# Regression Discontinuity And Kink Designs With Multiple Running Variables – A Design Having Potential For Health System Research

Alden Cheng

NBER Postdoctoral Research Associate
Gies Health Initiatives Research Scholar

November 2, 2023

# Regression Discontinuity (RD) Designs

- One of the most promising tools that have emerged out of the "credibility revolution" (Angrist and Pischke 2010).
  - Gold standard for studying causal relationships is RCT, but often infeasible.
  - Among other methods to study causal relationships, RD designs are arguably the most transparent, and best mimics an RCT.
- Simple example to illustrate: effect of college-going on long-run health outcomes.
  - Can run a regression of health on college-going, controlling for characteristics, but we may be worried about omitted variables bias.
  - Suppose college admission depends on whether students' test scores pass a threshold.
  - Compare outcomes of students who barely passed versus barely missed the threshold.
  - Convincing because we expect the two sets of students to be very similar on both observables *and unobservables.*

# Setup

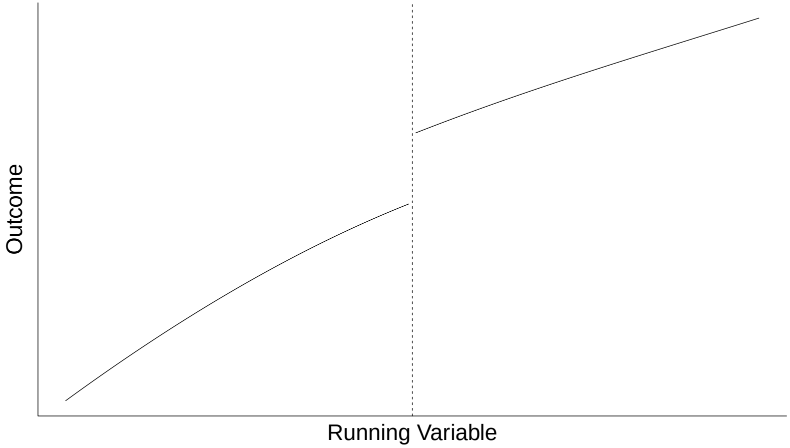
- Notation:
  - Treatment $W_i$, outcome $Y_i$.
  - Potential outcomes: $Y_i(1)$, $Y_i(0)$. $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$.
  - Running variable $X_i$ (assume normalized so that threshold is zero).
- Sharp RD: $W_i = \mathbb{I}[X_i \geq 0]$.
  - Estimand:

$$\lim_{x \to 0^+} \mathbb{E}[Y_i | X_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i | X_i = x]$$

- Fuzzy RD: $\lim_{x \to 0^+} Pr(W_i = 1 | X_i = x) > \lim_{x \to 0^-} Pr(W_i = 1 | X_i = x)$.
  - Estimand:

$$\frac{\lim_{x \to 0^+} \mathbb{E}[Y_i | X_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i | X = x]}{\lim_{x \to 0^+} Pr(W_i = 1 | X_i = x) - \lim_{x \to 0^-} Pr(W_i = 1 | X_i = x)}.$$

# Visual Illustration (Sharp RD)



# Testing the Identification Assumptions

- The most primitive assumption is that the conditional expectation functions of potential outcomes are continuous at the threshold: for $w = 0, 1$,

$$\lim_{x \to 0^+} \mathbb{E}[Y_i(w)|X_i = x] = \lim_{x \to 0^-} \mathbb{E}[Y_i(w)|X_i = x].$$

- However, this is fundamentally untestable, so Lee (2008) provide a more easily interpretable set of conditions:
  - Specifically, individuals should not be able to precisely manipulate their running variable.
- Two intuitive tests:
  1. "McCrary test": Density of the running variable $f_X(x)$ should be continuous at the threshold (McCrary 2008).
  2. "Placebo" RDs: if we estimate RDs with baseline characteristics of individuals as the "outcome", the effect should be zero.

# Many Application in Healthcare

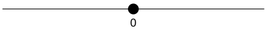
- Card, Dobkin, Maestas (2009): does Medicare save lives?
    - Medicare eligibility depends primarily on whether age exceeds 65.
    - Running variable = Age; Threshold = 65.
- Almond et al. (2010): effect of medical expenditure on at-risk newborns' health.
    - Newborns classified as "very low birth weight" (<1500g) receive more medical treatment.
    - Running variable = Birthweight; Threshold = 1500g.
- Almond et al. (2011): effect of longer hospital stay for mother and newborns.
    - Hospitals are reimbursed based on how many "midnights" newborns stay.
    - Running variable = Clock time of birth; Threshold = 12:00am (those born just after midnight stay for longer on average).

