

Discontinuity Designs vs Propensity Score Matching for Evaluating Regional Economic Development and Business Incentive Policies⁺

Daniele Bondonio⁺⁺

Abstract

Regional economic development and business incentive programs have a prominent role in the European Union (EU). For evaluating these programs, in recent years, a growing number of studies have exploited either spatial discontinuities, set by boundaries of the targeted areas, or ranking discontinuities, based on EU-fund eligibility indexes or firm-level application scores. In light of this literature, impact evaluations are being increasingly commissioned and designed under an a-priori assumptions that discontinuity designs have superior impact identification properties. This paper argues that in a number of frequently encountered, but often unrecognized, circumstances this assumption does not hold ground. When the running variable has a weak influence on the outcome of the analysis, discontinuity designs are at risk of either unnecessarily reduce external validity or, in the presence small sample sizes, failing to achieve the complete balancing of relevant controls. In this scenario, ensuring the common support for the crucial confounders and adopting propensity score matching approaches can mimic the features of block randomization designs, constituting a more viable empirical option.

JEL classification: O1; R5; C23

Keywords: Discontinuity Designs, Propensity Score Matching, Regional economic development, Business incentives, EU cohesion Policy,

⁺ This paper is drawn from the plenary talk *Evaluating Regional Economic Development Policies: A Second Look at Discontinuity Designs* presented at “Rigorous Impact Evaluation in Europe. A Research Conference in Honor of Alberto Martini” (Evaluation Review; University of Maryland –School of Public Policy; Collegio Carlo Alberto; International Network for Social Policies. May 20-22, 2018, Torino, Italy). The author thanks Jeffrey Smith and other participants to the conference for useful comments.

⁺⁺ (Corresponding Author) University of Piemonte Orientale, Via Cavour 84, 15100 Alessandria, Italy (e-mail: daniele.bondonio@uniupo.it)

1. Introduction

Regional economic development and business incentive programs have a prominent role in the European Union (EU), with €196 billion (more than 18% of the total EU budget) allocated to these policies in the current 2014-2020 programming period by means of the European Regional Development Fund. For evaluating these programs, in recent years, a growing number of studies have exploited either spatial discontinuities set by boundaries of the targeted areas (e.g. Einiö & Overman 2012; Giacomelli & Menon 2013; Jofre-Monseny 2014; Keele & Titiunik 2015; DeBlasio & Poy 2017; Giuia 2017; Crescenzi & Guia 2018) or ranking discontinuities based on cut-offs in the regional EU-fund eligibility indexes or in firm-level application scores (e.g. Becker et al. 2010; 2012; 2013; 2016; Pellegrini et al. 2013; Ferrara et al. 2016; Cerqua & Pellegrini 2014, 2017; Percoco 2017).

In light of this recent body of literature, impact evaluation studies of European regional economic development and business incentive policies are being increasingly commissioned and designed under an a priori assumption that discontinuity designs have superior impact identification properties: whenever an eligibility boundary, geographical border or application cut-off is identifiable, a nearly idiosyncratic reflex tends to prevail in assuming that cross-border/cross-cutoff comparisons or regression discontinuity designs are the best option for the analysis.

The goal of this paper is to highlight that in a number of frequently encountered, but often unrecognized, circumstances this assumption does not hold ground and favorable impact identification conditions could be also effectively exploited by other empirical strategies that ensure a more transparent and explicit balancing of the relevant controls than discontinuity designs. Cross-border, cross-cutoff comparisons or regression discontinuities, moreover, may be at risk of reducing external validity and/or failing to achieve the complete balancing of the truly relevant controls when the running variable has a weak influence on the outcome variable and the sample sizes are small. Under these circumstances, ensuring the common support for the crucial confounders and adopting propensity score matching approaches can mimic the features of block randomization designs, constituting a more viable empirical option

The reminder of this paper is organized as follows: section 2 focuses on spatial discontinuities, section 3 discusses ranking discontinuities, while section 4 offers some concluding remarks.

2. Spatial Discontinuities

The Importance of Theory of Change: a Forgotten Lesson of Holmes 1998, Black 1999 and Dube et. al 2010

The recent literature on European regional economic development programs (e.g. Einiö & Overman 2012; Giacomelli & Menon 2013; Jofre-Monseny 2014; Keele & Titiunik 2015; DeBlasio & Poy 2017; Giuia 2017; Crescenzi & Guia 2018) acknowledges Holmes (1998) as a first truly seminal work that inspires the adoption of spatial discontinuity designs. In his 1998 paper, Holmes lays out, upfront, a convincing theory of change linking the policy of interest (pro-business /right-to-work state law) to the observed outcome (long-term –postwar period- changes in the location of

manufacturing). In this way he identifies the main threats to the validity to be controlled for in the analysis, in terms of the major forces that affect the outcome variable independently from the policy of interest and that are differently distributed between the “treated” and “non-treated” states. It is such clearly defined theory of change that in Holmes (1998) supports the rationale for considering cross-border comparisons a superior impact identification strategy that enables the analysis to control for the major confounding factors in ways that do not rely on strong functional form assumptions: “At state borders, the geographic determinants of the distribution of manufacturing - for example, climate of a location; fertility of the soil; access to transportation, an ocean, river, or lake; proximity to raw materials; attitudes of people toward unions; and level of agglomeration benefits—are approximately the same on both sides of the border.” (Holmes 1998, p. 671).

Similarly to Holmes (1998), also Black (1999) is a second seminal precedent that is widely cited by the literature on European regional economic development programs. As for Holmes (1998), also in Black (1999) the adoption of a spatial discontinuity approach is thoroughly justified by the developing of a detailed theory of change that explicitly underlies the impact identification advantages of cross-border comparisons (within a same school district but across the boundaries of the specific schools’ attendance sub-districts) when estimating the impact of school quality on house prices.

Theory of change, finally, is also what justify the adoption of spatial discontinuity designs in a third study (Dube et al. 2010) that is widely cited by the European impact evaluation literature. Dube et. al. (2010) estimate short-run employment impacts of minimum wage raises, following the seminal precedents that exploit firm-level data on the fast-food industry (Card & Kreuger 1994, 2000 and Neumark & Wascher 2000). In this literature, the focus on fast-food restaurants is motivated by advantages in controlling for the main threats to the validity of the analysis, in terms of the confounding factors that may affect low-wage employment independently from the policy change.

Dube et. al. (2010) build on such theory of change to also justify the adoption of cross-borders comparisons as the empirical strategy of choice. This is because the difference in difference (DD) scheme adopted in Card & Kreuger (1994, 2000) and Neumark & Wascher (2000) controls for unobservable differences in the factors that affect the employment outcome of “treated” and “untreated” restaurants only under strict parallel trend assumptions. Comparing restaurants within same local markets crossed by a policy-change state border ensures instead that most of the relevant confounding factors are spontaneously controlled for (without relying on any functional form assumption). This rationale for adopting spatial discontinuity designs, however, is closely related to the same theory of change that justifies the focus on an industry, such as fast-food restaurants, that sells services and acquires labor inputs exclusively in the local market. There would be little or no basis for cross-border comparisons if the focus of the analysis were on different industries that do not sell products or services on the local market.

