REGRESSION DISCONTINUITY DESIGNS

- Principle
- Identification
- Estimation
- Discussion

Principle

- **Principle**
  - Non-experimental techniques introduced by Thistlethwaite and Campbell (1960) in psychology
  - Rediscovered late 1990s: Angrist and Kruger (1999), Hahn et al. (2001, Econometrica)
  - Simple idea: compare what happens to people below and above a certain threshold
  - Very widespread because: thresholds found in many institutional schemes (benefits, student grants, business support, etc.)
    - → similar individuals around threshold level
    - Method for observation data that has the closest internal validity compared to RCT

Principle

- **Formally:**
  - Method applicable when choice of treated and untreated subjects is made by applying an observable **eligibility criterion** (selection variable *S*) that cannot be manipulated by the treated (or untreated) subjects
    - No endogenous sorting intro treatment and control status
  - Assuming that subjects on either side of the eligibility for treatment boundary, but close to it are similar and can be compared.
  - There must be a **sharp break** with an arbitrary separation criterion between the treated and untreated subjects

- **Examples**
  - Alcohol consumption may be authorized at a certain age
  - Student grants may be reserved for students scoring a minimum mark in a particular examination

Principle

- **Sharp design *vs* fuzzy design**
  - 1. **Sharp design**: receiving treatment depends **deterministically** on S:

    $$T_i = T(S_i) = 1(S_i > \underline{S})$$

  - Usual framework:
    - Potential outcome with treatment Y_{i1},
    - Potential outcome without treatment: Y_{i0}.
    - Observed outcome exhibits a discontinuity in $\underline{S}$
  - S is called the score or forcing variable and $\underline{S}$ is the cut-off point.

Principle

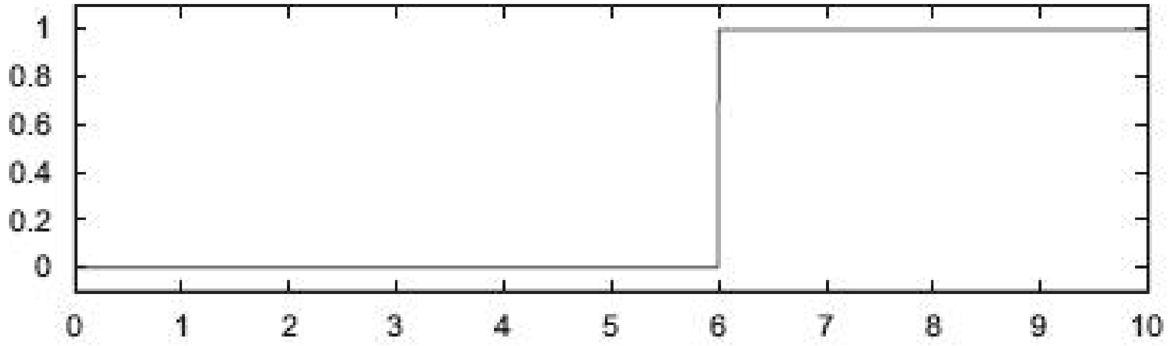


Fig. 1. Assignment probabilities (SRD).

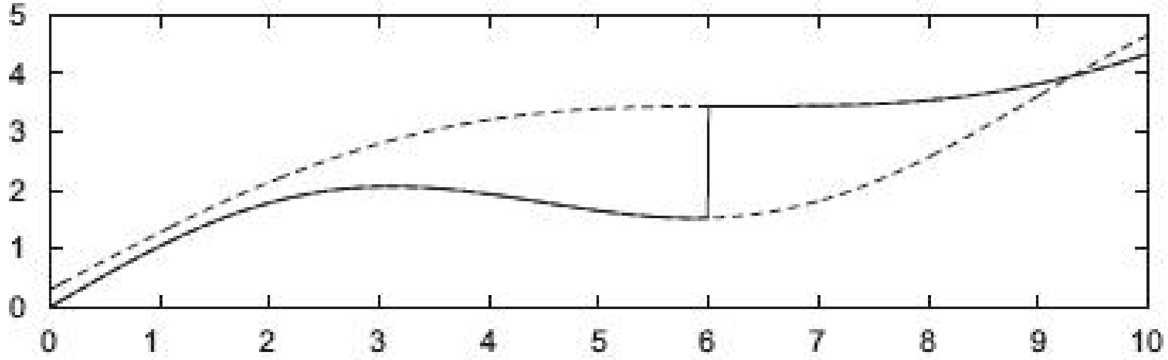
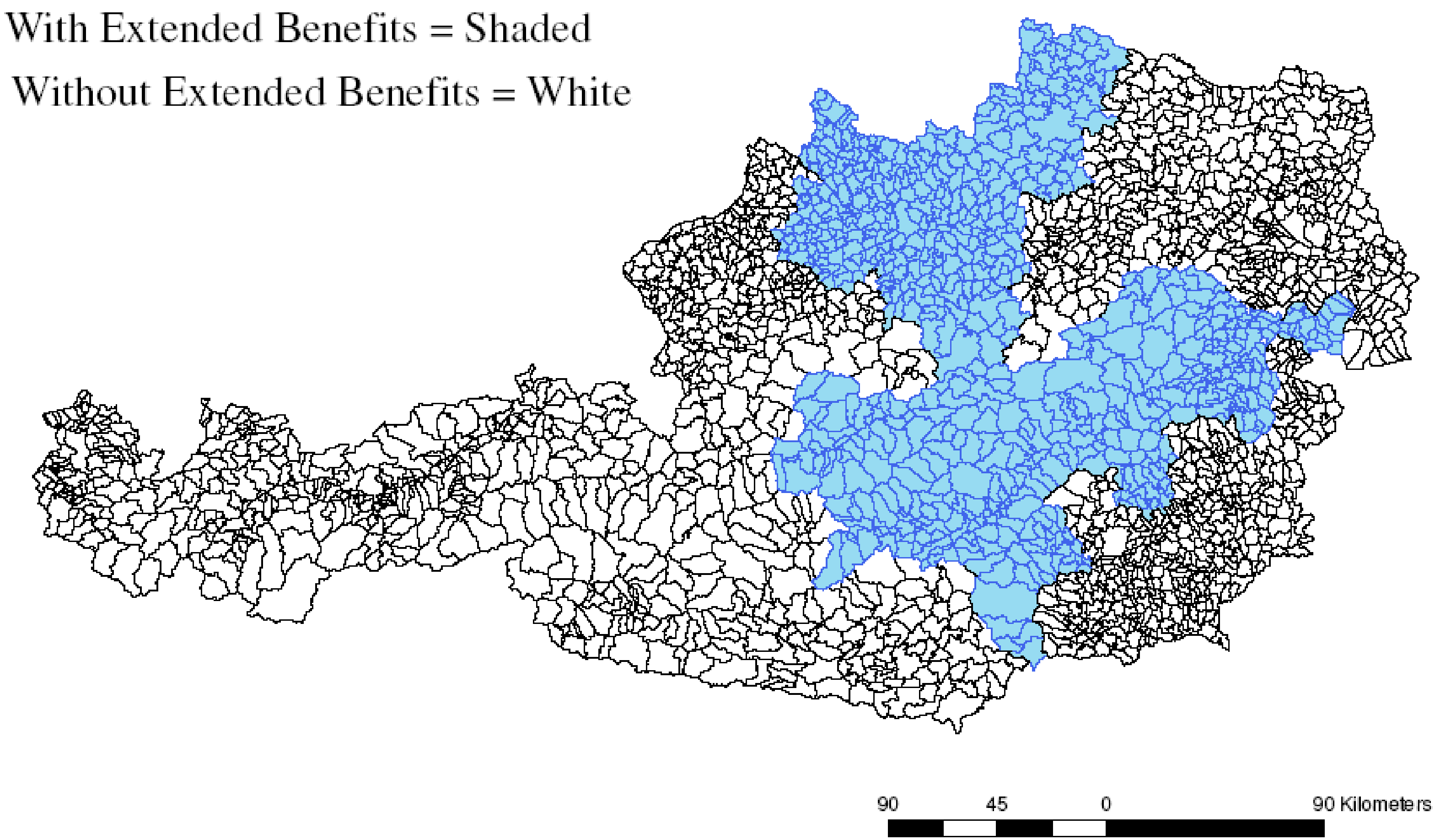


