CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

The effect of EU Structural Funds on regional performance

Master's thesis

Author: Bc. Barbora Žďárská

Study program: Economics and Finance

Supervisor: Petr Janský, Ph.D.

Year of defense: 2019

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 6, 2019

Barbora Zdarska

Abstract

The regional policy of the European Union is financed through a system of structural and investment funds, which allocates substantial amounts during each programming period to boost the regional growth. Currently, the regional policy uses almost a half of the European Union's budget. According to such an extensive investment plan it is believed that the structural funding has a positive impact on the regional performance. This thesis provides an analysis of the Objective 1 (Convergence strategy) treatment effect on the regional GDP and employment growth during two last programming periods 2000-2006 and 2007-2013 using mostly nonparametric estimation method of the regression discontinuity design. The thesis contributes to existing literature since the current research studies do not provide conclusive results. Based on the estimation results we did not find statistically significant effect of the Objective 1 treatment on the GDP per capita growth nor employment growth. These findings are robust to various model specifications and estimation methods.

JEL Classification	R11, R58, C21, C31
Keywords	the European Union, regional policy, the Objective 1, regression discontinuity design

Title	The effect of EU Structural Funds on regional performance

Author's e-mail	`zdarskabarbora@seznam.cz`
Supervisor's e-mail	`jansky@fsv.cuni.cz`

Abstrakt

V každém programovém období Evropská unie investuje významné částky na podporu regionálního růstu prostřednictvím strukturálních a investičních fondů regionální politiky. V současné době má regionální politika k dispozici téměř půlku rozpočtu Evropské unie. Tento rozsáhlý investiční plán je založen na předpokladu, že strukturální financování má pozitivní vliv na ekonomický vývoj regionů. Tato práce analyzuje vliv Cíle 1 (Konvergenční strategie) na růst HDP a zaměstnanosti v evropských regionech v průběhu období 2000-2006 a 2007-2013. Tato studie aplikuje primárně neparametrickou metodu regression discontinuity design a rozšiřuje tak současný výzkum, jehož zjištění se rozcházejí. Výsledek práce ukazuje, že nebyl nalezen statisticky významný vliv Konvergenční strategie regionální politiky na regionální růst. Tato zjištění jsou stabilní vůči změně různých parametrů modelu a metod estimace.

Klasifikace JEL	R11, R58, C21, C31
Klíčová slova	Evropská unie, regionální politika, Cíl 1, regression discontinuity design

Název práce	Vliv strukturálních fondů Evropské unie na regionální rozvoj

E-mail autora	zdarskabarbora@seznam.cz
E-mail vedoucího práce	jansky@fsv.cuni.cz

Acknowledgments

I would like to thank to my supervisor, Petr Janský, Ph.D., for his patient guidance and valuable advice. I would also like to express my grattitude to my family for their support during my studies.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Žďárská, Barbora: *The effect of EU Structural Funds on regional performance.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2019, pages 99. Advisor: Petr Janský, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

AIC Akaike information criterion

CER Coverage error

CF Cohesion Fund

EAFRD European Agricultural Fund for Rural Development

EAGGF European Agricultural Guidance and Guarantee Fund

EMFF European Maritime and Fisheries Fund

ERDF European Regional Development Fund

ESA The European System of National and Regional Accounts

ESF European Social Fund

ESI European Structural and Investment

EU European Union

Eurostat European Statistical Office

FRD Fuzzy Regression discontinuity

GDP Gross domestic product

GMM Generalized method of moments

IMSE Integrated Mean Squared Error Method

ITZZ Italian Extra-Regio

LATE Local average treatment effect

MSE Mean Squared Error

MV Mimicking Variance Method

NUTS Nomenclature of Territorial Units for Statistics

NZZ Dutch Extra-Regio

OLS Ordinary least squares

PCA Principal component analysis

PPS Purchasing power standard

RBC Robust bias corrected

RDD Regression discontinuity design

SEDI Socio-economic development

SME Small and medium-sized enterprises

SRD Sharp Regression discontinuity

YEI Youth Employment Initiative

2SLS Two-stage least squares

Master's Thesis Proposal

Author	Bc. Barbora Žďárská
Supervisor	Petr Janský, Ph.D.
Proposed topic	The effect of EU Structural Funds on regional performance

Motivation Regional policy of the EU is a policy focusing on the boost of the economic growth and on dealing with regional disparities. The Structural Funds and the Cohesion Fund, the main financial tools of the regional policy, support less developed regions using almost a third of the whole EU budget. The main objectives of the regional policy are Convergence, Regional Competitiveness and Employment, and European Territorial Cooperation. About 82% of the regional policy funding is concentrated on the Objective 1 (Convergence) to help regions whose GDP per capita is less than 75% of the EU average to revive.

