Nonparametric Causal Inference in Dynamic Thresholding Designs

Aditya Ghosh

Statistics Ph.D. student, Stanford University

Joint work with Stefan Wager

Motivation

- In healthcare, patients are often monitored over time and an intervention is applied whenever a biomarker crosses a clinical threshold.

Motivation

- In healthcare, patients are often monitored over time and an intervention is applied whenever a biomarker crosses a clinical threshold.
- For example, Iizuka et al. (2021): Observational study from Japan.
  - Patients visit primary-care provider *once a year*.
  - At each visit, measure fasting blood sugar (FBS), BMI etc.
  - If $FBS \geq 110$, the patient is *declared pre-diabetic* and prescribed lifestyle interventions.

Motivation

- In healthcare, patients are often monitored over time and an intervention is applied whenever a biomarker crosses a clinical threshold.
- For example, Iizuka et al. (2021): Observational study from Japan.
  - Patients visit primary-care provider *once a year*.
  - At each visit, measure fasting blood sugar (FBS), BMI etc.
  - If $FBS \geq 110$, the patient is *declared pre-diabetic* and prescribed lifestyle interventions.

Q: What is the **long-term effect** of these lifestyle interventions (or threshold-based treatment policies in general) on health outcomes?

Reaction #1:

- Treatment rule (i.e., the FBS threshold) is a seemingly clean RDD
- Run local linear regression $BMI_{t+1} \sim FBS_t * (FBS_t \geq 110)$?

Reaction #1:

- Treatment rule (i.e., the FBS threshold) is a seemingly clean RDD
- Run local linear regression $BMI_{t+1} \sim FBS_t * (FBS_t \geq 110)$?

⚠ This only captures a short-term effect (and we want long-term)!

Reaction #1:

- Treatment rule (i.e., the FBS threshold) is a seemingly clean RDD
- Run local linear regression $BMI_{t+1} \sim FBS_t * (FBS_t \geq 110)$?

⚠ This only captures a short-term effect (and we want long-term)!

⚠ **Carryover effects:** Actions taken in one period may influence outcomes and treatment assignments in subsequent time periods.

Reaction #1:

- Treatment rule (i.e., the FBS threshold) is a seemingly clean RDD
- Run local linear regression $BMI_{t+1} \sim FBS_t * (FBS_t \geq 110)$?

⚠ This only captures a short-term effect (and we want long-term)!

⚠ **Carryover effects:** Actions taken in one period may influence outcomes and treatment assignments in subsequent time periods.

⚠ **Dynamic treatment switching:** Patients just below the threshold today may cross it and receive treatment in the future (and vice-versa).

Reaction #2:

- Treat this as a dynamic/reinforcement learning (RL) problem.
- RL offers good proxies for the long-run health outcomes

Reaction #2:

- Treat this as a dynamic/reinforcement learning (RL) problem.
- RL offers good proxies for the long-run health outcomes

⚠ Do we need complicated RL algorithm to make progress?

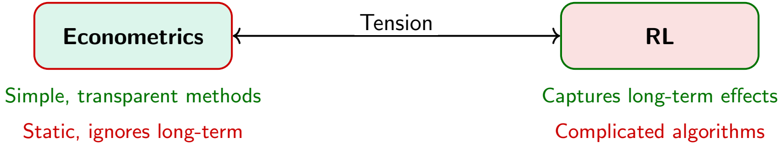
⚠ Accounting for full history → exponential blow-up in dimensionality!

Reaction #2:

- Treat this as a dynamic/reinforcement learning (RL) problem.
- RL offers good proxies for the long-run health outcomes

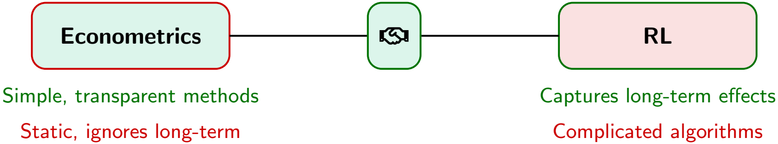
⚠ Do we need complicated RL algorithm to make progress?

⚠ Accounting for full history $\rightarrow$ exponential blow-up in dimensionality!



This Talk

- Re-interpret the standard RD estimand as a **policy-gradient**.
- Use this connection to define an **RD-like estimand** in the dynamic setting.
- Develop a simple method for estimation and inference.

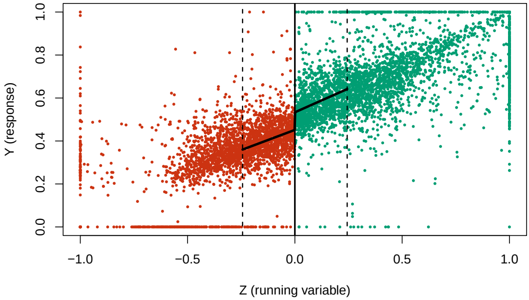


Standard RDD

- Running variable Z decides treatment: $A = \mathbf{1}(Z \geq c)$.
- Causal parameter: CATE at the treatmend threshold c, defined as

$$\tau_c := \mu_1(c+) - \mu_0(c-),$$

where $\mu_a(z) := \mathbb{E}[Y|Z = z, A = a]$ denote conditional response functions.

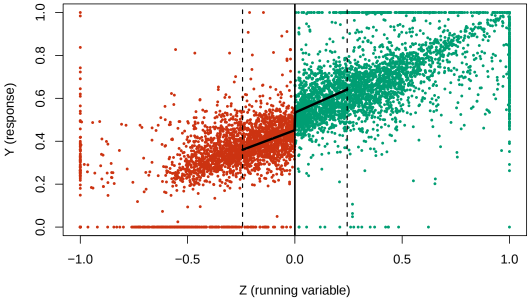


Standard RDD

- Running variable Z decides treatment: $A = \mathbf{1}(Z \geq c)$.
- Causal parameter: CATE at the treatmend threshold c, defined as

$$\tau_c := \mu_1(c+) - \mu_0(c-),$$

where $\mu_a(z) := \mathbb{E}[Y|Z = z, A = a]$ denote conditional response functions.



