# When Do Common Time Series Estimands Have Non-parametric Causal Meaning?

Ashesh Rambachan[1] and Neil Shephard[1]

[1]Harvard University

RuhrMetrics Reading Group
Discussed by: Christian Schulz

November 14, 2024

# Motivation

- We should have read Bojinov and Shephard (2023) :)
- The paper that we actually read is basically a reference not “simply” a paper about causality in Macro.
- I focus on the general idea and the two main parts of the paper in my view:
  - We obeserve only the outcomes. Structure is needed – We will read Sims (1980) next, so it is a good primer on the VAR.
  - We observe outcomes and treatments. We will look at the VAR and LP estimands.
- This paper does not provide new solutions to identification problems but structures different ideas about identification in a unified framework that is well established in microeconometrics already.

# Motivation

**Stock and Watson (2018)** write on pg. 922: "The macroeconometric jargon for this random treatment is a 'structural shock:' a primitive, unanticipated economic force, or driving impulse, that is unforecastable and uncorrelated with other shocks. The macroeconomist's shock is the microeconomists' random treatment, and impulse response functions are the causal effects of those treatments on variables of interest over time, that is, dynamic causal effects."

**Ramey (2016))** writes on pg. 75, "the shocks should have the following characteristics: (1) they should be exogenous with respect to the other current and lagged endogenous variables in the model; (2) they should be uncorrelated with other exogenous shocks; otherwise, we cannot identify the unique causal effects of one exogenous shock relative to another; and (3) they should represent either unanticipated movements in exogenous variables or news about future movements in exogenous variables."

## Sims (1980) – VAR I

- Main challenge: causality without dynamic effects, illustrated with a static example.
- Assume:

$$A_0 Y_t(w_{1:t}) = a + w_t, \quad t = 1, 2, \ldots$$

  where:
  - $A_0$ is a non-stochastic, square matrix.
  - The potential outcome process is deterministic and allows linear combinations of potential outcomes.
- If $A_0$ is invertible, define:

$$Y_t(w_{1:t}) = A_0^{-1}(a + w_t),$$

  implying:

$$\mathbb{E}[Y_t(W_{1:t-1}, w) - Y_t(W_{1:t-1}, w')] = A_0^{-1}(w - w').$$

# Sims (1980) – VAR II

- With observed data $(W_t, Y_t) = (W_t, Y_t(w_{1:t}))$:

$$\mathrm{Cov}(Y_t, W_t)\,\mathrm{Var}(W_t)^{-1} = A_0^{-1}$$
$$\mathrm{Var}(Y_t) = A_0^{-1}\mathrm{Var}(W_t)\left(A_0^{-1}\right)^T$$

- To untangle $A_0$ and $\mathrm{Var}(W_t)$, assumptions on the structure of $A_0$ are needed.
- Example of a recursive constraint by Sims (1980):

$$A_0 = \begin{pmatrix} 1 & 0 \\ a_{21} & 1 \end{pmatrix}, \quad A_0^{-1} = \begin{pmatrix} 1 & 0 \\ -a_{21} & 1 \end{pmatrix}$$

$$\mathrm{Var}(W_t) = \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{pmatrix}$$

- With constraints, $A_0$ can be identified from $\mathrm{Var}(Y_t)$ and $\mathrm{Var}(W_t)$.

# What are potential outcomes in the cross-section?

| $Y_t(u)$ | $Y_c(u)$ | $Y_t(u) - Y_c(u)$ |
|---|---|---|
| 130 | ? | ? |
| ? | 125 | ? |
| 100 | ? | ? |
| ? | 130 | ? |
| ? | 120 | ? |
| 115 | 125 | -10 |

| $Y_t(u)$ | $Y_c(u)$ | $Y_t(u) - Y_c(u)$ |
|---|---|---|
| 130 | 140 | -10 |
| 115 | 125 | -10 |
| 100 | 110 | -10 |
| 120 | 130 | -10 |
| 110 | 120 | -10 |
| 115 | 125 | -10 |

# What are potential outcomes for time series?

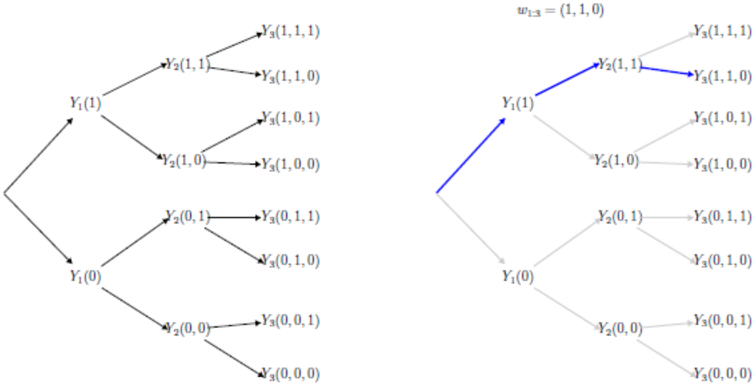


Figure: Time Series Outcomes and Treatments: $Y_{1:3}(w_{1:3})$ for $w_{1:3} = (1, 1, 0)$.

