

Выражаю благодарность Георгию Калашнову за использование части материалов (2020) в слайдах. Также использованы материалы из «A Practical Introduction to Regression Discontinuity Designs» Matias D. Cattaneo (2019), материалы «Applied metrics» Paul Goldsmith-Pinkham (2021)

Практическая эконометрика. Разрывная регрессия

авторы: Георгий Калашнов, Ольга Сучкова,
преподаватели 2021: Ольга Сучкова, Алексей Замниус,
Анна Ставнийчук

11 и 18 ноября 2022 г.

План на сегодня

Разрывная регрессия

Четкая разрывная регрессия

Инструментальные переменные в экспериментах

Размытая разрывная регрессия

Table of Contents

Разрывная регрессия

Четкая разрывная регрессия

Инструментальные переменные в экспериментах

Размытая разрывная регрессия

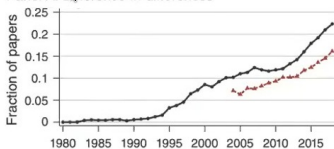
Применимость метода RDD

VOL. 110

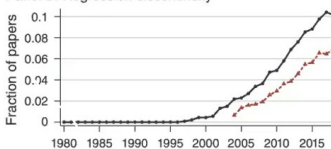
TECHNOLOGY AND BIG DATA ARE CHANGING ECONOMICS

45

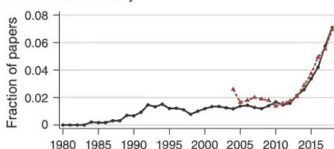
Panel A. Difference-in-differences



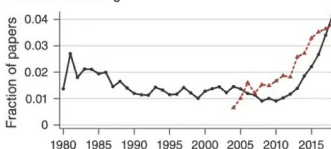
Panel B. Regression discontinuity



Panel C. Event study



Panel D. Bunching



— NBER working papers - - - Top-five journals

Source: *Technology and Big Data Are Changing Economics: Mining Text to Track Methods* by Janet Currie, Henrik Kleven, and Esmee Zwiers

Впервые - в статье Thistlewaite and Campbell (1960)

Пример RDD: Black (1999)

DO BETTER SCHOOLS MATTER?

581



FIGURE I
Example of Data Collection for One City: Melrose
Streets, and Attendance District Boundaries

Table of Contents

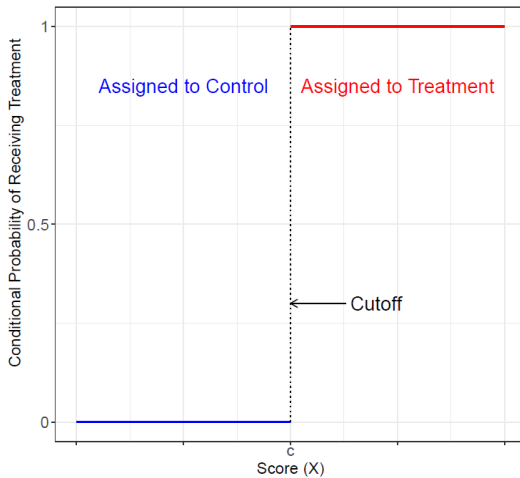
Разрывная регрессия

Четкая разрывная регрессия

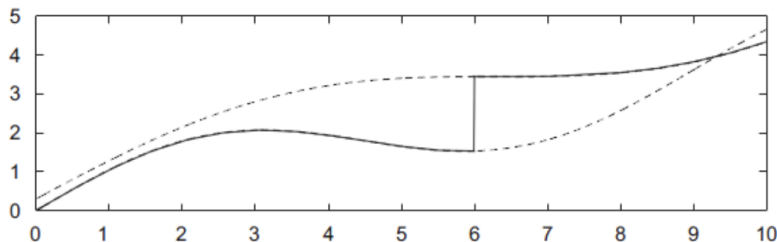
Инструментальные переменные в экспериментах

