Assessing the impacts of Cohesion Policy on EU regions

Ferrara, Antonella; McCann, Philip; Pellegrini, Guido; Stelder, Dirk; Terribile, Flavia

Published in:
Papers in Regional Science

DOI:
10.1111/pirs.12234

doi:10.1111/pirs.12234

Assessing the impacts of Cohesion Policy on EU regions: A non-parametric analysis on interventions promoting research and innovation and transport accessibility*

Antonella Rita Ferrara[1,2], **Philip McCann**[1], **Guido Pellegrini**[3], **Dirk Stelder**[4], **Flavia Terribile**[5]

[1] Faculty of Spatial Sciences, Department of Economic Geography, University of Groningen, Groningen, the Netherlands (e-mail: antonellarita.ferrara@unical.it, p.mccann@rug.nl)
[2] Department of Economics, Statistics and Finance, University of Calabria, Rende, Italy
[3] Department of Social and Economic Sciences, Sapienza University of Rome, Rome, Italy (e-mail: guido.pellegrini@uniroma1.it)
[4] Faculty of Economics and Business, Department of Economics, Econometrics & Finance, University of Groningen, the Netherlands (e-mail: t.m.stelder@rug.nl)
[5] Gran Sasso Science Institute (GSSI), Social Science Evaluation Unit, L'Aquila, Italy (e-mail: flavia.terribile@gmail.com)

Received: 25 May 2015 / Accepted: 14 March 2016

Abstract. Traditionally, the effectiveness of European Cohesion Policy has been evaluated in terms of GDP growth rate. In this paper, we consider the effect of the regional policy in terms of its impacts on two specific fields of intervention, namely 'research, technological development and innovation', and 'transport infrastructure'. Our econometric approach involves the use of a non-parametric regression discontinuity design technique to a uniquely-disaggregated Cohesion Policy dataset broken down according to the specific objectives of each stream of funding. The analysis considers different time intervals and sub-samples. Our results demonstrate a positive impact of Cohesion Policy interventions in these two specific fields of intervention.

JEL classification: O18, O47, C21, R11

Key words: EU Cohesion Policy, regression discontinuity design, research and innovation, transport infrastructure

1 Introduction

The aim of this paper is to assess the effectiveness of European regional policy interventions, or EU Cohesion Policy, during the 'programming' period 2000–2006, in improving both research and innovation activities and transport accessibility. Transport and research and innovation in Objective 1 regions accounted, respectively, for around 26 per cent and 5.4 per cent of total Cohesion Policy expenditure in the period 2000–2006 and they represent two of the key thematic

* We wish to thank the editor and two anonymous reviewers for their valuable comments on a previous version of this paper.

objectives of the new programming cycle 2014–2020 (European Commission 2007, 2014). Our objective is twofold. First, we demonstrate the use of non-parametric or semi-parametric techniques in assessing the effectiveness of Cohesion Policy in different domains in which it operates. Second, our objective is also to account for the lagged effects of the policy, taking into consideration that spending resulting from policy over this programming period has been completed only in 2008 (European Commission 2007) and that Structural Funds payments might be effective after some time lag (Rodríguez-Pose and Fratesi 2004; Mohl and Hagen 2010; Hagen and Mohl 2011). In dealing with this issue, we will consider longer time intervals for the outcome variables in order to keep trace of both the short and the long run impacts. Moreover, for making the sample more stable, we consider the eligibility status of the regions for two programming periods (1994–1999 and 2000–2006), this means that each region remained in the same group for at least two programming periods.

Previous researches employing non-parametric or semi-parametric techniques in order to assess the impacts of EU Cohesion Policy have focused primarily on the effects on the growth rate of regional GDP (Hagen and Mohl 2008; Manzella and Mendez 2009; Becker et al. 2010, 2012; Pellegrini et al. 2013), leaving largely unexplored the particular impacts on specific fields of intervention. However, recent research emphasising a broader multidimensional understanding of social progress (Sen 1999, 2006; Acemoglu et al. 2005; Stiglitz et al. 2009; Tabellini 2010; Fitoussi 2013; UNDP 2013) justifies an approach which employs other measures of economic development which are closely linked to the objectives of the specific interventions. For the first time, we are able to exploit an original dataset which provides comparable intervention-specific information at the NUTS 2 regional level for EU-15 member states in order to investigate a wider range of impacts of Cohesion Policy than those typically considered in literature so far. The areas of intervention we investigate are 'research, technological development and innovation' (RTDI), and 'transport infrastructure' (TI) and our analysis explores the impacts of policy interventions for the period 1999–2010 for RTDI and 2000–2012 for TI. In order to do this we apply the Regression Discontinuity Design technique (Thistlethwaite and Campbell 1960; Hahn et al. 2001), which is a non-experimental method for comparing the performance of different groups of observations. The rules governing the eligibility for the Objective 1 of EU Cohesion Policy funding and interventions provide a natural discontinuity,[1] the features of which can be examined using non-parametric techniques. Our results regarding specific types of policy actions suggest that the impacts of specific policy interventions do generate positive impacts of a type intended by the policy interventions. However, the strength of these results appears to differ across the policy domains. The rest of the paper is structured as follows. Section 2 provides a brief background to the structure and logic of EU Cohesion Policy and also provides an explanation of how the policy is amenable to regression discontinuity design method we employ. Section 3 discusses the methodological strategy and defines the response variables. Section 4 presents the results of our analysis and section 5 provides some brief conclusions and directions for future research. Details on the construction of the dataset and the technique are described in the Appendix.

2 The structure and logic of EU Cohesion Policy and its suitability for RDD approaches

The priorities of EU Cohesion Policy are fixed by means of the definition of the Structural Funds objectives which were classified prior to 2007 according to different regional categories denoted

[1] The eligibility for this specific policy objective allows us to identify two different groups of regions – 'treated' (with a *per capita* GDP level, measured in purchasing power standards, just below the 75% threshold) and 'untreated' regions (those just above the 75% threshold) – under the assumption that regions close to the cut-off point share the same characteristics and a difference in the growth rate of the outcome is considered as a causal effect of the treatment.

as: Objective 1, Objective 2 and Objective 3. The Objective 1 regions were the least economically developed regions in the EU and the policy logic aimed to promote the development and the structural adaptation of these lagging regions (McCann and Ortega-Argilés 2013a, 2013b). The funding allocated to these regions consisted of almost 70 per cent of total allocations for the Structural Funds for the period 2000–2006 (it was 68% in 1994–1999), amounting to some €136 billion. The recipients of this aid are identified by the Commission through the 'GDP criteria' whereby the Objective 1 aid is devolved only to the regions that have a *per capita* GDP (in purchasing power standards) which is lower than the 75 per cent of the community average. Much smaller allocations are made to the other types of regions, as they are economically more developed. The subsequent programming periods 2007–2013 and 2014 onwards, all experienced some changes in these allocation rules, but basically this system is still intact. Importantly, for our purposes, the 75 per cent Objective 1 funding allocation rule means that the policy is amenable to regression discontinuity techniques.

In terms of the impacts of the policy, there are now more than fifty studies analysing the effects of European Cohesion Policy on EU regions, of which between approximately two thirds and three quarters of these papers find either positive effects or positive but mixed effects on the recipient regions, while the remaining quarter find either negligible or even negative effects (McCann 2015). Almost all of these existing studies use regional GDP growth as a synthetic indicator of regional performance and most of these empirical assessments typically fall into one of two kinds of approaches, namely: a classical regression approach where growth equation models are estimated; and the more recent literature based on the treatment effect techniques. These treatment effects techniques aim to set up counter-factual frameworks, and recent years have witnessed an explosion of studies that evaluate public policies with these counterfactual types of methods. These studies generally adopt non-experimental methodologies based on the idea that the eligibility criteria to a specific objective of the policy could itself be considered as a treatment effect – in a manner analogous to the treatments given to medical patients. In these types of cases for such an approach to be workable it is necessary for us to be able to identify two different groups of regions with comparable characteristics and to assign them as being 'treated' or 'untreated', according to whether they have received policy assistance or not. Where this is possible, this distinction itself allows us to evaluate the causal effect of the treatment using the design and logic of policy as the basis for the assignment of the treatment. The 75 per cent Objective 1 financial allocation rule provides such a discontinuity and thereby makes the policy amenable to these types of testing procedures (Hagen and Mohl 2008; Becker et al. 2010, 2012, 2013; Pellegrini et al. 2013; Gagliardi and Percoco 2013; Accetturo et al. 2014). Our aim is to identify whether the regions which qualified for Objective 1 support experienced a greater growth in certain specific policy-outcome dimensions than the non-Objective 1 which did not quality for financial support. Two important assumptions underlie the application of RDD in our analysis: the regions close to the cut-off point share the same characteristics, also in terms of the level of the concurrent policies, except for the (binary) treatment[2]; differences in the growth rates of the outcome variable between Objective 1 and non-Objective 1 regions are higher than differences in the amount of transfers they received.[3] This could occur not only because the expenditure is – on average – higher for treated regions (see, for instance Tables A1 and A2 in the Appendix), but also because the programming framework and expenditure procedures under Structural Funds can be more effective.

New techniques of regression discontinuity design have been developed recently in other fields of geographically-related research dealing with issues of education (Black 1999), labour markets (Dell 2010), real estate markets (Dachis et al. 2012), firm size (Giacomelli and Menon

[2] Accordingly, close to the cut-off, one can easily put apart any confounding factor by comparing the units belonging to the treated and non-treated groups.

[3] It is a binary treatment setting, independent of treatment intensity.

2012) and firm incentives (Einiö and Overman 2012). These approaches are commonly known as 'spatial regression discontinuity design' or 'spatial RDD' approaches and they consider the geographical location as the key forcing variable. In these cases, the discontinuity which is to be exploited by the econometric technique is given by the administrative or geographical boundaries and the sub-samples to be examined are the spatial units on either side of the geographical boundary. In the case of EU regional policy evaluation, in some countries the regions falling into the Objective 1 and in the non-Objective 1 groups, respectively, can be simply identified by looking at the geographical boundaries. However, this is not true for all countries, with the consequence that the effect of the policy for the treated regions that have a good performance but which are located far from the geographical boundaries may be rather underestimated. For this reason, in our analysis, we decided to use a classical RDD approach.