# The Direct Potential Outcome System

**Definition 1 (Direct Potential Outcome System).** Any $\left\{W_t, \left\{Y_t\left(w_{1:t}\right) : w_{1:t} \in \mathcal{W}^t\right\}\right\}_{t\geq 1}$ satisfying Assumptions 1-4 is a direct potential outcome system.

**Assumption 1 (Assignment and Potential Outcome).** The assignment process $\{W_t\}_{t\geq 1}$ satisfies $W_t \in \mathcal{W} := \bigotimes_{k=1}^{d_w} \mathcal{W}_k \subseteq \mathcal{R}^{d_w}$. The potential outcome process is, for any deterministic sequence $\{w_s\}_{s\geq 1}$ with $w_s \in \mathcal{W}$ for all $s \geq 1$, $\left\{Y_t\left(\{w_s\}_{s\geq 1}\right)\right\}_{t\geq 1}$, where the time-t potential outcome satisfies $Y_t\left(\{w_s\}_{s\geq 1}\right) \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$. Allows for standard continuous macro treatments.

# The Direct Potential Outcome System

**Assumption 2 (Non-anticipating Potential Outcomes)** i.e. non-interference. For each $t \geq 1$, and all deterministic $\{w_t\}_{t\geq 1}, \left\{w_t'\right\}_{t\geq 1}$ with $w_t, w_t' \in \mathcal{W}$, $Y_t\left(w_{1:t}, \{w_s\}_{s\geq t+1}\right) = Y_t\left(w_{1:t}, \left\{w_s'\right\}_{s\geq t+1}\right)$ almost surely. Drop references to future assignments.

**Assumption 3 (Output).** The output is $\{W_t, Y_t\}_{t\geq 1} = \left\{W_t, Y_t\left(W_{1:t}\right)\right\}_{t\geq 1}$. The $\{Y_t\}_{t\geq 1}$ is called the outcome process.

**Assumption 4 (Sequentially probabilistic assignment process).** The assignment process satisfies $0 < P\left(W_t = w \mid \mathcal{F}_{t-1}\right) < 1$ with probability one for all $w \in \mathcal{W}$. Here the probabilities are determined by a filtered probability space of $\left\{W_t, \left\{Y_t\left(w_{1:t}\right), w_{1:t} \in \mathcal{W}^{\mathrm{t}}\right\}\right\}_{t\geq 1}$. Overlap condition – ensure there is always support for treatment and control.

# Estimands and Causal Interpretation

$$IRF_{k,t,h}\left(w_k, w_k'\right) = \mathbb{E}\left[Y_{t+h} \mid W_{k,t} = w_k\right] - \mathbb{E}\left[Y_{t+h} \mid W_{k,t} = w_k'\right].$$

$$\mathbb{E}(Y_{t+h}(w_k) - Y_{t+h}(w_k'))$$

---

$$LP_{1,t,h} := y_{t+h} = \beta_h w_{1,t} + \text{controls} + \text{error}$$

$$LP_{k,t,h} := \frac{\operatorname{Cov}\left(Y_{t+h}, W_{k,t}\right)}{\operatorname{Var}\left(W_{k,t}\right)}$$

$$\frac{\int_{\mathcal{W}_k} \mathbb{E}[Y_{t+h}'(w_k)]\mathbb{E}[G_t(w_k)]\, dw_k}{\int_{\mathcal{W}_k} \mathbb{E}[G_t(w_k)]\, dw_k}$$

# Impulse Response Function I

**Theorem 1.** Assume a direct potential outcome system, consider some $(k = 1, \ldots, d_w)$, $(t \geq 1)$, $(h \geq 0)$, fix $(w_k, w_k' \in \mathcal{W}_k)$, and that $(\mathbb{E}\left[|Y_{t+h}(w_k) - Y_{t+h}(w_k')|\right] < \infty)$. Then,

$$\mathrm{IRF}_{k,t,h}(w_k, w_k') = \mathbb{E}[Y_{t+h}(w_k) - Y_{t+h}(w_k')] + \Delta_{k,t,h}(w_k, w_k'),$$

where

$$\Delta_{k,t,h}(w_k, w_k') := \frac{\mathrm{Cov}\left(Y_{t+h}(w_k), \mathbf{1}\{W_{k,t} = w_k\}\right)}{\mathbb{E}\left[\mathbf{1}\{W_{k,t} = w_k\}\right]} - \frac{\mathrm{Cov}\left(Y_{t+h}(w_k'), \mathbf{1}\{W_{k,t} = w_k'\}\right)}{\mathbb{E}\left[\mathbf{1}\{W_{k,t} = w_k'\}\right]}.$$

$$Y_{t+h}(w_k) := Y_{t+h}(W_{1:t-1}, W_{1:k-1,t}, w_k, W_{k+1:d_w,t}, W_{t+1:t+h})$$