Fig. 2. Potential and observed outcome regression functions.

Principe

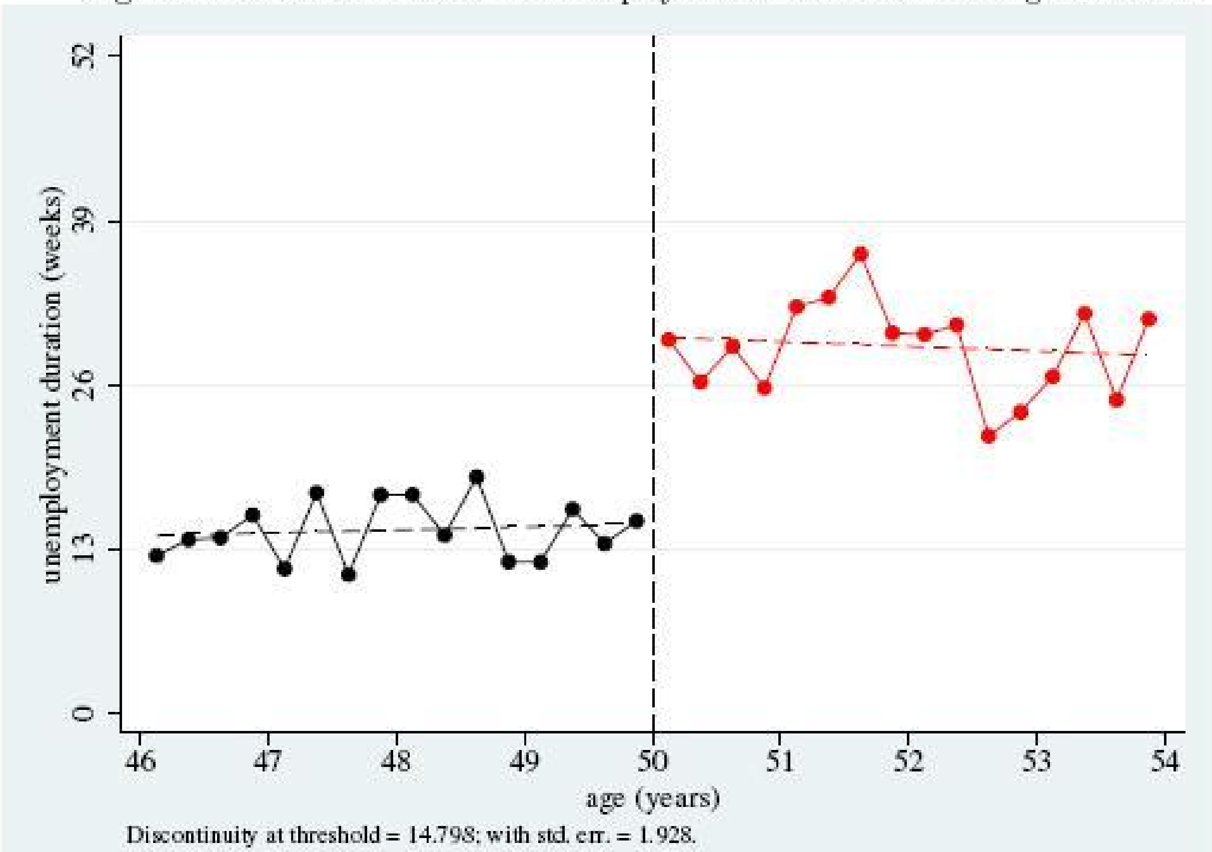
- **Example**
  - Lalive (2008) How do extended benefits affect unemployment duration? A regression discontinuity approch, JE
    - Effect of duration of unemployment benefit in Austria
    - Reform in some Austrian regions: longer duration of unemployment benefits for over 50 year olds
      - Related to collapse of steel industry
    - Therefore introduces discontinuity in entitlements
    - Impact on time to find new employment?

