Lecture 8: Causality in Time Series and Panel Data Models

Vitor Possebom

EESP-FGV

Econometrics 2

Administrative

- Optional Reading: Rambachan and Shephard [2021], Bojinov et al. [2021] and Heckman et al. [2016]
- Even though there are only optional readings, you should read my lecture notes carefully.

Outline

1. Introduction

2. When Do Common Time Series Estimands Have a Causal Meaning?

3. Dynamic Treatment Effects

Introduction

Introduction

- Hamilton's textbook is quite old.
- It predates two important changes in economic research.
  1. Micro-founded Macroeconomics
  2. Credibility Revolution

Introduction

- Hamilton's textbook is quite old.

- It predates two important changes in economic research.
    1. Micro-founded Macroeconomics
    2. Credibility Revolution

- This lecture is an attempt to connect the classic time series and panel data models with the modern literature on causality.

Introduction

- Rambachan and Shephard [2021] analyzes common time series estimands and discuss which assumptions are necessary for them to have a nonparametric causal interpretation.

Introduction

- Rambachan and Shephard [2021] analyzes common time series estimands and discuss which assumptions are necessary for them to have a nonparametric causal interpretation.

- We will focus on two time series estimands that are closely connected to the ones we saw during the course.
  1. Impulse Response Functions.
  2. Local Projections.

Introduction

- Rambachan and Shephard [2021] analyzes common time series estimands and discuss which assumptions are necessary for them to have a nonparametric causal interpretation.
- We will focus on two time series estimands that are closely connected to the ones we saw during the course.
    1. Impulse Response Functions.
    2. Local Projections.
- This paper is very readable and we will go through some of their proofs.

Introduction

- Bojinov et al. [2021] discuss a specific type of randomized control trials (RCTs): Panel Experiments.

Introduction

- Bojinov et al. [2021] discuss a specific type of randomized control trials (RCTs): Panel Experiments.
- It illustrates that even experiments can have interesting econometric problems.
    - For more problems like this, check Susan Athey's interview on *Mixtape: The Podcast*: `https://open.spotify.com/episode/226ISyTNqr7jTwaCafcX1S?si=01173248d1724962`.

Introduction

- Their approach to uncertainty is completely different from what you were exposed to in Econometrics 1 and 2.
  - They focus on design-based uncertainty instead of sampling uncertainty.
  - Consequently, they do not rely on asymptotic approximations. They derive inferential methods based on a finite population perspective.

Introduction

- Their approach to uncertainty is completely different from what you were exposed to in Econometrics 1 and 2.
    - They focus on design-based uncertainty instead of sampling uncertainty.
    - Consequently, they do not rely on asymptotic approximations. They derive inferential methods based on a finite population perspective.

- Since a lot of things are new for you here, I will just emphasize two of their main points.
    - Inverse Probability Weighting Estimators work.
    - Fixed effects models may be problematic.

- If you are interested in this type of work, we can talk more about relevant readings.

Introduction

- Heckman et al. [2016] connect dynamic decision problems with dynamic treatment effects models.

Introduction

- Heckman et al. [2016] connect dynamic decision problems with dynamic treatment effects models.

- It can still identify many interesting economic objects related to structural models but it does not require strong parametric assumptions.

Introduction

- Heckman et al. [2016] connect dynamic decision problems with dynamic treatment effects models.
- It can still identify many interesting economic objects related to structural models but it does not require strong parametric assumptions.
- I always learn something new when I read this paper.
- Since it is a lot to digest, I will focus on two topics.
  - Defining causal parameters based on a flexible economic model.
  - Classic IV estimands are hard to interpret.

When Do Common Time Series Estimands Have a Causal Meaning?

Outline

1. Introduction

2. When Do Common Time Series Estimands Have a Causal Meaning?

3. Dynamic Treatment Effects

When Do Common Time Series Estimands Have a Causal Meaning?

Goal: Rambachan and Shephard [2021] introduce the nonparametric, direct potential outcome system as a foundational framework for analyzing dynamic causal effects of assignments on outcomes in observational time series settings.

When Do Common Time Series Estimands Have a Causal Meaning?

Goal: Rambachan and Shephard [2021] introduce the nonparametric, direct potential outcome system as a foundational framework for analyzing dynamic causal effects of assignments on outcomes in observational time series settings.

Setting:

- Single unit (e.g., macroeconomy or market) observed over time.
- At each time period $t \geq 1$, the unit receives a vector of assignments W_t and an associated vector of outcomes Y_t are generated.

When Do Common Time Series Estimands Have a Causal Meaning?

Goal: Rambachan and Shephard [2021] introduce the nonparametric, direct potential outcome system as a foundational framework for analyzing dynamic causal effects of assignments on outcomes in observational time series settings.

Setting:

- Single unit (e.g., macroeconomy or market) observed over time.
- At each time period $t \geq 1$, the unit receives a vector of assignments W_t and an associated vector of outcomes Y_t are generated.
- The outcomes are causally related to the assignments through a *potential outcome process*.

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Potential Outcome Process

A potential outcome process is a stochastic process that describes what would be observed along counterfactual assignment paths.

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Potential Outcome Process

A potential outcome process is a stochastic process that describes what would be observed along counterfactual assignment paths.

Notation: $\left\{ Y_t \left(\{w_s\}_{s \geq 1} \right) \right\}_{t \geq 1}$

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Potential Outcome Process

A potential outcome process is a stochastic process that describes what would be observed along counterfactual assignment paths.

Notation: $\left\{ Y_t \left(\{w_s\}_{s \geq 1} \right) \right\}_{t \geq 1}$

Intuition: Marvel's Multiverse (*Loki* and *What if?* are particularly useful to understand potential outcomes.)

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Dynamic Causal Effect

A dynamic causal effect is generically defined as the comparison of the potential outcome process along different assignment paths at a fixed point in time.

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Dynamic Causal Effect

A dynamic causal effect is generically defined as the comparison of the potential outcome process along different assignment paths at a fixed point in time.

Notation: $Y_t\left(\{w_s\}_{s\geq 1}\right) - Y_t\left(\{w'_s\}_{s\geq 1}\right)$

- For each time period $t \geq 1$ and assignment paths $\{w_s\}_{s\geq 1}$ and $\{w'_s\}_{s\geq 1}$, $Y_t\left(\{w_s\}_{s\geq 1}\right) - Y_t\left(\{w'_s\}_{s\geq 1}\right)$ is a random variable.

When Do Common Time Series Estimands Have a Causal Meaning?

Definition: Dynamic Causal Effect

A dynamic causal effect is generically defined as the comparison of the potential outcome process along different assignment paths at a fixed point in time.

Notation: $Y_t\left(\{w_s\}_{s\geq 1}\right) - Y_t\left(\{w'_s\}_{s\geq 1}\right)$

- For each time period $t \geq 1$ and assignment paths $\{w_s\}_{s\geq 1}$ and $\{w'_s\}_{s\geq 1}$, $Y_t\left(\{w_s\}_{s\geq 1}\right) - Y_t\left(\{w'_s\}_{s\geq 1}\right)$ is a random variable.
- We will focus on moments of this random variable.

