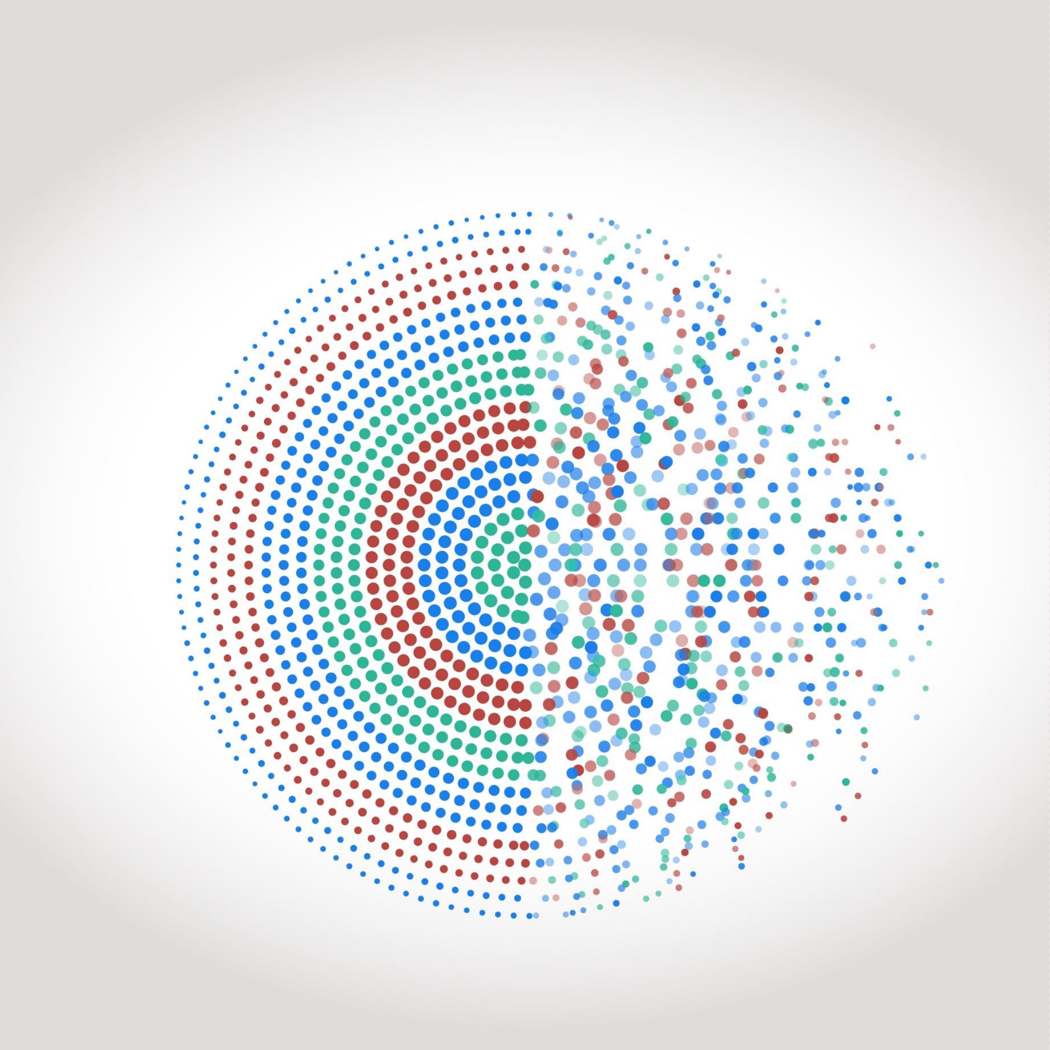
Sixth STATA lab

The Research Discontinuity Design (RDD) Method



Introduction

- The Research Discontinuity Design (RDD) method can be used for causal inference in a particular case...
- There is a policy, law, or particular intervention that is administered/enters into force based on the attainment of a threshold
- For example: all individuals with an ISEE-wealth score (which we denote by Z) below threshold $\underline{z}$ receive a basic income. All those for whom $Z>\underline{z}$ do not qualify for income.
- Formally, in this case we can denote our treatment as.

$$X = \begin{cases} 1 & if\ Z \leq \underline{z} \\ 0 & if\ Z > \underline{z} \end{cases}$$

In the RDD jargon we call the variable Z a ***running variable***.

