Causal mediation analysis: a new methodology for the identification of causal mechanisms through an RD approach

Ph.D candidate:

Viviana Celli

Supervisor: Prof. Guido Pellegrini

Department of Methods and Models for Economics, Territory and Finance

Sapienza University of Rome

Contents

Introduction: scientific context and motivation

This Ph.D thesis is comprised of three self-contained, but related essays, corresponding to the three different chapters, on the causal mediation analysis. Causal mediation analysis is a statistical framework used to study causal mechanisms. In such a context, a mechanism is defined as a process in which a causal variable of interest, known in literature as treatment, affects an outcome through one or more intermediate variables, called mediators, that lie in the causal pathway between treatment and outcome. This methodology has been developed above all in sociology, psychology and epidemiology. Surprisingly, few studies used this approach in economics, despite the great importance in knowing causal mechanisms of policy interventions and of economic phenomena. In fact, the main limit of the traditional policy evaluation approaches is that the causal effects can be estimated but without knowing nothing about the causes of the effects. In other words, we can estimate how large is an impact and if it is positive or negative, but we cannot know what is due that impact, leaving the causal effect as a "black box".

This thesis tries to make new developments in this direction. First of all, we try to use this approach in the economic field: causal mediation analysis is an important tool with a great potential, that permits to go deeper with the analysis and know more about economic phenomena. It is no more sufficient to know if a policy intervention worked or not, but it's becoming more and more important to know "why", in order to design more efficient policies. Secondly, we propose a new estimator trying to go beyond the limits imposed by structural models: until a few years ago, researchers

used predominantly structural equation models to study causal mechanisms. Only recently, some researchers moved towards new approaches, like counterfactual methods and quasi-experimental designs. Following these studies, we propose a new estimator that takes advantage of Regression Discontinuity Design (RDD) to solve the limits of the traditional mediation framework. Thirdly, we propose an interesting application to validate this model. In particular, we estimate the EU 2007-2013 Regional Policy on the 2006-2015 GDP per capita growth rate at NUTS 3 level, investigating if part of this effect is driven by Research and Development (R&D). The results suggest that the EU Regional Policy has a positive and significant impact on the per capita GDP growth rate, estimating a total treatment effect of 9.4%. A little part of this effect, 1.5%, is driven by R&D investments, confirming to be a mechanism of transmission of EU Regional Policy, even if not statistically significant.

To conclude, the idea of this thesis is to give a contribution to causal analysis in order to have better interpretations of the results and, then, to better know the phenomena that surround us. After a critical survey of the literature reported in the Chapter 1, the aforementioned issues are directly faced in chapter 2 and 3.

Chapter 1: Causal Mediation Analysis in Economics: objectives, assumptions, models

The aim of mediation analysis is to identify and evaluate the mechanisms through which a treatment affects an outcome. Its goal is to disentangle the total treatment effect into two components: the indirect effect that operates through one or more intermediate variables, called mediators, and the direct effect that captures all other possible explanation for why a treatment works. This paper reviews the methodological advancements in causal mediation literature in economics, in particular focusing on quasi-experimental designs. It defines the parameters of interest under the counterfactual approach, the assumptions and the identification strategies, presenting the Instrumental Variables (IV), Difference-in-Differences (DID) and the Synthetic Control (SC) methods.

.

Chapter 2: Identification of causal mechanisms through an RD approach
Causal Mediation Analysis has important implications, because it helps to better understand the policy makers' decisions. However, the identification process is not easy and analyzing causal mechanisms requires stronger assumptions than evaluating the classical average treatment effect. The main problem consists in the endogeneity of the mediator with the consequence that the simple randomization of the treatment does not imply the randomness of the mediator. Several methods have been developed, based on different set of assumptions and with different strategies for the estimation. In this chapter we propose a new identification strategy for the estimation of the direct and the indirect effect, through an implementation of a Regression Discontinuity Design. In this chapter, we present two different models. The first one follows the traditional identification strategy based on linear equation models. The second model follows the most recent literature based on non-parametric identification procedures. We show the consistency of this last estimator, validating the results through a Monte Carlo simulation study.

.

Chapter 3: Causal Mediation Analysis in Economics: an empirical application
Given the increasing share of the EU budget devoted to Regional Policy, several studies have tried to identify the impact of structural funds on the economic growth. In this chapter, we estimate the impact of EU 2007-2013 Regional Policy on the 2006-2015 economic growth rate, measured by per capita GDP. For this purpose, we implement the estimation strategy developed in chapter 2. In addition to that, we exploit the geographical distribution of the funds, using a Spatial Regression Discontinuity Design. Thanks to this new estimator, we are able also to estimate the indirect effect of the Policy. In particular, we focus on R&D as a channel of transmission: the estimator tells us how much of the total effect is due to R&D investments. Firstly, the results show the positive average impact, confirming the importance of

Regional Policy as a tool to counteract the crisis. Secondly, the results suggest also that, among treated regions, i.e. regions defined Objective "Convergence", the ones investing an high intensity of funds in R&D grow more the ones not investing in this priority theme. These findings confirm R&D to be an important driver for the economic recovery.

Chapter 1

Causal Mediation Analysis in Economics: objectives, assumptions, models

1.1 Introduction

Causal analysis has prooved to be a powerful approach to measure the causal effects of a variable of interest on the outcome. Causal analysis answers questions like: “Are public subsidies effective?” or “Are these effects positive or negative?”. Nevertheless, this kind of analysis cannot answer to another important question: “Why are these treatments effective?”. As pointed out by Gelman and Imbens (2013) not only the "effect of a cause", i.e. the treatment effect, seems relevant in many problems, but also "the cause of the effect", i.e. the mechanisms through which the total effect materializes. To use the words of Imai, Tingley and Yamamoto (2015): “A standard analysis of data [...] can only reveal that a program had such impacts on those who participated into it. It means that we can quantify the magnitude of these impacts, we can know how much a treatment affects an outcome, but these estimates tell us nothing about how. We know something about the causal effects, but nothing about causal mechanisms”.

To overcome these limits a solution can be found in the causal mediation analysis, i.e. a formal statistical framework that can be used to study causal mechanisms. Following the definition given by Imai, Tingley, Yamamoto (2013) a mechanism is a process where a causal variable of interest, that is a treatment, influences an outcome through an intermediate variable, the mediator, that lies in the causal pathway between the treatment and the outcome variables. Studying causal mechanisms permits to know something more about social and economic policy implications than the total effect alone. This allows policy makers to optimize decisions, making them more efficient. The main fields in which mediation has been developed are psychology and sociology. For instance, Brader, Valentino and Suhay (2008) go beyond estimating the framing effects of ethnicity-based media cues on immigration preferences and ask: "Why the race of ethnicity of immigrants, above and beyond arguments about the consequences of immigration, drives opinion and behavior?". That is, instead of simply asking whether media cues influence opinion, they explore the mechanisms through which this effect operates. Consistent with earlier work suggesting the emotional power of group-based politics (Kinder and Sanders, 1996), the authors find that the influence of group-based media cues arises through changing individual levels of anxiety.

Another example is in the electoral politics literature. Gelman and King (1990) found the existence of a positive incumbency advantage in the election. A few years later, in 1996, Cox and Kats lead the incumbency advantage literature in a new direction by considering possible causal mechanisms that explain why incumbents have an electoral advantage. They decomposed the incumbency advantage into a "scare off/quality effect" and effects due to other causal mechanisms such as name recognition and resource advantage.

Mediation is playing an increasing important role also in educational studies. Following the words of A. Gamoran "the next generation of policy research in education will advance if it offers more evidence on mechanisms so that the key elements of programs can be supported and the key problems in programs that fails to reach their goals can be repaired" (A. Gamoran, 2013, President of the William T. Grant-Foundation). Also in a recent special issue of the Journal of Research on Educational

Effectiveness focused on mediation and it has been noted that "such efforts in mediation analysis are fundamentally important to knowledge building, hence should be a central part of an evaluation study rather than an optional 'add-on'" (Hong, 2012). We can find some empirical researches in the educational field like in Bijwaard and Jones (2018), who study the impact of education on mortality via cognitive ability, or Heckman, Pinto and Savelyev (2013), who study the effect of Perry Preschool Program through cognitive and non cognitive mechanisms.

Surprisingly, in economics, mediation analysis has been much less contemplated, notwithstanding it has interesting and important implications. We can find few examples in Simonsen and Skipper (2006), who evaluate the direct effect of motherhood's wage, Flores and Flores-Lagunes (2009), who evaluate the direct effect on earnings of the Job Corps program through work experience. Other contributions are given by Huber (2015), who used causal mediation framework to decompose the wage gap using data from the U.S. National Longitudinal Survey of Youth 1979, or by Huber, Lechner and Mellace (2017), who investigate whether the employment effect of more rigorous caseworkers in the counselling process of job seekers in Switzerland is mediated by placement into labor market programs. The common approach used to study causal mechanisms in economics is structural equation model (SEM), see for instance the seminal work by Baron & Kenny (1986). But, as demonstrated by Imai, Keele, Tingley and Yamamoto (2011), SEM is not the appropriate method to study and to identify causal mechanisms. They showed that structural models rely upon untestable assumptions and are often inappropriate even under the validity of those assumptions. In particular, conventional exogeneity assumptions alone are insufficient for identification of causal mechanisms[1], while it can be a sufficient condition for identification of the classical average treatment effect. In addition to that, the mediator could be interpreted as an intermediate outcome: in such a model we should control for a large set of covariates (pre and post treatment), risking to have different results depending on the covariates chosen and then increasing the sensitiv-

[1]Structural models are misused also in the traditional causal analysis, because of the presence of strong assumptions to justify the causal interpretation of mathematical results. See for example James, Mulaik and Brett (1982), Pearl (1998) and many others.

ity of the estimates. Therefore, the use of mediation in economics can be useful and efficient, and this is the main motivation of this essay.

To overcome these problems, relaxing the structural restrictions, over the last decades, some authors have moved mediation analysis in the potential outcome framework. Some examples are Robins and Greenland (1992); Pearl (2001); Petersen, Sinisi and van der Laan (2006); VanderWeele (2009); Imai, Keele and Yamamoto (2010); Hong (2010); Albert and Nelson (2011); Tchetgen Tchetgen and Shpister (2012); Vansteelandt, Bekaert and Lange (2012) from many others. As in the classical treatment analysis, using the counterfactual approach, rather than structural models, allows us to formalize the concept of causality without making assumptions on the functional form of the parameters and, then, to have more flexible identification procedures. Moreover, in this kind of models, it is not necessary to know the entire set of covariates that could affect the design. Most of this literature handles identification by assuming that the treatment and the mediator are conditionally exogenous given observed characteristics, an assumption known as Sequential Ignorability. Nevertheless, this assumption sometimes is hardly satisfied, above all in economics, because of the presence of post-treatment confounders, that can confound the relations between variables. To handle this problem, recently some researchers have used quasi-experimental designs inside the mediation framework. These procedures are particularly attractive in this context also because the gold standard of causal analysis, i.e. randomization of the treatment, is not a sufficient condition for the identification of causal mechanisms, a requirement that make the counterfactual approach more appropriate than structural models.

Causal mechanism is an important issue to better understand why a policy works and go beyond the limits of this approach is one of the aim of the current research fields. Mediation analysis seems to be one of the fittest frameworks to describe these relations and many researchers have developed new methods or have readapted the classical ones to go deep with the analysis. This is a promising methodology in economics because it permits to study causal mechanisms and to analyze the causal steps between treatment and outcomes and, then, it permits to give a causal interpretation to the changes that occur in between. In addition to that, these new methods

that are emerging allow to do this kind of analysis without making too restrictive assumptions, a key issue in economic studies; mediation turns out to be a precious tool for policy makers. Thus, following the words of Imai, Keele and Yamamoto, also economics is trying to open his black box[2].

This paper reviews the methodological advancements in causal mediation literature in economics, in particular focusing on quasi-experimental designs, a recent perspective in the mediation panorama. The remainder of this paper is organized as follows: section 2 shows the counterfactual approach in mediation analysis and defines the parameters of interest; section 3 analyzes the assumptions required in mediation analysis; section 4 focuses on quasi-experimental designs, in particular showing instrumental variables (IV), difference-in-differences (DID) and synthetic control approaches; section 5 concludes.

1.2 Counterfactual approach

1.2.1 Definition of counterfactual mediation framework

Most recent research in mediation analysis uses counterfactual approach commonly exploited in causal inference, basing on the potential outcome framework proposed by Neyman (1923) for randomized experiments and then generalized to observational studies by Rubin (1974).

According to the main literature, formally we denote with D a binary treatment,[3] with M the mediator variable, that is assumed to have a boundary support and may be discrete or continuous, and with Y the outcome of interest. In this framework the potential outcome is defined as $Y(d', m)$ and the potential mediator is $M(d)$ with $d, d' \in \{0, 1\}$. We can write the realized outcome and mediator values as:

[2]Imai, K., L. Keele, D. Tingley, T. Yamamoto (2011): "Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies", *American political science review*, 105(4), 765-789.

[3]We here focus on binary treatment indicator for simplicity, but the methods can be extended easily to non-binary treatment, see for instance Imai, Keele & Tingley, 2010a.

$$Y_i = D_i \cdot Y_i(1) + (1 - D)_i \cdot Y_i(0)$$
$$M_i = D_i \cdot M_i(1) + (1 - D)_i \cdot M_i(0)$$

where the subscripted i is the unit observation.

It is easy to see that for each unit i only one of the two potential outcomes or mediator states is observed. Thus, also in mediation analysis we have to face the so called missing values problem (Holland, 1986). Because of the presence of two driver variables we must also take into account the potential presence of an interaction between them, making the analysis more challenging.

The goal of mediation analysis is to decompose the total treatment effect of D on Y into the indirect and the direct effect. The first one reflects one possible explanation for why treatment works, explicitly defining a particular mechanism behind the causal impact and it answers the following counterfactual question: what change would occur to the outcome if the mediator changed from what would be realized under the treatment condition, that is $M_i(1)$, to what would be observed under the control condition, that is $M_i(0)$, while holding the treatment status at d?[4] The second one, the direct effect, represents all other possible explanations through which a treatment affects an outcome and it corresponds to the change in the potential outcome when exogenously varying the treatment but keeping the mediator fixed at its potential value $M_i(d)$. These methods assess what portion of the effect of the treatment operates through a particular intermediate variable and what portion operates through other mechanisms in order to prescribe better policy alternatives. Finally, mediation analysis is the set of techniques by which a researcher assesses the relative magnitude of these direct and indirect effects.

[4]See for instance Keele, Tingley and Yamamoto, 2015

1.2.2 Definition of parameters

Using the potential outcome notation, we can define three quantities of interest, see for instance VanderWeele (2015):

1. CDE(m): $[Y_i(1, m) - Y_i(0, m)]$ is the controlled direct effect and it expresses how much the outcome would change between treated and control groups but keeping fix $M = m$. It quantifies the effect not mediated by M, but it is defined for every strata of m. If the effect changes across different level of m, then we are in presence of an interaction effect between D and M on Y.

2. NDE(d): $[Y_i(1, M(d)) - Y_i(0, M(d))]$ is the natural direct effect and it expresses how much the outcome would change if the treatment was exogenously set from 1 to 0 but, for each individual, the mediator was kept at the level it would have taken in treatment status d. It captures what the effect of the treatment on the outcome would remain if we were to disable the pathway from the treatment to the mediator.

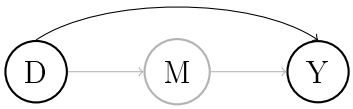


Figure 1.1: Natural direct effect

3. NIE(d): $[Y_i(d, M(1)) - Y_i(d, M(0))]$ is the natural indirect effect and it expresses how much the outcome would change if the treatment were set equal d but the mediator were changed from the level it would take if D=1 to the level it would take if D=0. It captures the effect of the treatment on the outcome that operates through the mediator.

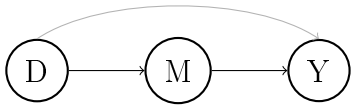


Figure 1.2: Natural indirect effect

These effects are defined at the unit level, implying that they are not observed for each observation i with the consequence that we cannot directly identify them without stronger assumptions. The reason is that they are defined with respect to multiple potential outcomes for the same individual and only one of those potential outcomes is observed in reality. So, we use the population averages for the identification of all the effects of interest. Basing on the potential outcome framework (Glynn 2012; Imai, Keele, Yamamoto 2010; Pearl 2001; Robins & Greenland 1992), we can identify these quantities of interest to disentangle the average total effect (ATE) given by $\Delta = E[Y(1) - Y(0)]$. First, we define the average indirect effect (ACME)[5] as:

$$\bar{\delta}(d) = E[Y(d, M(d)) - Y(d, M(1-d))] \qquad \forall\ d \in \{0, 1\} \tag{1.1}$$

The ACME corresponds to the change in mean potential outcome when exogenously shifting the mediator to its potential values under treatment and non treatment state but keeping the treatment fixed at $D = d$. Note that only one component of the right side equation is observable, whereas the other one is by definition unobservable (under treatment status d we never observe the value of M that it naturally would have under the opposite treatment state, i.e. $M(1-d)$).
In the same way, we define the average direct effect (ADE) as:

$$\bar{\theta}(d) = E[Y(d, M(d)) - Y(1-d, M(d))] \qquad \forall\ d \in \{0, 1\} \tag{1.2}$$

[5]Also known as Average Causal Mediation Effect.

It represents the average causal effect of the treatment on the outcome when the mediator is set to the potential value that would occur under treatment status d. It can be easily shown that ATE can be rewritten as the sum of the natural direct and indirect effect defined on the opposite treatment status:

$$\begin{aligned}\Delta &= E[Y_1 - Y_0] \\ &= E[Y(1, M(1)) - Y(0, M(0))] \\ &= E[Y(1, M(1)) - Y(0, M(1))] + E[Y(0, M(1)) - Y(0, M(0))] = \bar{\theta}(1) + \bar{\delta}(0) \\ &= E[Y(1, M(0)) - Y(0, M(0))] + E[Y(1, M(1)) - Y(1, M(0))] = \bar{\theta}(0) + \bar{\delta}(1)\end{aligned}$$

We obtaine these results simply adding and subtracting the counterfactual quantity $E[Y(0, M(1))]$ after the second equality, and adding and subtracting $E[Y(1, M(0))]$ after the third equality. More in general, we can write this result as:

$$\Delta = \bar{\delta}(d) + \bar{\theta}(1 - d) \qquad \forall\ d \in \{0, 1\} \tag{1.3}$$

Obviously, neither effect is identified without further assumptions: only one of $Y(1, M(1))$ and $Y(0, M(0))$ is observed for any unit, because both outcomes cannot be observed at the same time as stated in the fundamental problem of causal inference. The counterfactual quantities $Y(1, M(0))$ and $Y(0, M(1))$ are never observed for any individual, because we never observe the potential value of M defined under the opposite treatment state, but we only know the factual M that follows a particular treatment state. To face this identification issue we need to define a proper set of assumptions.

