



HAL
open science

The regression discontinuity design

Denis Fougère, Nicolas Jacquemet

► **To cite this version:**

Denis Fougère, Nicolas Jacquemet. The regression discontinuity design. LIEPP Methods Brief n°14, 2023. hal-04103359

HAL Id: hal-04103359

<https://sciencespo.hal.science/hal-04103359>

Submitted on 23 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

LIEPP METHODS BRIEF n°14

2023

This brief is part of a set of methods briefs published by LIEPP. As such, it is supported by the ANR and the State under the "Investissements d'avenir" programme within the framework of the IdEx Université Paris Cité (ANR-18-IDEX-0001).

The regression discontinuity design

Denis FOUGÈRE (Sciences Po, CRIS, CNRS, LIEPP, CEPR and IZA)
denis.fougere@sciencespo.fr

Nicolas JACQUEMET (Université Paris 1, CES, PSE)
nicolas.jacquemet@univ-paris1.fr



Distributed under a Creative Commons Paternité.
Attribution-NonCommercial-NoDerivatives | 4.0 International License

<https://www.sciencespo.fr/liepp/en.html>

How to cite this publication:

FOUGERE, Denis, and JACQUEMET Nicolas, **The regression discontinuity design**,
LIEPP Methods Brief n°14, 2023-05

Publication initialement rédigée en français :

FOUGERE, Denis, et JACQUEMET Nicolas, **La régression sur discontinuité**,
LIEPP Fiche méthodologique n°13, 2023-05

IN A NUTSHELL:

The regression discontinuity design is a quasi-experimental quantitative method that assesses the impact of an intervention by comparing observations that are close to an eligibility threshold fixed by the authorities in charge of the policy under study. The existence of such a threshold (for instance, becoming eligible to a policy at a certain age or below a certain income level) generates a treatment group and a control group, in a manner similar to an experimental approach.

Keywords: Quantitative methods, quasi-experimental methods, eligibility threshold, forcing variable, sharp vs fuzzy regression discontinuity design, optimal bandwidth, monotonicity, compliers

I. How is this method useful for policy evaluation?

When one wishes to perform a quantitative evaluation of the effects of a public policy, the main difficulty consists in finding a comparison group (called a *control group*) whose situation can serve as a reference (i.e., as a *counterfactual*; see the sheet devoted to the difference-in-differences method) for the beneficiaries of the intervention (the so-called *treatment group*). The randomised experiment, in which beneficiaries and non-beneficiaries are randomly selected from a given eligible population, is the reference framework for defining a valid control group: by construction, if a large enough sample is available, the distribution of relevant characteristics in the control group (gender, age, education level, etc.) is the same as in the treatment group.

Quasi-experimental methods aim to compensate for the lack of randomised controlled experiments by relying on variations that occur exogenously (usually due to a decision of local or national authorities) and produce observations that approximate an experimental situation. Matching or difference-in-differences estimation methods exploit cases in which the implementation of a public policy produces two groups whose comparison allows, under certain conditions, to measure its average effect. On the other hand, the regression discontinuity design exploits the existence of an eligibility threshold to conduct a statistical evaluation which is the equivalent of a local randomised experiment in the neighbourhood of the threshold.

II. What does this method consist of?

When the access to a public intervention or policy is conditioned by a threshold set by the authorities in charge of that policy, the intervention produces mechanically two groups, of which only one benefits from the intervention. But these groups are not directly comparable since they differ by construction because of the value of the variable defining the threshold (sometimes called the *cutoff*). This threshold can be an age condition (for instance, the statutory retirement age), a firm size condition (for instance, a tax reduction policy for firms with less than 20 employees) or a level of resources giving access to a grant scholarship or a tax credit. As these examples show, the assumption that the variable to which the threshold applies (e.g., the age, the firm size), commonly referred to as the *forcing variable*, would not influence the outcome variable of the intervention, is generally not credible. Retirement goes hand in hand with an increase in age, which in itself has many consequences on health status, consumption habits, social life, etc. Large firms operate within industries that are generally distinct from those in which SMEs operate, and their structure and activity are often very different. Income level obviously has a major impact on many household decisions. In these circumstances, the two groups thus formed do not allow for an evaluation of the effect of the intervention by directly comparing the value of the outcome variable between beneficiaries and non-beneficiaries.