-Z can have an influence on our dependent variable of interest (Y). The basic assumption of this method is that, in the absence of intervention X, the relationship between Y and Z would be continuous around threshold $\underline{z}$.

Intuition of the RDD estimator

- The presence of a threshold for access to the policy or a particular intervention creates the conditions for a "natural experiment."
- We expect those with ISEE-income score slightly above or slightly below <u>z</u> to be similar. If we observe a discontinuity in Y at point <u>z</u>, we can attribute this effect to intervention X
- There are plenty of concrete cases where this method can be applied. What is the running variable in each of these examples?
  - ➢ Scholarships whose access depends on income (income)
  - ➢ Minimum score in the undergraduate grade for access to certain professions (undergraduate grade)
  - ➢ Policies applied in only one of two neighboring regions (distance from regional boundary, this method is also called "Spatial RDD")
  - ➢ Application of Article 18 of the Workers' Statute only to enterprises with more than15 employees (number of employees)
  - ➢ Decontribution for newly hired young people who are under 30 years old (age)

An example: Hijzen et al. 2017

- Hijzen et al (2017) study whether there are differences in the proportion of fixed-term contracts in firms above/below the 15-employee threshold in Italy.
- Above 15 employees, Article 18 of the Workers' Statute applies to all permanent contracts, resulting in greater restrictions on the ability to dismiss. If permanent contracts are more “expensive” in large firms, are fewer of them used?
- The answer seems to be positive, can you tell why by looking at this graph?

