# Lecture 5: Distance, Buffers, and Map Algebra

Paul A. Raschky,

University of St Gallen, October 2017

- Geographic Regression Discontinuity (GRD)
- A research design in which a geographic, administrative or other boundary splits units into treated and control areas.
- Identifying assumption is that the division into treated and control areas occurs in an as-if random fashion.

- Standard Regression Discontinuity (RD) has cut-off (i.e. test results cut-off) and forcing variable that measures the difiference to the cut-off.
- Geographic Regression Discontinuity uses a border as the cut-off and the project coordinates into distance to the cut-off (D<sub>i</sub>)
- *D<sub>i</sub>* is positive for treatment group and positive for control group.

$$y_i = \alpha T_i + \beta D_i + \gamma T_i D_i + \epsilon_i \tag{1}$$

where

- ► *T<sub>i</sub>* is the binary treatment variable
- $D_i$  is the distance to the border.

- 1. Compound Treatment
  - A situation in which two ore more treatments that affect the outcome of interest occur simultaneously.
  - In geographic applications, compound treatments typically arise when two or more geographically defined borders are located at the sample place.
  - For example, county boundaries often coincide exactly with school district boundaries.

1. Compound Treatment



1. Compound Treatment



2. Naive Distance



#### 2. Naive Distance - Potential Solution

$$y_i = \alpha T_i + \beta D_i + \gamma T_i D_i + B_i + \epsilon_i$$
(2)

where

► *B<sub>i</sub>* is a vector of border segment fixed effects

3. Geographic treatment heterogeneity



3. Geographic treatment heterogeneity

$$y_i = \alpha T_i + \mathbf{x}'_i \beta D_i + T_i \mathbf{x}'_i \gamma + N_i + u_i$$
(3)

where

- x<sub>i</sub> are coordinates.
  - Coordinates are used instead of distance as RD polynomials
- Pick (equally-spaced) points on boundary (denoted by N)
- Assign each observation to its nearest boundary point, to define N.

- Question: Are there persistent effects of the labor system, Mita during Spanish colonial rule affect contemporary development in Peru?
- Identification strategy exploits clearly defined mita boundaries.
  - ► A forced labor system instituted by the Spanish government in Peru and Bolivia in 1573 and abolished in 1812.
  - The system The required over 200 indigenous communities to send one-seventh of their adult male population to work in the silver and mercury mines.



Image sources: Dell (2010)

The effect of Mita

- 1. In districts without Mita, large landowners (*haciendas*) could develop.
- 2. Non-Mita districts had more clearly defined land property rights (Mita districts adopted communal land tenure)
- 3. Hacienda elite was able to provided more public goods (security, roads).
- 4.  $\rightarrow$  positive effects on contemporary economic outcomes.

Outcome variables:

- Household consumption
- Road length
- Height of children (6-9)

Estimation Strategy:

 $c_{idb} = \alpha + \gamma mita_d + X'_{id}\beta + f(\text{geographic location}_d) + \phi_b + \varepsilon_{idb},$ 

Source: Dell (2010)

#### Main Results:

	Dependent Variable						
Sample Within:	Log Equiv. Hausehold Consumption (2001)			Stunted Growth, Children 6-9 (2005)			
	<100 km of Bound.	<75 km of Bound.	<50 km of Bound.	<100 km of Bound.	<75 km of Bound.	<50 km of Bound.	Border District
		Panel A	. Cubic Polynomial in	Latitude and Longitu	de		
Mita	-0.284	-0.216	-0.331	0.070	0.084*	0.087*	0.114**
	(0.198)	(0.207)	(0.219)	(0.043)	(0.046)	(0.048)	(0.049)
$R^2$	0.060	0.060	0.069	0.051	0.020	0.017	0.050
		Pane	B. Cubic Polynomial	in Distance to Potosí			
Mita	-0.337***	-0.307***	-0.329***	0.080***	0.078***	0.078***	0.063*
	(0.087)	(0.101)	(0.096)	(0.021)	(0.022)	(0.024)	(0.032)
$R^2$	0.046	0.036	0.047	0.049	0.017	0.013	0.047
		Panel C.	Cubic Polynomial in I	Distance to Mita Bound	darv		
Mita	-0.277***	-0.230**	-0.224**	0.073***	0.061***	0.064***	0.055*
	(0.078)	(0.089)	(0.092)	(0.023)	(0.022)	(0.023)	(0.030)
$R^2$	0.044	0.042	0.040	0.040	0.015	0.013	0.043
Geo. controls	yes	yes	yes	yes	yes	yes	yes
Boundary F.E.s	yes	yes	yes	yes	yes	yes	yes
Clusters	71	60	52	289	239	185	63
Observations	1478	1161	1013	158,848	115,761	100,446	37,421

Source: Dell (2010)

### 2. Application Dell (2010) Main Results:



### 2. Application Dell (2010) Main Results:



Source: Dell (2010)

### 3. Map Algebra

#### Exercise 5 - Preparing Geographic Data for GRD