1.2.3 Controlled direct effect versus natural direct effect

An important advantage of the counterfactual notation is that it allows for the potential presence of heterogeneity. Such heterogeneity is important both in practice and theory, as it is often the motivation for the endogeneity problems that concerns

economists (Imbens and Wooldridge, 2009). In structural models the effects are assumed to be constant, implying that the effect of various policies could be captured by a single parameter. In mediation this heterogeneity is even more important, because it implies not only that the direct effect of the treatment on the outcome could be different across individuals, but also that this effect can be different for different values of the mediator. With the counterfactual notation, then, the presence of non linearities and interactions is not a problem, because we don't need to specify the functional form and we don't need to model the relations between variables. But if the effect of the treatment is the same for the entire population, meaning that it doesn't change for different level of the mediator, then there is no interaction between treatment and mediator. In this particular case, $CDE(m) = CDE(m')$, for $m \neq m'$, implying that the controlled direct effect is equal to the natural direct one, $CDE = NDE$ (Baron & Kenny, 1986). Formally:

$$\bar{\delta}(1) = \bar{\delta}(0) = \bar{\delta}$$

$$\bar{\theta}(1) = \bar{\theta}(0) = \bar{\theta}$$

In this situation, the difference between the total effect and the controlled direct effect gives the indirect effect, or more formally: $\Delta - \bar{\theta} = \bar{\delta}$.
Usually, in empirical analysis the controlled direct effect and the natural direct effect do not coincide and then the difference between the total effect and the controlled direct effect does not generally give an indirect effect (Kaufman et al., 2004; Vanderweele, 2009) because there may simply be interaction between the effects of the exposure and mediator on the outcome, not guaranteeing the additional linearity functional form of the effects.

1.3 Assumptions

1.3.1 Classical Assumptions

Usually, in economics we can't manage a controlled experiment. In this situation we must rule out the presence of confounders. But, in mediation analysis, because of

the particular structure of the variables' relations, it is important to point out what kind of confounders we have.
Consider a classical mediation framework, in which X is a set of pre-treatment observable covariates and W is a set of post-treatment observable confounders, like in figure 1.3.1.

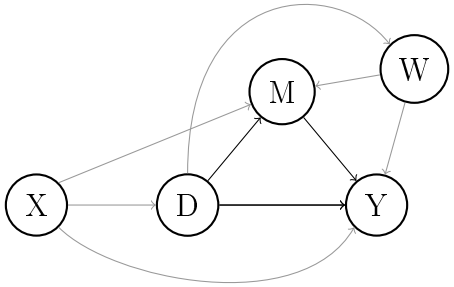


Figure 1.3: Mediation framework with: D=treatment; M=mediator; Y=outcome; X=pre-treatment covariates; W=post-treatment confounders

For example, suppose we want to assess whether a program of subsidies (D) increases firms' productivity (Y) and whether the share invested in R&D (M) may mediate part of this effect. In this example, investments in R&D may be a potential mediator because is affected by subsidies treatment and in turn may affect, at least partially, productivity outcome. But to interpret this association as a causal effect, we need to think carefully about and control for variables that may be confounders of the treatment-outcome relationship (X) and/or of the mediator-outcome relationship (W). For example, there might be a firms' size or firms' performance variables (X) that affect the participation in the program (D) and the firms' productivity (Y) or other factors, such as the quality of administration or the presence of a network (W), that affect both the level of investments in R&D and productivity. It is important to note that these W confounders could be affected by the treatment itself.
In such a context, we need to distinguish two situations: the identification of con-

trolled effects and the identification of natural effects. Following VanderWeele (2015) to estimate the CDE we need two assumptions:

- A1. There must not be confounders between treatment and outcome relationship
- A2. There must not be confounders between mediator and outcome relationship

For the satisfaction of the first assumption is sufficient to randomize the treatment, but even with randomized treatment the second assumption might not hold. If we refer to the previous example, to satisfy A1 we need to adjust for common causes of the exposure and the outcome - for example information about firms' size or firms' performance or any other factor (X) that can confound this relation in the analysis; or we can give subsidies randomly, implying the same distribution of X for treated and non-treated firms. At the same time, to satisfy A2 we need to adjust for common causes of the mediator-outcome relation - for example information about administration's quality or other factors (W) that can confound this relation. In this case, we need to think carefully to all possible post-treatment confounders and include them in the analysis, because the randomization of the treatment is not a sufficient condition to control for W.
To identify natural direct and indirect effects we need two more assumptions. In particular:

- A3. There must not be confounders between treatment and mediator relationship
- A4. There must not be confounders affected by the treatment between mediator and outcome relationship

Also in this case, to satisfy A3 is sufficient to randomize the treatment, but again for the fourth assumption this is not enough. In particular, A4 is a strong assumption, because it requires that there is nothing on the pathway from the treatment to the mediator that also affects the outcome. This assumption is more plausible if the mediator occurs shortly after the treatment (VanderWeele and Vansteelandt,

2009). If we consider again the previous example, the size of the firm could be a confouder of the treatment-mediator relation and then it must be included in the set of covariates (X) or we can randomly assign D. But if we consider A4, we have to take into account possible factors that could be affected by the treatment and that in turn affect the mediator and/or the outcome. For example, firms that receive subsidies could have more benefits (tax, bureaucratic) that could in turn affect R&D investments and productivity. The problem with this assumption is that, even if we have in mind these factors before the analysis, we can't have the exact measure of these confounders, because we don't know before the value that they will take after the treatment.

Another example could be the effect of a job training program (D) on the probability to find a job (Y). It could be possible that the program is designed in two steps: the first part of the program in which we can find different activities and a second part in which there is, for example, a PC course (M). In this kind of design, we can study how much the probability to find a job increases thanks to the PC course and/or thanks to the other components of the job training program. Also in this case, to correctly identify the direct and the indirect effect we must be sure to satisfy the previous four assumptions. In other words, we have to control for all possible pre-treatment confounders, like gender, age, education, kind of job, how long the individual is unemployed and so on, and we have to control for all possible post-treatment confounders, eventually affected also by the treatment, like the previous knowledge of PC, attitude and so on.

It is important to note that assumptions 1-4 implicitly imply an assumption of temporal ordering (Cole & Maxwell, 2003). If the temporal ordering assumptions were not satisfied, then neither would the no unmeasured confounding assumptions, and then the association would not represent the causal effect. For this last assumption it is important to use panel data to measure the various factors at different time: such framework consisting of an initial treatment, an intermediate mediator, a final outcome and, possibly, observed covariates. Differently, with cross sectional data, we cannot determine the direction of causality or the relative magnitude of the two possible directions that causality may operate and we cannot distinguish between

mediation and confounding (see, for instance, Baron & Kenny, 1986). In addition, it is important to have in mind a strong theory to give the right causal interpretation. Another issue is that when the treatment, the mediator and the outcome vary over time, we have to control for the prior values of these variables to make no confounding assumptions more plausible and to rule out the possibility of reverse causality (Vanderweele, 2015): even if we know the temporal ordering, it is possible that prior values of the variables serve as the most important confounding variables.

1.3.2 Identification under Sequential Ignorability

The key insight is that under randomized designs ATE is identified, but direct and indirect effect are not. Even in the presence of a double randomization of the treatment and the mediator the effects of interest are not identified without further assumptions. In fact, even if both treatment and mediator are exogenous, and then the conventional exogeneity assumption is satisfied, simply combining the effect of T on M and the effect of M on Y is not sufficient for the identification of the indirect effect. The assumption called "Sequential Ignorability" is a partial solution to this problem and so far the most used. There are different interpretations of this assumption, with different implications and different formalizations. The most used version and maybe the most flexible is the one given by Imai, Keele and Yamamoto (2010). Formally, it is expressed as:

$$\{Y_i(d', m), M_i(d)\} \perp D_i | X_i = x \tag{1.4}$$

$$Y_i(d', m) \perp M_i(d) | D_i = d, X_i = x \tag{1.5}$$

where:

$$Pr(D_i = d | M_i = m, X_i = x) > 0 \tag{1.6}$$

$$\forall\ d \in \{0, 1\} \text{ and } m, x \text{ in the support of } M, X^6$$

[6]Imai, Keele, Tingley and Yamamoto (2011) wrote this common support assumption as: $0 < Pr(D_i = d | X_i = x)$ and $0 < P(M_i = m | D_i = d, X_i = x)$ for $d = 0, 1$ and all x and m in the support

The first part of the sequential ignorability assumption, equation (4), is the classical conditional independence of the treatment, also known as no-omitted variable bias, conditional exogeneity or unconfoundedness, see for instance Imbens (2004). By equation (4), there are no unobserved confounders jointly affecting the treatment and the mediator and/or the outcome given X, meaning that we can consistently identify the effect of D on Y and D on M. In non-experimental designs, the validity of this assumption hinges on the richness of pre-treatment covariates, while in experimental designs, this assumption holds if the treatment is either randomized within strata defined by X or randomized unconditionally[7]. The second part of sequential ignorability assumption, equation (5), states that there are no unobserved confounders jointly affecting the mediator and the outcome once we condition on D and X. It means that there are no unobserved confounders between mediator and outcome, ruling out the presence of post-treatment confounders not captured by X. This is a strong assumption because randomizing both treatment and mediator does not suffice for this assumption to hold; in addition to this, it is more plausible if treatment and mediator are measured at a short distance, as we mentioned in the previous subsection. The last part of sequential ignorability, equation (6), is the common support assumption. It states that the conditional probability to receive or not receive the treatment given M and X, recalling the propensity score literature, is larger than zero[8]. By Bayes' theorem, this version of common support implies that $Pr(M_i = m|D_i = d, X_i = x) > 0$ if M is discrete or that the conditional density of M given D and X is larger than 0 if M is continuous. The main implication of the equation (6) is that conditional on X, the mediator state must not be a deterministic function of the treatment, otherwise no comparable units in terms of the mediator are available across different treatment states (Huber, 2019). In other words, there must be different values of M once we condition on D and X, in order to compare different mediator states inside the same group defined by the treatment status. Under sequential ignorability (equations 4-6), it is possible to identify causal mechanisms,

of X and M.

[7] In this case, the stronger version of the assumption $\{Y_i(d', m), M_i(d), X_i\} \perp D_i$ is satisfied.

[8] In the classical causal analysis to identify the ATE we face the weaker common support assumption: $Pr(D_i = d|X_i = x) > 0$

in particular, we can get the nonparametric identification of the counterfactual quantity $E[Y_i(d, M(d')|X_i = x]$, proved by Imai, Keele and Yamamoto (2010), implying the nonparametric identification of the average natural direct (ADE) and the average natural indirect effect (ACME). In the standard causal mediation analysis the nonparametric identification of the counterfactual quantity is the following:

$$
\begin{aligned}
&E[Y_i(d, M_i(d'))|X_i = x] = \\
&= \int E(Y_i(d, m)|M_i(d') = m, X_i = x]\, dF_{M_i(d')|X_i=x}(m) \\
&= \int E(Y_i(d, m)|M_i(d') = m, D_i = d', X_i = x]\, dF_{M_i(d')|X_i=x}(m) \\
&= \int E(Y_i(d, m)|D_i = d', X_i = x]\, dF_{M_i(d')|X_i=x}(m) \\
&= \int E(Y_i(d, m)|D_i = d, X_i = x]\, dF_{M_i(d')|D_i=d',X_i=x}(m) \\
&= \int E(Y_i(d, m)|M_i = m, D_i = d, X_i = x]\, dF_{M_i(d')|D_i=d',X_i=x}(m) \\
&= \int E(Y_i|M_i = m, D_i = d, X_i = x]\, dF_{M_i(d')|D_i=d',X_i=x}(m) \\
&= \int E(Y_i|M_i = m, D_i = d, X_i = x]\, dF_{M_i|D_i=d',X_i=x}(m)
\end{aligned}
$$

where, assuming a continuous mediator, the first equality follows from the law of iterated expectation; equation (4) is used to establish the second, the fourth and the last equalities; equation (5) is used to establish the third and the fifth equalities, while the sixth equality follows from the fact that $M_i = M_i(D_i)$ and $Y_i = Y_i(D_i, M_i(D_i))$, also known as observational rule (T. VanderWeele, 2015) or consistency assumption (Imai, Keele, Tingley and Yamamoto, 2011).

Sequential ignorability used in the counterfactual analysis is crucially different w.r.t. the classical exogeneity assumption used in the structural models. In particular, as we said before, to identify causal mechanisms, and then the indirect effect that goes from T to Y through M, is not sufficient randomize both treatment and mediator. Differently, if we use structural models, it is required to satisfy only the exogene-

ity assumption, meaning that it's sufficient the double randomization of T and M. Nevertheless, the resulting estimation is consistent only if there is not heterogeneity effect. In particular, in the first case we can identify the causal mediation effect (T $\rightarrow$ M $\rightarrow$ Y) in which we are interested in, while in the second case we can just identify the causal effect of the mediator (T $\rightarrow$ M and M $\rightarrow$ Y). These two quantities coincide only in the absence of heterogeneity. Under exogeneity assumption and in the absence of heterogeneity, then, we can consistently estimate only CDE, because in this particular case this quantity is equal to NDE. The interesting fact is that, in the presence of heterogeneity, the exogeneity assumption still holds if treatment and mediator are randomized, but the correlation between the error terms of M and Y is different from 0, implying biased estimations of the effects, that structural models are not able to capture.

1.3.3 Other interpretations of Sequential Ignorability

The main limit of this result is that the nonparametric identification works only if we don't condition on post-treatment confounders, implying that the set of pre-treatment observable confounders must be sufficient to control for them, requirement not always credible. This issue has been addressed by Robins (2003). In his fully randomized causally interpreted structured tree graph model (FRCISTG), he used a different version of sequential ignorability: the first part is the same of equation (4), while equation (5) is replaced by $Y_i(d', m) \perp M_i(d)|D_i = d, Z_i = z, X_i = x$, where Z is a vector of post-treatment confounders. This is an important practical advantage because permits to control for observable variables that could confound the relationship between the mediator and the outcome. But it comes at the cost of adding the parametric assumption of non-interaction between direct and indirect effect: $Y_i(1, m) - Y_i(0, m) = B_i$, where B_i is a random variable independent of m. This condition has two implications: (i) absence of heterogeneity; (ii) the same value of the direct effect regardless the level of the mediator, i.e. the independence between the direct and the indirect effect. Therefore, it exists an important trade-off: if we condition on post-confounders, we need to assume a non-interaction assumption

to identify natural effects, which is very restrictive condition and it doesn't permit a nonparametric identification. On the other hand, if we don't condition on post-treatment confounders, but assuming that all the X's are sufficient to control for them, we can identify the effects nonparametrically without any parametric restrictions.

Another formalization of Sequential Ignorability is given by Pearl (2001). In particular, in his Theorem 1 and Theorem 2 for the identification of the average natural direct effect and in Theorem 4 for the identification of the average natural indirect effect, he used a different set of assumptions arriving anyway at the same expression of ADE and ACME given by Imai, Keele and Yamamoto (2010). It is important to note that sequential ignorability implies Pearl's assumptions, while the converse in not always true, but in practice, the difference is only technical. Another advantage of sequential ignorability is that it is easier to interpret than Pearl's assumptions, in which we have an independence between two potential quantities[9]. This difficulty in the interpretation is pointed out also by Pearl himself: "Assumptions of counterfactual independencies can be meaningfully substantiated only when cast in structural form"[10]. In contrast, in the second part of sequential ignorability, eq. (5), we have the observed value $M_i(d)$ independent of potential outcome, in other words M_i is effectively randomly assigned given $D_i = d$ and $X_i = x$, a concept that is easier to understand.

A further version of sequential ignorability is given by Petersen, Sinisi and Van der Laan (2006). They split equation (4) into two parts: $Y_i(d, m) \perp D_i | X_i = x$ and $D_i \perp M_i(d) | X_i = x$, while equation (5) is the same[11]. This is just a mathematical difference, because in experimental designs, in which treatment is randomized, equation (4) is equivalent to them. To identify the natural direct effect they also assume that the potential value of mediator under non-treatment state is independent of the potential outcome. Formally, $E[Y_i(d, m) - Y_i(0, m) | M_i(0) = m, X_i = x] = E[Y_i(d, m) - Y_i(0, m) | X_i = x]$, meaning that the potential value of the mediator under

[9] The assumption given in Pearl(2001) is: $Y_i(d', m) \perp M_i(d) | X_i = x$

[10] See Pearl (2001), pag. 416

[11] In particular, they use $Y_i(d, m) \perp D_i | X_i = x$ and $D_i \perp M_i(d) | X_i = x$ to identify controlled direct effect and they add equation (5) to identify natural direct effect.

non treatment state, $M_i(0)$, doesn't give us any additional information on the effect of the treatment. This additional assumption is necessary to identify the counterfactual quantity $Y(d, M(0))$. Anyway, if treatment is randomized this last assumption is not necessary for the nonparametric identification given by Imai, Keele and Yamamoto (2010), making their sequential ignorability a preferable solution once again.

1.4 Quasi-experimental designs

As mentioned in the previous section, most recent research in mediation analysis considers more general identification approaches based on the potential outcome framework, commonly used in treatment evaluation (Rubin, 1974) to overcome the limits of structural models. The gold standard of this approach is the randomness of the treatment, a condition that is easily met in experiments. When treatment or mediator cannot be determined exogenously, the only way to estimate the parameters of interest and give them a causal interpretation is to use quasi-experimental designs, in which endogeneity can be controlled under particular assumptions. Mediation analysis borrowed these methods from causal literature in order to identify and estimate causal mechanisms, but, nowadays, there are only few studies using these approaches. We can find some examples in Instrumental variables (See for example Robins and Greenland, (1992); Geneletti (2007); Imai et el. (2013); Powdthavee et al. (2013); Burgess et al. (2015); Jhun (2015); Frölich and Huber (2017)), Difference-in-differences (see Deuchert, Huber and Schelker (2018); Huber and Steinmayr (2017)) and synthetic control (see Mellace and Pasquini (2019)), while, at the best of our knowledge, there are not still studies using regression discontinuity design[12]. In the next section we will discuss some of them.

[12]See M. Angelucci, V. Di Maro (2010): they provide a practical guide for the identification of treatment effect on eligibles and the indirect effect on ineligibles based on conditional independence, RD and IV assumptions

1.4.1 Instrumental variables

Recently, part of the literature tried to study causal mechanisms through instrumental variables (IV) methods (see Robins and Greenland (1992); Imai et al. (2013) from many others). The reason is that, in some empirical applications, sequential ignorability is not a credible assumption to rule out the presence of post-treatment confounders and an instrument could be an important tool to solve the problem of the mediator's endogeneity. In other cases, also the treatment is not exogenous even after conditioning on a set of pre-treatment covariates and a second instrument could be used for this kind of endogeneity. We can find two different ways in which mediation analysis with IV has been dealt. Some authors identified direct and indirect effects through structural models. For example, Powdthavee, Lekfuangfu and Wooden (2013) studied the impact of education on subjective well-being (SWB) through the mediator income. They used different timing of education laws across states of Australia and shocks in personal income (such as lottery wins etc.) as instruments respectively of treatment and mediator. Assuming the independence between instruments, they estimate the direct and indirect effect using the structural equation model (see Baron & Kenny, 1986), inside a 2SLS framework. Other studies used two instruments and a parametric identification such as Burgess et al. (2015) and Jhun (2015). Ten Have et al. (2002) used treatment-covariates interactions as instruments for the mediator, but imposing the absence of the treatment-mediator, mediator-covariate and treatment-covariate interactions in the outcome model, implying an identification based on strong structural restrictions. The limit of this structural methods is that they don't allow for the existence of a heteorgeneous effect between direct and indirect effect.

