and I-additivity

- **Proposition**: Assume $Y_{1:t}(\bullet)$ is a POTS with SVAR form, where Γ_0 is invertible. Then,
  1. $Y_{1:t}(\bullet)$ can be written as a VAR with $A_s = \Gamma_0^{-1}\Gamma_s$ while $\Theta_0 = \Gamma_0^{-1}$.
  2. The time-t causal effects of Y_t follow $A(L)_t\{Y_t(w_{1:t}) - Y_t(w'_{1:t})\} = \Theta_0(w_t - w'_t)$. The time-$t$ lag-p causal coefficients follow the recursion

$$\beta_p = A_1\beta_{p-1} + ... + A_p\beta_0, \quad \beta_0 = \Theta_0.$$

- **Proposition**: Assume a POTS $Y_{1:t}$ has I-additive causal effects and β_0 is invertible. Then, the time-t causal effects follow a VAR with $A(L)_t\{Y_t(w_{1:t} - w'_{1:t})\} = \beta_0(w_t - w'_t)$, where $A_p = (\beta_p - A_1\beta_{p-1} - \ldots - A_{p-1}\beta_1)\beta_0^{-1}$. Proof

Causal equivalence and SVAR

- **Theorem**: A SVAR with invertible Γ_0 is causally equivalent to POTS with I-additive causal effects and invertible β_0.
  - Follows immediately from previous propositions and holds with no assumptions on the treatment path.
- This causal equivalence is a *foundational point* for existing macroeconometric work.
  - We argued that the shocked POTS is a modern version of the Frisch-Slutzky paradigm.
  - If additionally assume I-additivity and an invertible matrix of contemporaneous causal coefficients β_0, then the shocked POTS is causally equivalent to an SVAR with invertible Γ_0.

Conclusion

- Use a nonparametric, potential outcome time series framework for defining dynamic causal effects in macroeconomics. Defined several new concepts:
  - Time-t causal effects, weighted causal effects, causal response function,
  - Additive & I-additive causal effects,
  - Causal equivalence,
  - Shocked potential outcome time series.
- Introduce non-anticipation as sufficient condition for non-parametric estimation of dynamic causal effects.
  - Provided a Horvitz-Thompson style estimator for dynamic causal effects and analyzed its properties.
- Used potential outcome time series as common backbone to provide causal interpretation to existing estimation strategies in macroeconomics:
  - Local projections
  - Local projection-IV
  - Structural vector autoregressions

Is the IRF causal?

- Suppose $Y_{k,t} = Y_{k,t}(W_{1:t})$, $W_{j,t}$ joint stationary. Then,

$$IRF_{k,j}(w, w')(p) \equiv \mathbb{E}[Y_{k,t}|W_{j,t-p} = w] - \mathbb{E}[Y_{k,t}|W_{j,t-p} = w']$$

  is an impulse response function (IRF).
  - Without additional assumptions – *not causal*. Just the difference between two conditional expectations.

- **Theorem**: Assume POTS and $Y_{k,t}, W_{j,t}$ are jointly stationary. Then, if the expectations exist,

$$\mathbb{E}^{W_{1:t-p-1}, Y_{1:t-p-1}}[CRF_{k,j,t}(p)] = IRF_{k,j}(w, w')(p).$$

  - Application of iterated expectations.