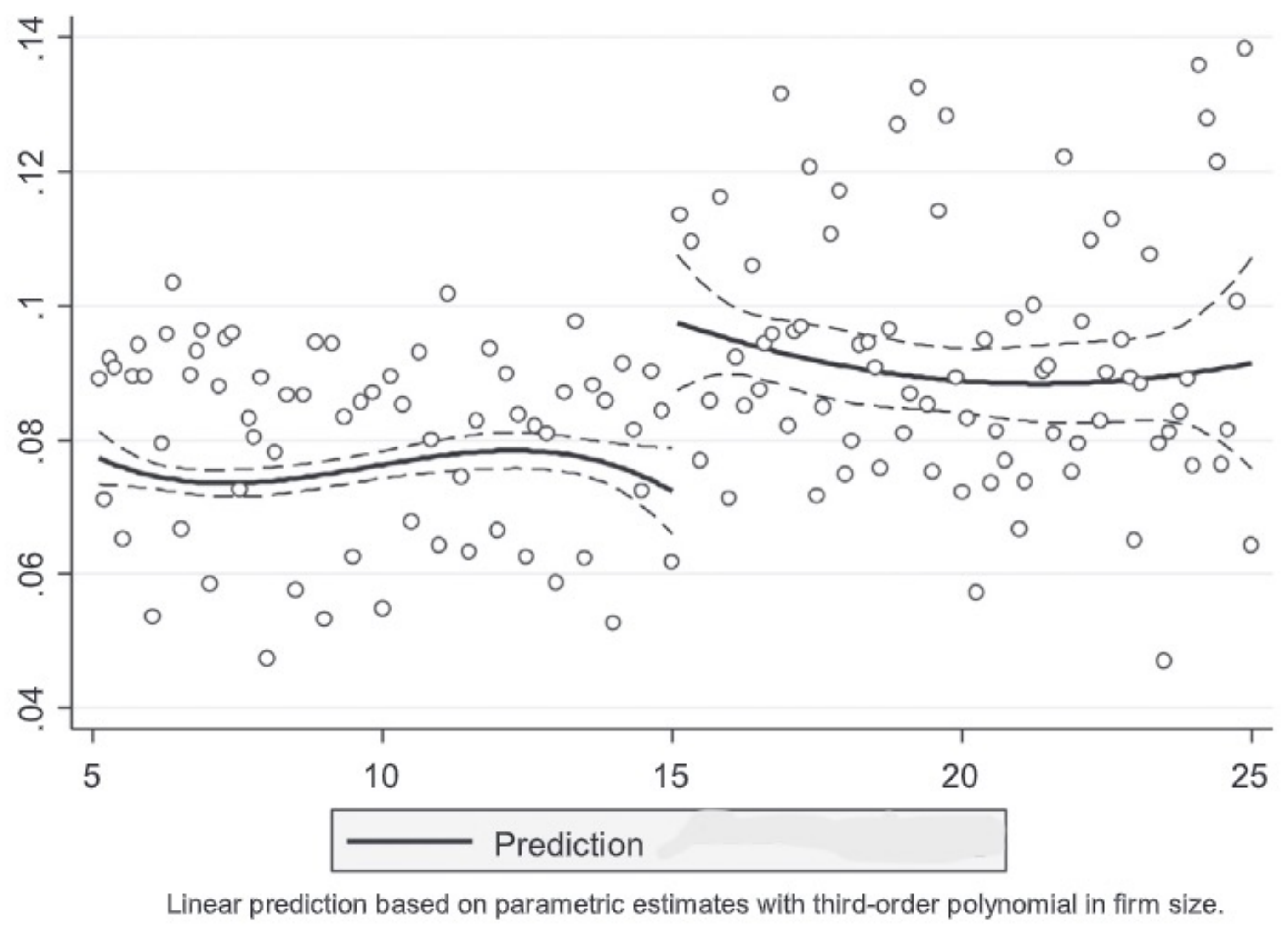


Fig. 4. The impact of employment protection on the incidence of temporary employees.

Specifications of the RDD estimator

Formally, there are two methods for estimating the effect of intervention X using the RDD approach

-Based on the parametric method, the regression of interest is

$$Y = \alpha + \beta_1 X + \gamma_1 Z + \gamma_2 Z^2 + \cdots + \gamma_k Z^k + e$$

$\gamma_1, \ldots, \gamma_k$: estimate of the (nonlinear) effect of Z on Y using a polynomial specification of degree *k*

β_1: estimate of the effect of intervention X. If statistically significant, it means that there is a discontinuity in outcome Y upon reaching eligibility threshold $\underline{z}$

-Based on the nonparametric method, the regression of interest is

$$Y = \alpha + \beta_1 X + \gamma_1 Z + \gamma_2 ZX + e$$

γ_1: estimate of the (linear) effect of Z on Y when X=0

γ_2: estimate of the (linear) effect of Z on Y when X=1

β_1: estimate of the effect of intervention X. If statistically significant, it means that there is a discontinuity in outcome Y upon reaching eligibility threshold $\underline{z}$

It is also possible to extend the nonparametric approach to the polynomial case, for example, by including $\gamma_3 Z^2$ and $\gamma_4 Z^2 X$ …

Assumptions of the RDD estimator

Formally, two basic assumptions must be met:

-The conditional function of Y with respect to Z is continuous around the eligibility threshold $\underline{z}$

-Individuals cannot manipulate Z, that is, they cannot decide to acquire more/less Z to self-select in/out of intervention X

There are some implications arising from these hypotheses that can be tested empirically:

-You can test whether the distribution of the variable Z is continuous around the $\underline{z}$-threshold (ideally it should be, otherwise you may suspect manipulation if there is an excess mass of observations and/or a deficit of observations just above/below the threshold)

-You can test whether the conditional distribution with respect to Z of the variables that are predetermined at intervention X is continuous around threshold $\underline{z}$ (ideally it should be, otherwise there are significant differences in the composition of the sample above/below the threshold)