When Do Common Time Series Estimands Have a Causal Meaning?

Advantages:

When Do Common Time Series Estimands Have a Causal Meaning?

Advantages:

- No functional form restrictions on the potential outcome process.

When Do Common Time Series Estimands Have a Causal Meaning?

Advantages:

- No functional form restrictions on the potential outcome process.
- No restriction on the extent to which past assignments may causally affect outcomes.

When Do Common Time Series Estimands Have a Causal Meaning?

Advantages:

- No functional form restrictions on the potential outcome process.
- No restriction on the extent to which past assignments may causally affect outcomes.
- Most leading econometric models used to study dynamic causal effects in time series settings (e.g., Structural VAR) can be cast as special cases of the direct potential outcome system.

When Do Common Time Series Estimands Have a Causal Meaning?

Advantages:

- No functional form restrictions on the potential outcome process.
- No restriction on the extent to which past assignments may causally affect outcomes.
- Most leading econometric models used to study dynamic causal effects in time series settings (e.g., Structural VAR) can be cast as special cases of the direct potential outcome system.
- The potential outcome system provides a flexible, nonparametric foundation upon which to analyze dynamic causal effects in time series settings.

Lecture's goal:

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

To do so, we will analyze two cases.

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

To do so, we will analyze two cases.

- **Case 1:** The researcher observes Y_t and W_t through time.

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

To do so, we will analyze two cases.

- **Case 1:** The researcher observes Y_t and W_t through time.
  - Impulse Response Functions and Local Projections identify, respectively, a dynamic average treatment effect and a weighted average of marginal treatment effects if the assignments are independent of the potential outcome process and over time.

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

To do so, we will analyze two cases.

- **Case 1:** The researcher observes Y_t and W_t through time.
  - Impulse Response Functions and Local Projections identify, respectively, a dynamic average treatment effect and a weighted average of marginal treatment effects if the assignments are independent of the potential outcome process and over time.
- **Case2:** The researcher only observes Y_t.

When Do Common Time Series Estimands Have a Causal Meaning?

Lecture's goal: We will analyze conditions under which predictive time series estimands (Impulse Response Function and Local Projection) have a nonparametric causal interpretation in terms of dynamic causal effects of assignments on outcomes.

To do so, we will analyze two cases.

- **Case 1:** The researcher observes Y_t and W_t through time.
  - Impulse Response Functions and Local Projections identify, respectively, a dynamic average treatment effect and a weighted average of marginal treatment effects if the assignments are independent of the potential outcome process and over time.
- **Case2:** The researcher only observes Y_t.
  - Researchers can recover familiar model-based approaches by introducing functional form restrictions on the potential outcome process.

Assumptions

Assumptions

There is a single unit. At each time period $t \geq 1$, the unit receives a d_w-dimensional assignment $\{W_t\}_{t\geq 1}$. Associated with this assignment process, we observe a d_y-dimensional outcome $\{Y_t\}_{t\geq 1}$.

Assumption 1 (Assignment and Potential Outcome)

The assignment process $\{W_t\}_{t\geq 1}$ satisfies $W_t \in \mathcal{W} := X_{k=1}^{d_w} \mathcal{W}_k \subseteq \mathcal{R}^{d_w}$. 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 $Y_t \left(\{w_s\}_{s\geq 1} \right) \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$.

Assumptions

Assumptions

Our next assumption restricts the potential outcome to only depend on past and contemporaneous assignments.

Assumptions

Our next assumption restricts the potential outcome to only depend on past and contemporaneous assignments.

Assumption 2 (Non-anticipating Potential Outcome)

For each $t \geq 1$ and all deterministic $\{w_t\}_{t\geq 1}$ and $\{w'_t\}_{t\geq 1}$ with $w_t, w'_t \in \mathcal{W}$,

Assumptions

Our next assumption restricts the potential outcome to only depend on past and contemporaneous assignments.

Assumption 2 (Non-anticipating Potential Outcome)

For each $t \geq 1$ and all deterministic $\{w_t\}_{t \geq 1}$ and $\{w'_t\}_{t \geq 1}$ with $w_t, w'_t \in \mathcal{W}$,

$$Y_t\left(w_{1:t}, \{w_s\}_{s \geq t+1}\right) = Y_t\left(w_{1:t}, \{w'_s\}_{s \geq t+1}\right).$$

Under Assumption 2, we drop references to the future assignments in the potential outcome process and write $\left\{Y_t\left(\{w_s\}_{s \geq 1}\right)\right\}_{t \geq 1} = \{Y_t(w_{1:t})\}_{t \geq 1}$.

Assumptions

Assumptions

Assumption 3 (Output)

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

Assumptions

Assumption 3 (Output)

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

The outcome process is the potential outcome process evaluated at the assignment process.

Assumptions

Assumption 3 (Output)

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

The outcome process is the potential outcome process evaluated at the assignment process.

The *Fundamental Problem of Causal Inference* is that we only observe the potential outcome evaluated at the observed assignment path.

Assumptions

Assumptions

Assumption 4 (Sequentially Probabilistic Assignment Process)

The assignment process satisfies $0 < \mathbb{P}\left[W_t = w \middle| W_{1:t-1}, Y_{1:t-1} \right] < 1$ for any $w \in \mathcal{W}$.

Definitions

Definitions

For each time period $t \geq 1$ and assignment paths $w_{1:t}$ and $w'_{1:t}$,

$$Y_t\left(w_{1:t}\right) - Y_t\left(w'_{1:t}\right)$$

is a dynamic causal effect.

Definitions

For each time period $t \geq 1$ and assignment paths $w_{1:t}$ and $w'_{1:t}$,

$$Y_t\left(w_{1:t}\right) - Y_t\left(w'_{1:t}\right)$$

is a dynamic causal effect.

- Enormous class of dynamic causal effects.
- We focus on causal parameters that average over these dynamic causal effects along various underlying assignment paths.

Definitions

For each time period $t \geq 1$ and assignment paths $w_{1:t}$ and $w'_{1:t}$,

$$Y_t\left(w_{1:t}\right) - Y_t\left(w'_{1:t}\right)$$

is a dynamic causal effect.

- Enormous class of dynamic causal effects.
- We focus on causal parameters that average over these dynamic causal effects along various underlying assignment paths.
- **Shorthand Notation:** For $t \geq 1$, $h \geq 0$ and any $w \in \mathcal{W}$, write the time-$(t+h)$ potential outcome at the assignment process $(W_{1:t-1}, w, W_{t+1:t+h})$ as

$$Y_{t+h}(w) = Y_{t+h}\left(W_{1:t-1}, w, W_{t+1:t+h}\right).$$

Definitions

Definitions

Definition 1 (Impulse Causal Effect)

For $t \geq 1$, $h \geq 0$ and any $w, w' \in \mathcal{W}$, the time-t, h-period ahead impulse causal effect is

$$Y_{t+h}(w) - Y_{t+h}(w').$$

Definitions

Definition 1 (Impulse Causal Effect)

For $t \geq 1$, $h \geq 0$ and any $w, w' \in \mathcal{W}$, the time-t, h-period ahead impulse causal effect is

$$Y_{t+h}(w) - Y_{t+h}\left(w'\right).$$

Note: The time-t, h-period ahead impulse causal effect is a random variable.

