Flexible Covariate Adjustments in Regression Discontinuity Designs

Claudia Noack (Oxford)

Tomasz Olma (UCL)

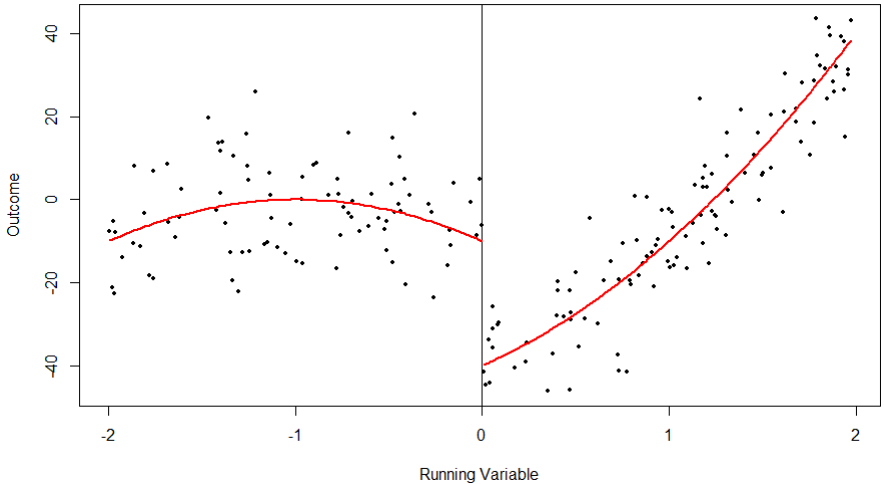
Christoph Rothe (Mannheim)

June 2022

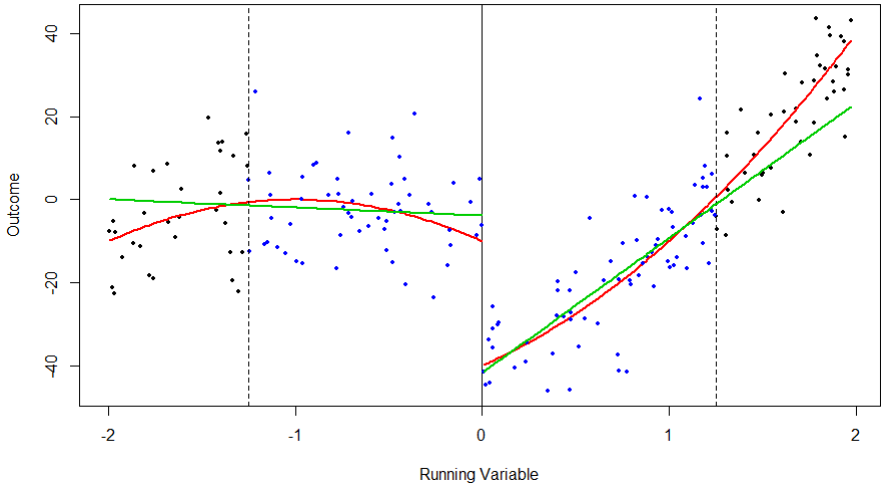
Regression Discontinuity (RD) Designs

- RD designs are widely used for causal inference in empirical microeconomics (labor, education, development, public economics, etc.).
- Units are assigned to treatment group if and only if a “running variable” falls above known cutoff value.
- Jump in CEF identifies causal effect local to cutoff if units there differ only in their treatment assignment.
- RD parameter is typically estimated via local linear regressions.

Illustration



Illustration



Predetermined Covariates

- Like in randomized experiments, controlling for predetermined covariates can yield more efficient estimates.
- Common practice is to include such covariates linearly and without localization in the local linear RD regression [e.g., Calonico et al., 2019].
- RD regression with linear adjustments yields consistent estimate of RD parameter even if linear functional form is not correct.
- Approach is simple to implement, but might not fully exploit the available covariate information, and cannot handle high-dimensional settings.