Moreover, in the analyses of fast-food restaurant data and minimum-wage state policies, it is also important to notice that the rationale for adopting spatial discontinuity designs does not stem from a scenario in which an increased distance from the policy-change state border signals a monotonic increase in the likelihood of comparing “treated” and “untreated” firms with unbalanced unobservables. This is because these analyses focus on short-run changes of an outcome variable

recorded in units of observation that were already in existence at the time of the policy change. In the short run, and with respect to the unobservables of the existing restaurants, the “treatment assignment” (i.e. becoming affected by an higher minimum wage) is at random also away from the policy-change state border. Considering for example the well know case of the New Jersey (NJ) and Pennsylvania (PA) fast-food industry at the time of the 1992 New Jersey minimum wage raise (Card & Krueger 1994, 2000 and Neumark & Wascher 2000), such random nature of the “treatment assignment” with respect to restaurant-specific unobservables, is due to the fact that there was no self-selection into treatment in terms of relocation efforts of the existing NJ restaurants into moving to PA, in order to avoid a forthcoming minimum-wage raise. This is because the NJ minimum-wage raise was not such a dramatic and well-anticipated event to spur massive preventive relocation attempts of the existing restaurants away from NJ into PA. In this circumstance, the distance from the states’ border is not correlated to the managerial abilities of the restaurants (i.e. restaurants did not change their location in anticipation of the minimum-wage change and, consequently, no restaurants with higher managerial abilities relocated deeper into the PA territory, while the low-ability restaurants remained located deeply into the NJ territory, Figure 1).

FIGURE 1
Distribution of Fast-Food Restaurants in Card & Krueger (1994)’s Data

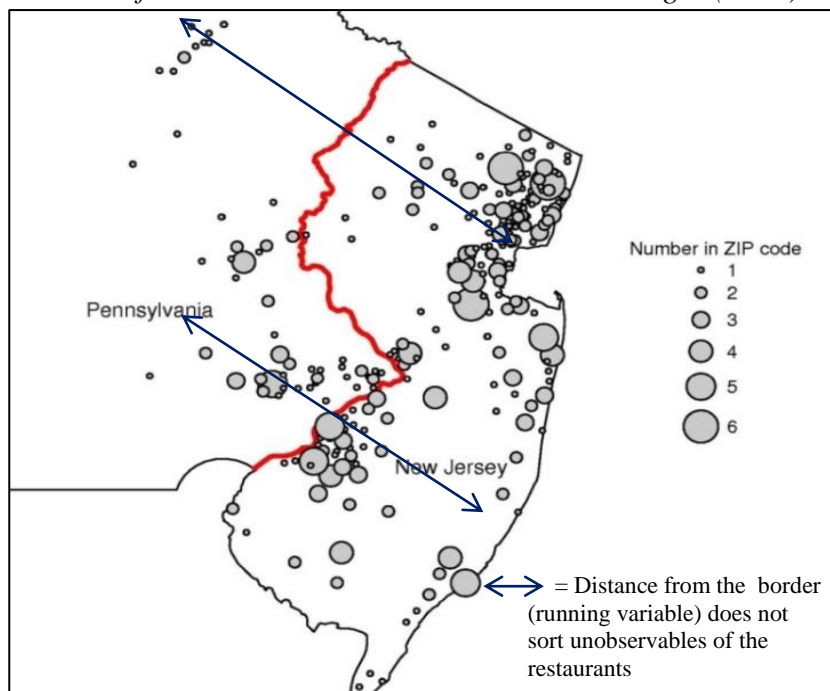
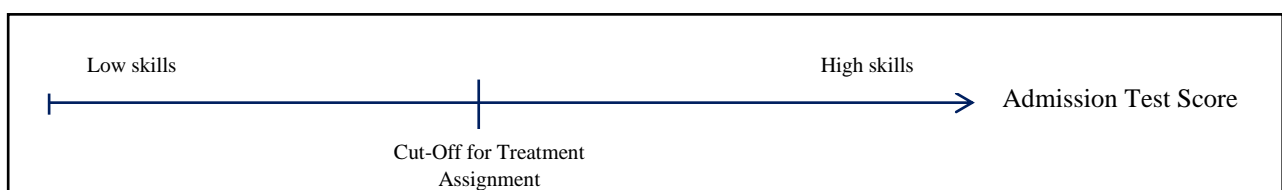


FIGURE 2
*Example of Classical Set-Up for Discontinuity Designs:
Admission-Test Scores (Running Variable) Sort Crucial Unobservables*



Such lack of sorting of the unobservables along the distance from the policy-change borders is very common in many spatial discontinuity set-ups (including Holmes 1998) and is very different from the cases of the classical scenarios for discontinuity designs related, for example, to admission tests for job training programs. In these latter cases the running variable (i.e. the admission score) sorts the unobserved abilities of the applicants in a way that the higher is the distance from the cutoff for admission, the higher is the unbalance of crucial unobservable controls (Figure 2).

In the recent applications of discontinuity designs to European programs, the lesson of Holmes (1998), Black (1999) and Dube et al. (2010), about justifying the adoption of spatial discontinuity designs by means of a specific and detailed theory of change, is often forgotten. Rather, there is an increasing trend in postulating that spatial discontinuity designs have superior impact identification properties based on a-priori assumptions or mere citations of precedents in the literature. For at least two important types of European regional economic development and business incentive programs, such postulated assumptions do not hold ground.

Regional-based Policies with Discontinuities of Treatment Intensity at the Borders

A first type of programs, for which superior impact identification properties of discontinuity designs cannot be postulated, is that of either regional-based cohesion policies supported by the EU structural funds (e.g. Becker et al. 2010; 2012; 2013; 2016 Pellegrini et al. 2013; Ferrara et al. 2016; Cerqua & Pellegrini 2017; Percoco 2017) or other types of regional-based interventions for which the focus of the evaluation is on short-run regional outcomes and spatial discontinuities are encountered at the policy-change borders. These latter policy interventions are, for example, local wage regulations (De Blasi & Poy 2017) or regional unemployment protection programs (Jofre-Monseny 2014), for which cross-border comparisons or spatial regression discontinuity designs can be adopted using data at either the municipality or the province level. For the case of EU cohesion policies, the “treated” regions with the higher intensities of EU structural funds are referred to as “Objective 1”, “Convergence” or “Less-developed” regions¹, and when spatial discontinuity approaches are adopted (e.g. Guia 2017, Crescenzi & Guia 2018), the outcome variables of the analysis are derived from firm-level or individual-level data aggregated at the municipality-level.