Definitions

Definition 1 (Impulse Causal Effect)

For $t \geq 1$, $h \geq 0$ and any $w, w' \in \mathcal{W}$, the time-t, h-period ahead impulse causal effect is

$$Y_{t+h}(w) - Y_{t+h}\left(w'\right).$$

Note: The time-t, h-period ahead impulse causal effect is a random variable.

Intuition: It measures the ceteris paribus causal effect of intervening to switch the time-t assignment from w' to w on the h-period ahead outcomes holding all else fixed along the assignment process.

Definitions

Definition 2 (Average Treatment Effect)

For $t \geq 1$, $h \geq 0$ and any $w, w' \in \mathcal{W}$, the time-t, h-period ahead average treatment effect is

$$\mathbb{E}\left[Y_{t+h}(w) - Y_{t+h}(w')\right].$$

Definitions

Definition 2 (Average Treatment Effect)

For $t \geq 1$, $h \geq 0$ and any $w, w' \in \mathcal{W}$, the time-t, h-period ahead average treatment effect is

$$\mathbb{E}\left[Y_{t+h}(w) - Y_{t+h}(w')\right].$$

Intuition: It measures the unconditional expectation of the impulse causal effect $Y_{t+h}(w) - Y_{t+h}(w')$.

Definitions

Definitions

We further define analogous versions of the dynamic causal effects for a particular scalar assignment. For any $w_k \in \mathcal{W}_k$, define

$$Y_{t+h}(w_k) = Y_{t+h}\left(W_{1:t-1}, W_{1:k-1,t}, w_k, W_{k+1:d_W,t}, W_{t+1:t+h}\right).$$

Definitions

We further define analogous versions of the dynamic causal effects for a particular scalar assignment. For any $w_k \in \mathcal{W}_k$, define

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

The corresponding time-t, $h-$period ahead impulse causal effect and average treatment effect for the k-th assignment are, respectively,

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

and

$$\mathbb{E}\left[Y_{t+h}(w_k) - Y_{t+h}(w_k')\right].$$

Definitions

Definitions

The previous definitions summarize causal effects of discrete interventions to switch the time-t assignment on the outcomes. We introduce derivatives to summarize causal effects of marginally varying the time-t assignment.

The previous definitions summarize causal effects of discrete interventions to switch the time-t assignment on the outcomes. We introduce derivatives to summarize causal effects of marginally varying the time-t assignment.

Definition 3 (Marginal Impulse Causal Effect and Average Marginal Treatment Effect)

The time-t, h-period ahead Marginal Impulse Causal Effect and Average Marginal Treatment Effect are, respectively,

$$Y'_{t+h}(w_k) = \frac{\partial Y_{t+h}(w_k)}{\partial w_k}$$

and

$$\mathbb{E}\left[Y'_{t+h}(w_k)\right].$$

Definitions

Definitions

From now on, we assume that the outcome Y_t is univariate. There is no loss in generality in doing so because the more general case is covered by running the analysis equation by equation.

Definitions

From now on, we assume that the outcome Y_t is univariate. There is no loss in generality in doing so because the more general case is covered by running the analysis equation by equation.

Definition 4 (Impulse Response Function)

For $h \geq 0$ and deterministic $w_k, w_k' \in \mathcal{W}$, the impulse response function is defined by

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

Definitions

Definitions

Nonparametrically estimating impulse response functions may be challenging. It is much easier to use local projections, which directly regress the h-step ahead outcome on a constant and the assignment.

Nonparametrically estimating impulse response functions may be challenging. It is much easier to use local projections, which directly regress the h-step ahead outcome on a constant and the assignment.

Definition 5 (Local Projection)

For $h \geq 0$, the Local Projection estimand is defined by

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

Interpreting IRFs

Interpreting IRFs

Theorem 6

Consider some $k = 1, \dots, d_w$, $t \geq 1$, $h \geq 0$, *fix* $w_k, w_k' \in \mathcal{W}_k$ *and that* $\mathbb{E}\left[\left|Y_{t+h}(w_k) - Y_{t+h}(w_k')\right|\right] < \infty$. *Then, under Assumptions 1-4,*

$$IRF_{k,t,h}\left(w_k, w_k'\right) = \underbrace{\mathbb{E}\left[Y_{t+h}\left(w_k\right) - Y_{t+h}\left(w_k'\right)\right]}_{Average Treatment Effect} + \underbrace{\Delta_{k,t,h}\left(w_k, w_k'\right)}_{Selection Bias},$$

Interpreting IRFs

Theorem 6

Consider some $k = 1, \ldots, d_w$, $t \geq 1$, $h \geq 0$, *fix* $w_k, w_k' \in \mathcal{W}_k$ *and that* $\mathbb{E}\left[\left|Y_{t+h}(w_k) - Y_{t+h}(w_k')\right|\right] < \infty$. *Then, under Assumptions 1-4,*

$$IRF_{k,t,h}\left(w_k, w_k'\right) = \underbrace{\mathbb{E}\left[Y_{t+h}\left(w_k\right) - Y_{t+h}\left(w_k'\right)\right]}_{AverageTreatmentEffect} + \underbrace{\Delta_{k,t,h}\left(w_k, w_k'\right)}_{SelectionBias},$$

where

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

Interpreting IRFs

Interpreting IRFs

Proof: Note that

$$
\begin{aligned}
\mathbb{E}\left[\left. Y_{t+h} \right| W_{k,t} = w_k\right] &= \frac{\mathbb{E}\left[Y_{t+h} \cdot 1\left\{W_{k,t} = w_k\right\}\right]}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \frac{\mathbb{E}\left[Y_{t+h}\left(w_k\right) \cdot 1\left\{W_{k,t} = w_k\right\}\right]}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \frac{\mathbb{E}\left[Y_{t+h}\left(w_k\right)\right] \cdot \mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right] + \operatorname{Cov}\left(Y_{t+h}\left(w_k\right), 1\left\{W_{k,t} = w_k\right\}\right)}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \mathbb{E}\left[Y_{t+h}\left(w_k\right)\right] + \frac{\operatorname{Cov}\left(Y_{t+h}\left(w_k\right), 1\left\{W_{k,t} = w_k\right\}\right)}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]}.
\end{aligned}
$$

Proof: Note that

$$
\begin{aligned}
\mathbb{E}\left[\, Y_{t+h} \right| W_{k,t} = w_k] &= \frac{\mathbb{E}\left[Y_{t+h} \cdot 1\left\{W_{k,t} = w_k\right\}\right]}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \frac{\mathbb{E}\left[Y_{t+h}\left(w_k\right) \cdot 1\left\{W_{k,t} = w_k\right\}\right]}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \frac{\mathbb{E}\left[Y_{t+h}\left(w_k\right)\right] \cdot \mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right] + \operatorname{Cov}\left(Y_{t+h}\left(w_k\right), 1\left\{W_{k,t} = w_k\right\}\right)}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]} \\
&= \mathbb{E}\left[Y_{t+h}\left(w_k\right)\right] + \frac{\operatorname{Cov}\left(Y_{t+h}\left(w_k\right), 1\left\{W_{k,t} = w_k\right\}\right)}{\mathbb{E}\left[1\left\{W_{k,t} = w_k\right\}\right]}.
\end{aligned}
$$

The result is then immediate by (i) applying the same calculation to $\mathbb{E}\left[\, Y_{t+h} \right| W_{k,t} = w_k']$ and (ii) taking the difference.