In this paper, our goal is to try to move the policy impact attention away from the traditional measures of effectiveness to some more specific indicators of the efficacy of policy interventions. With this aim in mind, we decided to take into account, in addition to *per capita* GDP, two different aspects of the social and economic development of regions, namely research and innovation and also transport accessibility. RTD and innovation and transport infrastructure were two of the main policy areas during the programming period 2000–2006 and they still represent two of the key thematic objectives for the programming period 2014–2020, as will be seen hereafter. The main challenge is the identification of possible outcome variables and the availability of data at NUTS 2 level. In particular, rather than observing *per capita* GDP growth as is typically the case, instead here we examine the regional impacts on patent applications as a proxy for regional innovation improvements in response to research and innovation funding support, and also the impacts on potential road accessibility in response to the funding of regional infrastructure improvements (Stelder 2016).

Considering the structure of spending of Cohesion Policy in the period 2000–2006, investment was concentrated in three main areas: infrastructure (mainly transport and the environment), productive investment (largely small and medium-sized enterprises (SMEs) and RTDI) and investment in people. In particular, transport in Objective 1 accounted for by 26 per cent of total expenditure, whereas RTDI was slightly lower and accounted for approximately 5.4 per cent (European Commission 2007). Transport infrastructure is one of the main areas of investment of Cohesion Policy: an efficient transport system is a key factor underlining regional competitiveness and growth. Over the period 2000–2006, the 47 per cent of the total spending on transport went on motorways and other roads. Analogously, consolidation of regional innovation system is a potentially important factor in fostering the competitiveness of regions. In Objective 1 regions, Cohesion Policy represents an important contribution in strengthening national R&D and innovation systems. Overall, Cohesion Policy can boost development by investing in second nature determinants of growth: public capital stock, accessibility, human capital, innovation, institutional quality and agglomerations (European Commission 2014).

Considering the Cohesion Policy funding by broad policy areas in EU-15, in the programming period 2000–2006, for the less developed regions, the highest share of funding was allocated to infrastructure (transport, energy, telecom and social infrastructure) and to business support (including RTDI) with a share over the total funding of 30.9 per cent and 28 per cent, respectively. In the new programming period 2014–2020, European Regional Development Fund (ERDF) resources are concentrated on support for R&D and Innovation (about 22% of ERDF the total allocations) and transport and energy infrastructure (14%; European Commission 2014).

Regional accessibility is also generally considered to be an essential prerequisite for regional economic development. According to the Territorial Agenda of the European Union mobility and accessibility are key prerequisites for economic development of all regions of the EU and transport infrastructure improvement is a key policy instrument to promote regional economic development (ESPON 2006). During the first 15 years of its existence the European Regional Development Fund devoted 80 per cent of its funding to infrastructure projects (Vickerman 1991) and over the period 2000–2006 about 35 per cent of the Structural Funds and 50 per cent of the Cohesion Fund has been spent on infrastructure projects (Crescenzi and Rodríguez-Pose 2008). In his theoretical and empirical overview, Ottaviano (2008) stresses the importance of the network character of the spatial economy in which accessibility and market potential are decisive for innovation, regional spill-overs and productivity.

3 The methodology and the response variables

The fundamental hypothesis underpinning RDD method is that the units just above (or under) the threshold that do not receive the treatment, represent a very good comparison group for those just under (or above) the threshold that do receive the treatment. Therefore, any discontinuity in the conditioned expected value of the outcome in the immediate proximity of the cut-off point can be interpreted as an evidence of a causal effect of the treatment. As such, an advantage for geographical research is that these techniques allow us to gain a deeper understanding of causality relationships (Overman 2013) in ways which are not always possible with spatial econometric frameworks.

For the application of the RDD technique to work, four basic assumption need to be complied with (Lee and Lemieux 2009):

- the treatment is not randomly assigned, but there is at least one observable variable known as an ‘assignment variable’ or a ‘forcing variable’;
- the assignment variable presents a discontinuity corresponding to a particular threshold;
- the assignment variable cannot be manipulated in that agents cannot modify it in order to move from one side to the other of the threshold;
- the other variables are regular functions without any discontinuities corresponding to the cut-off point, such that the only factor that produces a jump at the threshold is the discontinuity in the treatment effect itself.

The impacts of European regional policy are captured here by means of a sharp RDD technique that can help to isolate it from other factors that may affect the analysis’ results (Thistlethwaite and Campbell 1960) such as the effects of geographical location or externalities. This methodology considers a discontinuity in the treatment related to some observations in order to obtain an estimation of the average treatment effect (ATE) by comparing some units eligible for the treatment (Objective 1 regions) with others which are not eligible (non-Objective 1 regions). The effect of the treatment estimated is localized in the point of discontinuity.

In our analysis, the statistical units examined are the NUTS 2 regions of the EU-15 member states all of which were part of the European Union both in the programming periods 1994–1999 and in 2000-2006.[4] Even though, the Eastern enlargement of the EU (in 2004) by 10 countries, matters for the programming period 2000–2006 (Becker et al. 2010), for the new members it is too early to determine trends in the expenditure and in the evaluation of the transfers’ impacts (European Commission 2007). This policy framework provides a good context for the application of the RDD approach. In our analyses, the forcing variable we adopt is the *per capita* GDP (in PPS) of the region and the cut-off point is the 75 per cent threshold which defines whether a region is eligible for the ‘treatment’ of Objective 1 funding support. For example,

[4] EU-15 includes: Germany, France, Italy, Netherlands, Belgium, Luxembourg (founding countries); Denmark, Ireland and United Kingdom (1973); Greece (1981); Spain and Portugal (1986); Austria, Finland and Sweden (1995).

if a NUTS 2 region A exhibits a *per capita* GDP level which is equal to 74.99 per cent of EU average it will be eligible for Objective 1 policy funding and will receive the treatment whereas if a NUTS 2 region B with a *per capita* GDP which is equal to 75.01 per cent of the EU average will not be eligible for the treatment. Given that these two regions are so close in terms of productivity levels, it is perfectly reasonable to suppose that these two regions have very similar characteristics except for the treatment, and as such, they are much better comparators than other regions which are more distant from the cut-off threshold (Becker et al. 2010).

Following the approach of Pellegrini et al. (2013) we adopt a sharp version of the RDD, since the assignment to the treatment is assumed to be binary and to depend only from the assignment rule of 75 per cent.[5] Moreover, to support this assumption we will exclude from the sample the regions that receive any aid for other reasons. Consider c the cut-off point and X_i the forcing variable. We can denote the potential outcome of the region i with $Y_i(0)$ and $Y_i(1)$, where $Y_i(1)$ is the outcome obtained in presence of the treatment (Objective 1 regions) and $Y_i(0)$ is the outcome obtained by the non-treated regions (non-Objective 1). Given the covariates, and corresponding to the discontinuity point, the conditioned expectation of the outcome underlining the causal effect of the treatment is given as (Imbens and Lemieux 2008):

$$\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]. \tag{1}$$

If the average causal effect of the treatment is taken into consideration the above relationship become:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c], \tag{2}$$

where τ is the discontinuity of the outcome variable estimated in proximity of the cut-off point.

In order to increase the robustness of the results, the estimation will be obtained using both a parametric and non-parametric approach and verifying the results for different samples, specifications, kernels and confidence intervals. The aim here is to avoid any problems related to the limited number of observations in proximity of the cut-off point, which could reduce the accuracy of the estimations. For the non-parametric estimation we use the local linear regression method with standard errors obtained via the bootstrap method, whereas we use the parametric regression as a robustness check, applying the ordinary least square (OLS) estimation with standard errors robust to heteroscedasticity.

In the RDD approach three statistical choice issues are important to consider. First, the choice of the kernel is important, with some authors preferring certain types of kernels over others. Here we will opt for more than one specification for the kernel and will employ the Epanechnikov kernel, the Gaussian kernel, Rectangular kernel and the Triangular kernel. Second, another important element is the choice of the bandwidth (Bw). There are many rules of thumb regarding the choice of the optimal bandwidth. Different bandwidths produce different estimations, so it is important to estimate more than one bandwidth and across at least three scales, namely, the optimal bandwidth, its double and its half. The wider the bandwidth the stronger will be the discontinuity, because the impact of possible erratic observations close to the threshold will become smaller. For the choice of the optimal bandwidth, the index of Imbens and Kalyanaraman (2009) is calculated and this index determines the asymptotic optimal interval for the regression discontinuity. Third, it is also important to test that there are no jumps in the levels of the treatment and of the outcome and that other covariates (not affected by the treatment) do not have any discontinuities in the cut-off point. In order to verify the first point, the effect is estimated for different thresholds and with different kernels and bandwidths, and in

[5] Sharp RDD is opposed to fuzzy RDD which supposes to have a continuous treatment variable.

order to verify the second point, we consider the population average using a local linear regression with different kernels.

As a first robustness check a parametric approach is applied, the equation of a generic polynomial model of m order is[6]:

$$Y = \alpha + \tau D + \sum_{i=1}^{m} \beta_i X^i + \sum_{i=1}^{m} \delta_i D X^i + \varepsilon, \tag{3}$$

Y is the annual average growth rate of the outcome variable considered, D is a binary variable identifying the Objective 1 regions, τ is the coefficient of the estimated discontinuity and X is the forcing variable.

In the parametric approach, the equivalent of bandwidth's choice is the definition of the polynomial order (i) of the regressions (3) (Lee and Lemieux 2009). Different specifications are considered in order to analyse how the polynomial degree affects the results. The best polynomial order is chosen by looking at the Akaike information criterion (AIC): the best model is the one with the lowest AIC.

