Causal Analysis

Impact Evaluation and Causal Machine Learning with Applications in R

Chapter 9: Regression Discontinuity, Kink, and Bunching Designs

Table of Contents

Introduction

Regression discontinuity design (RDD) (Thistlethwaite and Campbell, 1960):

- Treatment effect estimation relies on assumption that at some threshold of a running variable, the treatment status changes from 0 to 1 for everyone (sharp RDD) or for a share of the population (fuzzy RDD).
- Requires that individuals slightly above and below the threshold are similar in unobserved characteristics (which may affect the outcome).
- Aims to evaluate the treatment effect locally for the subpopulation at the threshold.

Example

Example

- Treatment effect of extended eligibility of unemployment benefits that individuals aged 50 and older are entitled to, see Lalive (2008).
- Sharp RDD: all individuals above a specific age threshold are treated (eligible), while no individual below the threshold is treated.
- Compare treated and untreated outcomes (e.g., unemployment duration) slightly above and below the threshold of 50 years.
- This comparison relies on the assumption that the individuals slightly above and below the threshold are similar in background characteristics.

Sharp RDD - Notation and Identifying Assumptions in

Notation:

- Running variable R with threshold value r_0.
- In sharp RDDs, the treatment is deterministic in R and takes the value one whenever the threshold is reached or exceeded, $D = I\{R \geq r_0\}$.

Sharp RDD identifying assumptions (Hahn, Todd, and van der Klaauw, 2001)

- Mean potential outcomes given the running variable, $E[Y(1)|R]$ and $E[Y(0)|R]$, must be continuous and sufficiently smooth around $R = r_0$.
- The density of the running variable R must be continuous and larger than zero around the threshold $R = r_0$.

Sharp RDD - Effect Identification

- Under these assumptions, a local average treatment effect at the threshold, Δ_{r_0}, is identified:

$$
\begin{aligned}
&\lim_{\epsilon \to 0} E[Y|R \in [r_0, r_0 + \epsilon)] - \lim_{\epsilon \to 0} E[Y|R \in [r_0 - \epsilon, r_0)] \\
&\quad = \lim_{\epsilon \to 0} E[Y(1)|R \in [r_0, r_0 + \epsilon)] - \lim_{\epsilon \to 0} E[Y(0)|R \in [r_0 - \epsilon, r_0)] \\
&\quad = E[Y(1) - Y(0)|R = r_0] = \Delta_{r_0}. \qquad (9.1)
\end{aligned}
$$

- Identification based on treated and nontreated outcomes in a neighborhood $\epsilon > 0$ around the threshold when letting ϵ go to zero, $\epsilon \to 0$.
- The treatment effect corresponds to the difference in potential outcomes under treatment and nontreatment at the threshold.

Sharp RDD - Graphical Illustrations

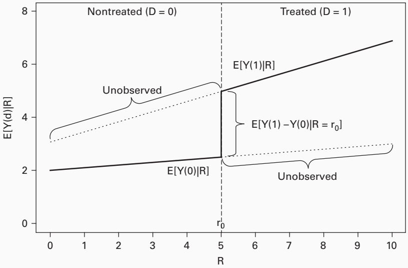

Figure 1: Sharp regression discontinuity design.

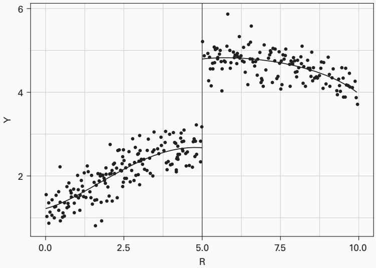

Figure 2: Observations and regression functions above and below the threshold.

Fuzzy RDD - Notation and Identifying Assumptions

Additional notation:

- Potential treatment state $D(Z)$ given binary threshold indicator $Z = I\{R \geq r_0\}$, which serves as instrument for treatment.

Fuzzy RDD identifying assumptions (Dong, 2014)

- Share of treated units must change discontinuously at the threshold. $\Rightarrow$ There are compliers whose treatment status changes from 0 to 1 at the threshold.
- There are no defiers (monotonicity) around the threshold.
- Shares of compliers, always takers, and never takers, and their potential outcomes under treatment and nontreatment are continuous around the threshold.

