UNIVERSITY OF ILLINOIS

AT URBANA-CHAMPAIGN

Regression discontinuity and heterogeneous effects

illinois.edu

Outline

- Regression discontinuity design
- Accounting for heterogeneity of treatment
- Summary

Regression discontinuity

- Looking for exogenous assignment into treatment
- Consider a policy where there is a threshold to receive treatment
- Households or individuals are similar, but get very different treatment depending on which side of the threshold they sit
- Cut-off is exogenous to households

Examples of thresholds

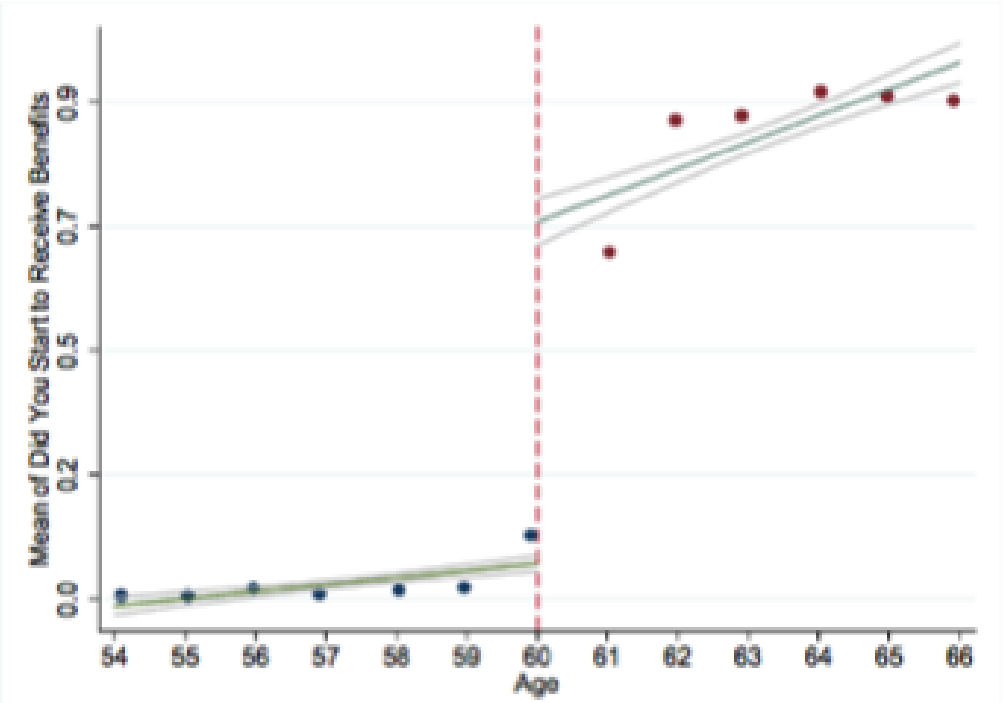
- Pensions available to senior citizens in China at age 60
- Scholarships and school admissions given to top n applicants
- Extension services targeted to those with landholdings of less than 5 hectares
- Rural roads ranking for priority for repair—Sierra Leone
- Effect of public research funding in Chile on researcher productivity

When can you use RD?

- Discontinuity in program eligibility
- Relationship between variable that determines program eligibility and the outcome of interest in smooth
- People do not change their behavior to become eligible (no age heaping at eligibility age)
- Sufficient number of observations around the discontinuity
- Valid even if selection on unobservables