Growing expenditures spent on narrowing of disparities between regions have attracted many researchers over last 20 years. Interestingly current surveys present different methods with different results. One possible explanation of these contradictions could be complicated separation of causal effect of Objective 1 treatment and other factors. Fortunately, modern approaches deal with this problem and present convincing results. Mohl and Hagen (2008) found positive but statistically insignificant impact on regional growth using a generalized propensity score estimator. Becker et al. (2012) also used the method of GPS but they came with a positive significant effect. A method of the Fuzzy Regression Discontinuity (FRD) design was used by Becker et. al (2010) resulting in higher GDP per capita growth of treated regions by about 1.6% while the employment effects have not been confirmed. Pellegrini et al. (2013) also found a positive impact on regional GDP per capita growth (0.6-0.9%) using a method of the Sharp Regression Discontinuity (SRD) design. The recent research published by Becker et al. (2016) contains also the last programming period 2007-2013 using only limited application of the method.

Most of aforementioned researches cover periods up to 2006. The aim of this thesis is to extend the analysis by another programming period 2007-2013 and to try to improve the analysis by parametric and non-parametric estimation with adjusting various characteristics of the model.

Hypotheses

Hypothesis #1: The cohesion policy of the EU has a positive impact on the economic growth of regions.

Hypothesis #2: The cohesion policy of the EU has a positive impact on employment growth of regions (using an improved model).

Hypothesis #3: The use of Objective 1 transfers is effective on average.

Methodology For measuring the effect of the cohesion policy the Regression Discontinuity Design (RDD) model would be used. The RDD is a quasi-experimental design dividing the observations into two groups based on their position- above or below a certain cut-off point (in this case the 75% of GDP per capita average in EU). The main idea is that observations close to the cut-off point could be used as a "treatment" and "control" group for an experiment. (Lee and Lemieux 2010) The RDD methodology was successfully used by Becker et. al (2010) and Pellegrini et al. (2013). Pellegrini et al. (2013) used data of EU15 and excluded non-eligible regions receiving Objective 1 transfers, whereas Becker et. al (2010) estimated panel data of all eligible and non-eligible regions.

As Becker et. al (2010) the FRD design would be used to measure the causal effect of the treatment as precisely as possible. Based on the European Commission's documents several formally non-eligible regions received the Convergence support for different reasons. On the contrary some eligible regions were left without the transfers. The structure of data corresponds to the FRD design methodology. The model would be estimated by 2SLS using parametric (a multi-ordered polynomial function) and non-parametric (a local linear regression) specifications. The idea behind that is the robustness of results to different types of estimation. Finally, the predicted GDP growth caused by the Objective 1 treatment would be compared with the costs spent to measure the effectivity of the regional policy.

To check a validity of the method various sensitivity tests would be performed. According to Lee and Lemieux (2010) baseline covariates influencing the forcing variable (initial GDP per capita) would be graphically and formally estimated to prove the adequacy of the FRD design. Moreover, the graphical analysis of the dependent variable should detect possible problems of the method. The analysis would be also corrected from the bias caused by spillovers between regions.

For the purposes of this study the data from different sources would be used. The variables of the main interest would be the annual growth rate of GDP per capita and the annual employment growth for 1999-2013 at NUTS2 regional levels. For testing the validity of methodology other variables like population, population density, agriculture share, employment share would be used. Above mentioned variables would be obtained from the Eurostat database. The information about the Objective 1 treatment would be collected from the European Commission's documents. The list of regions receiving the Objective 1 support for certain programming period is specified in Council regulations, specifically for period 2000-2006 in Council Regulation 1999/502/EC and for period 2007-2013 in Council Regulation 2006/595/EC.

Expected Contribution To the best of my knowledge only few existing studies cover the latest period 2007-2013. Moreover, the RDD method was used only in few cases despite the fact that the decision about the Objective 1 suitability fits for the analysis. I will extend the analysis by another programming period 2007-2013 using panel data structure of all eligible and non-eligible regions at NUTS2 levels. The thesis should complement the gap in current researches by robust results of the impact of EUs cohesion policy using various estimation methods.

