

Econometrics Problem Set 4

Conte, Eustacchi, Kakunuri, Khoso, Wu

March 2022

A fundamental question in labor economics is whether a longer duration of unemployment benefits improve the quality of the job after finding employment. However, providing a causal answer to this question is difficult because unemployment duration is correlated with individual characteristics that may also affect job quality. In this exercise, we use a regression discontinuity design for causal identification. Under regular circumstances, an unemployed person receives unemployment benefits for 30 weeks. However, the benefit duration increases to 39 weeks if the person is at least 40 years old when he/she becomes unemployed. The dataset *data_ps3.dta* is a sample of workers who have been laid off and subsequently found employment. The sample contains the following variables: *age* age at layoff *nonemp* non-employment duration in weeks *jobfind* a dummy that equals one if a person had a new job after 39 weeks *lwage0* log monthly wage in previous job *lwage1* log monthly wage in new job

1) Explain intuitively how this discontinuity in benefit duration can be exploited to estimate the causal effect of benefit duration on wages after re-employment.

The running variable is the age of unemployed individuals, where the discontinuity of the sharp RDD is at age of 40, from which the unemployment benefit duration will increase from 30 weeks to 39 weeks. Given the choice of age at 40 has no subjective dependence from unemployed, we may assume that the choice of discontinuity is as good as random.

To estimate the causal effect of the change in benefit duration, the observable characteristics age, also our discontinuity running variable, is confounding both the treatment status as well as the potential outcome, while age presumably is not changing how benefit duration affects the wage in re-employment. In other words, backdoor path from this confounder is closed, the running variable can be considered as independent of how the treatment would affect the potential outcome, and the choice of the discontinuity cannot be endogenous to treatment assignment. Therefore, we limit the selection bias in the sense that the treatment is exogenous, and individuals will not self-select into the treatment, instead the longer benefit duration is an on-and-off treatment.

To confirm that there is no self-selection, data should show no bunching around the cut-off threshold, thus there is no manipulation around the discontinuity for running variable. The c It's also necessary to check the running variable is continuous around the threshold, which further implies that the age as running variable itself is not a source of heterogeneity of unemployed. Where the heterogeneous treatment effect may be endogenously affected by this individual characteristics. Moreover, with regression discontinuity, we limit the selection bias in the sense of whom are more prone to lose their jobs and whom will have higher wage after re-employment, with such random choice of age.

The criterion for longer benefit duration depends only on age which is independent of other individual characteristics, and presumably not a subject of manipulation.

Ideally, around the age threshold it's only the benefit duration that changes, thus, the difference in wage when re-employed comparing those who just above and below the age cut-off will be considered as the treatment effect of receiving the longer benefit duration. Additionally, this sharp regression discontinuity design will be informative for the individual characteristics that may also affect job quality.

First, the age cut-off itself with smoothness across the threshold will imply whether the age is or not one source of heterogeneity for individuals having different wage when re-employed; the cut-off provides the indication of effect of longer benefit duration, given that the characteristics of individual across the discontinuity are assumed to be "smooth", in the sense that individuals will have drastic change in behaviours from age 39 to age 40, but instead, the only change in their environment is the benefit duration.

Second, the cut-off at age provides a comparison of individuals who aged below 40 and above, thus also provides the difference between those who receive the treatment and those who not, by which enables us to exploit the heterogeneity in characteristics other than age, within the group as well as between the two groups.

2) Write down an estimating equation for a sharp regression discontinuity design whereby you control for the running variable with a second-order polynomial that is allowed to differ above and below the discontinuity. State and explain the identifying assumption that is necessary to interpret your coefficient of interest as causal.

We denote the cutoff age for longer benefit duration, age_c

$$D_i = \begin{cases} 1 & \text{if } X_i \geq age_c \\ 0 & \text{if } X_i \leq age_c \end{cases} \quad (1)$$

Given the constant treatment effect, we have the potential outcomes,

$$Y_i^0 = \alpha + \beta X_i \quad (2)$$

$$Y_i^1 = Y_i^0 + \delta \quad (3)$$

and,

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0)D_i \quad (4)$$

$$Y_i = \alpha + \beta X_i + \delta D_i + v_i \quad (5)$$

where the δ is the discontinuity in the conditional expectation function:

$$\delta = \lim_{X_i \rightarrow X_0} E[Y_i^1 | X_i = X_0] - \lim_{X_i \leftarrow X_0} E[Y_i^0 | X_i = X_0] \quad (6)$$

$$= \lim_{X_i \rightarrow X_0} E[Y_i | X_i = X_0] - \lim_{X_i \leftarrow X_0} E[Y_i | X_i = X_0] \quad (7)$$

Define the local treatment as follows,

$$\delta = E[Y_i^1 - Y_i^0 | X_i = age_c] \quad (8)$$

suppose a nonlinear relationship with second-order polynomial,

$$E[Y_i^0 | X_i] = f(X_i) \quad (9)$$

Thus we have,

$$Y_i = f(X_i) + \delta D_i + \eta_i \quad (10)$$

Since $f(X_i)$ is counterfactual for $X_i > age_c$, we can approximate it to:

$$Y_i = +\beta_1 X_i + \beta_2 X_i^2 + \delta D_i + \eta_i \quad (11)$$

3) Carry out two tests for the validity of the RD design: 1) plot the density of age at layoff; 2) plot the log previous wage against the age at layoff. For each test, produce a scatter plot with bin size 4 months (i.e. each point summarizes the average value on the vertical axis for workers whose age at layoff falls within that bin). For easier visual inspection, the plot should contain a vertical line at the discontinuity and lines for the second-order polynomials above and below the cutoff. What do those graphs tell us about the validity of the RD design?

4) Produce the main results graphically. We focus here on three outcomes, namely non-employment duration, the probability of finding a job within 39 weeks and log wages in the new job. Use the same binned scatters as in exercise 3). Interpret your results.

5) Focusing on the same three outcomes, report the coefficient of interest (i.e. the coefficient of a dummy for being above or below the discontinuity) of the regression outlined in exercise 2). Produce a regression table with five panels

- The reduced-form effect at the discontinuity in the full sample using the regression from exercise
- The reduced-form effect at the discontinuity with a bandwidth of ± 5 years in age at layoff.
- The reduced-form effect at the discontinuity in the full sample using a linear control for the running variable, allowing for different slopes above and below the discontinuity.
- The reduced-form effect at the discontinuity in the full sample, controlling for the running variable with a fourth-order polynomial and, allowing for different parameters above and below the discontinuity
- The reduced-form effect at the discontinuity based on the optimal bandwidth computed based on the procedure by Calonico et al. (2014). You can use their Stata/R package *rdrobust*. The package offers multiple procedures to calculate the bandwidth. Use the default procedure. Report the estimated bandwidth along with each coefficient. Interpret and discuss the differences between the panels.