This Paper

- New class of RD estimators that allow for flexible covariate adjustments.
- Approach based on RD regressions with a "covariate-adjusted" outcome $Y - \widehat{\mu}(Z)$.
- Optimal adjustment function can be estimated by various methods (simple regression, nonparametric methods, machine learning, etc).
- Procedure is "very insensitive" to estimation error and misspecification of chosen adjustment function.
- Approach is very easy to implement; works directly with existing methods for bandwidth choice and inference.
- In our empirical application, nonlinear adjustments reduce the standard error by twice as much as linear adjustments.

Flexible Covariate Adjustments

Sharp RD Design

- Y_i is outcome, X_i is running variable, $T_i = \mathbf{1}\{X_i \geq 0\}$ is treatment status, and Z_i are pre-treatment covariates.
- Parameter of interest: $\tau = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0^-]$.
- The standard sharp RD estimator is based on local linear regression:

$$\widehat{\tau}(h) = e_1^\top \operatorname*{argmin}_{\beta \in \mathbb{R}^4} \sum_{i=1}^{n} K(X_i/h)(Y_i - (T_i, X_i, T_i X_i, 1)^\top \beta)^2,$$

where $K(\cdot)$ is a kernel, $h > 0$ is a bandwidth, and $e_1 = (1, 0, 0, 0)^\top$.

Current Practice: Linear Adjustments

- Covariates are often included linearly in the RD regression [Calonico et al., 2019]:

$$\widehat{\tau}_{CCFT}(h) = e_1^\top \operatorname*{argmin}_{\beta,\gamma} \sum_{i=1}^{n} K(X_i/h)(Y_i - (T_i, X_i, T_i X_i, 1)^\top \beta - Z_i^\top \gamma)^2.$$

- This estimator is asymptotically equivalent to standard RD estimator with adjusted outcome variable: $Y_i - Z_i^\top \gamma_0$

- Consistent if $\mathbb{E}[Z_i | X_i = x]$ is continuous around cutoff, because then:

$$\tau = \mathbb{E}[Y_i - Z_i^\top \gamma_0 | X_i = 0^+] - \mathbb{E}[Y_i - Z_i^\top \gamma_0 | X_i = 0^-].$$

Our Contribution: Flexible Adjustments

- Consider a general class of RD estimators, $\widehat{\tau}(h;\mu)$, with modified outcome variables of the form

$$M_i(\mu) = Y_i - \mu(Z_i).$$

- With pre-determined covariates the distribution of $Z_i|X_i = x$ changes smoothly around the cutoff so that $\mathbb{E}\left[\mu(Z_i)|X_i = x\right]$ is continuous in x for "any" function μ.

- It then holds that

$$\tau = \mathbb{E}[M_i(\mu)|X_i = 0^+] - \mathbb{E}[M_i(\mu)|X_i = 0^-].$$

- What is the optimal function μ for estimating τ?

Optimal Adjustment Function

- Under standard conditions, we have that

$$\widehat{\tau}(h;\mu) - \tau \overset{a}{\sim} N\left(h^2 B, \frac{1}{nh} V(\mu)\right), \text{ where}$$

$$B = \frac{\bar{\nu}}{2}\left(\partial_x^2 \mathbb{E}[Y_i|X_i = x]|_{x=0^+} - \partial_x^2 \mathbb{E}[Y_i|X_i = x]|_{x=0^-}\right),$$

$$V(\mu) = \frac{\bar{\kappa}}{f_X(0)}(\mathbb{V}[Y_i - \mu(Z_i)|X_i = 0^+] + \mathbb{V}[Y_i - \mu(Z_i)|X_i = 0^-]).$$

- Bias does not depend on μ, and variance is minimized by the function

$$\mu_0(z) = \frac{1}{2}(\mu_0^+(z) + \mu_0^-(z)),$$

where $\mu_0^+(z) = \mathbb{E}[Y_i|X_i = 0^+, Z_i = z]$ and $\mu_0^-(z) = \mathbb{E}[Y_i|X_i = 0^-, Z_i = z]$.