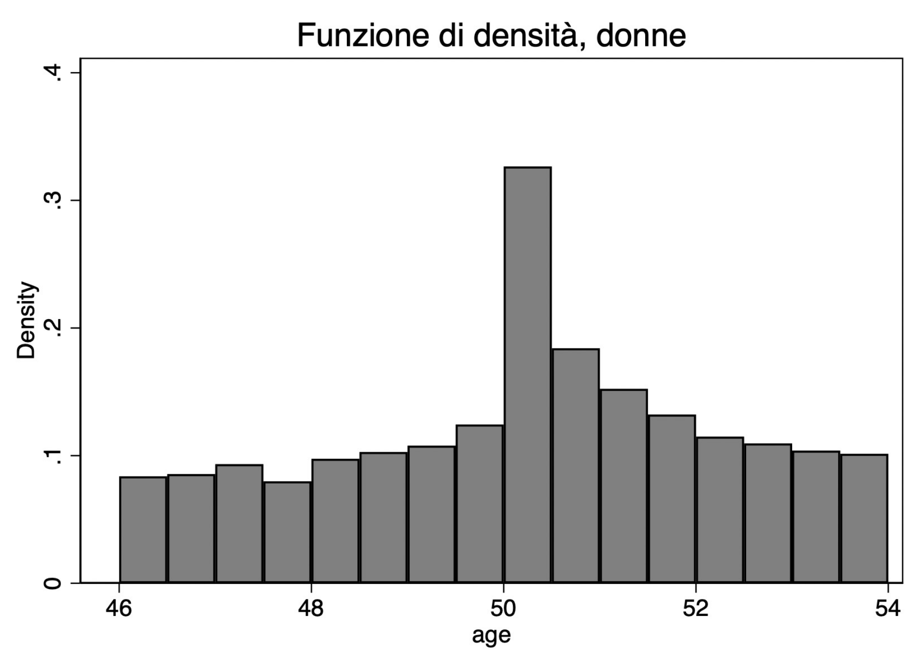
Application: Lalive (2008) on unemployment duration

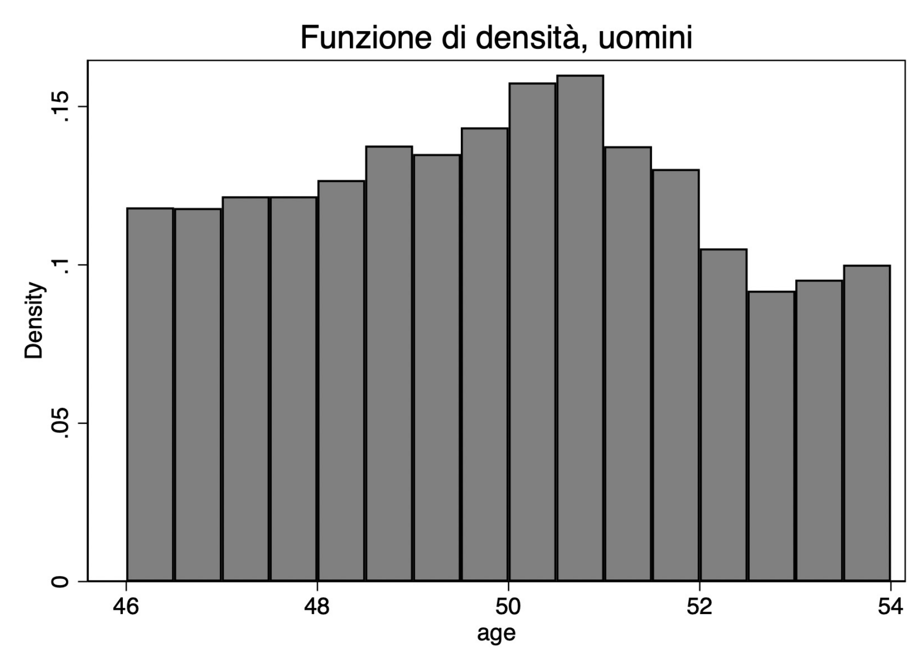
- The study by Lalive (2008) that we replicate in STATA examines a reform that extended the duration of unemployment benefits in Austria from 30 weeks to 229 weeks.
- However, this extension of duration affected only people over the age of 50 and some Austrian regions. Lalive (2008) exploits two types of RDDs (called sharp RDDs-terminology we will explain shortly):
  - Discontinuity in the duration of unemployment benefits at age 50 (age is the running variable) -> we replicate with STATA
  - Discontinuity by region of residence (distance from regional boundary is the running variable) -> we do not replicate