Without the development of a policy-specific and credible theory of change, however, for these programs the rationale for adopting a spatial discontinuity approach cannot stem from a mere referral to Holmes (1998) or (Black 1999) in postulating, a priori, that spatial units at the threshold have homogeneous characteristics in everything except for the treatment and that the confounding unobservables evolve smoothly over space². This can be clearly illustrated considering, for

¹ “Objective 1”, “Convergence” or “Less-developed” regions in the EU are areas with, in general, a per-capita GDP lower than the 75% of the EU average. “Objective 1” was the denomination in place until the 2000-06 programming period, “Convergence” was the denomination for the 2007-13 programming period, while “Less developed” is the denomination in place for the current 2014-20 period.

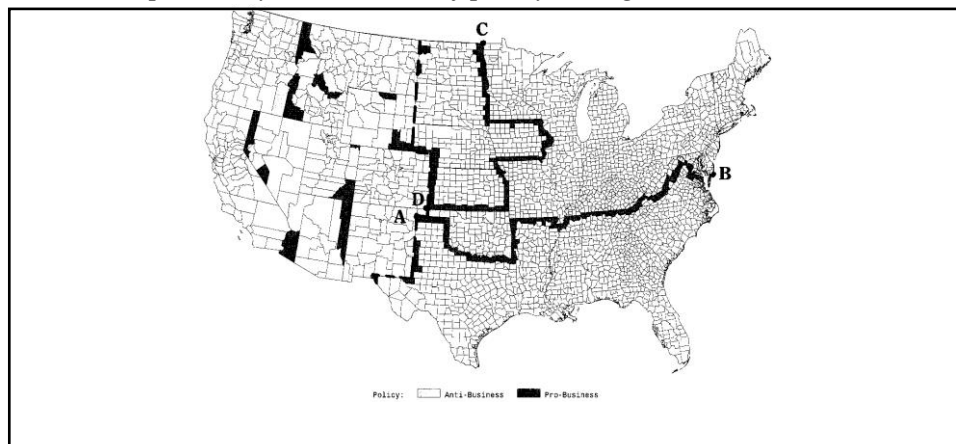
² E.g. Jofre-Monseny (2014, p. 78): [“Following Holmes (1998) and Black (1999), we choose neighboring municipalities that are on opposite sides of the SIPTA -i.e. the regional unemployment-protection policy of interest- regional border as our treatment and control groups. This identification strategy is appealing because it controls for confounding unobservables that evolve smoothly over space”].

Guia (2017, p. 110): [“This spatial approach has not previously been applied in the counterfactual literature on EU Regional Policy but it is increasingly used to evaluate spatially-targeted policies in different fields. The simple idea on

example, the case of the EU cohesion-policy interventions supported by the structural funds, in comparison with Holmes (1998)³. Holmes (1998) justify the adoption of a cross-border comparison (Figure 3) by explicitly describing all the specific circumstances that support the assumption of a monotonic increase in the unbalancing of the relevant controls when moving away from adjacent counties across the border. These circumstances are as follows. First, the pre-treatment period is set in the early post-world-war-II era, with very little (or complete absence) of anthropic density (in terms of pre-existing infrastructures, industrial activity, urban settlements, historical sites). Second, the North-America terrain has smoothly and slow-changing morphological features, considering, in particular, the potential for industrial development in terms of flat versus steep-land. In most cases this is coupled by a complete absence of abrupt changes, or short-distance alternations in pre-existing transportation infrastructures and tourism attractiveness. Third, the outcome variable of interest is in terms of long-term changes (i.e. over a span of 40+ years) in manufacturing locations. This entails the fact that all the infrastructural and urbanization trends, that occurred after the early post post-world-war-II era (i.e. the pre-treatment period in Holmes 1998), are to be considered as part of the treatment effect, and they do not have to be controlled for. Such specific and unique circumstances make a very strong rationale for arguing that, by means of cross-border comparisons between adjacent counties with different treatment status, all major confounding factors are naturally controlled for (without relying on any functional form assumptions).

FIGURE 3

Counties in proximity (<25 miles) of policy change border in Holmes (1998)



which it relies is that across space only the treatment assignment changes with a discontinuous jump, whereas all the other observable characteristics are smoothly distributed. Provided that unobservable characteristics are distributed similarly to the observables, the different spatial patterns of the treatment (sharply changing) with respect to all the other aspects (randomly distributed and changes smoothly) makes it possible to separate its effect from everything else (Black, 1999)"].

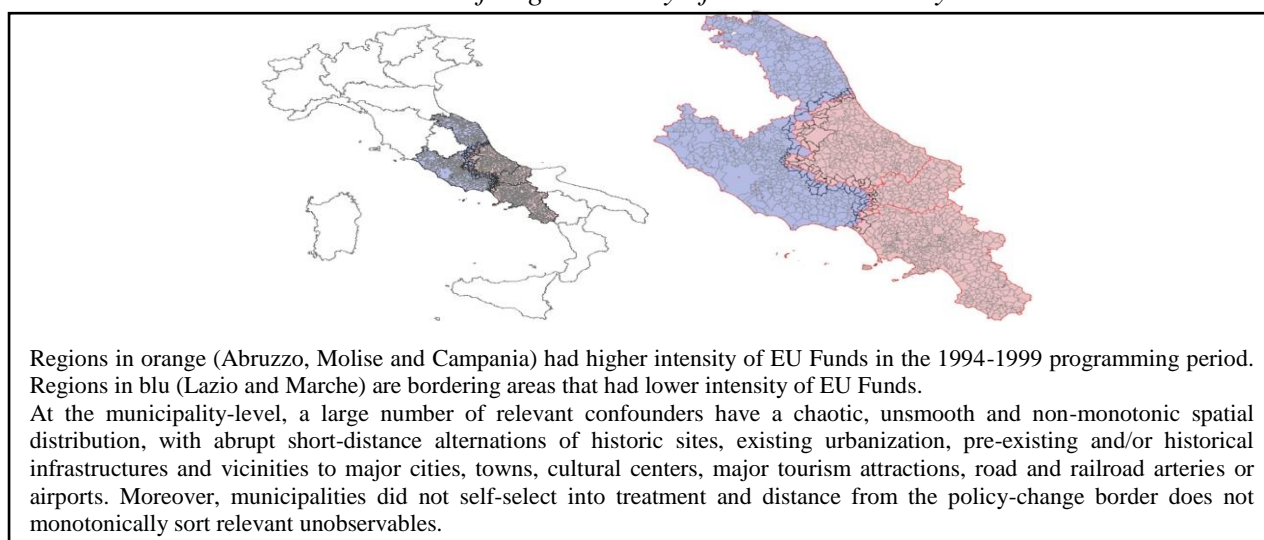
De Blasio & Poy (2017, p. 49): ["We compare the employment performance of municipalities similar on the basis of geographic and sociodemographic characteristics, but different in terms of the provincial wage zone they belong to. As in the groundbreaking paper of Holmes (1998), we focus on what happens when one crosses province borders. This helps to isolate the effect of wage zones from that of other characteristics, which might confound identification: geographically close municipalities are likely to be homogeneous in terms of many observed and unobserved local features."]