# Limitations

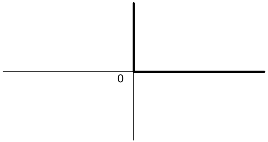
- While RD provides clean identification, its biggest limitation is that it only estimates the effect for individuals with running variable values local to the threshold.
  - Let $\tau(x) \equiv \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. RD only identifies $\tau(0)$.
- In reality, we may be interested in other quantities, e.g.,
  - Average Treatment Effect: $\mathbb{E}[\tau(X_i)] = \int_x \tau(x) dF_X(x)$.
  - Treatment effect heterogeneity: $\tau'(x)$.

## Multidimensional RD (MRD) Designs Presents an Opportunity

- Suppose that treatment $W$ is determined by more than one running variable.
  - Suppose two for simplicity: $X_1$, $X_2$.
  - Example: financial aid eligibility may depend on academic performance ($X_1$) and family income ($X_2$).
- Key insight: the threshold is no longer a point, but a "frontier", $\mathbb{F}$.
  $\rightarrow$ Can estimate heterogeneous treatment effects $\tau(x)$ for $x \in \mathbb{F}$.

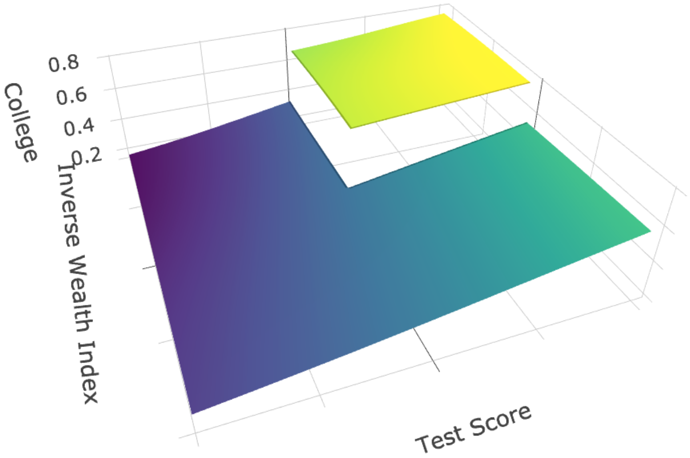



(a) Single-Dimensional RD




(b) MRD

# Visual Illustration (MRD)



# Most Applied Work Do Not Take Advantage of This

- Most applied work do not take advantage of this feature of MRD.
  - Typically, limit the sample to individuals eligible on one dimension, and estimate single-dimensional RD on the other dimension.
- Disadvantages of applied approach:
  - Does not make full use of heterogeneity: cannot answer questions such as "is the treatment effect increasing/decreasing in income"?
  - Inefficient: throws away (often a large) part of the sample.
- In today's talk, I will describe a simple and more efficient way to estimate MRD designs, which also allows us to estimate heterogeneous treatment effects.

# Examples of MRD

- With the growing availability of richer datasets, MRD designs can be used in an increasing number of settings.
  - Financial aid eligibility may depend on academic performance ($X_1$) and family income ($X_2$).
  - Medicaid eligibility may depend on age and wealth.
  - Extra attention is paid to VLBW (<1500g) *and* premature (gestation age $< 37$ weeks) newborns.

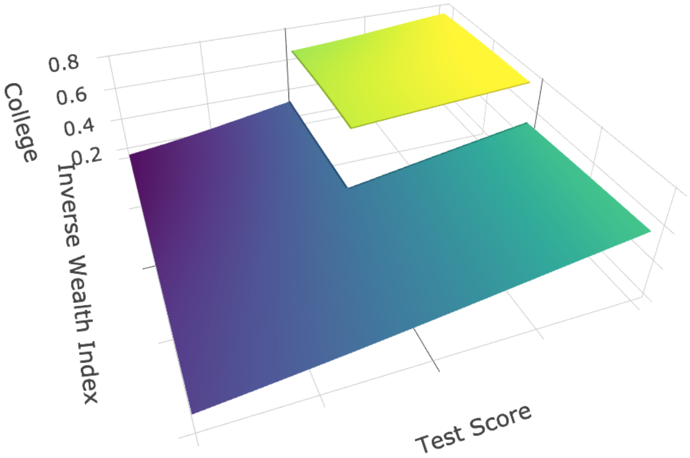
# MRD Identification

For MRD, the conditional average treatment effect (CATE) at any point along the frontier $\mathbb{F}$ is identified:

$$\begin{aligned}\tau_{MRD}(x) &\equiv \mathbb{E}\left[Y_i(1) - Y_i(0) | X_i = x\right] \\ &= \lim_{\epsilon \to 0} \mathbb{E}\left[Y_i | X_i \in B_{\epsilon}^{1}(x)\right] - \lim_{\epsilon' \to 0} \mathbb{E}\left[Y_i | X_i \in B_{\epsilon'}^{0}(x)\right],\end{aligned}$$

for any $x \in \mathbb{F}$, where $B_{\epsilon}^{w}(x) \equiv B_{\epsilon}(x) \cap \Omega_w$, $B_{\epsilon}(x)$ is the $\epsilon$-ball around $x$, and $\Omega_w$ is the region in running variable space where units are eligible for treatment $w \in \{0, 1\}$.

# Essentially, We Just Need to Estimate Two CEFs. But How?



# Single-Dimensional RD Estimation

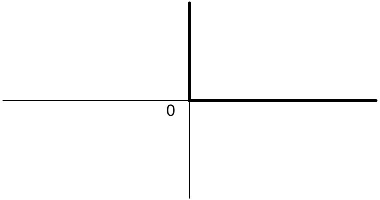
- Before discussing MRD estimation, let's first review how it's done for single-dimensional RDs.

- A common approach in applied work for single-dimensional RDs is to fit global polynomials on both sides of the threshold.
  - Econometricians have shown that this is generally a bad idea: $\hat{\tau}(x)$ may be affected by observations far from the threshold.

- RD estimation using local linear regressions has theoretically optimal properties:
  - RD estimate is only determined based on observations "close" to the threshold.
  - How close? Need to estimate the optimal bandwidth $h^{opt}$, which depends on parameters like the second derivative of the conditional expectation function (CEF) that need to be estimated nonparametrically.

## Challenge for MRD Estimation

- Can we extend the local linear regression estimation method to MRD?
  - In principle yes, but in practice data requirements will be very high. Why?
  - Need to estimate a *continuum* of optimal bandwidth $h^{opt}(x)$ for each point $x$ along the treatment frontier $\mathbb{F}$.
  - Optimal bandwidth formula at each point also more complicated: "second derivative" is now the Hessian matrix for the two-dimensional case.



# What Do We Want in a Good Estimator?

1. Flexible: estimates close to $\mathbb{F}$ should only be affected by nearby points (nonparametric estimation).

2. Not *too* flexible: don't want to overfit, so need to control the smoothness (regularization, e.g., bandwidth in local linear regressions).

3. Feasible/easy to compute: while estimating MRD designs via local linear regressions has optimality properties, estimating second derivative of a multivariate function nonparametrically requires a lot of data.

## Thin Plate Splines

- The solution I propose is to use a method called thin plate splines.
- Specifically, for $z = 0, 1$, we estimte:

$$\hat{g}_z = argmin_{u \in \Omega_z} \sum_{i=1}^{n_z} (y_i^z - u(x_i^z))^2 + \lambda_z J_{md}^z(u),$$

where the penalty $J_{md}^z(u)$ is given by:

$$J_{md}^z(u) \equiv \int_{\Omega_z} \sum_{|\alpha| \leq m} |D^\alpha u|^2 dx.$$

- For our purposes, main difference from local linear regressions is that the penalty parameter is a scalar $\lambda_z$, instead of a continuum of optimal bandwidths $h^{opt}(x)$.

# Inference

- Under suitably chosen priors for the thin plate splines' parameters, it is easy to construct Bayesian standard errors for the estimates $\hat{\tau}(x)$.
  - Alternatively, we can use nonparametric bootstrap to obtain standard errors.
- A potential issue is that MSE-optimal estimates generally have non-negligible bias in the asymptotic distribution.
  - Common issue for nonparametric estimation.
- Solution here: undersmoothing.
  - Either ad hoc: divide MSE-optimal $\hat{\lambda}_z$ by two, or;
  - Use MSE-optimal $\hat{\lambda}_z$ from a higher-order thin plate spline.