The main research question is whether benefiting from an unemployment benefit that potentially lasts longer (X=1) has an effect on the length of the unemployment period (Y).

Lalive (2008): test on threshold manipulation

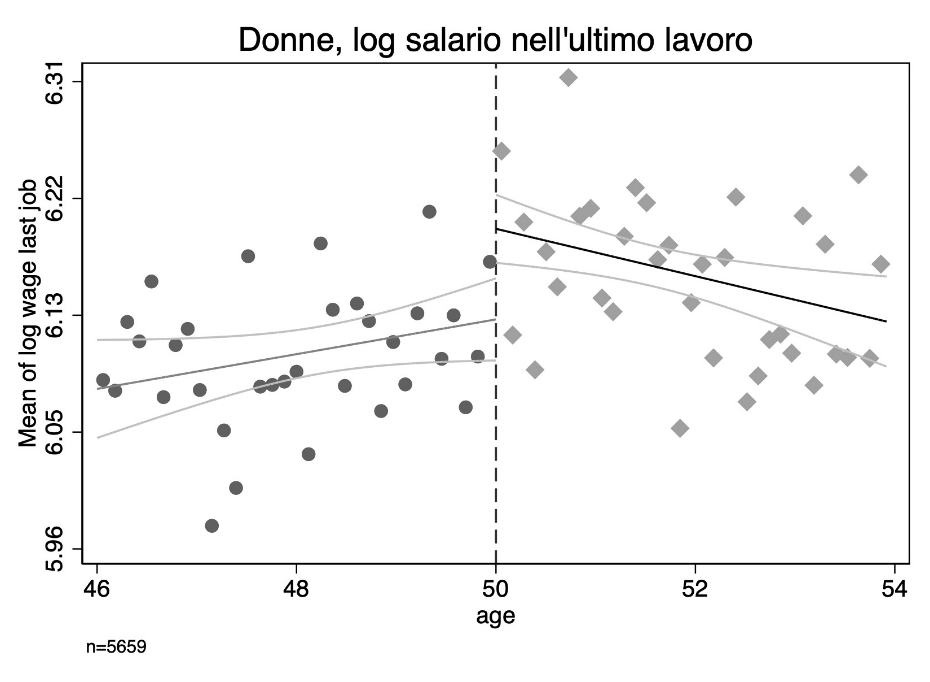
- For men, a continuous age distribution is observed around age 50.
- For women, an excess mass is observed above age 50.
- In fact, for women, the reform is equivalent to a pre-retirement: with a duration of 229 weeks of unemployment benefits, they can reach the age to qualify for the old-age pension without having to return to work.
- As a result, many more women enter unemployment upon reaching age 50: it is as if they are retiring, and they are therefore likely to be statistically different from women who enter unemployment before age 50.