Размытая разрывная регрессия

Как присваивается treatment



Четкая разрывная регрессия: Обозначения

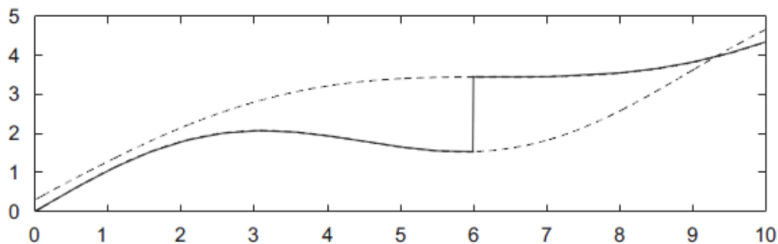


- ▶ Потенциальные исходы $Y(0)$, $Y(1)$. Ковариаты: X
- ▶ R – Running variable (или Score). По ней проходит граница
- ▶ $T = \mathbb{I}(R > c)$. c – отсечка (cutoff)
- ▶ $Y = TY(1) + (1 - T)Y(0)$ – наблюдаемый исход

⁰Картинка из Imbens и Lemieux 2008.

⁰Также на русском можно почитать Ниворожкин 2009.

Четкая разрывная регрессия: Предположения



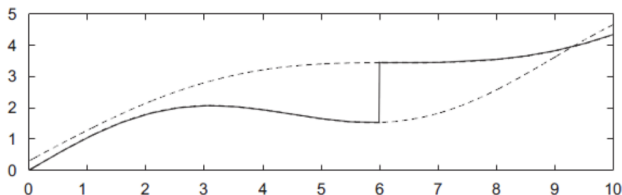
Для начала упражнение:

- ▶ $(Y(0), Y(1), R) \perp T$ – экзогенность. Выполнена ли? **нет**
- ▶ $(Y(0), Y(1)) \perp T | R$ – unconfoundedness. Выполнена ли? **верим**
- ▶ $1 > P(T|R) > 0$ – overlap. Выполнена ли? **нет**

Предположим непрерывность

$E(Y(1)|R), E(Y(0)|R), E(X|R)$ вместо независимости T и $(Y(0), Y(1))$

Четкая разрывная регрессия: Эффект локальный!



$$\tau_{SRD} = E[Y(1) - Y(0) | R = utoff] = \\ \lim_{R \rightarrow cutoff^+} E[Y | R] - \lim_{R \rightarrow cutoff^-} E[Y | R]$$

⁰Картинка из Imbens и Lemieux 2008.

⁰Также на русском можно почитать Ниворожкин 2009.

Четкая разрывная регрессия: оценка

- ▶ Выбираем ширину окна h
- ▶ Берем данные $R \in [c - h, c + h]$
- ▶ Оцениваем (простая версия)

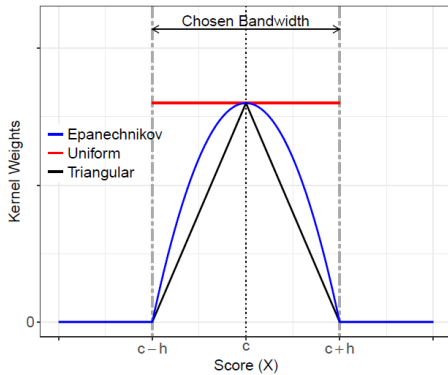
$$\hat{\tau}_{\text{rdd}} = \frac{1}{\sum_{(c+h > R > c)} \sum_{c+h > R > c} Y_i} - \frac{1}{\sum_{(c-h < R < c)} \sum_{c-h < R < c} Y_i}$$

Или с помощью регрессии на данных $R \in [c - h, c + h]$:

$$Y = a + \tau T$$

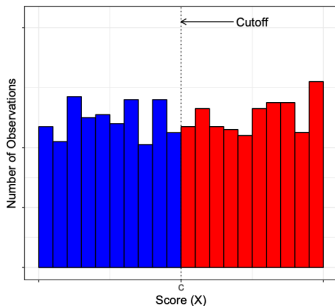
- ▶ Таким образом мы получаем локальный эффект
($\text{LATE} = \mathbb{E}(\tau | R = c)$)

Ядро

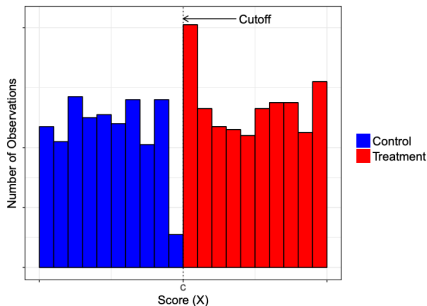


Проверка баланса¹

1. Проверить распределение данных по R (McCray test)
2. Проверить непрерывность X