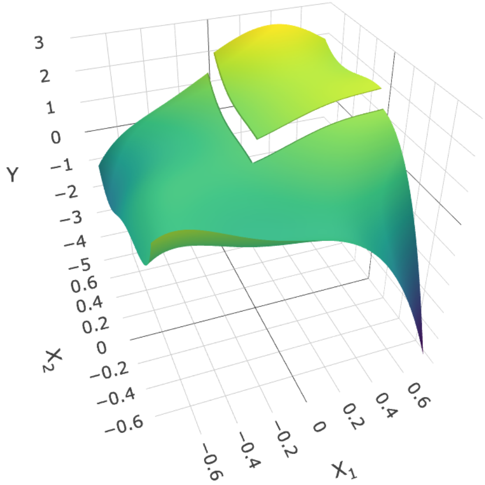
## So, Does It Really Work?

- To test the performance of the estimator, I compare its performance with single-dimensional RD methods.
- Data-generating process (DGP):

$$Y_i = \begin{cases} \sum_{p+q\leq 5} a_{p,q} X_{1i}^p X_{2i}^q + \tau(X_{1i}, X_{2i}) + \epsilon_i & X_{1i} \geq 0, X_{2i} \geq 0, \\ \sum_{p+q\leq 5} a_{p,q} X_{1i}^p X_{2i}^q + \epsilon_i & \text{otherwise.} \end{cases}$$

- Assume $X_{1i}$, $X_{2i}$ drawn independently from $2Beta(3,3) - 1$, and $\epsilon_i \sim^{i.i.d.} N(0,1)$.
- Two specifications:
  - Constant treatment effects: $\tau(X_{1i}, X_{2i}) = 0.5$.
  - Heterogeneous treatment effects: $\tau(X_{1i}, X_{2i}) = 0.5 + X_{1i} - X_{2i}$.
- 100 simulations, 10,000 observations in each.

# CEF for DGP with Constant Treatment Effects



# MRD Estimator Performs Comparably and is Slightly More Precise

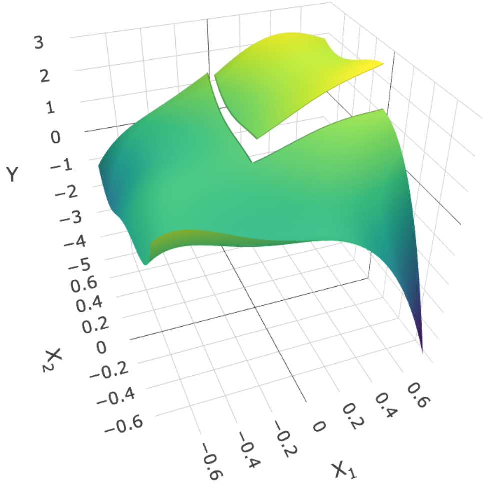
*Panel A. Estimates of the Average Treatment Effect Over $\{X_1=0, X_2\geq 0\}$*

| Estimator | Bias | MSE | Coverage | Average CI Length |
|---|---|---|---|---|
| IK | 0.002 | 0.009 | 0.97 | 0.334 |
| CCT | -0.015 | 0.017 | 0.87 | 0.397 |
| KR | 0.001 | 0.009 | 0.96 | 0.372 |
| MRD (MSE-Optimal) | -0.030 | 0.006 | 0.94 | 0.318 |
| MRD (Bias-Corrected) | -0.031 | 0.007 | 0.94 | 0.329 |
| MRD (Undersmoothing) | -0.038 | 0.008 | 0.94 | 0.333 |

*Panel B. Estimates of the Average Treatment Effect Over $\{X_1\geq 0, X_2=0\}$*

| Estimator | Bias | MSE | Coverage | Average CI Length |
|---|---|---|---|---|
| IK | -0.005 | 0.010 | 0.93 | 0.350 |
| CCT | 0.022 | 0.014 | 0.91 | 0.397 |
| KR | -0.007 | 0.008 | 0.98 | 0.361 |
| MRD (MSE-Optimal) | -0.018 | 0.006 | 0.95 | 0.320 |
| MRD (Bias-Corrected) | -0.018 | 0.007 | 0.95 | 0.331 |
| MRD (Undersmoothing) | -0.015 | 0.007 | 0.95 | 0.335 |