The second way in which mediation analysis with IV can be studied is using the potential outcome framework. An important contribution is given by Chen, Chen and Liu (2017), who studied the gender of the second born on the first born education outcome, through the sibling size (also interpreted as fertility choice) mediator. In their study, they assume a randomized sibling gender and they use a twinning indicator at the second birth as instrument for the mediator (following the studies of

Rosenzweig and Wolpin (1980); Black, Devereux and Salvanes (2005); Angrist, Lavy and Schlosser (2010)). Their IV estimates give a causal interpretation limited only to complying families, whose sibling size would rise with twinning at the second birth, i.e. $M(Z = 1) > M(Z = 0)$, but, on the other hand, allowing for heterogeneous effect, i.e. interaction between treatment and mediator. In particular, they found that having a younger brother lowers the potential sibling size of a first-born girl to a degree that the positive indirect effect cancels out the negative direct effect on her education outcomes, resulting in a near zero total effect. These results offer new evidence about gender bias in family settings that has not been detected in the previous literature. This was possible thanks to the decomposition of the total effect and thanks to the presence of heterogeneity captured by the interaction between sibling size and sibling gender. A second contribution using the potential outcome approach is given by Frölich and Huber (2017). They used a counterfactual framework and join a nonparametric identification using two different instruments respectively for treatment and mediator, allowing, then, for the endogeneity of them. In addition, both instruments and mediator can be discrete or continuous. The main advantage of their result is that they identify natural and controlled effects for all treatment compliers, overcoming the limit of identification only of the controlled direct effect for subpopulations defined on compliance in either endogenous variable (see Miquel, 2002). They applied this method on two empirical studies. One of them is about the effect of education on the social life outcome through income. Treatment is instrumented by an increase in the UK minimum school leaving age in 1971 from 15 to 16 years (see also Oreopoulos, (2006) and Brunello et al., (2013)), while the annual individual income is instrumented by windfall income (Lindhal, (2005) and Gardner and Oswald, (2007)). They found a positive effect of education on social life functioning, but disentangling the total effect on compliers (LATE) showed a positive direct effect, while the indirect effect is close to 0 and not significant. They then conclude that education affects social functioning, but through different mechanisms than income (Huber and Frölich, 2017).

1.4.2 Difference-in-differences

The first contribution that deals the identification of direct and indirect effect using a different framework than sequential ignorability and instrumental variables approach is given by E. Deuchert, M. Huber and M. Schelker (2018). They disentangle the total effect basing on a difference-in-differences (DID) approach within subpopulation or strata (Frangakis and Rubin, 2002) defined upon the reaction of a binary mediator to treatment, implying the presence of four subpopulations: always takers, never takers, compliers and defiers (see for instance Angrist, Imbens and Rubin, 1996). In particular, they identify the direct effect on always takers and never takers, whose mediator doesn't react to treatment, i.e. treatment doesn't change the mediator's state, corresponding to the controlled direct effect, and then they identify the indirect effect and the direct effect on compliers, whose mediator reacts to treatment. The main assumptions that they use are the classical random treatment assignment; the second one is the monotonicity assumption that comes from the local average treatment effect (LATE) literature (see Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996), ruling out the presence of defiers. The last important set of assumptions is the common trend assumptions, which come from the DID literature, but now defined across strata. This fact permits to control for post-treatment confounders and it allows for differences in the effects of unobservable confounders on specific potential outcomes across strata, as long as these differences are time constant. As discussed in this paper, the identification of the effects of interest under principal strata in mediation has been criticized for not permitting a decomposition of direct and indirect effect on compliers in a DID framework and focussing on subgroups that may be less interesting than the entire population (VanderWeele 2008). But thanks to previous set of assumptions the authors identify the effects on compliers and they present an empirical application in which the effect on subgroups is relevant for political decision making[13]. A second critique is about confusion made in the literature between mediation and principal stratification causal effects (VanderWeele 2012). In

[13]The empirical application is about the Vietnam draft lottery in the US (1969-1972) on political preferences and personal attitudes. The mediator of interest is military service during the Vietnam War.

particular, it is important to note that $E[Y_1 - Y_0|M(1) = 1, M(0) = 0]$ is the total causal effect of treatment on the outcome for the compliers subgroup and it doesn't always correspond to the mediated effect. To notice this fact, we can observe that this effect can be nonzero even if the intermediate variable has no effect on the outcome, meaning that M is not a mediator. This happens whenever M is a surrogate for the effect of the treatment on the outcome: surrogacy concerns whether the effect of a treatment on an outcome can be predicted by the effect of a treatment on an intermediate variable, whereas mediation concerns whether the effect of treatment goes to the outcome through the mediator. A good surrogate may be often a mediator, but it need not be (Vander Weele, 2012). Principal stratification is a good framework to capture surrogacy, while natural effects (Pearl 2001, from many others) are the appropriate concept to study mediation. An intuitive example is given by Lindsay Page (2012), who provides evidence that Career Academies program (D) had a substantial effect on subsequent earnings (Y) those for whome the program would change exposure to the world-of-work (M) but not those for whome it would not change exposure to the world-of-work. In her analysis, she used a Bayesian approach to principal stratification and she used covariates to attempt to predict which principal stratum different individuals belong to. But, even if these assumptions hold, it could happen that there are still some unmeasured confounders of the mediator-outcome relationship, like motivation (U), that make M a surrogate rather than a mediator, like in Figure 1.4.2.

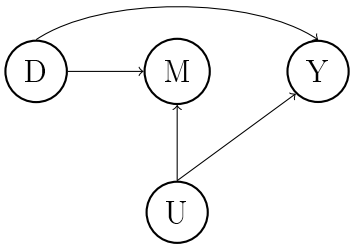


Figure 1.4: Principal strata causal effect with no mediation

A possible solution is to study mediation with principal strata approach, but adding

the sequential ignorability assumption to rule out the potential presence of post-treatment confounders.

1.4.3 Synthetic control

To the best of our knowledge, the only contribution that uses synthetic control method (SCM) to study causal mechanisms is given by Mellace and Pasquini (2019). The main advantage of this method is that it estimates total causal effects, even in presence of only one treated unit and few control units (Abadie et al. (2010)). They develop a generalization of SCM that allows disentangling the total effect into its direct and indirect component defining a Mediation Analysis Synthetic Control (MASC). The procedure that they use consists in re-weighting control unit post-intervention outcomes by choosing weights that minimize the distance between treated and control in pre-intervention observable characteristics as well as in post-intervention values of mediator. This allows to mimic what would have happened to the treated in the absence of the intervention if her mediator were set to her potential mediator under treatment (Mellace and Pasquini, 2018). In particular, they use a dynamic factor model with interactive fixed effects as in Abadie et al. (2010).

1.5 Conclusions

Mediation analysis is a promising methodology in economics, because it allows to study causal mechanisms of transmission of a policy without making unreliable and often restrictive assumptions: it permits to know not only if a policy is working or not, but also why, going into a deeper level of analysis. In the literature, there are not many economic applications and this could be due to technical difficulties and to the absence of clear methodological developments. We reviewed the pillars of this methodology, presenting current results and advancements and providing that it's a validated method, that can be used to investigate the changes that occur between inputs and outputs, answering the opened questions of economic studies. Causal mediation analysis is the statistical tool to understand causal mechanisms and it may

bring to an improvement in the power of quantitative analysis of economic phenomena.

This paper provides a survey of methodological developments in causal mediation analysis in economics, with a specific focus on quasi-experimental designs. We presented several methods, often used by economists and statisticians, that are clearly useful and fruitful for economic causal analysis. In the first part, we defined direct and indirect effects, both formally and mathematically. Next, we discussed the main assumptions needed for the identification of the counterfactual quantities of interest, with particular attention to the sequential ignorability assumption. In the fourth section we reviewed the main studies that use quasi-experimental designs, a new frontier in this field, discussing in particular instrumental variables, difference-in-differences and synthetic control approaches.

Chapter 2

Identification of causal mechanisms through an RD approach

2.1 Introduction

In economics causal analysis is a fundamental instrument to study the average effect (ATE) of a policy or, more in general, the effects of a treatment on an outcome of interest. But, in the last decades, many studies are going beyond the ATEs, studying the causal mechanisms through which the treatment transmits the effects to the outcome. In this case, the researchers want to disentangle the total effect into a direct effect of the treatment on the outcome and into an indirect effect, that operates through one or more intermediate variables, called mediators, that lie in the causal pathway between treatment and outcome. This kind of analysis has important and interesting implications, because it helps to better understand the policymakers' decisions and to implement changes in the policies' designs to make them more efficient. However, even in experimental designs, the causal mechanisms are not easy to identify. The main problem is that analyzing causal mechanisms requires stronger identifying assumptions than evaluating the classical ATE. First of all, the randomization of the treatment does not imply the randomness of the mediator, as discussed in Robins and Greenland (1992). The consequence is that the total effect cannot be

disentangled by simply conditioning on the mediator, because this generally introduces selection bias coming from variables influencing both the mediator and the outcome (Rosembaum, 1984). For this reason, several methods have been developed, based on different sets of assumptions and with different strategies for the estimation. In particular, while earlier works often relied on tight linear specifications, as in Judd and Kenny (1981) and Baron and Kenny (1986), more recent studies focus on nonparametric and semiparametric identification, allowing for nonlinearities and heterogeneity in the effects of interest, as in Pearl (2001), Robins (2003), VanderWeele (2009), Imai, Keele and Yamamoto (2010) and Huber (2014), among many others.

The main contribution of this paper is to show the identification and the estimation of the natural direct and indirect effects with an implementation of a Regression Discontinuity Design (RDD) to solve the problem of the endogeneity of the mediator. Thanks to the presence of a continuous, observable forcing variable Z that generates variation in the state of the mediator, we can rule out the presence of unobservable and observable post-treatment confounders that jointly affect the mediator and the outcome. In particular, we explain two different models in wchich Z has a different role, relying on two different sets of assumptions. In the first Model, Z is affected by the treatment and in turn deterministically affects the mediator.[1] Under these structural relations only a parametric identification is possible, recalling the linear structural equation model of an influential article by Baron and Kenny (1986). In the second model, Z is an exogenous variable, but still deterministically affects the mediator.[2] Under these relations, we have a non-parametric identification and then a more flexible framework, in which we do not need the specification of the functional form and, at the same time, the heterogeneity of the effects is allowed. In both cases, then, we use a sharp RDD, see for instance Trochim (1984), Imbens and Lemieux (2008), Lee (2008), Lee and Lemieux (2010), with the consequence that we have only

[1] We could think about treatment like a job training program in which there are PC lectures. Z is a PC test score to measure the knowledge after the training and the mediator is a PC course that people have to attend only if they do not get the sufficiency at the test.

[2] Following the example in the previous footnote, there are not PC lectures in the job training, but the rule to attend the PC course is the same.

compliers as population of reference.
So, it is the first study that uses RD method to solve the problem of the endogeneity of the mediator and then it is the first methodological contribution that join these two literatures.[3] Moreover, there are very few studies using quasi-experimental designs to disentangle the total treatment effect: an important contribution is given by E. Deuchert, M. Huber and M. Schelker (2018), who use a difference-in-differences approach for disentangling the total treatment effect, providing an empirical application based on the Vietnam draft lottery. Secondly, the estimation procedure is easy to implement, basing on a local regression to get the effect of the treatment on the outcome weighted by the treatment propensity scores that are straightforward to implement by a probit (or logit) estimation to get the potential values of the mediator. The estimation is computed in a bandwidth defined by $\bar{z}$.
The remainder of this paper is organized as follows. Section 2 defines the parameters of interest. Section 3 presents model 1 and discusses the identifying assumptions, the parametric identification and gives a graphical interpretation. Section 4 presents model 2 with its assumptions, the non-parametric identification and the graphical interpretation. Section 5 shows the estimation procedure and in section 6 we present a simulation study which shows the behavior of the estimators. Section 7 concludes.

2.2 Definition of parameters

The aim of mediation analysis is to decompose the average treatment effect (ATE) of a binary treatment D on the outcome Y into a direct effect and into an indirect effect that operates through an intermediate variable, the mediator M, that lies in the causal pathway between treatment and outcome. Estimate these two effects is not too easy in empirical designs, because of the presence of post confounders that could occur after the treatment, see for instance Imai, Keele, Tingley and Yamamoto (2011). Even in the presence of a double randomization respectively of the treatment and the mediator, we cannot identify a mechanism (Imai, Tingley and Yamamoto,

[3] An intuition about using RD to study indirect effects is given by Angelucci and Di Maro (2015).

2013). To solve this problem we use a continuous and observed forcing variable Z[4] that can induce an exogenous change in the mediator state, depending if Z exceeds a known cutoff point z^*, recalling the regression discontinuity (RD) literature (Lee, 2008; Lee & Lemieux, 2010).

2.2.1 Parameters of interest

To define the parameters of interest in this new setting that combines mediation framework and RD design we make use of potential outcome notation, see for instance Neyman (1923) and Rubin (1974). We denote by $Y(d', m)$ and $M(d)$ the potential outcome and the potential mediator state, with $d, d', m \in \{0, 1\}$. Furthermore, we denote by $Z = z^*$ the cutoff point at which the mediator state changes sharply, according to the following deterministic rule: $M_i = \{1[Z \geq z^*]\}$, where the subscripted i is the individual observation. We have two important implications thanks to this rule.

The first one is that, because of the status of M depends deterministically on Z, there is no an error term in the selection into M, implying the absence of unobservable factors that could create the presence of Always takers, Never takers and Defiers in the behavior of M with respect to Z. In our sharp setting, we have only Compliers[5], meaning that who is above (or below[6]) z^* will have $M = 1$ ($M = 0$) and vice versa. In this way we can identify the potential value of the mediator, because we know for that population (Compliers) what would be the value of M under the opposite treatment status, simply looking at the control group.

The second key point is that, because we take only individuals just above and below the threshold, defined as $\bar{z} \in [z^* - \epsilon, z^* + \epsilon]$ according to the RD literature, the value of the mediator is like randomized in this window, meaning that units in our population of interest will have comparable observables and unobservables characteristics.

In this context, we can locally define our parameters of interest as:

[4] In literature known also as "running" variable.

[5] The compliance is defined with respect to the forcing variable Z rather than to the treatment.

[6] The interpretation depends on the definition of the score.

$$\theta(d) = E[Y(1, M(d)) - Y(0, M(d))|Z = z^*], \qquad d \in \{0, 1\} \tag{2.1}$$

$\theta(d)$ is the average natural direct effect (Pearl, 2001)[7] for the population near the threshold and it expresses how much the mean potential outcome would change if the treatment was set from 1 to 0 but the mediator was kept at the potential level it would have taken in treatment status equal d. It captures what the effect of the treatment on the outcome would remain if we were to disable the pathway from the treatment to the mediator for the local population.
In the same way, we can define the local natural average indirect effect as:

$$\delta(d) = E[Y(d, M(1)) - Y(d, M(0))|Z = z^*], \qquad d \in \{0, 1\} \tag{2.2}$$

$\delta(d)$ corresponds to the change in mean potential outcome for the population near the threshold when exogenously shifting the mediator to its potential values under treatment and non-treatment state but keeping the treatment fixed at $D = d$ to switch off the direct effect.
It can be easily shown that the ATE, even for the local population, is the sum of the natural direct and indirect effects defined upon opposite treatment states, like in the traditional mediation framework, but looking only at the individuals just above and below the cutoff point:

$$\begin{aligned}
\Delta &= E[(Y_1 - Y_0)|Z = z^*] \\
&= E[Y(1, M(1)) - Y(0, M(0))|Z = z^*] \\
&= E[Y(1, M(1)) - Y(0, M(1))|Z = z^*] + E[Y(0, M(1)) - Y(0, M(0))|Z = z^*] \\
&= [\theta(1) + \delta(0)|Z = z^*] \\
&= E[Y(1, M(0)) - Y(0, M(0))|Z = z^*] + E[Y(1, M(1)) - Y(1, M(0))|Z = z^*] \\
&= [\theta(0) + \delta(1)|Z = z^*]
\end{aligned} \tag{2.3}$$

[7] Robins and Greenland (1992) and Robins (2003) denominated these parameters as "pure" direct and indirect effects.

where the third equality comes from adding and subtracting the quantity $E[Y(0, M(1))]$ and the fifth equality comes from adding and subtracting the quantity $E[Y(1, M(0))]$. The main problem with this analysis is identifying the counterfactual quantities $E[Y(d,\ M(d'))]$, never observed for each individual and hardly identified in non-experimental designs with the classical assumptions. A second issue is that only one of $Y(1, M(1))$ and $Y(0, M(0))$ is observed for any unit, which is known in literature as the fundamental problem of causal inference (Holland, 1986).
Identification of direct and indirect effect hinges on exploiting exogenous variation in the treatment and the mediator, as follows in the next section.

2.2.2 Natural and controlled effects

Another parameter taken into account from the mediation literature is the controlled direct effect (CDE). Formally:

$$\gamma(m)^{\star} = E[Y(1, m) - Y(0, m)], \qquad \text{for } m \text{ in the support of } M \tag{2.4}$$

and it expresses how much the mean potential outcome would change if the mediator were fixed at a particular value m uniformly in the population but the treatment was exogeneously changed from 1 to 0. Usually it is easier identify this parameter because it is not necessary to know the potential value of the mediator and then the analyst needs less assumptions. At the same time, for the policy implications most of the time is useful to know the natural effects. Unfortunately, the CDE is equal to NDE only if there is no interaction effect between treatment and mediator to the outcome and, then, in the absence of heterogeneity.
But, as before, we menage a local direct effect, defined just for the population near the threshold:

$$\gamma(m) = E[Y(1, m) - Y(0, m)|Z = z^{*}], \qquad \text{for } m \text{ in the support of } M \tag{2.5}$$

The main implication of this local analysis is that, because we are in a sharp RD, meaning that M is deterministically determined by Z, everyone is complier in our

population of interest, implying that the CDE fixing the value of m at 0 is equal to the NDE fixing the potential value of M at the value it would have under treatment state equal 0, fomrally: $[\gamma(m{=}0)|Z{=}z^*] = [\theta(0)|Z{=}z^*]$. In fact, the direct effects reflect the difference in the outcomes between treated and non-treated groups, mantaining a fixed level of M. But, at the threshold, we know that the entire population under analysis is complier, so everyone who has $Z \leq z^*$ has $M{=}0$ and this permits to identify not only $Y(1,\, 0)$ but also $Y(1,\, M(0))$, because we know for everyone what would be the potential mediator defined on the opposite status of the treatment. If in reality the behavior is like in the threshold we can identify the natural effects because, in this setting, the CDE coincides with the NDE. It would no longer be true if we were in a Fuzzy RD context, because we would have the presence of Never Takers, Always takers and Defiers. In this case the CDE is no longer equal to the NDE, because the populations of referement will be different, and we couldn't know the potential value of the mediator simply looking at the value of Z.
In our framework the parameters of interest are:

$$NDE(d) = E[Y(1, M(d)) - Y(0, M(d))|Z = z^*] \quad \forall\, d \in \{0, 1\} \tag{2.6}$$

$$NIE(d) = E[Y(d, M(1)) - Y(d, M(0))|Z = z^*] \quad \forall\, d \in \{0, 1\} \tag{2.7}$$

In the next sections, we will discuss two different models.