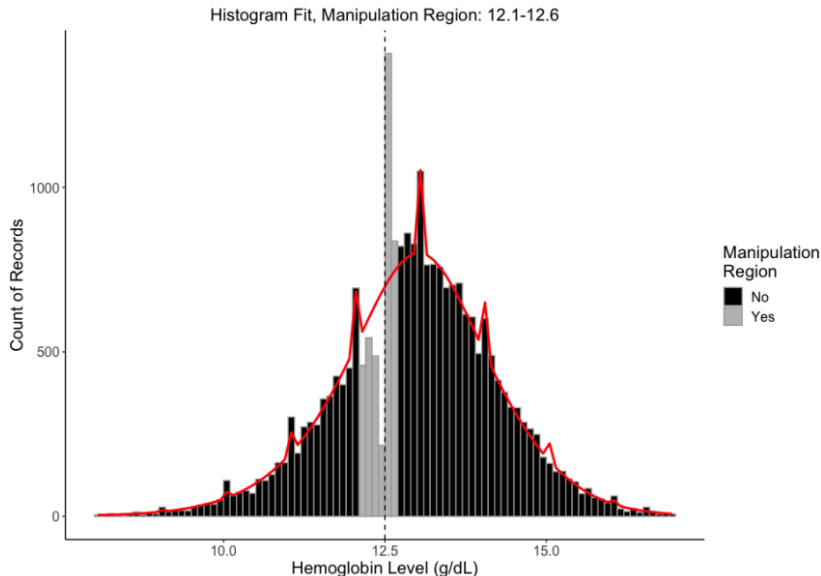
(a) No Sorting



(b) Sorting

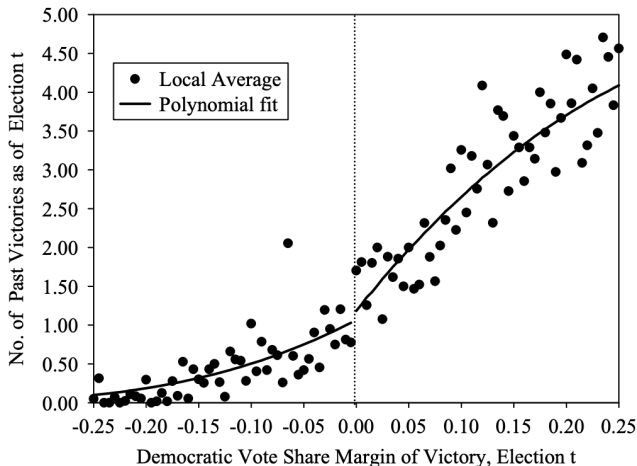
¹Картинка из [Matias D Cattaneo, Nicolás Idrobo и Rocio Titiunik \(2019\). A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge University Press](#), Тут также есть хороший обзор метода.

Манипуляция R - делаем donut hole пример

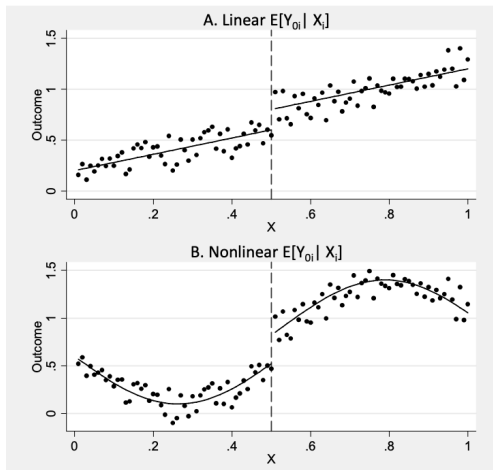


Проверка предположения непрерывности

1. Проверить непрерывность $Y(1)$, $Y(0)$ мы не можем
2. Но можем проверить непрерывность X

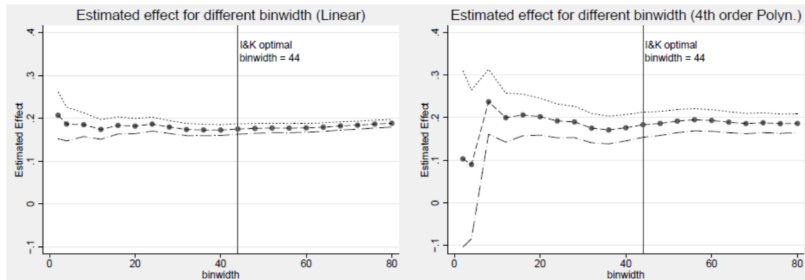


Контроль на тренд



¹Angrist и Pischke 2008, Глава 6.