Our Proposed Estimator

1. Generate adjustment function using cross-fitting [Chernozhukov et al., 2018].
   - Choose an estimation method for μ_0.
   - Randomly split the data $\{(Y_i, X_i, Z_i)\}_{i\in[n]}$ into S folds of equal size.
   - For all $s \in [S]$, use data outside the s-th fold to generate estimator $\widehat{\mu}_s$ of μ_0.
2. Compute local linear RD estimator, $\widehat{\tau}_{CF}(h; \hat{\mu})$, with the outcome variable

$$M_i(\widehat{\mu}_{s(i)}) = Y_i - \widehat{\mu}_{s(i)}(Z_i),$$

   where $s(i)$ denotes the fold containing unit i.

Examples of Covariate Adjustments

- Can use a wide range of methods to estimate the CEFs

$$\mu_0^+(z) = \mathbb{E}[Y_i | X_i = 0^+, Z_i = z] \text{ and } \mu_0^-(z) = \mathbb{E}[Y_i | X_i = 0^-, Z_i = z].$$

- Depending on the setting, one could use:
  - simple parametric methods, e.g. linear regression.
  - classic nonparametric methods, e.g. kernel or series regression.
  - machine learning methods, e.g. (post-)Lasso regression, random forests, (deep) neural networks, or ensemble combinations thereof.

- Our theory does not require fast rates of convergence, and even consistent estimation is only needed for full efficiency.

Theory

Assumptions

1. There exist a set $\mathcal{T}_n$ and a function $\bar{\mu} \in \mathcal{T}_n$ such that: $\widehat{\mu}_s$ belongs to $\mathcal{T}_n$ with probability approaching 1 for all $s \in [S]$; and

$$\sup_{\mu \in \mathcal{T}_n} \sup_{x \in \mathcal{X}_h} \mathbb{E}\left[(\mu(Z_i) - \bar{\mu}(Z_i))^2 | X_i = x\right] = O(r_n^2) \text{ for some } r_n = o(1).$$

   - $\widehat{\mu}_s$ essentially concentrates in mean square around deterministic $\bar{\mu}$.
   - $\bar{\mu}$ could be different from μ_0, and r_n can converge arbitrarily slowly.

Assumptions

1. There exist a set $\mathcal{T}_n$ and a function $\bar{\mu} \in \mathcal{T}_n$ such that: $\widehat{\mu}_s$ belongs to $\mathcal{T}_n$ with probability approaching 1 for all $s \in [S]$; and

$$\sup_{\mu \in \mathcal{T}_n} \sup_{x \in \mathcal{X}_h} \mathbb{E}\left[(\mu(Z_i) - \bar{\mu}(Z_i))^2 | X_i = x\right] = O(r_n^2) \text{ for some } r_n = o(1).$$

   - $\widehat{\mu}_s$ essentially concentrates in mean square around deterministic $\bar{\mu}$.
   - $\bar{\mu}$ could be different from μ_0, and r_n can converge arbitrarily slowly.

2. For all $n \in \mathbb{N}$ and $\mu \in \mathcal{T}_n$, the function $\mathbb{E}[\mu(Z_i)|X_i = x]$ is twice continuously differentiable around $x = 0$.
   - Formalizes the notion that covariates are predetermined.
   - Can check plausibility in practice by running RD with outcome $\widehat{\mu}_{s(i)}(Z_i)$.

Assumptions

1. There exist a set $\mathcal{T}_n$ and a function $\bar{\mu} \in \mathcal{T}_n$ such that: $\widehat{\mu}_s$ belongs to $\mathcal{T}_n$ with probability approaching 1 for all $s \in [S]$; and

$$\sup_{\mu \in \mathcal{T}_n} \sup_{x \in \mathcal{X}_h} \mathbb{E}\left[(\mu(Z_i) - \bar{\mu}(Z_i))^2 | X_i = x\right] = O(r_n^2) \text{ for some } r_n = o(1).$$

   - $\widehat{\mu}_s$ essentially concentrates in mean square around deterministic $\bar{\mu}$.
   - $\bar{\mu}$ could be different from μ_0, and r_n can converge arbitrarily slowly.