³ Similar arguments apply also for the case of other regional-based policy intervention, such as regional unemployment-protection (e.g. Jofre-Monseny 2014) or wage-zone policies (e.g. De Blasio & Poy 2017).

In many EU countries, in proximity of policy-change borders, and considering shorter-term outcomes, this same rationale for cross-border comparisons is completely absent. This is because, very often (and this is the case, for example, of the borders -Figure 4- for the Italian regions with high intensities of EU funded cohesion policy interventions), the territories across adjacent municipalities offer an extraordinary degree of variations and sharp and abrupt changes in terms of entropic density and morphological features of the terrains. Very commonly there are municipalities (or provinces) with prevalingly flat terrain (suitable for development), alternated with adjacent municipalities with prevalingly steep terrain (not suitable for development). Continuous and abrupt short-distance alternations, across a series of adjacent municipalities (or provinces), are also very common in terms of presence of historic sites, existing urbanization, pre-existing and/or historical infrastructures and vicinities to major cities, towns, cultural centers, major tourism attractions, road and railroad arteries or airports. All these elements, in many European regions, generate a chaotic, unsmooth and non-monotonic spatial distribution of a large number of relevant confounding factors across all the municipalities (or provinces) located anywhere in a group of adjacent regions. Moreover, it is also important to notice that, likewise in the analyses of the fast-food industry at the time of minimum-wage changes, when the units of observation are at the municipality (or province) level, there is no self-selection into treatment (as the treatment assignment is set at the regional level, with well-established and pre-existing administrative borders). Under these circumstances, unobservables are not monotonically distributed along the distance from the policy-change regional border. As a consequence, comparing municipalities (or provinces) in proximity of the policy-change border does not ensure, per-se, any better balancing of crucial observable and unobservable confounders than comparing municipalities (or provinces) further away from the border.

FIGURE 4

*Example of Municipalities in Proximity of Policy-Change Borders in Europe.
The case of High Intensity of EU Funds in Italy*



Under this commonly-encountered European scenario, adopting a cross-border comparison design is the equivalent of arbitrarily restricting the focus on a subset of municipalities (those close to the border) when exogenous treatment assignment conditions (in terms of lack of self-selection into treatment based on unobservables) apply also for the municipalities away from the border. Such

restricted focus reduces the sample size of the analysis without obtaining any clear advantage in better controlling for crucial confounders. Because of a reduced sample size, cross-border comparisons have a higher likelihood of hiding some unbalance between treated and non-treated units in important controls (for which data are often at risk of not being even collected when discontinuity designs are implemented). This is similarly to what would happen in actual randomized treatment assignments, if the sample sizes are not large enough to guarantee an adequate balancing of the crucial controls. For example, Kernan et al. (1999) highlight that, even with pure random assignment, the risk that the treatment and control group differ significantly along some important dimensions becomes very relevant with sample sizes smaller than 200 units. In such cases, a block randomization would be required to achieve an adequate balancing of the major confounders (e.g. Bruhm and McKenzie, 2008).

For these reasons, applying spatial discontinuity designs to evaluate these types of regional-based European policies may be at higher risk both of unnecessarily reducing external validity and failing to ensure the proper balancing of the crucial controls, compared to other impact identification strategies that, for example, mimics the features of a block randomization design. These identification strategies could be, for example, in terms of the following steps. First, the focus of the analysis can be enlarged to units of observations farther away from the border, and theory of change can be used to explicitly identify the crucial observable controls. Second, on the base of these controls, units of observation may be excluded based on the lack of common support (instead on the mere distance from the border), and statistical matching procedures may be used to ensure an actual balancing of the crucial confounders between the treatment and the comparison group (e.g. Bondonio 2016).

Non-Exogenous Borders of the Treatment Areas: The case of EZ, Science Parks, ZFU, Local-Partnership Programs

A second types of European programs, for which the postulated assumptions of superior properties of spatial discontinuity designs do not hold ground, are the geographically-targeted interventions in which the borders of the treatment areas are specifically shaped as part of the policy intervention. These types of interventions are quite widespread throughout most of the European countries in terms of Industrial Parks and Science Parks, Enterprise Zones in UK (Einio & Overman 2012) and France (“Zones Franches Urbaines”, Rathelot & Sillard 2009, Gobillon et al. 2012), and Local Partnerships Areas in Italy (“Patti Territoriali”, Magnatti et al. 2005, Accetturo & De Blasi 2012).

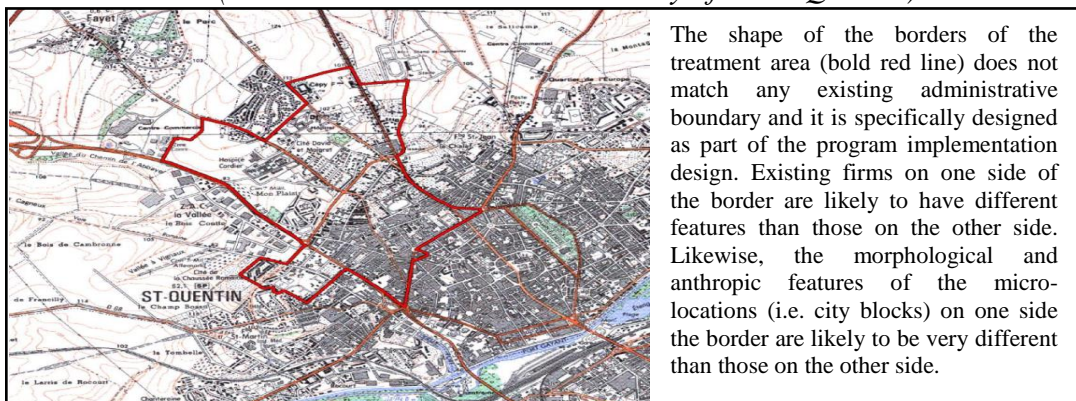
Adopting cross-border comparisons or spatial regression discontinuity designs for evaluating these types of programs is particularly problematic because when the borders of the treatment areas are specifically designed as part of the policy implementation process, and do not coincide with pre-existing municipal/province/state or regional boundaries, near random treatment assignments in proximity of the edge of the targeted areas is not granted. This is because the decision on where exactly to draw the boundary lines for the treatment areas is very likely to be endogenous to the characteristics of each firm in existence in the specific micro-locations (e.g. city blocks) taken into consideration for inclusion (or exclusion) in each targeted areas at the time of the policy implementation phase. Moreover, for the same reason, also the morphological and anthropic

features of the terrain in proximity of the treatment-assignment border is likely to be endogenous, with systematic differences between adjacent micro-locations across the opposite sides of the border (Figure 5).