Проверка чувствительности к ширине окне



Чем меньше окно, тем выше дисперсия (т.к. мало наблюдений), но меньше смещение (так как наблюдения более похожие друг на друга). Два случая нарисованы для 2 разных трендов.

Table of Contents

Разрывная регрессия

Четкая разрывная регрессия

Инструментальные переменные в экспериментах

Размытая разрывная регрессия

Терминология

- ▶ Инструмент Z (предложение)
- ▶ Переменная интереса T (реализация, принятие)
- ▶ Исход Y

Эффекты

Как устроен мир: $Z \rightarrow T \rightarrow Y$

- ▶ Reduced form (Intention to treat): $Z \rightarrow Y$
- ▶ First stage: $Z \rightarrow T$
- ▶ Second stage: $T \rightarrow Y$

Обозначения

- ▶ $Y(00), Y(10), Y(01), Y(11)$ – Зависимая переменная (**potential outcomes**)
- ▶ $T(1), T(0)$ – Переменная интереса (теперь тоже с потенциальными исходами)
- ▶ Z – Инструментальная переменная (**instrumental variable**)
- ▶ X – Независимые переменные (**Covariates**)
- ▶ Мы наблюдаем только (Y, Z, T, X) , где
 $T = ZT(1) + (1 - Z)T(0)$ – **observed treatment**
 Y = длинная формула – **observed outcomes**

¹Angrist и Pischke 2008, Глава 4.4.

Предпосылки

- ▶ $(Y_{ij}, T_i, X) \perp Z$ – рандомизация
- ▶ $Y(T, Z = 1) = Y(T, Z = 0)$ – exclusion restriction:
напрямую T влияет на Y , а не Z напрямую не влияет на Y
- ▶ SUTVA
- ▶ $T|Z = 1 \geq T|Z = 0$ – монотонность
 - ▶ Мы хотим, чтобы никогда не было $T_1 < T_0$, но бывало такое, что $T_1 > T_0$
 - ▶ Бывает одностороннее неповиновение тритменту (one-sided noncompliance)
 - ▶ Бывает двустороннее неповиновение тритменту (two-sided noncompliance)

Когда выполнены все эти предпосылки, мы говорим, что эффект **идентифицирован**.

Подробнее про предпосылки

- ▶ $P(T_1 \geq T_0) = 1$ – монотонность. На самом деле возможных случаев в только 4
 - ▶ Always takers: $T_1 = 1, T_0 = 1$
 - ▶ Compliers: $T_1 = 1, T_0 = 0$
 - ▶ Never takers: $T_1 = 0, T_0 = 0$
 - ▶ Defiers: $T_1 = 0, T_0 = 1$
- ▶ Two-sided noncompliance: no Defiers
- ▶ One-sided noncompliance: no Defiers and no Always takers

Пример одностороннего
неповиновения: encouragement design (Sommer
Zeger, 1991)

Table I. Mortality rates in control and programme villages, months
4–12, stratified by compliance

Study group	Complied	Children	Deaths	Mortality (per 1000)
Control	—	11,588	74	6·4
Treatment	—	12,094	46	3·8
	Yes	9,675	12	1·2
	No	2,419	34	14·1

Пример двустороннего неповиновения (Hirano et al., 2000)

1. В 1978-80 гг. в США рассылали письма с предложением сделать прививку от гриппа.
2. В чём отличие от предыдущего примера?

	Grand mean	Means		<i>t</i> -stat.	Means		<i>t</i> -stat.
		No letter $Z_i^{\text{obs}} = 0$	Letter $Z_i^{\text{obs}} = 1$		No flu shot $D_i^{\text{obs}} = 0$	Flu shot $D_i^{\text{obs}} = 1$	
Letter (Z_i^{obs})	0.514	0	1	—	0.475	0.631	-7.5
Flu Shot (D_i^{obs})	0.250	0.190	0.307	-7.3	0	1	—
Hospitalization (Y_i^{obs})	0.085	0.092	0.078	1.4	0.085	0.084	0.1

Формула

$$ITT_T = \bar{T}_{Z=1} - \bar{T}_{Z=0} = P(\text{compliers}) \quad (1)$$

$$ITT_Y = \bar{Y}_{Z=1} - \bar{Y}_{Z=0} \quad (2)$$

$$TE = \frac{ITT_Y}{ITT_T} = E(\tau | \text{Compliers}) = \text{LATE} \quad (3)$$

Откуда брать инструментальные переменные?