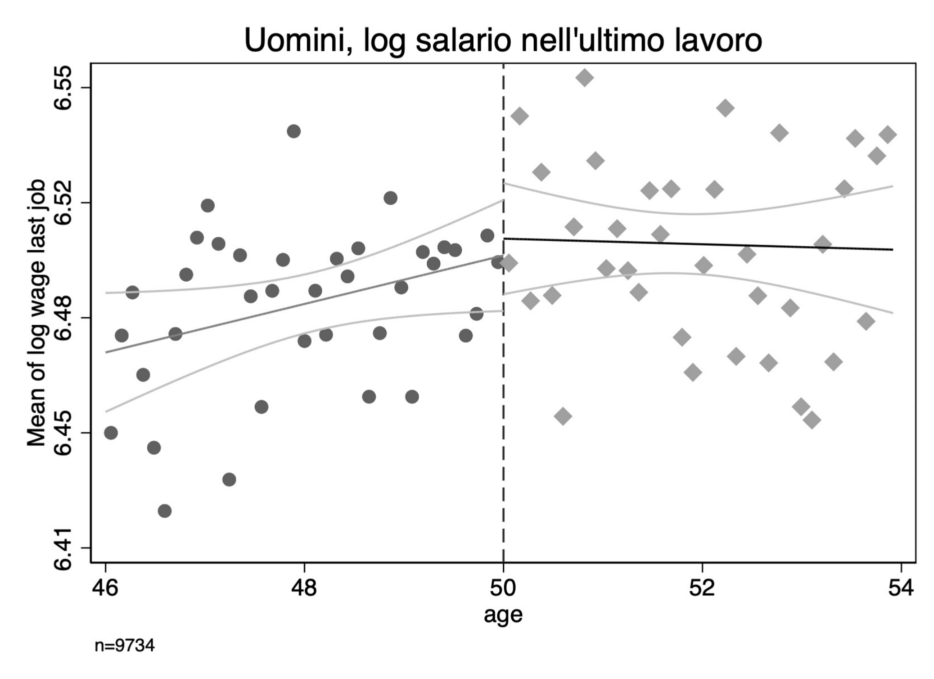




Lalive (2008): test on the continuity of predetermined variables

- For men, no significant differences are observed in the last wage earned before unemployment above/below age 50
- For women, differences in this pre-determined variable above/below 50 are visible.





Lalive (2008): Effect of maximum duration of unemplotment benefit on duration of unemployment

- Lalive's (2008) study shows (for men) a positive effect of maximum benefit duration on unemployment duration (increase of 11-14 weeks depending on estimates)
- An important issue is the choice of bandwith (how many observations to include in the regression based on their distance from the z-threshold?) There are some methods for an endogenous (i.e., data-driven) choice of bandiwth -> see rdrobust command

```
. rdrobust unemployment_duration x if female == 0, c(0) p(2)
Mass points detected in the running variable.

Sharp RD estimates using local polynomial regression.

     Cutoff c = 0 | Left of c  Right of c            Number of obs =       9734
-------------------+----------------------            BW type       =      mserd
     Number of obs |      4975        4759            Kernel        = Triangular
Eff. Number of obs |      2131        2452            VCE method    =         NN
    Order est. (p) |         2           2
    Order bias (q) |         3           3
       BW est. (h) |     1.655       1.655
      BW bias (b) |     2.127       2.127
         rho (h/b) |     0.778       0.778
        Unique obs |        48          48

Outcome: unemployment_duration. Running variable: x.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.     z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  13.625     5.8464    2.3304   0.020    2.16588     25.0835
            Robust |     -          -      2.0899   0.037    .876196     27.3157
--------------------------------------------------------------------------------
Estimates adjusted for mass points in the running variable.
```

```
. reg unemployment_duration d x i.d#c.x if female == 0, robust

Linear regression                               Number of obs     =          9
                                                F(3, 9730)        =          5
                                                Prob > F          =         0.
                                                R-squared         =         0.
                                                Root MSE          =         58

------------------------------------------------------------------------------
             |               Robust
unemployme~n | Coefficient  std. err.      t    P>|t|     [95% conf. inter
-------------+----------------------------------------------------------------
           d |   14.79848   2.234337     6.62   0.000     10.41872     19.1
           x |   .2293574   .4480373     0.51   0.609    -.6488888     1.10
             |
       d#c.x |
          1  |  -.6028858    .981324    -0.61   0.539    -2.526485     1.32
             |
       _cons |   14.97024    1.08089    13.85   0.000     12.85147     17.0
------------------------------------------------------------------------------
```

```
. reg unemployment_duration d x x2-x5 if female == 0, robust

Linear regression                               Number of obs     =      9,734
                                                F(6, 9727)        =      26.27
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0160
                                                Root MSE          =     58.009

------------------------------------------------------------------------------
             |               Robust
unemployme~n | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   11.78048   3.738889     3.15   0.002     4.451477    19.10948
           x |    2.75757   2.339296     1.18   0.239    -1.827938    7.343077
          x2 |  -.1443008   .4385676    -0.33   0.742    -1.003984    .7153828
          x3 |  -.5916064   .3782731    -1.56   0.118      -1.3331    .1498874
          x4 |   .0034525    .031997     0.11   0.914    -.0592684    .0661733
          x5 |   .0309217   .0185955     1.66   0.096    -.0055294    .0673729
       _cons |   16.48416   1.915503     8.61   0.000     12.72937    20.23894
------------------------------------------------------------------------------
```