Outline

1. Introduction: the regional policy, funding, historical background, Objective 1 treated regions
2. Existing researches: different methods and results
3. Data description and adjustment
4. Methodology: The RD design, The FRD design, a model, sensitivity tests
5. Results: the impact on GDP growth and employment growth, effectiveness of the policy
6. Conclusion: discussion of the results and possible improvements

Core bibliography

BECKER, Sascha O.; EGGER, Peter H.; VON EHRLICH, Maximilian. 2010. Going NUTS: The effect of EU Structural Funds on regional performance. Journal of Public Economics, 94.9: 578-590.

BECKER, Sascha O.; EGGER, Peter H.; VON EHRLICH, Maximilian. 2012. Too much of a good thing? On the growth effects of the EU's regional policy. European Economic Review, 56, 648668.

BECKER, Sascha O.; EGGER, Peter H.; VON EHRLICH, Maximilian. 2016. Effects of EU Regional Policy. Warwick Economics Research Paper Series, No. 1118.

BLACK, Dan; GALDO, Jose; SMITH, Jeffrey A. 2007. Evaluating the bias of the regression discontinuity design using experimental data. University of Chicago, Working paper.

CUARESMA, Jesus Crespo; DOPPELHOFER, Gernot; FELDKIRCHER, Martin. 2014. The determinants of economic growth in European regions. Regional Studies, 48.1: 44-67.

IMBENS, Guido W.; LEMIEUX, Thomas. 2008. The regression discontinuity design theory and applications. Special Issue, Journal of Econometrics 142, no. 2: 611-614.

LEE, David S.; LEMIEUX, Thomas. 2010. Regression discontinuity designs in economics. Journal of economic literature, 48.2: 281-355.

MOHL, Philipp; HAGEN, Tobias. 2009. Do EU structural funds promote regional growth? Evidence from various panel data approaches.

MOHL, Philipp; HAGEN, Tobias. 2008. Which is the Right Dose of EU Cohesion Policy for Economic Growth? ZEW Discussion Papers, 09-052.

NEUMARK, David; SIMPSON, Helen. 2014. Place-based policies. National Bureau of Economic Research.

PELLEGRINI, Guido, et al. 2013. Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach. Papers in Regional Science, 92.1: 217-233.

Author Supervisor

Chapter 1

Introduction

The regional policy of the European Union is a cooperation between the EU, member states, regions and individuals with the common aim of reducing social and economic disparities between regions and improving people's quality of life in each one of them. The EU influences the regional investment plans by setting objectives and strategies, that should be followed during each programming period. The regional policy is financed through a system of structural and investment funds (ESI), which allocates about a half of the EU's budget. Over the past years, the EU invested larger amount with each new programming period to boost economic growth of disadvantaged regions and to improve economic situation in the society. Judging from this behavior it is believed that the EU's treatment has a positive impact on the regional performance. In this thesis, we examine whether the Objective 1 treatment truly affects the regional growth, specifically GDP per capita and employment growth using the regression discontinuity design approach.

It is not surprising, that the EU's regional policy has attracted great attention of researchers over the past years. Many authors tried to evaluate the impact of structural funding on the regional performance using various methods and datasets. Despite that, there does not seem to be a consensus about their findings. To the best of our knowledge, not many studies include the last programming period 2007-2013 and almost none of them use the nonparametric estimation method of the regression discontinuity design (RDD). In this thesis we follow the study written by Becker *et al.* (2010) and we extend the analysis for another programming period. Unlike other authors we focus on the nonparametric approach of the homogeneous treatment effect estimation and we supplement current debate about efficiency of the EU's regional policy with a valid analysis using estimation method that is believed to be more suitable for the problem of this nature due to its convenient properties. Moreover, we analyse the development of the treatment during the programming period and we compare the parametric and nonparametric approaches to provide an insight into discussion about the suitability of these methods.