Following Imbens and Lemieux (2008), Lee and Lemieux (2009) and Pellegrini et al. (2013), two additional robustness checks are implemented here. We verify if in the density function of X, for $X = c$, there are other discontinuities that may reveal an alteration in the control variable and we also investigate the presence of other discontinuities in the outcome variable. Moreover, in order to exclude any gerrymandering (Menon 2012) type of manipulation in the proximity of the threshold with respect to the continuity of the density function of the forcing variable, the McCrary test is used (McCrary 2008). The McCrary test estimates the density function of *per capita* GDP for a confidence interval of 95 per cent, and McCrary (2008) suggests that a jump in the conditional density of the forcing variable can be considered as a test on its manipulability. Under such conditions when regions are sorted around the threshold, the RDD approach is not applicable.

In order to undertake our estimations we can exploit a uniquely-detailed dataset on the certified expenditure from 1999 to 2007 of the EU-15 regions. As such, with these data we are able to know which regions received the funding transfers and which specific fields of policy interventions (FOI) these funds were allocated and used for. The advantages of using these types of actual expenditure data on specific funding activities over and above purely eligibility criteria is emphasized by Aiello and Pupo (2012) and De La Fuente (1995) who point out that the potential impacts will be related to the funds actually spent and not simply to those which are programmed or committed. Our use of certified expenditure data avoids all of these difficulties. Moreover, in each specific area of intervention, we refer to those specific outcome variables closely related to the intended objectives of the interventions, and the samples we refer for each of these outcome variables are different because we analyse only regions which certified transfers in these specific fields of policy interventions (FOI). In order to test the robustness of the results our analysis is also conducted with different specifications of the outcome variables, including growth rates and differences in levels. The construction of the dataset and the samples used in the analysis are described in details in the Appendix. Moreover, the use of the regression discontinuity design approach allows us to rule out the problems associated with the choice of a specific functional form, which typically occurs in classical growth equation-type models.

For our certified expenditure data we consider the field of intervention (FOI) 'research, technological development and innovation (RTDI)' for Structural Funds and 'technical assistance' (TA) for Cohesion Funds. All the regions that have a positive TA are already included in the RTDI sample. The identification of the best time frame to consider is not an easy task, however we believe, following the main literature (Rodríguez-Pose and Fratesi 2004; Hagen and Mohl

[6] We consider $m = 3$.

2008, 2011) and the European Commission's Reports (2007, 2014), that even though there are some interventions with an immediate impact, the policy requires longer time intervals than the programming period to be effective. For this reason, we considered at least three years over the end of the programming period and when possible we split the whole period in different sub-periods. In the case of the patent applications, the whole period considered in the analysis covers from 1999 to 2010, although we also split the time interval of the analysis into three sub-periods: 1999–2007; 2002–2010; 2002–2007.[7] For the certified expenditure relating to transport infrastructure we consider the FOI 'transport infrastructure' both for Structural and Cohesion Funds. All the NUTS 2 who received the Cohesion Funds also received Structural Transport Funds. However, the period covered cannot be split into sub-periods because data on potential road accessibility (hereafter POT) are available just for some specific years (1955, 1970, 1980, 1990, 2000, 2012). We therefore decide to consider the growth rate 2000–2012. Since it has a growth rate equal to zero and looks like an outlier, we exclude Reunion from the sample. The next step is the identification of an appropriate outcome variable for each field of intervention.

In terms of an appropriate outcome response variable for regional research and innovation interventions, we choose to employ data on patent application counts. Patent applications are probably the most widely used variable for assessing progress in research and innovation activities, and both the strengths and the limitations of this variable are also very well understood.[8] In particular, we choose patent applications per million regional inhabitants from the OECD *Regpat* dataset as our response variable. Following the OECD (2009) Patent Manual recommendations, we use a fractional accounting system for patents which attributes to each region its actual contribution to the invention and when summed over all regions gives a total of 100 per cent. Patent data can be regionalized considering the address of either the inventor or the holder, although the inventor's address usually indicates where the invention was made. The priority year is the year of first filing for a patent as it is the closest to the actual date of invention, and should therefore be used as the reference date when compiling patent indicators aimed at reflecting technological improvements (Maraut et al. 2008). In constructing our outcome response variable we consider fractional count, by inventor and priority year patent data. The *Regpat* database used includes patent applications to the European Patent Office (EPO), to the Patent Cooperation Treaty (PCT) and to the United States Patent and Trademark Office (USPTO).[9] In the first step, we consider the whole sample and the outcome variable is expressed as both growth rate and difference in levels. In a second step, we also consider some restricted samples. In our estimations: we use parametric (OLS) and non-parametric estimations (local linear polynomial estimation with standard errors estimated with bootstrap method for 500 replications).

In terms of an appropriate outcome variable associated with the transport infrastructure expenditure we employ a measure of accessibility changes for each region at each time period. Following Batty (2009) and Reggiani (2012) the concept of accessibility is generally viewed as the relative nearness or proximity of one place or person to all other places and persons and we use it in the same way from the accessibility database constructed by Stelder (2016)

[7] We excluded alternatively the first three years, the last three years and both, in order to assess if the results are related to the time span considered.

[8] See, among others Saxenian (1996), Storey and Tether (1998), Malecki (2007) and Paci and Marrocu (2013).

[9] For OECD patent data missing values are treated as being equal to zero. However, when data are not available at NUTS2 level we used the Eurostat variable 'Employment in technology and knowledge-intensive sectors by NUTS regions and sex' (1994–2008, NACE Rev. 1.1) for the calculation of the weight (countries involved: Greece, Belgium, France Outre-Mer, Germany, Netherlands, England) which let us to transform the national statistics in data suitable for imputation at regional level. Only for Greece and Cumbria the Eurostat data are not available: in these cases the data imputed are, respectively, the average NUTS 1 value (NUTS 1 value/nr. of NUTS 2) and the mean of the other NUTS 2.

which is also adopted by the European Commission (2014) in its assessments of regional road transport accessibility.

Typically, accessibility indicators take the form of a two-variable function *f(a,b)* with *a* being the activity to be reached elsewhere and *b* indicating the costs to reach that activity. In spatial economics, the most used functional form is inspired by the original gravity approach of Reilly (1931):

$$A_i = \sum_j P_j D_{ij}^{-\beta}, \tag{4}$$

with A for accessibility, P for the local activity to be reached, D for distance or any other definition of transport costs, and parameter β indicating the distance decay intensity. In Stelder (2016) this is implemented with population P and travel time D_{ij} in minutes using a newly constructed database of historical European road networks over the period 1960–2012.

The absolute accessibility A_j is scaled to relative accessibility a_j:

$$a_j = \frac{A_i}{\sum_j A_j}. \tag{5}$$

For each location accessibility may be increasing at the same ratio, which may cause additional economic growth, but uniform for all locations, with the consequence that no one is benefiting more than others from infrastructure improvement. Therefore, we use the change in relative accessibility α_i derived as:

$$\Delta\alpha_i(t) = \frac{\alpha_i(t)}{\alpha_i(t-1)}, \tag{6}$$

and with this transformation, we are also able to eliminate the usual geographical bias that gives central locations the highest accessibility (Stelder 2016). In our estimations for the change in accessibility associated with transport infrastructure investments we use non-parametric estimation techniques for a local linear polynomial estimation with standard errors estimated with bootstrap method for 1,000 replications and parametric (OLS) estimations with robust standard errors.

In both of the cases of innovation and transport infrastructure funding our goal is to identify whether the treated regions that received and spent the EU transfers for these specific fields of intervention were associated with a greater growth in the specific outcome variables relating to the impacts of these transfers. As mentioned before, the samples used are different for each specific FOI, because not all the units received transfers in both sectors of intervention.[10]

Our analysis and its robustness checks are undertaken by referring to two main issues, namely the time intervals being considered and also the sample composition. In terms of timing issues, the behaviour of the outcome variable, when looking at patent applications, is considered for the whole period (1999–2010) and also for other three sub-periods 1999–2007, 2002–2010 and 2002–2007. Our initial screening of the dataset has shown the presence of some possible outliers, so we need to see if the results found for the whole sample are associated with the inclusion or exclusion of outliers.[11] As we see below our results are robust

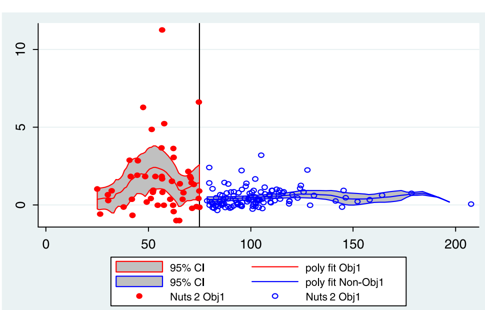
[10] Some regions are eliminated from the transport analysis because their values are missing: Notio Aigaio, Kriti, Ciudad Autónoma de Ceuta, Ciudad Autónoma de Melilla, Canarias, Região Autónoma dos Açores, Região Autónoma da Madeira.

[11] We consider a first sub-sample R1 that excludes Martinique, Guyane, Regio-Autonoma des Açores, Ciudad Autonoma de Melilla and Ciudad Autonoma de Ceuta that have some missing values and negative growth rate. In the second sub-sample, R2 we decide to exclude Alentejo that seems to be an outlier in the whole period, because it has the highest growth rate, though it has a clear increasing trend. Its elimination may give more stability to the results obtained.

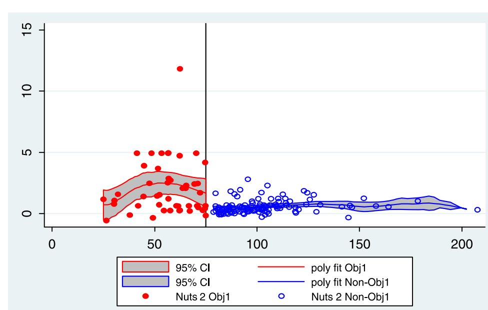
to both the time intervals being considered and also the sample compositions. First, we look at the graphic analysis of the discontinuity, then we estimated the discontinuity for different kernels and different bandwidths with a non-parametric approach and finally we control for different robustness checks. Once we have looked at the discontinuity in different thresholds, we estimate the polynomial regressions with OLS and we control, as a further robustness check, for the presence of a discontinuity considering a different outcome variable – the average population – that should not be affected by the treatment.

