

Historical Persistence

Applied Economics Research Course

Bas Machielsen

Regression Discontinuity

Introduction

- ▶ In the previous classes, I have focused on data collection and spatial data wrangling
- ▶ In this class, I'll focus on *analyzing* these data
- ▶ In principle, there are various ways in which you can creative use spatial variation to identify a parameter of interest
- ▶ I'll just focus on one: a *geographic* regression discontinuity design
 - ▶ Before I do that, I'll recapitulate the basic intuitions and terminology for an ordinary regression discontinuity design

Regression Discontinuity

Basic Philosophy

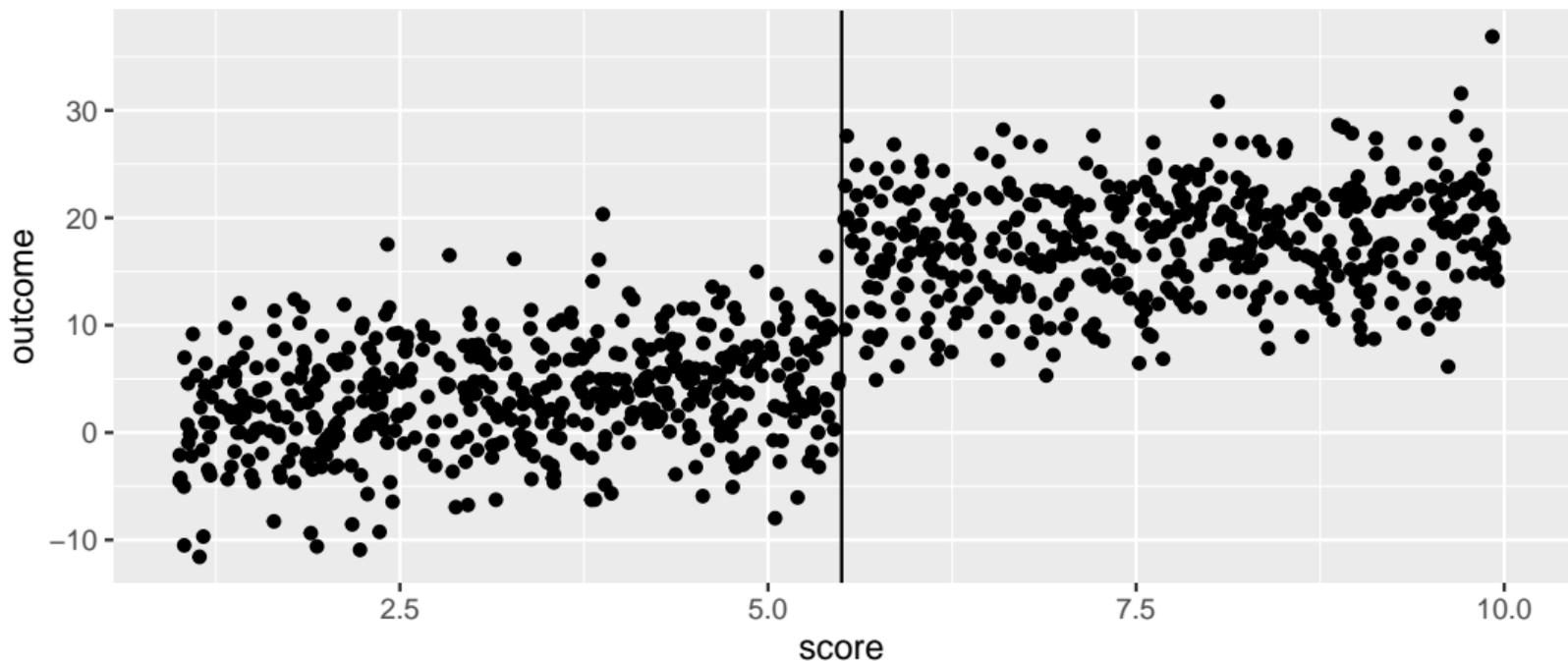
- ▶ A *discontinuity* occurs whenever some treatment is assigned discontinuously
 - ▶ People just on one side of a line get it, and people just on the other side of the line don't
 - ▶ Historical example: test admission requirement in schools
 - ▶ Students with some score $s > t$ can be admitted to a school
 - ▶ Students with a score $s < t$ cannot
- ▶ A regression discontinuity analysis compares people just on either side of a cutoff.
- ▶ Without the cutoff they'd likely be pretty similar. So if they're different, we can probably attribute the difference to whatever it is that happens at that cutoff.