Interpreting IRFs

Interpreting IRFs

Our last theorem states that the impulse response function is equal to the average treatment effect if and only if the selection bias term

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

is equal to zero.

Interpreting IRFs

Our last theorem states that the impulse response function is equal to the average treatment effect if and only if the selection bias term

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

is equal to zero.

These covariace terms depend on how the assignment $W_{k,t}$ covaries with the potential outcome $Y_{t+h}\left(w_k\right) = Y_{t+h}\left(W_{1:t-1}, w_k, W_{-k,t}, W_{t+1:t+h}\right)$. Consequently, the selection bias term depends on how the assingment $W_{k,t}$ relates to past assignments, other contemporaneous assignments, future assignments and the potential outcome process.

Interpreting IRFs

Interpreting IRFs

By placing further restrictions on the assignment process, we arrive at sufficient conditions for $\Delta_{k,t,h}(w_k, w'_k) = 0$. In particular, we impose that the assignment $W_{k,t}$ is randomized in the sense that it is independent of all other assignments and the time-$(t + h)$ potential outcomes.

By placing further restrictions on the assignment process, we arrive at sufficient conditions for $\Delta_{k,t,h}(w_k, w_k') = 0$. In particular, we impose that the assignment $W_{k,t}$ is randomized in the sense that it is independent of all other assignments and the time-$(t+h)$ potential outcomes.

We formalize this intuition in the next slide.

Interpreting IRFs

Theorem 7

Under Assumptions 1-4, if $\mathrm{Cov}\left(Y_{t+h}\left(w_k\right), 1\left\{W_{k,t}=w_k\right\}\right)=0$ *and* $\mathrm{Cov}\left(Y_{t+h}\left(w_k'\right), 1\left\{W_{k,t}=w_k'\right\}\right)=0$, *then* $\Delta_{k,t,h}\left(w_k, w_k'\right)=0$. *Moreover, this condition is satisfied if*

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

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

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

Interpreting Local Projections

Interpreting Local Projections

Theorem 8

Under Assumptions 1-4, further assume that:

Interpreting Local Projections

Theorem 8

Under Assumptions 1-4, further assume that:

1. *The support of $W_{k,t}$ is a closed interval, $\mathcal{W}_k := [\underline{w}_k, \bar{w}_k] \subset \mathbb{R}$.*

Interpreting Local Projections

Theorem 8

Under Assumptions 1-4, further assume that:

1. *The support of $W_{k,t}$ is a closed interval, $\mathcal{W}_k := [\underline{w}_k, \bar{w}_k] \subset \mathbb{R}$.*
2. *Differentiability: $Y_{t+h}(w_k)$ is continuously differentiable in w_k, as is $\mathbb{E}\left[Y'_{t+h}(w_k)\right]$.*

Theorem 8

Under Assumptions 1-4, further assume that:

1. *The support of* $W_{k,t}$ *is a closed interval,* $\mathcal{W}_k := [\underline{w}_k, \bar{w}_k] \subset \mathbb{R}$.
2. *Differentiability:* $Y_{t+h}(w_k)$ *is continuously differentiable in* w_k*, as is* $\mathbb{E}\left[Y'_{t+h}(w_k)\right]$.
3. *Independence:* $W_{k,t} \perp\!\!\!\perp \{Y_{t+h}(w_k) : w_k \in \mathcal{W}_k\}$.

Then,

Interpreting Local Projections

Theorem 8

Under Assumptions 1-4, further assume that:

1. *The support of* $W_{k,t}$ *is a closed interval,* $\mathcal{W}_k := [\underline{w}_k, \bar{w}_k] \subset \mathbb{R}$.
2. *Differentiability:* $Y_{t+h}(w_k)$ *is continuously differentiable in* w_k*, as is* $\mathbb{E}\left[Y'_{t+h}(w_k)\right]$.
3. *Independence:* $W_{k,t} \perp\!\!\!\perp \{Y_{t+h}(w_k) : w_k \in \mathcal{W}_k\}$.

Then,

$$LP_{k,t,h} = \frac{\int_{\mathcal{W}_k} \mathbb{E}\left[Y'_{t+h}(w_k)\right] \mathbb{E}\left[G_t(w_k)\right] dw_k}{\int_{\mathcal{W}_k} \mathbb{E}\left[G_t(w_k)\right] dw_k},$$

Interpreting Local Projections

Theorem 8

Under Assumptions 1-4, further assume that:

1. *The support of $W_{k,t}$ is a closed interval, $\mathcal{W}_k := [\underline{w}_k, \bar{w}_k] \subset \mathbb{R}$.*
2. *Differentiability: $Y_{t+h}(w_k)$ is continuously differentiable in w_k, as is $\mathbb{E}\left[Y'_{t+h}(w_k)\right]$.*
3. *Independence: $W_{k,t} \perp\!\!\!\perp \{Y_{t+h}(w_k) : w_k \in \mathcal{W}_k\}$.*

Then,

$$LP_{k,t,h} = \frac{\int_{\mathcal{W}_k} \mathbb{E}\left[Y'_{t+h}(w_k)\right] \mathbb{E}\left[G_t(w_k)\right] dw_k}{\int_{\mathcal{W}_k} \mathbb{E}\left[G_t(w_k)\right] dw_k},$$

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

Interpreting Local Projections

Proof: First, note that

$$\begin{aligned} Y_{t+h} &= Y_{t+h}\left(W_{k,t}\right) \\ &= Y_{t+h}\left(\underline{w}_k\right) + \int_{\underline{w}_k}^{W_{k,t}} \frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k}\, d\tilde{w}_k \\ &\quad \text{by the Fundamental Theorem of Calculus} \\ &= Y_{t+h}\left(\underline{w}_k\right) + \int_{\underline{w}_k}^{\overline{w}_k} \frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k} \cdot 1\left\{\tilde{w}_k \leq W_{k,t}\right\} d\tilde{w}_k. \end{aligned} \tag{1}$$

Interpreting Local Projections

Moreover, observe that

$$
\begin{aligned}
\operatorname{Cov}\left(Y_{t+h}, W_{k,t}\right) &= \mathbb{E}\left[Y_{t+h}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] \\
&= \mathbb{E}\left[Y_{t+h}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] - 0 \\
&= \mathbb{E}\left[Y_{t+h}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] - \mathbb{E}\left[Y_{t+h}\left(\underline{w}_k\right)\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] \\
&\qquad \text{because } W_{k,t} \perp\!\!\!\perp \left\{Y_{t+h}\left(w_k\right) : w_k \in \mathcal{W}_k\right\} \\
&= \mathbb{E}\left[\left(Y_{t+h} - Y_{t+h}\left(\underline{w}_k\right)\right)\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] \\
&= \mathbb{E}\left[\left(\int_{\underline{w}_k}^{\overline{w}_k} \frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k} 1\left\{\tilde{w}_k \leq W_{k,t}\right\} d\tilde{w}_k\right)\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] \\
&\qquad \text{by Equation (1).}
\end{aligned}
$$