Figure 1: Regional distribution of REBP



Principle

Figure 2: The effect of REBP on unemployment duration for men: age threshold



Principle

- **2. Fuzzy design**
  - The rule is not applied stringently: the variable S affects only the *probability* of receiving treatment:
  - $P(T = 1|S)$ exhibits a discontinuity in $\underline{S}$
  - However: some below threshold may receive treatment and some above may not
  - So the selection variable affects the probability of receiving treatment, but the attribution is imperfect

Principle

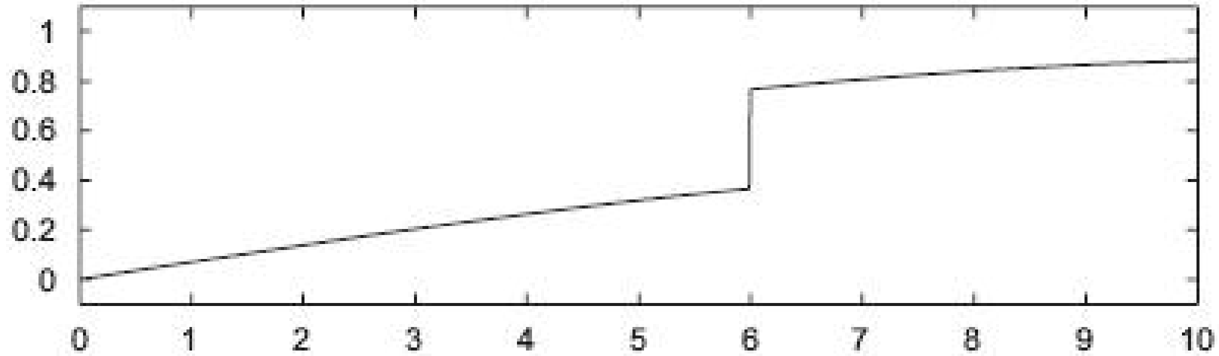


Fig. 3. Assignment probabilities (FRD).

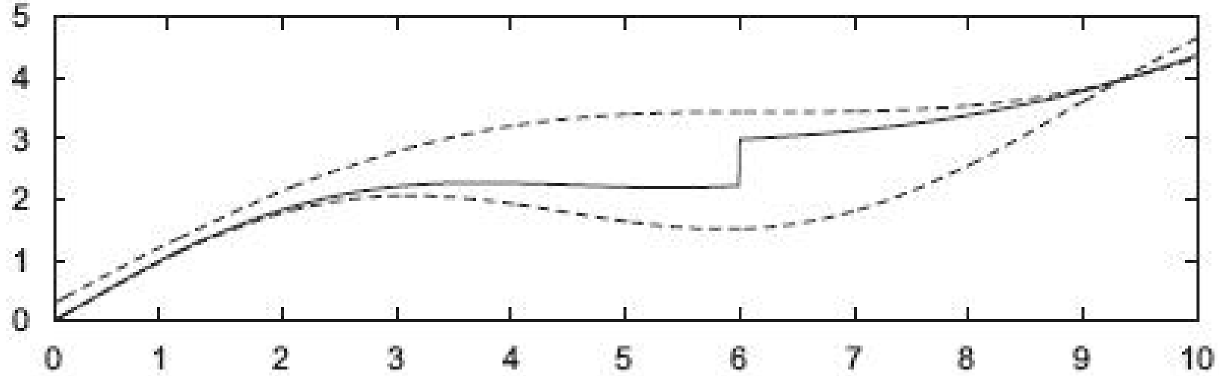


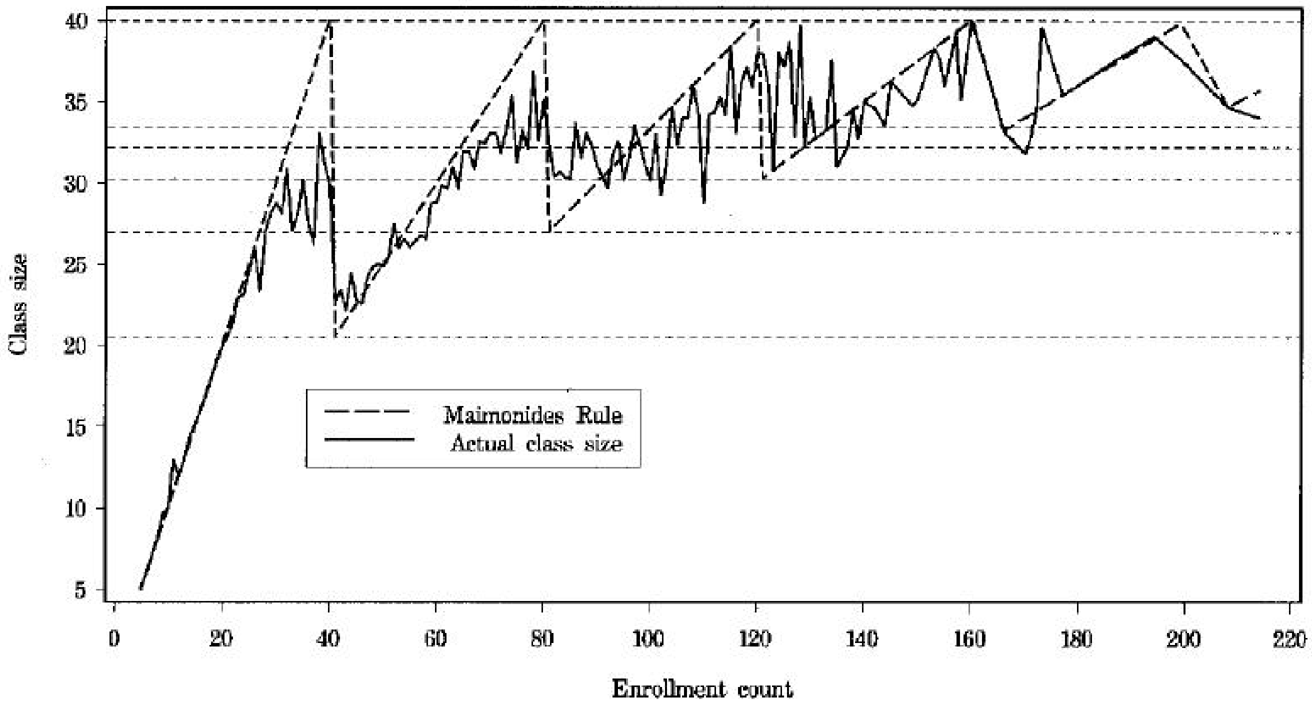
Fig. 4. Potential and observed outcome regression (FRD).