Overall, the thesis is structured as follows. Chapter 2 introduces the EU's regional policy in details. We describe the development of the policy focusing on the differences between programming periods and objectives of the EU. Specifically, we present the background to assignment process of the Objective 1 treatment, which is the objective of our main interest. We also mention various programs and projects, that are financed through the system of ESI funds. Next, Chapter 3 provides literature review regarding the researches exploring the impacts of the structural funding on the regional performance. Since many authors studied the effects of EU's regional policy using various methods over the years, we focus on the papers analysing the last two programming periods and we pay close attention to research studies applying the regression discontinuity design

In Chapter 4, we describe the process of creating our datasets. We had to face to many problems such as the redefinition of borders of regions at NUTS2 level, development of definitions of variables of our main interests or adjustments and corrections of the data while preparing the data files. These issues forced us to create two datasets with several adjustments. Methodological background is provided in Chapter 5. First, we outline the idea behind the regression discontinuity design mechanism and its suitability for our estimation problem with a special emphasis on the fuzzy design. Then, we describe the model and we point at the problem of choosing between parametric and nonparametric estimation method.

In Chapter 6, we present estimation methods and results. We start with the graphical analysis of the Objective 1 treatment effect on the GDP per capita growth using both a smooth and a non-smooth approximation of regression function. Then, we proceed to estimation of the model using parametric and nonparametric methods with different specifications such as different bandwidths, model parameters and polynomial orders of the forcing variable.

This section also provides validity tests of suitability of the RDD mechanism. Specifically, we analyse the density function of the assigning variable and the comparability of the groups at each side of the threshold before the Objective 1 treatment. Moreover, we add various model extensions to reveal other characteristics of the Objective 1 treatment. Finally, Chapter 7 provides a summarizing discussion of the findings and comparison with other research studies supplemented by sensitivity tests of the bandwidth and kernel function selection.

Chapter 2

Regional policy of the European Union

The regional policy of the European Union is an investment policy focusing on the boost of the economic growth and on dealing with regional disparities. The main goal of the EU regional policy is to help regions to improve their economic performances and people's well-being. The regional policy currently uses almost a half of the whole EU's budget (about EUR 461.1 billion out of total EUR 1082 billion during current programming period 2014-2020[1]) to invest in regional growth, transport and communication infrastructure, energy production, job creation, education systems and research and innovation. It also supports small and medium-sized enterprises (SMEs) and it helps to build an eco-friendly economy. Moreover, the regional policy is also "an expression of solidarity between EU countries as it dedicates the bulk of its funding to the EU's less developed regions." (European Commission 2014b) In a union, in which regions like Inner London with GDP per capita 3 times higher than EU average and Romanian Nord-East region with GDP per capita at 20% of EU average are "next to each other", there is a need for a policy dealing with such variation. Generally, the regional policy helps disadvantaged regions to eliminate economic and social disparities and to converge to common values. The idea of solidarity is important in current society, where a belief in an indispensability of common European integration is weakening. (European Commission 2014b)

[1]Information about the EU budget was collected from official website `https://ec.europa.eu/regional_policyandhttps://cohesiondata.ec.europa.eu/overview`.

The EU regional policy is financed through the European Regional Development Fund (ERDF), the European Social Fund (ESF), the European Agricultural Fund for Rural Development (EAFRD), the European Maritime and Fisheries Fund (EMFF) and the Cohesion Fund (CF) (also a complementary resource The Youth Employment Initiative (YEI)). Together, they form a system of Structural and Investment funds (ESI), which covers various investment areas contributing to the regional growth.

2.1 General view on programming periods

The EU regional policy supports member states in a specific way, since it focuses on the regions and their diversities. It is not surprising that there are large differences between regions from different countries, since each country is in various economic, social and political situation. However, in each state there are significant divergencies between its regions- for example, in Spain the region Extremadura had GDP per capita at 60% level of EU average and the region Community of Madrid exceeded the EU average by about 33% over 2000-2002 (reference years for eligibility decision for period 2007-2013). In these two regions the unemployment rate differed by about 10% and the agriculture sector in Extremadura employed by about 15% of inhabitants more. Such situation is not an exception and it demands a special treatment. Although the member states have the final say in an investment program for each region, they must take into consideration the guidelines of the EU as a co-financer of regional development. The EU regional policy sets a programming period, during which member states prepare Community support frameworks (translated into Operational Programmes), describing the whole concept of boosting their disadvantaged regions, and Single programming documents, containing the data and shorter notes from Community support frameworks.

