VOLUME 2, 2011

ISSN 2045-3345

REGRESSION DISCONTINOUS DESIGN AS POLICY ASSESSMENT TOOL - PROBLEMS AND POSSIBILITIES

Jan Čadil, David Hartman, Petr Hájek
Unicorn College, Prague, Czech Republic

ABSTRACT

The contemporary economic policy is using variety of tools. Some of them focus on support of chosen groups of agents. This is the case of EU policies, namely the Cohesion Policy. Therefore it is necessary to analyze the impact of such support on affected groups. One of the widely used analytical tools is Regression Discontinous Design (RDD). European Union has started to utilize the RDD in combination with Propensity Score Matching (PSM) quite recently and is trying to use this tool as a basic method for Contrafactual Impact Analysis (CIA). Although the RDD is a sophisticated and well documented method its application might be quite difficult. We face standard problems related to this method like the proper groups matching but regarding the Cohesion Policy we face some new problems as well. One of the most serious ones is the issue of different regional price levels which affects the assessment process and should be dealt with. Otherwise we could get spurious results.

JEL CLASSIFICATION & KEYWORDS

■ C40 ■ Regression Discontinous Design ■ Propensity Score Matching ■ Contrafactual Impact Analysis ■ Cohesion Policy

INTRODUCTION

Impact analysis of intervention connected with specific support programs, called also counterfactual impact evaluation (CIE), is rapidly expanding area of research on both theoretical and practical level. This development is driven by existence of many applications that can be solved by this approach. Assessing impact of intervention is realized by comparison of groups of subjects participating and not participating in program. Time comparison before and after application of program within one group, called also as a *reflexive method*, brings many problems for a limited number of subjects, while many aspects of solved problems are neglected incorrectly with respect to mentioned incompleteness. For these reasons this method can be used only for a sufficiently complete set of subjects (Ravallion 2008). On the other hand, a comparison of groups of participating and non-participating subjects brings problems with separation of impacts of interventions from other side effects. Nevertheless most of the problems originate from using this basic approach combined with suitable modification of subject's selection methods or with prior modification of subject's selection methods in such a way, that particular groups are very similar to corresponding program perspective. What follows is a brief description of used characteristics and principles of the propensity score matching method. This part follows notations and some ideas from wonderful survey (Caliendo and Kepeinig 2008) that are more deeply described there and therefore it can be recommended for detailed studies.

Results of particular subject depend on its potential participation in program and possibly many others observed and/or not observed characteristics. This relation can be represented by following equation (Khandker S.R. et al. 2010):

$$Y_i = \alpha X_i + \beta T_i + \varepsilon_i \ ,$$

where $T_i \in \{0,1\}$ is participation indicator variable for subject *i* and X_i is a set of other observable characteristics and ε_i corresponds to all other unobservable characteristics having influence on subject's results. Result of program participation is represented by parameter β. However the dependent variable Y_i is a function of many other factors. If, for simplicity, it is considered only as a function of indicator variable having all other conditions constant, it is possible to introduce individual effect of intervention as (Roy 1951; Rubin 1974):

$$\tau_i = Y_i(1) - Y_i(0).$$

The problem with such an equation is that we have only one of its components in hand and therefore usually one has to use alternative group characteristic called *average treatment effect* (ATE):

$$\tau_{ATE} = E[Y(1) - Y(0)].$$

There still exist problems how to include an influence of subjects that have not been part of corresponding program (Heckman 1997). Therefore an alternative characteristic called average treatment effect on treated (ATET) is used:

$$\tau_{ATET} = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 1].$$

Term $E[Y(0) \mid T = 1]$ is usually not a part of input observation and it should be therefore substituted by known value. Frequent arrangement is based on adding a term

$$E[Y(0) \mid T = 0]$$

to both sides of mentioned equation resulting in:

$$\tau_{ATET} = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0],$$

providing that

$$E[Y(0) \mid T = 1] - E[Y(0) \mid T = 0] = 0.$$

If this condition is not true corresponding analysis has nontrivial bias called *selection bias*.

Solution to this statistical weakness is, in ideal case, random participation choice. Where randomness in participation is hard or even impossible to achieve one has to assume the *conditional independence assumption* (Lechner 1999), called also *uncounfoundedness* (Rosenbaum and Rubin 1983), which suppose that for a given set of observables X not influenced by potential participation the results of subjects are independent of this participation, i.e.:

$$Y(1), Y(0) \perp T \mid X \quad ,$$

It is obvious that for random subject assignment this condition comes true. However for most real studies this is not the case and one has to introduce some correction methods to deal with such non-random processes.