2.3 Model 1

We consider a first general model in which a random binary treatment D affects the outcome Y and the forcing variable Z. This last one deterministically affects the mediator M, inducing a sharply change in the mediator state depending on the particular value of Z, and in turn it affects the outcome Y. In this model a causal effect between Z and Y is allowed, because to estimate the effects of interest we have to look just at the population near the threshold defined by $Z{=}\bar{z}$, controlling, then, for the direct effect of the forcing variable on the outcome. So, in this model, Z is a continuous, observed and endogenous variable.

The general model is given by:

$$Y = \phi(D, Z, M, u)$$
$$M = 1[Z \geq z^*]$$
$$Z = \xi(D, v)$$
$$D = \lambda(\epsilon)$$

where ϕ, ξ, λ are linear functions and u, v, ϵ are unobservable components. In the model's notation we did not include the set of covariates X for sake of simplicity. The general outcome equation is:

$$Y = \phi\{D(\epsilon), Z[D(\epsilon), v], M[Z(D(\epsilon), v)], u\}$$

2.3.1 Identifying assumptions of Model 1

For the first model, we assume that the forcing variable Z is function of the treatment D. To identify our parameters of interest, the first assumption we need is the classical conditional independence of the treatment, see for instance Imbens (2004):

ASSUMPTION 1. Conditional randomness of the treatment:

$$\{Y(d', m), M(d)\} \perp D | X = x, \qquad \forall\ d, d', m \in \{0, 1\}$$

By assumption 1 we state that there are no unobserved confounders between treatment and mediator and/or outcome conditioning on pre-treatment covariates X, implying the independence of the potential outcome and the potential mediator from D. With this assumption we can identify the direct effect from D to Y and the effect from D to M. In non-experimental data, the plausibility of this assumption depends on the richness of variables available. In experimental data, this assumption holds if the treatment is randomized within strata defined on X.[8]

[8] If treatment is randomized unconditionally, the stronger assumption $\{Y(d', m), M(d)\} \perp D$ holds as well.

ASSUMPTION 2. Continuity of the potential outcome at the threshold:

$$E\{Y(d', m)|Z, X\} \quad \text{is continuous in } Z = z^*$$

This assumption states that in the counterfactual quantities there is no discontinuity due to selection bias and that conditioning on Z and X, M is like randomized at the threshold, implying then the absence of unobserved confounders jointly affecting the mediator and the outcome, if we condition it on the set of pre-treatment covariates and on the value of the threshold, recalling the assumptions' RD literature. It means that near the threshold we can correctly identify the effect of M on Y. The difference with the classical mediation framework is that now we have to look at a local population. Looking at the threshold also permits to don't take into account the relation between Z and Y, implying an exlusion restriction for Z, such that $corr(Z, Y) = 0|Z = z^*$. In this way the indirect effect is not confounded for the local population. Assumption 2 is violated if unobserved pre-treatment confounders affect both M and Y directly, or if unobserved post-treatment variables influence M and Y and are not fully determined by X and/or D.

ASSUMPTION 3. Perfect compliance of the mediator:

$$P(M = 1|Z = z^*_+) = 1$$
$$P(M = 0|Z = z^*_-) = 1$$

This assumption comes from the Sharp RD design (SRDD) and it states that the mediator is deterministically and fully determined by the value of Z, meaning that the effect of D goes to M only through Z, implying an exclusion restriction w.r.t. the treatment. In particular, every observation with a score just above z^* will have M=1 and every observation with a score just below z^* will have M=0.[9] It is important to note that there is no error term, implying the presence only of Compliers in our population of interest. In other words, for this population, we know for sure what will be the value of the mediator, simply looking at the score of Z.

[9]This deterministic law can be written as: $M_i = 1[Z \geq z^*]$.

ASSUMPTION 4. Homogeneity effect:
We can add a parametric assumption of effect homogeneity:

$$\theta(0) = \theta(1) = \theta$$

$$\delta(0) = \delta(1) = \delta$$

This assumption states that direct and indirect effects do not vary as functions of treatment status. In other words, the direct and indirect effects are independent of each other and the direct effect is constant regardless the level of the mediator and the indirect effect is constant regardless the level of the treatment removing the possibility of non-linearity, see for instance VanderWeele (2015). This assumption is necessary for the identification strategy, as we explain later. Assumptions 1-4 imply:

ASSUMPTION 5. Conditional indipendence between treatment and forcing variable Z:

$$Z(d) \perp D|X = x, \qquad \forall\ d \in \{0, 1\}$$

This assumption holds thanks to assumption 1 and 4. Without the parametric assumption this would no longer be true. This assumption states that if there are no confounders between treatment and mediator, then there are no confounders between treatment and Z if at the threshold the entire effect of D goes to M through Z. This implies that we can correctly identify the homogeneous effect from D to Z.

2.3.2 Parametric identification of Model 1

By Assumptions 1-5, we can parametrically identify model 1. According to the literature, the parametric (linear) estimation implies the absence of interactions between the effect of D and M on Y and it imposes additivity between the observed and

unobserved terms, implying that the effects are constant across individual characteristics.[10] So, by Assumption 4, we can rewrite a parametric, but less flexible, model like:

$$\begin{cases} Y = \beta_0 + \beta_1 D + \beta_2 M + u \\ M = 1[Z \geq \bar{z}] \\ Z = \alpha_0 + \alpha_1 D + v \end{cases}$$

This system recalls the linear structural equation model (LSEM), see for instance Baron & Kenny (1984).
Assuming $\beta_0 = \alpha_0 = 0$ for sake of simplicity, and solving the system, we can rewrite the linear outcome equation as:

$$\begin{aligned} Y &= \beta_1 D + \beta_2(Z) + u \\ Y &= \beta_1 D + \beta_2[\alpha_1 D + v] + u \\ Y &= D(\beta_1 + \beta_2 \alpha_1) + [\beta_2 v + u] \end{aligned} \tag{2.8}$$

where, in (2.8), β_1 represents the direct effect of D on Y, $\beta_2\alpha_1$ is the indirect effect that goes from D to Y through Z and M if $Z \geq z^*$, otherwise the indirect effect will be null. It is worth to note that, in this setup, the total effect is given simply by summing the direct and indirect effect. In the classical policy analysis we observe only one coefficient for D that can be unbiased if correctly specified but it can't explain the "causes of the effect" but only the "effects of the causes".
We can identify these effects because by Assumption 1 the direct effect is unconfounded; always by Assumption 1 and 5 we can still correctly estimate the effect from D to Z; by Assumption 3 we know that M is deterministically determined

[10] The model can be augmented adding the interaction term between mediator and treatment. This, at least, allows for an heterogeneity in $\theta(d)$ and $\delta(d)$ w.r.t. d

by Z, implying unconfoundedness of M; even if the corr$(u,v)\neq 0$ it doesn't matter because we have to look at the threshold by Assumption 2; then, because we are in a LSEM we can estimate the indirect effect simply multiplying every single causal parameter. We can't obtain a non-parametric identification because without the homogeneity assumption we are not able to correctly estimate the effects, because of the correlation between D and v when we condition on a particular state of treatment and on the value of Z. Because of this correlation, Assumption 5 does not hold and the non-parametric identification is impossible to get.

2.3.3 Graphical interpretation of Model 1

The first Model can be represented like in figure 1: it illustrates the framework based on a direct acyclic graph, in which the arrows represent causal observable effects and the dashed arrows represent unobservable effects. We didn't take into account the set of covariates X for ease of exposition.

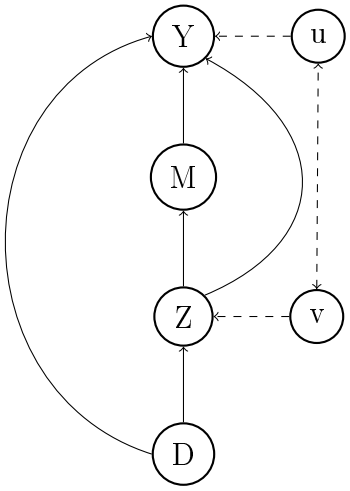


Figure 2.1: Model 1

This causal diagram satisfies Assumotions 1-5. In particular, it shows the effects that can be parametrically identified for the local population. An important point is that

actually Z is a post-confounder variable because it is caused by the treatment and in turn it causally affects both M and Y, but, because we are interested only on the local population, fixing $Z = \bar{z}$, we can correctly estimate the causal effect that goes from D to Y through Z and M, allowing also for a correlation between the error terms of the forcing variable and the outcome, respectively u and v. The limit of this setup is that fixing both the treatment and the forcing variable we can't nonparametrically identify the effects, because even if we have had the randomization of the treatment, we would have had selection bias for the estimation of the effects.

2.4 Model 2

We consider a second general model in which the treatment D affects the outcome and the mediator, but now the forcing variable Z[11] is not affected by treatment but still deterministically affects the mediator, like in figure 2.2. In particular, D is randomly assigned and Z is an exogenous variable, implying a zero correlation between the error terms of these two variables. So, the mediator is a function of the treatment and it sharply changes state depending on the value of Z. The general model is given by:

$$\begin{aligned} Y &= \phi(D, M, Z, u) \\ M &= \zeta(D, Z) \\ Z &= \xi(v) \\ D &= \lambda(\epsilon) \end{aligned}$$

where ϕ, ζ, ξ, λ are unknown functions and u, v, ϵ are unobservable components. We did not include in the model's notation the set of covariates X for ease of exposition, but the assumptions discussed later on are more plausible after conditioning on observable characteristics.
The general outcome equation is given by:

$$Y = \phi[D(\epsilon), Z(v), M(D(\epsilon), Z(v), u)]$$

[11] The forcing variable Z must be always a continuous and observable variable.

2.4.1 Identifying assumptions of Model 2

If we are in a framework in which Z is an exogenous forcing variable we need a different set of assumptions for the identification of the effects of interest. In particular, in addition to assumptions 1 and 2 we have:

ASSUMPTION 7. Conditional indipendence between treatment and forcing variable:

$$Z \perp D|X = x$$

meaning that now Z is not a function of the treatment and it is still orthogonal to the treatment conditional on X. But now, this assumption holds even without the homogeneity assumption required in Model 1[12], because we don't have correlation between v and ϵ. This means that we can identify the effects even if they are not constant across units, allowing for a more flexible design.
Assumptions 1 and 7 imply:

ASSUMPTION 8. Conditional randomness of the treatment at the threshold:

$$\{Y(d', m), M(d)\} \perp D|Z = z^*, X = x \qquad \forall\ d, d', m \in \{0, 1\}$$

This assumption is implied by Assumption 7. This one permits to have a non-parametric identification of the natural effects, because now we don't have correlation between treatment and the error term of Z and then we have the independence between treatment and the potential outcome at the threshold. Assumption 8 is weaker than assumption 1, because now the treatment can be randomized only at the threshold.

ASSUMPTION 9. Compliance of the mediator:

$$Pr(M = 1|Z^+, D = 1) = 1$$

[12]See Assumption 4

$$Pr(M = 0|Z^{-}, D = 1) = 1$$
$$Pr(M = 0|D = 0) = 1$$

In this model the mediator is a deterministic function of D and Z. In particular, in the treated group we can observe two different values of M depending on the cutoff point z^*. On the contrary, in the control group we observe just the mediator status equal zero. This implies that we cannot identify all the parameters of interest, but still we can identify some effects under analysis.

ASSUMPTION 10. Common support:
$0 < Pr(D = d|Z = \bar{z}, X = x) < 1, \qquad \forall\ d \in \{0, 1\}$ and x in the support of X
By assumption 9, the conditional probability to receive or not receive the treatment given Z and X is between 0 and 1, meaning that we can observe a particular value of Z and X both in the treated and non treated group. This assumption is stronger than the standard common support in policy evaluation.[13] By Bayes' theorem, Assumption 9 also implies that $0 < Pr(Z = \bar{z}|D = d, X = x) < 1$, meaning that conditional on X, the forcing variable must not be a deterministic function of the treatment , otherwise no comparable units would be available across treatment states.

2.4.2 Non-Parametric identification of Model 2

Now, Assumption 1 and Assumption 7 imply Assumption 8 and this allows for a local non-parametric, and then more flexible, identification of the natural effects. In particular, we can identify the counterfactual quantity $E[Y(d, M(d'))]$:

[13] $0 < Pr(D = d|X = x) < 1$, for each value of X there is a positive probability of being both treated and untreated.

$$
\begin{aligned}
&E[Y(d, M(d'))|Z = \bar{z}] = \\
&= \iint E[Y(d,m)|M(d') = m, Z = \bar{z}, X = x]\, dF_{M(d')|Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\
&= \iint E[Y(d,m)|M(d') = m, D = d', Z = \bar{z}, X = x]\, dF_{M(d')|Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\
&= \iint E[Y(d,m)|D = d', Z = \bar{z}, X = x]\, dF_{M(d')|Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\
&= \iint E[Y(d,m)|D = d, Z = \bar{z}, X = x]\, dF_{M(d')|D=d',Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\
&= \iint E[Y(d,m)|M = m, D = d, Z = \bar{z}, X = x]\, dF_{M|D=d',Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\
&= \iint E[Y|M = m, D = d, Z = \bar{z}, X = x]\, dF_{M|D=d',Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \qquad (2.9) \\
&= \iint E[Y|M = m, D = d, Z = \bar{z}, X = x] \cdot \frac{Pr(D = d'|M = m, Z = \bar{z}, X = x)}{Pr(D = d'|Z = \bar{z}, X = x)} \\
&\quad dF_{M|Z=\bar{z},X=x}(m)\, dF_{X|Z=\bar{z}}(x) \\
&= \int E[Y|M = m, D = d, Z = \bar{z}, X = x] \cdot \frac{Pr(D = d'|M = m, Z = \bar{z}, X = x)}{Pr(D = d'|Z = \bar{z}, X = x)}\, dF_{M=m,X=x|Z=\bar{z}}(m, x) \\
&= E\left[E\Big[Y|M = m, D = d, Z = \bar{z}, X = x\Big] \cdot \frac{Pr(D = d'|M = m, Z = \bar{z}, X = x)}{Pr(D = d'|Z = \bar{z}, X = x)} \middle| Z = \bar{z} \right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.10)
\end{aligned}
$$

The first equality follows from the law of iterated expectations and from replacing the outer expectations by integrals, the second from Assumption 1 and 7, the third from Assumption 2, the fourth from Assumtion 1 and 7 again, the fifth from Assumption 2, the sixth from Assumption 1, the seventh and the eighth equality follows from Bayes' theorem and the last one from the law of iterated expectations. (9) recalls the so called mediation formula for identifying the direct and indirect effect, see for instance Pearl (2001) and Imai, Keele and Yamamoto (2010), with the

difference that now we are looking only at the population near the threshold.
With weaker restrictions we can identify the observable quantity $E[Y(d, M(d))|Z = \bar{z}]$:

$$\begin{aligned} & E[Y(d, M(d))|Z = \bar{z}] = \\ =& E\Big[E[Y(d, M(d))|Z = \bar{z}, X = x]\Big|Z = \bar{z}\Big] \\ =& E\Big[E[Y|D = d, Z = \bar{z}, X = x]\Big|Z = \bar{z}\Big] \end{aligned} \tag{2.11}$$

where the first equality follows from the law of iterated expectation and the second from Assumption 1 and 7.
Therefore, $\theta(d)$ and $\delta(d)$ are identified by either subtracting (2.10) from equation (2.11) or vice versa, depending on whether d is one or zero. In particular, the average direct effect $\theta(d)$ is given by:

$$\begin{aligned} \theta(d) =& \iint \Big[E[Y|D = d, M = m, Z = \bar{z}, X = x] - E[Y|D = d', M = m, Z = \bar{z}, X = x]\Big] \\ & dF_{M|D=d,Z=\bar{z},X=x}(m)\, dF_{x|Z=\bar{z}}(x) \\ =& \iint \Big[E[Y|D = d, M = m, Z = \bar{z}, X = x] - E[Y|D = d', M = m, Z = \bar{z}, X = x]\Big] \\ & \cdot \frac{Pr(D = d|M = m, Z = \bar{z}, X = x)}{Pr(D = d|X = x, Z = \bar{z})}\, dF_{M=m,X=x|Z=\bar{z}}(m, x) \\ =& E\Bigg[\Big[E[Y|D = d, M = m, Z = \bar{z}, X = x] - E[Y|D = d', M = m, Z = \bar{z}, X = x]\Big] \\ & \cdot \frac{Pr(D = d|M = m, Z = \bar{z}, X = x)}{Pr(D = d|Z = \bar{z}, X = x)}\Bigg| Z = \bar{z}\Bigg] \end{aligned} \tag{2.12}$$

while the indirect effect $\delta(d)$ is given by:

$$\delta(d) = \iint E[Y|D=d, M=m, Z=\bar{z}, X=x]\cdot$$
$$\{\, dF_{M=m|D=d,Z=\bar{z},X=x}(m) - \, dF_{M=m|D=d',Z=\bar{z},X=x}(m)\}\, dF_{x|Z=\bar{z}}(x)$$

$$= E\Bigg[E\Big[Y|D=d, M=m, Z=\bar{z}, X=x\Big]\cdot$$
$$\left(\frac{Pr(D=d|M=m, Z=\bar{z}, X=x)}{Pr(D=d|Z=\bar{z}, X=x)} - \frac{Pr(D=d'|M=m, Z=\bar{z}, X=x)}{Pr(D=d'|Z=\bar{z}, X=x)}\right)\Bigg| Z=\bar{z}\Bigg] \tag{2.13}$$

Following the identification results and assuming the availability of an i.i.d. sample of size n, we can estimate the natural direct effect under control group $\theta(0)$ and the natural indirect effect under treated group $\delta(1)$ and the total effect given by the sum of the previous two effects[14]. In general, they can be estimated by various strategies. In literature, parametric methods have been commonly used, like in Pearl (2011) and VanderWeele (2009), but they have some drawbacks like a restrictive functional form and a difficult interpretability in case of nonlinearities. Most recent nonparametric estimation has been developed by Imai et al. (2010). These methods avoid the before mentioned shortcomings but they might be cumbersome in empirical application in case of high dimensionality of X or if M is continuous. In this case the estimation is based on a combination of them. In particular, we estimate the conditional mean of Y by local regression and by a weighting formula the density of M, once we conditioned on a particular window defined in Z, recalling the RD estimation strategy.