Over each programming period the EU sets main goals and objectives, which point the direction of investments and funding in each region. Historically, the programs last 5 years over period 1988-1992 and 6 years over period 1994-1999. Starting with period 2000-2006 the EU regional policy is established for a period of 7 years. In our analysis we focus on the last two programming periods due to lack of the data and due to comparability reasons (explained in next sections). The EU defines regional areas based on the Nomenclature of Territorial Units for Statistics (NUTS) system, which was defined by European

Statistical Office (Eurostat) for purposes of "the production of regional statistics and for targeting political interventions at a regional level." (European Union 2018) The NUTS standards were redefined over years, starting with a version from 1988. These changes in definitions of regional areas may complicate the analyses. For our situation this problem is commented on in more details in section Data description.

As it was already stated, we focus on the last two programming period in our analysis. Also, the list of priorities and goals of European regional policy is very long. For this reason, we present only a sample of activities to outline the main ideas of the policy. For period 2000-2006 the EU prepares guidelines according to three main priorities - regional competitiveness, economic and social cohesion and the development of urban and rural areas. The EU insists on certain conditions of the economic environment that must be met for companies in each area in order to reach the regional competitiveness. For example, based on the official legislation of the EU, member states must include investments in transport networks into their programs. (European Commission 2005a) Specifically, states should improve effectiveness and modernization of the network by connecting the main lines with side road system from smaller cities. The states should also support the environment-friendly networks and try to reduce the importance of the road transport in favour of other types of transport. The EU states that the key to regional convergence also lies in support of the small and medium-sized enterprises and in investment in the energy networks, including the renewable energy resources, which is related to eco-friendly thinking such as responsible treating of waste water and recycling or disposal of waste. Moreover, the EU points out that in current society the development of telecommunication infrastructure and investments in research and development are crucial. Since information has great value nowadays, the authority should assist SMEs to reach knowledge in field of innovative technical developments by providing training and mobility opportunities for workers of various professions. (European Commission 1999)

Under the European employment strategy, the authority focuses on improving skills and knowledge of human capacity. This strategy is mainly financed by the European Social Fund (ESF), which allocates the funds "to improve employment opportunities in the internal market and to contribute thereby to raising the standard of living". (European Commission 1999) The employment strategy is closely related to the Objective 3 of the European regional policy, which focuses on the employment and education of workers and is described in more details in the next section. Generally, the EU proposes various policy fields in the spirit of the three main elements: a support of equal opportunities regardless of the gender, a promotion of the employment potentials in the information society and focusing on the local development through specific pacts. Based on these policy fields, the member states should implement practical processes with concentration on the groups at risk such as young people, disabled individuals or groups of individuals, that may be subjects of sex or racial discrimination. Moreover, the EU promotes the lifelong learning with a special attention to the information technologies and communication. Unlike in the previous programming periods the responsibility for the employment development is partially transferred to local authorities, which should relate activities of the ESF to other Structural Funds. (European Commission 1999)

Finally, the EU issues recommendations about the common development of urban and rural areas, which should lead to deeper European integration. In other words, the goal of the EU is to "to reduce disparities between the core and peripheral regions" (European Commission 1999) and to weaken the continuing concentration to metropolitan areas to reach more balanced economic situation. Generally, the EU introduces a plan of a system of areas with deep economic integration more evenly distributed across the union. On the other hand, the authority understands differences of each area and it proposes specific treatments to be in conformity with its characteristics. Specifically, the urban areas are a centre of communication, innovation or economic and technological development, but they are also great consumers of energy and natural resources. Moreover, some urban regions are also classified as eligible for Objective 1 treatment and they should be supported based on the plan which considers their specifications. Many regions with rural characteristics suffer from economic problems caused by the structural changes of society followed by falling importance of income from agriculture activities. These regions often experience lack of job opportunities and of resources for the development of other sectors. The development of these regions represents an important part of the regional policy, also because of its relation to the European agriculture model. For this reason, the authority promotes the modernization of agriculture, the improvement of product quality and marketing. Another strategy to boost the rural areas is to diversify their income resource by developing another sectors or services and creating new job opportunities. Moreover, the authority requires an environmental- friendly attitude to all activities in rural areas. (European Commission 1999)

The priorities of regional policy across programming periods are similar in many aspects. So, for period 2007-2013 we describe only ideas and activities, which were not mentioned in previous paragraphs. Like in previous programming periods the process of investment and the progress of the projects is managed by decentralized authorities as well as the EU and social partners and organizations. The member states and regions present the National Strategic Reference Frameworks and the Operational Programmes, which serve as a base for the investment decision, made by regional and national authorities in cooperation with the EU. Based on the Council decision (Council of the European Union 2006a) over period 2007-2013 the main investment fields are knowledge and innovation, transport, human resources and environmental protection. Besides these areas the EU regional policy focuses on the energy networks development, tourism, administrative processes or regional culture.