- $\widehat{\tau}_{\mathrm{LLR}} \leftarrow$ local linear regression of Y on $Z * 1\{Z \geq c\}$.

Dynamic Thresholding Designs

How should we generalize standard RD to the dynamic setting?

Dynamic Thresholding Designs

How should we generalize standard RD to the dynamic setting?

In a multi-period setting, potential outcomes take on a branching structure (Robins, 1986):

- Current outcomes can be affected by past and current treatments
- Current treatment can be affected by past treatments

Dynamic Thresholding Designs

How should we generalize standard RD to the dynamic setting?

In a multi-period setting, potential outcomes take on a branching structure (Robins, 1986):

- Current outcomes can be affected by past and current treatments
- Current treatment can be affected by past treatments

⚠ Attempting to define RD parameters by conditioning on a time-varying running variable would suffer from an exponential curse of dimension with the horizon. Is there a better way?

Related Work

- There is a large and active literature on RD designs (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012; Calonico et al., 2014; Armstrong and Kolesár, 2018), but largely ignores dynamics.

Related Work

- There is a large and active literature on RD designs (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012; Calonico et al., 2014; Armstrong and Kolesár, 2018), but largely ignores dynamics.
- Cellini, Ferreira & Rothstein (2010) apply RD designs to study effect of funding on schools via bond elections. They consider dynamics, but rely on a strict parametric model

$$\underbrace{Y_{it}}_{outcome} = \sum_{t' \leq t} \theta_{t'-t} \underbrace{A_{it}}_{treatment} + \underbrace{\varepsilon_{it}}_{noise}$$

Related Work

- There is a large and active literature on RD designs (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012; Calonico et al., 2014; Armstrong and Kolesár, 2018), but largely ignores dynamics.
- Cellini, Ferreira & Rothstein (2010) apply RD designs to study effect of funding on schools via bond elections. They consider dynamics, but rely on a strict parametric model

$$\underset{outcome}{Y_{it}} = \sum_{t' \leq t} \theta_{t'-t} \underset{treatment}{A_{it}} + \underset{noise}{\varepsilon_{it}}$$

  Hsu & Shen (2024) show that this approach cannot accommodate deviations from this outcome model.
- Hsu & Shen (2024) propose an alternative approach, but they require a conditional mean independence assumption, which essentially precludes existence of time-varying latent confounders.

Viewing Regression Discontinuities as Marginal Policy Effects

RDs as Marginal Policy Effects

- An RDD with cutoff c induces a treatment policy $\pi_c(z) = 1\{z \geq c\}$, with associated policy value

$$V(\pi_c) = \mathbb{E}[Y_i(\pi_c(Z_i))] = \mathbb{E}_{\pi_c}[Y_i].$$

RDs as Marginal Policy Effects

- An RDD with cutoff c induces a treatment policy $\pi_c(z) = 1\{z \geq c\}$, with associated policy value

$$V(\pi_c) = \mathbb{E}[Y_i(\pi_c(Z_i))] = \mathbb{E}_{\pi_c}[Y_i].$$

- Using calculus, we can verify that

$$\underbrace{(\mu_1(c) - \mu_0(c))}_{\text{standard RD parameter}} = \underbrace{\frac{\partial}{\partial c} V(\pi_c)}_{\text{policy gradient for outcome}} \Big/ \underbrace{\frac{\partial}{\partial c} \mathbb{E}_{\pi_c}[A_i]}_{\text{policy gradient for treatment}}$$

$$\begin{aligned}
\frac{\partial}{\partial c} V(\pi_c) &= \frac{\partial}{\partial c} \int \mathbb{E}_{\pi_c}[Y \mid Z = z]\, f(z)\, dz \\
&= \frac{\partial}{\partial c} \left(\int_c^{\infty} \mu_1(z)\, f(z)\, dz + \int_{-\infty}^{c} \mu_0(z)\, f(z)\, dz \right) \\
&= -(\mu_1(c) - \mu_0(c)) f(c) = (\mu_1(c) - \mu_0(c)) \frac{\partial}{\partial c} \mathbb{E}_{\pi_c}[A_i]
\end{aligned}$$

RDs as Marginal Policy Effects

- An RDD with cutoff c induces a treatment policy $\pi_c(z) = 1\{z \geq c\}$, with associated policy value

$$V(\pi_c) = \mathbb{E}[Y_i(\pi_c(Z_i))] = \mathbb{E}_{\pi_c}[Y_i].$$

- Using calculus, we can verify that

$$\underbrace{(\mu_1(c) - \mu_0(c))}_{\text{standard RD parameter}} = \underbrace{\frac{\partial}{\partial c} V(\pi_c)}_{\text{policy gradient for outcome}} \Big/ \underbrace{\frac{\partial}{\partial c} \mathbb{E}_{\pi_c}[A_i]}_{\text{policy gradient for treatment}}$$

✔ Thus, the RD parameter **is** a marginal policy effect from lowering the treatment-eligibility cutoff (Carneiro, Heckman, and Vytlacil, 2010).

♻ This perspective generalizes better to dynamic settings.

RDs as Marginal Policy Effects?

Is the RD parameter really *always* a marginal policy effect?

RDs as Marginal Policy Effects?

Is the RD parameter really *always* a marginal policy effect?