[14] By Assumption 9, it is not possible to identify $\theta(1)$ and $\delta(0)$

2.4.3 Graphical interpretation of Model 2

Model 2 can be represented by the following acyclical graph of causal relations between observed and unobserved variables, in which for sake of simplicity we neglected the set of covariates X:

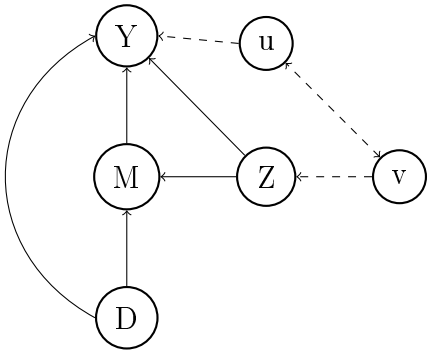


Figure 2.2: Model 2

The assumptions before discussed satisfy our structural system represented in figure 2. In particular, by assumption 1 and 7 we can correctly identify the effect of the treatment on the mediator and by assumption 8 we can control for the direct effect from Z to Y, recalling the exclusion restriction for the forcing variable. In other words, we can identify the natural effects, even in presence of a correlation between u and v. On the contrary, in this model, the correlation between v and the error term of the treatment is not allowed, as stated by assumption 7.

2.5 Estimation

Estimation based on the mediation formula (2.9) requires plug-in estimates for the conditional mean outcomes and the conditional mediator densities. In our model, the estimators of direct and indirect effects are given by:

$$\hat{\theta}(0) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\hat{\mu}_Y(1, M_i, Z_i, X_i) - \hat{\mu}_Y(0, M_i, Z_i, X_i)\right]\left(\frac{\hat{\rho}(m_i, x_i)}{1 - \hat{p}(x_i)}\right)\middle| Z = \bar{z}\right\} \quad (2.14)$$

$$\hat{\delta}(1) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}_Y(1, M_i, Z_i, X_i)\left(\frac{\hat{\rho}(m_i, x_i)}{\hat{p}(x_i)} - \frac{1 - \hat{\rho}(m_i, x_i)}{1 - \hat{p}(x_i)}\right)\middle| Z = \bar{z}\right\} \quad (2.15)$$

where we define the bandwidth by a local linear regression performed to either side of the cutpoint using the Imbens-Kalyanamaran optimal bandwidth calculation and $\hat{\rho}(m_i, x_i)$ and $1 - \hat{p}(x_i)$ denote the respective estimates of the propensity scores $Pr(D = 1|M = m_i, X = x_i)$ and $Pr(D = 1|X = x_i)$. Treatment propensity scores might be estimated by probit or logit specifications, see for instance Huber (2014) and Tchetgen Tchetgen (2013). The model can be better specified adding the interaction term between all variables' combinations in the conditional means outcome.

2.6 Simulation Study

This section presents a simulation study that provides some intuition for the identification result. We consider a data generating process (DGP) based on the following equations:

$$Y = 5 \cdot D + 3 \cdot M + 0.5 \cdot Z + \beta_1 \cdot DZ + \beta_2 \cdot ZM + 2 \cdot X + \epsilon_Y \quad (2.16)$$

$$M = I(D \cdot Z > 0) \quad (2.17)$$

$$D = I(0.5 \cdot x + 2 \cdot \epsilon_D > 0) \quad \text{with} \quad \epsilon_D, \epsilon_Y \sim N(0, 1) \quad i.i.d. \quad (2.18)$$

Equation (2.16) is the outcome equation, in which the observed Y is function of the observed variables D, M, Z, X and of the unobserved term ϵ_Y. β_1 and β_2 capture the

interaction between respectively D and Z and Z and M. Equation (2.17) describes the mediator behavior under Assumption 9. In the simulations, we set $\beta_1 = 1.3$ and $\beta_2 = 1.5$[15]. Table I provides the true direct and indirect effects.

Table I. True effects

$\theta(0)$	5
$\delta(1)$	1.5
Δ	6.5

We run two simulation studies with a bandwidth chosen using the Imbens-Kalyanaraman optimal bandwidth calculation[16]. In the first scenario we have 1000 simulations, whereas in the second scenario we have 2000 simulations.
Table II and Table III present the bias, variance (VAR) and root mean squared error (RMSE) of the estimators in the two scenarios. As tables show, augmenting the number of simulations $\hat{\theta}(0)$ performs much better, reaching zero bias. The behavior of $\hat{\delta}(1)$ is slower, but it respects the asymptotic characteristics. Moreover, in each simulation we applied the trimming rule (with trim=0.05) to discard observations with extreme propensity scores to improve overlap. The default is to discard observations with treatment propensity score smaller than 0.05 (5%) or larger than 0.95 (95%).

Table II

	Δ	$\theta(0)$	$\theta(0)$ trim	$\delta(1)$	$\delta(1)$ trim
Bias	-0.041	0,001	0,001	-0.043	-0.043
VAR	0,088	0,111	0,111	0,0871	0,0871
RMSE	0,096	0,111	0,111	0,095	0,095
nsim = 1000; nobs = 1000					

[15]We chose these values following other simulation studies in the mediation literature.
[16]See the RDestimate R package.

Table III

	Δ	$\theta(0)$	$\theta(0)$ trim	$\delta(1)$	$\delta(1)$ trim
Bias	0,024	0,000	0,000	0,024	0,024
VAR	0,090	0,063	0,063	0,030	0,030
RMSE	0,080	0,060	0,060	0,037	0,037
nsim = 2000; nobs = 1000					

2.7 Conclusions

One of the main problems in the evaluation of causal effects is the presence of causal chains between treatment and outcomes. As we show in the chapter, the estimation of these direct and indirect effects is a complex issue, due to the presence of an endogenous variable, the mediator, that lies in the causal pathway between treatment and outcome. In this chapter, we present a possible solution, demonstrating how to disentangle the total effect of a binary treatment using an RD approach. In particular, we developed two different models: (i) in the first one, the forcing variable Z is causally affected by the treatment and in turn deterministically affects the mediator and (ii) in the second model, Z is no more influenced by the treatment but still deterministically influences the mediator. We focused above all on this second model, because of its flexibility due to the exogeneity of Z. The results of this work suggest that the identification procedure is possible under hypothesis not too restrictive and often used in empirical studies. Secondly, we demonstrate the consistency of the estimator, validating the results through a Monte Carlo simulation study. Lastly, the estimation procedure is relatively easy to implement. In the following chapter, we present an empirical study using this new estimator, for analyzing the causal mechanisms.

Chapter 3

Causal Mediation Analysis in Economics: an empirical application

3.1 Introduction

The aim of this paper is to assess the causal effect of the EU Regional Policy with respect to the 2007-2013 programming period on the economic growth. More specifically, we investigate the role of research and development (R&D)expenditure as a causal channel of transmission of the policy. For this purpose, we use a statistical method called mediation analysis (see for instance Baron & Kenny, 1986; Pearl, 2001; Imai, Tingley and Yamamoto, 2015), that, by disentangling the total treatment effect into the direct effect and the indirect effect, permits to study the causal mechanisms through which a policy works, as deeply explained in chapters 1 and 2. To identify these effects and solve the endogeneity of the mediator that lies in the causal pathway between treatment and outcome[1], we use a spatial regression discontinuity design (SRDD), a quasi-experimental approach that exploits the geographical borders as discontinuity (see for instance Keele et al., 2015).
The EU Regional Policy, also known as Cohesion Policy, is a system of public transfers to subnational regions. This system is delivered through two main funds: the *Euro-*

[1]See chapter 2 for more details.

pean Regional Development Fund (ERDF) and the *Cohesion Fund* (CF). Together with the *European Social Fund* (ESF), the *European Agricultural Fund for Rural Development* (EAFRD) and the *European Maritime and Fisheries Fund* (EMFF), they make up the *European Structural and Investment Fund* (ESIF). The goal of EU Regional Policy is stated in Article 174 of the Treaty on the Functioning of the European Union (TFEU), that states: "the Union shall aim at reducing disparities between the levels of development of the various regions and the backwardness of the least favored regions or islands, and particular attention is to be paid to rural areas, areas affected by industrial transition, and region which suffer from severe and permanent natural or demographic handicaps".

The main areas of interest of these funds are: R&D, digital technologies, supporting the low-carbon economy, sustainable management of natural resources and small business. The promotion of R&D has a central role in the Cohesion Policy programs. As stated by the European Commission (2019)[2], sustainable growth is increasingly related to the capacity of regional economies to innovate and transform, adapting to a more competitive environment. For these reasons, Europe decided to put greater effort into the creation of the eco-system that encourage innovation, R&D and entrepreneurship (Europe 2020 strategy). The promotion of R&D is, therefore, a central feature in the Cohesion Policy for the 2007-2013 programming period, where about 86.4 billion or nearly 25% of the total allocation go toward R&D in a wider sense.

Given the stated objectives, Cohesion Policy has an important role in the European panorama. Since the 1990s, resources allocated for this Regional Policy have nearly doubled and nowadays, they represent one third of the total EU budget. More and more regions benefit from this policy and it has been a fundamental policy instrument for dampening the negative economic effects of the Great Recession (Cerqua et al., 2018). Despite its great importance, the EU Cohesion Policy's evolution has not occurred without difficulties and a final consideration about its effectiveness is not so clear. First of all, the EU is characterized by an asymmetric spatial distribution of the benefits coming from the process of economic integration (Fingleton et al., 2015). In addition to that, after the enlargement of the Union, the EU Regional Policy had

[2] `https://ec.europa.eu/regional_policy/en/policy/themes/research-innovation/`

to address sizable regional disparities, from the social, economic and cultural point of view. A second difficulty comes from the complexity of the internal administrative rules and procedures: the high institutional and managerial capacity needed for the implementation of the financed projects often causes delays. A third problem comes from the political scene: the departure from the EU of the UK, that is a net contributor to the EU budget, and the progress of the euro-skeptic parties pushed for a different vision of all EU policies (Crescenzi R., Giua M., 2019). These difficulties and this complicated panorama made the evaluation of this policy more difficult. The existing literature does not give a univocal answer on the overall impact of Cohesion Policy in less developed regions and no consensus has been reached (Barca, 2009). For example, De La Fuente and Vives (1995), Cappelen et al. (2003), Beugelsdijk and Eijffinger (2005), and Mohl and Hagen (2008) found a positive and significant impact of EU Regional Policy, while Fagerberg and Verspagen (1996), Boldrin and Canova (2001) and Dall'Erba and Le Gallo (2008) found not significant or even negative impact. An important contribution is given by counterfactual studies, that try to identify the net causal effect of the policy. They confirm the positive and significant impact of the regional policies, see for example Pellegrini et al. (2013), Cerqua and Pellegrini (2018) and Becker et al. (2013).

In this paper, we want to estimate not only the effect of the ERDF and CF on the economic growth, but also a possible causal mechanism through which this policy works, in particular investigating the role of R&D in this process. We propose a causal mediation framework to estimate these effects, using a SRDD for the identification of the quantities of interest. For this purpose, we use a regional dataset stems from the European Commission and the spatial grid defined by the EU27 regions at level 3 of the 2006 NUTS classification[3]. At the best of our knowledge this is the first paper using the mediation framework applied to the European dataset and it is the

[3]The "nomenclature of territorial units for statistics" (NUTS) was created by the European Office for Statistics (Eurostat) in order to apply a common statistical standard across the European Union. NUTS levels are geographical areas used to collect harmonized data in the EU. They have been used in the Structural Funds since 1988 and play an important role in allocating Structural Funds. The current nomenclature subdivides the Member States into three categories, according to specific population thresholds

first time that a SRDD is used together with causal mediation analysis. In particular, we implement the estimator developed in chapter 2. The analysis suggests that EU Cohesion Policy has a positive and significant impact on the economic growth, confirming the literature results. It is important to take into account the presence of the crisis in that period: the EU Regional Policy seems then to be an important tool to counteract the economic and financial crisis. In addition to that, we find also that investments in R&D are an important driver of the economic recovery: treated regions investing a large amount of funds in R&D experienced a better economic performance than treated regions that choose to invest in other priority themes.

3.2 EU regional policy

3.2.1 Programming period 2007-2013

The EU Cohesion Policy over the 2007-2013 programming period aimed to promote harmonious and sustainable development across the EU and to reduce socioeconomic disparities among regions. In fact, social and economic disparities are substantial in Europe, especially when we consider NUTS 3 regions. Just to give an overview, in 2007 the richest region was Inner London with 290% of the EU-27's average gross domestic product (GDP) per capita, while the poorest region was Nord-East in Romania with 23% of the EU average (EU Regional Policy, 2008).
In particular, the challenge of this programming period was to adapt the actions to the new necessities and to face changes in the job-market and globalization, expand research and innovation, create a more dynamic business environment, sustain a greener economy and combat climate change. For this purpose, the programming period 2007-2013 Cohesion policy made use of two financial instruments: ESIF and CF. ESIF is divided into the European Regional Development Fund (ERDF) and the European Social Fund (ESF). The first one, supports programs on regional development, economic change, enhanced competitiveness and territorial cooperation throughout the EU. The second one, provides support to anticipate and manage economic and social change, in particular increasing adaptability of workers and en-

terprises, enhancing access to employment and participation in the labor market, reinforcing social inclusion by combating discrimination and promoting reform in employment and inclusion. On the other hand, the CF is considered complementary to them (Malta's Strategy for Cohesion Policy 2007-2013). In fact, the objectives of CF are the same of ERDF: reduce regional disparities and strengthening economic and social cohesion. In particular, CF focuses on transport and environment infrastructures, as well as on energy efficiency and renewable energy in Member States with a Gross National Income (GNI) lower than 90 % of the EU average. For this reason, unlike in the 2000-2006 period, it is considered to be an explicit part of Cohesion Policy. Because of that, we follow several other studies on the overall effectiveness of the Cohesion Policy, and we do not distinguish between the supports given by different funds and they are all treated as Cohesion Policy. A total of €346.5 million of EU funds, corresponding to the 35% of the total budget, has been allocated for Cohesion Policy in the 2007-2013 programming period: the ERDF and Cohesion Fund accounted for €269.9 billion of it, corresponding to 78% of the total, and the ESF accounted for €76.67 billion.

Analyzing this period is also particularly interesting because it is the first full period in which the Central and Eastern European Countries are in receipt of Cohesion Policy funding. It is important to investigate the performance of the policy over this period, given the particular needs of the countries concerned to strengthen their endowment of infrastructure and to overcome other constraints on development, like the competitiveness of their firms and the relatively low expenditure on R&D. At the same time, it is important to evaluate the way in which the funding they received was invested and what the results were, as well as analyzing how well the policy was managed by the New-Accession countries given their administrative with large-scale funding.

After the enlargement of the EU, the role of Cohesion Policy becomes even more important to counteract the structural differences between regions: most of these disparities are long-term ones which have existed for several decades and which cannot be expected to be reduced quickly. The main areas concern economic performance, GDP per head and the rate of job creation, which are important determinants of

social disparities, differences in living conditions and the quality of life and the incidence of poverty and social exclusion (WP1: synthesis report, EU Commission, 2016).

In the period under analysis, Cohesion Policy had three different objectives to achieve its goals. The first one is the "Convergence" Objective (the ex Objective 1, with respect to the 2000-2006 period), that uses ERDF, ESF and CF. In EU-27 it concerns 17 Member States and 84 regions, with a per capita GDP less than 75% of the EU average. In addition to that, there is a transitional support for the phasing-out regions.[4] They are regions that were eligible for Objective 1 support during the period 2000-2006 but were no longer eligible for the period 2007-2013. The second one is the Competitiveness and employment Objective (the ex Objective 2, w.r.t. the 2000-2006 period). It covers all areas not eligible for the Convergence Objective, for a total of 19 Member States and 168 regions. Within these, 15 regions[5] are phasing-in areas: they receive a transitional support because they were covered by Objective 1 in 2000-2006, but had a GDP above 75% of the EU-27 average. The third Objective is the European Territorial Cooperation (ETC): it covers NUTS 3 regions on land-based internal borders and some regions on external borders as well as on maritime borders separated by a maximum distance of 150 km.

3.2.2 The Great Recession

The new Millennium is characterized by an increase in regional inequalities worldwide. Just to give an overview, the inequality in income per person among US metropolitan areas was 30% higher in 2016 than in 1980 (Ganong and Shoag, 2015). The so called "Great inversion" was triggered by a combination of technological

[4]The regions concerned are: Hainaut in Belgium; Brandenburg-Südwest, Lüneburg, Leipzig and Halle in Germany; Kentriki Makedonia, Dytiki Makedonia and Attiki in Greece; Principado de Asturias, Región de Murcia, Ciudad Autónoma de Ceuta and Ciudad Autónoma de Melilla in Spain; Basilicata in Italy; Burgenland in Austria; Algarve in Portugal; and Highlands and Islands in the UK.

[5]The regions concerned are Border, Midland and Western in Ireland; Sterea Ellada and Notio Aigaio in Greece; Castilla y León, Comunidad Valenciana and Canarias in Spain; Sardegna in Italy; Cyprus; Közép-Magyarország in Hungary; Região Autónoma da Madeira in Portugal; Itä-Suomi in Finland; and Merseyside and South Yorkshire in the UK.

change, globalization and policy choices, with deep consequences on the geographical economy of the World. In particular, some rural regions and medium metropolitan areas that were once quite prosperous have been characterized by job loss, declining labor-force participation and declining in per capita income, while the suburbs of rural areas are characterized by income stagnation. In the EU the situation is even more complex. There are social, political and economic inequalities between states and regions, within regions, between core areas and peripheral areas, between prosperous metropolitan regions and less-prosperous ones (Iammarino et al., 2018). In addition to that, the Great Recession and the related tight fiscal policies have generated an interruption in the historical trend towards decreasing inter-regional disparities (Crescenzi et al., 2016). In this situation, place-based policies played an important role. The idea of this public intervetion is to transfer large amounts of resources to underperforming areas and disadvantaged regions, providing them with infrastructure investment, incentives to increase labor market participation and subsidies to firms, trying to favor the establishment of new businesses and the growth of already existing ones to foster economic activity and tap into under-utilized resources in localities and regions (Pike et al. 2016). Some examples of place-based policy are enterprise zones and industrial cluster policies, but the most famous, extensive and long-lasting place-based policy worldwide is the EU Regional Policy. Even if place-based policies move billions of public resources, their impact on the economic performance is ambiguous. Some theories predict their ineffectiveness (see Glaeser and Gottlieb (2008); Dall'Erba and Fang, 2017) and other economists expressed less support for these programs, fearing they will generate large distortions in economic behavior (Busso et al., 2013). On the other hand, some contributions support the continuation of such policies, see for example Esposti & Bussoletti (2008). However, as expressed by Neumark and Simpson (2015), we need to know more about what features of these policies make them more effective or less effective, who gains and who loses from these policies, and how we can reconcile the existing findings, we need to know more about what policies create self-sustaining longer run gains. At this end, it is important to note that the impact of place-based policies has most of the time being estimated in periods of economic expansion, while there are rare

studies about these policies under recession periods. Several theoretical reasons show that the multipliers of place-based policies under times of crisis are different compared to periods of growth. Austin et al. (2018) stated that the presence of unused resources reduces the likelihood of tensions on the market for goods and labor prices, accentuating the real effects of regional policies. Following Filippetti et al. (2019), the consequences of the crisis on the social ground reinforces the effects of policies in places where, instead, perverse spirals of high-unemployment levels. Then, the presence of negative shocks common to neighboring areas diminishes the shooting effects of workers' mobility and therefore amplifies the positive effects of local policy interventions.