Back

Proof: Shocks and causal effects

- **Lemma**: Assume POTS $Y_{1:t}(\bullet)$ with additive causal effects. Then,
  i. Weighted causal effect is
     $\bar{\tau}_{k,j,t}(w,w')(p) = \beta_{p,k,j,t}(w-w') + \sum_{s=1}^{p}\sum_{i=1}^{n_w} \beta_{s,k,i,t}\bar{V}_{s,i,j,t-p}(w,w')$, with
     $\bar{V}_{s,i,j,t}(w,w') = \mathbb{E}^{W\,|\,Y(\bullet)}[W_{i,t+s} \mid w_{1:t-1}^{obs}, W_{j,t} = w] - \mathbb{E}^{W\,|\,Y(\bullet)}[W_{i,t+s} \mid w_{1:t-1}^{obs}, W_{j,t} = w']$.
  1. Causal response function is
     $CRF_{k,j,t}(w,w')(p) = \beta_{p,k,j,t}(w-w') + \sum_{s=1}^{p}\sum_{i=1}^{n_w} \beta_{s,k,i,t}\dot{V}_{s,i,j,t-p}(w,w')$, with
     $\dot{V}_{s,i,j,t}(w,w') = \mathbb{E}[W_{i,t+s} \mid y_{1:t-1}^{obs}, w_{1:t-1}^{obs}, W_{j,t} = w] - \mathbb{E}[W_{i,t+s} \mid y_{1:t-1}^{obs}, w_{1:t-1}^{obs}, W_{j,t} = w']$.

- **Proof**: Recall the form of $\tau_{k,j,t}(w,w')(p)$ from earlier results on additive causal effects. Applying definitions of weighted causal effect and causal response function produces results.

- Apply martingale property of treatments – all $\dot{V}_{s,i,j,t}(w,w')$ and $\bar{V}_{s,i,j,t}(w,w')$ terms are zero. Back

Proof: LP under additive causality

- Consider fixed known counterfactual $\bar{w}_{1:t} = 0$. Then,

$$Y_t(W_{1:t}) = Y_t(\bar{w}_{1:t}) + \tau_t(W_{1:t}, \bar{w}_{1:t}).$$

- Assumption A of additive causal effects + Assumption B of existence of the mean of the counterfactual outcome $\implies$

$$Y_{k,t}(W_{1:t}) = \mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t})] + \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t},$$

where

$$\eta_{k,j,t} = Y_{k,t}(\bar{w}_{1:t}) - \mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t})] + \sum_{i=1; i \neq j}^{n_w} \beta_{p,k,i} W_{i,t-p}$$
$$+ \sum_{s=0; s \neq p}^{t-1} \sum_{i=1}^{n_w} \beta_{s,k,i,t} W_{i,t-s}.$$

Proof: LP under additive causality

- Shock property $\implies$ the treatments are martingale differences $\implies$ if it exists, $\mathbb{E}^{W,Y}(\eta_{k,j,t}) = 0$ for all k, j, t.

- Finally,

$$\begin{aligned}
&\mathbb{E}^{Y,W}(Y_t(\bar{w}_{1:t})W_{t-p}) \overset{(1)}{=} \mathbb{E}^{Y}\{Y_t(\bar{w}_{1:t})\mathbb{E}^{W|Y(\bullet)}(W_{t-p}|Y_t(\bar{w}_{1:t}), Y_{t-p-1}(\bar{w}_{1:t}))\}, \\
&\overset{(2)}{=} \mathbb{E}^{Y}\{Y_t(\bar{w}_{1:t})\mathbb{E}^{W|Y(\bullet)}(W_{t-p}|Y_{t-p-1}(\bar{w}_{1:t}))\}, \\
&\overset{(3)}{=} 0,
\end{aligned} \tag{6}$$

where (1) iterated expectations, (2) non-anticipating treatments and (3) MD assumption.

- Properties of shocks imply $Cov^{W,Y}(W_{j,t-p}, \eta_{k,j,t}) = 0$, if this covariance exists. Existence of the moments is guaranteed by Assumption C. Back

Proof: LP-IV under additive causality

- **Lemma**: Consider proxy POTS. If $W_t, \hat{W}_t$ is a joint mds w.r.t. $W_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}$, then, if they exist, the superpopulation moments obey:
  i. $\mathbb{E}^{W,\hat{W}}[W_t \hat{W}_s] = 0$, for all s, t where $s \neq t$.
  ii. $\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}] = 0$, for all $w_{1:t}$ and $p \geq 0$.