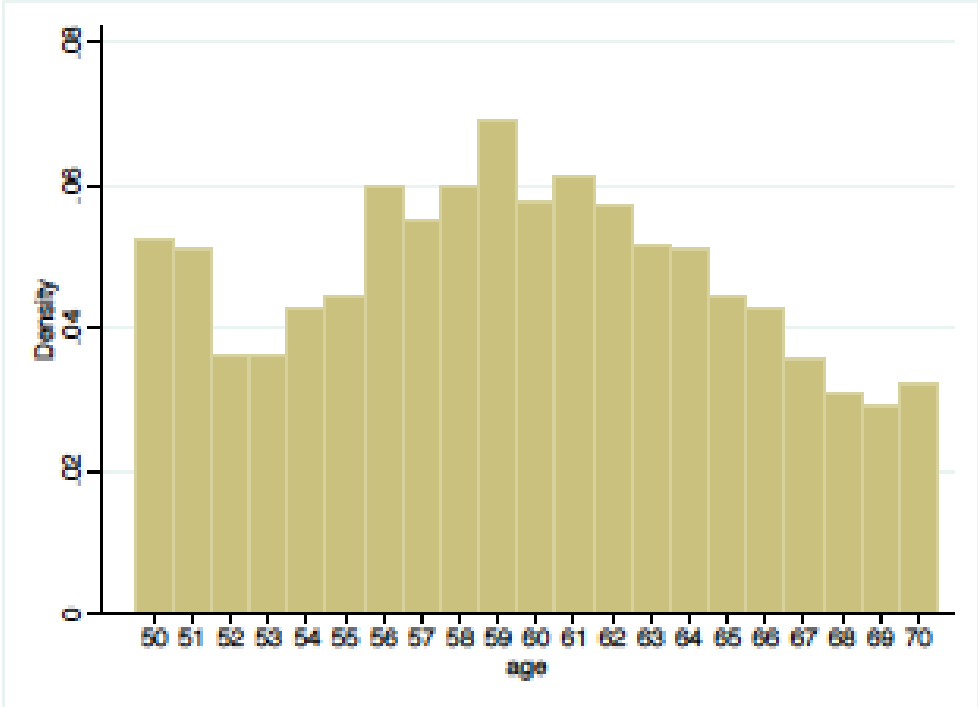
RD—Pension receipt in China

Figure 4: Incomplete Compliance at Age Threshold

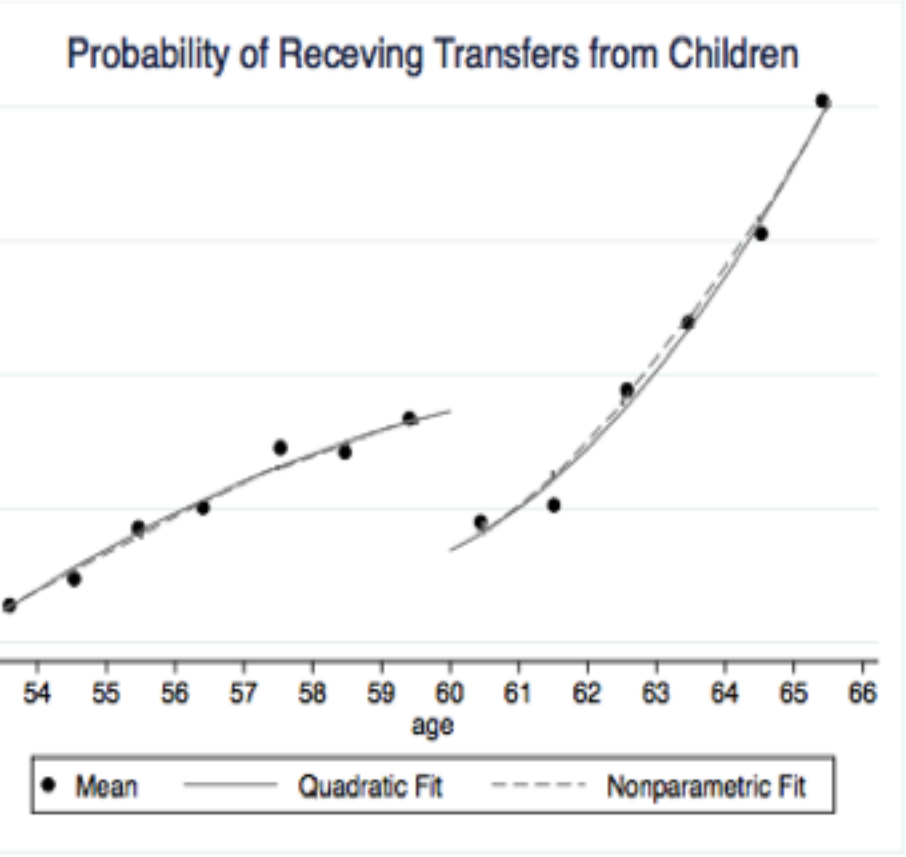