Fuzzy RDD - LATE Identification

- These assumptions are related to the IV setup in chapter 6, see expression (6.5).
- The first-stage effect of instrument Z (passing the threshold) on treatment D at the threshold $R = r_0$, denoted by γ_{r_0}, is:

$$\begin{aligned}
&\lim_{\epsilon \to 0} E[D|R \in [r_0, r_0 + \epsilon)] - \lim_{\epsilon \to 0} E[D|R \in [r_0 - \epsilon, r_0)] \\
&\quad = \lim_{\epsilon \to 0} E[D(1)|R \in [r_0, r_0 + \epsilon)] - \lim_{\epsilon \to 0} E[D(0)|R \in [r_0 - \epsilon, r_0)] \\
&\quad = E[D(1) - D(0)|R = r_0] = \gamma_{r_0}, \qquad (9.2)
\end{aligned}$$

while equation (9.1) identifies the ITT effect, θ_{r_0}.

- The LATE for compliers at the threshold is identified as the ratio of the ITT effect to the first-stage effect:

$$\Delta_{D(1)=1, D(0)=0, R=r_0} = \frac{\theta_{r_0}}{\gamma_{r_0}}. \qquad (9.3)$$

Fuzzy RDD - Graphical Illustration

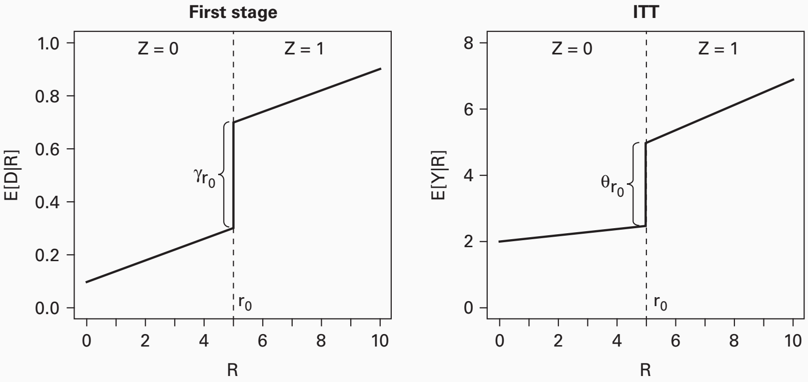

Figure 3: The fuzzy regression discontinuity design.

Regression Approach to RDD

- In empirical applications (considering the sharp RDD), OLS regression is frequently used for estimating $E[Y|D=0, R<r_0]$ and $E[Y|D=1, R \geq r_0]$ within some bandwidth ϵ around r_0.
- This approach estimates Δ_{r_0} as the difference of the regression functions at $R=r_0$.
- This may be implemented based on the following regression equation:

$$
\begin{aligned}
&E[Y|R \in [r_0-\epsilon, r_0+\epsilon)], D = I\{R \geq r_0\}] \\
&\quad = \alpha + \beta_R R + \underbrace{\beta_D I\{R \geq r_0\}}_{E[Y|R=r_0,D=1]-E[Y|R=r_0,D=0]} + \beta_{R,D} R \cdot I\{R \geq r_0\}, \qquad (9.4)
\end{aligned}
$$

  with β_D capturing the causal effect.
- Interaction term $R \cdot I\{R \geq r_0\}$ allows for regression lines with different slopes above and below the threshold.

Higher-Order Terms and Bandwidth Selection

Bias-variance trade-off when selecting model parameters:

- By adding higher order terms of R and interacting them with the treatment D we can increase model flexibility and decrease bias.
- However, this may come at the cost of increased variance.
- Similarly, decreasing the bandwidth ϵ may decrease bias at the cost of increased variance.

Optimal Bandwidth Selection

Some approaches for selecting the optimal bandwidth ϵ:

1. Compute the MSE as a function of ϵ using the analytic formula of Imbens and Kalyanaraman (2012).
2. Leave-one-out cross-validation (Ludwig and Miller, 2007):
   - For each observation i close enough to the threshold, regress Y on R, leaving out R_i, and predict Y for R_i.
   - To mimic RDD regression at the threshold, estimate regression function locally: $R_i - \epsilon \leq R < R_i$ if R_i is below the threshold, $R_i < R \leq R_i + \epsilon$ if R_i is above the threshold.
   - Calculate the MSE.
   - Repeat these steps for various ϵ and select bandwidth which minimizes the MSE.
3. Estimate placebo treatment effects for different model configurations (Kettlewell and Siminski, 2020).