- **Proof**: (i) is follows from the martingale difference assumption. (ii) follows from

$$
\begin{aligned}
&\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}] \\
&\overset{(1)}{=} \mathbb{E}^{Y}\{Y_t(w_{1:t})\mathbb{E}^{\hat{W}|Y}[\hat{W}_{t-p}|Y_{1:t}(w_{1:t})]\} \\
&\overset{(1)}{=} \mathbb{E}^{Y}\{Y_t(w_{1:t})\mathbb{E}^{W_{1:t-p},\hat{W}_{1:t-p-1}|Y}\mathbb{E}^{\hat{W}|Y,W}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p}, Y_{1:t}(w_{1:t})]\} \\
&\overset{(2)}{=} \mathbb{E}^{Y}\{Y_t(w_{1:t})\mathbb{E}^{W_{1:t-p},\hat{W}_{1:t-p-1}|Y}\mathbb{E}^{\hat{W}|Y,W}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p}, Y_{1:t-p-1}(w_{1:t-p-1})]\} \\
&\overset{(3)}{=} \mathbb{E}^{Y}\{Y_t(w_{1:t})\mathbb{E}^{W_{1:t-p-1},\hat{W}_{1:t-p-1}|Y}\mathbb{E}^{\hat{W}|Y}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p-1}, Y_{1:t-p-1}(w_{1:t-p-1})]\} \\
&\overset{(4)}{=} 0,
\end{aligned}
$$

  where (1) iterated expectations, (2) non-anticipation of $\hat{W}_t$, (3) integrates out W_{t-p}, (4) martingale differences assumption.

Proof: LP-IV under additive causality

- Under additive causal effects, for any non-stochastic counter-factual $\bar{w}_{1:t}$,

$$Y_{k,t}(W_{1:t}) = \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}, \quad \text{where}$$
$$\eta_{k,j,t} = Y_{k,t}(\bar{w}_{1:t}) + \sum_{i=1;i\neq j}^{n_w} \beta_{p,k,i,t} W_{i,t-p} + \sum_{s=0;s\neq p}^{t-1} \sum_{i=1}^{n_w} \beta_{s,k,i,t} W_{i,t-s}.$$

Then, the result follows by checking off all the terms using Assumption D and using the results from Lemma. Back

SVAR and I-additivity: Proof – Proposition 1

- Claim (i) is immediate from SVAR-VAR definition.

- Next, claim (ii). From definition of the VAR,

$$A(L)_t\{y_t(w_{1:t}) - \mu\} - A(L)_t\{y_t(w'_{1:t}) - \mu\} = \Theta_0(w_t - w'_t).$$

The left hand side simplifies as μ cancels. We are left with

$$A(L)_t\{y_t(w_{1:t}) - y_t(w'_{1:t})\} = \Theta_0(w_t - w'_t)$$

as claimed. To be explicit about the solution, write $\tau_t = y_t(w_{1:t}) - y_t(w'_{1:t})$

$$\begin{aligned}
\tau_1 &= \Theta_0(w_1 - w'_1) \\
\tau_2 &= A_1\{y_1(w_1) - y_1(w'_1)\} + \Theta_0(w_2 - w'_2) = \Theta_0(w_2 - w'_2) + A_1\Theta_0(w_1 - w'_1) \\
\tau_3 &= A_1\{y_2(w_{1:2}) - y_2(w'_{1:2})\} + A_2\{y_1(w_1) - y_1(w'_1)\} + \Theta_0(w_3 - w'_3) \\
&= \Theta_0(w_3 - w'_3) + A_1\{\Theta_0(w_2 - w'_2) + A_1\Theta_0(w_1 - w'_1)\} + A_2\{\Theta_0(w_1 - w'_1)\} \\
&= \Theta_0(w_3 - w'_3) + A_1\beta_0(w_2 - w'_2) + (A_1\beta_1 + A_2\beta_0)(w_1 - w'_1).
\end{aligned}$$

SVAR and I-additivity: Proof – Proposition 2

- This follows from the recursion in previous Proposition. We now write

$$A_p\beta_0 - \beta_p = A_1\beta_{p-1} - \ldots - A_{p-1}\beta_1.$$

With β_0 invertible, we arrive at the result. Back