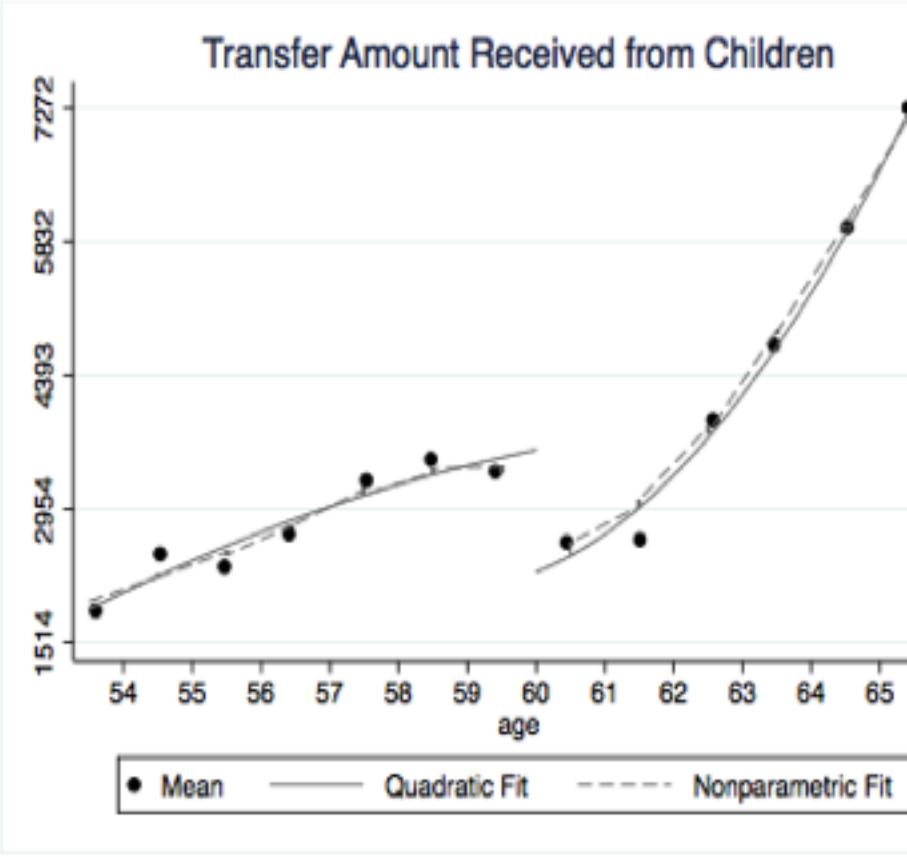
Check balancing

Figure 3: Age Distribution of Chinese Rural Elders

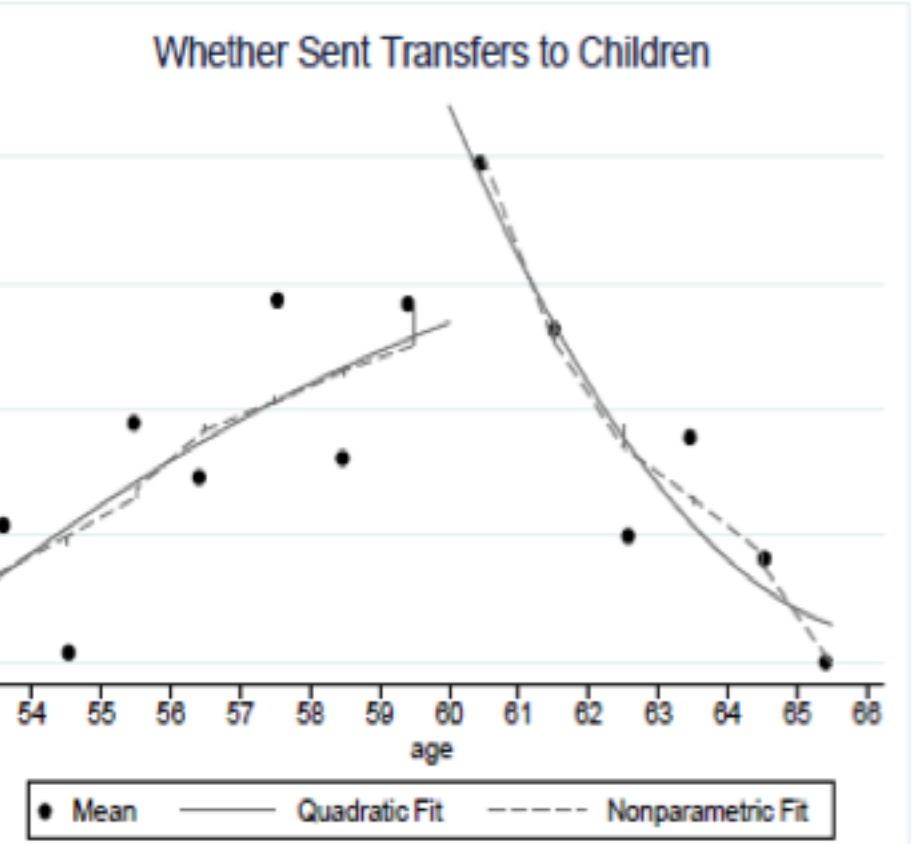


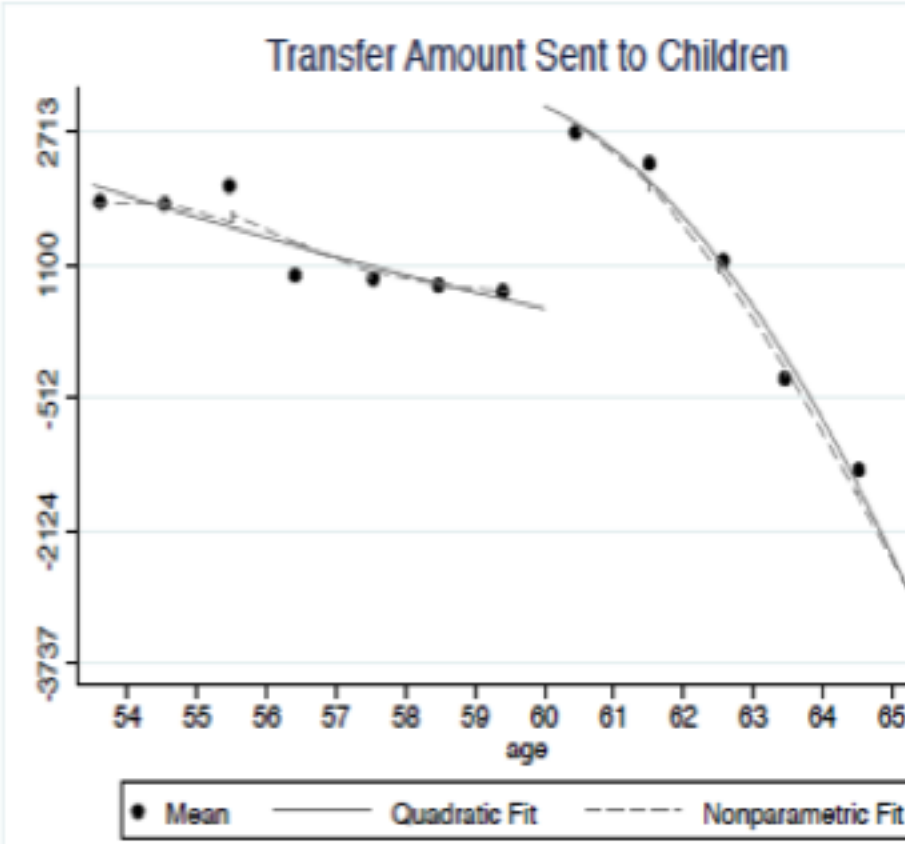
RD—transfers from children





RD—transfers to children





Sharp regression discontinuity

- Household eligible for irrigation program if landholding is below a threshold s*