# CEF for DGP with Heterogeneous Treatment Effects



# Vastly Outperforms at Estimating Heterogeneous Treatment Effects

*Panel A. Estimates of the Treatment Effect Over $\{X_1=0, X_2\geq 0\}$*

| Estimator | Bias | IMSE | Coverage | CI Length |
|---|---|---|---|---|
| IK | 0.008 | 0.044 | 0.741 | 0.982 |
| CCT | -0.006 | 0.080 | 0.751 | 1.216 |
| KR | 0.019 | 0.029 | 0.786 | 0.933 |
| MRD (MSE-Optimal) | -0.028 | 0.017 | 0.927 | 0.498 |
| MRD (Bias-Corrected) | -0.030 | 0.021 | 0.922 | 0.522 |
| MRD (Undersmoothing) | -0.037 | 0.020 | 0.930 | 0.532 |

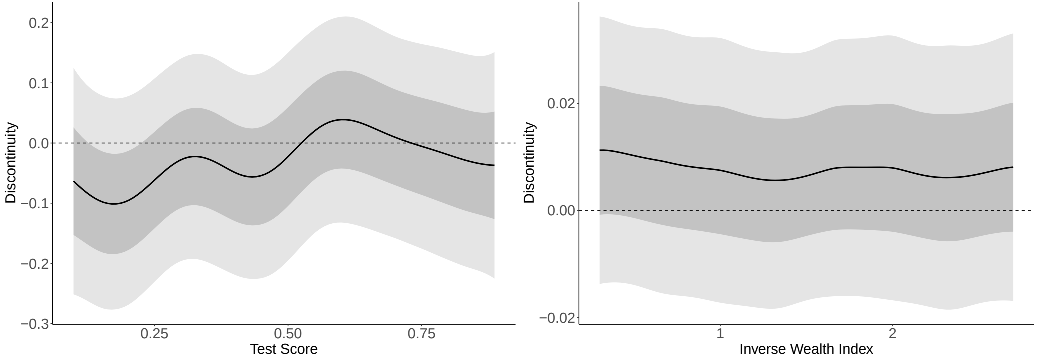
*Panel B. Estimates of the Treatment Effect Over $\{X_1\geq 0, X_2=0\}$*

| Estimator | Bias | IMSE | Coverage | CI Length |
|---|---|---|---|---|
| IK | -0.010 | 0.043 | 0.507 | 0.998 |
| CCT | 0.004 | 0.077 | 0.564 | 1.225 |
| KR | -0.020 | 0.032 | 0.526 | 0.949 |
| MRD (MSE-Optimal) | -0.019 | 0.015 | 0.944 | 0.502 |
| MRD (Bias-Corrected) | -0.020 | 0.018 | 0.942 | 0.527 |
| MRD (Undersmoothing) | -0.016 | 0.018 | 0.945 | 0.536 |

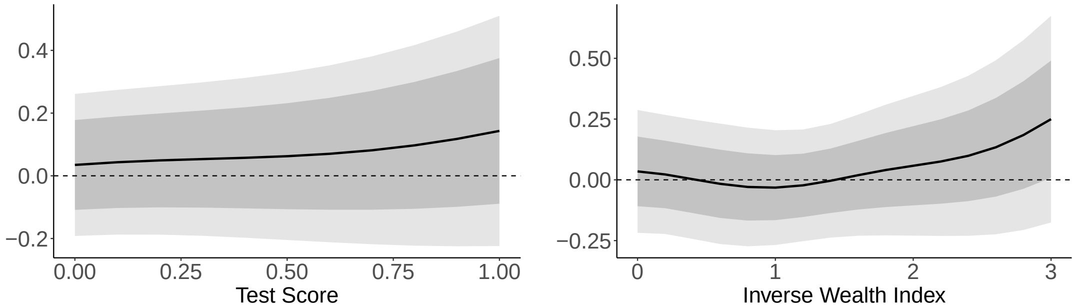
# Empirical Application 1: Effect of Financial Aid on College Enrollment

- The first empirical example here based on a financial aid program in Colombia.
  - However, the same MRD design can be applied to study later life health outcomes once enough time has passed.
- Students in Colombia are eligible for financial aid if their test scores exceed a certain threshold, and family wealth is low enough.
  - Hence, we have an MRD design with test scores and an inverse wealth index as the running variables.
  - We will use this to study the effect of financial aid on college enrollment.