Empirical example	**Running variable**	**Cutoff**	**MPE?**
Incumbency advantage (Lee, 2008)	Two-party vote share	50%	?
Effects of summer school on future grade (Jacob & Lefgren, 2004)	Test score	Pass-fail cutoff	?
Effect of antiretrovial therapy on HIV-positive patients (Bor et al., 2017)	Immune cell count	Clinical threshold	?

RDs as Marginal Policy Effects?

Is the RD parameter really *always* a marginal policy effect?

Empirical example	**Running variable**	**Cutoff**	**MPE?**
Incumbency advantage (Lee, 2008)	Two-party vote share	50%	No
Effects of summer school on future grade (Jacob & Lefgren, 2004)	Test score	Pass-fail cutoff	?
Effect of antiretrovial therapy on HIV-positive patients (Bor et al., 2017)	Immune cell count	Clinical threshold	?

- Implicit in our derivation is that changing the cutoff is a meaningful policy intervention

RDs as Marginal Policy Effects?

Is the RD parameter really *always* a marginal policy effect?

Empirical example	**Running variable**	**Cutoff**	**MPE?**
Incumbency advantage (Lee, 2008)	Two-party vote share	50%	No
Effects of summer school on future grade (Jacob & Lefgren, 2004)	Test score	Pass-fail cutoff	Maybe?
Effect of antiretrovial therapy on HIV-positive patients (Bor et al., 2017)	Immune cell count	Clinical threshold	?

- Implicit in our derivation is that changing the cutoff is a meaningful policy intervention, and that the running variable is exogenous to the policy.

RDs as Marginal Policy Effects?

Is the RD parameter really *always* a marginal policy effect?

Empirical example	**Running variable**	**Cutoff**	**MPE?**
Incumbency advantage (Lee, 2008)	Two-party vote share	50%	No
Effects of summer school on future grade (Jacob & Lefgren, 2004)	Test score	Pass-fail cutoff	Maybe?
Effect of antiretrovial therapy on HIV-positive patients (Bor et al., 2017)	Immune cell count	Clinical threshold	Plausibly Yes

- Implicit in our derivation is that changing the cutoff is a meaningful policy intervention, and that the running variable is exogenous to the policy.
- We'll focus on the last kind: RD parameter $\leftrightarrow$ policy gradient

Marginal Policy Effects in
Dynamic Thresholding Designs

- At time $t = 0, 1, \ldots, T$, we observe running variable Z_{it}, take action A_{it} and observe an outcome Y_{it}, $i = 1, \ldots, n$.
- We collect data under a thresholding policy π_c: $A_{it} = \mathbf{1}\{Z_{it} \geq c\}$.

Dynamics Thresholding Designs: A General Model

- At time $t = 0, 1, \dots, T$, we observe running variable Z_{it}, take action A_{it} and observe an outcome Y_{it}, $i = 1, \dots, n$.
- We collect data under a thresholding policy π_c: $A_{it} = \mathbf{1}\{Z_{it} \geq c\}$.
- We allow for general temporal propagation of causal effects (Robins, 1986)

$$\mathbb{P}_{\pi_c}\Big[Z_{i0}, Y_{i0}, \dots, Z_{iT}, Y_{iT}\Big] = \prod_{t=0}^{T} \mathbb{P}\Big[Z_{it} \mid S_{it}\Big]\, \mathbb{P}\Big[Y_{it} \mid S_{it}, Z_{it}, A_{it} = \mathbf{1}\{Z_{it} \geq c\}\Big]$$

where $S_{it} = \{Z_{it'}, A_{it'}, Y_{it'}\}_{t' < t}$ is history up to time t $(S_{i0} = \emptyset)$.

- We consider both finite-horizon $(T < \infty)$ and infinite-horizon $(T = \infty)$.

The Target: Dynamic Marginal Policy Effect

- We are interested in understanding how our actions drive total lifetime discounted rewards,

$$V(\pi_c) := \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, Y_{it}\right].$$

with discount rate $0 < \gamma \leq 1$ (with $\gamma < 1$ if $T = \infty$).

- We also get a (discounted) treatment intensity

$$V^A(\pi_c) := \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, A_{it}\right].$$

The Target: Dynamic Marginal Policy Effect

- We are interested in understanding how our actions drive total lifetime discounted rewards,

$$V(\pi_c) := \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, Y_{it}\right].$$

with discount rate $0 < \gamma \leq 1$ (with $\gamma < 1$ if $T = \infty$).

- We also get a (discounted) treatment intensity

$$V^A(\pi_c) := \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, A_{it}\right].$$

- Our previous discussion of the marginal policy effect suggests defining

$$\tau_{\mathrm{RD}} := \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

The Dynamic Marginal Policy Effect

$$V(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, Y_{it}\right], \quad V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, A_{it}\right].$$

Causal parameter of interest for the Dynamic RD setting:

$$\tau_{\text{RD}} = \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

Captures the marginal welfare gain per unit increase in treatment exposure.

- ✓ Reduces to the standard RD estimand in the single-period setting.

The Dynamic Marginal Policy Effect

$$V(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, Y_{it}\right], \quad V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, A_{it}\right].$$

Causal parameter of interest for the Dynamic RD setting:

$$\tau_{\text{RD}} = \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

Captures the marginal welfare gain per unit increase in treatment exposure.

- ✓ Reduces to the standard RD estimand in the single-period setting.
- ✓ Naturally captures long-term effects.

The Dynamic Marginal Policy Effect

$$V(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, Y_{it}\right], \quad V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, A_{it}\right].$$

Causal parameter of interest for the Dynamic RD setting:

$$\tau_{\text{RD}} = \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

Captures the marginal welfare gain per unit increase in treatment exposure.

- Reduces to the standard RD estimand in the single-period setting.
- Naturally captures long-term effects.
- Enables direct comparison with treatment costs for a cost-benefit analysis.