The programming period 2007-2013 is characterized by the most important global recession since the Second World War, designing a more complex situation in Europe, in which Cohesion Policy implemented. In this context, almost all countries experienced a fall in GDP with a consequence of a decline in the tax revenues, pushing up public expenditure on income support to counter the downturn in economic activity. The budget deficits and the large amount of government debt, led to tighter fiscal policies, diminishing public expenditure with much of the reduction being concentrated on public investment directly or on central government transfers to regional and local authorities which resulted in public investment being reduced indirectly. In this situation, Cohesion Policy became even more important. In 2008 and 2009, only one year after the beginning of the programming period, because of the Recession, GDP fell on average by 2% a year. In the following two years, 2009-2011, there was some recovery in output in the EU as a whole, averaging 2% a year, less than in the pre-recession period. Over the next four years, 2011-2015, the growth was on average less than 1% a year in the EU. In such a situation characterized by economic and financial crisis, Cohesion Policy represented a fundamental source of finance for development expenditure in many parts of the EU, especially for periphery regions. The goals for which the policy was initially designed has changed, shifting away from tackling long-term structural problems, like strengthening the development of economies to more short-term aims of counteracting the economic downturn, looking at more immediate and more direct effect on growth and jobs to stimulate economic

activity (WP 1: synthesis report, European Commission (2016)). In this particular context and given the absence of works in times of crisis, is interesting to investigate how this policy works during the Recession. The aim of this paper is to determine whether, and the extent to which, R&D expenditure had a role in counteracting the negative impact of the financial crisis in underdeveloped regions in order to know more about the processes of this regional program.

3.2.3 State of art on the evaluation of EU Cohesion Policy

There are many studies that try to assess what is the role of EU Cohesion Policy. Surprisingly, the findings are not homogeneous and until a few years ago the outcomes were ambiguous. Another limit that makes the evaluation of the effectiveness of Cohesion Policy sensitive, is the data availability and comparability at regional level. The main reason for this ambiguity is that the results are sensitive to the model specification and to the identification strategy. In addition to that, in this field is important to take into account the different regional level factors (Ederveen et al (2002); Rodriguez-Pose and Fratesi (2004); Percoco, (2005); Mohl and Hagen, (2010); Mancha-Navarro and Garrido-Yserte (2008); Crescenzi and Giua (2016); Crescenzi et al (2017)) that can make the identification of the policy effect more challenging. To solve this problem, the most recent studies are based on conterfactual approaches, that are able to identify the EU Cohesion Policy causal effect, regardless of the model specification. The idea is to study and try to investigate what would happen to the less developed areas of Europe in the absence of EU Cohesion Policy. The main characteristic of these approaches, based on quasi-experimental designs, consists in relaxing the structural restrictions needed in the classical models, allowing to isolate the causal effect without knowing the entire set of covariates that could affect the design. This kind of studies concludes that the EU Cohesion Policy has a positive impact on disadvantaged areas, in particular Ferrara et al. (2017) find a positive effect on innovation and transport infrastructure and Becker et al., (2010) and Pellegrini et al., (2013) estimate a positive effect on economic growth and employment. Other studies go deeper with the analysis, investigating the role of lo-

cal quality of government (Accetturo et al., 201)), the policy's expenditure intensity (Pellegrini et al., 2017), the regional contextual conditions (Bachtrögler et al., 2017) and the sectoral structure of the local economy (Percoco, 2017). These studies take advantage of the eligibility threshold, corresponding to 75% of the average EU GDP, for the assignment of the status of disadvantaged region, that permits to have access to a major part of the EU Cohesion Policy funds. The idea is to compare regions with level of GDP just above and just below the eligibility threshold. In this way, regions near the threshold are assumed to be similar in everything except for the fact that regions with a GDP just below the 75% of the EU average receive the policy and regions with a GDP just above the 75% of the EU average don't receive the policy, recalling the RDD literature. Following this approach, the difference in the outcome for this local population can be attributed only at the policy effect. This counterfactual approach deals with the problem of isolating the policy effect in a growth model, given the unknown functional form and unknown control variables: if "treated" regions are similar to "non-treated" regions around the threshold, we do not need any controls to consistently detect the growth effects of EU Regional Policy (Pellegrini et al., 2013). Other counterfactual studies estimate the impact of EU Cohesion Policy in one single EU Country: Mitze et al., (2012) looked at the effect of regional subsidies on labor-productivity growth in Germany, concluding that such policies are effective, but only up to a certain maximum treatment intensity; Bondonio and Greenbaum (2014) use firm-level data of the Northern Italian region to evaluate the impact of EU Regional Policies on the employment, finding that the effects of the programs are increasingly larger the higher the economic value of the incentives and that the most generous incentives come with a much higher cost per each additional new job; Di Cataldo (2017) studied the impact of EU funding in the U.K.'s most subsidized regions, providing evidence of a positive effect of EU Objective 1 funds on the regional labor market and economic performance; Barone et al (2016) looked at the case of Abruzzi (Italy) to study the long term effects of the EU regional policies, showing that the policies fail to move the treated regions towards a permanently higher GDP growth path; Giua (2017) has focused on the Italian Mezzogiorno estimating positive effects of the EU regional policies for the

regional employment.

Another interesting approach to solve the problem of the identification of the control group is to use a Spatial Regression Discontinuity Design (SRDD). It leverages the geographical distance to the physical boundary between eligible and non-eligible areas as a forcing variable for the identification of the policy impact (Crescenzi and Giua (2018))[6].

All these studies focus on the estimation of the causal effect of EU Cohesion Policy, answering to questions like: "Is this policy effective?", "How large is the impact?" or "Is it positive or negative?". But this kind of analysis does not answer to other important questions, like: "Why is this policy effective?" or "Which is the main driver of a positive outcome?". In other words, it is important to analyze and investigate which are the channels of transmission of this policy, what is the role played by some components of these funds to better design policy conclusions and policy actions. For this purpose, this paper uses the causal mediation analysis to disentangle the classical total treatment effect into two components: the indirect effect, that identifies a causal channel of transmission of the policy and a direct effect that captures all other possible explanations for why a treatment works. Following Model 2, explained in the second chapter, we use a SRDD to identify the counterfactual quantities and, then, estimate what is the effect of Regional Policy on the economic growth through R&D investments. Following the literature and EU studies, in fact, the rate of innovation is an important determinant of a region's economic growth and expenditure on R&D is a major way in which this can be stimulated, as recognized in Europe 2020. This approach takes advantage from the assumption that at the cut-off, that corresponds to the geographical boundaries, the Convergence Regions are similar in everything except for the fact that some regions invest a high quantity of funds in R&D and others not. Comparing these two types of regions we can estimate what is the effect of CP that affects the outcome through R&D investments, as explained in the following sections.

[6]This approach was born in other fields of policy evaluation, see for instance Holmes (1998), Black (1999), Gibbons et al (2013), Dell (2010), Menon and Giacomelli (2012), Einio and Overman (2012), Jofre-Monseny (2014), Papaioannu and Michalopoulos (2014), de Blasio and Poy (2017)

3.3 Data

This study is based on a new, reliable and comparable dataset, stemming from the European Commission. The spatial grid used in this work is defined by EU-27 regions al level NUTS 3 2006 classification[7]. In this study, we used data based on geographical expenditure work package (WP 13), which collected data from Managing Authorities (MAs) on expenditure and allocations in the different NUTS 3 regions within Member States by category of expenditure and broken down by the 86 priority themes (defined in Commission Regulation no. 1828/2006), estimating the data that were missing on the basis of the most relevant indicator available. The database covers the Convergence, Regional Competitiveness and Employment (RCE) as well as the European Territorial Cooperation (ETC) Objectives for the period 2007-2013, as shows Table 3.5. For the empirical analysis, we link these data with information on various sub-regional pre-treatment characteristics stemming from Cambridge Econometrics' Regional Database. In particular, we use at NUTS 3 level: per capita GDP growth rate between 2001 and 2006, the 2006 per capita GDP, the 2006 total employment level, the 2006 Gross Value Added (GVA), the 2006 GVA services sector, the 2006 total population, the 2006 ratio between the total employment and the active population, the 2006 population density and the 2006 per capita funds expenditure. As outcome we used the per capita GDP growth rate between 2006 and 2015. The idea is to use the per capita GDP as a summary indicator of development and prosperity of the regions. It is a good indicator of many key characteristics of the regions: economies at similar income levels often share many structural attributes, including education levels, science and technology endowments, infrastructure and institutional quality (Iammarino et al. (2019)). In our analysis, because of the interest in the R&D channel, we focus on the priority themes 01-09 corresponding to the Research and Technological development (R&D)[8], Innovation and entrepreneurship over the program period 2007-2013. The detail of the scomposition of R&D investment broken down by priority codes is reported in Table 3.7 of the Appendix. In particular, we

[7] We use the NUTS 2006 classification. Regions - Nomenclature of territorial units for statistics NUTS 2006/EU27.

[8] Commission Regulation no. 1828/20.

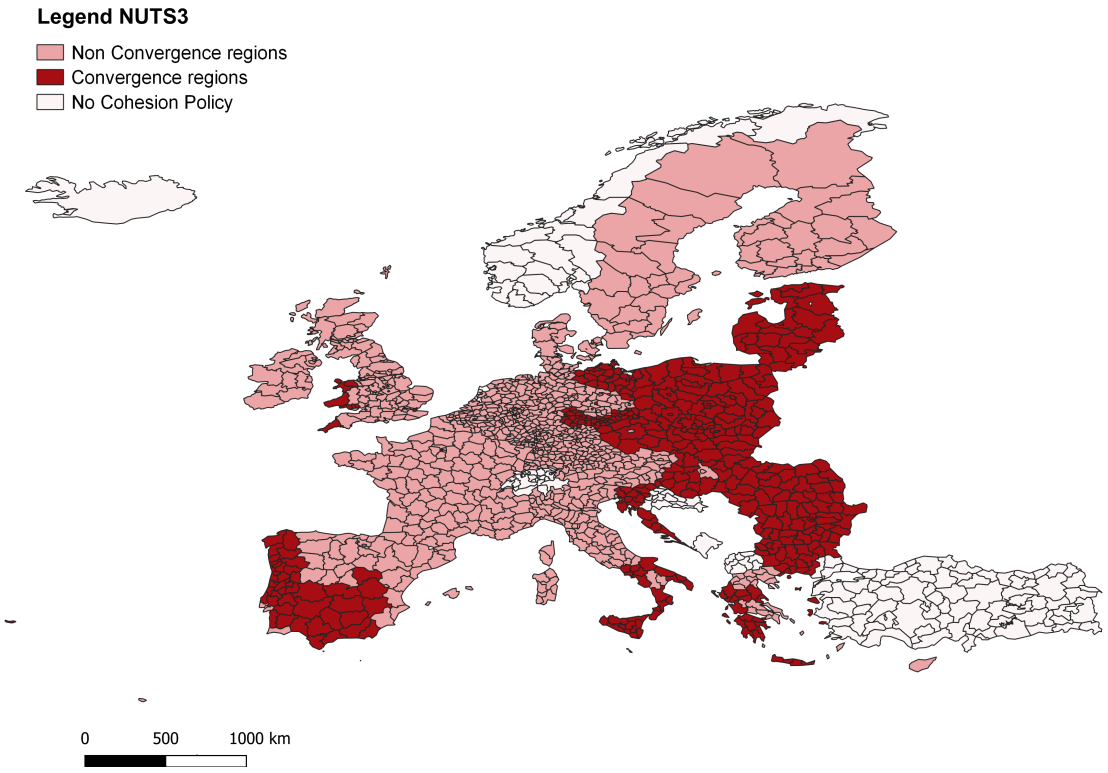
used as mediator a dummy variable that takes the value 1 for NUTS 3 regions investing al least 20%[9] of the total expenditure in R&D, and takes the value 0 otherwise, as showed in Table 3.6 of the Appendix.

3.3.1 Some descriptive statistics

In line with the RDD approach, we selected a restricted sample, which includes the closest regions to the 75% Convergence region discontinuity (see for example, Pellegrini et al., 2013). For this purpose, we exclude from the analysis all the regions with a per capita GDP greater than 150% of the average EU per capita GDP[10]. Because of the presence of a spatial analysis, we also exclude the islands of France, Guadeloupe, Martinique, Guyane and Reunion, the Canarias inslands of Spain and the islands of Portugal, Madeira and Azores.
Croatia entered in EU in 2013, was not eligible for the ERDF or Cohesion Fund in this period, but received pre-accession support for a total of €707 million. Because of the much smaller amount of support received than other countries with similar level of GDP per head and because of the absence of any data about pre-treatment variables, Croatia is excluded from the analysis.
Therefore, at the end of this process represented in figure 3.3 of the Appendix, we compile data on 1129 NUTS 3 regions: 385 are defined as treated regions (i.e. receiving Objective Convergence funds), while 744 as non-treated, as showed in Figure 1. This map presents the geographical position of treated and non-treated regions in the EU: the standard core-periphery picture is clearly outlined.

[9] It corresponds to the median of R&D expenditure distribution among treated NUTS3 regions.
[10] We don't exclude from the analysis regions with a per capita GDP lower than 50% of the average EU because in our analysis we are interested in the estimation of the indirect effect among treated regions. In other words, it is important to know what is the effect of R&D investments among poorest regions. Nevertheless, it is important as well to take into account this selection process in the interpretation of the results.

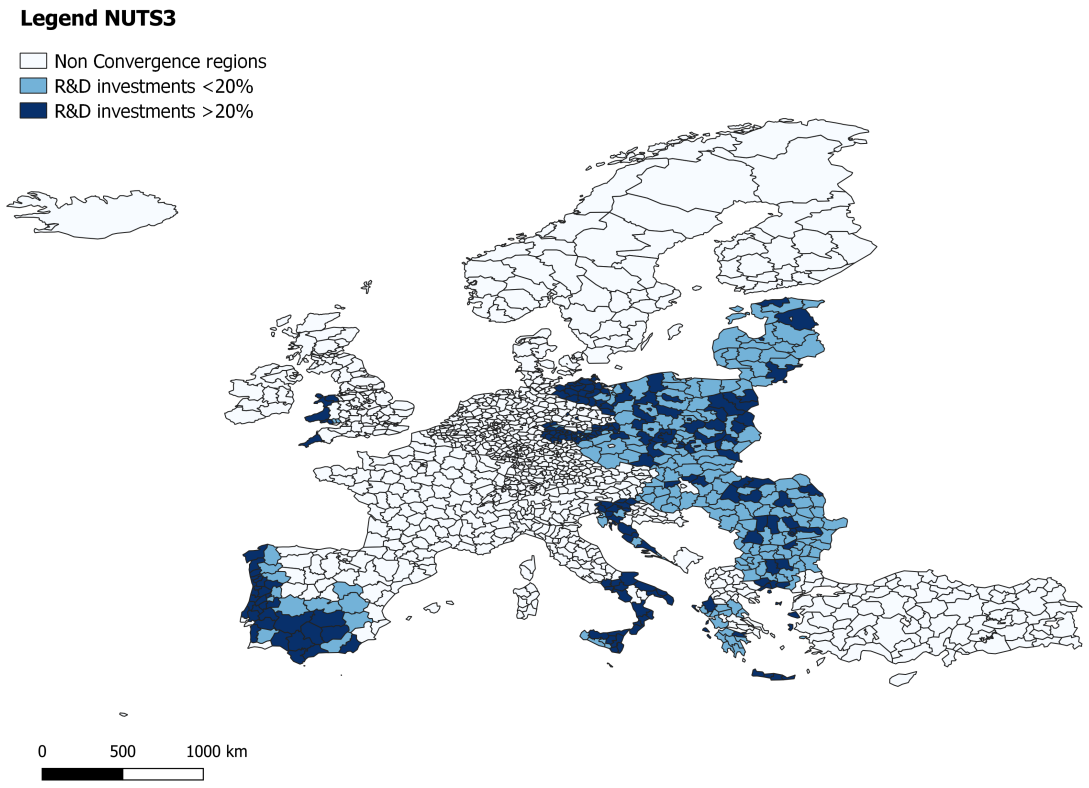
Figure 3.1: Geographical distribution of Convergence and Non-Convergence regions for the 2007-2013 Programming Period.



Among treated regions, we are interested to know which of them have invested a high share in R&D (at least 20% of received funds, corresponding to the median of the R&D investments distribution) and which of them have invested less than 20%. Figure 2 shows the geographical position of treated NUTS 3 regions with high and low intensity of R&D investments. The distribution of these regions results less clustered, with larger variance even within Member States.

.

Figure 3.2: Geographical distribution of R&D investments among Convergence regions



In Table 3.3.1 we compare treated and non-treated regions with respect to different pre-treatment characteristics. As suggested by the table, treated regions are generally more populated than the non-treated ones. Of course, they are richer and more productive. Still, the average per capita GDP growth is lower than that of the treated regions. The employment rate is equal between the two groups.
In Table 3.3.2, we summarize the main characteristics of treated regions divided by mediator status. Regions that invest more in R&D are generally richer, less populated and with a lower growth rate with respect to non-mediated regions. The behavior of mediated and non-mediated regions seems to replace the behavior of the entire

sample.