Principle

- **Angrist and Lavy (1996):** Using Maimonides' rule to estimate the effect of class size on scholarship achievement, QJE
  - Class size limit in Israel: 40 (Maimonides' rule)
  - Small demographic shock on school may lead to creation of new class and so reduced numbers:
    - 78 children registered in year group: 2 classes of 39
    - 84 children registered in year group: 3 classes of 28

Principle

- Class size by number of children in year group in school



Principle

- **Van der Klauw (2002),** Estimating the effect of financial aid offers on college enrolment: a regression discontinuity design, IER
  - US universities may offer grants to students and compete over size of grant
  - Question: What is the causal impact on the decision to go to the university in question?
  - Difficult to estimate: generally, good candidates have attractive offers elsewhere
  - Mode of attribution: grant size depends on academic level with four separate groups
  - Very fuzzy because of other determinants

Principle

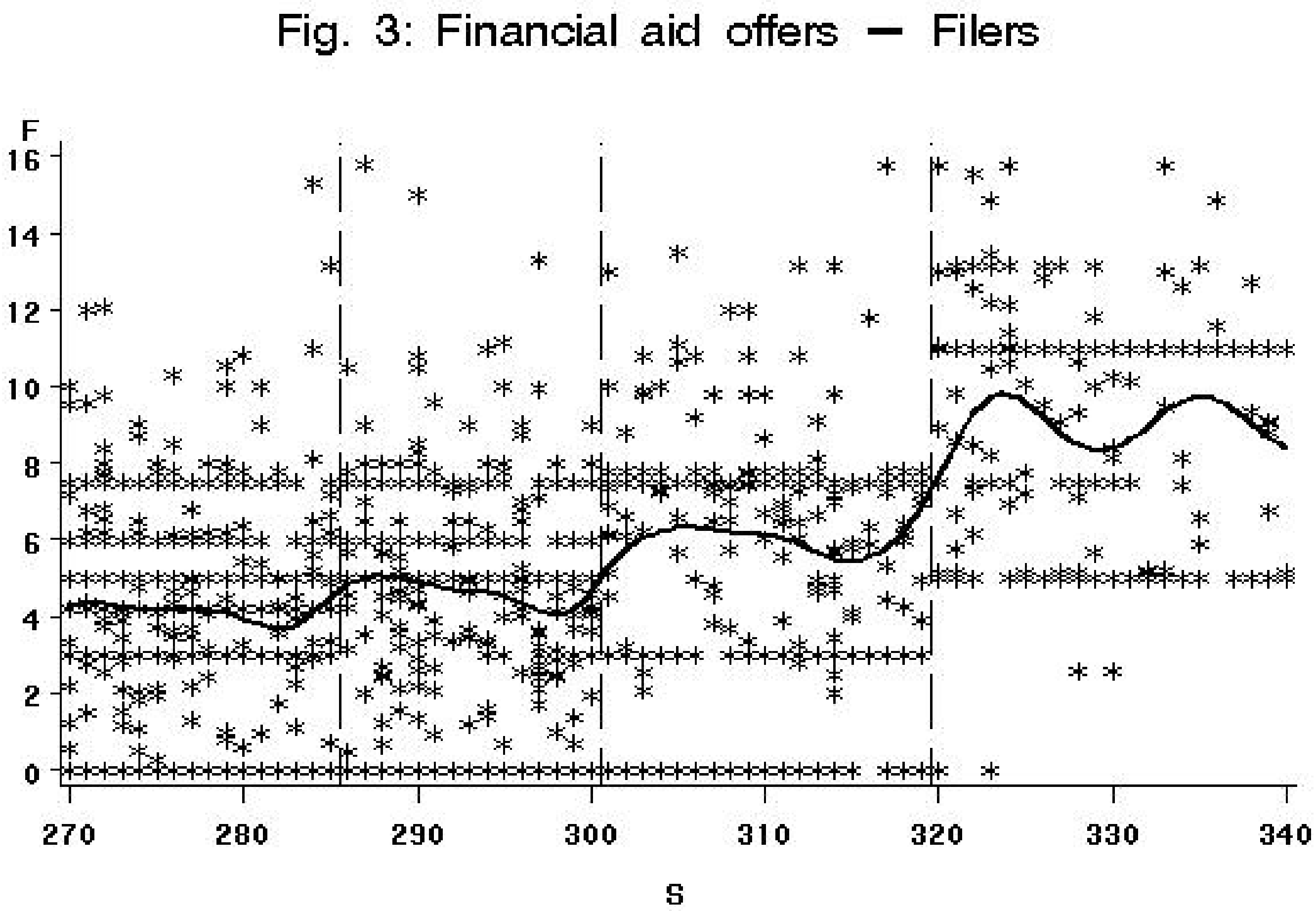
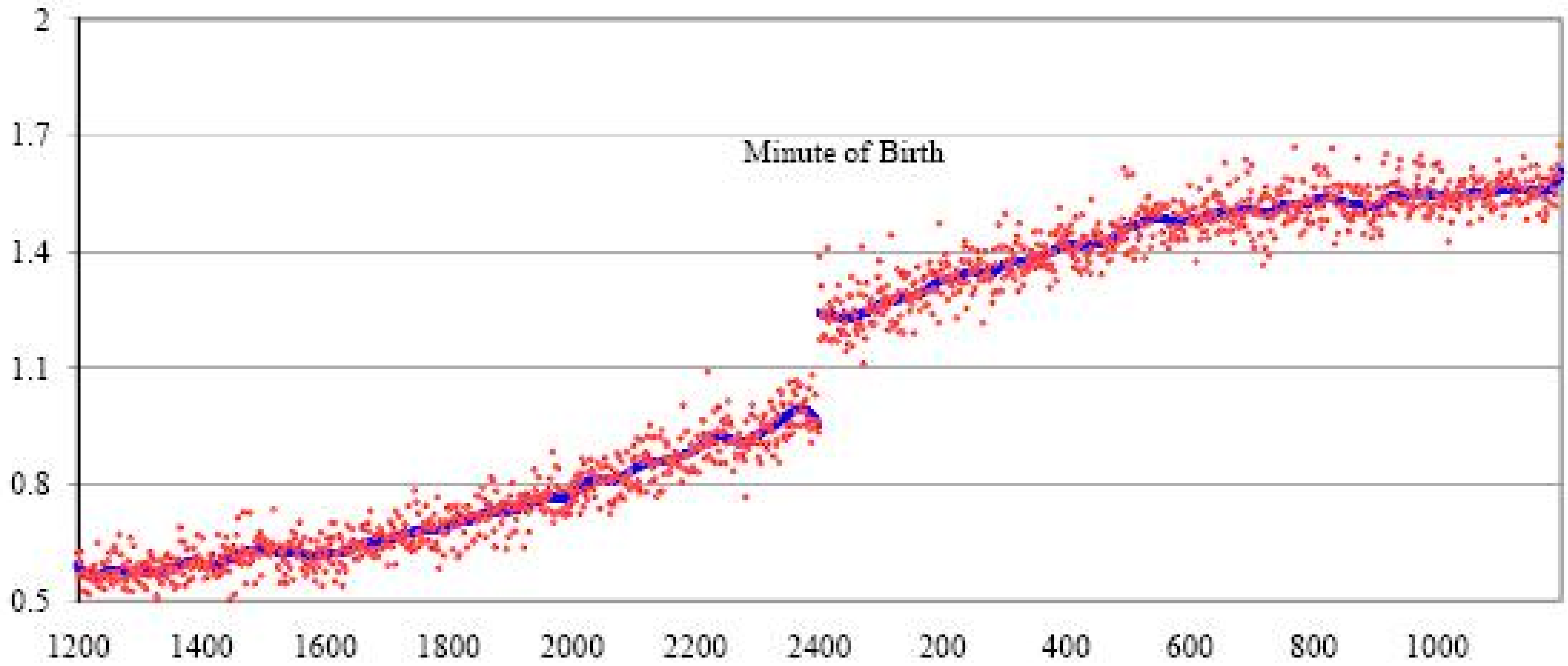