Propensity score matching

One of the methods that can be used here is *propensity score matching* (PSM), which is based on determination of probability that subject participate in program computed from given set of observables $P(X)$called *propensity score* (Rosenbaum a Rubin 1983). Participated subjects are then matched with their non-participating counterparts based on agreement in this score using matching algorithms. Resulting comparison can be determined as ATET between obtained groups. However a necessary condition is that observables have influence on participation. Again the assumption of conditional independence is applied extended in addition by propensity score matching (Rosenbaum and Rubin 1983):

$$Y(1), Y(0) \perp T \mid P(X)$$

Second important condition for PSM implementation is a condition of *common support* which state, that participated subjects have their controls near in a sense of propensity score measure, i.e. subjects with the same have positive probabilities of participation and non-participation, i.e.

$$0 < P(T = 1 \mid X) < 1\,.$$

In fact, the condition says that there exists sufficiently large set of participants and corresponding set of non-participants with the similar size. These methods are indeed fair candidates for comparisons. Although there are some automatic methods to estimate common support, most of the applications use visual inspection of propensity score distribution coupled from both comparison groups (Lechner 2001).

For determination of propensity score one has to choose a model that is able to sum corresponding variables. For binary indicator a frequent choice is logit function. There are extensions of this model having scalable participation values based mainly on multinomial probit (Caliendo and Kepeinig 2008). Since its application brings many problems it seems to be valuable to start with binary approach connected with careful interpretation of results providing arguments for possible extension to non-binary model. An important part of this process is also a choice of variables as inputs to this model, so they take all principal effects into account, while omitting some of them can lead to considerable bias (Heckman et al. 1997). Parallel to that there exist a problem with over-parameterization of used model (Bryson et al. 2002).

While matching algorithm is one of the crucial step in PSM there exist many approaches to perform this algorithm. As the most widely known can be considered: *nearest neighborhood matching*, *caliper or radius matching*, *stratification and interval matching*, *kernel and local matching* (Caliendo and Kepeinig 2008). The choice of suitable algorithm is strongly dependent on specific dataset and it is therefore connected with careful interpretation.

The last step is statistical test of results (Caliendo and Kepeinig 2008). Except possible testing of matching quality the main part of testing is estimation of variance of ATE(T). One of the choice is *bootstrapping*, which is based on repetition of a process for several samples. This method is widely accepted for these purpose despite some criticism (Imbens 2004).

General conception of RDD

Assess the implication of direct company support is a task, which in practical applications confronted with many problems and limitations. Since the application of PSM leads to relatively well defined support and control groups it is possible to use *Regression Discontinuity Design* (RDD). This method belongs to category of so called “pretest-posttest” methods. It is based on relatively simple regression analysis with easily interpretable results, however there are some drawbacks, presented below. The advantage of RDD is that it meets conditional independence assumption.

The core tool in implementation of RDD is so called *cut-off criterium* separating sample into supported and not supported subjects. Above this threshold there are companies that are supported, i.e. selected by PSM, and below this threshold there are those that has not been supported according to PSM. Adopting PSM assure that both groups are approximately same in size – in ideal case in each group every company has the same characteristics in PSM sense. In cut-off point there is a change of probability of obtaining the support and therefore it is also a point of discontinuity, i.e. generally (Barristin and Rettore 2008):

$$Pr\left\{I = 1 | s^{\pm}\right\} \neq Pr\left\{I = 1 | s^{--}\right\}$$

where *s+-* and *s--* are limit values above and under cut off value, further denotes as . If this acute boundary is applied resulting method is called *sharp RDD*. Within this case it holds that values of changes from 0 to 1 as it passes through. Let us denote a concrete value of company's evaluation *S*, and then it holds (Becker, 2009):

$$I = 1(S_i \geq \hat{s})$$

The effect of support can be consequently expressed as:

$$Y = Y_0 + \beta I(s)$$

where *β* is again effectiveness of granting a support respectively elasticity of output parameter Y with respect to support. For average difference in output parameters for a neighborhood of , the average effectiveness of support for supported companies is (modification of ATET):

$$E\left\{\beta | s^{\pm}\right\} = \left[E\left\{Y | s^{\pm}\right\} - E\left\{Y_0 | s^{\pm}\right\}\right] - \left[E\left\{Y | s^{--}\right\} + E\left\{Y_0 | s^{--}\right\}\right]$$

Providing a continuity of output variable, where limit values are equal in neighborhood of, one can rewrite above equation as (Becker, 2009):

$$E\left\{\beta | s^{\pm}\right\} = E\left\{Y | s^{\pm}\right\} - E\left\{Y | s^{--}\right\}$$

The assumption of continuity leads to possibility of regression analysis application which is able to directly quantify the impact of support. Ideally, we apply regression only for a point or close neighborhood (see *fuzzy RDD* below), which significantly increase quality of corresponding estimation. In practice there is usually a lack of data for this area and it is therefore necessary to extend this basic approach. Generally, the parameters of following function are estimated:

$$Y = c + f(X_i) + \beta Z_i$$

where *X* is a value of observed parameter before support application and *Y* is a value after this event. *Z* is an auxiliary binary variable indicating possible support. Again, parameter *β* represents dependency of observed parameter on support.