Terminology

- ▶ *Running variable*: The running variable, also known as a forcing variable, is the variable that determines whether you're treated or not. For example, if schools admit students with a score higher than t , but not with a score lower than t , then score is your running variable.
- ▶ *Cutoff*: The cutoff is the value of the running variable that determines whether you get treatment. Using the test score example, if you get a score above t , you get the treatment, if you get something below t , you don't.
- ▶ *Bandwidth*: It's reasonable to think that people just barely to either side of the cutoff are basically the same other than for the cutoff. But people farther away (say, students with a very high score or a very low score) might be different for reasons other than the cutoff. The bandwidth is how much area around the cutoff you're willing to consider comparable.

RDD Graphically

- ▶ Let's make a simulated RDD example:
 - ▶ The influence of a test score on later wages through the treatment, admission to a school



RDD Interpretation

- ▶ In this example, we have simulated the *cut-off* to be 5.5.
 - ▶ The idea is that individuals who score higher than a 5.5 get access to the treatment.
 - ▶ And if you get access to the treatment, you get a higher outcome, e.g. wage, simulated from a normal distribution with $\mu = \text{score} + 10$ rather than $\mu = \text{score}$ if you don't receive the treatment
- ▶ The most basic version of RDD allows for a jump but forces the slope to be the same on both sides of the cut-off:

$$Y_i = \beta_0 + \beta_1 \times \text{Treated}_i + \beta_2 \times C_i + U_i$$

where $\text{Treated}_i = 1$ if you are above the cut-off, C_i is the running variable centered around the cut-off $X_i - t$, and our coefficient of interest is β_1 , how the intercept changes as a result of treatment status

RDD Estimation

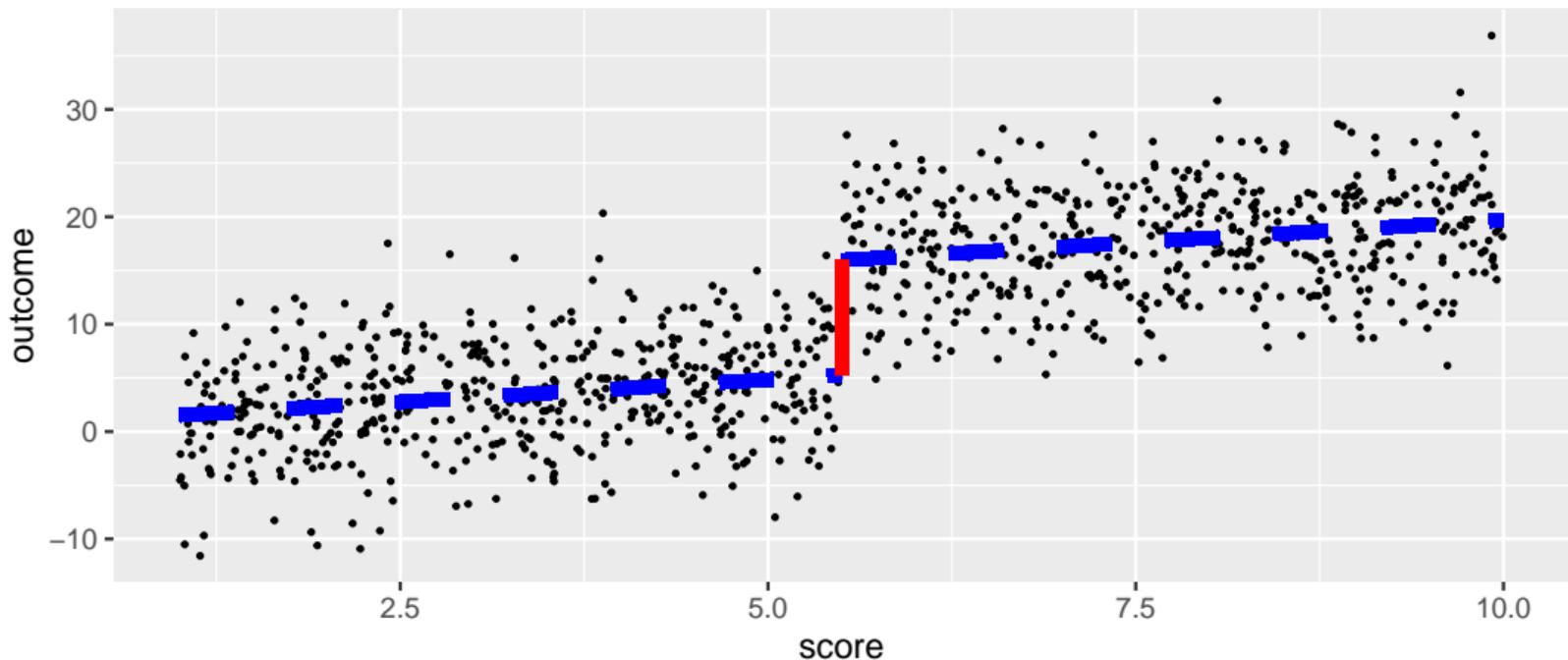
- ▶ This previous equation can be estimated on our data, for example, using the `feols` function from the `fixest` package:

<hr/>	
(1)	
<hr/>	
(Intercept)	5.247***
	(0.345)
treatment	10.706***
	(0.621)
xc	0.829***
	(0.120)
<hr/>	
Num.Obs.	1000
<hr/>	

+ p < 0.1, * p < 0.05, **
p < 0.01, *** p < 0.001

RDD Graphically

- ▶ Graphically, what we've done is this:
 - ▶ Blue is the *slope* we've estimated
 - ▶ Red is the *treatment effect, β_1 , we've estimated*



RDD Estimation: Choices

- ▶ We find a coefficient pretty close to 10, which is the number we had simulated as the true treatment effect
- ▶ This is however but a very raw approach.
- ▶ In reality, we have to make choices regarding:
 - ▶ Bandwidth
 - ▶ Functional form
 - ▶ Control variables

Choices: Bandwidth

- ▶ The idea of RDD is that people *close to* the cut-off are comparable, but people farther away might not be
- ▶ For example, your test score might be determined by your ability
 - ▶ So people with a low score have a very different ability than people with a high score
 - ▶ This is problematic, because then a comparison of outcomes between people with a high score and a low score might not only be due to the treatment, but also due to differences in ability
- ▶ However, we can assume that people within a narrow margin of the cut-off may be more similar to each other, therefore neutralizing the potential confounding influence of ability and other potential variables
 - ▶ People with a score of 5.4 and 5.6 are pretty similar, and it can be argued that their precise score is essentially due to randomness

Choices: Bandwidth

- ▶ Pay attention to the following sample sizes, effect sizes and standard errors:
 - ▶ As you can see, the effect approaches the true effect as we condition our sample to observations closer to the threshold

	(1)	(2)	(3)	(4)
(Intercept)	4.202*** (0.267)	4.459*** (0.321)	4.676*** (0.470)	4.909*** (0.700)
treatment	13.155*** (0.376)	12.399*** (0.462)	11.950*** (0.682)	11.875*** (0.999)
Num.Obs.	666	450	219	112

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Choices: Bandwidth

- ▶ The idea of RDD is that observations just around the cut-off are comparable
- ▶ So we might limit our analysis to just a narrow window around the cut-off
- ▶ However, how narrow? And how many observations do we then retain?
 - ▶ There is a literature on *optimal bandwidth selection*
 - ▶ As we will see, there is an R package that implements these optimal bandwidths automatically
 - ▶ And you can choose whether you want to have a different bandwidth on both sides of the threshold
 - ▶ This can be useful if you have much more observations in the treatment or control group

Choices: Functional Form

- ▶ Why would we only fit a straight line on both sides of the threshold?
- ▶ The relationship between the running variable and the outcome is potentially more complex
- ▶ A misspecified line can impact the RD estimates
- ▶ Hence, we want to estimate the relationship as well as possible
- ▶ A common way to do this is to use *non-parametric* regression
 - ▶ These are flexible lines that do not presuppose any shape
 - ▶ This is generally thought to be better than polynomials

Non-Parametric Regression: The Idea

- ▶ A **global** parametric model fits one \hat{f} for all x — e.g., $Y = \beta_0 + \beta_1 X$
- ▶ A **local** approach instead fits a *separate* weighted least squares problem at each target point x_0 :

$$\hat{\beta}(x_0) = \arg \min_{\beta} \sum_{i=1}^n K_h(X_i - x_0) \left(Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right)^2$$

- ▶ The fitted value at x_0 is $\hat{f}(x_0) = \hat{\beta}_0(x_0)$
- ▶ For RDD we use $p = 1$: **local linear regression**
- ▶ Observations far from x_0 receive low (or zero) weight via the kernel K_h