4 The results

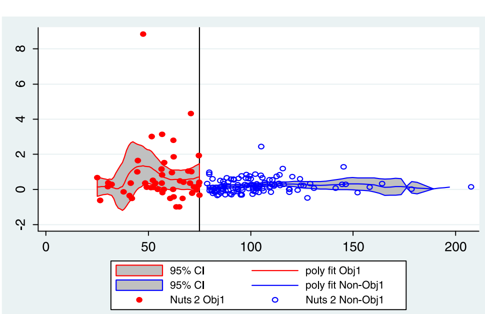
The effects of regional policy expenditures on research and innovation activities in Objective 1 regions in comparison to the equivalent performance of non-Objective 1 regions is captured by our outcome indicator of patent applications per million. These effects are depicted in Figures 1a–d each which refer to the different time periods, and in which the outcome variable is represented as a function of the forcing variable which is the level of *per capita* GDP in PPS (EU-15 = 100, average 88–90) for both the groups. The vertical line plotted in the graphs is the cut-off point at 75 per cent threshold; the units on the left are the Objective 1 regions; the units on the right are the non-Objective 1 regions. As we see Figure 1a provides the first evidence of the presence of a discontinuity when the whole sample is considered, for the period 1999–2010. This means that the average growth rate in patent applications for the Objective 1 regions is indeed greater than the growth rate of the untreated regions. As we see in Figures 1b–d a similar picture arises when we consider each



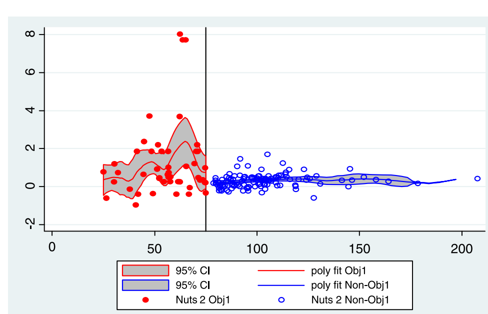
(a) 1999–2010



(b) 1999–2007

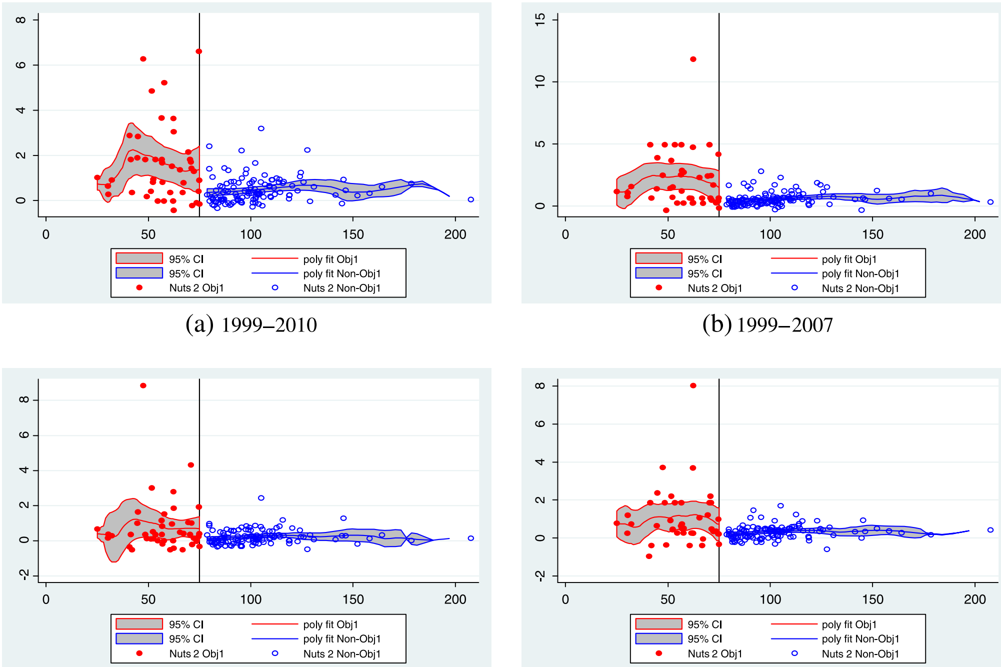


(c) 2002–2010



(d) 2002–2007

Fig. 1. Comparison of the growth rate in patent applications between the Objective 1 and non-Objective 1 regions, whole sample



(a) 1999–2010 (b) 1999–2007

(c) 2002–2010 (d) 2002–2007

Fig. 2. Comparison of the growth rate in patent applications between the Objective 1 and non-Objective 1 regions, sample R2

of the time sub-periods.[12] The graphic analysis confirms the discontinuity also when focusing on the restricted samples R1 and R2. In particular, for the sample R2, Figures 2a–d confirm these results: once again, the regions on the left exhibit a higher growth rate than the regions on the right, independently of the time interval considered.

All of these results strongly support the presence of discontinuity in favour of the treated regions (Lee and Lemieux 2009) which is independent of the time-period being considered, although there appears to have been a slightly greater impact during the earlier periods of the policy interventions. The observed discontinuity is now estimated using the RDD approach with a local linear regression estimation and standard errors estimated with bootstrap (500 replications). Tables 1–4 report the results of these estimations on the whole sample for four different types of kernels (triangle, rectangular, Gaussian and Epanechnikov) and three bandwidths (optimal, half and double). The optimal bandwidth is obtained through the index of Imbens and Kalyanaraman (2009) that provides the optimal trade-off between precision (greater number of observations) and distortion (wider interval, greater differences among treated and untreated regions). For all the four time periods considered, the results are statistically significant for both the Gaussian and the Epanechnikov kernels.

In particular, when the whole period is considered (Table 1) the discontinuity is around 1 percentage point and is statistically significant at 10 per cent for the optimal bandwidth. It

[12] Preliminary evidence of discontinuity can be obtained by considering a *naïve* estimation of the difference between the annual average growth rate of the treated and non-treated regions. For the whole period there is a statistically significant difference in favour of the Objective 1 regions (at the 1% level) represented by a positive coefficient equal to 1.07 with standard error 0.22. This value becomes 1.45 with standard error 0.21 and is still significant at 1 per cent if the last three years are excluded (1999–2007) and decreases to 0.49 (standard error 0.16) and 0.92 (standard error 0.19), respectively, if the periods 2002–2010 and 2002–2007 are considered.

Table 1. Growth rate of patent applications, whole sample, period 1999–2010, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
6.12 (optimal)	–2.942	–7.093	–0.997*	–0.997*
	(41.22)	(50.18)	(0.574)	(0.583)
3.06	0	0	–0.781	–0.636
	(0)	(0)	(0.745)	(0.825)
12.25	–0.414	0.231	–1.269**	–1.331***
	(1.492)	(1.326)	(0.517)	(0.504)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 2. Growth rate of patent applications, whole sample, period 1999–2007, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
6.67 (optimal)	–0.635	–2.776	–1.439**	–1.426**
	(27.21)	(22.14)	(0.574)	(0.606)
3.34	0	0	–1.089*	–1.106*
	(0)	(0)	(0.586)	(0.623)
13.35	–0.625	–0.705	–1.736***	–1.803***
	(0.955)	(0.807)	(0.573)	(0.559)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3. Growth rate of patent applications, whole sample, period 2002–2010, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
3.19 (optimal)	0	0	–0.480	–0.467
	(0)	(0)	(0.351)	(0.342)
1.59	0	0	–0.446	–0.422
	(0)	(0)	(0.413)	(0.432)
6.39	0.297	–0.631	–0.573*	–0.603**
	(7.869)	(13.71)	(0.309)	(0.285)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. **$p < 0.05$, *$p < 0.1$.

becomes approximately 1.3 percentage points and statistically significant at 1 per cent (kernel Epanechnikov) and 5 per cent (kernel Gaussian), when the bandwidth is doubled.

If the last three years are excluded so we focus only on 1999–2007 in Table 2 the discontinuity increases both in size and significance such that it is 1.4 percentage points and statistically significant at 5 per cent with an optimal bandwidth, and increases to 1.8 (significant at 1%) with double the bandwidth. Furthermore, for the half bandwidth we observe a significant (10%) discontinuity of about 1 percentage point.

When we focus on the period 2002–2010 (Table 3) a significant discontinuity is found only for double the bandwidth and it is equal to 0.6 percentage points and statistically significant at 5 per cent for the Epanechnikov kernel.

Table 4. Growth rate of patent applications, whole sample, period 2002–2007, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
6.72 (optimal)	0.355	–0.216	–1.279**	–1.343**
	(3.979)	(5.885)	(0.512)	(0.531)
3.36	0	0	–0.775*	–0.861*
	(0)	(0)	(0.433)	(0.461)
13.45	0.106	0.163	–1.402***	–1.424***
	(0.292)	(0.568)	(0.517)	(0.518)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

As we see in Table 4 if the period 2002–2007 is considered, the discontinuity is statistically significant for all the bandwidths and it is equal to 1.3 percentage points (with 5 percent significance) for the optimal bandwidth, and 0.8 percentage points (1 percent significant) for the half bandwidth and 1.4 percentage points (1 percent significance) for the double bandwidth.

Tables 5–8, referred to the restricted sample R2, confirm that our results are also found to be robust to the inclusion or exclusion of specific individual outliers.[13]

The discontinuity trend related to bandwidth dimensions can be analysed by looking at Figures 3a–c for the Epanechnikov kernel and Figures 4a–c for the Gaussian kernel. In both cases, figures show a clear jump of the outcome variable in proximity of the threshold.

Table 9 shows the parametric estimations (OLS with robust standard errors) on the restricted sample R2. Model 5 was chosen as the best model using the AIC. The effect of the Regional Policy was positive and statistically significant at 5 percent and equal to 3.6 annual percentage points. The selected model presents one linear term and one quadratic term. The most similar results to the non-parametric regression was the estimation of model number 4, in which the effect was of 1.15 percentage points.