Statistical Inference

Optimal parameter choices for estimation might not be optimal for inference:

- Bandwidth optimal for effect estimation is generally too large for statistical inference.

Improved methods for inference:

- Calonico, Cattaneo, and Titiunik (2014): Use one order higher for inference. Example: if Δ_{r_0} is estimated based on (local) linear regression, compute standard errors and confidence intervals using a quadratic regression.
- Armstrong and Kolesár (2018): Compute worst-case bias given a specific bandwidth.
- Cattaneo, Frandsen, and Titiunik (2015): Propose randomization inference, related to methods for synthetic controls (chapter 8).

Testable Assumptions

McCrary (2008) test:

- Test continuity of the running variable at the threshold.
- Rejection indicates discontinuity of the running variable and points to selective bunching at one side of the threshold.

Lee (2008) test:

- Explore if observed characteristics X are balanced above and below the threshold.
- Any X affecting the outcome must be balanced under the continuity assumption of the potential outcomes.

Weakening continuity assumption:

- Continuity assumption might hold conditional on covariates X, implying that all potential confounders are observed.

Extrapolating Local Effects

RDD estimates are local:

- Estimates from RDD are local, meaning they only represent the effect on a subpopulation.
- Sharp RDDs yield the effect for individuals at the threshold.
- Fuzzy RDDs yield the effect only for compliers at the threshold.

Methods to extrapolate effects:

- Compute the derivative of the treatment effect at the threshold (Dong and Lewbel, 2011).
- Test if running variable's association with the outcome vanishes conditional on covariates X (Angrist and Rokkanen, 2015).
- Test for equality in mean outcomes of treated compliers and always takers, as well as of untreated compliers and never takers (Bertanha and Imbens, 2019).

Table of Contents

Regression Kink Designs

Regression kink designs (RKD) (Card, Lee, Pei, and Weber, 2015):

- While RDDs exploit discontinuities in the treatment at the threshold, RKD exploits changes in the slopes of the treatment.
- RKDs suitable for continuous treatments with a kink at threshold $R = r_0$ (while RDDs are based on binary treatments).
- First derivative of treatment variable D with respect to R must be discontinuous at the threshold.

Regression Kink Design Example

Example

Example inspired by Landais (2015):

- Consider unemployment benefits (D) which are a percentage (e.g., 80%) of the previous wage R up to a maximum threshold r_0 (e.g., 5,000 EUR).
- Beyond r_0, benefits D remain constant.
- The derivative of D with respect to R is 0.8 for $R < r_0$ and zero for $R \geq r_0$.
- This deterministic relationship between D and R can be exploited by a sharp RKD.

Sharp RKD (1)

- Identification in sharp RKD requires continuously distributed potential outcomes and a smooth density of the running variable around the threshold (similar to sharp RDD).
- The RKD consists of scaling the (reduced form) change in the first derivative of the mean outcome Y with regard to R at the threshold by the (first-stage) change in the first derivative of D with regard to R at the threshold.
- Identifies the marginal treatment effect at the threshold:

$$\Delta_{r_0}(d_0) = \frac{\partial E[Y(d_0)|R = r_0]}{\partial d_0} = \frac{\lim\limits_{\epsilon \to 0} \frac{\partial E[Y|R \in [r_0, r_0+\epsilon)]}{\partial r_0} - \lim\limits_{\epsilon \to 0} \frac{\partial E[Y|R \in [r_0-\epsilon, r_0)]}{\partial r_0}}{\lim\limits_{\epsilon \to 0} \frac{\partial D|R \in [r_0, r_0+\epsilon)}{\partial r_0} - \lim\limits_{\epsilon \to 0} \frac{\partial D|R \in [r_0-\epsilon, r_0)}{\partial r_0}}, \tag{9.5}$$

Sharp RKD (2)

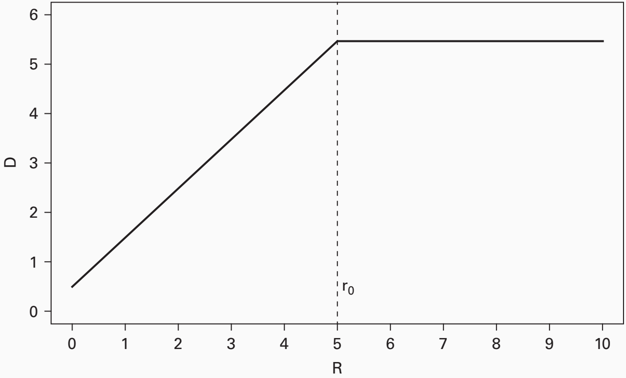

Figure 4: Sharp regression kink design.