Fig. 3: Financial aid offers — Filers
F
16
14
12
10
8
6
4
2
0
270
280
290
300
310
320
330
340
S

Principle

- **Example 3: Dylie and Almond**
  - Impact of length of post-partum hospital stay
  - US: very often limited to 1 or 2 days
  - Reimbursement of insureds: number of nights counted by presence at midnight
  - Threshold effect:
    - Child born 5 minutes to midnight: one extra night counted compared with child born 5 minutes after
    - Incentive to keep mother and child for shorter time

Principle



Minute of Birth
2
1.7
1.4
1.1
0.8
0.5
1200
1400
1600
1800
2000
2200
2400
200
400
600
800
1000

Principle

- **Extensions**
  - **Multicutoff RD** : settings where the treatment assignment rule depends on more than one cutoff point
  - **Multiscore and geographic RD**: the treatment assignment rule depends on more than one score variable

QUIZ 1

- Situation: You must evaluate the impact of student grants on students' academic careers. Grants were introduced in 2010. An eligibility score S is calculated for each student based on family income. All students whose score S exceeds the threshold S^* are eligible for a grant.
- Which of the following is/are a regression discontinuity method?
  - Compare students just above the threshold S^* after 2010 with students just below the threshold before 2010.
  - Compare students just above the threshold S^* with students below the threshold, before and after 2010
  - Compare students just above the threshold S^* after 2010 with students just below the threshold after 2010.

QUIZ 2

- Situation: You must evaluate the impact of student grants on students' academic careers. Grants were introduced in 2010. An eligibility score *S* is calculated for each student based on family income. All students whose score *S* exceeds the threshold *S** are eligible for a grant.
- Not all eligible students necessarily apply for a student grant. Is it possible to use the regression discontinuity method?
    - No, because all students just above the threshold do not receive treatment.
    - Yes, by selecting only those students just above the threshold who apply for a grant.
    - Yes, by using the difference in the probability of receiving treatment just below and above the threshold.

QUIZ 3

- Situation: You must evaluate the impact of student grants on students' academic careers. Grants were introduced in 2010. An eligibility score *S* is calculated for each student based on family income. All students whose score *S* exceeds the threshold *S** are eligible for a grant.
- Not all eligible students necessarily apply for a student grant. Moreover, some students just below the threshold receive a grant even so: e.g. students with disabilities, or those with outstanding academic performance. Is it possible to use the regression discontinuity method?
  - No, because there are treated and untreated students on both sides of the threshold.
  - Yes, because we can select just the untreated students below the threshold and the treated students above it.
  - Yes, because we can use the difference in the probability of being treated just below and just above the threshold.