We also check how the observed responses of the outcome variable change around the discontinuity if the outcome variable is expressed in terms of a levels variable reflecting the absolute number in patent applications per million rather than as a growth rate in patents per million inhabitants as considered above. Figure 5 depicts the results for the whole sample, and as we see there is no observable discontinuity and no difference between the levels of patents per million which are applied for between Objective 1 and the non-Objective 1 regions closely situated around the 75 per cent threshold. Not surprisingly, as we see in Table 10 there is also no statistical difference when estimated using a flexible polynomial for different kernels and these results all hold irrespective of the time-period being considered.[14]

Figure 6 represents the conditional density discontinuity of the forcing variable regional GDP *per capita* computed with the method of McCrary (2008) for a 95 per cent confidence interval. As we see, the discontinuity around the cut-off is not statistically significant at 5 per cent, so the assignment to the treatment determined by the eligibility for the Objective 1 status cannot be easily predicted on the basis of gerrymandering types of issues and there are good reasons for this.[15] This gives us further confidence that our results reported above are robust to the sample composition.

[13] Results for sample R1 are available on request from the authors.
[14] Results are available on request from the authors.
[15] Although we might suspect that some countries may behave opportunistically by maintaining their *per capita* GDP below the threshold in order to capture funds. However, in reality this cannot happen because the threshold is fixed at 75 per cent of *per capita* GDP community average and this is known only after the publication of all the regional data. Moreover, Eurostat ensures that there are very strict controls on the procedures for the estimation of regional accounts.

Table 5. Growth rate of patent applications, sample R2, period 1999–2010, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
3.44 (optimal)	0	0	–1.061*	–1.072**
	(0)	(0)	(0.548)	(0.509)
1.72	0	0	–0.957	–0.969
	(0)	(0)	(0.717)	(0.676)
6.88	–0.443	–4.191	–1.161**	–1.184**
	(28.07)	(26.17)	(0.481)	(0.460)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. **$p < 0.05$, *$p < 0.1$.

Table 6. Growth rate of patent applications, sample R2, period 1999–2007, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
6.77 (optimal)	–0.484	–2.776	–1.397***	–1.433**
	(27.90)	(24.64)	(0.539)	(0.586)
3.39	0	0	–1.149**	–1.213**
	(0)	(0)	(0.563)	(0.592)
13.54	–0.607	–0.725	–1.551***	–1.586***
	(0.991)	(0.945)	(0.534)	(0.566)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$.

Table 7. Growth rate of patent applications, sample R2, period 2002–2010, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
3.08 (optimal)	0	0	–0.571	–0.597*
	(0)	(0)	(0.352)	(0.354)
1.54	0	0	–0.530	–0.576
	(0)	(0)	(0.392)	(0.429)
6.16	–0.0499	–0.631	–0.600*	–0.613**
	(10.37)	(11.91)	(0.320)	(0.308)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. **$p < 0.05$, *$p < 0.1$.

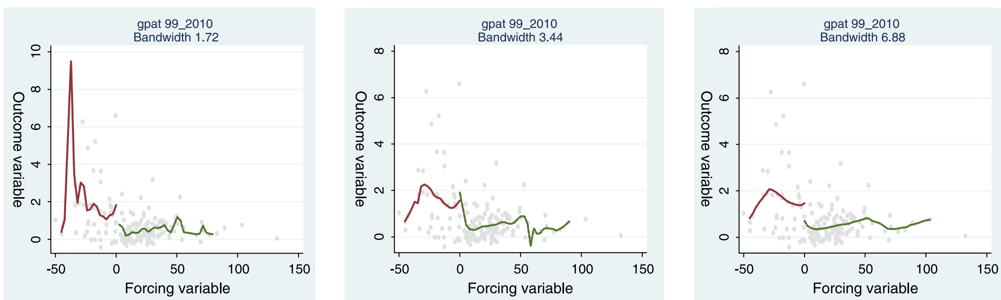
Taken together our results suggest that there has been a strong and statistically significant effect of the research and innovation policy expenditures in Objective 1 regions which has increased their growth rates in innovation-related activities to the extent that these weaker regions now perform at the more or less same levels as those regions which are economically stronger than they are. Although there is some evidence to suggest a greater effect in earlier years, these results are robust to the time period being considered and to the samples being considered.[16]

[16] The analysis was also conducted by considering the number of people employed in technology and knowledge-intensive sectors as outcome variable and looking to the field of expenditure on human resources but not significant results have been obtained; in particular, for the latter there were not enough units in proximity of the threshold.

Table 8. Growth rate of patent applications, sample R2, period 2002–2007, non-parametric estimations with different kernels and bandwidths

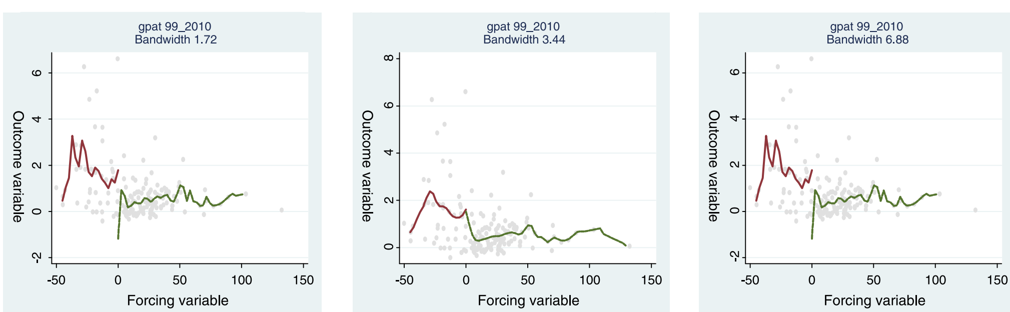
Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
5.55 (optimal)	–0.353	0	–0.792**	–0.838**
	(4.408)	(0)	(0.348)	(0.382)
2.77	0	0	–0.639**	–0.698**
	(0)	(0)	(0.290)	(0.349)
11.09	0.0168	–0.0879	–0.819**	–0.827**
	(0.329)	(4.389)	(0.366)	(0.387)

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. $^{**}p < 0.05$.



(a) half bandwidth (b) optimal bandwidth (c) double bandwidth

Fig. 3. Robustness check: Epanechnikov kernel, different bandwidths, cut-off = 0
Notes: Outcome variable: patent applications growth rate (1999–2010), forcing variable (GDP *per capita* in PPS (75% EU15 = 0), 1988–1990).



(a) half bandwidth (b) optimal bandwidth (c) double bandwidth

Fig. 4. Robustness check: Gaussian kernel, different bandwidths, cut-off = 0
Notes: Outcome variable: patent applications growth rate (1999–2010), forcing variable (GDP *per capita* in PPS (75% EU15 = 0), 1988–1990).

If we now consider the case of transport infrastructure investments, from the descriptive statistics presented in Table A2 in the Appendix we see that the transport infrastructure results are likely to be different from the case of the patent applications, because the difference between the mean values of the two groups is lower and the treated group is characterized by a much higher variability. A first evidence of the discontinuity is given from a *naïve* estimation in the annual average growth rate of the outcome variable, which is equal to 1.01 (standard error 0.23) and is statistically significant at 1 per cent. This means that the Objective 1 regions in average grow more than the non-Objective 1 of one percentage point per year.

Table 9. Parametric estimations with different polynomial orders

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	–9.02e–05***		1.77e–05	7.18e–07	0.000258**	0.000215**	0.000396	0.000260
	(2.52e–05)		(1.70e–05)	(0.000149)	(0.000108)	(0.000102)	(0.000623)	(0.000617)
X^2				1.45e–10	–6.94e–09**	–5.72e–09**	–1.59e–08	–8.26e–09
				(4.14e–09)	(2.84e–09)	(2.62e–09)	(3.33e–08)	(3.29e–08)
X^3							0	0
							(0)	(0)
Obj1		1.107***	2.081*	1.149**	3.624**	0.917	1.855	–4.468
		(0.258)	(1.165)	(0.451)	(1.415)	(1.948)	(3.866)	(6.770)
DX			–0.000104		–0.000234	0.000437	0.000290	0.00280
			(0.000139)		(0.000154)	(0.000493)	(0.000687)	(0.00253)
DX^2						–4.09e–08	–3.51e–08	–3.57e–07
						(3.32e–08)	(3.68e–08)	(3.19e–07)
DX^3								0
								(0)
Constant	1.923***	0.473***	0.219	0.431	–1.738*	–1.392	–2.407	–1.645
	(0.379)	(0.0542)	(0.258)	(1.265)	(0.948)	(0.905)	(3.718)	(3.684)
Observations	160	160	160	160	160	160	160	160
R–squared	0.096	0.195	0.204	0.198	0.212	0.219	0.220	0.227
AIC	478.30712	461.66263	462.02673	461.31083	460.47844	460.89608	462.86836	461.36874

Source: Estimations on European Commission and OECD data.
Notes: Standard errors in parentheses. The dependent variable is the annual average growth rate in patent applications (1999–2010); X = Gdp *per capita* in PPS (EU-15 = 100, average 1988–1990), D = Objective 1 dummy variable; robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

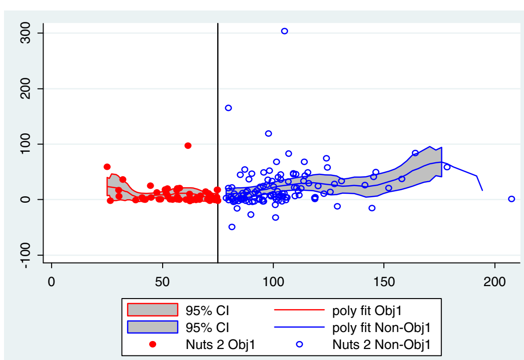


Fig. 5. Comparison of the difference in levels in patent applications between the Objective 1 and non-Objective 1 regions, whole sample (1999–2010)

This result is illustrated in Figure 7. As we see here is no unambiguous discontinuity jump in the proximity of the cut-off point is observable, although the patterns of dispersion do point very much towards this conclusion.[17] These observations are confirmed in Table 11 which shows that

[17] The position of the dots in the scatter plot implies that this finding is due to a heterogeneous composition of the treated group. Inspection of the data reveals that the patterns of dispersion of the Objective 1 group is comprised of two sub-groups of regions, one of which consists mainly of Spanish and Portuguese regions and which has a markedly higher growth rate than the other group which is mainly comprised of Germany, Italy and Greece. These findings are in line with the results of the Fourth Report on Economic and Social Cohesion (European Commission 2007) which underline that over this period there was a considerable road construction in Spain and Portugal.