Under these circumstances, comparing treated and untreated firms located in proximity of the border of a treatment area is at high risk of hiding some relevant unbalance of crucial unobservables. Further away from border, instead, it is more likely that exogenous treatment exclusions may be in place due, for example, to political reasons that prevented the implementation of treatment areas in other parts of a same province or region. For these reasons, a better balance of crucial unobservables is likely to be achieved by other empirical strategies, than discontinuity designs. These could be in terms of excluding non-treated firms in proximity of the border of the treatment areas, gathering data on crucial observable confounders, and adopting a statistical matching estimator to compare the outcomes of the treated firms to those of similar non-treated firms located elsewhere within a same province or region⁴.

FIGURE 5

*Example of Non-Exogenous Borders: An Enterprise Zone Area in France
("Zone Franche Urbain" in the City of Saint-Quentin)*



3. Ranking Discontinuities

Ranking discontinuity designs are being increasingly adopted in Europe for evaluating the impact of mainly either regional-based cohesion policy interventions supported by the EU Structural Funds or business incentive programs to support firm-level investment and innovation projects.

In the first case, the focus of the analysis is on estimating the impact of EU funded interventions on regional growth. During the EU programming periods 1994-99; 2000-06 and 2007-2013, higher intensities of EU Funds were allocated to regions with per-capita GDP levels below 75% of the EU average. This 75% cut-off for the eligibility to the higher intensity of EU Funds⁵ determines a sharp ranking discontinuity exploited in the literature (e.g. Becker et al. 2010; 2012; 2013; 2016; Pellegrini et al. 2013;; Ferrara et al. 2016; Cerqua & Pellegrini 2017; Percoco 2017).

⁴ In the presence of longitudinal data, a difference in difference scheme can be also added to the estimation procedure, in order to control for unobservable fixed effects that may remain accidentally unbalanced between the treated and non-treated firms (e.g. Bondonio and Martini 2012 and 2018).

⁵ The regions with higher intensities of EU Funds are referred to as either "Objective 1", "Convergence" or "Less-developed" areas (see Footnote 1 for additional details).

In the second case, the focus is on estimating the impact of business incentive programs on short-run firm-level outcomes. In some noticeable circumstances, for example in the case of the Italian Law 488/92 (e.g. Bondonio and Martini 2012, Cerqua & Pellegrini 2014) and the Portuguese POE/PRIME program (e.g. Bondonio et al. 2016), these policy interventions are implemented by means of calls for applications with scores assigned to the submitted investment projects. Firms are ranked based on their application score, with a cut-off for treatment assignment that is often determined exogenously (for example, due to limits in the program budget) and that can be exploited in the context of a discontinuity design (e.g. Cerqua & Pellegrini 2014).

Rankings of EU Regions versus Rankings of Firms

A strong difference, that is often unrecognized, is in existence between the ranking of regions for the allocation of the EU funds and the ranking of firms within business incentive programs. In the latter case (similarly, for example, to many education or job-training programs), the selection into treatment is based on a two-stages process. In the first stage, firms choose to apply or not to apply to the program, and the status of applicant versus non-applicant firm is primarily the result of self-selection. In the second stage, applicant firms are sorted into treated and non-treated applicants based on cut-offs in the rankings of applicants, determined by program rules and procedures.

In the case of the allocation of EU funds, instead, stage one does not apply: all regions within the EU countries are automatically taken into consideration for receiving EU funds. An higher intensity of the funds (i.e. the treatment to be evaluated) is determined by a centralized rule based solely on an observable variable (the ranking of the per-capita GDP indicator). For this reason, no distinction is in existence between the general population of the units of observation and the applicant units, and the running variable for the allocation of the funds is not a proxy of crucial unobservables (e.g. motivation to apply for the support, administration abilities, institutional qualities etc.). With respect to such unobservables, natural experiment conditions are in place also away from the cut-off. Thus, restricting the focus on regions in proximity of the treatment-assignment discontinuity ensures the balancing solely of one observable covariate (the running variable), and it is not a necessary step to ensure the balancing of crucial unobservables. This is unlike the case of business incentive programs in which, instead, a strong difference is in place between applicant and non-applicant firms in terms of crucial unobservables (such as, for example, position of the firm on its investment cycle, desire to invest and managerial abilities, in terms of awareness of the incentives and availability of human resources capable of preparing the application package).

Cohesion Policies Supported by the EU Structural Funds

For evaluating regional-based cohesion policy interventions supported by the EU structural funds, superior impact identification properties of ranking discontinuity designs are increasingly postulated with reference to a literature that focus on different policy domains (see, for example, the Lee and Lemieux's (2010) guide to discontinuity-design applications in economics). These domains are, for example, education, welfare, unemployment insurance, and disability programs, in which there is often a strong rationale in assuming that the running variable is a proxy for (often unobservable)

crucial confounders. In these circumstances, pseudo-experimental treatment assignment conditions hold only in the proximity of the cut-off. In the case of the EU cohesion-policy interventions, instead, no credible theory of change can convincingly support the assumption that the running variable is an important determinant of the future regional growth outcomes that are commonly of interest in the evaluation design. This is because the indicator used for determining the intensity of the EU funds is based on GDP levels (and not growth) and it is measured years before the beginning of the actual policy interventions⁶.

This is an important aspect because the more the running variable has a negligible effect on the outcome variable (and it is not a proxy of unobservables to be controlled for), the more the treatment assignment mimic an actual randomization also away from the cut-off point, with no monotonic increase in the unbalancing of the relevant controls when moving away from the cut-off (e.g. Lee and Lemieux 2010). In these circumstances, the actual balancing between the treatment and the comparison group of the control variables relies on sample sizes that have to be large enough to grant that no accidental clustering within one of the two groups took place for any of the confounders⁷. Similarly to what previously highlighted for spatial discontinuity designs, this is in the same way as it can occur in actual randomized treatment assignments, for which block randomization on crucial confounders would be required in the presence of small sample sizes (e.g. Kernan et al. 1999).

For these reasons, when commissioning or planning evaluations of EU-funded cohesion policies in the presence of running variables that are negligible confounders, particularly with small sample sizes, there is no rationale in assuming that ranking discontinuity designs have clearly superior impact identification properties. Other empirical strategies, such as statistical matching estimators, may ensure a more straightforward and transparent balancing of the relevant confounders, mimicking the role of block randomization between regions above and below the cut-off point. Thus, if the running variable is a negligible confounder, and sample sizes are small, evaluation designs should contemplate to acquire data also on the major observable confounders and enlarge the focus to all regions above and below the cut-off for the treatment assignment. Then, units of observation can be excluded based on the lack of common support for the crucial observable confounders (not just based on the distance from the cut-off in the running variable), and matching estimators can be used to ensure the equivalent of a block randomization in the presence of small-sample-driven residual unbalancing of confounding factors.

Is the Running Variable for the Allocation of EU Funds a Relevant Confounder?

Some empirical evidence on whether or not the running variable for the allocation of EU funds is actually a proxy of crucial confounders can be found by correlating the GDP indicator, used to

⁶ The per-capita GDP indicator, expressed in terms of percentage of the EU average, was computed on 1988-90 GDP-level data for allocating the EU funds of the 1994-99 policy interventions; on 1994-96 GDP-level data for the funds pertaining to the 2000-06 interventions; and on 2000-02 GDP-level data for the funds pertaining to the 2007-13 interventions).