The Dynamic Marginal Policy Effect

$$V(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, Y_{it}\right], \quad V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t \, A_{it}\right].$$

Causal parameter of interest for the Dynamic RD setting:

$$\tau_{\text{RD}} = \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

Captures the marginal welfare gain per unit increase in treatment exposure.

- ✔ Reduces to the standard RD estimand in the single-period setting.
- ✔ Naturally captures long-term effects.
- ✔ Enables direct comparison with treatment costs for a cost-benefit analysis.
- ? Can we estimate this without a full-blown RL algorithm?

The Dynamic Marginal Policy Effect

$$V(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, Y_{it}\right], \quad V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t\, A_{it}\right].$$

Causal parameter of interest for the Dynamic RD setting:

$$\tau_{\mathrm{RD}} = \frac{\partial}{\partial c} V(\pi_c) \Big/ \frac{\partial}{\partial c} V^A(\pi_c).$$

Captures the marginal welfare gain per unit increase in treatment exposure.

- ✓ Reduces to the standard RD estimand in the single-period setting.
- ✓ Naturally captures long-term effects.
- ✓ Enables direct comparison with treatment costs for a cost-benefit analysis.
- ? Can we estimate this without a full-blown RL algorithm?
  - Yes! Using a carefully designed local linear regression.

Theory for Estimation and Inference

Key Assumptions

- Causal queries can be resolved using the g-formula of Robins (1986):

$$\mathbb{P}_{\pi_c}\Big[Z_{i0}, Y_{i0}, \ldots, Z_{iT}, Y_{iT}\Big] = \prod_{t=0}^{T} \mathbb{P}\Big[Z_{it} \mid S_{it}\Big]\,\mathbb{P}\Big[Y_{it} \mid S_{it}, Z_{it}, A_{it} = \mathbf{1}\{Z_{it} \geq c\}\Big]$$

  where $S_{it} = \{Z_{it'}, A_{it'}, Y_{it'}\}_{t'<t}$ is history up to time t ($S_{i0} = \emptyset$).

- Conditional on the history S_{it}, the running variable Z_{it} has a density $f_t(\cdot \mid S_{it})$ continuous and strictly positive at c.

A Policy Gradient Theorem for the Dynamic RD

- Dynamic analog of the conditional response functions are the *Q*-functions:

$$Q_{c,t}(s, z, a) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T-t} \gamma^t Y_{it} \,\middle|\, S_{it} = s, Z_{it} = z, A_{it} = a\right].$$

A Policy Gradient Theorem for the Dynamic RD

- Dynamic analog of the conditional response functions are the *Q-functions*:

$$Q_{c,t}(s, z, a) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T-t} \gamma^t Y_{it} \,\middle|\, S_{it} = s, Z_{it} = z, A_{it} = a\right].$$

Theorem. Under our general dynamic model (and regularity assumptions, e.g., the Q-functions are continuous in z at the threshold c),

$$-\frac{\partial}{\partial c} V(\pi_c) = \sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{it}, c, 1) - Q_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right],$$

provided that either $\gamma < 1$ or $T < \infty$. A similar result holds for $V^A(\pi_c)$.

Recall: $V(\pi_c)$ and $V^A(\pi_c)$ respectively denote the net-present expected welfare and treatment frequency: $V(\pi_c) := \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t Y_{it}\right]$, and $V^A(\pi_c) = \mathbb{E}_{\pi_c}\left[\sum_{t=0}^{T} \gamma^t A_{it}\right]$.

Back to the MPE

The policy-gradient theorem yields the following characterization of the dynamic MPE parameter

$$\tau_{\mathsf{RD}} = \frac{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{it}, c, 1) - Q_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q^A_{c,t}(S_{it}, c, 1) - Q^A_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}$$

Back to the MPE

The policy-gradient theorem yields the following characterization of the dynamic MPE parameter

$$\tau_{\mathrm{RD}} = \frac{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{it}, c, 1) - Q_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q^A_{c,t}(S_{it}, c, 1) - Q^A_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}$$

✔ Naturally incorporates carryover effects and dynamic treatment switching.

- The Q-functions automatically incorporate carryover effects.
- The density $f_t(c \mid S_{it})$ combined with the discount factor γ^t accounts for the dynamically evolving frequency of units near the threshold.

Back to the MPE

The policy-gradient theorem yields the following characterization of the dynamic MPE parameter

$$\tau_{\mathrm{RD}} = \frac{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{it}, c, 1) - Q_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q^A_{c,t}(S_{it}, c, 1) - Q^A_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}$$

- ✔ Naturally incorporates carryover effects and dynamic treatment switching.
- ✔ Enables cost-benefit analysis via the interpretation $\tau_{\mathrm{RD}} = \frac{\partial}{\partial c} V(\pi_c) \big/ \frac{\partial}{\partial c} V^A(\pi_c)$. (Increase c if $\tau_{\mathrm{RD}} >$ cost per treatment)
- ✔ Reduces to the standard RD estimand in the single-period setting.

Back to the MPE

The policy-gradient theorem yields the following characterization of the dynamic MPE parameter

$$\tau_{\mathsf{RD}} = \frac{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{it}, c, 1) - Q_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q^A_{c,t}(S_{it}, c, 1) - Q^A_{c,t}(S_{it}, c, 0)\right) f_t(c \mid S_{it})\right]}$$

- Naturally incorporates carryover effects and dynamic treatment switching.
- Enables cost-benefit analysis via the interpretation $\tau_{\mathsf{RD}} = \frac{\partial}{\partial c} V(\pi_c) \big/ \frac{\partial}{\partial c} V^A(\pi_c)$. (Increase c if $\tau_{\mathsf{RD}} >$ cost per treatment)
- Reduces to the standard RD estimand in the single-period setting.
- Despite the apparent complexity, particularly from the growing-dimensional history S_{it}, estimation and inference remains tractable via the ratio of carefully designed LLRs that account for the temporal structure.