2. For all $n \in \mathbb{N}$ and $\mu \in \mathcal{T}_n$, the function $\mathbb{E}[\mu(Z_i)|X_i = x]$ is twice continuously differentiable around $x = 0$.
   - Formalizes the notion that covariates are predetermined.
   - Can check plausibility in practice by running RD with outcome $\widehat{\mu}_{s(i)}(Z_i)$.

3. Other regularity conditions, including those for standard RD, hold.

Theoretical Results

1. Asymptotic equivalence: It holds that $\widehat{\tau}_{CF}(h;\widehat{\mu}) = \widehat{\tau}(h;\bar{\mu}) + O_p(r_n(h^2 + (nh)^{-1/2}))$.
   - Remainder is small if r_n converges fast.
   - $r_n = o(1)$ is sufficient for first-order equivalence.

Theoretical Results

1. Asymptotic equivalence: It holds that $\widehat{\tau}_{CF}(h;\widehat{\mu}) = \widehat{\tau}(h;\bar{\mu}) + O_p(r_n(h^2 + (nh)^{-1/2}))$.
   - Remainder is small if r_n converges fast.
   - $r_n = o(1)$ is sufficient for first-order equivalence.

2. Asymptotic normality: $\widehat{\tau}_{CF}(h;\widehat{\mu}) - \tau \overset{a}{\sim} N\left(h^2 B, (nh)^{-1} V(\bar{\mu})\right)$.
   - Bias does not depend on the choice of $\widehat{\mu}$ and variance depends only on $\bar{\mu}$.
   - Inference is valid using standard methods with generated data $\{X_i, M_i(\widehat{\mu}_{s(i)})\}_{i\in[n]}$.

Theoretical Results

1. Asymptotic equivalence: It holds that $\widehat{\tau}_{CF}(h;\widehat{\mu}) = \widehat{\tau}(h;\bar{\mu}) + O_p(r_n(h^2 + (nh)^{-1/2}))$.
   - Remainder is small if r_n converges fast.
   - $r_n = o(1)$ is sufficient for first-order equivalence.

2. Asymptotic normality: $\widehat{\tau}_{CF}(h;\widehat{\mu}) - \tau \overset{a}{\sim} N\left(h^2 B, (nh)^{-1} V(\bar{\mu})\right)$.
   - Bias does not depend on the choice of $\widehat{\mu}$ and variance depends only on $\bar{\mu}$.
   - Inference is valid using standard methods with generated data $\{X_i, M_i(\widehat{\mu}_{s(i)})\}_{i\in[n]}$.

3. Optimality: For "any" functions $\mu^{(a)}$ and $\mu^{(b)}$,

$$V(\mu^{(a)}) < V(\mu^{(b)}) \text{ iff } \mathbb{V}[\mu_0(Z_i) - \mu^{(a)}(Z_i)|X_i = 0] < \mathbb{V}[\mu_0(Z_i) - \mu^{(b)}(Z_i)|X_i = 0].$$

   - $\bar{\mu} = \mu_0$ minimizes the asymptotic variance.
   - Even if $\bar{\mu} \neq \mu_0$, our covariate adjustments still yields efficiency gains in most settings.

Discussion

Analogy with Randomized Experiments

- Close to the cutoff, sharp RD designs resemble randomized experiments (RE) with known and constant propensity score $p = 1/2$.
- In such REs, the efficient influence function (EIF) estimator is sample analogue of

$$\mathbb{E}[Y_i - m_0(Z_i)|T_i = 1] - \mathbb{E}[Y_i - m_0(Z_i)|T_i = 0]$$

  where $m_0(z) = \frac{1}{2}(\mathbb{E}[Y_i|Z_i = z, T_i = 1] + \mathbb{E}[Y_i|Z_i = z, T_i = 0])$.
- EIF estimator with true propensity score is known to be very insensitive to the first-stage estimation error [e.g, Wager et al., 2016, Chernozhukov et al., 2018].
- Our estimator is fully analogous to this EIF estimator, and our minimal variance is analogous to the RE's semiparametric efficiency bound [Hahn, 1998].

