Advancing Econometric Methods for Analyzing Data from Regression Discontinuity Designs



## Advancing Econometric Methods for Analyzing Data from Regression Discontinuity Designs

### **Results in Brief**

# Machine learning improves the estimation of the causal effects in regression discontinuity designs (RDD)

An EU-funded project has extended the range of methodological tools for researchers working with regression discontinuity designs through use of modern machine learning techniques.





A valuable methodological tool to study cause and effect relationships, the regression discontinuity design (RDD) plays a significant role in a growing number of research programmes. To further refine RDD and make its analysis more accurate, the RD-ADVANCE project examined and developed new methods to be applied in studies using this evaluation design.

"RDDs allows researchers to learn about

causal relationships in certain settings where randomised controlled trials are not feasible and observational data has to be used instead," explains <u>Christoph Rothe</u> , professor of economics and RD-ADVANCE project coordinator.

"The newly developed methods will allow researchers in areas, such as economics,

education and public health, to better quantify the causal impact of different types of policy measures," adds Rothe.

### **Developing new methods for RDD**

RD-ADVANCE, coordinated by the University of Mannheim in Germany, was divided into three parts to evaluate different elements of RDD. In the first part, team members developed methods to incorporate covariates (independent variables that can influence the outcome of a given statistical trial) into the analysis of regression discontinuity designs. This was carried out with the help of AI to allow for more precise conclusions.

"Specifically, machine learning was used to extract information from a potentially large number of covariates, which in turn is then used to reduce the variance of regression discontinuity estimates of causal effects," notes Rothe. With the expanded techniques, Rothe says researchers will be able to reduce the uncertainty resulting from limited data and, therefore, give better policy advice.

In the second part of the project, the team studied the confidence intervals (CIs) commonly used in statistical analysis. These CIs are based on standard errors clustered by a certain variable, called the running variable. The aim of these common CIs is to quantify the uncertainty related to the treatment effects that are being studied.

However, the researchers found that, when dealing with a <u>discrete</u> running variable, these commonly used CIs may not perform well and may not accurately represent the true uncertainty related to the estimated treatment effects. To address this issue and provide a more reliable approach for researchers, the project team developed two new CIs that can help researchers make more accurate assessments of causal effects.

#### Accuracy ensured

The RDD approach works with two groups of individuals - a treatment group that receives the intervention and a control group that does not. Participants are grouped based on a <u>cutoff point</u> in the running variable. Comparing outcomes on either side of the cutoff point helps infer the treatment's causal effect.

However, in RDD, there is a risk that individuals might deliberately change certain aspects of themselves or their behaviour in order to influence which treatment group they are assigned to. This could undermine the credibility and validity of the RDD approach.

Therefore, to overcome this potential problem and ensure the accuracy of RDD analysis, the third part of the project was dedicated to developing methods for estimation and drawing conclusions that can account for manipulation in RDD studies. The team established a broad framework to address the issue of manipulation, using nonparametric statistical methods, while also considering additional aspects of the manipulation scenario.

For more details on the methods developed, there are three publications available on the <u>results page</u> of the project.

### Keywords

RD-ADVANCE, regression discontinuity design, causal effects, AI, economics, observational data, machine learning, policy advice

#### Discover other articles in the same domain of application



The future of food through the 'virtual tongue'





Improving the collection and management of data in citizen science





Quantum processing scales up and bursts out from theory to reality





#### Advancing theories of how pathogens spread from animals to humans



**Project Information** 

**RD-ADVANCE** 

Grant agreement ID: 772021

Project website 🛃

DOI 10.3030/772021

Project closed

**EC signature date** 23 February 2018

Start date 1 September 2018 End date 28 February 2023

## This project is featured in...

Funded under EXCELLENT SCIENCE - European Research Council (ERC)

**Total cost** € 878 970,00

**EU contribution** € 878 970,00

Coordinated by UNIVERSITAET MANNHEIM Germany





Last update: 25 August 2023

**Permalink:** <u>https://cordis.europa.eu/article/id/446040-machine-learning-improves-</u> the-estimation-of-the-causal-effects-in-regression-discontinuit

European Union, 2025