How the Local Problems Form a Curve

- ▶ We solve one $\arg \min$ problem **per evaluation point** $x_0 \in \mathcal{X}$
- ▶ Each solution yields a single predicted value $\hat{f}(x_0)$
- ▶ Sweeping x_0 across the range of X and collecting all $\hat{f}(x_0)$ traces out the smooth curve
- ▶ Intuition: imagine sliding a weighted window along the x -axis
 - ▶ At each position, fit a local line using only nearby observations
 - ▶ Record just the intercept (the predicted value at x_0)
 - ▶ The collection of intercepts **is** your smooth curve
- ▶ This is sometimes called a **moving regression**: nearby points contribute more (via the kernel) than distant ones
- ▶ Bandwidth h controls how wide the window is

Kernel Functions

- ▶ A kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ assigns weights by distance; $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ rescales by bandwidth h
- ▶ Three common choices:

Kernel	Formula	Notes
Uniform	$K(u) = \frac{1}{2}\mathbf{1}(\ u\ \leq 1)$	Equal weight within bandwidth
Triangular	$K(u) = (1 - \ u\)\mathbf{1}(\ u\ \leq 1)$	Linearly declining; default in rdrobust
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}(\ u\ \leq 1)$	MSE-optimal kernel

- ▶ All three assign zero weight to observations outside the bandwidth
- ▶ Results are typically robust to kernel choice; **bandwidth selection matters more**

Optimal Bandwidth Selection

- ▶ Large $h \Rightarrow$ more data \Rightarrow lower variance, higher bias (over-smoothing)
- ▶ Small $h \Rightarrow$ fewer observations \Rightarrow lower bias, higher variance (under-smoothing)
- ▶ MSE-optimal bandwidth minimises:

$$\text{MSE}(\hat{\tau}) = \text{Bias}^2(\hat{\tau}) + \text{Var}(\hat{\tau})$$

- ▶ Calonico, Cattaneo & Titiunik (CCT, 2014) show the MSE-optimal bandwidth satisfies:

$$h^* = C_n \cdot n^{-1/5}$$

where C_n depends on the curvature of the regression function and the kernel

- ▶ Implemented in `rdrobust`:

```
rdrobust::rdbwselect(y = df$outcome, x = df$score, c = 5.5) |> summary()
```

Local Linear Regression in Practice

Switching kernels in rdrobust:

```
# triangular kernel (default)
library(rdrobust)
rdrobust(y = df$outcome, x = df$score, c = 5.5, kernel = "triangular")

# Epanechnikov kernel
rdrobust(y = df$outcome, x = df$score, c = 5.5, kernel = "epanechnikov")

# uniform kernel
rdrobust(y = df$outcome, x = df$score, c = 5.5, kernel = "uniform")
```

- ▶ Results typically robust to kernel choice
- ▶ The bandwidth (and its selection method) has far more influence on the estimate