If program targeted to poor based on landholding, receive program if

$$S_i \leq s^*$$

Look at a small bandwidth around s* called ε.

$$E[y_i \mid s^* - \varepsilon] - E[y_i \mid s^* + \varepsilon] = E[\beta S_i \mid s^* - \varepsilon] - E[\beta S_i \mid s^* + \varepsilon]$$

$$\lim_{\varepsilon \to 0} E[y_i \mid s^* - \varepsilon] - \lim_{\varepsilon \to 0} E[y_i \mid s^* + \varepsilon] = y^- - y^+ = \beta(S^- - S^+)$$

$$\Rightarrow \beta = \frac{y^- - y^+}{S^- - S^+}$$

Fuzzy regression discontinuity

- Substitute for s with probability of participating, P(S) = E(T|S) where T=1 if treated and T=0 otherwise.

$$\Rightarrow \beta_{frd} = \frac{y^- - y^+}{E(T \mid S^-) - E(T \mid S^+)}$$

- This estimator is a LATE estimator and analogous to IV.

How to estimate?

- Local linear regressions on outcome y, given a set of covariates, on both sides of the threshold
- Use parametric and non-parametric methods
- Problematic to include higher order polynomials
- Choice of bandwidth and the optimal bandwidth
- State of the art is changing very rapidly
- New Stata program rdrobust
- Afternoon exercise—use locpoly

Resources for RD regression

- Articles by Calonico, Cattaneo, Titiunik
- McCrary (2008)—test for smoothness of forcing variable around the threshold
- Imbens and Lemieux (2008)
- Gelman and Imbens (2014)
- Stata Journal article about rdrobust

Pair and share--RD

- Can you think of a research study you could do using RD?
- Start by thinking about strict eligibility requirements

Should you use RD?

- Advantages
  - Unbiased estimate of treatment at the discontinuity
  - It's common to have eligibility rules for participants to receive benefits
  - Unlike randomization, you don't have to exclude eligibles from treatment

Should you use RD?

- Disadvantages
  - Produces local average treatment effects that may not be generalizable
  - Effect estimated at discontinuity, so you may have very few observations
  - Specification can be sensitive to functional form.

Heterogeneity in impact

- The average treatment effect may not be the right effect to examine
- Examine effects in other parts of the distribution
- Effects on poverty—program helps very poor, or those close to poverty line?

Nutrition example

- Kandpal (2011) looks at impact Indian Integrated Child Development Services on child health
- Previous studies found no effect
- Matching—programs targeted in poor areas
- Found no effect over the distribution, but effects for severely stunted boys in 1992-93.
- In 1998-99, found effect for moderately stunted boys

Agricultural examples

- A new drought resistant variety of wheat only shows effects if it is a drought year
- Varying slopes and intercepts based on crops
- Impacts of CA are really crop specific
- Interact CA dummy by rainfall shocks (resiliency)
- Normal times, CA has zero to negative effect
- CA builds resiliency to drought for maize, sorghum, cowpea

Heterogeneity in RCTs

- Growing awareness in the field about ex post hypothesis testing
- Suppose I do an RCT, and I find no significant effect
- I start testing subsamples—gender, age, land, etc.
- If I run enough regressions, by chance, I would expect that some would be statistically significant
- Type 1 error

Solutions for RCTs

- Registries such as 3ie
- I indicate which groups I am going to analyze in advance
- Significance levels—adjust for multiple hypothesis testing
- Bonferroni method

Blogs to read

- Development Impact blog of the World Bank
- http://blogs.worldbank.org/impactevaluations/
- Evidence Matters—3ie
- http://blogs.3ieimpact.org
- Innovations for Poverty Action
- http://www.poverty-action.org/blog
- Chris Blattman
- http://chrisblattman.com