- ▶ Из опыта чужих исследований
- ▶ Миру известны десятки хороших инструментов.
Каждый из них вызывает много дебатов
- ▶ Если вы придумали свой инструмент это тянет на хорошее оригинальное исследование.

Table of Contents

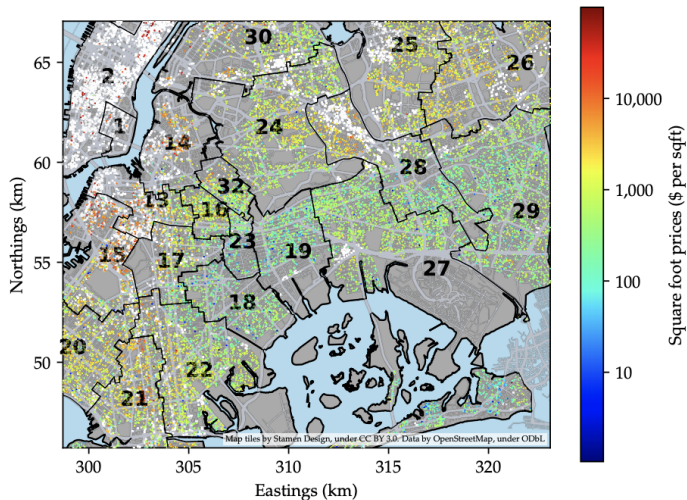
Разрывная регрессия

Четкая разрывная регрессия

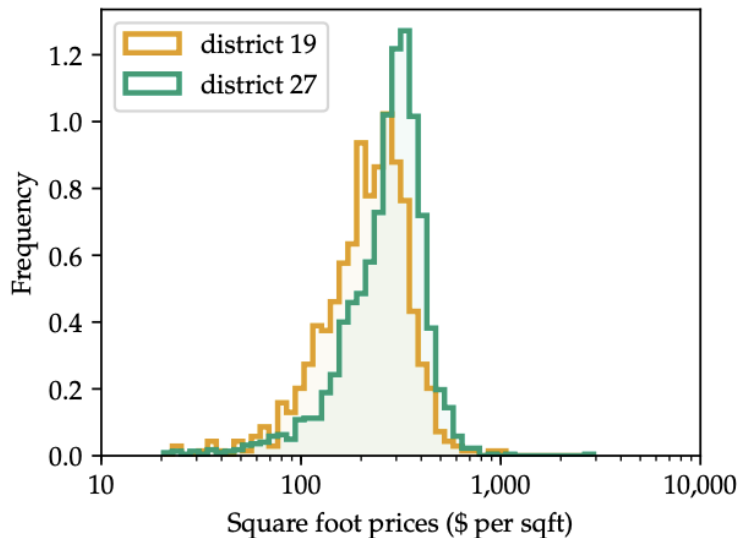
Инструментальные переменные в экспериментах

Размытая разрывная регрессия

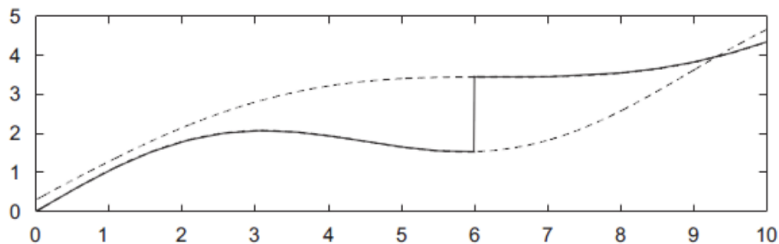
Пример: оценка спроса на образование в Нью-Йорке