Estimation via Local Linear Regression

Define the discounted sum of future outcomes and treatment assignments:

$$\Gamma^Y_{i,t} = \sum_{j=0}^{T-t} \gamma^j \, Y_{i,t+j}, \quad \text{and} \quad \Gamma^A_{i,t} = \sum_{j=0}^{T-t} \gamma^j \, A_{i,t+j}. \tag{1}$$

Note,

$$\mathbb{E}_{\pi_c}\left[\Gamma^Y_{i,t} \mid S_{i,t},\, Z_{i,t},\, A_{i,t}\right] = Q_{\pi_c}(S_{i,t},\, Z_{i,t},\, A_{i,t}),$$

$$\mathbb{E}_{\pi_c}\left[\Gamma^A_{i,t} \mid S_{i,t},\, Z_{i,t},\, A_{i,t}\right] = Q^A_{\pi_c}(S_{i,t},\, Z_{i,t},\, A_{i,t}),$$

Recall:

$$\tau_{\text{RD}} = \frac{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q_{c,t}(S_{i,t}, c, 1) - Q_{c,t}(S_{i,t}, c, 0)\right) f_t(c \mid S_{i,t})\right]}{\sum_{t=0}^{T} \gamma^t \, \mathbb{E}_{\pi_c}\left[\left(Q^A_{c,t}(S_{i,t}, c, 1) - Q^A_{c,t}(S_{i,t}, c, 0)\right) f_t(c \mid S_{i,t})\right]}$$

Define the discounted sum of future outcomes and treatment assignments:

$$\Gamma_{i,t}^{Y} = \sum_{j=0}^{T-t} \gamma^{j}\, Y_{i,t+j}, \quad \text{and} \quad \Gamma_{i,t}^{A} = \sum_{j=0}^{T-t} \gamma^{j}\, A_{i,t+j}. \tag{1}$$

Run the following twice-discounted local linear regressions (with bandwidth h):

$$\widehat{\tau}_Y(h) := e_2^\top \underset{\theta=(\alpha,\tau,\beta_0,\beta_1)}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T} \boxed{\gamma^t}\, K\left(\frac{|Z_{i,t}-c|}{h}\right)\left(\boxed{\Gamma_{i,t}^{Y}} - r(Z_{i,t})^\top\theta\right)^2,$$

$$\widehat{\tau}_A(h) := e_2^\top \underset{\theta=(\alpha,\tau,\beta_0,\beta_1)}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T} \boxed{\gamma^t}\, K\left(\frac{|Z_{i,t}-c|}{h}\right)\left(\boxed{\Gamma_{i,t}^{A}} - r(Z_{i,t})^\top\theta\right)^2,$$

where $r(z) = (1, \mathbf{1}\{z \geq c\}, (z-c)_+, (z-c)_-)$. Finally, $\widehat{\tau}_{\mathrm{RD}} := \widehat{\tau}_Y(h)\,/\,\widehat{\tau}_A(h)$.

Looks like standard LLR except: Discounted outcomes to capture long-term effects and temporal weighting to mirror the policy gradient result.

Estimation via Local Linear Regression

Define the discounted sum of future outcomes and treatment assignments:

$$\Gamma^{Y}_{i,t} = \sum_{j=0}^{T-t} \gamma^{j}\, Y_{i,t+j}, \quad \text{and} \quad \Gamma^{A}_{i,t} = \sum_{j=0}^{T-t} \gamma^{j}\, A_{i,t+j}. \tag{1}$$

Run the following twice-discounted local linear regressions (with bandwidth h):

$$\widehat{\tau}_Y(h) := e_2^\top \operatorname*{argmin}_{\theta=(\alpha,\tau,\beta_0,\beta_1)} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T} \boxed{\gamma^t} K\left(\frac{|Z_{i,t}-c|}{h}\right)\left(\boxed{\Gamma^{Y}_{i,t}} - r(Z_{i,t})^\top\theta\right)^2,$$

$$\widehat{\tau}_A(h) := e_2^\top \operatorname*{argmin}_{\theta=(\alpha,\tau,\beta_0,\beta_1)} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T} \boxed{\gamma^t} K\left(\frac{|Z_{i,t}-c|}{h}\right)\left(\boxed{\Gamma^{A}_{i,t}} - r(Z_{i,t})^\top\theta\right)^2,$$

where $r(z) = (1, \mathbf{1}\{z \geq c\}, (z-c)_+, (z-c)_-)$. Finally, $\widehat{\tau}_{\mathrm{RD}} := \widehat{\tau}_Y(h)\,/\,\widehat{\tau}_A(h)$.

! This is computationally equivalent to a fuzzy RD estimator (Imbens & Lemieux, 2008)

Define the discounted sum of future outcomes and treatment assignments:

$$\Gamma_{i,t}^{Y} = \sum_{j=0}^{T-t} \gamma^j \, Y_{i,t+j}, \quad \text{and} \quad \Gamma_{i,t}^{A} = \sum_{j=0}^{T-t} \gamma^j \, A_{i,t+j}. \tag{1}$$

Run the following twice-discounted local linear regressions (with bandwidth h):

$$\widehat{\tau}_Y(h) := e_2^\top \operatorname*{argmin}_{\theta=(\alpha,\tau,\beta_0,\beta_1)} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=0}^{T} \boxed{\gamma^t} K\left(\frac{|Z_{i,t}-c|}{h}\right) \left(\boxed{\Gamma_{i,t}^{Y}} - r(Z_{i,t})^\top \theta\right)^2,$$

$$\widehat{\tau}_A(h) := e_2^\top \operatorname*{argmin}_{\theta=(\alpha,\tau,\beta_0,\beta_1)} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=0}^{T} \boxed{\gamma^t} K\left(\frac{|Z_{i,t}-c|}{h}\right) \left(\boxed{\Gamma_{i,t}^{A}} - r(Z_{i,t})^\top \theta\right)^2,$$

where $r(z) = (1, \mathbf{1}\{z \geq c\}, (z-c)_+, (z-c)_-)$. Finally, $\widehat{\tau}_{\mathrm{RD}} := \widehat{\tau}_Y(h) \, / \, \widehat{\tau}_A(h)$.