On the other hand, the application of an eligibility threshold produces a sudden discontinuity in the distribution of observations **near the threshold**: for example, observations with a forcing variable whose

value is just below the threshold could benefit from the intervention while their “neighbours” with a forcing variable whose value is just above the threshold could not. The regression discontinuity design exploits this property by assuming that small variations in the forcing variable around the threshold are the result of pure randomness, similar to a coin toss, which determines the access to the intervention of otherwise identical observations. Near the threshold, the assignment of a person or a firm to the treatment group is thus similar to what happens in a randomised experiment. Under this assumption, when observations are ranked in ascending order according to the value of the forcing variable, any discrepancy in the average value of the outcome variable once the threshold is crossed can be interpreted as a measure of the effect of the intervention.

In its simplest form, the regression discontinuity approach therefore measures the effect of a policy by comparing the average value of the outcome variable in the group of people eligible to the intervention, for example those with an income or an age just below the eligibility threshold, with the average value of that variable in the comparable control group, made up of people with an income or an age just above the threshold. The underlying assumption is that among people with otherwise similar characteristics in terms of skills, education, or gender, those just below and above the threshold are potentially identical. This implementation of the method therefore requires defining the interval (called the *bandwidth*) within which observations are kept for the analysis. This bandwidth choice is based on a trade-off between the quality of the statistical analysis permitted by a larger sample size and the weakening of the hypothesis of similarity that results from a wider interval. Imbens and Kalyanaraman (2012) propose a method to choose the magnitude of the optimal bandwidth.

The regression discontinuity design is said to be *sharp* when the assignment to the group eligible to the intervention is mandatory and strictly triggered by the value of the forcing variable. If eligibility is based, for example, on an age criterion, and applied by an authority that has access to an exhaustive census of the population, then the probability of benefiting from the intervention is equal to 1 when the age condition is met; and this probability is equal to 0 otherwise, so that assignment according to the threshold is a certain event. Let us take the example of a training programme for jobseekers aged 25 or over. The principle is then to compare the average value of the outcome variable (e.g., the hiring wage at the time of return to work) for jobseekers who are just above the age threshold, e.g., aged 25 or 26, with the average hiring wage for those aged 23 or 24, who could not benefit from this programme.

The fuzzy regression discontinuity design corresponds in contrast to situations where this threshold is less binding, so that there are observations on both sides of the threshold that are, or are not, beneficiaries of the intervention. In the example of the training programme for jobseekers aged 25 and over introduced above, let us assume that, in a given locality, this training can only be provided to 100 people aged 25 or 26 due to budgetary constraints, and that this training is not compulsory, so that only 80 of these 100 eligible people (i. e., 80%) actually agree to participate in the programme. The local employment agency then offers the remaining 20 places to 100 unemployed people aged 23 or 24; among these 100 persons, only 10 (10%) agree to participate in the programme. Rather than a sudden change in the treatment status, the notion of discontinuity here refers to the ‘jump’ in the probability of benefiting from the intervention when the eligibility threshold (age 25) is crossed. The objective is then to measure the average effect of the intervention by restricting the approach to the variation in the outcome variable that results from this “jump” in the probability of benefiting from the intervention. This procedure is based on a strong assumption, called the *monotonicity* assumption: this assumption implies that among the unemployed who do not participate in the training programme because their age is below 25, there is a subgroup of individuals who would accept to participate if their age were 25 (or above). In technical terms, these individuals are called the *compliers*. By construction, the fuzzy regression discontinuity design allows to estimate the average effect of the intervention for this subgroup only. In addition to the fact that this subgroup can sometimes be very small, it excludes two important groups, namely individuals who are

always willing to participate in the programme regardless of the value of the forcing variable (the *always takers*), and those who do not wish to participate under any circumstances (the *never takers*).