# Impulse Response Function II

$$W_{k,t} \perp\!\!\!\perp Y_{t+h}\left(w_k\right), \quad \text{and} \quad W_{k,t} \perp\!\!\!\perp Y_{t+h}\left(w_k'\right),$$

which is in turn implied by

$$W_{k,t} \perp\!\!\!\perp \left\{Y_{t+h}\left(w_k\right) : w_k \in \mathcal{W}_k\right\}$$

which is in turn implied by

$$W_{k,t} \perp\!\!\!\perp \left(W_{1:t-1}, W_{1:k-1,t}, W_{k+1:d_w,t}, W_{t+1:t+h}, \left\{Y_{t+h}\left(w_{1:t+h}\right) : w_{1t+h} \in \mathcal{W}^{t+h}\right\}\right)$$

Does this mean: $\mathbb{E}[W_t \mid W_{1:t-1}, Y_{1:t-1}] = 0$?

# Example I: Dynamic Causal Effect of Federal Funds Rate on Unemployment

- Suppose the FED raises the Federal Funds rate by 25 basis points.
- Question: What is the dynamic causal effect of this change on unemployment, $Y_t$?
- Treat the Federal Funds rate as the treatment variable, $W_t$.
- **Non-Anticipating Potential Outcomes (Our Assumption 1):**
  - Unemployment in the current period, $Y_t$, does not depend on future values of the Federal Funds rate $W_{t+1:T}$.
  - This is conditional on current and past realizations of $W_{1:t}$.
  - However, expectations about future values of the Federal Funds rate may affect the unemployment rate today.

# Example II

**Non-Anticipating Treatment (Assumption from Bojinov and Shephard (2023))**

- The choice of the Federal Funds rate at time $t$ depends only on:
  - Past potential outcomes of the unemployment rate, $Y_{1:t-1}$.
  - Past values of the Federal Funds rate, $W_{1:t-1}$.
- In observational data, treatments may not satisfy Assumption 2:
  - The Federal Reserve may have private information (e.g., about the financial system's health) that helps predict future outcomes.
- To enforce Assumption 2, expand the dimension of $Y_t$ to include observable financial information.

# Local Projection

$$LP_{k,t,h} := \frac{\text{Cov}\left(Y_{t+h}, W_{k,t}\right)}{\text{Var}\left(W_{k,t}\right)}$$

**Theorem 3.** Under the same conditions as Theorem 1, further assume that:

1. The support of $W_{k,t}$ is a closed interval, $\mathcal{W}_k := [\underline{w}_k, \overline{w}_k] \subset \mathbb{R}$.
2. Differentiability: $Y_{t+h}(w_k)$ is continuously differentiable in $w_k$, as is $\mathbb{E}[Y_{t+h}(w_k)]$.
3. Independence: $W_{k,t} \perp\!\!\!\perp \{Y_{t+h}(w_k) : w_k \in \mathcal{W}_k\}$.

Then, if it exists,

$$\text{LP}_{k,t,h} = \frac{\int_{\mathcal{W}_k} \mathbb{E}[Y_{t+h}(w_k)]\mathbb{E}[G_t(w_k)]\, dw_k}{\int_{\mathcal{W}_k} \mathbb{E}[G_t(w_k)]\, dw_k},$$

where

$$G_t(w_k) = \mathbf{1}\{w_k \leq W_{k,t}\}(W_{k,t} - \mathbb{E}[W_{k,t}]), \quad \text{noting } \mathbb{E}[G_t(w_k)] \geq 0.$$

# Take-away

- Structure how you think about shocks.
- Developed a nonparametric, direct potential outcome system to study causal inference in observational time series settings.
- Places no functional form restrictions on the potential outcome process.
- Allows for unrestricted causal effects of past assignments on outcomes.
- Does not require common time series assumptions, such as ”invertibility” or ”recoverability.”
- The direct potential outcome system includes most leading econometric models used in time series analysis as a special case.

# How to estimate a time series causal effect? – An example.

Define a Horvitz and Thompson (1952) style estimator

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

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

# References I

Bojinov, I. and Shephard, N. (2023). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, 2:71–162.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610):917–948.