The infinite-horizon ($T = \infty$) case is similar, except there we replace the $\Gamma_{i,t}^{Y}$ in (1) with $\Gamma_{i,t}^{Y}(\ell) = \sum_{j=0}^{\ell-1} \gamma^j \, Y_{i,t+j}$ (similarly for $\Gamma_{i,t}^{A}$), where ℓ is a truncation window.

Theory

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.

Theory

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.
- We derive the limiting distribution of $\widehat{\tau}_{\mathrm{RD}}$ and explicitly characterize its bias and variance. Similar to classical RD, it has $O(h_n^2)$ curvature bias.

Theory

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.
- We derive the limiting distribution of $\widehat{\tau}_{\mathrm{RD}}$ and explicitly characterize its bias and variance. Similar to classical RD, it has $O(h_n^2)$ curvature bias.
- When $h_n = O(n^{-1/5})$, the estimator $\widehat{\tau}_{\mathrm{RD}}$ achieves the standard nonparametric rate of $n^{-2/5}$ under appropriate smoothness conditions.

Theory

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.
- We derive the limiting distribution of $\widehat{\tau}_{\mathrm{RD}}$ and explicitly characterize its bias and variance. Similar to classical RD, it has $O(h_n^2)$ curvature bias.
- When $h_n = O(n^{-1/5})$, the estimator $\widehat{\tau}_{\mathrm{RD}}$ achieves the standard nonparametric rate of $n^{-2/5}$ under appropriate smoothness conditions.

$$\sqrt{nh}\left(\widehat{\tau}_{\mathsf{RD}} - \tau_{\mathsf{RD}} - \frac{1}{2}h^2\xi_1\frac{\Delta\mu_Y''(c) - \tau_{\mathsf{RD}}\Delta\mu_A''(c)}{\Delta\mu_A(c)}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RD}}\right)$$

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.
- We derive the limiting distribution of $\widehat{\tau}_{\mathrm{RD}}$ and explicitly characterize its bias and variance. Similar to classical RD, it has $O(h_n^2)$ curvature bias.
- When $h_n = O(n^{-1/5})$, the estimator $\widehat{\tau}_{\mathrm{RD}}$ achieves the standard nonparametric rate of $n^{-2/5}$ under appropriate smoothness conditions.

$$\sqrt{nh}\left(\widehat{\tau}_{\mathrm{RD}} - \tau_{\mathrm{RD}} - \frac{1}{2}h^2\xi_1\frac{\Delta\mu''_Y(c) - \tau_{\mathrm{RD}}\Delta\mu''_A(c)}{\Delta\mu_A(c)}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RD}}\right)$$

- We construct asymptotically valid CIs using sandwich standard errors.

- $\widehat{\tau}_{\mathrm{RD}} \to_p \tau_{\mathrm{RD}}$ as $n \to \infty$ provided the bandwidth $h_n \to 0$ such that $nh_n \to \infty$.
- We derive the limiting distribution of $\widehat{\tau}_{\mathrm{RD}}$ and explicitly characterize its bias and variance. Similar to classical RD, it has $O(h_n^2)$ curvature bias.
- When $h_n = O(n^{-1/5})$, the estimator $\widehat{\tau}_{\mathrm{RD}}$ achieves the standard nonparametric rate of $n^{-2/5}$ under appropriate smoothness conditions.

$$\sqrt{nh}\left(\widehat{\tau}_{\mathrm{RD}} - \tau_{\mathrm{RD}} - \frac{1}{2}h^2\xi_1\frac{\Delta\mu_Y''(c) - \tau_{\mathrm{RD}}\Delta\mu_A''(c)}{\Delta\mu_A(c)}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RD}}\right)$$

- We construct asymptotically valid CIs using sandwich standard errors.
- Using time fixed-effects doesn't change the target and improves variance.

Numerical experiments to illustrate the proposed method

Back to the Motivating Example

- Iizuka et al. (2021) report data from an observational study in Japan:
  - Patients visit primary-care provider *once a year*.
  - At each visit, doctor measures fasting blood sugar (FBS), BMI etc.
  - If $FBS \geq 110$, the patient is *declared pre-diabetic* and prescribed lifestyle interventions.

Question (original, qualitative): What is the **long-term effect** of these lifestyle interventions on health outcomes?

Back to the Motivating Example

- Iizuka et al. (2021) report data from an observational study in Japan:
  - Patients visit primary-care provider *once a year*.
  - At each visit, doctor measures fasting blood sugar (FBS), BMI etc.
  - If $FBS \geq 110$, the patient is *declared pre-diabetic* and prescribed lifestyle interventions.

Question (original, qualitative): What is the **long-term effect** of these lifestyle interventions on health outcomes?

Question (proposed interpretation): What is the **marginal policy effect** of reducing the FBS cutoff for declaring a patient pre-diabetic?

An Autoregressive Simulator

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$
\begin{aligned}
Z_{t+1} &= 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t, \\
A_t &= \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},
\end{aligned}
$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0, 1)$ for $t = 0, 1, \ldots, 11$.

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$
\begin{aligned}
Z_{t+1} &= 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t, \\
A_t &= \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},
\end{aligned}
$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0,1)$ for $t = 0, 1, \ldots, 11$.

- We contrast our method with the following baseline methods.