Interpreting Local Projections

Moreover, observe that

$$\operatorname{Cov}\left(Y_{t+h}, W_{k,t}\right) = \mathbb{E}\left[\left(\int_{\underline{w}_k}^{\overline{w}_k} \frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k} 1\left\{\tilde{w}_k \leq W_{k,t}\right\} d\tilde{w}_k\right)\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right]$$

$$= \int_{\underline{w}_k}^{\overline{w}_k} \mathbb{E}\left[\frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k} 1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] d\tilde{w}_k$$

because our differentiability assumption ensures

that we can apply Fubini's Theorem

$$= \int_{\underline{w}_k}^{\overline{w}_k} \mathbb{E}\left[\frac{\partial Y_{t+h}\left(\tilde{w}_k\right)}{\partial \tilde{w}_k}\right] \mathbb{E}\left[1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] d\tilde{w}_k$$

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

Interpreting Local Projections

Analogously, we have that

$$\mathrm{Var}\left(W_{k,t}\right) = \int_{\underline{w}_k}^{\overline{w}_k} \mathbb{E}\left[1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] d\tilde{w}_k.$$

Analogously, we have that

$$\operatorname{Var}\left(W_{k,t}\right)=\int_{\underline{w}_k}^{\overline{w}_k}\mathbb{E}\left[1\left\{\tilde{w}_k\leq W_{k,t}\right\}\left(W_{k,t}-\mathbb{E}\left[W_{k,t}\right]\right)\right]d\tilde{w}_k.$$

The result follows immediately.

Interpreting Local Projections

Analogously, we have that

$$\mathrm{Var}\left(W_{k,t}\right) = \int_{\underline{w}_k}^{\overline{w}_k} \mathbb{E}\left[1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] d\tilde{w}_k.$$

The result follows immediately.

To see that the weights are non-negative,

Interpreting Local Projections

Analogously, we have that

$$\mathrm{Var}\left(W_{k,t}\right) = \int_{\underline{w}_k}^{\overline{w}_k} \mathbb{E}\left[1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] d\tilde{w}_k .$$

The result follows immediately.

To see that the weights are non-negative, observe that, for $\tilde{w}_k \in [\underline{w}_k, \bar{w}_k]$,

$$\begin{aligned}
&\mathbb{E}\left[1\left\{W_{k,t} \geq \tilde{w}_k\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)\right] \\
&\quad = \mathbb{E}\left[1\left\{W_{k,t} \geq \tilde{w}_k\right\} W_{k,t}\right] - \mathbb{E}\left[1\left\{W_{k,t} \geq \tilde{w}_k\right\}\right]\mathbb{E}\left[W_{k,t}\right] \\
&\quad = \left(\mathbb{E}\left[W_{k,t} \mid W_{k,t} \geq \tilde{w}_k\right] - \mathbb{E}\left[W_{k,t}\right]\right)\mathbb{P}\left(W_{k,t} \geq \tilde{w}_k\right) \geq 0
\end{aligned}$$

since $\mathbb{E}\left[W_{k,t} \mid W_{k,t} \geq \tilde{w}_k\right] \geq \mathbb{E}\left[W_{k,t}\right]$ for $\tilde{w}_k \in [\underline{w}_k, \bar{w}_k]$.

Interpreting Local Projections

Interpreting Local Projections

According to our last theorem, we know that

$$LP_{k,t,h} = \frac{\int_{\mathcal{W}_k} \mathbb{E}\left[Y'_{t+h}\left(w_k\right)\right] \mathbb{E}\left[G_t\left(w_k\right)\right] dw_k}{\int_{\mathcal{W}_k} \mathbb{E}\left[G_t\left(w_k\right)\right] dw_k}.$$

Intuitively, the local projection estimand is a weighted average of average marginal treatment effects of $W_{k,t}$ on Y_{t+h}, where the weights are non-negative and sum to one.

Interpreting Local Projections

According to our last theorem, we know that

$$LP_{k,t,h} = \frac{\int_{\mathcal{W}_k} \mathbb{E}\left[Y'_{t+h}(w_k)\right] \mathbb{E}\left[G_t(w_k)\right] dw_k}{\int_{\mathcal{W}_k} \mathbb{E}\left[G_t(w_k)\right] dw_k}.$$

Intuitively, the local projection estimand is a weighted average of average marginal treatment effects of $W_{k,t}$ on Y_{t+h}, where the weights are non-negative and sum to one.

Thus, if the assignment $W_{k,t}$ is an exogenous shock, the local projection estimand has a nonparametric causal interpretation.

Interpreting Local Projections

Interpreting Local Projections

Food for Thought: Are the weights,

$$\frac{1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)}{\int_{\mathcal{W}_k} 1\left\{\check{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right) d\check{w}_k},$$

policy-relevant? Do the weights capture a economically meaningful parameter?

Food for Thought: Are the weights,

$$\frac{1\left\{\tilde{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right)}{\int_{\mathcal{W}_k} 1\left\{\check{w}_k \leq W_{k,t}\right\}\left(W_{k,t} - \mathbb{E}\left[W_{k,t}\right]\right) d\check{w}_k},$$

policy-relevant? Do the weights capture a economically meaningful parameter?

In some contexts, linearity may have relevant costs.

Case 2: Estimands Based Only on Outcomes

Case 2: Estimands Based Only on Outcomes

- The dominant approach to causal inference in macroeconometrics is a model-based approach.

Case 2: Estimands Based Only on Outcomes

- The dominant approach to causal inference in macroeconometrics is a model-based approach.
- Researchers introduce parametric models to study the dynamic causal effects of *unobservable* structural shocks.

Case 2: Estimands Based Only on Outcomes

- The dominant approach to causal inference in macroeconometrics is a model-based approach.
- Researchers introduce parametric models to study the dynamic causal effects of *unobservable* structural shocks.
- These structural shocks must be inferred from the outcomes.

Case 2: Estimands Based Only on Outcomes

- The dominant approach to causal inference in macroeconometrics is a model-based approach.
- Researchers introduce parametric models to study the dynamic causal effects of *unobservable* structural shocks.
- These structural shocks must be inferred from the outcomes.
- Rambachan and Shephard [2021] illustrate that this approach can be nested in the direct potential outcome system framework.