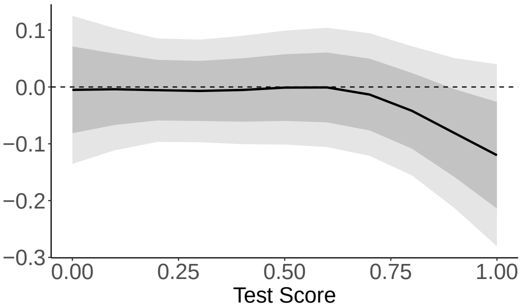
# No Discontinuity in the Multivariate Density Along $\mathbb{F}$ (Extension of McCrary Test to Multiple Dimensions)



# Placebo RDs: No Discontinuity in Age



# Placebo RDs: No Discontinuity in Gender




# Comparison with RD Methods from Applied Work (Test Score Threshold)

*Panel A. MRD Estimates*

| | High Quality Institutions | | | | Low Quality Institutions | | |
|---|---|---|---|---|---|---|---|
| | Any | Any | Private | Public | Any | Private | Public |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Treatment Effect Estimate | 0.334*** | 0.475*** | 0.478*** | -0.01** | -0.139*** | -0.062*** | -0.07*** |
| | (0.012) | (0.013) | (0.012) | (0.005) | (0.007) | (0.004) | (0.005) |
| Number of Observations | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 |

*Panel B. Original Estimates from LRS*

| | High Quality Institutions | | | | Low Quality Institutions | | |
|---|---|---|---|---|---|---|---|
| | Any | Any | Private | Public | Any | Private | Public |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Treatment Effect Estimate | 0.320*** | 0.465*** | 0.466*** | 0.000 | -0.154*** | -0.063*** | -0.087*** |
| | (0.012) | (0.012) | (0.011) | (0.007) | (0.011) | (0.007) | (0.009) |
| Number of Observations | 299,475 | 299,475 | 299,475 | 299,475 | 299,475 | 299,475 | 299,475 |

Notes: Standard errors are shown in parentheses.

# Comparison with RD Methods from Applied Work (Wealth Threshold)

*Panel A. MRD Estimates*

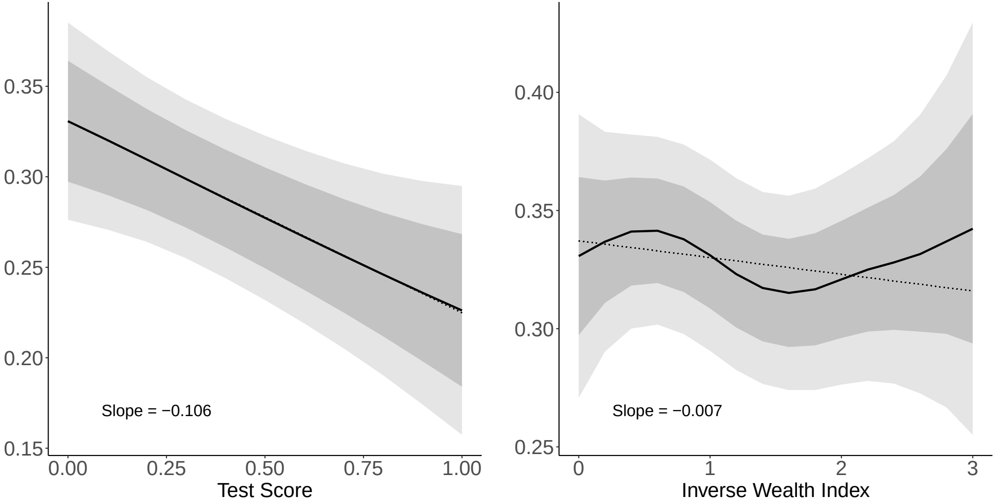
| | | High Quality Institutions | | | Low Quality Institutions | | |
|---|---|---|---|---|---|---|---|
| | Any | Any | Private | Public | Any | Private | Public |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Treatment Effect Estimate | 0.288*** | 0.429*** | 0.478*** | -0.02*** | -0.147*** | -0.074*** | -0.077*** |
| | (0.021) | (0.023) | (0.019) | (0.005) | (0.011) | (0.006) | (0.005) |
| Number of Observations | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 | 349,015 |

*Panel B. Original Estimates from LRS*

| | | High Quality Institutions | | | Low Quality Institutions | | |
|---|---|---|---|---|---|---|---|
| | Any | Any | Private | Public | Any | Private | Public |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Treatment Effect Estimate | 0.274*** | 0.396*** | 0.477*** | -0.079*** | -0.120*** | -0.052*** | -0.076*** |
| | (0.027) | (0.024) | (0.020) | (0.018) | (0.022) | (0.015) | (0.016) |
| Number of Observations | 23,132 | 23,132 | 23,132 | 23,132 | 23,132 | 23,132 | 23,132 |

Notes: Standard errors are shown in parentheses.