QUIZ 4

- Situation: You must evaluate the impact of student grants on students' academic careers. Grants were introduced in 2010. An eligibility score *S* is calculated for each student based on family income. All students whose score *S* exceeds the threshold *S** are eligible for a grant.
- In 2015, the rule changes. All applications, regardless of the score *S* are now reviewed and eligible. However, applications with high scores are more likely to be accepted. Is it possible to use the regression discontinuity method?

  - Yes, because the groups $S < S^*$ and $S > S^*$ still have different probabilities of receiving grants.
  - No, because there is no longer any discontinuity in receiving grants.

QUIZ 5

- You have to evaluate the effect of having attended the country's leading university on people's career path *Y*. The university selects students after reviewing their applications based on the average mark *M* in the national school-leaving examination.
- You see from the available data that students having scored less than 70% are almost systematically rejected. Probability of admission rises steadily and rapidly between 70 and 80%, and students scoring more than 80% are almost systematically admitted.
- Is it possible to use the regression discontinuity method?

  - Yes, by comparing students scoring less than 70% and those scoring more than 80%, because their probability of receiving treatment differs discontinuously.
  - Yes, by comparing students scoring between 70 and 75% with those scoring between 75 and 80%, because the probability of receiving treatment rises very rapidly in these intervals.
  - No, because the probability of being admitted only rises steadily between 70 and 80%.

Identification

- **1. Case of a sharp design**
  - We can write $Y_i = \alpha + \beta_i T_i + u_i$ with:

  $\alpha = E(Y_{i0})$,

  $\beta_i = Y_{i1} - Y_{i0}$

  $u_i = Y_{i0} - E(Y_0)$

  - β_i measures the effect of treatment on individual i

  - **Identifying assumption**: potential incomes $E(Y_0|S)$ and E($Y_1|S$) are continuous in $\underline{S}$
  - Thus, E($u_i|S$) and E($\beta_i|S$) will be continuous in $\underline{S}$

Identification

- **This assumption:**
  - Formalizes the idea that individuals just below and just above the threshold are comparable.
  - Does not exclude outcomes being dependent on the selection variable.
  - Is the minimum: if the unobservable determinants are also discontinuous in $\underline{S}$, it is impossible to identify a causal effect.

Identification

- **Effect of the treatment at the cut-off point**

$$E(\Delta_i | S = \underline{S}) = \lim_{S \searrow \underline{S}^+} E(Y|S) - \lim_{S \nearrow \underline{S}^-} E(Y|S)$$

- **Important comment:**
  - Only a local effect of treatment is estimated, at the cut-off point. If treatment in the population is not constant, the interpretation of this estimation will be limited.
  - → limit of external validity

Identification

- **2. Fuzzy design**
  - If the effect of treatment is constant in the vicinity of the cut-off point, the previous assumption is sufficient. The effect of the treatment at the cut-off point is:

$$\frac{\lim_{S \nearrow \underline{S}} E(Y|S) - \lim_{S \searrow \underline{S}} E(Y|S)}{\lim_{S \nearrow \underline{S}} E(T|S) - \lim_{S \searrow \underline{S}} E(T|S)}$$

  - Otherwise, if the effect of treatment varies in the vicinity of the cut-off point, it must be considered a random variable. In this case, a monotonic assumption is sufficient (Hahn et al., 2001): $T_i(S)$ is non decreasing in S, for each individual i around the cut-off point
    - The probability of receiving treatment must rise when the threshold is crossed

Identification

- In this case, the effect of treatment is estimated in the same way, but the interpretation differs slightly.
  - **LATE estimator** (Local Average Treatment Effect): solely for individuals for whom being below or above the threshold makes a difference – the ‘compliers’.

QUIZ 6

- Situation: You have to evaluate the effect of having attended the country's leading university on people's career path *Y*. The university selects students after reviewing their applications based on the average mark *M* in the national school-leaving examination.
- All applicants scoring more than M^* = 80% in the exam are automatically accepted and those scoring less than 80% are automatically rejected.
- You want to use a regression discontinuity method. What are the necessary identification assumptions?
  - *Students on either side of the threshold but close to it must react to treatment in the same way.*
  - *Apart from their exam mark, students on either side of the threshold but very close to it have identical average characteristics.*
  - *The probability of being selected must remain identical when the threshold is crossed.*
  - *Measurement of the career path must be constant on either side of the threshold.*

Estimation

- **Evaluating the validity of the identification assumption**
  - The method's success lies in its graphics
  - Several useful graphics (Imbens and Lemieux, 2008)
    - The treatment variable exhibits a discontinuity at the expected point (otherwise it is pointless continuing)
    - The income variable exhibits a discontinuity at the same point and that point alone
    - The other possible determinants of income do not exhibit a jump at point $\underline{S}$

Estimation

- **Several approaches for the treatment effect**
  - Non-parametric estimation: kernel estimators
    - No good properties on cut-off points (Porter, 2003)
  - Semi-parametric estimation: local polynomial regression
    - Preferable if many data close to threshold
    - No assumption about variation of variable of interest with eligibility criterion
  - Parametric estimation: in some cases, not enough observations around cut-off point, or discrete data
    - Stronger assumptions

Estimation

- **Implementing parametric estimation**
  - We posit a model of how *Y* behaves depending on distance to the borderline
  - For a sharp design, we can write:

$$y_i = m(S_i) + T_i\delta + e_i$$

with $e_i = y_i - E(y_i \mid S)$, $T_i = 1(S_i > \underline{S})$,

$m(S) = \alpha + E(u_i \mid S) + (E(\beta_i \mid S) - E(\beta_i \mid \underline{S})$

and $\delta = E(\beta_i \mid \underline{S})$

  - Local continuity hypothesis, continuous $m(S)$
  - Can be approximated by polynomial functions

Estimation

- **Problems with parametric estimators**
  - Properties are true in vicinity of the cut-off point $\underline{S}$
  - The function $m(S)$ is there to correct dependence, but the validity of the estimator depends on proper specification

Estimation

- **Semi-parametric method**
  - We predict the value of *Y* on each side of the threshold by modelling the way *Y* varies with distance from the threshold, using observations very close to the threshold only
  - The model gives more weight to data close to the threshold
  - This method may also be used in a fuzzy design by estimating the difference at the threshold in the probability of receiving treatment

Discussion

- **Summary: implementing the method**
  - Have detailed data and large enough samples to be able to keep to observations close to the cut-off point. Check the discontinuity assumption is relevant.
  - Estimation: In practice, studies test robustness by varying the size of the neighbourhood
  - Interpretation: **local effect**, obtained solely for individuals around the cut-off point. Ascertaining whether the results can be generalized to a wider population is discussed on a case by case basis.