According to results of previous programming periods the investment in research and innovation is crucial for a sustainable growth of regions. That is why the EU allocates about EUR 60 billion to this field (about 2-3 times more than for previous programming periods)[2]. Out of this about EUR 27 billion is used for research and development for firms, about EUR 15 billion to information technologies and the rest is used for research infrastructure. In programming period 2007-2013 the EU stresses the importance of sustainable transport connected between cities and member states. Specifically, about EUR 76 billion (two times more than in previous period) is allocated to the trans-European transport (TEN-T) projects. To provide more balanced transport network, the authority invests not only to road infrastructure (about EUR 41 billion) but also to ports and inland waterways (about EUR 4 billion), airports (about EUR 2 billion) or to rail network (about EUR 24 billion). More than ever, there is a need for investment in the road infrastructure due to growing number of cars. Moreover, this problem is closely related to environmental issues. The Structural and Cohesion Funds invest about EUR 100 billion in environmental programmes, focusing on pollution reduction or water and waste treatment. A large part of the funding is spent on the renewable sources of

[2]Information about the EU budget was collected from the official guideline on cohesion Council Decision (2006/702/EC) and from the magazines prepared by the European Commission in 2014.

energy, which should boost the competitiveness of the whole European area. (European Commission 2014b)

All these developments are an important step in arranging of conditions for creating new enterprises and attracting new investors. In this spirit the EU helps small and medium-sized enterprises to overcome the difficult first years and to stay in the business. Specifically, the EU allocates up to EUR 43 billion to new companies as direct investments with higher concentration on firms involved in technology and innovation of environmentally-friendly production. The authority invests about EUR 94 billion in human capital and services related to business. Behind each company there are human resources with necessary knowledge and know-how. The EU regional policy is prepared to invest about EUR 95 billion to human resources (EUR 76 billion through the ESF and EUR 19 billion through the ERDF) to create new job opportunities and to improve people's knowledge and skills. While the ESF focuses on "soft development" such as education, training, social inclusion and the equality of opportunities, the ERDF promotes the infrastructure behind the improvement of people's knowledges like schools and educational centres. (European Commission 2014b)

For technical reasons explained later, we analyse data from two last complete programming periods 2000-2006 and 2007-2013. On the other hand, it seems right to also present current information about the funding and objectives of the regional policy. Over the current programming period 2014-2020 the EU has focused on 4 main themes, which represent 4 objectives financed through the ERDF: research and innovation, information and communication technologies, SME competitiveness and low carbon economy. In addition, the EU promotes fields like in previous programming periods such as transport, education and training, employment, dealing with climate changes, tourism, health, energy sustainability projects and so on[3]. Generally, the regional policy supports projects, which may help to fulfil goals defined in the Europe 2020 strategy. The main targets for 2020 are 75% employment, 3% of the EU GDP to be invested in research and development, to lower number of inhabitants in or in risk of poverty, to lower the greenhouse gas emissions by 20% compared to 1990 and to reduce the number of early school leavers to 10%. (European Statistical Office 2017)

[3]Most of these priorities are covered by the 11 objectives that are described in the next section.

2.2 Objectives across programming periods

The regional policy intervenes in many areas, which leads to a complex system of programs, funds and goals. To make the system more transparent, the EU set main objectives, which cover a range of programs and investments used to fulfil these goals. Over the period 2000-2006 the EU set three objectives: Objective 1, Objective 2 and Objective 3[4], which were redefined to Convergence objective, Regional Competitiveness and Employment and European Territorial Cooperation during period 2007-2013. The Convergence objective remained very similar to the form of Objective 1 from previous programming periods. Based on the Convergence objective the EU regional policy supports less developed regions to reach a convergence in economic and social areas. To assign the support to regions as objectively as possible, the EU define an official rule for eligibility: the regions with GDP per capita below 75% of EU average during a certain period are classified as eligible for Objective 1 treatment. Naturally, over each programming period there were few exceptions receiving (or not receiving) the treatment regardless of the eligibility rule. The examples of such exception are Swedish regions Norra Mellansverige, Mellersta Norrland, Ovre Norrland and Finnish regions North, Central and East Finland, which received the treatment in period 2000-2006 due to low population density. Regions Canary Islands, Guadeloupe, Martinique, Reunion, French Guiana, the Azores and Madeira also received special treatment due to their unique position separated from the EU[5]. (European Commission 2005b) Moreover, in few cases the regional data were not available during the decision-making period, which prevented some candidates from the support.