⁷ When instead the running variable is an important proxy for crucial confounders, such unbalancing concerns tend to lose relevance. This is because in proximity of the cut-off, treated and untreated units with similar values of the running variable would be likely to also have similar values of the crucial observable and/or unobservable confounders.

identify the regions with high intensity of EU funds, with some obvious pre-intervention observable controls. When the outcome of interest is the regional GDP growth recorded during the EU programming periods in which the cohesion-policy interventions took place (similarly as in Becker et al. 2010; 2012; 2013; 2016; Pellegrini et al. 2013; Ferrara et al. 2016; Cerqua & Pellegrini 2017; Percoco 2017), these pre-intervention controls can be, for example, the growths of per-capita GDP, worker compensation, fixed capital formation, labor productivity and employment rate.

As shown in Table 1, the balance of the above-mentioned controls between the treated and non-treated regions that are in the proximity of the cut-off of the running variable (e.g. the regions with values of the GDP indicator within $\frac{1}{4}$ s.d. of distance to the 75% threshold for the higher intensity of the funds, upper part of Table 1)⁸ is overall no better than the balance of the controls between the full population of all treated and not treated EU regions (lower part of Table 1)⁹.

TABLE 1
Balance of Observable Controls between Treated and Non-treated EU regions

	Mean		%bias ^b	t-test	
	Treated ^a	Non-treated		t	p> t
Regions within ¼ SD away from cut-off of the running variable					
N	13	13			
Δ% GDP	1.1565	1.5455	-28.8	-0.74	0.469
Δ% Worker compens.	-.24538	1.8341	-86.7	-2.21	0.037
Δ% Gross fixed cap. form.	-1.4274	-.83096	-15.5	-0.39	0.696
Δ% Lab. Product.	2.6028	2.9452	-22.1	-0.56	0.579
Δ% Employment rate	-1.3534	-1.3031	-3.2	-0.08	0.935
All EU regions					
N	54	133			
Δ% GDP	1.7761	1.0389	17.7	1.36	0.174
Δ% Worker compens.	3.1352	1.24	35.9	2.86	0.005
Δ% Gross fixed cap. Form.	-1.042	.02656	-21.7	-1.39	0.166
Δ% Lab. Product.	2.7872	2.1785	14.0	1.08	0.279
Δ% Employment rate	-.90036	-1.0859	15.1	0.98	0.326

Source: Eurostat data

a. Regions with higher intensity of EU funds for policy interventions in the 1994-99 programming period (i.e. "Objective 1 areas"). Non-treated regions are those with lower intensity of EU funds.

b. Difference of the means of treated and non-treated regions as a percentage of the square root of the average of the sample variances in the two groups (Rosenbaum and Rubin, 1985).

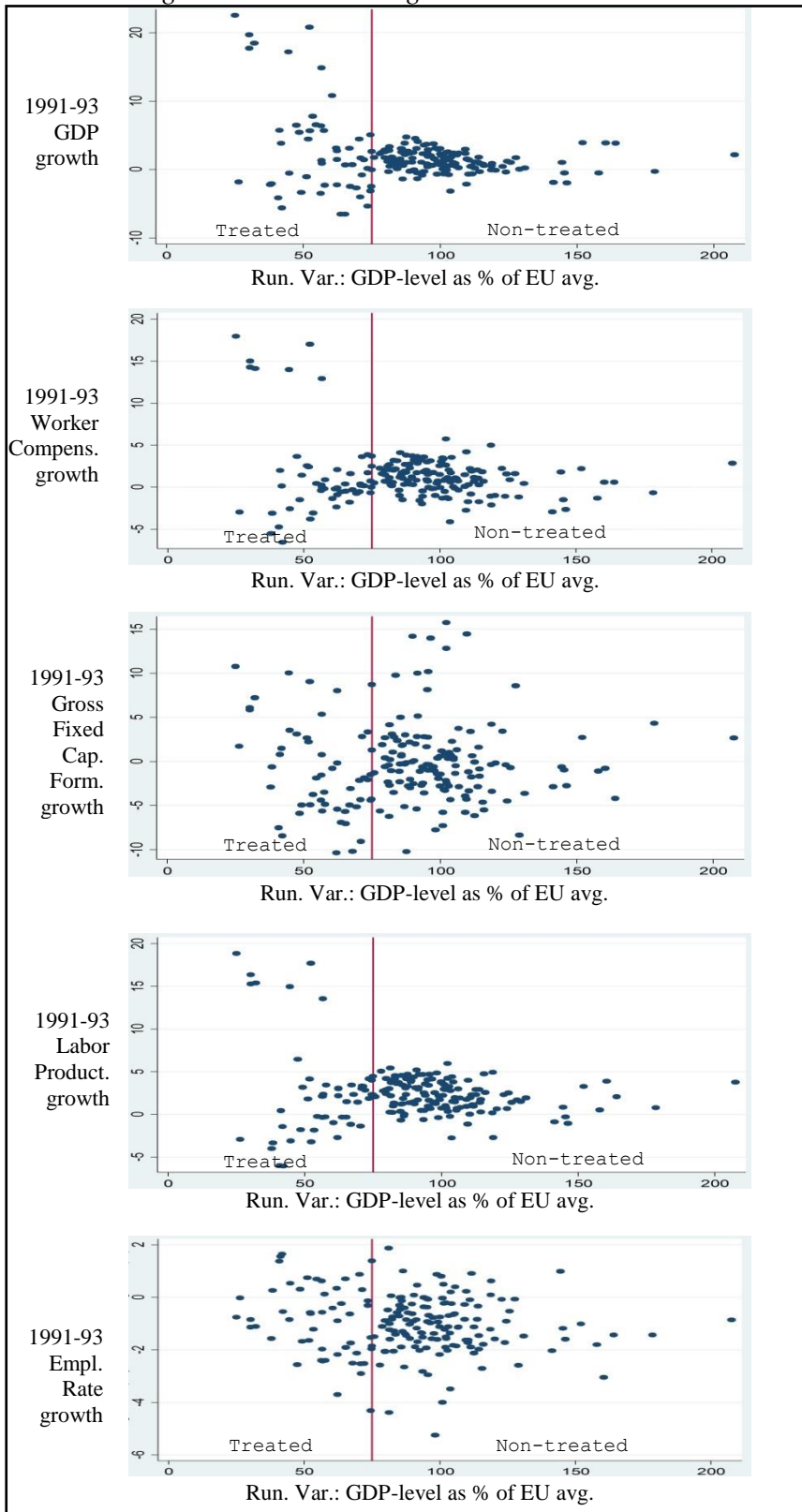
Moreover, plotting the running variable against the different pre-intervention controls (Figure 6) reveals, in nearly all cases, a chaotic and seemingly random distribution. This is highly consistent with the hypothesis that the GDP-level indicator for the allocation of the EU structural funds is indeed unlikely to be a proxy of crucial confounders.