Table 3.1: Descriptive statistics (mean) of NUTS 3 regions by treatment status in 2006

	Restricted RDD sample (1129 NUTS 3)	
	Treated (385 NUTS 3)	Non-treated (744 NUTS 3)
GDP per capita	14259	23500
GDP per capita growth rate (2001-2006)	0.31	0.16
Total population	382381	348962
Active population	177153	172222
Total employment	153290	153634
Employment/active pop	0.85	0.86
Total GVA	3425929	8186512
GVA services sectors	2239910	5681702
Area (Km2)	4193	3392
Population density (inhabitants per km2)	225	404

Table 3.2: Descriptive statistics (mean) of NUTS 3 treated regions by mediator status in 2006

	Treated regions (385 NUTS 3)	
	Mediated (192 NUTS 3)	Non-mediated (193 NUTS 3)
GDP per capita	16793	11737
GDP per capita growth rate (2001-2006)	0.27	0.36
Total population	370422	394298
Active population	171656	182621
Total employment	151856	154716
Employment/active pop.	0.86	0.85
Total GVA	4424014	2433015
GVA services sectors	3067488	1416620
Area (Km2)	3039	5341
Population density (inhabitants per km2)	328	123

3.4 Econometric approach

Traditional methods identify the causal effects without explaining the link between policy interventions and the outcomes of interest. They cannot detect whether results are driven by some particular policy's components, leaving the interpretation of causal effect as a "black box". The aim of our empirical strategy is to estimate the causal effect of EU 2007-2013 Cohesion Policy on the 2006-2015 per capita GDP growth rate at NUTS 3 level, estimating what part of this effect is due to funds invested in R&D. For this purpose, we use a new estimator that is able to capture the

causal direct and indirect effect of a treatment, as explained in detail in chapter 2. In particular, $\hat{\theta}(0)$ is the direct effect and it estimates the effect of EU Regional Policy on the per capita GDP growth rate, controlling for the share invested in R&D. In other words, it captures all the mechanisms, not explicitly identified, as in the traditional causal analysis, through which the policy works, without the contribution given by R&D. On the other hand, $\hat{\delta}(1)$ is the indirect effect and it estimates the single effect of Regional Policy on the outcome brought by the R&D investments. As showed in chapter 2, the sum of this two effects gives the totale effect, called Δ, that is the traditional parameter, known in literature as average tretatment effect (ATE). In addition to that, we look at the spatial distribution of EU Cohesion Policy (see for example Giua, (2017)) and, in order to identify the effects, we use a spatial RDD: thanks to the geo-referenced data, we exploit the geographical discontinuities in funds to identify the quantities of interest, in particular comparing neighboring regions with high and low intensity of R&D among treated regions. In particular, we use as forcing variable the coordinates of the centroids of NUTS 3 regions (Eurostat). The idea behind the spatial RDD is to interpret the distance to the regional border as an assignment variable that decides about intensity in R&D investments: in other words, location acts as the forcing variable allowing us to exploit the discontinuities change in R&D intensity at the geographical border.
To do that, we use the following estimators:

$$\hat{\theta}(0) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\hat{\mu}_Y(D_i, M_i, Z_i, X_i) - \hat{\mu}_Y(1-D_i), M_i, Z_i, X_i)\right]\left(\frac{1-\hat{\rho}(m_i, x_i)}{1-\hat{p}(x_i)}\right)\middle| Z = \bar{z}\right\} \tag{3.1}$$

$$\hat{\delta}(1) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}_Y(D_i, M_i, Z_i, X_i)\left(\frac{\hat{\rho}(m_i, x_i)}{\hat{p}(x_i)} - \frac{1-\hat{\rho}(m_i, x_i)}{1-\hat{p}(x_i)}\right)\middle| Z = \bar{z}\right\} \tag{3.2}$$

where the outcome variable Y is the GDP per capita 2006-2015 growth rate for the NUTS 3 region i, D is the binary indicator variable for treatment which is unit in case of Convergence regions during the 2007-2013 programming period and zero otherwise, M is the binary indicator variable which is 1 if regions invest at least 20%

of EU Cohesion Fund in R&D[11], X is a set of pre-treatment variables to control for differences in treated and non-treated regions. We use Z as forcing variable, specifying the function as the two-dimensional RDD latitude-longitude space proposed by Dell (2010). In particular, we use the RDD polynomial, which controls for smooth functions of geographic location. We employ a 2^{nd} order polynomial which allows comparison of units which are very close to each other and absorbs all smooth variations in the outcome. The key identification assumption behind the SRDD strategy is that the potential outcomes are independent of treatment assignment[12] for regions that are close to the boundary that separates regions with high and low intensity of R&D investments, conditional on pre-treatment characteristics. With this approach we are able to estimate the direct effect net of R&D investments, and the indirect effect, i.e. the effect that goes from the Policy to the per capita GDP growth rate through investments in R&D, among treated regions. The main implication of this approach is that we can say something more about a policy, investigating why or why not is working, allowing to study not only the effect of a cause, but also the causes of the effect.

3.5 Empirical results

In table 3.3, we reported the empirical results of the previously mentioned identification strategy. Looking at the whole EU-results, we estimate a positive and significant average effect of the policy on the 2006-2015 GDP per capita growth rate. This result confirms the fundamental role of EU Regional Policy as instrument to counteract the crisis. Regions receiving a high share of Structural funds seem to slowly converge towards more developed regions. In addition, our estimates suggest that among treated regions, the ones investing more in R&D have better economic performance than the ones that don't invest in this priority theme. These statistical results show that structural funds have a positive impact on the economic growth and a little part of this impact goes through R&D channel. It is interesting to note that without decompos-

[11] Corresponding to the Priority code 01-09

[12] In our case mediator assignment.

ing the total average effect, the only conclusion is that the EU Cohesion Policy has a positive impact on the economic growth, without knowing what could be the policy implication to improve the policy direction and optimize the choice of investments. Table 3.4 shows results for a simple spatial RDD regression: it confirms the positive coefficient of the treatment, but this policy conclusion remains a black box, without the possibility to know which are the mechanisms that drive the process towards a faster economic recovery.

Table 3.3: Results

Δ	$\theta(0)$	$\delta(1)$
0.094***	0.079**	0.015
(0.028)	(0.030)	(0.033)

Table 3.4: Spatial Regression Discontinuity Design by regression

Objective 1	0.086***
	(0.021)
R-squared	0.502
Polynomial degree	2
Degrees of freedom	1113

Notes: Robust standard errors: * significant at 10%; ** significant at 5% level; *** significant at 1% level.

3.6 Conclusions

In this paper we evaluate the effect of the European Regional Policy on the economic growth using a causal model based on the mediation framework together with a quasi-experimental design, the SRDD. This estimation procedure permits to disentangle the total treatment effect into two components: the indirect effect that consists in the effect that goes from the Policy to the outcome only through investments in R&D; and the direct effect of the Policy, in which lie all other possible explanations for why the Policy works. Our findings show a positive and significant impact of Cohesion Policy on economic growth for the 2006-2015 period. In particular, the average treatment effect is 8.6% larger in regions which received a larger amount of Structural Funds. Analyzing the composition of this effect, it is possible to see that R&D investments have a positive impact, even if it is not statistically significant. Convergence Regions investing at least 20% of their funds in R&D grow more than the other similar regions (which invested more in other priorities), but without a strong evidence. Probably, this is due to the fact that this kind of investments have an impact on the economic performance in the middle-long period. It is important to note that the context in which we run the analysis is characterized by the economic and financial crisis that followed the Great Recession. The difficulty of this panorama confirms the importance of place-based policies as tools to counter the economic downturn and the fundamental role of R&D in the recovery process.

3.7 Appendix A. Sensitivity check

We check the sensitivity of the results and summarize the outcomes of interest in the next table. In particular, we replicate the analysis changing the RDD thredhold and then the number of observations that are mediated and not. The following table shows the estimates of the total effect, the direct and the indirect effect. Each block shows the outcomes for different number of units, in particular augmenting (diminishing) the RDD threshold up to 60° (40°) percentile. The results confirm what we obtained in the main analysis, in which the RDD threshold corresponds to the distribution's median of R&D expenditure[13].

Table 3.5: Sensitivity check

	Δ	$\theta(0)$	$\delta(1)$
Main analysis	0.094***	0.079**	0.015
(RDD threshold: median)	(0.028)	(0.030)	(0.03)
RDD threshold:	0.094**	0.081**	0.012
55° percentile	(0.035)	(0.03)	(0.036)
RDD threshold:	0.09**	0.074*	0.016
57.5° percentile	(0.035)	(0.032)	(0.038)
RDD threshold:	0.094**	0.075*	0.019
60° percentile	(0.036)	(0.022)	(0.039)
RDD threshold:	0.095***	0.071*	0.024
45° percentile	(0.027)	(0.03)	(0.031)
RDD threshold:	0.092***	0.074*	0.018
42.5° percentile	(0.026)	(0.029)	(0.029)
RDD threshold:	0.091***	0.08**	0.008
40° percentile	(0.025)	(0.028)	(0.027)

[13]See par. 3.5

3.8 Appendix B

Table 3.7: Division of ERDF and CF for 2007-2013 period between Member States by Objective (EUR million).

Country	**Obj_1**	**Obj_2**	**Obj_3**	**Obj_4**	**Total**
AT	96	371	0	0	467
BE	318	398	0	0	716
BG	4391	0	0	0	4391
CB	0	0	4314	0	4314
CY	0	0	0	368	368
CZ	15075	186	0	1140	16401
DE	9436	3655	0	0	13091
DK	0	216	0	0	216
EE	2572	0	0	0	2572
ES	15075	4352	0	2956	22382
FI	0	866	0	0	866
FR	1334	3771	0	28	5133
GR	11851	0	0	1905	13755
HR	220	0	0	0	220
HU	16325	1486	0	315	18125
IE	0	289	0	0	289
IT	12151	2392	0	0	14543
LT	4928	0	0	0	4928
LU	0	21	0	0	21
LV	3409	0	0	0	3409
MT	729	0	0	0	729
NL	0	741	0	0	741
PL	44331	0	0	0	44331
PT	7885	515	0	3624	12023
RO	7469	0	0	0	7469

Table 3.7 continued from previous page

Country	Obj_1	Obj_2	Obj_3	Obj_4	Total
SE	0	814	0	0	814
SI	2608	0	0	0	2608
SK	5672	60	0	759	6491
UK	1341	2919	0	0	4261

Notes: the Objectives 1-4 correspond respectively to: Convergence, Competitiveness, Cooperation and Multi-Objectives.

Table 3.8: Country share R&D expenditure

Country	R&D share	Country	R&D share
AT	0.800	IE	0.468
BE	0.607	IT	0.345
BG	0.140	LT	0.168
CB	0.170	LU	0.532
CY	0.221	LV	0.204
CZ	0.192	MT	0.114
DE	0.502	NL	0.455
DK	0.836	PL	0.198
EE	0.229	PT	0.318
ES	0.253	RO	0.156
FI	0.608	SE	0.659
FR	0.370	SI	0.287
GR	0.224	SK	0.121
HR	0.228	UK	0.576
HU	0.167		

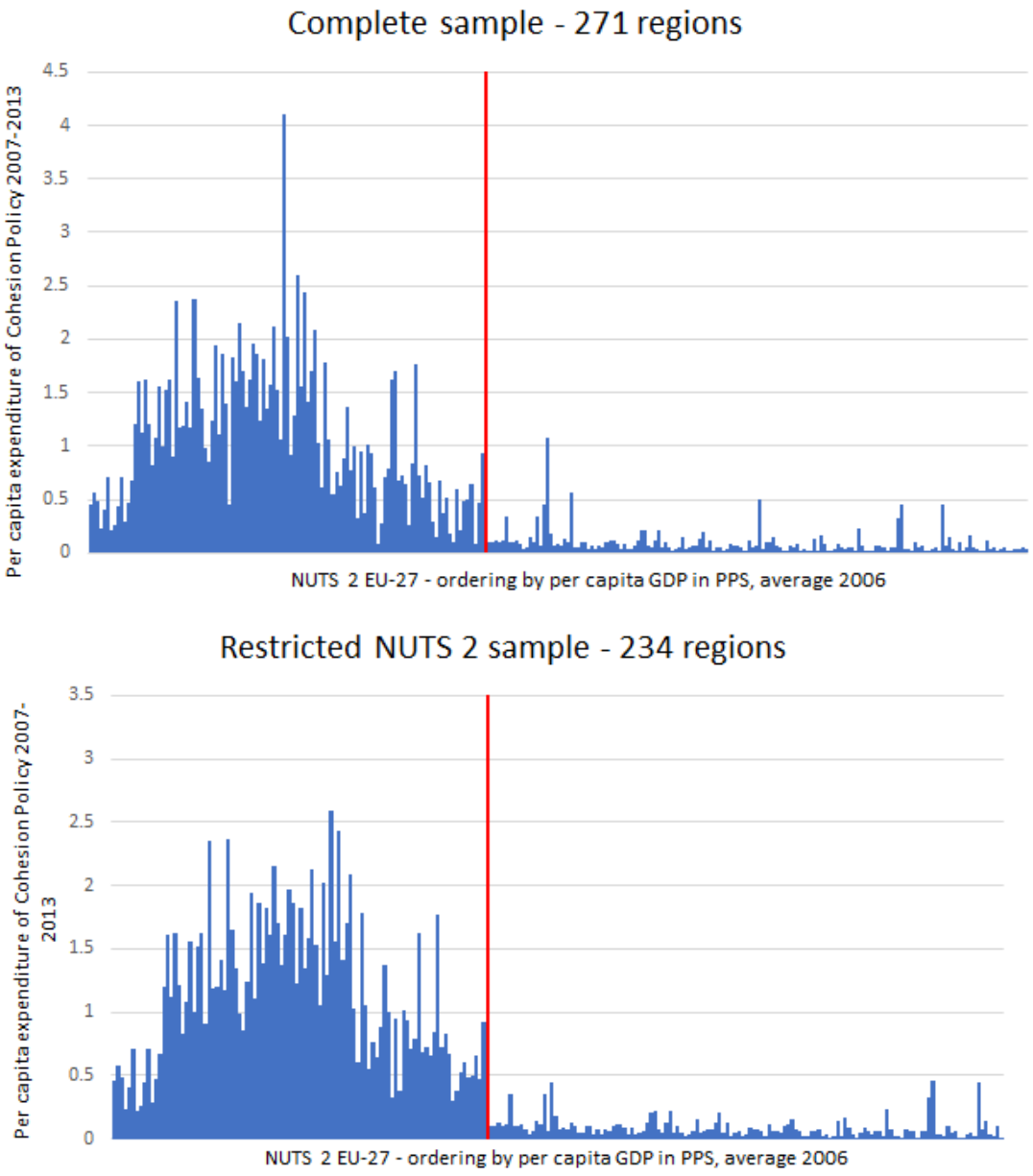
Notes: The table shows the R&D share invested in % with respect to the total amount of the expenditure at NUTS 0 level.

Table 3.9: Country scomposition of R&D share expenditure by priority codes.

Country	code_01	code_02	code_03	code_04	code_05	code_06	code_07	code_08	code_09
AT	0.061	0.072	0.044	0.068	0.051	0.023	0.136	0.320	0.025
BE	0.045	0.095	0.020	0.017	0.069	0.005	0.001	0.271	0.083
BG	0.000	0.003	0.001	0.014	0.016	0.000	0.005	0.081	0.021
CB	0.023	0.019	0.059	0.005	0.016	0.008	0.002	0.004	0.034
CY	0.031	0.028	0.001	0.011	0.005	0.000	0.000	0.143	0.001
CZ	0.019	0.067	0.005	0.011	0.013	0.008	0.027	0.037	0.006
DE	0.029	0.108	0.040	0.052	0.015	0.003	0.020	0.201	0.036
DK	0.042	0.016	0.282	0.067	0.140	0.079	0.035	0.094	0.081
EE	0.021	0.096	0.012	0.004	0.000	0.000	0.000	0.060	0.033
ES	0.032	0.042	0.005	0.050	0.002	0.002	0.022	0.083	0.014
FI	0.091	0.061	0.095	0.036	0.056	0.024	0.011	0.143	0.090
FR	0.055	0.085	0.037	0.032	0.041	0.017	0.014	0.048	0.041
GR	0.000	0.001	0.002	0.001	0.071	0.002	0.006	0.096	0.044
HR	0.045	0.064	0.005	0.000	0.031	0.020	0.000	0.008	0.056
HU	0.002	0.002	0.003	0.013	0.005	0.001	0.009	0.113	0.019
IE	0.176	0.131	0.000	0.000	0.000	0.000	0.000	0.161	0.000
IT	0.054	0.029	0.004	0.021	0.036	0.020	0.069	0.022	0.091
LT	0.000	0.047	0.002	0.008	0.011	0.000	0.009	0.027	0.065
LU	0.099	0.235	0.145	0.000	0.051	0.000	0.000	0.000	0.003
LV	0.042	0.000	0.012	0.000	0.005	0.000	0.049	0.023	0.073
MT	0.000	0.012	0.002	0.004	0.024	0.002	0.005	0.007	0.014
NL	0.043	0.039	0.101	0.037	0.019	0.020	0.061	0.023	0.114
PL	0.016	2.328	0.009	0.008	0.020	0.001	0.060	0.034	0.010
PT	0.017	0.005	0.024	0.015	0.023	0.000	0.150	0.053	0.020
RO	0.005	0.021	0.000	0.002	0.000	0.004	0.008	0.100	0.004
SE	0.036	0.004	0.104	0.035	0.147	0.026	0.005	0.114	0.159
SI	0.011	0.066	0.096	0.014	0.000	0.032	0.026	0.026	0.062
SK	0.025	0.090	0.020	0.000	0.004	0.003	0.004	0.022	0.007
UK	0.044	0.043	0.031	0.049	0.065	0.039	0.009	0.190	0.084

Notes: in the columns is showed the R&D composition and the corresponding share invested by regions in each priority code. Priority codes 01-09 correspond respectively to: R&TD activities in research centres, R&TD infrastructure and centres of competence in a specific technology, Technology transfer and improvement of cooperation networks, Assistance to R&TD, particularly in SMEs (including access to R&TD services in research centres), Advanced support services for firms and groups of firms, Assistance to SMEs for the promotion of environmentally-friendly products and production processes, Investment in firms directly linked to research and innovation, Other investment in firms, Other measures to stimulate research and innovation and entrepreneurship in SMEs.

Figure 3.3:


Complete sample - 271 regions
Per capita expenditure of Cohesion Policy 2007-2013
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
NUTS 2 EU-27 - ordering by per capita GDP in PPS, average 2006
Restricted NUTS 2 sample - 234 regions
Per capita expenditure of Cohesion Policy 2007-2013
3.5
3
2.5
2
1.5
1
0.5
0
NUTS 2 EU-27 - ordering by per capita GDP in PPS, average 2006

Bibliography

[1] Abadie, A., A. Diamond, J. Hainmueller (2010): "Synthetic control methods for comparative case studies: estimating the effect of california's tobacco control program", *Journal of the american statistical association*, 105(490), 493-505.

[2] Accetturo, A., G. de Blasio, and L. Ricci (2014): "A Tale of an Unwanted Outcome: Transfers and Local Endowments of Trust and Cooperation", *Journal of Economic Behavior & Organization*, 102, 74–89.

[3] Albert, J. M., S. Nelson (2011): "Generalized causal mediation analysis", *Biometrics*, 67, 1028-1038.

[4] Angelucci, M., V. Di Maro (2015): "Program evaluation and spillover effects", *Working paper, University of Michigan.*

[5] Angrist, J., G. Imbens, D. Rubin (1996): "Identification of causal effects using Instrumental Variables", *Journal of American statistical association*, 91, 444-472.

[6] Angrist, J., Lavy V., Schlosser A. (2010): "Multiple experiments for the causal link between the quantity and quality of children", *Journal of labor economics*, 28(4), 773-824.