Table 10. Difference in levels of patent applications, whole sample, period 1999–2010, non-parametric estimations with different kernels and bandwidths

Bw/Kernel	(1) tri	(2) rect	(3) gau	(4) epa
12.98 (optimal)	36.64	38.41	–2.332	–4.752
	(60.91)	(43.73)	(10.92)	(10.40)
6.49	1.954	–167.4	6.955	11.16
	(3,407)	(2,406)	(20.09)	(21.51)
25.96	6.930	12.02	1.035	2.493
	(22.51)	(16.41)	(7.826)	(7.391)

Source: Estimations on DG Regional Policy data and Stelder (2016) data.
Notes: Standard errors in parentheses. $*p < 0.01$.

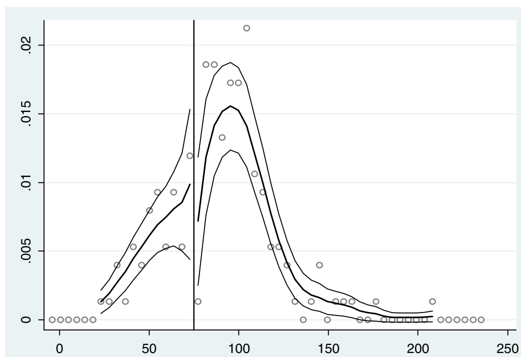


Fig. 6. Estimation of the density function of the forcing variable (Gdp *per capita* in PPS, average 1988–1990) at the threshold, whole sample

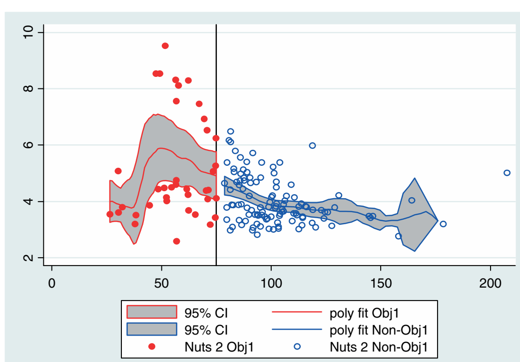


Fig. 7. Comparison of the growth rate in potential road accessibility between the Objective 1 and non-Objective 1 regions (2000–2012)

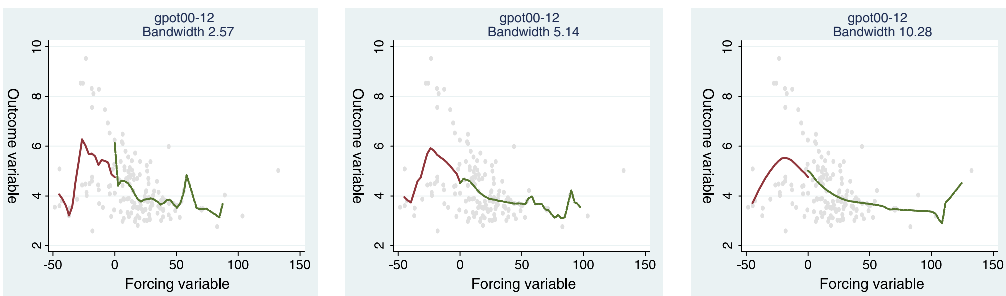
the results are statistically significant (10%) using both the Gaussian kernel and the Epanechnikov kernel and for double the bandwidth.

In both cases the discontinuity is equal to 0.9 percentage points, and the standard errors are estimated using bootstrap with 1000 replications. Figures 8a–c show the discontinuity trend with the Epanechnikov kernel in relation to bandwidth size. It is clear in sections *b* and *c* of the graph. Figures 9a–c display the discontinuity in relation to bandwidth size when the Gaussian kernel is considered. In this case, the jump in proximity with the cut-off point is visible also with the half bandwidth, but the high variability of the treated regions does not allow for any

Table 11. Growth rate of potential road accessibility, period 2000–2012, non-parametric estimations with different kernels and bandwidths

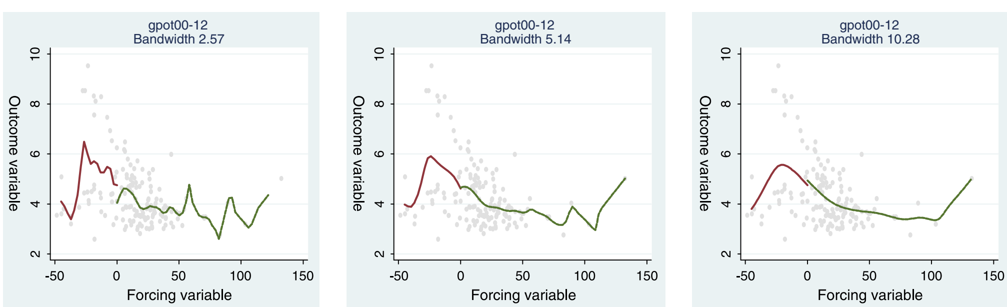
	(1) tri	(2) rect	(3) gau	(4) epa
5.14 (optimal)	1.567	0	–0.462	–0.604
	(8.411)	(0)	(0.523)	(0.560)
2.57	0	0	0.0945	0.192
	(0)	(0)	(0.570)	(0.630)
10.28	–0.381	0.236	–0.839*	–0.901*
	(0.891)	(22.78)	(0.471)	(0.489)

Source: Elaborations on DG Regional Policy data and Stelder (2016) data.
Notes: Standard errors in parentheses. *$p < 0.1$.



(a) half bandwidth (b) optimal bandwidth (c) double bandwidth

Fig. 8. Robustness check: Epanechnikov kernel, different bandwidths, cut-off = 0
Notes: Outcome variable: potential road accessibility growth rate (2000–2012), forcing variable (GDP *per capita* in PPS (75%EU15 = 0), 1988–1990).



(a) half bandwidth (b) optimal bandwidth (c) double bandwidth

Fig. 9. Robustness check: Gaussian kernel, different bandwidths, cut-off = 0
Notes: Outcome variable: potential road accessibility growth rate (2000–2012), forcing variable (GDP *per capita* in PPS (75%EU15 = 0), 1988–1990).

significant estimation. As a robustness check, we also ran the parametric estimation with a different polynomial order (Table 12).[18]

As with the patent data, for our regional accessibility measure we also performed a levels analysis rather than a growth rate analysis, and again this also produces no statistically significant differences around the threshold levels.[19] Similarly an estimation of the density function of

[18] The results show a problem of strong multicollinearity; indeed, from model 3 onwards, the variance inflation factor (VIF) assumes a value higher than 20 and blows up in models 5 and 6.
[19] Results are available on request from the authors.

Table 12. Parametric estimations with different polynomial orders

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	–0.000121***		–7.49e–05*	–0.000129	–0.000518***	–0.000633***	–0.000563	–0.000800
	(3.73e–05)		(3.87e–05)	(0.000200)	(0.000157)	(0.000140)	(0.000845)	(0.000847)
X^2				2.12e–09	1.28e–08***	1.61e–08***	1.22e–08	2.56e–08
				(5.95e–09)	(4.30e–09)	(3.63e–09)	(4.50e–08)	(4.52e–08)
X^3							0	–0
							(0)	(0)
Obj1		1.144***	–0.334	0.629	–3.182*	–10.91***	–10.55*	–22.79**
		(0.313)	(1.334)	(0.445)	(1.684)	(2.714)	(5.500)	(9.119)
DX			0.000121		0.000361**	0.00229***	0.00223**	0.00709**
			(0.000137)		(0.000161)	(0.000732)	(0.00104)	(0.00331)
DX^2						–1.18e–07**	–1.16e–07**	–7.36e–07*
						(4.90e–08)	(5.68e–08)	(4.10e–07)
DX^3								0
								(0)
Constant	5.832***	4.014***	5.083***	5.411***	8.690***	9.630***	9.240*	10.56**
	(0.531)	(0.0789)	(0.561)	(1.634)	(1.369)	(1.259)	(5.068)	(5.077)
Observations	150	151	150	150	150	150	150	150
R–squared	0.126	0.151	0.166	0.160	0.187	0.237	0.237	0.255
AIC	481.43973	481.62804	478.50243	477.47501	474.61663	467.13464	469.1317	465.54912

Source: Estimations on DG Regional Policy data and Stelder (2016) data.

Notes: Standard errors in parentheses. The dependent variable is the annual average growth rate in potential transport accessibility (2000–2012); X = Gdp *per capita* in PPS (EU-15 = 100, average 1988–1990), D = Objective 1 dummy variable; robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

the forcing variable (McCrary 2008) provides no evidence of manipulation, and alongside the fact that our sub-samples provide basically the same results gives us confidence that our observed results are indeed associated with the policy interventions, and are not related to other issues.[20]

Both for patent applications and for potential road accessibility (POT), another robustness test was undertaken to verify whether there are no jumps in the level of the outcome when the threshold is not identified. The model was tested for a null effect for different values of the forcing variable, with different kernels (Epanechnikov and Gaussian) and the optimal bandwidth for different thresholds (50, 60, 70, 90). The results confirm that there are no significant discontinuities. Moreover, we verified that there is no discontinuity at the cut-off point for another covariate that should not be affected by the treatment: we considered the average population. The estimations were carried out with a non-parametric local linear regression with three kernels (Gaussian, Epanechnikov and Rectangular) with the optimal bandwidth and standard errors computed with bootstrap. The results confirm the absence of any significant discontinuity.