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$Z_{t+1} = 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t,$$
$$A_t = \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0, 1)$ for $t = 0, 1, \ldots, 11$.

- We contrast our method with the following baseline methods.
- Baseline #1: Standard local linear reg (LLR). Targets short-term effect $\tau_{\text{partial eq.}} = \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c+] - \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c-] = \tau(c - 100) = 1$.

An Autoregressive Simulator

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$Z_{t+1} = 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t,$$
$$A_t = \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0, 1)$ for $t = 0, 1, \ldots, 11$.

- We contrast our method with the following baseline methods.
- Baseline #1: Standard local linear reg (LLR). Targets short-term effect $\tau_{\text{partial eq.}} = \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c+] - \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c-] = \tau(c - 100) = 1$.
- By Monte Carlo, we verify that $\tau_{\text{RD}} = 4.308$ for $\gamma = 1$ (equivalent to mean fasting blood sugar), 4x the partial equilibrium effect!

An Autoregressive Simulator

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$Z_{t+1} = 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t,$$
$$A_t = \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0,1)$ for $t = 0, 1, \ldots, 11$.

- We contrast our method with the following baseline methods.
- Baseline #1: Standard local linear reg (LLR). Targets short-term effect $\tau_{\text{partial eq.}} = \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c+] - \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c-] = \tau(c - 100) = 1$.
- By Monte Carlo, we verify that $\tau_{\text{RD}} = 4.308$ for $\gamma = 1$ (equivalent to mean fasting blood sugar), 4x the partial equilibrium effect!
- Baseline #2: A naive long-run LLR ($\sum_{t=0}^{12} \gamma^t Y_{it}$ on Z_{i0}, and $\sum_{t=0}^{12} \gamma^t A_{it}$ on Z_{i0}, then take the ratio).

An Autoregressive Simulator

Consider a simple simulation setting motivated by Iizuka et al. (2021):

$$Z_{t+1} = 1 + Z_t - \rho(Z_t - 100) - \tau A_t (Z_t - 100)_+ + 4\varepsilon_t,$$
$$A_t = \mathbf{1}\{Z_t \geq c\}, \quad \text{and} \quad Y_t = -Z_{t+1},$$

where $c = 110$, $\rho = 0.1$, $\tau = 0.1$, and $\varepsilon_t \sim \mathcal{N}(0, 1)$ for $t = 0, 1, \ldots, 11$.

- We contrast our method with the following baseline methods.
- Baseline #1: Standard local linear reg (LLR). Targets short-term effect $\tau_{\text{partial eq.}} = \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c+] - \mathbb{E}_{\pi_c}[Y_t \mid Z_t = c-] = \tau(c - 100) = 1$.
- By Monte Carlo, we verify that $\tau_{\text{RD}} = 4.308$ for $\gamma = 1$ (equivalent to mean fasting blood sugar), 4x the partial equilibrium effect!
- Baseline #2: A naive long-run LLR ($\sum_{t=0}^{12} \gamma^t Y_{it}$ on Z_{i0}, and $\sum_{t=0}^{12} \gamma^t A_{it}$ on Z_{i0}, then take the ratio). This does not exactly target τ_{RD} (unless stationary), and usually produces CIs wider than our method.

Autoregressive Simulator: Results

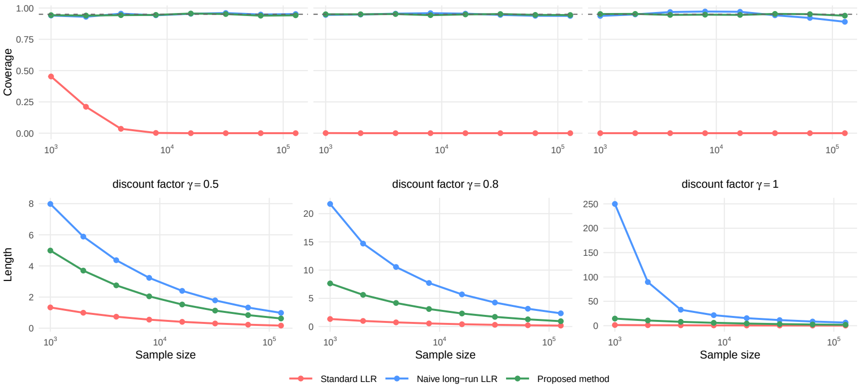


Figure 1: Empirical coverage and average length of 95% CIs based on standard LLR, a naive long-run LLR of $\sum_{t=0}^{T} \gamma^t Y_{it}$ and $\sum_{t=0}^{T} \gamma^t A_{it}$ on Z_{i0}, and our proposed method.

A More Realistic Simulator

- FDA-approved Type-1 Diabetes simulator (Man et al., 2014)
- CGM (continuous glucose monitoring) observed at every 30 min
- A bolus dose (insulin injection) is given every time the CGM reading crosses a threshold before taking a meal (meals are random, endogeneous).

$$A_{i,t} = \mathrm{Meal}_{i,t} \cdot \mathbf{1}\{\mathrm{CGM}_{i,t} \geq c\}.$$

- Reward: Negative risk index (Clarke and Kovatchev, 2009)

$$Y_{i,t} = -10\left(1.509\left(\left(\log \mathrm{CGM}_{i,t+1}\right)^{1.084} - 5.381\right)\right)^2.$$

- Threshold $c = 150$ when the simulator was built (now revised to 180).

The Type-1 Diabetes Simulator (Contd.)