[7] Austin, B.A., E.L. Glaeser, L.H. Summers (2018): "Jobs for the heartland: Place-based policies in 21st century America", NBER Working Paper No. 24548.

[8] Bachtrögler, J., U. Fratesi, G. Perucca (forthcoming): "The Influence of the local context on the implementation and impact of EU Cohesion Policy", *Regional Studies*, 2017.

[9] Barca, F. (2009): "An agenda for a reformed Cohesion Policy. Independent Report prepared at the request of Danuta Hubner, the Commissioner for Regional Policy," available at: `http://ec.europa.eu/regional_policy/archive/policy/future/pdf/report_barca_v0306.pdf` (accessed 05 November 2017).

[10] Baron R.M., Kenny D.A. (1986): "The moderator-mediator variable distiction in social psychological research: conceptual, strategic and statistical considerations", *Journal of personality and social psychology*, 51, 1173-1182.

[11] Becker, S. O., P. H. Egger, and M. von Ehrlich (2010): "Going NUTS: The Effect of EU Structural Funds on Regional Performance", *Journal of Public Economics*, 94(1–2), 578–590.

[12] Becker, S. O., P. H. Egger, and M. von Ehrlich (2013): "Absorptive capacity and the growth and investment effects of regional transfers: a regression discontinuity design with heterogeneous treatment effects", *American Economic Journal*, 5(4), 29–77.

[13] Beugelsdijk, M., S.C.W. Eijffinger (2005): "The effectiveness of structural policy in the European Union: an empirical analysis for the EU 15 in 1995–2001", *Journal of Common Market Studies*, 43, 37–51.

[14] Bijwaard, G. E., A. M. Jones (2018): "An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality", *Empirical economics*, pp. 1-47.

[15] Black, S. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education", *Quarterly Journal of Economics*, 114(2), 577–599.

[16] Black, S. E., P. J. Devereux, K. J. Salvanes (2005): "The more the merrier? The effect of family size and birth order on children's education", *The quarterly journal of economics*, 120(2), 669-700.

[17] Boldrin, M., F. Canova (2001): "Inequality and convergence in Europe's regions: Reconsidering European regional policies", *Economic Policy*, 0(32), 205–245.

[18] Brader, T., N. A. Valentino and E. Suhay (2008): "What triggers public opposition to immigration? Anxiey, group cues and immigration", *American Journal of Political Sciences*, 52(4), 959-978.

[19] Burgess, S., R. M. Daniel, A. S. Butterworth, S. G. Thompson (2015): "Network mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways", *International journal of epidemiology*, 44, 484-495.

[20] Busso, M., J. Gregory, P. Kline (2013): "Assessing the incidence and efficiency of a prominent place based policy", *American Economic Review*, 103, 897–947.

[21] Cappelen, A., F. Castellacci, J. Fagerberg, B. Verspagen (2003): "The impact of EU regional support on growth and convergence in the European Union", *Journal of Common Market Studies*, 41, 621–644.

[22] Cerqua, A. and G. Pellegrini (2018): "Local policy effects at a time of economic crisis", MPRA, *Working paper 85621*.

[23] Cerqua, A. and G. Pellegrini (2017): "Are we spending too much to grow? The case of Structural Funds", *Journal of Regional Science*, 1-36.

[24] Cerqua, A., G. Pellegrini (2014): "Do subsidies to private capital boost firm's growth? A multiple regression discontinuity design approach", *Journal of public economics*,109, 114-126.

[25] Chen, S. H.,Y. C. Chen, J. T. Liu (2017): "The impact of family composition on educational achievement", *Journal of Human Resources*, 54(1), 122-170.

[26] Cole, D. A., S. E. Maxwell (2003): "Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling", *Journal of abnormal pshychology*, 112, 558-577.

[27] Cox, G. W., J. N. Kats (1996): "Why did the incumbency advantage in U.S. House elections grow?", *American journal of political science*, 20(2), 478-497.

[28] Crescenzi, R., U. Fratesi and V. Monastiriotis (2017): "The EU cohesion policy and the factors conditioning success and failure: evidence from 15 regions," *Regions Magazine*, 305 (1). pp. 4-7. ISSN 1367-3882.

[29] Crescenzi, R., M. Giua (2019): "One or Many Cohesion Policies of the European Union? On the Diverging Impacts of Cohesion Policy across Member States", *Regional Studies*, 54, 10-20.

[30] Crescenzi, R., D. Luca, S. Milio (2016): "The geography of the economic crisis in Europe: National macroeconomic conditions, regional structural factors and short-term economic performance", *Cambridge Journal of Regions, Economy and Society*, 9, 13–32.

[31] Dall'Erba, S., F. Fang (2017): "Meta-analysis of the impact of European Union Structural Funds on regional growth", *Regional Studies*, 51(6), 822–832.

[32] Dall'erba, S., J. Le Gallo (2008): "Regional convergence and the impact of European Structural Funds over 1989–1999: A spatial econometric analysis", *Papers in Regional Science*, 87, 219–244.

[33] de Blasio, G. and S. Poy (2017): "The Impact of Local Wage Regulation on Employment: A Border Analysis from Italy in the 1950s", *Journal of Regional Science*, Vol. 57(1) 48-74.

[34] de la Fuente, A., X. Vives (1995): "Infrastructure and education as instruments of Regional Policy: Evidence from Spain", *Economic Policy*, 10, 11–51.

[35] ell, M. (2010): "The Persistent Effects of Peru's Mining Mita", *Econometrica*, 78, 1863–1903.

[36] Deuchert, E., M. Huber, M. Schelker (2019): "Direct and indirect effects based on difference-in-differences with an application to political preferences following the Vietnam draft lottery", *Journal of Business & Economic statistics*, 37(4), 710-720.

[37] Di Cataldo, M. (2017): "The long-term impact of Objective 1 funding on unemployment and labour market disparities", *Journal of Regional Science*, Vol. 57, Issue 5, 814-839.

[38] Ederveen, S., J. Gorter, R. de Mooij, R. Nahuis (2002): "Funds and Games: The economics of European Cohesion Policy", Special publication 41. CPB, The Hague.

[39] Einio, E. and H. Overman (2012): "The Effects of Spatially Targeted Enterprise Initiatives: Evidence from UK LEGI," ERSA Conference Papers, European Regional Science Association.

[40] Esposti, R., S. Bussoletti (2008): "Impact of Objective 1 Funds on Regional Growth Convergence in the European Union: A Panel-data Approach", *Papers in Regional Studies*, 42(2), 159-173.

[41] Fagerberg, J., B. Verspagen (1996): "Heading for divergence? Regional growth in Europe reconsidered", *Journal of Common Market Studies"*, 34, 431–438.

[42] Ferrara, A. R., P. McCann, G. Pellegrini, D. Stelder and F. Terribile (2016): "Assessing the impacts of Cohesion Policy on EU regions: A non-parametric analysis on interventions promoting research and innovation and transport accessibility", *Papers in Regional Science*, doi: 10.1111/pirs.12234.

[43] Filippetti, A., F. Guy, S. Iammarino (2019): "Regional disparities in the effect of training on employment", *Regional Studies*, 53(2), 217–230.

[44] Fingleton B., H. Garretsen, R. Martin (2015): "Shocking aspects of monetary union: the vulnerability of regions in Euroland", *Journal of Economic Geography*, 15(5), 907-934.

[45] Flore C. A., A. Flores-Lagunes (2009): "Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness", *IZA DP No. 4237.*

[46] Frangakis, C., D. Rubin (2002): "Principal stratification in causal inference", *Biometrics*, 58, 21-29.

[47] Frölich, M., M. Huber (2017): "Direct and indirect treatment effect - causal chains and mediation analysis with instrumental variables", *Journal of the royal statistical society: series B*, 79(5), 1645-1666.

[48] Ganong, P., D. Shoag (2015): "Why has regional income convergence in the US declined?" Harvard Kennedy School Working Paper, Cambridge.

[49] Gardner, J., A. J. Oswald (2007): "Money and mental wellbeing: a longitudinal study of medium-sized lottery wins", *Journal of health economics*, 26(1), 49-60.

[50] Gelman, A., G.W. Imbens (2013): "Why ask Why? Forward causal inference and reverse causal questions", *NBER Working paper No. 19614.*

[51] Gelman, A., G. King (1990): "Estimating incumbency advantage without bias", *American Journal of Political Science*, 34(4), 1142-64.

[52] Gibbons, S., S. Machin and O. Silva (2013): "Valuing School Quality Using Boundary Discontinuity Regressions", *Journal of Urban Economics*, 75, 15–28.

[53] Giua, M. (2017): "Spatial discontinuity for the impact assessment of the EU Regional policy: the case of the Italian Objective 1 regions", *Journal of Regional Science*, Vol. 57(1), 109-131.

[54] Glaeser, E., J.D. Gottlieb (2008): "The economics of place-making policies", *Brookings Papers on Economic Activity*, 155–239.

[55] Glynn, A. N. (2012): "The product and difference fallacies for indirect effects", *Journal of political science*, 56, 257-269.

[56] Havelmo, T. (1943): "The statistical implications of a system of simultaneous equations", *Econometrica*, 11, 1-12.

[57] Heckman, J., R. Pinto, P. Savelyev (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes", *American economic review*, 103, 2052-2086.

[58] Holland, P. W. (1986): "Statistics and causal inference", *Journal of the american statistical association*, 81, 945-60.

[59] Holmes, T. (1998): "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders", *Journal of Political Economy*, 106(4), 667–705.

[60] Hong M. (2010): "Ratio of mediator probability weighting for estimating natural direct and indirect effects", in *Proceedings of the American statistical association, Biometrics section*, p. 2401-2415. Alexandria, VA: American Statistical Association.

[61] Hong M. (2012): "Editorial comments", *Journal of educational effectiveness*, 5, 213-214.

[62] Huber M. (2014): "Identifying causal mechanisms (primarly) based on inverse probability weighting", *Journal of Applied Econometrics*, 29, 920-943.

[63] Huber, M. (2015): "Causal pitfalls in the decomposition of wage gap", *Journal of business and economic statistics*, 33, 179-191.

[64] Huber, M. (2019): "A review of causal mediation analysis for assessing direct and indirect treatment effects", *Working paper No. 500*.

[65] Huber, M., M. Lechner, G. Mellace (2017): "Why do tougher caseworkers increase employment? The role of program assignment as a causal mechanism", *the Review of economics and statistics*, 99, 180-183.

[66] Iammarino, S., A. Rodríguez-Pose, M. Storper (2019): "Regional inequality in Europe: Evidence, theory and policy implications", *Journal of Economic Geography*, 19(2), 273–298.

[67] Imai, K., L. Keele, D. Tingley, T. Yamamoto (2011): "Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies", *American political science review*, 105(4), 765-789.

[68] Imai, K., L. Keele, T. Yamamoto (2010): "Identification, inference and sensitivity analysis for causal mediation effects", *Statistical Science*, 25(1), 51-71.

[69] Imai, K., D. Tingley, T. Yamamoto (2013): "Experimental designs for identifying causal mechanisms", *Journal of the Royal statistical society, Series A*, 176(1), 5-51.

[70] Imbens, G. W. (2004): "Nonparametric estimation of average treatment effect under exogeneity: a review", *The review of economics and statistics*, 86(1), 4-29.

[71] Imbens, G. W., J. Angrist (1994): "Identification and estimation of local average treatment effects", *Econometrica*, 62, 467-475.

[72] Imbens G.W., Lemieux T. (2008): "Regression discontinuity designs: a guide to practice", *Journal of econometrics*, 142(2), 615-635.

[73] Imbens, G. W., J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation", *Journal of economic literature*, 47, 5-86.

[74] Jofre-Monseny, J. (2014): "The Effects of Unemployment Protection on Migration in Lagging Regions", *Journal of Urban Economics*, 83, 73–86.

[75] Judd C.M., Kenny D.A. (1981): "Process analysis: estimating mediation in treatment evaluations", *Evaluation review*, 5(5), 602-619.

[76] Keele L.J., R. Titiunik (2015): "Geographic Boundaries as Regression Discontinuities", *Political Analysis*, 23(1), 127-155.

[77] Lee, D.S. (2008): "Randomized experiments from non-random selection in U.S. House Election", *Journal of econometrics*, 142(2), 675-697.

[78] Lee, D.S., Lemieux T. (2010): "Regression discontinuity designs in economics", *Journal of economic literature*, 48(2), 281-355.

[79] Lindsay, C. Page (2012): "Principal stratification as a framework for investigating mediational processes in experimental sttings", *Journal of Research on educational effectiveness*, 5(3), 215-244.

[80] Kaufman, J. S., R. F. Maclehose, S. Kaufman (2004): "A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation", *Epidemiologic Perspectives & innovations*, 1, 4.

[81] Keele, L., D. Tingley, T. Yamamoto (2015): "Identifying mechanisms behind policy interventions via causal mediation analysis", *Journal of policy analysis and management*, 34, 937-963.

[82] Kinder, D.R., L. Sanders (1996): "Divided by color: racial politics and democratic ideals", *Chicago: University of Chicago Press.*

[83] Kopmans, T. C., H. Rubin, R. B. Leipnik (1950): "Measuring the equation systems of dynamic economics". In T. C. Koopmans (Ed.), statistical inference in dynamic economic models, Cowles Commission monograph 10. New York: Wiley, 1950. 53-237.

[84] Mackinnon, D. (2008): "Introduction to statistical mediation analysis", New York: Routledge.

[85] Mancha-Navarro, T., R. Garrido Yserte (2008): "Regional policy in the European Union: The cohesion-competitiveness dilemma", *Regional Science*, 1(1), 47-66.

[86] Mellace, G., A. Pasquini (2019): "Mediation analysis synthetic control", *mimeo.*

[87] Menon, C. and S. Giacomelli (2012): “Firm Size and Judicial Efficiency in Italy: Evidence from the Neighbour’s Tribunal”, SERC Discussion Papers 0108. London: Spatial Economics Research Centre, London School of Economics.

[88] Miquel, R. (2002): "Identification of dynamic treatment effects by instrumental variables", *University of St. Gallen economics discussion paper series*, 2002-11.

[89] Mohl, P., T. Hagen (2008): "Does EU Cohesion Policy promote growth? Evidence from regional data and alternative econometric approaches" Discussion Paper 08-086. ZEW, Mannheim.

[90] ohl, P. and T. Hagen (2010): “Econometric Evaluation of EU Cohesion Policy—A Survey,” ZEW—Centre for European Economic Research Discussion Paper No. 09-052. Manheim, Germany: ZEW—Centre for European Economic Research.

[91] Oreopoulos, P. (2006): "Estimating average and local treatment effects of education when compulsory schooling laws really matter", *American economic review*, 96(1), 152-175.

[92] Papaioannu, S. and E. Michalopoulos (2014): “National Institutions and Subnational Development in Africa”, *The Quarterly Journal of Economics*, 129(1), 151–213.

[93] Pearl, J. (2001): "Direct and indirect effcts", in *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pp. 411-420, San Francisco. Morgan Kaufman.

[94] Pearl J. (2011): "The causal mediation formula: a practitioner guide to assessment of causal pathways. Technical Report R-379, University of California, LA.

[95] Pellegrini, G., T. Muccigrosso (2017): "Do subsidized new firms survive longer? Evidence from a counterfactual approach", *Regional Studies*, 51(10), 1483-1493.

[96] Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso, F. Busillo (2013): "Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach", *Papers in Regional Science*, 92(1), 217-233.

[97] Percoco, M. (2005): "The Impact of Structural Funds on the Italian Mezzogiorno, 1994-1999", *Région et Développement*, 21, 141–152.

[98] Percoco M. (2017): "Impact of European Cohesion Policy on regional growth: Does local economic structure matter?", *Regional Studies*, 51 (6), 833-843.

[99] Petersen, M. L., S. E. Sinisi, M. J. Van der Laan (2006): "Estimation of direct causal effects", *Epidemiology*, 17, 276-284.

[100] Pike, A., A. Rodríguez-Pose, J. Tomaney (2016): "Local and regional development" (2nd edition). Routledge, London.

[101] Powdthavee, N., W. N. Lekfuangfu, M. Wooden (2013): "The marginal income effect of education on happiness: estimating the direct and indirect effect of compulsory schooling on well-being in Australia", *IZA discussion paper No. 7365*.

[102] Robins, J. M. (2003): "Semantics of causal DAG models and the identification of direct and indirect effects", in *In highly structured stochastic systems*, ed. by P. Green, N. Hjort and S. Richardson, 70-81, Oxford. Oxford University Press.

[103] Robins J. M., S. Greenland (1992): "Identifiability and exchangeability for direct and indirect effects", *Epidemiology*, 3(2), 143-155.

[104] Rodríguez-Pose, A., U. Fratesi (2004): "Between development and social policies: The impact of European Structural Funds in objective 1 regions", *Regional Studies*, 38, 97–113.

[105] Rosembaum P. (1984): "The consequences of adjustment for a concomitant variable that has been affected by the treatment", *Journal of Royal Statistical Society, Series A*, 147(5), 656-666.

[106] Rosenzweig, M. R., K. Wolpin (1980): "Testing the quantity-quality fertility model: the use of twins as a natural experiment", *Econometrica*, 48(1), 227-240.

[107] Rubin, B. (1974): "Estimating causal effects of treatment in randomized and non nonrandomized studies", *Journal of educational psychology*, 66(5), 688-701.

[108] Shadish, W. R., T. D. Cook, D. T. Campbell (2001): "Experimental and quasi-causal designs for generalized causal inference", Boston, Houghton Mifflin.

[109] Simonsen, M., L. Skipper (2006): "The costs of motherhood: an analysis using matching estimators", *Journal of applied econometrics*, 21, 919-934.

[110] Tchetgen Tchetgen, E. J., I. Shpitser (2012): "Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis", *The annals of statistics*, 40, 1816-1845.

[111] Trochim W. M. K. (1984): "Research desig for program evaluation: the regression-discontinuity approach", Sage publications, Beverly Hills, CA.

[112] VanderWeele, T. J. (2008): "Simple relations between principal stratification and direct and indirect effects", *Statistics & Probability letters*, 78, 2957-2962.

[113] VanderWeele, T. J. (2009): "Marginal structural models for the estimation of direct and indirect effects", *Epidemiology*, 20(1), 18-26.

[114] VanderWeele, T. J. (2012a): "Comments: should principal stratification be used to study mediational processes?", *Journal of research on educational effectiveness*, 5(3), 245-249.

[115] VanderWeele T.J. (2015): "Explanation in causal inference. Methods for mediation and interaction", Oxford University Press.

[116] VanderWeele, T. J., S. M. Vansteelandt (2009): "Conceptual issues concerning mediation, interventions and composition", *Statistics and its inference*, 2, 457-468.

[117] Vansteelandt, S., M. Bekaert, T. Lange (2012): "Imputation strategies for the estimation of natural direct and indirect effects", *Epidemiologic Methods*, 1, 129-158.