Numerical Examples

Implementation Details

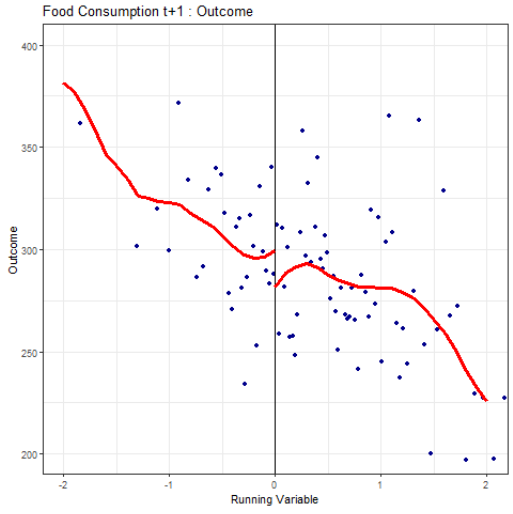
- In the first stage, we estimate μ_0 by the following methods:
  - linear regression; shallow neural nets; random forest; boosted trees; post-lasso estimator; ensemble combination of these methods.
- We use default values of the tuning parameters in the respective R-packages.
- We consider 10 folds for cross-fitting.
- In the second stage, final bandwidth choice and inference are conducted using robust bias corrections [Calonico et al., 2014].

Empirical Application

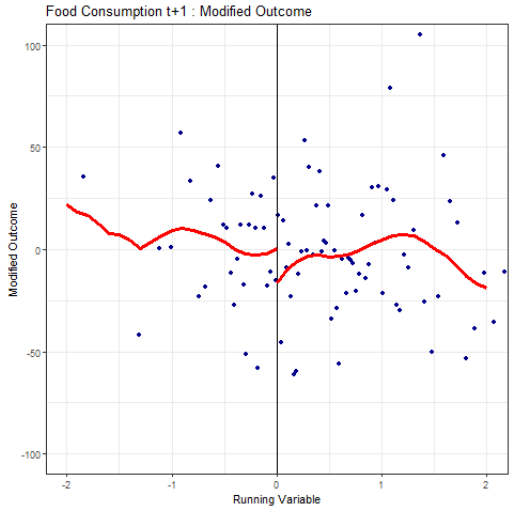
Empirical Application

- Progresa/Oportunidades is an anti-poverty program in Mexico that provides conditional cash transfers to poor households.
- Data is a sample of 1824 households living in urban areas, where the program started in 2003 [Calonico et al., 2014] .
- Eligibility for the transfer was based on a household poverty index.
- We estimate a sharp RD with poverty index as running variable and food consumption expenditures one year after its implementation as outcome.
- We consider the following pre-determined households characteristics:
  - Housholds size, age, education, number of children, house characteristics (number of rooms, etc.), pre-intervention consumption, and location indicators.

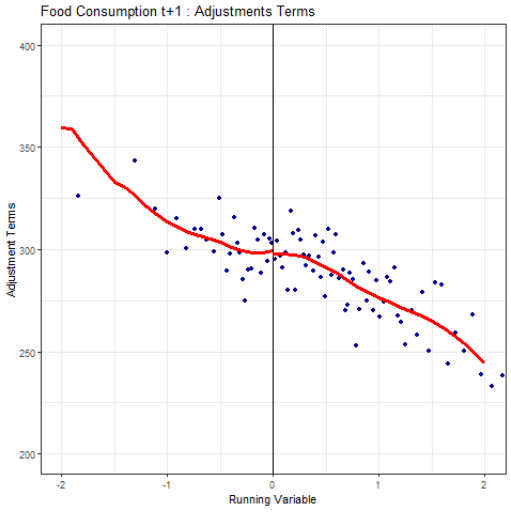
Outcome vs. Running Variable (binned)



Adjusted Outcome vs. Running Variable (binned)



Adjustment vs. Running Variable (binned)



Empirical Illustration - Results

Table: Estimation results for the empirical application using rdrobust.