Fuzzy RKD (1)

- In contrast to the sharp RKD, the fuzzy RKD permits random deviations of the treatment values from the kinked function.
- However, treatment D changes on average as a function of R, which can be characterized by a regression model.

Example

Example considered in Simonsen, Skipper, and Skipper (2016):

- Consumer price of prescription drugs (D) is a kinked function of the drug's actual costs R due to a reimbursement scheme.
- Actual consumer price might deviate from the kinked function due to unobserved factors, such as private health insurance arrangements.
- Fuzzy RKD identifies causal effects at the threshold among individuals with nonzero kinks, similar to compliers in IV settings.

Fuzzy RKD (2)

- Fuzzy RKD relies on continuity assumptions (like the sharp RKD) and additional monotonicity assumptions.
- Monotonicity: the kink in the association between D and R of any subject is either in the same direction or zero.
- Requires that kink is not downward sloping for some individuals and upward sloping for others.
- These assumptions are related to IV assumptions and permit identifying treatment effects for compliers at the threshold r_0 by ruling out defiers:

$$\frac{\partial E[Y(d_0)| \frac{\partial D|R\in[r_0+\epsilon,r_0)}{\partial r_0} - \frac{\partial D|R\in[r_0-\epsilon,r_0)}{\partial r_0} \neq 0, R = r_0]}{\partial d_0}$$

$$= \frac{\lim\limits_{\epsilon\to 0} \frac{\partial E[Y|R\in[r_0,r_0+\epsilon)]}{\partial r_0} - \lim\limits_{\epsilon\to 0} \frac{\partial E[Y|R\in[r_0-\epsilon,r_0)]}{\partial r_0}}{\lim\limits_{\epsilon\to 0} \frac{\partial E[D|R\in[r_0,r_0+\epsilon)]}{\partial r_0} - \lim\limits_{\epsilon\to 0} \frac{\partial E[D|R\in[r_0-\epsilon,r_0)]}{\partial r_0}}. \tag{9.6}$$

Fuzzy RKD (3)

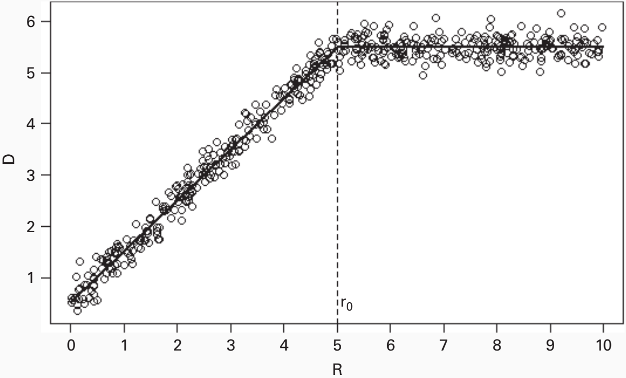

Figure 5: Fuzzy Regression kink design.

Comments on RKDs

- The fuzzy RKD (in contrast to the sharp RKD) may also be applied to a binary D, because the expectation of a treatment may be continuous even if the treatment itself is not.
- Calonico, Cattaneo, and Titiunik (2014) provide robust methods for computing confidence intervals and p-values for the sharp and fuzzy RKD similar to approaches discussed in chapter 9.1.
- Ganong and Jäger (2018) propose a permutation method based on placebo treatments in the spirit of randomization inference (as discussed in chapters 8.1 and 9.1).

Table of Contents

Bunching designs:

- ...exploit discontinuities or kinks in an assignment/running variable at a specific threshold, see e.g. Saez (2010) or Chetty, Friedman, Olsen, and Pistaferri (2011).
- ...aim to evaluate to which extent bunching above or below the threshold occurs, implying that subjects can choose the level of the running variable and thus whether they are located above or below the threshold.
- In contrast, RDDs and RKDs rule out such self-selection around the threshold by assuming continuity of the running variable, which is violated in the case of bunching.