After accession of countries from Central and Eastern Europe the average GDP per capita dropped significantly, which changed standard for Objective 1 eligibility. For this reason, regions that were originally eligible for the treatment, received the support transitionally. Over period 2000-2006 the Objective 1 was mostly financed by the European Regional Development Fund (ERDF), the European Social Fund (ESF), the Guidance section of the European Agricultural Guidance and Guarantee Fund (EAGGF) and the Financial Instrument for Fisheries Guidance (together known as Structural Funds) by about EUR 137 billion. Moreover, about EUR 128 billion of the Objective 1 budget

[4]In period 1994-1999 there were two more objectives: Objective 4 and Objective 5.

[5]In 2000-2006 Northern Ireland received special Community assistance to help establish a stable society.

is allocated from member states' contribution. (European Commission 2005b) Over period 2007-2013 the Convergence objective was financed by the ERDF, the ESF and the Cohesion fund, allocating together about EUR 251 billion, which corresponds to 81.5% of the total budget of regional policy[6]. (Council of the European Union 2006b)

For period 2000-2006 the EU defined the Objective 2 as a combination of former Objective 2, focusing on the conversion of regions which were on the downgrade and of former Objective 5(b), representing the support of rural regions. Unlike the Objective 1, the Objective 2 eligibility depends on "a population ceiling, and on criteria specific to each area." (European Commission 2005c) Generally, the population in a region cannot exceed two thirds of the population previously covered by Objective 2 and Objective 5b, to be eligible for the new Objective 2 treatment. Due to accession of new member states in 2004, the rules were redefined for each member state separately, which makes the structure even more complex. Moreover, there were specific criteria for each member state based on the development in economic sectors. Despite the broad range of interests of the Objective 2, only about EUR 22.5 billion was allocated to Objective 2 from the ERDF and the ESF over period 2000-2006. For period 2007-2013 the EU set a new objective Regional Competitiveness and Employment, which concentrates on various area related to labour market. Under this objective the EU tried to boost competitiveness of regions that were falling behind but were not eligible for the Convergence treatment. Specifically, this objective provides help to regions which were eligible for Objective 1 support in previous programming period, but not in the period 2007-2013 and not because of the accession of the new member states. Over period 2007-2013 the ERDF and ESF allocated about 50 billion EUR to achieve the Regional Competitiveness and Employment objective. (European Commission 2008)

The Objective 3 under period 2000-2006 seems rather to correspond to the Regional Competitiveness and Employment than to European Territorial Cooperation, since it focuses on the themes related to human resources. Originally, it was split into two separate goals: the Objective 3, which focuses on problems of young people without jobs and on the long-term unemployment, and the Objective 4, concentrating on the conceptual unemployment due to changes in the economy and production. Generally, under the Objective 3

[6]About EUR 69.6 billion was reserved for the Cohesion fund.

from the period 2000-2006 the EU invested in training the future workers and promoted education and employment. The Objective 3 can be considered as a complement of the Objective 1, since it covered regions excluded from the Objective 1 treatment. The budget was solely covered by the ESF, which allocated about EUR 24.1 billion for original member states (about 12.3% of the total Structural funds budget) plus about EUR 110 million for new member states. (European Commission 2005d)

Over period 2007-2013 the ERDF allocated about EUR 8 billion to the objective European Territorial Cooperation, which supported small and medium-sized enterprises and promoted a cooperation between regions in innovation, research and environmental issues. Specifically, about EUR 6.44 billion was earmarked for cross-border cooperation, EUR 1.83 billion for transnational cooperation and EUR 445 million for interregional cooperation and networks. (European Commission 2008) In current programming period 2014-2020 the EU defines eleven thematic objectives, which are divided into three subgroups according to their funding. The list of the objectives with their funding is presented in Appendix (the highlighted marks represent the main priorities of each fund)[7].