	Estimate	SE	ΔSE	CI	h	b
No covariates	-27.2	23.7	-	[-73.6, 19.3]	0.37	0.61
Linear	-24.7	21.4	-9.7%	[-66.6, 17.3]	0.38	0.66
Neural Nets	-23.4	21.1	-11.2%	[-64.7, 17.8]	0.38	0.64
Boosted Trees	-21.5	19.4	-18.2%	[-59.5, 16.6]	0.39	0.68
Random Forest	-19.1	19.5	-17.7%	[-57.3, 19.2]	0.39	0.67
rLasso	-19.7	19.5	-18.0%	[-57.8, 18.4]	0.38	0.62
SuperLearner	-20.9	19.2	-19.1%	[-58.6, 16.7]	0.39	0.65

Summary

Summary

- We propose a new class of two-step, cross-fitted RD estimators that allow us to flexibly incorporate covariates.
- We require only very mild conditions on the first-stage estimator.
- The final estimator is more efficient than existing procedures.
- It is very easy to implement using standard software in empirical applications.
- We extend our framework to fuzzy RD designs.

Thank you

Simulations

Simulation Setup

- We consider three different setups with four continuous covariates that enter the outcome equation with different degrees of complexity.
- We compare “no covariates” estimator to different types of adjustments.
  - Two infeasible ones: the optimal and the best linear.
  - The feasible adjustments described above.
- Sample size is 2,000 and we consider 10,000 Monte Carlo draws.

Simulations - Data Generating Processes

- We generate data as

$$Y_i = T_i + \mu_X(X_i) + \mathbf{1}\{L > 0\} \cdot \left(\bar{\iota}_L(\rho) \sum_{l=1}^{L} b_l(Z_i) + \sum_{l=1}^{4} b_l(Z_i) \cdot (T_i + \mu_X(X_i)) \right) + \varepsilon_i,$$

  with $\mu_X(X_i) = \text{sign}(X_i)(X_i + X_i^2 - 2 \cdot (|X_i| - 0.1)_+^2)$, and $b_l(Z_i)$ are Hermite polynomials of Z_i.

- We let $\varepsilon_i \sim \mathcal{N}(0, 0.2)$, and $X_i, Z_i \sim U[-1, 1]$.

- We set $\bar{\iota}_L(\rho)$ to equalize the signal to noise ratio of covariates across models.

Simulation Results: L=0

- If μ_0 is independent of covariates, all adjustment methods yield similar estimates.

	Cov	Bias	Std	CI Len.	Rel. CI Len. (%)	Avg. h
No covariates	94.1	0.8	4.4	16.7	0.0	29.6
Infeasible Adjustments						
Oracle	94.1	0.8	4.4	16.7	0.0	29.6
Linear Oracle	94.1	0.8	4.4	16.7	0.0	29.6
Feasible Adjustments						
Linear	94.0	0.8	4.5	16.8	0.3	29.6
Neural Nets	94.0	0.8	4.5	16.8	0.2	29.6
Boosted Tree	93.9	0.8	4.5	16.9	1.1	29.6
Random Forest	94.0	0.8	4.6	17.2	2.6	29.7
rLasso	94.0	0.8	4.4	16.7	0.1	29.6
SuperLearner	94.0	0.8	4.4	16.7	0.1	29.6

Simulation Results: L=4

- If μ_0 is linear in covariates, linear adjustments yield substantial improvements.

	Cov	Bias	Std	CI Len.	Rel. CI Len. (%)	Avg. h
No covariates	94.2	0.8	39.3	143.0	0.0	30.2
Infeasible Adjustments						
Oracle	94.7	0.8	14.9	56.9	-60.2	26.8
Linear Oracle	94.7	0.8	14.9	56.9	-60.2	26.8
Feasible Adjustments						
Linear	94.6	0.8	15.4	58.3	-59.3	27.0
Neural Nets	94.5	0.7	15.6	58.9	-58.8	27.1
Boosted Tree	94.0	0.7	18.2	67.1	-53.1	27.9
Random Forest	94.3	0.7	18.9	69.6	-51.3	28.1
rLasso	94.7	0.7	15.4	58.4	-59.2	27.0
SuperLearner	94.5	0.7	15.5	58.5	-59.1	27.0

Simulation Results: L=16

- If μ_0 is nonlinear in covariates, nonparametric adjustments substantially improve upon linear adjustments.