Taken together, therefore, we can conclude on the basis of the results obtained here, that transport infrastructure investments under Cohesion Policy have indeed increased the accessibility of recipient regions to a greater level than non-treated regions. However, the discontinuity observed regarding the accessibility improvements afforded by transport infrastructure is less robust than the results obtained for the patent applications. This may be due to the fact that our outcome variable considers only road accessibility and associated improvements in other transport infrastructures are not examined here.[21]

5 Conclusions

The regression discontinuity design (RDD) approach we have employed in this paper is increasingly argued to be very well adapted to answering these types of policy performance-evaluation questions, and especially in a regional context (Overman 2013; Gibbons et al. 2014). However, the novelty of the analysis undertaken here is that it also demonstrates that the RDD technique is also appropriate for analysing different sub-fields of intervention within a single broad policy framework. Such an exercise has not previously been undertaken before, and the results provided here suggest that different individual policy domains exhibit different performance outcomes, even when the individual policy domains operate under a common umbrella programming logic. Our results provide robust evidence that Cohesion Policy interventions do lead to types of results desired by the policy, however in order to calibrate the scale of these different effects across different policy domains additional analyses will be required. For example, these types of techniques could be applied using different outcome indicators for intervention fields such as innovation (Capello and Lenzi 2013), transportation or any of the other domains in which the policy operates. Further research extending this analysis will also involve the use of dose response functions (Imbens 2000; Becker et al. 2012) allowing for the fact that differences in the amount of resources devoted to the various specific fields of intervention may be one of the main elements leading to heterogeneity in the treatment effects (Dotti 2013). Such approaches may also help to better identify differing optimal resource allocations in the individual policy sub-fields covered by EU Cohesion Policy, although doing this properly will also

[20] Results are available on request from the authors.

[21] However, in the period 2000–2006 about the 47 per cent of the total spending on transport went to roads (European Commission 2007). Variables that consider the accessibility in other transport networks are not available, but we expect that without considering those variables we are just underestimating the discontinuity results. The analysis was also done considering the 'kilometres of road, railway and navigable way', but without any significant results.

involve translating marginal discontinuity effects into monetary values of the outcome variables, issues which require a great deal of additional complexity and analysis.

Appendix

Source and the dataset

The construction of the dataset can be divided into three steps. Following Pellegrini et al. (2013), the first step aims at the definition of a sample that satisfies the hypothesis of the sharp RD design and allow us to have regions included in the same group for two consecutive 'programming' periods (1994–1999 and 2000–2006). The second and the third steps are aimed at obtaining a panel structure for the dataset of the certified expenditure for the NUTS 2 regions and the transformation of the outcome variables. The dataset consists of EU-15 regions at NUTS 2 level with the Objective 1 recipient regions of the transfers being those NUTS 2 regions with a *per capita* GDP (in PPS) lower than the 75 per cent of the community average. For the programming period 1994–1999, the Commission computed the eligibility threshold considering the data on *per capita* GDP of the period 1988–1990 (per-capita GDP in PPS, ESA79 criteria). Therefore, we consider the *per capita* GDP of the period 1988–1990 when constructing the forcing variable. Our initial sample includes 213 regions classified as NUTS 2 in 2003, of which 61 were Objective 1 regions in the programming period 1994–1999 and the remaining 152 were non-Objective 1 regions. In order to make the sample more homogeneous over the two programming periods, we also excluded from the initial group of Objective 1 regions four of the NUTS 2 regions, namely Hainaut (BE), Corse (FR), Molise (IT), Lisboa (PT). The reason is that in the period 1988–1990 (which is the referring period of the Commission for establishing the eligibility to the funds) these regions each experienced a level of *per capita* GDP which was greater than 75 per cent of the EU average, but for other political reasons they became eligible for the funds. The remaining 57 regions also stayed eligible for the Objective 1 status in the following programming period 2000–2006. At the same time, in order to have a more comparable and stable control group we also decided to exclude from our sample those regions that were eligible for Objective 1 funding in the period 2000–2006, but which were not eligible in the previous period[22] and we also need to be cognizant of the fact that some regions which were not eligible for Objective 1 funding also benefited of the Cohesion Policy transfers because they fell into other categories of policy objectives.

Following Pellegrini et al. (2013), we take into account the *per capita* intensity of the financial resources among the different regions, distinguishing between regions which were hard-financed (Objective 1, treated regions) and regions which were soft-financed (non-treated regions). As many sources of EU financing exist (Structural Funds, Cohesion Fund, national co-financing, private financing) both in the programming periods 1994–1999 and 2000–2006, we need to identify a threshold value of *per capita* transfer intensity. We fixed this threshold at €1,960, which is the minimum value of certified *per capita* expenditure in Objective 1 regions (Pellegrini et al. 2013). Our dataset show that nine non-Objective 1 regions had a level of *per capita* expenditure higher than this threshold, and in particular we excluded the non-Objective 1 Spanish regions that received aid from the Cohesion Fund and also the Finnish regions that benefited of other funds.[23] Finally, we excluded the regions that did not receive transfer in the

[22] These are: the five regions who were non treated in 1994–1999, but they become eligible for Objective 1 in 2000–2006: Burgenland (AT), Itä-Suomi (FI), South Yorkshire (UK), Cornwall and Isles of Scilly (UK), West Wales and the Valleys (UK); and also the five regions which were non-Objective 1 in the period 1994–1999, but they become partially eligible in 2000–2006. These are: Länsi-Suomi (FI), Pohjois-Suomi (FI), Norra Mellansverige (SE), Mellersta Norrland (SE), vre Norrland (SE).

[23] Pais Vasco, Comunidad Foral de Navarra, La Rioja, Aragòn, Comunidad de Madrid, Cataluña, Illes Balears and also the Finnish regions of Etelä-Suomi and Åland.

FOI of certified expenditures selected.[24] Thus, our final cleaned sample consists of 180 NUTS 2 regions – of which 54 are treated and 126 are untreated regions – which stayed in the same group for both of the programming periods considered in the analysis. Moreover, these also represent homogeneous groups of soft-financed (untreated) or hard-financed (treated) funds in terms of the amounts of per-capita transfers received. Thus, our sample fits with the features required for the application of the regression discontinuity design in the sharp version.

Our data on the certified expenditure comes directly from the European Commission offices (DG-Regional and Urban Policy) and from the Italian Ministry of Economic Development (Department for the Development and Economic Cohesion). We selected two specific FOI (level 2) for the Structural Funds, namely research and innovation and also transport infrastructure[25]; as regards the Cohesion Fund we choose technical assistance project and transport project. The data did not originally have a panel structure as the Structural Funds and the Cohesion Fund are reported in two different tables and identifying regions from funding codes was a laborious task. Thus it was necessary to transform the data before undertaking any econometric analysis. The dataset was constructed manually, observing the following rules: (*i*) the total amount is fully imputed to the region if the name of the region is expressly and univocally specified in the identification name of the program; (*ii*) the expenditure of programs for NUTS at a lower level than NUTS 2 are imputed to the respective NUTS 2 region; (*iii*) the expenditure of national programmes are shared between all the regions of the country, using the population at the beginning of the programming period as a distribution criteria[26]; (*iv*) the imputation of the expenditure of municipality union programmes, natural regions and consortium to the NUTS 2 involved in the group (when identifiable), using the same criteria of the previous point[27]; (*v*) the expenditures for which recipient regions cannot be identifiable from the name of the programme, were deleted; and (*vi*) data about cross border and interregional cooperation were not considered. After these preliminary transformations, the dataset exhibited a panel structure containing for each NUTS 2 region the certified expenditure by year and the fund and field of intervention.

Descriptive statistics

Table A1 shows the descriptive statistics for the patent applications and for RTDI *per capita* expenditure for the whole sample and the two groups (treated and untreated regions). The sample consists of 167 units, of which 50 regions are Objective 1 and 117 are non-Objective 1. The maximum growth rate in patent applications is 11.25 (Alentejo) and the minimum is –1 (Ceuta and Melilla). Its average value is 0.80 (standard deviation 1.39). If we look at each group separately the result is quite different, the average growth rate in patent applications for the treated regions is 1.55 while for the untreated is 0.47; both the maximum and the minimum values of the growth rate of patent applications are accounted for by Objective 1 regions. Looking at the *per capita* expenditure, there is still a sharp demarcation amongst the two groups in that the mean value is three times higher for the Objective 1 than the non-Objective 1. The descriptive statistics confirm that the two groups show different performances both in the outcome variable and in the certified expenditure levels.

Meanwhile, Table A2 provides the descriptive statistics for the growth rate of potential road accessibility index (POT), for both the groups and for the certified expenditure in transport infrastructure per area expressed in square kilometres.[28] For the potential road accessibility data,

[24] Bruxelles, Provincia di Trento, Prov. Brabant Wallon, Prov. Vlaams Brabant, Bedfordshire, Hertfordshire, East Anglia, Eastern Scotland, Usimaa-Helsinki

[25] In particular: 18. Research, technological development and innovation, RTDI; 31. Transport Infrastructure.

[26] Otherwise, the first year available is used.

[27] For the municipalities associations of Portugal, for which there were specific web sites, the expenditure is attributed to the NUTS 2 of the headquarter of the association.

[28] Given its location far outside of Europe we have removed Reunion from the sample.

Table A1. Descriptive statistics: growth rate of patent applications and RTDI

	N	Minimum	Maximum	Mean	Std. Deviation
Growth patent applications (1999–2010)	167	–1.00	11.25	0.80	1.39
Growth patent applications (1999–2010) – Non Objective 1	117	–0.34	3.19	0.47	0.59
Growth patent applications (1999–2010) – Objective1	50	–1.00	11.25	1.55	2.22
RTDI *per capita* expenditure	167	9.08	1903403.00	47370.94	168203.20
RTDI *per capita* expenditure – Non Objective 1	117	9.08	728218.60	28949.93	76922.34
RTDI *per capita* expenditure – Objective 1	50	44.78	1903403.00	90476.08	281306.90

Source: Elaborations on DG Regional Policy data and OECD regpat data.