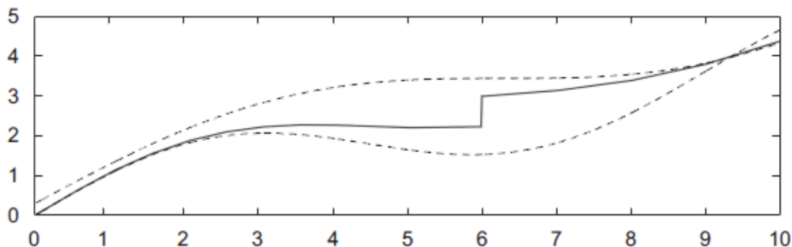
Плацебо тест



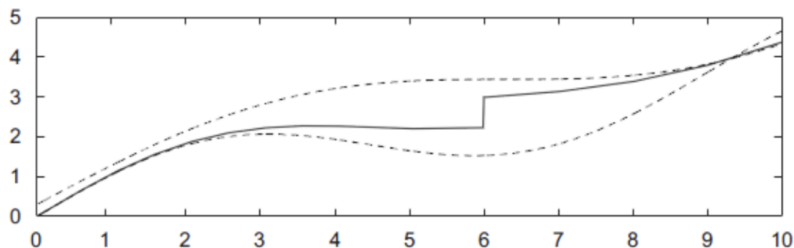
Fuzzy regression discontinuity



IV + Sharp RDD = Fuzzy RDD



Размытая разрывная регрессия: Предположения



- ▶ Что у нас инструментальная переменная? $Z = R > c$
- ▶ Что у нас переменная интереса? T
- ▶ Главное предположение разрывной регрессии:

Предположим непрерывность всего:

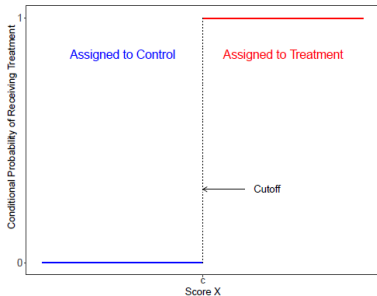
$$E(Y_{00}|R), E(Y_{01}|R), E(Y_{10}|R), E(Y_{11}|R),$$

$$E(T_0|R), E(T_1|R), E(X|R)$$

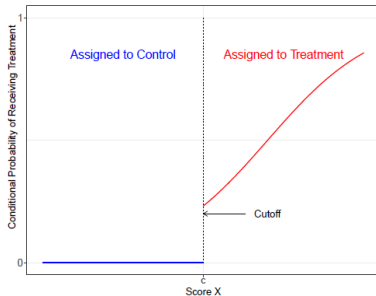
¹Angrist и Pischke 2008, Глава 6.

¹Картинка из Imbens и Lemieux 2008.

Попадание в тритмент sharp vs fuzzy

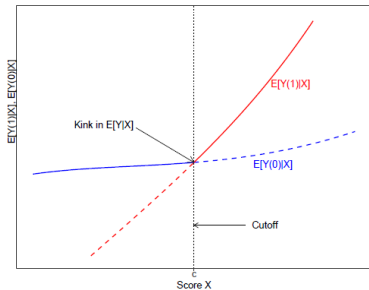


(a) Sharp RD

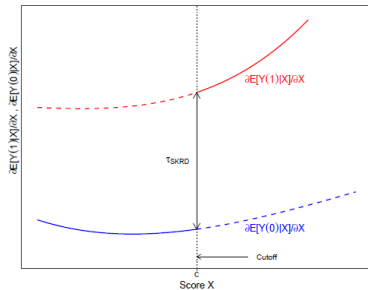


(b) Fuzzy RD (one-sided compliance)

RDD kink

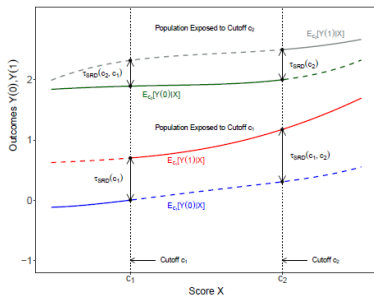


(a) Kink RD (levels)



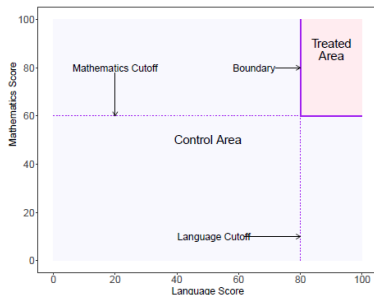
(b) Kink RD (derivatives)

Расширения



(a) Multi-cutoff:

$$\tau_{\text{SRD}}(x, c) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x, C_i = c]$$



(b) Multi-score:

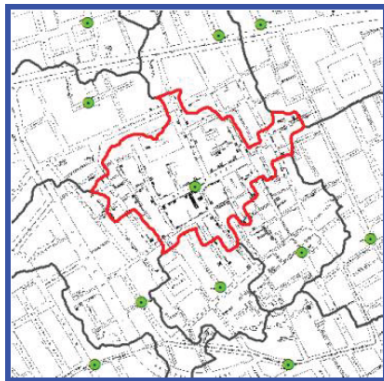
$$\tau_{\text{SRD}}(x_1, x_2) = \mathbb{E}[Y_i(1) - Y_i(0) | X_{1i} = x_1, X_{2i} = x_2]$$

Пример – холера в Лондоне 1854 (Ambrus 2020)

Panel A. John Snow's 1854 cholera map

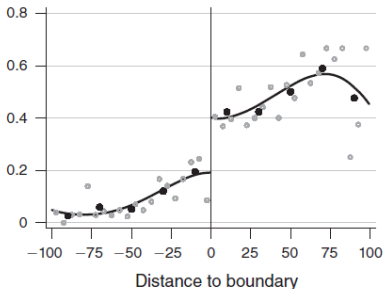


Panel B. Pumps' catchment areas

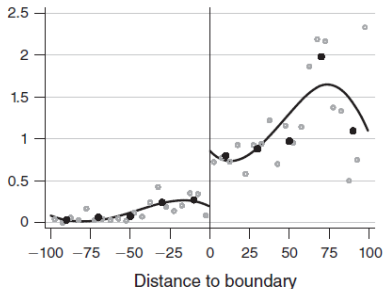


Смерти от холеры в Лондоне 1854

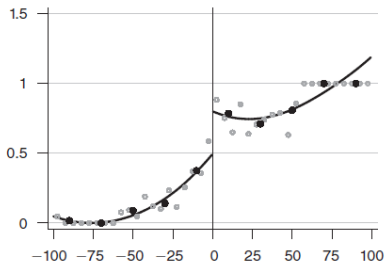
Panel A. House has at least one death



Panel B. Number of household deaths

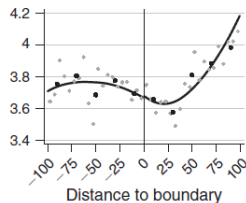


Panel C. At least 25% of houses in block have a death

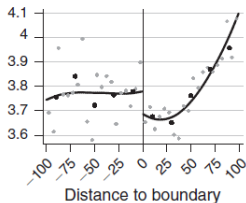


Стоимость аренды жилья в Лондоне

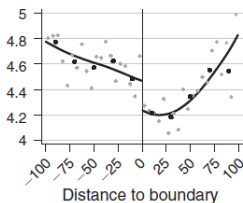
Panel A. Rental values, 1853



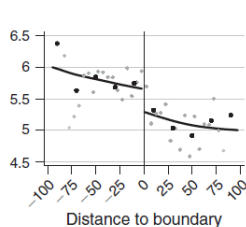
Panel B. Rental values, 1864



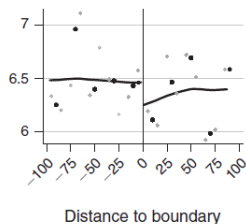
Panel C. Rental values, 1894



Panel D. Rental values, 1936



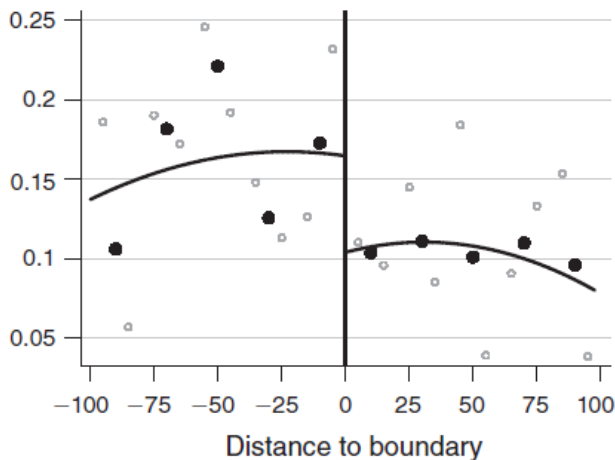
Panel G. Rental prices, 2015



Состав жителей

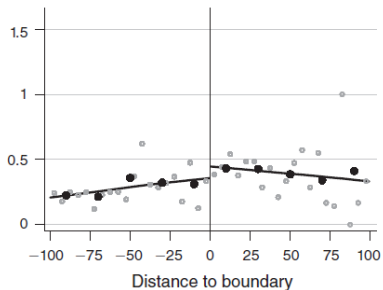
Механизм - «дурная слава» района и проблемы с водой