	Cov	Bias	Std	CI Len.	Rel. CI Len. (%)	Avg. h
No covariates	94.1	0.1	28.3	105.5	0.0	30.0
Infeasible Adjustments						
Oracle	94.6	0.8	14.4	54.9	-48.0	28.0
Linear Oracle	94.6	0.5	19.5	73.5	-30.3	29.1
Feasible Adjustments						
Linear	94.6	0.5	19.7	74.1	-29.8	29.1
Neural Nets	94.5	0.6	15.9	59.9	-43.2	28.3
Boosted Tree	94.5	0.5	18.4	69.2	-34.4	28.9
Random Forest	94.3	0.6	16.7	62.3	-40.9	28.6
rLasso	94.3	0.7	14.9	56.2	-46.7	28.1
SuperLearner	94.3	0.7	15.0	56.4	-46.6	28.1

Theory

Standard RD Assumptions I

Assumption 4

1. X_i is continuously distributed with density f_X, which is continuous and bounded away from zero over an open neighborhood of the cutoff;

2. The kernel function K is a bounded and symmetric density function that is continuous on its support, and equal to zero outside some compact set, say $[-1, 1]$;

3. The bandwidth satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

Standard RD Assumptions II

Assumption 5

There exist constants C and L such that the following conditions hold for all $n \in \mathbb{N}$.

1. $\mathbb{E}[M_i(\bar{\mu})|X_i = x]$ is twice continuously differentiable on $\mathcal{X} \setminus \{0\}$ with L-Lipschitz continuous second derivative bounded by C;

2. For all $x \in \mathcal{X}$ and some $q > 2$, $\mathbb{E}[(M_i(\bar{\mu}) - \mathbb{E}[M_i(\bar{\mu})|X_i])^q|X_i = x]$ exists and is bounded by C;

3. $\mathbb{V}[M_i(\bar{\mu})|X_i = x]$ is L-Lipschitz continuous and bounded from below by $1/C$ for all $x \in \mathcal{X} \setminus \{0\}$.

Back

Constants in Asymptotic Distribution

$$B(\bar{\mu}) = \frac{\bar{\nu}}{2}\left(\partial_x^2 \mathbb{E}[M_i(\bar{\mu})|X_i = x]\big|_{x=0^+} - \partial_x^2 \mathbb{E}[M_i(\bar{\mu})|X_i = x]\big|_{x=0^-}\right) + o(1),$$
$$V(\bar{\mu}) = \frac{\bar{\kappa}}{f_X(0)}\left(\mathbb{V}[M_i(\bar{\mu})|X_i = 0^+] + \mathbb{V}[M_i(\bar{\mu})|X_i = 0^-]\right),$$

where

$$\bar{\nu} = (\bar{\nu}_2^2 - \bar{\nu}_1\bar{\nu}_3)/(\bar{\nu}_2\bar{\nu}_0 - \bar{\nu}_1^2)$$
$$\bar{\kappa} = \int_0^\infty (k(v)(\bar{\nu}_1 v - \bar{\nu}_2))^2 dv, (\bar{\nu}_2\bar{\nu}_0 - \bar{\nu}_1^2)^2,$$
$$\bar{\nu}_j = \int_0^\infty v^j k(v) dv.$$

Relation to the Literature

- RD Literature
  - Calonico et al. [2019] include covariates linearly in the regression equation.
  - Frölich and Huber [2019] incorporate covariates in a fully nonparametric fashion, but their method suffers from the curse of dimensionality.
  - Arai et al. [2021], Kreiß and Rothe [2021] consider lasso regularization.
- Two-stage estimation with nonparametric first stage
  - The combination of locally robust moment conditions and cross-fitting is used for estimation of regular parameters [Belloni et al., 2017, Chernozhukov et al., 2018].
  - Estimation of conditional average treatment effects in models with unconfoundedness using orthogonal moments [Kennedy et al., 2017, Kennedy, 2020, Fan et al., 2020, Colangelo and Lee, 2020].
  - Randomized experiments [Wager et al., 2016, Chernozhukov et al., 2018].