Lalive (2008): Fuzzy RDD

- If we want to study what is the effect of unemployment duration on the level of wages earned in the first job upon exiting unemployment, we can use the discontinuity at age 50 as an instrumental variable.
- The reform, by increasing the maximum duration of the benefit, has an influence on the duration of unemployment -> it is a <u>relevant instrument</u> (first-stage)
- If having a subsidy with a longer maximum duration influences the post-unemployment wage only because it influences the duration of unemployment, the instrument will also be <u>valid</u> (exclusion restriction)
- This estimation approach is called **fuzzy RDD**, different than the sharp RDD we saw earlier. Discontinuity in this case is used to estimate a relationship between the variable influenced by reform X (our variable Y) and a third variable (W).
- Structural equation: W = a + bY + cZ
- First-stage: Y = a + bX + cZ
- Reduced form: W = a + bX + cZ
- The IV estimate of the fuzzy RDD effect is given by b_iv = b_reduced / b_first

Lalive (2008): Does a longer unemployment duration help finding a better job?

- Based on fuzzy RDD estimates, an extra week spent in job search has a positive effect of 2.5 percent on the wage you earn once you find a new job.
- Under the F-test, the instrument is relevant. Do you consider the exclusion restriction credible in this setting?

```
. ivreg2 rwage (unemployment_duration = d) x 1.d#c.x if female == 0 & abs(x)<2, robust

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity

                                                      Number of obs =      5382
                                                      F(  3,  5378) =     18.72
                                                      Prob > F      =    0.0000
Total (centered) SS     =  698.6729936                Centered R2   =  -18.3336
Total (uncentered) SS   =  218082.9663                Uncentered R2 =    0.9381
Residual SS             =  13507.86405                Root MSE      =     1.584

-----------------------------------------------------------------------------------------
                      |               Robust
                rwage | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
----------------------+------------------------------------------------------------------
unemployment_duration |   .0249508   .0062968     3.96   0.000     .0126093    .0372923
                    x |   .0093276   .0389873     0.24   0.811     -.067086    .0857413
                      |
                d#c.x |
                    1 |  -.1009598   .0870967    -1.16   0.246    -.2716662    .0697467
                      |
                _cons |   5.835988   .1192393    48.94   0.000     5.602283    6.069692
-----------------------------------------------------------------------------------------
Underidentification test (Kleibergen-Paap rk LM statistic):              16.468
                                                   Chi-sq(1) P-val =     0.0000
-------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                14.752
                         (Kleibergen-Paap rk Wald F statistic):          16.569
Stock-Yogo weak ID test critical values: 10% maximal IV size              16.38
                                         15% maximal IV size               8.96
                                         20% maximal IV size               6.66
                                         25% maximal IV size               5.53
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
-------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):          0.000
                                                 (equation exactly identified)
-------------------------------------------------------------------------------
Instrumented:         unemployment_duration
Included instruments: x 1.d#c.x
Excluded instruments: d
-------------------------------------------------------------------------------
```