III. Two examples of the use of this method in education

Variations in housing prices across neighbourhoods reflect the willingness of households to pay for the set of services and amenities (i.e., the benefits delivered by the living environment) to which a house or an apartment gives access. One such amenity is of course the quality of the local school to which children of the residents have access. Attempts to estimate the effect of school quality on housing prices are often unconvincing, as the best schools tend to be located in the best neighbourhoods. Valuations that do not take sufficient account of neighbourhood characteristics therefore tend to overestimate the value of schools located in such areas. To overcome this difficulty, Black (1999) uses a particularly original application of the sharp regression discontinuity design, based on a threshold corresponding to the contours of the Boston school map. The study estimates the value that parents place on the quality of the local public school by comparing prices of dwellings that are located on both sides of the geographic boundaries of a school district. The fact that the average scores of students in schools in different but neighbouring sectors sometimes vary greatly, while the characteristics of dwellings on either side of school divisions change relatively little, allows to identify the relationship between educational outcomes (interpreted as the school quality) and housing prices thanks to the spatial discontinuities. The estimates suggest that a one-point increase in the average school test score leads to a 1.3% to 1.6% increase in the housing prices near the geographical limit of a school district.

The study by Matsudaira (2008) is an example of the implementation of a fuzzy regression discontinuity design, also applied to educational attainment. The study uses an administrative data set from a large school district in the United States. In this district, students advance to the next grade if their grades are above predefined thresholds. Students with grades below these thresholds are required to attend a four- to six-week summer school to avoid repeating a grade. Since the observed characteristics of students near the thresholds are almost identical, the differences in subsequent academic achievement between students just below and just above the thresholds can be attributed to the causal impact of the summer school. The sample is restricted to students enrolled in the third grade of elementary school (at the age of about eight years) and the fifth grade (at the age of about 10 years). Student scores were recorded for math and reading tests in the spring of 2001 and 2002, giving rise to a sample of 338,608 students. However, the regression discontinuity design is fuzzy: the relationship between the end-of-year test scores and summer school attendance is not deterministic. Some students whose scores were below the thresholds did not attend the summer school, while some students whose scores were above the thresholds did. For instance, only 38% of the students in third and fifth grades whose math grades were below the prerequisites at the end of the 2000-2001 school year were enrolled in the 2001 summer school. Estimates from the fuzzy regression discontinuity design method suggest that the scores of 3rd grade compliers increased by 12.8% the following year, while those of 5th grade compliers attending the summer school increased by 24.1%.

IV. What are the criteria for judging the quality of the mobilisation of this method?

For the regression discontinuity technique to mimic a local randomised experiment, it is important that the forcing variable is an exogenous covariate that is beyond the control of the population involved in the intervention. If people or firms can manipulate the value of the threshold, then assignment to the treatment group becomes a choice variable. The classic example is that of a public policy that offers employment subsidies to firms with less than 20 employees. The natural reaction of some firms whose employment level is approaching the threshold is to recruit more temporary workers, in order to increase the firm's labour force without this increase appearing in the tax returns to which they are subject, so as to continue

to benefit from employment subsidies. To detect such a manipulation of the threshold, McCrary (2008) proposes a simple statistical test, based on an aggregate reasoning. Firms that actually employ more than 20 employees (e.g., 21 or 22 employees), but whose reported size is less than 20 employees (i.e., 19 or 20), will artificially increase the proportion of firms with less than 20 employees and simultaneously decrease the proportion of firms with 21 or 22 employees. The existence of manipulations in response to the eligibility threshold therefore has a direct consequence on the distribution of firm sizes, which can be checked using a histogram. In theory, this histogram should not show a discontinuity just before and just after the threshold of 20 employees. However, if this were the case, and this can be tested statistically, then the manipulative behavior of some firms could be suspected.