Assumptions

Further Regularity Conditions

Assumption 3

For $j \in \{1, 2\}$, it holds that:

$$\sup_{\mu \in \mathcal{T}_n} \sup_{x \in \mathcal{X}_h \setminus \{0\}} \left| \partial_x^j \mathbb{E}\left[\mu(Z_i) - \bar{\mu}(Z_i) | X_i = x\right] \right| = O(r_n).$$

- We also assume that the first two derivatives of the conditional expectations of our adjustment terms converge.
- This condition is implied by Assumption 2 and additional smoothness assumptions on the conditional distribution of the covariates.
- One sufficient condition is that $f_{Z|X}(z|x)$ is bounded away from zero and $\partial_x^j f_{Z|X}(z|x)$ is bounded for $j \in \{1, 2\}$.
- We further impose standard RD assumptions.

Theoretical Results

Asymptotic Equivalence

Theorem 1

(i) Under our assumptions, it holds that

$$\widehat{\tau}_{CF}(h;\widehat{\mu}) = \widehat{\tau}(h;\bar{\mu}) + O_P(r_n(h^2 + (nh)^{-1/2})),$$

where $\widehat{\tau}(h;\bar{\mu})$ is the RD estimator using $\bar{\mu}$ as adjustment function.

- The accuracy of this approximation increases with the rate of r_n, but the first-order asymptotic equivalence holds even if r_n converges arbitrarily slowly.
- $\widehat{\tau}_{CF}(h;\widehat{\mu})$ is insensitive to sampling variation in $\widehat{\mu}$ as it is a sample analogue of

$$\tau = \mathbb{E}[M_i(\mu)|X_i = 0^+] - \mathbb{E}[M_i(\mu)|X_i = 0^-].$$

$\Rightarrow$ $\widehat{\tau}_{CF}(h;\widehat{\mu})$ is not very sensitive to the choice of tuning parameters in the first stage.

Asymptotic Distribution

Theorem 1

(ii) Under our assumptions and for some functions $B(\bar{\mu})$ and $V(\bar{\mu})$, it holds that

$$\sqrt{nh}\, V(\bar{\mu})^{-1/2} \left(\widehat{\tau}_{CF}(h; \widehat{\mu}) - \tau - B(\bar{\mu}) h^2 \right) \to \mathcal{N}(0, 1).$$

- If $\partial_x^2 \mathbb{E}[\bar{\mu}(Z_i) | X_i = x]$ is also continuous at the cutoff, then the bias is identical to that of the "no covariates" RD estimator's bias.
- As the sampling uncertainty in $\widehat{\mu}$ can be essentially ignored, under additional assumptions, existing methods for bandwidth choice and inference can be readily applied to the data $\{(X_i, M_i(\widehat{\mu}_{s(i)}))\}_{i \in [n]}$.

Constants

Result: Minimum Variance

Theorem 2

Consider arbitrary adjustment functions $\mu^{(a)}$ and $\mu^{(b)}$, and assume that $\text{Var}(m^\star(Z_i) - \mu^{(j)}(Z_i)|X = x)$ is continuous for $\star \in \{+, -\}$ and $j \in \{a, b\}$. Then

$$V(\mu^{(a)}) < V(\mu^{(b)}) \Leftrightarrow \mathbb{V}[\mu_0(Z_i) - \mu^{(a)}(Z_i)|X_i = 0] < \mathbb{V}[\mu_0(Z_i) - \mu^{(b)}(Z_i)|X_i = 0]$$

- Under such a "smooth variance" condition, our estimator has smallest possible variance if $\bar{\mu} = \mu_0$.
- Even if $\bar{\mu} \neq \mu_0$, our proposed covariate adjustments still yield efficiency gains relative to "no covariates" estimator in typical settings.
- The "smooth variance" condition is sufficient but not necessary. Can sometimes exploit structure of $\mathcal{T}_n$ (example: linear adjustments).