Case 2: Estimands Based Only on Outcomes

Case 2: Estimands Based Only on Outcomes

Focusing on the one lag model with no intercept for brevity, the Structural VAR approach assumes that the potential outcome process satisfies

$$A_0 Y_t \left(w_{1:t}\right) = w_t + A_1 Y_{t-1} \left(w_{1:t-1}\right),$$

Case 2: Estimands Based Only on Outcomes

Focusing on the one lag model with no intercept for brevity, the Structural VAR approach assumes that the potential outcome process satisfies

$$A_0 Y_t (w_{1:t}) = w_t + A_1 Y_{t-1} (w_{1:t-1}),$$

implying that

$$Y_t (w_{1:t}) = A_0^{-1} w_t + A_0^{-1} A_1 Y_{t-1} (w_{1:t-1}).$$

Case 2: Estimands Based Only on Outcomes

Focusing on the one lag model with no intercept for brevity, the Structural VAR approach assumes that the potential outcome process satisfies

$$A_0 Y_t\left(w_{1:t}\right) = w_t + A_1 Y_{t-1}\left(w_{1:t-1}\right),$$

implying that

$$Y_t\left(w_{1:t}\right) = A_0^{-1} w_t + A_0^{-1} A_1 Y_{t-1}\left(w_{1:t-1}\right).$$

In this case, the *h*-period ahead average marginal treatment effect is

$$\mathbb{E}\left[\frac{\partial Y_{t+h}\left(w_{1:t+h}\right)}{\partial w_t'}\right] = \left(A_0^{-1} A_1\right)^h A_0^{-1}.$$

Case 2: Estimands Based Only on Outcomes

In this case, the h-period ahead average marginal treatment effect is

$$\mathbb{E}\left[\frac{\partial Y_{t+h}\left(w_{1:t+h}\right)}{\partial w_t'}\right] = \left(A_0^{-1} A_1\right)^h A_0^{-1}.$$

Case 2: Estimands Based Only on Outcomes

In this case, the h-period ahead average marginal treatment effect is

$$\mathbb{E}\left[\frac{\partial Y_{t+h}\left(w_{1:t+h}\right)}{\partial w_t'}\right] = \left(A_0^{-1}A_1\right)^h A_0^{-1}.$$

- $\left(A_0^{-1}A_1\right)$ can be determined from the dynamics of the observable outcomes if this process is stationary.

Case 2: Estimands Based Only on Outcomes

In this case, the h-period ahead average marginal treatment effect is

$$\mathbb{E}\left[\frac{\partial Y_{t+h}\left(w_{1:t+h}\right)}{\partial w_t'}\right] = \left(A_0^{-1} A_1\right)^h A_0^{-1}.$$

- $\left(A_0^{-1} A_1\right)$ can be determined from the dynamics of the observable outcomes if this process is stationary.
- A_0 and $\mathrm{Var}\left(W_t\right)$ cannot be separately identified from the observable outcomes. So further structural assumptions are needed.

Dynamic Treatment Effects

Outline

Dynamic Treatment Effects

- Heckman et al. [2016] develops models for estimating and interpreting treatment effects arising from both ordered and unordered multi-stage decision problems.

Dynamic Treatment Effects

- Heckman et al. [2016] develops models for estimating and interpreting treatment effects arising from both ordered and unordered multi-stage decision problems. (We will focus on the ordered case.)

Dynamic Treatment Effects

- Heckman et al. [2016] develops models for estimating and interpreting treatment effects arising from both ordered and unordered multi-stage decision problems. (We will focus on the ordered case.)
- Identification is secured through instrumental variable and conditional independence assumptions.
- They decompose treatment effects into direct effects and continuation values associated with moving to the next stage of a decision problem.
- IV estimand does not capture an economically interpretable parameter in dynamic discrete choice models unless policy variables are instruments.

Dynamic Treatment Effects

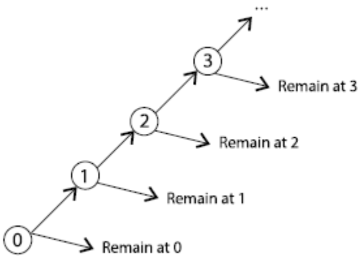

Fig. 1. An ordered multi-stage dynamic decision model.

- This figure illustrates one ordered decision problem.
- The stages could correspond to a sequence of educational choices.
- All agents start at stage 0. Some transit to 1, while other stay at 0 forever, and some of those who go to 1 stop there while others go on, etc.
- At each stage, agents update their information sets and decide whether to transit to the next stage.
- Associated with each stage is a set of potential outcomes.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j + 1$.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j+1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j+1$.

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j+1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j+1$. $D_j = 1$ if a person stops at j.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \dots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j+1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j+1$. $D_j = 1$ if a person stops at j.
- Let $S = s$ if $D_s = 1$.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j+1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j+1$. $D_j = 1$ if a person stops at j.
- Let $S = s$ if $D_s = 1$.
- The environment is time-stationary and decisions are irreversible.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j + 1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j + 1$. $D_j = 1$ if a person stops at j.
- Let $S = s$ if $D_s = 1$.
- The environment is time-stationary and decisions are irreversible.
- $Q_j = 1$ indicates that an agent gets to decision node j.

Dynamic Treatment Effects

- Let $\mathcal{J} = \{0, \ldots, \bar{s}\}$ denote an ordered set of possible terminal states.
- At each node, there are only two possible choices: remain at j or transit to $j + 1$.
- $D_j = 0$ if a person at j does not stop there and goes on to $j + 1$. $D_j = 1$ if a person stops at j.
- Let $S = s$ if $D_s = 1$.
- The environment is time-stationary and decisions are irreversible.
- $Q_j = 1$ indicates that an agent gets to decision node j. $Q_j = 0$ if the person never gets there.

Dynamic Treatment Effects

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

for $Q_j = 1$, $j \in \{0, \ldots, \bar{s} - 1\}$, where

Dynamic Treatment Effects

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

for $Q_j = 1$, $j \in \{0, \ldots, \bar{s} - 1\}$, where

- I_j is the perceived value at node j of going on to $j + 1$.

Dynamic Treatment Effects

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

for $Q_j = 1$, $j \in \{0, \ldots, \bar{s} - 1\}$, where

- I_j is the perceived value at node j of going on to $j + 1$.
- Heckman et al. [2016] do not take a position on the precise content of the information sets and preferences governing agent choices at different nodes or the exact decision rules used.

Dynamic Treatment Effects

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

for $Q_j = 1$, $j \in \{0, \ldots, \bar{s} - 1\}$, where

- I_j is the perceived value at node j of going on to $j + 1$.
- Heckman et al. [2016] do not take a position on the precise content of the information sets and preferences governing agent choices at different nodes or the exact decision rules used.
- Agents could make irrational choices and their choices could be governed by behavioral anomalies.

Dynamic Treatment Effects

The decision process at each node is assumed to be characterized by an index threshold-crossing property:

$$D_j = \begin{cases} 0 & \text{if } I_j \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

for $Q_j = 1$, $j \in \{0, \ldots, \bar{s} - 1\}$, where

- I_j is the perceived value at node j of going on to $j + 1$.
- Heckman et al. [2016] do not take a position on the precise content of the information sets and preferences governing agent choices at different nodes or the exact decision rules used.
- Agents could make irrational choices and their choices could be governed by behavioral anomalies.
- Thus, the model does not rule out dynamic inconsistency (e.g., ex-post regret).