⁸ The data reported in Table 1 are in terms of yearly average % change of the control variables recorded between 1991 and 1993. This 1991-93 interval represents the pre-treatment period that is the closest to the subsequent 1994-99 EU programming period, and it is chosen due to the fact that the policy interventions from 1994 onward are at the center of the discontinuity design impact evaluation literature related to the EU structural funds (e.g. Becker et al. 2010; 2012; 2013; Pellegrini et al. 2013; Becker & Egger 2016; Ferrara et al. 2016; Cerqua & Pellegrini 2017; Percoco 2017). The values of the running variable, finally, are those used for the allocation of the EU funds pertaining to this 1994-99 programming period (i.e. per-capita GDP level as % of the EU average in the years 1988-90).

⁹ Similar results (available upon request to the author) are also obtained when other boundaries are considered for operationalizing the proximity to the cut-off.

FIGURE 6

Running Variable Plotted against Observable Controls



Business Incentive Programs with Ranking of Applicants

For evaluating European business incentive programs, the availability of non-treated applicants is undoubtedly a desirable impact identification scenario. This is because, in most programs, all firms within contiguous ranges of industrial sectors, size classes and/or regional locations are typically fully eligible to apply for the incentives, with no exogenous exclusions¹⁰ that can be exploited for finding comparison groups composed by non-applicant firms that share with the treated firms the same crucial unobservables (in terms, for example, of the firm's position on its investment cycle, desire to invest and managerial abilities for preparing the application package). For this reason, the presence of cut-offs for treatment assignment in the application scores are extremely helpful for impact identification, because they enables the analysis to focus solely on treated and non-treated applicants that share the same crucial unobservables.

When commissioning and/or designing impact evaluations of programs with rankings of applicants, and cut-offs for treatment assignments, however, theory a change to identify the main confounding factors is still important, and, within empirical strategies that restrict the focus to applicant firms, no superior properties of discontinuity designs should be postulated simply based on references to the literature pertaining to other policy domains. This is because, similarly to the case of the EU structural funds interventions, in some, often unrecognized, circumstances it is possible that the application score is not among the relevant confounding factors identified with respect to the outcome of the analysis, and, often, the sample sizes of the applicant firms are relatively small.

Application scores that are weak or negligible confounders may happen because of two possible reasons. First, the computation of the score may be based on program rules that focus on factors that do not have any relevant influence on the outcome variables. This is the case, for example, of the POE-PRIME program in Portugal during the 2003-2006 period. For this program, Bondonio et al. (2016) show that the application score did not sort out the applicants based on characteristics that correlate with their pre-treatment trends of the outcome variables. Under these circumstances, and because of the concurring presence of small sample sizes, Bondonio et.al (2016) opted for gathering data also on crucial observable confounders. With these data they implemented an empirical strategy that, while restricting the focus solely on the applicant firms, ensured the actual balancing of the crucial confounders by means of a statistical matching estimator.

Second, in many business incentive programs¹¹, the application score is assigned with reference to the features of a submitted investment project, while the impact evaluation is carried with firm-level data. Thus, in many programs, it is quite possible that firms with very different crucial confounders (e.g. size, industrial sector, location, etc.) submit investment projects with features that, based on the program rules, determine similar application scores. In this scenario, the application score would not capture a number of crucial observable confounders, because, in essence, the analysis focus on firm-level outcomes (with units of observation represented by the firms), while the application scores are determined with reference to different units of observation (i.e. the investment projects), for which no specific outcomes can be recorded. This is unlike many other policy domains, in

¹⁰ A comprehensive summary of the features of business incentive programs in Europe can be found on the European Commission platform: http://ec.europa.eu/regional_policy/en/

¹¹ As previously mentioned, http://ec.europa.eu/regional_policy/en/ contains a comprehensive summary of the features of European programs.

which, instead, the application scores are determined with reference to the same units of observation (e.g. individuals, in the case of education, welfare, disability and job training programs) for which the outcome variables are measured.

When application scores are negligible confounders, as previously discussed for the EU structural fund interventions, desirable natural experiment conditions may extend further away from the cut-off, and cross-cut-off comparisons or regression discontinuity designs may suffer from small-sample limitations that may lead to failures in balancing crucial observable controls. Under these circumstances, other viable empirical strategies, can be adopted to exploit the presence of non-treated applicants. These strategies, similarly again to the case of the EU structural fund interventions, require to acquire data also on the major observable confounders (e.g. industrial sector, size, age and location of the firm) and to exclude from the analysis applicant firms based on the lack of common support for the crucial confounders. In the presence of small samples, matching estimators can finally ensure to the analysis a more transparent and explicit balancing of the crucial controls, mimicking a block randomization design (e.g. Bondonio et.al. 2016, Bondonio and Martini 2018).

4. Conclusions

Regional economic development and business incentive programs are very prominent policy intervention domains in Europe. In recent years, impact evaluation studies of these programs have been increasingly commissioned and designed under an a priori assumption that discontinuity designs have superior impact identification properties than other empirical strategies. This paper reviews and discusses some different types of widespread European policies for which, in frequently encountered, but often unrecognized circumstances, this assumption does not hold ground and other empirical strategies may be preferable for the analysis.

In one of these types of policies, the borders of the treatment areas are specifically shaped as part of the program intervention and do not follow pre-existing administrative boundaries. This is the case of a number of geographically-targeted programs for economic development, and/or revitalization of economically-depressed areas, such as, for example, Industrial Parks and Science Parks (widespread in many European countries), Enterprise Zones (in UK and France) and “Local Partnerships Areas” (in Italy). For these programs, the decision on where exactly to draw the boundary lines of the treatment areas is very likely to be endogenous to the characteristics of each firm in existence in the specific micro-locations taken into consideration for inclusion (or exclusion) in each targeted areas. Under these circumstances adopting spatial discontinuity designs is particularly problematic, because the assumption of nearly-experimental treatment assignment in proximity of the policy-change border is violated. As a consequence, firms with different treatment status in proximity of the border are likely to have systematically different unobservables, while further away from the border it is actually more likely that some exogenous treatment exclusions can be encountered.