Discussion

- **This method has very strong internal validity**
  - Compares very well with random experiments (Smith et al., 2007, AER)
- **Cases where regression discontinuity design is unsuitable**
  - Other programmes using the same threshold
  - Threshold manipulation: agents control the value of the assignation variable. The criterion must be imposed on individuals.

QUIZ 7

- Situation: You have to evaluate the effect of having attended the country's leading university on people's career path *Y*. The university selects students after reviewing their applications based on the average mark *M* in the national school-leaving examination.
- All applicants scoring more than M^* = 80% in the exam are automatically accepted and those scoring less than 80% are automatically rejected.
- Which of the following might undermine the validity of the regression discontinuity?
  - *Attending this university does not have the same effect for students with an 80% average as for others.*
  - *Data show applicants with less than 80% are accepted.*
  - *Data show applicants with more than 80% are rejected.*
  - *Data show a discontinuity at 80% in the probability of coming from a privileged background.*
  - *The national exam board could add points for students just below 80% to take them above, if students had positive reports from teachers.*

QUIZ 8

- Situation: You must evaluate the impact of student grants on students' academic careers. Grants were introduced in 2010. An eligibility score *S* is calculated for each student based on family income. All students whose score *S* exceeds the threshold *S** are eligible for a grant.
- Which of the following undermine the identification hypothesis of a regression discontinuity?
  - *Families may lie about their income to display a score S > S**
  - *Students below the threshold may manage to get other types of student grant.*
  - *The awards commission may decide to award grants to students below the threshold based on different criteria.*

Example

- **Giua M. (2017) Spatial discontinuity for the impact assessment of the EU regional policy: the case of Italian objective-1 regions, *Journal of Regional Science***
  - Analysis of the impact of European cohesion policies
  - Use of administrative boundaries as a spatial discontinuity for estimating the causal impact of these policies on employment in objective-1 regions in Italy
    - Objective-1 regions: regions whose GDP per capita is less than 75% of the European Union average
    - Treated regions and their counterfactuals are municipalities belonging to 5 objective-1 regions or otherwise of a given country
      - 3 objective-1 regions
      - 2 non-objective-1 regions

FIGURE 1: Italian Regions, Whole Sample and Contiguous Municipalities.

Example

- Y: variation in employment between 1991 and 2001 in corporate sector
- T = 1 for a district in an objective-1 region for the period 1988–1999 and 0 otherwise
- **Model 1:** effect of policy on variation in employment in municipalities contiguous to another treated/untreated district

$$Y_{it} = \beta_1\, Policy_{it} + \beta_2\, \mathrm{BS}_{it} + \beta_3 \mathrm{IC}_{it} + \beta_4\, \mathrm{X}_{it} + \varepsilon_{it}.$$

- **Model 2**: all districts in sample with decreasing function of distance

$$Y_{it} = \alpha + f\,(distance_i) + Policy_{it}\left[\beta_1 + f(distance_i)\right] + \beta_2 \mathrm{IC}_{it} + \beta_3\, \mathrm{X}_{it} + \varepsilon_{it},$$