Panel D. Middle class

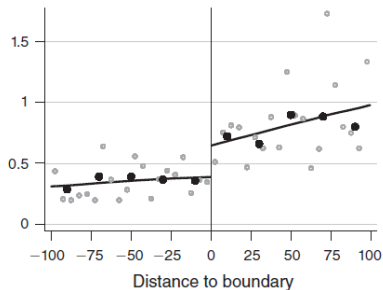


Состав жителей

Panel C. Number of immigrants, 1851

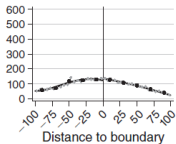


Panel D. Number of immigrants, 1861

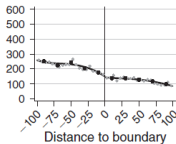


Проверка непрерывности ковариатов

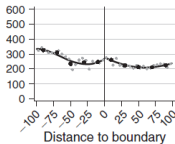
Panel A. Closest pump



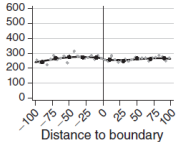
Panel B. Soho centroid



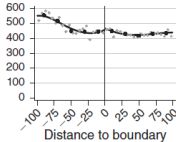
Panel C. Presumed plague pit



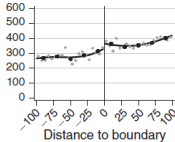
Panel D. Public square



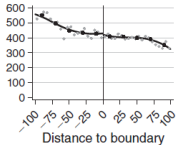
Panel E. Police station



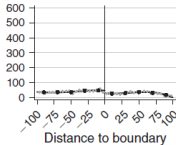
Panel F. Fire station



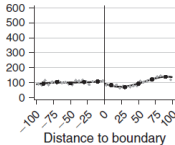
Panel G. Theater



Panel H. Pub

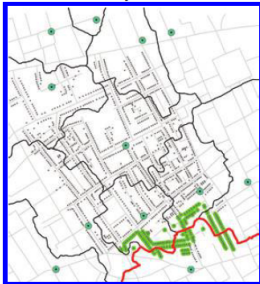


Panel I. Urinal

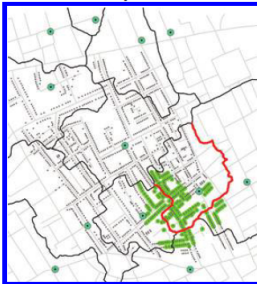


Плацебо-тест: ложные границы

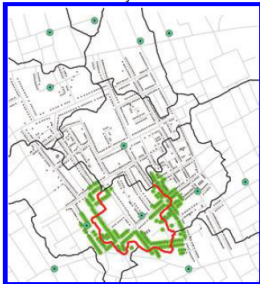
Panel A. Boundary 1



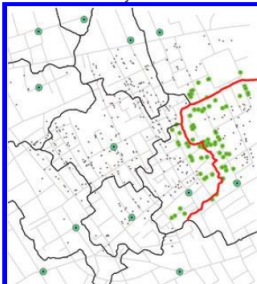
Panel B. Boundary 2



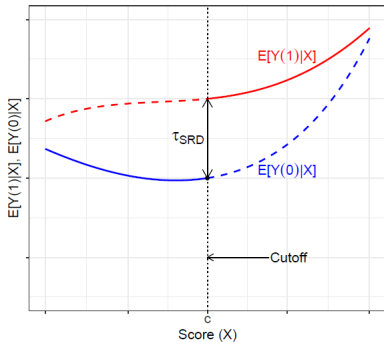
Panel C. Boundary 3



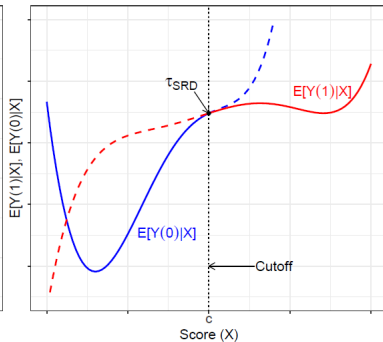
Panel D. Boundary 4



Внешняя валидность: RDD - локальная оценка



(a) Mild Heterogeneity



(b) Severe Heterogeneity