For other types of prominent regional-based policies, such as, for example, the cohesion-policy interventions supported by the EU structural funds, local wage-regulations or regional unemployment protection programs, the units of observation do not self-selected (based on crucial

unobservables) along the values of the running variable (represented by the distance from either the policy-change border or the cut-off in the ranking of the GDP indicator used for the allocation of the EU funds). Moreover, for evaluating these types of policies, no specific theory of change (and also no empirical evidence, in the case of the EU structural funds) can be often developed to support the assumption that the running variable is a relevant observable confounding factor for the analysis. Under these circumstances (that are unlike the case of classical set-ups for discontinuity designs in other policy domains), the more the running variable has a negligible effect on the outcome variable (and it is not a proxy of unobservables to be controlled for), the more the treatment assignment mimics an actual randomization also away from the cut-off, with no monotonic increase in the unbalancing of the relevant controls when moving away from the policy-change discontinuity. In this scenario, adopting cross-border/cross-cut-off comparisons or regression discontinuity designs in the presence of small sample sizes may lead to failures in the balancing of the crucial observable controls. This is similar to what would happen in actual randomized treatment assignments, for which, in the presence of small samples, block randomization would be required to achieve an adequate balancing of the major confounders. For these reasons, other empirical strategies may ensure a higher external validity to the analysis, and/or a better balancing of the crucial controls, than discontinuity designs. These strategies could be, for example, in terms of: enlarging the focus to units of observations farther away from the cut-off (or policy-change border); using theory of change to explicitly identify the crucial observable confounders; excluding units of observation based on the lack of common support for the relevant controls (instead on the distance from the cut-off or policy-change border); and adopting a statistical matching estimator to mimic a block randomization design that ensures the actual balancing of the crucial confounders also in the presence of small sample sizes.

In the case of business incentive programs, finally, the existence of discontinuities for the treatment assignment, based on cut-offs in the ranking of applicants, is undoubtedly a desirable impact identification scenario that enables the analysis to focus exclusively on (treated and non-treated) applicants that share the same crucial unobservables (in terms, for example, of the firm's position on its investment cycle, desire to invest and managerial abilities). Also for these policies, however, in noticeable circumstances, the sample sizes of applicant firms are often small and the running variable (represented by the application score) is a weak confounding factor. Consequently, natural experiment conditions may extend further away from the cut-off. In this scenario, similarly to the case of the regional-based policies, cross-cut-off comparisons or regression discontinuity designs may suffer from small-sample limitations that may lead to failures in the balancing of the crucial observable controls. While, instead, a preferable empirical strategy, also in this case, can be in terms of enlarging the focus to applicant firms farther away from the cut-off; using theory of change to identify the major observable confounders; excluding applicants based on the lack of common support of the control variables; and using matching estimators to ensure a more transparent and explicit balancing of the crucial controls, mimicking a block randomization design.

References

- Accetturo A. & De Blasio G. (2012). Policies for local development: An evaluation of Italy's Patti Territoriali. *Regional Science and Urban Economics*, 42 (1–2), 15–26
- Becker, S. O., Egger P. H., & von Ehrlich M. (2010). Going NUTS: The effect of EU Structural Funds on regional performance, *Journal of Public Economics*, 94, 578–590.
- Becker, S.O., Egger, P. H., & von Ehrlich, M. (2012). Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review*, 56(4), 648–668.
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (2013). Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy*, 5(4), 29–77.
- Becker, S. O., Egger, P. H., & von Ehrlich, M. (2016). Effects of EU regional policy: 1989–2013. *Working Paper Series No. 271. Coventry: Centre for Competitive Advantage in the Global Economy*, University of Warwick.
- Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2), 577–99.
- Bondonio D. (2016). *The impact of varying per capita intensities of EU Funds on regional growth: Estimating dose response treatment effects through statistical matching*. Technical Report , DG -Regional and Urban Policy, European Union, Bruxelles, 2016. ISBN: 978-92-79-58513-5; doi: 10.2776/256991.
- Bondonio D., Fernandes T.F. & Paes Mamede R. (2016). *Does EU Public Support to Firm Investments Boost High Quality Jobs? Evidence from Linked Employer-Employee Microdata and Natural-Experiment Conditions*. Dinamia'Cet, ISTCE-IUL working paper n. 2016/01.
- Bondonio, D. & Martini, A. (2018). Are they worth it? A Counterfactual Impact Evaluation of a Decade of Investment Subsidies to Italian Firms, *Italian Journal of Regional Science* ISSN1720-3929 (In press).
- Bruhm M. and McKenzie D. (2008). *In Pursuit of Balance: Randomization in Practice in Development Economics*. World Bank, Policy Research Working Paper n. 4752.
- Card D. & Kreuger A. (1994). Minimum Wages and Employment: A Case Study of the New Jersey and Pennsylvania Fast Food Industries. *American Economic Review*, 84(4), 772–793.
- Card D. & Kreuger A. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *American Economic Review*, 90(5)5, 1397–1420.
- Cerqua A. & Pellegrini G. (2014). Do subsidies to private capital boost firms' growth? A multiple regression discontinuity design approach. *Journal of Public Economics*, 109, 114–126.
- Crescenzi R. & Guia M. (2018). *One or Many Cohesion Policies of the European Union? On the Diverging Impacts of Cohesion Policy across Member States*. SERC –Spatial Economic Research Centre Discussion Paper n. 230.
- De Blasio G. & Poy S. (2017). The impact of local wage regulation on employment: a border analysis from Italy in the 1950s. *Journal of Regional Science*, 57(1), 48–74.

- Dube A., W. T. Lester & M. Reich (2010), Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties. *The Review of Economics and Statistics*, 92(4), 945–964.
- Einio E. & Overman, H. (2012). *The Effects of Spatially Targeted Enterprise Initiatives: Evidence from UK LEGI*. ERSA conference papers, European Regional Science Association.
- Ferrara A. R., McCann P., Pellegrini G., Stelder D. & Terribile F. (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*, 96(4), 817–841.
- Giacomelli S. & Menon C. (2013). *Firm size and judicial efficiency: evidence from the neighbour's Court*. Banca D'Italia Working Paper n. 898.
- Gobillon I. Magnac T., & Selod H. (2012), Do Unemployed Workers Benefit from Enterprise Zones? The French Experience. IZA DP No. 6357.
- Giulia M. (2017). Spatial discontinuity for the impact assessment of the EU regional policy: the case of Italian Objective 1 regions. *Journal of Regional Science*, 57(1), 109–131.
- 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.
- Jofre-Monseny J. (2014). The Effects of Unemployment Protection on Migration in Lagging Regions. *Journal of Urban Economics*, 83, 73–86.
- Keele L. J. & Titiunik R. (2015). Geographic Boundaries as Regression Discontinuities. *Political Analysis*, 23(1), 127–155.
- Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. and Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, 52, 19–26.
- Lee D.S. & Lemieux T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48, 281–355.
- Magnatti P., F. Ramella, C. Trigilia & G. Viesti (2005). *Patti Territoriali: lezioni per lo sviluppo*. Il Mulino, Bologna
- Neumark D. & Wascher W. (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment. *American Economic Review* 90(5), 1362–1396.
- Pellegrini G., Terribile F., Tarola O., Muccigrosso T. & Busillo F. (2013). Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach. *Papers in Regional Science*, 92(1), 217–233.
- Percoco M. (2017). Impact of European Cohesion Policy on regional growth: does local economic structure matter?. *Regional Studies*, 51(6), 833–843
- Rathelot R. and P. Sillard (2009). Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements?. *Economie et Statistique*, 415–416, 81–96.