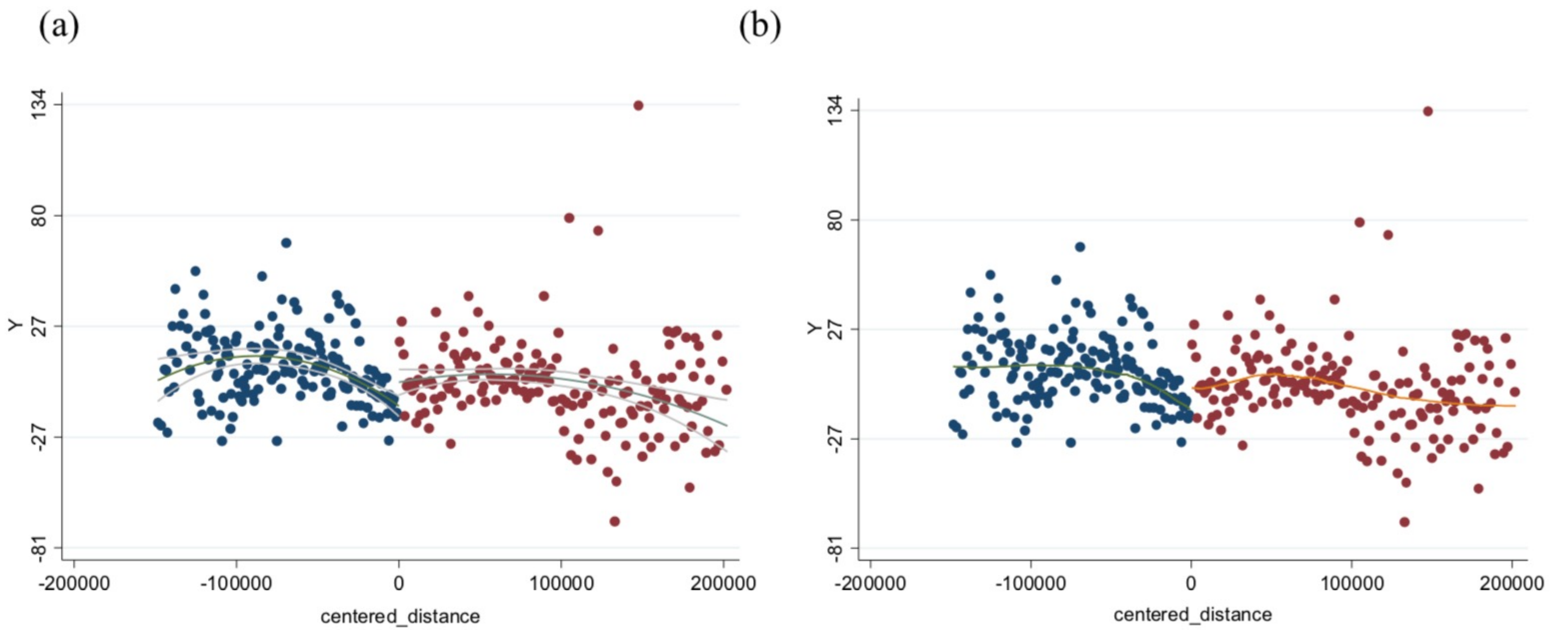
Example

TABLE 2: Differences in Mean for the Observable Variables

	Whole Sample			Contiguous Municipalities		
	Objective 1	Non-Objective 1	Diff	Objective 1	Non-Objective 1	Diff
Dependency ratio	57.13	54.37	-2.75^{***}	59.62	59.41	-0.21
Old population ratio	18.45	19.63	1.17^{**}	21.19	21.48	0.29
Uneducated population	5.33	2.70	-2.63^{***}	3.25	3.46	0.20
Highly educated pop.	2.05	2.04	-0.01	1.84	1.92	0.08
Regional transfers	2.41	3.68	1.27	1.17	1.64	0.47
Employment	1662.61	3331.91	1669.29	1361.08	1774.51	413.43
Plants	384.22	646.46	262.24^{*}	287.85	436.51	148.66
Log employment	6.08	6.29	0.21^{**}	6.00	5.89	-0.11
Log plants	5.00	5.22	0.22^{**}	4.86	4.88	0.02

Notes: *** Statistically significant at 1 percent level; ** statistically significant at 5 percent level; * statistically significant at 10 percent level. Left panel refers to the whole sample of municipalities (1,566 obs). Right panel refers to the municipalities that are contiguous to the *policy-change boundary* (99 obs). All the variables refer to the pretreatment year.

Example



Note: Quadratic and smooth local polynomial relations are respectively represented by the graph on the left and on the right. Each point represents the average employment variation by bins of 200 m. Treated side is in red color (right part of each graph).

FIGURE 3: (a) and (b) Graphical Analysis on Outcome and Forcing Variable.

TABLE 4: Effect of EU Regional Policy on Employment. Border Strategy Specification

	Left Panel—OLS Benchmark			Right Panel—OLS Border Strategy		
Objective 1 status	−1.7620	−1.4999	2.008	15.1803**	15.1896**	14.3379**
	(3.7287)	(0.7653)	(1.0245)	(6.5585)	(6.6592)	(5.8379)
Employment		−0.0017**	−0.0004**		0.0006	0.0004
		(0.0001)	(0.0001)		(0.0016)	(0.0036)
Plants		0.0121***	0.0030***		−0.0013	−0.0017
		(0.0063)	(0.0015)		(0.0094)	(0.0229)
Population density			0.0012***			−0.0051
			(0.0007)			(0.0303)
Dependency ratio			−0.5882**			−0.6117
			(0.0275)			(0.3664)
Uneducated population			−1.3310**			1.3459
			(0.0791)			(1.1854)
R^2	0.001	0.010	0.066	0.127	0.130	0.156
Obs	1,566	1,566	1,564	99	99	99

Notes: *** Statistically significant at 1 percent level; ** statistically significant at 5 percent level; * statistically significant at 10 percent level. Specifications are related to model (1) where the outcome variable is Y (relative employment variation 1991–2001) and all the other variables refer to the pretreatment year (1991). Regressions in the right panel include the set of boundary dummies and excluded the constant term. Standard errors (in brackets) are clustered at the regional level and corrected for small number of clusters via wild bootstrap (Cameron et al., 2008).

TABLE 5: Effect of EU Regional Policy on Employment. Parametric RDD Specification

	Forcing Variable: *distance*			Forcing Variable: *centroids' coordinates*		
Objective 1 status	9.0378*	9.4504***	12.0415***	10.7132***	10.5849***	9.4740***
	(4.6112)	(4.8217)	(6.1437)	(5.4660)	(5.4005)	(4.8337)
Employment		−0.0015**	−0.0004*		−0.0005*	−0.0001
		(0.0001)	(0.0001)		(0.0000)	(0.0001)
Plants		0.0103***	0.0028**		0.0040*	0.0012
		(0.0053)	(0.0015)		(0.0021)	(0.0007)
Population density			0.0012***			−0.0001
			(0.0006)			(0.0002)
Dependency ratio			−0.5487**			−0.5077***
			(0.0018)			(0.0259)
Uneducated population			−0.8672**			−0.5890
			(0.1449)			(0.2728)
Constant	−15.6554**	−16.7258**	24.0460***	−165904**	−168082**	−80285.87**
	(0.0000)	(0.0398)	(12.2684)	(20319.40)	(20479.65)	(40054.24)
Polynomial degree	3	3	2	3	3	3
R^2	0.035	0.042	0.078	0.062	0.063	0.089
Obs	1,566	1,566	1,564	1,566	1,566	1,564

Notes: *** Statistically significant at 1 percent level; ** statistically significant at 5 percent level; * statistically significant at 10 percent level. Specifications are related to model (2) where the outcome variable is Y (relative employment variation 1991–2001) and all the other variables refer to the pretreatment year (1991). Standard errors (in brackets) are clustered at the regional level and corrected for small number of clusters via wild bootstrap (Cameron et al., 2008). Parametric specifications with up to 3 polynomial order. We report coefficients of the best specification according to the Akaike Informative Criterion (AIC).