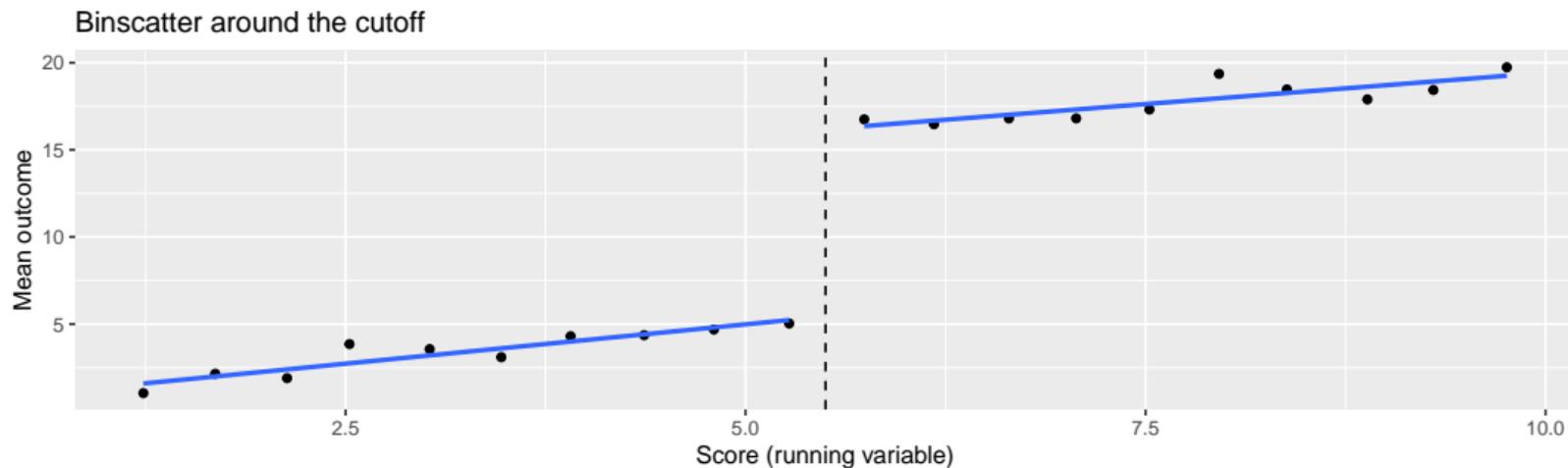
Binscatter: The Idea

- ▶ Binscatter **visualises** $E[Y | X]$ non-parametrically — no functional form assumption
- ▶ Algorithm:
 1. Partition the support of X into J equal-sized bins B_1, \dots, B_J
 2. Within each bin j , compute the bin mean \bar{X}_j and outcome mean \bar{Y}_j
 3. Plot (\bar{X}_j, \bar{Y}_j) for $j = 1, \dots, J$
- ▶ Formally, within bin j :

$$\bar{X}_j = \frac{1}{n_j} \sum_{i \in B_j} X_i, \quad \bar{Y}_j = \frac{1}{n_j} \sum_{i \in B_j} Y_i$$

- ▶ The non-parametric analogue of a scatter plot: reveals the *shape* of the relationship
- ▶ Often overlaid with a linear fit to compare parametric vs. non-parametric shape

Binscatter: Code and Plot



- ▶ The vertical dashed line marks the cutoff — the **jump** is the RD estimate

Choices: Control Variables

- ▶ In principle, because treatment assignment around the threshold is random, you don't need control variables
- ▶ In practice, they can make your estimate more precise
- ▶ Another reason for searching for control variables is that you can do **placebo tests, also called tests for covariate balance**
- ▶ This involves looking at other, pre-treatment variables than the outcome, and see whether they change discontinuously at the threshold
- ▶ If they do, that is a problem, because your treatment assignment is *correlated* with those variables
- ▶ If they don't, this makes the assumption of random treatment assignment stronger

Rdrobust Estimation

- ▶ In R, the package `rdrobust` implements the optimal bandwidths automatically, and it also takes care of the functional form
- ▶ It estimates local linear regression, hence evading the issue of polynomials

```
library(rdrobust)
res <- rdrobust(y=df$outcome, x=df$score, c=5.5)
```

- ▶ You can also add control variables in the following way:

```
res <- rdrobust(y=df$outcome, x=df$score,
               c=5.5,
               covs=cbind(df$var1, df$var2))
```

The rdrobust Package

Sharp RD estimates using local polynomial regression.

Number of Obs. 1000
BW type mserd
Kernel Triangular
VCE method NN

Number of Obs.	507	493
Eff. Number of Obs.	152	139
Order est. (p)	1	1
Order bias (q)	2	2
BW est. (h)	1.314	1.314
BW bias (b)	2.025	2.025
rho (h/b)	0.649	0.649
Unique Obs.	507	493

```
=====
```

Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
Conventional	11.941	1.385	8.624	0.000	[9.228 , 14.655]
Robust	-	-	7.295	0.000	[8.871 , 15.389]

```
=====
```

Covariate Balance

- ▶ The only thing it does not implement is the covariate balance tables
- ▶ You want to see that the treatment and control group are equal in terms of all observable pre-treatment characteristics:
 - ▶ Making it likely that they are also equal in terms of unobservables
- ▶ I demonstrate this using a toy data set:

	Control Group (N=16)		Treatment Group (N=16)		Diff. in Means	p
	Mean	Std. Dev.	Mean	Std. Dev.		
mpg	22.0	7.1	18.2	4.1	-3.8	0.078
cyl	5.9	2.0	6.5	1.5	0.6	0.331

Geographic Regression Discontinuity

Mining Mita

- ▶ Now, geographic regression discontinuity should be very simple for you to understand
- ▶ In geographic regression discontinuity, we use a geographic boundary as a *cut-off* and the distance to that boundary as a *running variable*
 - ▶ Seminal paper: Dell (2010, Econometrica):