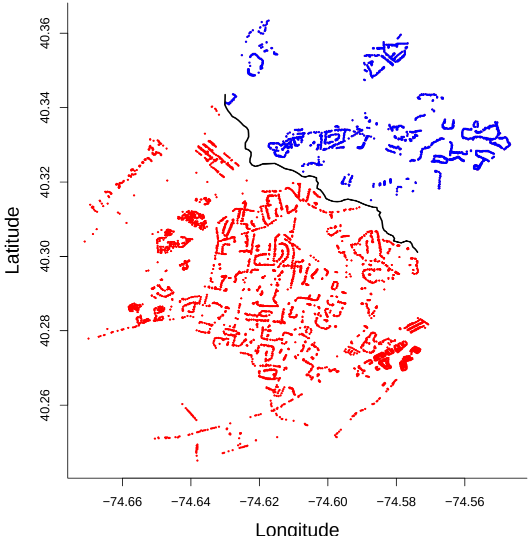
# Is there Treatment Effect Heterogeneity?



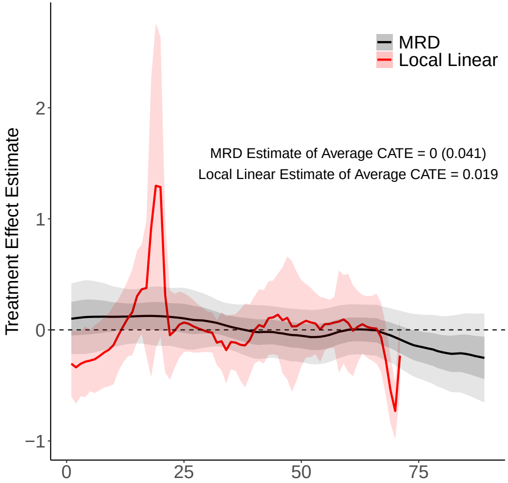
# Empirical Application 2: Geographical MRD Design

- The second empirical application comes from a political science study, based on a geographical RD design.
  - Possible applications of geographical RDs in health include time zone RD.
- In this application, authors study the effect of campaign advertisements on voter turnout in the 2008 US presidential elections.
  - Campaigns buy TV advertisements based on designated market areas (DMAs), which are chosen by Nielsen Media Research.
  - Authors argue that precise DMA designation is unrelated to political variables.
  - MRD exploiting the fact that neighboring DMAs receive very different levels of TV presidential advertisements in leadup to elections: average of 177 daily in one DMA, and zero in the other.

# Visualization of Treatment and Control Groups



# MRD Estimates Suggest No Effects Along Boundary, and are More Stable Than Local Linear Estimates



# Regression Kink Designs

- An extension to RD designs is regression kink (RK) designs (Card et al. 2016).
  - In RK designs, the endogenous variable $\mathcal{W}_i$ is continuous.
  - There is a kink in the first derivative of the treatment variable.
- Example: unemployment insurance (UI).
  - Replacement rate $\mathcal{W}_i$ is a fraction $\alpha$ of previous income $X_i$ up to a cap $\alpha\bar{X}$, constant after that.
  - Results in a kink at $\bar{X}$.
- Single-dimensional RK estimand:

$$\frac{\lim_{x\to 0^+} \mathbb{E}[dY_i/dX_i|X_i=\bar{X}] - \lim_{x\to 0^-} \mathbb{E}[dY_i/dX_i|X=\bar{X}]}{\lim_{x\to 0^+} \mathbb{E}[d\mathcal{W}_i/dX_i|X_i=\bar{X}] - \lim_{x\to 0^-} \mathbb{E}[d\mathcal{W}_i/dX_i|X=\bar{X}]}.$$