Example

Example

Income taxation:

- Consider gross earnings as running variable, assuming that earnings are not taxed up to a certain threshold.
- Entails a kinked function of net earnings, which are equal to the gross earnings below the threshold, but less than the gross earnings above the threshold due to taxation.
- Some individuals might feel that relative to earnings just below the threshold (without tax), working more is not attractive because part of the additional earnings are taxed away.
- Those who are discouraged to work more bunch together just below the threshold.

Estimation Approach

Bunching approach:

1. Estimate the density function of the running variable, excluding observations within a specific bandwidth around the threshold.
2. Predict/extrapolate the estimated density function to the area within the bandwidth around the threshold.
3. Difference in the observed and extrapolated density just below the threshold yields an estimate for the magnitude of bunching.

- The excess density below the threshold r_0 should be matched by a lack of density at or above the threshold.
- This requirement might be used as a specification test for density function estimation.

Graphical Illustration

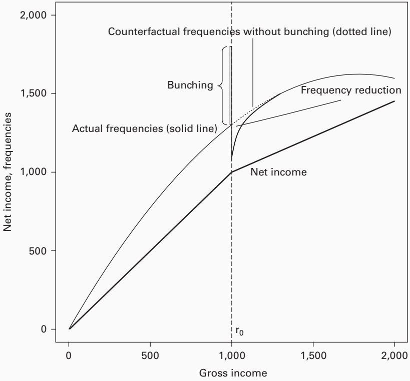

Figure 6: Bunching in the running (=outcome) variable.

Bunching in a Continuous Treatment

- Another, conceptually different design concerns bunching in a continuous treatment (rather than a running variable) as a consequence of censoring, see Caetano (2015).
- Censoring implies that the treatment value cannot be below or above a certain threshold.

Example

- Continuous measurement of education as the treatment variable, assuming 9 years of compulsory schooling by law.
- Children who would have chosen less than 9 years of education if the law were not in place, bunch exactly at the threshold.
- Hence, children bunching at the threshold might differ in unobservable characteristics compared to individuals slightly above the threshold.
- Bunching thus leads to a discontinuity in unobservable characteristics like ability and motivation U at the threshold.

Bunching in a Continuous Treatment - Illustration

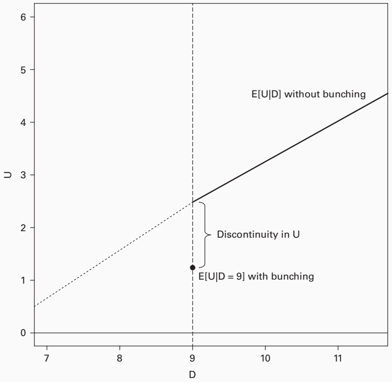

Figure 7: Bunching in the treatment variable: discontinuity in unobservables.

Treatment Selection Bias

This scenario can be exploited to estimate the treatment selection bias that arises if U jointly affects treatment D and outcome Y:

1. Estimate the regression function $E[Y|D]$ in a data window close to, but not including $D = 9$.
2. Predict the regression function at $D = 9$.
3. Subtract the average outcome among observations with $D = 9$ from the predicted conditional mean of Y at $D = 9$ to estimate the bias due to unobserved characteristics.

- One may apply this approach when controlling for covariates X to verify whether controlling for covariates tackles selection bias.
- Under specific assumptions, one may also use this approach to correct for selection bias in the estimation of treatment effects, see Caetano, Caetano, and Nielsen (2020).

Treatment Selection Bias - Illustration

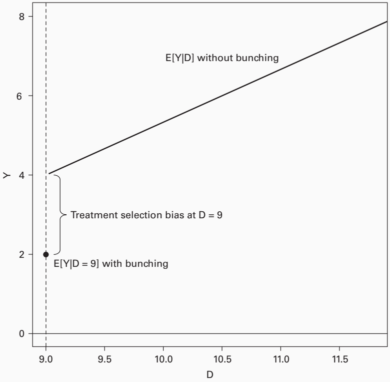

Figure 8: Bunching in the treatment variable: discontinuity in outcome.