This study utilizes regression discontinuity to examine the long-run impacts of the mita, an extensive forced mining labor system in effect in Peru and Bolivia between 1573 and 1812. Results indicate that a mita effect lowers household consumption by around 25% and increases the prevalence of stunted growth in children by around 6 percentage points in subjected districts today. Using data from the Spanish Empire and Peruvian Republic to trace channels of institutional persistence, I show that the mita's influence has persisted through its impacts on land tenure and public goods provision. Mita districts historically had fewer large landowners and lower educational attainment. Today, they are less integrated into road networks and their residents are substantially more likely to be subsistence farmers.

Warning

- ▶ Borders are generally *not* exogenous
 - ▶ Usually, borders are defined by some kind of geographical feature such as a river
 - ▶ In addition to the treatment variable, many things might change discretely at this boundary
- ▶ From Dell's (2010) paper:

[validity of the RD design] requires all relevant factors besides treatment to vary smoothly at the mita boundary (..), I focus exclusively on the portion that transects the Andean range in southern Peru. Much of the boundary tightly follows the steep Andean precipice, and hence has elevation and the ethnic distribution of the population changing discretely at the boundary. In contrast, elevation, the ethnic distribution, and other observables are statistically identical across the segment of the boundary on which this study focuses.

Dells' Strategy

- ▶ Dell considers the basic regression of the following form:

$$c_{idb} = \alpha + \gamma \times \text{mita}_d + X'_i\beta + f(\text{geographic location}_d) + \phi_b + \epsilon_{idb}$$

- ▶ where c_{idb} is an outcome of interest, X_i are control variables, and $f(\text{geographic location})$ are a set of geographical controls (longitude, latitude, elevation, temperature)
- ▶ Outcomes at level: village/household i in administrative district d along mita segment b
- ▶ The treatment is administered at the district level d
- ▶ Examples of c_{idb} : Household consumption, stunted growth
 - ▶ Mechanisms: Forced labor institutions vs. pro-market *haciendas*: rural estates with an attached labor force
 - ▶ Also: less integration in road networks, lowered educational levels

Seminal Papers - Wahl

- ▶ This theme is inspired by Wahl (2017)'s paper, *Does European development have Roman roots? Evidence from the German Limes*

This paper contributes to the understanding of the long-run consequences of Roman rule on economic development. In ancient times, the area of contemporary Germany was divided into a Roman and a non-Roman part. The study uses this division to test whether the formerly Roman part of Germany are more developed than the non-Roman part. This is done using the Limes Germanicus wall as geographical discontinuity in a regression discontinuity design framework. The results indicate that economic development—as measured by luminosity—is indeed significantly and robustly larger in the formerly Roman part of Germany. The study identifies the persistence of the Roman road network until the present as an important factor causing this developmental advantage of the formerly Roman part of Germany both by fostering city growth and by allowing for a denser road network.

Seminal Papers: Lowes & Montero

- ▶ Lowes & Montero also use a geographic regression discontinuity design by an arguably random border of the Congo Free State:

All colonial powers granted concessions to private companies to extract natural resources during the colonial era. In Africa, these concessions were characterized by indirect rule and violence. We use the arbitrarily defined borders of rubber concessions granted in the north of the Congo Free State to examine the causal effects of this form of economic organization on development. We find that historical exposure to the concessions causes significantly worse education, wealth, and health outcomes. To examine mechanisms, we collect survey and experimental data from individuals near a former concession boundary. We find that village chiefs inside the former concessions provide fewer public goods, are less likely to be elected, and are more likely to be hereditary. However, individuals in the concessions are more trusting, more cohesive, and more supportive of sharing income. We interpret our results as reflecting how institutions and culture may act as substitutes in the context of the widespread colonial concession system.

Conclusion

- ▶ We have seen a recapitulation of regression discontinuity designs in economics
- ▶ We have seen the importance of choices regarding bandwidth, functional form and control variables
- ▶ We have learned to use the `rdrobust` package for RD analyses and learn to analyze the output
- ▶ We have learned to validate the RD design's assumptions by means of balancing tables
- ▶ As examples, we have studied several geographic regression discontinuity designs in the economics literature

Resources

- ▶ You can have a look at the code underlying this presentation [here](#).
- ▶ Here is a function that allows you to create nice tables using `rdrobust`.