Problems with practical implementation – method, data and approach

Within practical application of RDD there are several problems. We typically face problems with research method itself, problems related to standardization of variables and problems with data relevancy. The first group of problems is related to specification of samples of supported and non-supported subjects itself. This trouble is usually solved by introducing PSM method. However the PSM method itself does not (and cannot) offer any "final" solution as it is naturally affected by researcher himself. A set of variables leading to matching the participant and non-participant groups seems absolutely vital but can substantially vary according to researcher – here we get to the second group. Therefore we may get quite different results dependent on the institution which is undertaking the research. Another very common problem is that variables are chosen on the data availability basis. Some indicators are not optimal but the optimal ones are not available. Again this might lead to serious problems when using PPS and RDD for policy and support assessments. Researchers have a tendency to use data available than to gather new data (as it is costly and sometimes even inefficient). Of course a standardization of variables should at least partially help to solve this.

Contemporary there is a strong tendency (mainly at the EU level) to develop standard assessing method based on RDD and PSM with common indicators involved coping with the indicator problem mentioned above. It is used mainly for the Cohesion Policy assessment. For example it is needed to analyze the impact of particular program on affected companies in a region. However here we get another problem right away. That is data relevancy. As we know it is very difficult to compare economic variables at regional (NUT 2 or NUTS 3) level because the PPS (Purchasing Power Standard) conversion works only at the national level. This is a serious problem as we cannot compare policy successfulness between the different nation´s regions (to say that the support in Greece was less efficient in comparison to France for example) and even among the regions within one country.

A classical problem is then hidden in the real relation of output variables before and after application of program, i.e. relation of *X* and *Y*. Although a usually assumption is a linear relation, the real situation can be significantly different. The function should be therefore correctly specified and its parameter robustly estimated (possibly with convenient transformation of input data). To achieve this one can adopt standard representation of regression model (F-statistic, t-test, residual analysis) or simply a visual inspection of dataset. As already mentioned one of the problems can be also a validity of model, or more specifically internal validity of model. Such validity can be affected for example by subjective approach of particular evaluator when deciding about subjects' application in a program. In this case it is advantageous to extend cut off to a wider interval that directly minimize subjectivity. Solving such problems leads to *fuzzy RDD* (Trochim, 1984). Generally, if there are enough observations in this interval, it is possible to apply regression analysis directly on data. On the other hand, if there is lack of observations it is apply several supporting methods like generating some auxiliary data by randomization of parameter and using this data together with real ones.

CONCLUSION

Regression Discontinuous Design (RDD) is contemporary analytical tool used widely for assessing the impact of support on supported agents. As it is necessary to get mostly similar groups of supported and not-supported units it is convenient to use PSM (Propensity Score Matching) for identification of the "matching" groups. Although this method is being used by researchers and institutions quite often and is even recommended by the European Union there are several caveats which should be further dealt with (if it is possible). First it is a selection of instrumental variables for PSM and the matching groups specification. Here we possibly should make a sort of standardization of variables to avoid the specific researcher´s or institution´s attitude. Nevertheless even if we use some widely accepted common variables we quite often face the data reliability problem. This problem is typical for regional indicators as almost all economic variables are affected by regional price level. A PPS (Purchasing Power Standard) does not help in this matter as it reflects an average national price level rather than region-specific price level. This imperfection is important especially for the case of Cohesion Policy assessment as this policy is regional-oriented and we are unable to make proper contrafactual impact analysis. Probably selecting the regional location as one of the PSM variables could help but it has not been often used yet.

REFERENCES

Becker S.O. 2009. *Methods to Estimate Causal Effects, Theory and Application*. University of Stirling.

Battistin,E., Rettore,E. 2008. Ineligible and Eligible Non-Participants as a Double Comparison Group in Regression Discountinous Design. *Journal of Econometrics* 2008. 715-730.

Bryson, A., Dorsett R., Purdon S. 2002. The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies. *Working Paper* 4, Department for Work and Pensions, London.

Caliendo, M., and Kopeinig, S. 2008. Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys* 22 (1): 31–72.

Heckman, J., Ichimura H., Todd, P. 1997. Matching as an economic evaluation estimator: evidence from evaluating a jov training programme. *Review of Economic Studies* 64(4): 605-654.

Imbens G.W. 2004 Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics*, 86(1): 4-29

Khandker S.R., Koolwal G.B., Samad H.A. 2010 Handbook on Impact Evaluation: Quantitative Methods and Practices. *The International Bank for Reconstruction and Development/ The World Bank*, Washington DC

Lechner M. 2001 A note on common support problem in applied evaluation studies. *Discussion Paper No.* 2001-01. University of St Gallen, SIAW.

Ravallion, M. 2008. Evaluating Anti-Poverty Programs. In *Handbook of Development Economics*,vol. 4, ed. T. Paul Schultz and John Strauss, 3787–846. Amsterdam: North-Holland.

Rosenbaum, P. R., and Rubin D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70 (1): 41–55.

Trochim, W.M.K. 1984. *Research Design for Program Evaluation – the Regression Discontinuous Approach*. Beverly Hills, Sage.