To avoid narrowing the bandwidth around the threshold too much, it is common to add explanatory variables other than the forcing variable, providing control over the variations in the outcome variable that are due to observed covariates. For instance, individual income tends to increase with age, so that widening the bandwidth around the age threshold leads to add observations for which the outcome variable changes with the level of the individual income. Taking this income effect into account in the statistical analysis undermines such confounding differences between groups. It is important to check that the distributions of covariates other than the forcing variable do not exhibit a discontinuity in the neighbourhood of the threshold considered. If this is the case, it means that the intervention to be evaluated has some effects not only on the outcome variable but also on some of these covariates. Incorporating these covariates into the statistical analysis generally generates a biased estimate of the average effect of the intervention on the outcome variable, since discontinuities in the distributions of these covariates are themselves explained by the implemented intervention.

V. What are the strengths and limitations of this method compared to others?

The main difficulty raised by most quasi-experimental methods is that they are based on strong assumptions, which are often questioned, such as the comparability of the control and treatment groups before the implementation of the intervention. When one wishes to apply the difference-in-differences method, this is for instance the reason why it is necessary to check that the outcome variable has previously followed the same evolution in the two groups and that their observable characteristics are similar. The difficulty is the same when one wishes to use a matching method: it requires to find observations serving as a control group which have similar observable characteristics to those of the treatment group, and which also have a non-zero probability of being eligible to the intervention being evaluated. The regression discontinuity design avoids this difficulty because it is based on a principle of quasi-random assignment for the subpopulation which is close to the exogenous threshold. As in a randomised controlled experiment, the comparability of the two groups is based on a statistical argument: if the sample size is sufficiently large, the distribution of all covariates that are relevant to significantly explain variations in the outcome variable is similar in the two groups.

This assimilation of the regression discontinuity design to a randomised experiment is all the more convincing as the interval within which it is supposed to be applied is narrow, which leads to restricting the measured effect to a very particular subpopulation, characterised by the proximity of its forcing variable to the threshold. The measure provided by this local quasi-randomised experiment is therefore specific to this sub-population. Since the effect of the intervention varies greatly across different sub-groups, the estimated average treatment effect is local and only valid in the neighbourhood of the exogenous threshold (this estimate corresponds to a local average treatment effect, or LATE). An extrapolation of the results obtained for observations far from the threshold (which would define the *external validity* of the LATE) is of little relevance. This limitation of the method is further amplified in the case of a fuzzy regression discontinuity design, where the local effect is specific to the compliers only. This lack of external validity is problematic since thresholds are often set according to the expected benefit

of the intervention for the eligible group. For example, a training programme for long-term unemployed aims to counteract the effects of human capital losses due to increased unemployment spells. Part of the rationale for setting a threshold between long- and short-term unemployment spells is that this human capital loss is minimal when spells are sufficiently short. Estimating the effect of such a programme based on a regression discontinuity design thus amounts to focus on the specific sub-population (unemployed workers experiencing relatively shorter unemployment spells) for which the program is likely to be the least effective.

The interested reader will find excellent surveys about the regression discontinuity design, for instance, in the article by Lee and Lemieux (2010), and in the textbook by Cattaneo, Idrobo and Titiunik (2019).

Some bibliographical references to go further

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education", *Quarterly Journal of Economics*, 114 (2): 577-99. <https://doi.org/10.1162/003355399556070>

Cattaneo, Matias D., and Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press. <https://doi.org/10.1017/9781108684606>

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator", *Review of Economic Studies*, 79 (3), 933-59. <https://doi.org/10.1093/restud/rdr043>

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics". *Journal of Economic Literature*, 48 (2): 281-355. <https://doi.org/10.1257/jel.48.2.281>

Matsudaira, Jordan D. 2008. "Mandatory Summer School and Student Achievement". *Journal of Econometrics*, 142 (2): 829-50. <https://doi.org/10.1016/j.jeconom.2007.05.015>

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test". *Journal of Econometrics*, 142 (2): 698-714. <https://doi.org/10.1016/j.jeconom.2007.05.005>

Resources to implement this method with Stata and R software

Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press: New Haven and London. Available in free access on the website <https://mixtape.scunning.com/index.html>

Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality*, Chapter 20. Chapman and Hall/CRC Press: Boca Raton, Florida. Available in free access on the website <https://theeffectbook.net/ch-RegressionDiscontinuity.html>