Dynamic Treatment Effects

- Associated with each final state $s \in \mathcal{J}$ is a set of K potential outcomes for each agent with indices $k \in \mathcal{K} = \{1, \dots, K\}$.

Dynamic Treatment Effects

- Associated with each final state $s \in \mathcal{J}$ is a set of K potential outcomes for each agent with indices $k \in \mathcal{K} = \{1, \ldots, K\}$.

- We denote each potential outcome by Y_s^k for $k \in \mathcal{K}, s \in \mathcal{J}$.

Dynamic Treatment Effects

- Associated with each final state $s \in \mathcal{J}$ is a set of K potential outcomes for each agent with indices $k \in \mathcal{K} = \{1, \ldots, K\}$.

- We denote each potential outcome by Y_s^k for $k \in \mathcal{K}, s \in \mathcal{J}$.

- The observed outcome Y^k is given by

$$Y^k = \sum_{s \in \mathcal{J}} D_s \cdot Y_s^k.$$

Parametrizations of decision rules and potential outcomes

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j(Z) + \boldsymbol{\theta}'\boldsymbol{\alpha}_j - \nu_j,$$

where

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j(Z) + \boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j,$$

where

- $\phi_j(Z)$ is a function of observed variables Z.

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j\left(Z\right) + \boldsymbol{\theta}'\boldsymbol{\alpha}_j - \nu_j,$$

where

- $\phi_j\left(Z\right)$ is a function of observed variables Z.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j(Z) + \boldsymbol{\theta}'\boldsymbol{\alpha}_j - \nu_j,$$

where

- $\phi_j(Z)$ is a function of observed variables Z.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.
- ν_j is an idiosyncratic transition-specific innovation.

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j(Z) + \boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j,$$

where

- $\phi_j(Z)$ is a function of observed variables Z.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.
- ν_j is an idiosyncratic transition-specific innovation.
- Endowments θ are not directly observed by the econometrician but are proxied by a measurement system.

Parametrizations of decision rules and potential outcomes

We represent I_j using a separable model:

$$I_j = \phi_j(Z) + \boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j,$$

where

- $\phi_j(Z)$ is a function of observed variables Z.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.
- ν_j is an idiosyncratic transition-specific innovation.
- Endowments θ are not directly observed by the econometrician but are proxied by a measurement system.
  - If we interpret θ as cognitive skill, we can proxy it with test scores.

Parametrizations of decision rules and potential outcomes

Parametrizations of decision rules and potential outcomes

Outcomes are also represented by a separable model:

$$Y_s^k = \tau_s^k(X) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k - \omega_s^k,$$

where

Parametrizations of decision rules and potential outcomes

Outcomes are also represented by a separable model:

$$Y_s^k = \tau_s^k(X) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k - \omega_s^k,$$

where

- $\tau_s^k(X)$ is a function of observed variables X.

Parametrizations of decision rules and potential outcomes

Outcomes are also represented by a separable model:

$$Y_s^k = \tau_s^k(X) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k - \omega_s^k,$$

where

- $\tau_s^k(X)$ is a function of observed variables X.
    - Z can include all variables in X. When using IV, it is assumed that there are some variable in Z not in X.

Parametrizations of decision rules and potential outcomes

Outcomes are also represented by a separable model:

$$Y_s^k = \tau_s^k(X) + \boldsymbol{\theta}' \boldsymbol{\alpha}_s^k - \omega_s^k,$$

where

- $\tau_s^k(X)$ is a function of observed variables X.
    - Z can include all variables in X. When using IV, it is assumed that there are some variable in Z not in X.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.

Parametrizations of decision rules and potential outcomes

Outcomes are also represented by a separable model:

$$Y_s^k = \tau_s^k(X) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k - \omega_s^k,$$

where

- $\tau_s^k(X)$ is a function of observed variables X.
  - Z can include all variables in X. When using IV, it is assumed that there are some variable in Z not in X.
- $\boldsymbol{\theta}$ is a finite dimensional vector of unobserved (by the economist) endowments.
- ω_s^k is an idiosyncratic innovation for the k-th outcome in state s.

Assumptions

Assumptions

$\boldsymbol{\theta}$ captures all sources of dependence across transitions apart from X and Z. Along with the observed variables, it generates dependence between choices and outcomes.

Formally, we assume that, conditional on X,

$$\nu_j \perp\!\!\!\perp \nu_l \quad \forall l \neq j \quad l, j \in \{0, \ldots, \bar{s} - 1\}$$

$$\omega_s^k \perp\!\!\!\perp \omega_{s'}^k \quad \forall s \neq s' \quad \forall k \text{ and } s, s' \in \mathcal{J}$$

$$\boldsymbol{\omega}_s \perp\!\!\!\perp \boldsymbol{\nu} \quad \forall s \in \mathcal{J}$$

$$\boldsymbol{\theta} \perp\!\!\!\perp Z$$

$$(\boldsymbol{\omega}_s, \boldsymbol{\nu}) \perp\!\!\!\perp (\theta, Z) \quad \forall s \in \mathcal{J}$$

Definitions: Treatment Effects

Definitions: Treatment Effects

Heckman et al. [2016] define treatment effects that take into account the direct effect of transiting to the next node in a decision tree plus the benefits associates with the options opened up by the additional choices made possible by such transitions.

Definitions: Treatment Effects

The person-specific treatment effect T_j^k for outcome $k \in \mathcal{K}$ for an individual selected from the population with $Q_j = 1$ with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ is the difference between the individual's outcomes under two possible actions:

Definitions: Treatment Effects

The person-specific treatment effect T_j^k for outcome $k \in \mathcal{K}$ for an individual selected from the population with $Q_j = 1$ with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ is the difference between the individual's outcomes under two possible actions:

$$\begin{aligned}
T_j^k & \left[Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}} \right] \\
& := \left(Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0 \right) \\
& \quad - \left(Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1 \right)
\end{aligned}$$

where

Definitions: Treatment Effects

The person-specific treatment effect T_j^k for outcome $k \in \mathcal{K}$ for an individual selected from the population with $Q_j = 1$ with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ is the difference between the individual's outcomes under two possible actions:

$$
\begin{aligned}
T_j^k & \left[Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}} \right] \\
& := \left(Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0 \right) \\
& \quad - \left(Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1 \right)
\end{aligned}
$$

where

- $\left(Y_k | X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0 \right)$ is the value of Y_k for a person with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ from the population that attains at least node j and for whom we fix $D_j = 0$ so that the agent is forced to go on to the next node.

Definitions: Treatment Effects

The person-specific treatment effect T_j^k for outcome $k \in \mathcal{K}$ for an individual selected from the population with $Q_j = 1$ with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ is the difference between the individual's outcomes under two possible actions:

$$
\begin{aligned}
&T_j^k \left[\, Y_k \middle| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}} \right] \\
&\quad := \left(Y_k \middle| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0 \right) \\
&\qquad - \left(Y_k \middle| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1 \right)
\end{aligned}
$$

where

- $\left(Y_k \middle| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 0 \right)$ is the value of Y_k for a person with characteristics $X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ from the population that attains at least node j and for whom we fix $D_j = 0$ so that the agent is forced to go on to the next node.
- $\left(Y_k \middle| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, \text{Fix } D_j = 1 \right)$ is defined for the same individual but forces the person with these characteristics not to transit to the next node.