Table A2. Descriptive statistics: growth rate of potential road accessibility and transport infrastructure expenditure

	N	Minimum	Maximum	Mean	Std. Deviation
Growth potential road accessibility (2000–2012)	151	2.60	9.50	4.30	1.28
Growth potential road accessibility (2000–2012) –Non Objective 1	113	2.80	6.50	4.01	0.84
Growth potential road accessibility (2000–2012) –Objective 1	38	2.60	9.50	5.15	1.87
Transport infrastructure expenditure per Km^2	151	2.36	1335975.00	41223.87	143833.20
Transport infrastructure expenditure per Km^2 –Non Objective 1	113	3.93	887265.40	39386.84	109579.20
Transport infrastructure expenditure per Km^2 –Objective 1	38	2.36	1335975.00	46686.60	217903.50

Source: Elaborations on DG Regional Policy data and Stelder (2016) data.

there are no complete time series data available, but rather just for some specific years, so it is not possible to consider different sub-periods. We therefore refer to the period 2000–2012. The minimum growth rate for POT is 2.6 and the maximum is 9.5 for the regions of Calabria (IT) and Norte (PT), respectively, both of which are Objective 1 regions. The mean value for the treated group is 5.15, whereas for the untreated it is 4.01, while for the treated group the standard deviation is almost double that of the untreated group. Looking at the expenditure, also for this FOI, the mean value is higher for the treated group (46,686) than for the untreated (39,386).

References

Accetturo A, de Blasio G, Ricci L (2014) A tale of an unwanted outcome: Transfers and local endowments of trust and cooperation. *Journal of Economic Behavior & Organization* 102: 74–89

Acemoglu D, Johnson S, Robinson JA (2005) Institutions as a fundamental cause of long-run growth. *Handbook of Economic Growth* 1: 385–472

Aiello F, Pupo V (2012) Structural funds and the economic divide in Italy. *Journal of Policy Modeling* 34: 403–418

Batty (2009) Accessibility: In search of a unified theory. *Environment and Planning B: Planning and Design* 36: 191–194

Becker SO, Egger PH, von Ehrlich M (2010) Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics* 94: 578–590

Becker SO, Egger PH, von Ehrlich M (2012) Too much of a good thing? On the growth effects of the EU's regional policy. *European Economic Review* 56: 648–668

Becker SO, Egger PH, 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: 29–77

Black SE (1999) Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114: 577–599

Capello R, Lenzi C (2013) Territorial patterns of innovation and economic growth in European regions. *Growth and Change* 44: 195–227

Crescenzi R, Rodríguez-Pose A (2008) Infrastructure endowment and investment as determinants of regional growth in the European Union. *European Investment Bank Papers* 13: 62–101

Dachis B, Duranton G, Turner MA (2012) The effects of land transfer taxes on real estate markets: Evidence from a natural experiment in Toronto. *Journal of Economic Geography* 12: 327–354

de la Fuente A, Vives X, Dolado JJ, Faini R (1995) Infrastructure and education as instruments of regional policy: Evidence from Spain. *Economic Policy* 10: 13–51

Dell M (2010) The persistent effects of Peru's mining mita. *Econometrica* 78: 1863–1903

Dotti NF (2013) The unbearable instability of Structural Funds' distribution. *European Planning Studies* 21: 596–614

Einiö E, Overman H (2012) The effects of spatially targeted enterprise initiatives: Evidence from UK LEGI. Working Paper, London School of Economics, London

ESPON (2006) *Accessibility, transport and communication networks – Thematic study of INTERREG and ESPON activities*. European Spatial Planning Observation Network (ESPON), Luxembourg

European Commission (2007) Fourth report on economic and social cohesion. *Growing regions, growing Europe.* Brussels (May)

European Commission (2014) *Sixth report on economic, social and territorial cohesion. Investment for jobs and growth*. Brussels, July

Fitoussi JP (2013) *Il teorema del lampione*. O come mettere fine alla sofferenza sociale, Einaudi, Torino

Gagliardi L, Percoco M (2013) *The impact of European cohesion policy in urban and rural regions: A regression-discontinuity approach*. University Bocconi, Milano, Working Papers

Giacomelli S, Menon C (2012) *Firm size and judicial efficiency in Italy: Evidence from the neighbour's tribunal*. Working paper, Spatial Economics Research Centre, LSE, London

Gibbons S, Nathan M, Overman HG (2014) Viewpoint: Evaluating spatial policies. *Town Panning Review* 85: 427–432

Hagen T, Mohl P (2008) Which is the right dose of EU Cohesion Policy for economic growth. Discussion Paper 08-104, ZEW, Mannheim

Hagen T, Mohl P (2011) Econometric evaluation of EU Cohesion Policy: A survey. In: Jovanovic MN (ed) *International Handbook on the Economics of Integration, Volume III*. Edward Elgar, Cheltenham

Hahn J, Todd PE, van der Klaauw W (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69: 201–209

Imbens GW (2000) The role of the propensity score in estimating dose-response functions. *Biometrika* 87: 706–710

Imbens G, Kalyanaraman K (2009) Optimal bandwidth choice for the regression discontinuity estimator. NBER Working Paper 14726

Imbens G, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142: 615–635

Lee DS, Lemieux T (2009) Regression discontinuity designs in economics. NBER Working Paper 14723

Malecki EJ (2007) Corporate organization of R&D and the location of technological activities. *Regional Studies* 41: S73–S88

Manzella GP, Mendez C (2009) *The turning points of EU Cohesion policy*. European Commission, Brussels

Maraut S, Dernis H, Webb C, Spiezia V, Guellec D (2008) The OECD Regpat database: A presentation. Sti Working Paper 2008/2, Statistical Analysis of Science, Technology and Industry

McCann P (2015) *The regional and urban policy of the European Union: Cohesion, results orientation and smart specialisation*. Edward Elgar, Cheltenham

McCann P, Ortega-Argilés R (2013a) Redesigning and reforming European regional policy. *International Regional Science Review* 36: 424–445

McCann P, Ortega-Argilés R (2013b) Transforming European regional policy: A results-driven agenda and smart specialisation. *Oxford Review of Economic Policy* 29: 405–431

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142: 698–714

Menon C (2012) The bright side of MAUP: Defining new measures of industrial agglomeration. *Papers in Regional Science* 91: 3–28

Mohl P, Hagen T (2010) Do EU structural funds promote regional growth? New evidence from various panel data approaches. *Regional Science and Urban Economics* 40: 353–365

OECD (2009) *Patent statistics manual*. OECD, Paris

Ottaviano GIP (2008) Infrastructure and economic geography: An overview of theory and evidence. *European Investment Bank Papers* 13: 8–35

Overman HG (2013) Geographical economics and policy. In: Fisher MM, Nijkamp P (eds) *Handbook of regional science*, volume 2. Springer, Berlin

Paci R, Marrocu E (2013) Knowledge assets and regional performance. *Growth and Change* 44: 228–257

Pellegrini G, Terribile F, Tarola O, Muccigrosso T, Busillo F (2013) Measuring the effect of European regional policy on economic growth: A regression discontinuity approach. *Papers in Regional Science* 92: 217–233

Reggiani A (2012) Accessibility, connectivity and resilience in complex networks. In: Geurs KT, Krizek KJ, Reggiani A (eds) *Accessibility analysis and transport planning*. Edward Elgar, Cheltenham

Reilly WJ (1931) *The law of retail gravitation*. Republished by Pilsbury Publishers, New York

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

Saxenian AL (1996) *Regional advantage*. Harvard University Press, Cambridge, MA

Sen A (1999) *Development as freedom*. Oxford University Press, New Delhi

Sen A (2006) Conceptualizing and measuring poverty. In: Grusky D, Kanbur R (eds) *Poverty and nequality*. Stanford University Press, Redwood City, CA

Stelder D (2016) Regional accessibility trends in Europe: Road infrastructure, 1957–2012. *Regional Studies* 50: 983–995

Stiglitz J, Sen A, Fitoussi JP (2009) The measurement of economic performance and social progress. URL: http://www.stiglitz-sen-fitoussi.fr/documents/overview-eng.pdf

Storey DJ, Tether BS (1998) Public policy measures to support new technology-based firms in the European Union. *Research Policy* 26: 1037–1057

Tabellini G (2010) Culture and institutions: Economic development in the regions of Europe. *Journal of the European Economic Association* 8: 677–716

Thistlethwaite DL, Campbell DT (1960) Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* 51: 309–317

UNDP (2013) 2013 Human Development Report. March. URL: http://www.undp.org

Vickerman R (1991) *Infrastructure and regional development*. Pion Limited, London

doi:10.1111/pirs.12234

Resumen. Tradicionalmente, la eficacia de la Política de Cohesión Europea ha sido evaluada en términos de la tasa de crecimiento del PIB. En este artículo se considera el efecto de la política regional en términos de sus impactos en dos ámbitos específicos de intervención: la "investigación, el desarrollo tecnológico y la innovación", y la "infraestructura de transporte". El planteamiento econométrico hizo uso de una técnica de diseño de regresión en discontinuidad no paramétrico aplicado a un conjunto de datos desagregado de política de cohesión, de acuerdo con los objetivos específicos de cada fuente de financiación. El análisis consideró diferentes intervalos de tiempo y submuestras. Los resultados muestran un impacto positivo de las intervenciones de la Política de Cohesión en estos dos ámbitos específicos de intervención.

抄録: European Cohesion Policy（EUの結束政策）の有効性は、慣例的に国内総生産の成長率で評価されてきた。本稿では、二つの介入領域、すなわち「研究、技術開発、イノベーション」および「輸送インフラ」に対するインパクトに関する地域政策の効果を考慮する。今回の計量経済学的アプローチでは、回帰不連続デザインによるノンパラメトリックな手法を、独自の方法でそれぞれの資金の目的によって内訳した結束政策のデータセットに用いる。分析では、複数の時間間隔とサブサンプルを考慮する。結果から、この二つの介入領域において結束政策による介入の正のインパクトが認められる。