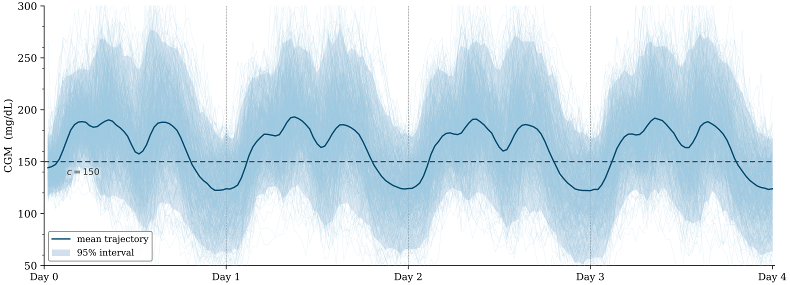


Figure 2: Sample trajectories of the CGM reading for $n = 500$ patients under the threshold-based poicy with threshold $c = 150$ mg/dL, observed for 4 days.

- Running variable: $Z_{i,t} = \text{CGM}_{i,t}$ if $\text{Meal}_{i,t} = 1$, $Z_{i,t} = -\infty$ otherwise. With this convention, $A_{i,t} = \text{Meal}_{i,t} \cdot \mathbf{1}\{\text{CGM}_{i,t} \geq c\} = \mathbf{1}\{Z_{i,t} \geq c\}$.

Oracle results for the Type-1 Diabetes Simulator

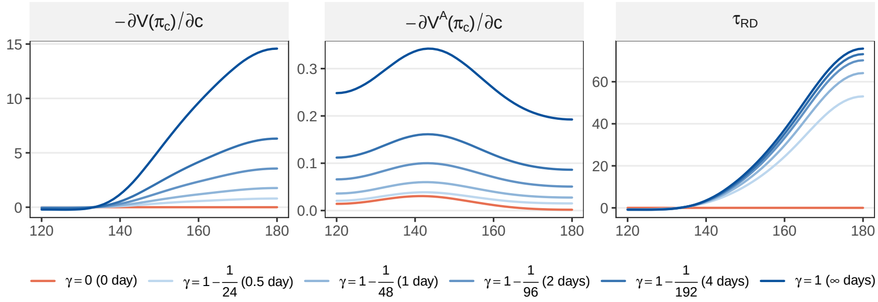


Figure 3: Oracle estimates across different discounting factors (with the effective horizons (in days) shown in parentheses). The first two panels plot the derivatives of the outcome and treatment value functions against the threshold. The third panel shows that the marginal policy effect τ_{RD}, which increases gradually from $\gamma = 0$ (no discounting, short-term reward) to $\gamma = 1$ (infinite horizon, long-term reward).

Evaluating CIs

γ	Eff. horizon	Standard LLR		Naive method		Proposed method	
		coverage	width	coverage	width	coverage	width
$1-\frac{1}{24}$	0.5 day	0.0%	0.71	94.7%	20.86	96.1%	15.96
$1-\frac{1}{48}$	1 day	0.0%	0.71	94.8%	33.22	96.1%	18.33
$1-\frac{1}{96}$	2 days	0.0%	0.71	93.9%	50.83	96.9%	21.48
$1-\frac{1}{192}$	4 days	0.0%	0.71	95.0%	72.18	96.0%	27.10
1	∞	0.0%	0.71	97.5%	139.04	95.8%	38.24

Table 1: Empirical coverage and average width of 95% confidence intervals for τ_{RD} for the threshold $c = 150$ using the Type-1 diabetes simulator. Results are aggregated across 1000 replications, each with $n = 1000$ independent trajectories, $T = 4$ days.

Oracle results for the Type-1 Diabetes Simulator (Contd.)

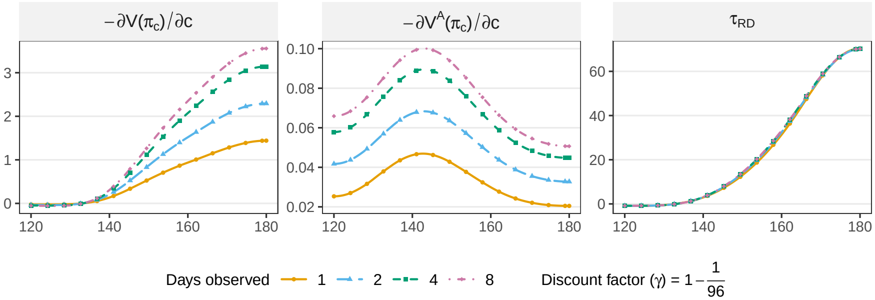


Figure 4: Oracle estimates across observation horizons of 1, 2, 4, and 8 days. The first two panels plot the derivatives of the outcome and treatment value functions against the threshold; both vary systematically with the horizon. The third panel shows that the marginal policy effect τ_{RD}, defined as the previous two, remains stable across all four horizons.

Discussion

- RDDs are among the most widely used observational study designs in social sciences, and dynamic settings are becoming increasingly common.
- In many applications, we would like to combine the thresholding structure of RDDs with general dynamics and long-term treatment effects.

Discussion

- RDDs are among the most widely used observational study designs in social sciences, and dynamic settings are becoming increasingly common.
- In many applications, we would like to combine the thresholding structure of RDDs with general dynamics and long-term treatment effects.
- We show that a carefully designed LLR can recover long-term effects of threshold-based policies, bridging standard RD and dynamic models *without materially increasing algorithm complexity over standard RD analyses!*

Discussion

- RDDs are among the most widely used observational study designs in social sciences, and dynamic settings are becoming increasingly common.
- In many applications, we would like to combine the thresholding structure of RDDs with general dynamics and long-term treatment effects.
- We show that a carefully designed LLR can recover long-term effects of threshold-based policies, bridging standard RD and dynamic models *without materially increasing algorithm complexity over standard RD analyses!*
- This highlights a broader theme where modeling formalisms from reinforcement learning (e.g., Markov decision processes) can *analytically justify* and *extend the scope* of standard causal inference methods.

Draft is available on ArXiv.

Thank You!