Definitions: Treatment Effects

Definitions: Treatment Effects

The person-specific treatment effect can be decomposed into two components.

Definitions: Treatment Effects

The person-specific treatment effect can be decomposed into two components.

- The **Direct Effect** of going from j to $j+1$:

$$DE_j^k = Y_{j+1}^k - Y_j^k.$$

Definitions: Treatment Effects

The person-specific treatment effect can be decomposed into two components.

- The **Direct Effect** of going from j to $j+1$:

$$DE_j^k = Y_{j+1}^k - Y_j^k.$$

- The **Continuation Value** of going beyond $j+1$:

$$C_{j+1}^k = \sum_{r=1}^{\bar{s}-(j+1)} \left[\prod_{l=1}^{r} (1 - D_{j+l}) \right] \left(Y_{j+r+1}^k - Y_{j+r}^k \right).$$

The person-specific treatment effect can be decomposed into two components.

- The **Direct Effect** of going from j to $j+1$:

$$DE_j^k = Y_{j+1}^k - Y_j^k.$$

- The **Continuation Value** of going beyond $j+1$:

$$C_{j+1}^k = \sum_{r=1}^{\bar{s}-(j+1)} \left[\prod_{l=1}^{r} (1 - D_{j+l}) \right] \left(Y_{j+r+1}^k - Y_{j+r}^k \right).$$

We have that the **Total Effect** of fixing $D_j = 0$ on Y^k satisfies

$$T_j^k = DE_j^k + C_{j+1}^k.$$

Definitions: Treatment Effects

Definitions: Treatment Effects

The associated population level average treatment effect conditional on $Q_j = 1$ is

$$ATE_j^k := \int \mathbb{E}\left(T_j^k\left[\left. Y_k \right| X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}\right]\right) dF_{X,Z,\boldsymbol{\theta}}\left(x, z, \left.\overline{\boldsymbol{\theta}}\right| Q_j = 1\right).$$

Definitions: Treatment Effects

The associated population level average treatment effect conditional on $Q_j = 1$ is

$$ATE_j^k := \int \mathbb{E}\left(T_j^k\left[\,Y_k\middle|\, X = x, Z = z, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}\right]\right) dF_{X,Z,\boldsymbol{\theta}}\left(x, z, \overline{\boldsymbol{\theta}}\middle|\, Q_j = 1\right).$$

The population continuation value at $j+1$ conditional on $Q_j = 1$ is

$$ACV_{j+1}^k = \mathbb{E}_{X,Z,\boldsymbol{\theta}}\left[\sum_{l=j+1}^{\bar{s}-1}\left\{\begin{array}{c}\mathbb{E}\left[\,Y_{l+1}^k - Y_l^k\middle|\, X, Z, \boldsymbol{\theta}, Q_{l+1} = 1, \text{Fix } Q_{j+1} = 1\right] \\ \cdot\mathbb{P}\left[\,Q_{l+1} = 1\middle|\, \text{Fix } Q_{j+1} = 1, X, Z, \boldsymbol{\theta}, Q_j = 1\right]\end{array}\right\}\middle|\, Q_j = 1\right].$$

Understanding the Wald (IV) Estimand

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y | Z = z_2] - \mathbb{E}[Y | Z = z_1]}{\mathbb{E}[S | Z = z_2] - \mathbb{E}[S | Z = z_1]}$$

$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1} | S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y | Z = z_2] - \mathbb{E}[Y | Z = z_1]}{\mathbb{E}[S | Z = z_2] - \mathbb{E}[S | Z = z_1]} = \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1} | S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

- $\mathbb{E}[Y_j - Y_{j-1} | S(z_2) \geq j > S(z_1)]$ is the mean gain of going from $j-1$ to j for a set of people who pass through j even though they may end up far above j and start far below j.

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y| Z = z_2] - \mathbb{E}[Y| Z = z_1]}{\mathbb{E}[S| Z = z_2] - \mathbb{E}[S| Z = z_1]}$$
$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1}| S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

- $\mathbb{E}[Y_j - Y_{j-1}| S(z_2) \geq j > S(z_1)]$ is the mean gain of going from $j-1$ to j for a set of people who pass through j even though they may end up far above j and start far below j. **Is this an interesting margin of choice?**

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y | Z = z_2] - \mathbb{E}[Y | Z = z_1]}{\mathbb{E}[S | Z = z_2] - \mathbb{E}[S | Z = z_1]}$$
$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1} | S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y | Z = z_2] - \mathbb{E}[Y | Z = z_1]}{\mathbb{E}[S | Z = z_2] - \mathbb{E}[S | Z = z_1]}$$
$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1} | S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

- What about different policies? Building a new school? Decreasing the tuition price?

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y|Z = z_2] - \mathbb{E}[Y|Z = z_1]}{\mathbb{E}[S|Z = z_2] - \mathbb{E}[S|Z = z_1]} = \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1}|S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y|Z = z_2] - \mathbb{E}[Y|Z = z_1]}{\mathbb{E}[S|Z = z_2] - \mathbb{E}[S|Z = z_1]}$$
$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1}| S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

- It does not disentangles the direct effect and the continuation value.

Understanding the Wald (IV) Estimand

Heckman et al. [2016] show that the Wald Estimand is a weighted average of treatment effects across many subpopulations and many margins of choice.

For an instrument vector Z set at two different values $Z = z_1$ and $Z = z_2$ such that $S(z_2) > S(z_1)$, the Wald Estimand is given by

$$\frac{\mathbb{E}[Y|Z = z_2] - \mathbb{E}[Y|Z = z_1]}{\mathbb{E}[S|Z = z_2] - \mathbb{E}[S|Z = z_1]}$$
$$= \frac{\sum_{j=1}^{\bar{s}} \mathbb{E}[Y_j - Y_{j-1}| S(z_2) \geq j > S(z_1)] \cdot \mathbb{P}[S(z_2) \geq j > S(z_1)]}{\sum_{j=1}^{\bar{s}} \mathbb{P}[S(z_2) \geq j > S(z_1)]}.$$

- It does not disentangles the direct effect and the continuation value. Wald Estimand is hard to interpret.

Thank you!

Contact Information:

Vitor Possebom

E-mail: vitor.possebom@fgv.br

Website: sites.google.com/site/vitorapossebom/

References

I. Bojinov, A. Rambachan, and N. Shephard. Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective. *Quantitative Economics*, 12:pp, 1171–1196, 2021.

J. J. Heckman, J. E. Humphries, and G. Veramendi. Dynamic Treatment Effects. *Journal of Econometrics*, 191(2):pp. 276–292, 2016.

A. Rambachan and N. Shephard. When Do Common Time Series Estimands Hace Nonparametric Causal Meaning? Working Paper. Available at `https://asheshrambachan.github.io/assets/files/arns_commontimeseries_causal.pdf`., Oct. 2021.