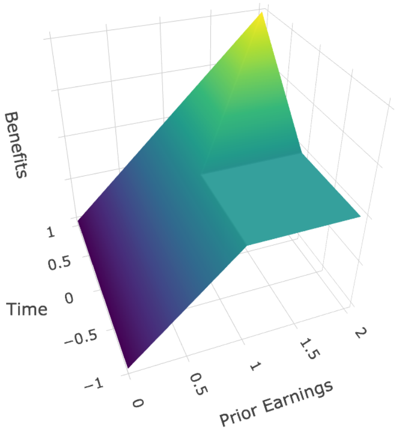
# Multidimensional RK (MRK) Designs

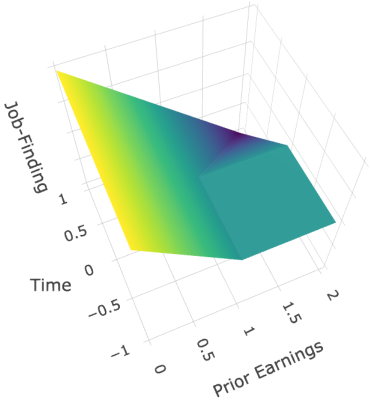
- Suppose that at a certain point in time $t_0$ onwards, the cap gradually raised, e.g., $\bar{X}_t = \bar{X}_{t_0} + \gamma(t - t_0)$ for $t \geq t_0$.
- Then, we have a kink not only at $\bar{X}$ and $t < t_0$, but also another kink at $\bar{X}_t$ and $t \geq t_0$.
  - Can be estimated using my R package.
- RKD estimand for $x \in \mathbb{F}$, and for any $v \in \mathbb{R}^d$ satisfying $x + \delta v \notin \mathbb{F}$ for sufficiently small $\delta > 0$.:

$$\frac{\lim_{\epsilon \to 0} D_v \mathbb{E}\left[Y_i | X_i = x + \epsilon \cdot v, X_i \in B^1_\epsilon(x)\right] - \lim_{\epsilon' \to 0} D_v \mathbb{E}\left[Y_i | X_i = x + \epsilon' \cdot v, X_i \in B^0_{\epsilon'}(x)\right]}{\lim_{\epsilon \to 0} D_v \mathbb{E}\left[\mathcal{W}_i | X_i = x + \epsilon \cdot v, X_i \in B^1_\epsilon(x)\right] - \lim_{\epsilon' \to 0} D_v \mathbb{E}\left[\mathcal{W}_i | X_i = x + \epsilon' \cdot v, X_i \in B^0_{\epsilon'}(x)\right]}.$$

- Treatment effect heterogeneity may be interesting:
  - E.g., is the effect of marginal dollar on UI greater/smaller when UI is less/more generous?

# Visual Illustration of MRK Design

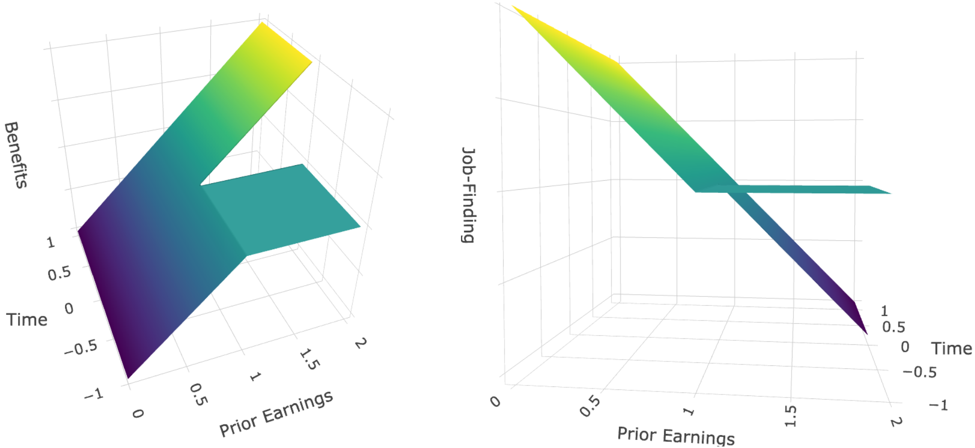


# Multidimensional Regression Discontinuity/Kink (MRDK) Designs

- Consider again the previous example, but instead of a gradual increase in the cap at time $t_0$, there is a one-time increase in the cap to a higher level $\alpha X^*$.
- Then, we have a kink at $\bar{X}$ and $t < t_0$, and a discontinuity at $X_i \in [\bar{X}, X^*]$ and $t = t_0$.

# Visual Illustration of MDRK Design



# Conclusion

- Today, I introduced multidimensional RD (MRD) designs.
- Provided a simple estimation procedure that allows users to receover heterogeneous treatment effects, and improves precision relative to single-dimensional methods.
- Methods can be extended to multidimensional kink designs.
- Numerous potential applications in healthcare research, e.g., to study the effects of Medicaid, effects medical spending on at-risk newborns, and effect of education on health.

# Thank you!

Questions or comments? Feel free to email me at alden15@nber.org.

Paper available at aldencheng.com.

R code available at: https://github.com/alden1505/rd_multiple_running_var_code.