Over period 2014-2020 the EU stresses the sustainable regional growth, requiring the ability to adapt to innovations and to promote the research. The Structural Funds allocate about 30% of the total budget of regional policy for implementation of the strategy for smart specialisation (RIS3), which is based on greater cooperation of policy makers and local players to benefit from the technical diversification. The idea of the first Objective is closely related to the improvement of the information and communication technologies, which is covered by the second Objective. The ICT development is mostly financed by the ERDF, which allocates over EUR 20 billion to support a formation of the EU's single digital market, that represents an online environment without certain barriers and regulations. The main strategy of the Digital Single market is to provide "access to online activities for individuals and businesses under conditions of fair competition, consumer and data protection, removing geo-blocking and copyright issues." (European Commission 2019) As in the previous programming periods, the EU sees economic potential in SMEs and it earmarks about 20% of the ERDF budget for their development. The EU regional pol-

[7] The source of information is official website of EU regional policy `https://ec.europa.eu/regional_policy/en/policy/how/priorities/`

icy pay close attention to the question of the ecology, carbon footprint and sustainable resources. The policy makers realize that reaching the sustainable growth is possible only if the environment is protected and regions invest in the renewable resources, environmental-friendly technologies and low-carbon economy. Low-carbon economy is represented by one of the Objectives, mostly supported by the ERDF. Starting with the current programming period, the member states must contribute to this objective by a certain proportion of their funding. Together with the Cohesion Fund, the ERDF invests about EUR 40 billion to build smart energy networks, sustainable transport network and more efficient buildings. Besides that, the EU points out that human capital is the key to sustainable growth. To reach 75% employment 2020, as one of the goals of Europe 2020 strategy, the EU invests in improving the job conditions and in creating of equal job opportunities. (European Union n.d.)

The ESF (with support of the ERDF, the EMFF and the EAFRD) finances the areas of employment, labour market and education. About EUR 30 billion is allocated to projects, which support employment and job creation and about EUR 25 billion to projects, promoting lifelong learning, modernization of educational systems and strengthening of connection between education systems and labour market. The ESF provides resources for boosting employment (especially for young or differently disabled people) and for building a stable society with equal opportunities. The ESF work is based on close cooperation of European, national, regional authorities and public or private organisations. To distribute responsibility and to evoke the feeling of participation, each project is partially financed by local authorities or organisations (usually between 50%-80% of the project's budget). Such a cooperation is also implemented into management of the project, since its program goes through series of negotiations with all actors. Currently, the ESF provides funding for 438 projects that improve workers' skills and knowledges and that create jobs opportunities for everyone. Specifically, in the Czech Republic there are about 20 programs such as *Training café brings an end of isolation* in Slaný, concentrating on the mentally disabled people, or *A way back into work for former drug addicts* in Prague. (European Commission n.d.) During period 2007-2013 the ESF also financed many projects such as running a gift shop in the *Dragonfly gallery in Prague*, where people with serious psychiatric illnesses are employed. (European Commission 2014a)

Moreover, the EU promote the social inclusion (about EUR 30 billion through the ESF and ERDF). Specifically, the EU's funding should not be spent for activities, which contribute to gender, racial or any other discrimination. (European Commission 2018) In 2013 the European Commission came with the Social investment package, "an integrated policy framework which takes account of the social, economic and budgetary divergences between Member States". (European Union 2013) The main ideas of this package are providing sustainable social protection systems, fighting the childhood poverty and to generally help people with difficulties through their lives. The last Objective is dedicated to the efficient public administration, that is based on building stable institutions, which put an emphasis on the flexibility and transparency for public. (European Union n.d.) Although there is no specific objective, covering less developed regions, the Structural Funds allocate about EUR 180 billion to areas with GDP per capita at level less than 75% of the EU average. Moreover, almost EUR 40 billion is allocated to regions with the transitional support.

The process of funds allocation starts with member states' presentation of regional development plans, in which they explain their economic and social situation and they suggest "the most appropriate strategy for achieving the stated development objectives and indications on the use and form of the financial contribution from the Structural Funds". (European Commission 2005b) For our purposes we analyse the impact of the Objective 1 treatment, since it is the only objective, whose eligibility definition is formally stated. Moreover, the largest part of the budget is allocated under this objective and the definition is stable over the programming periods, which makes it suitable for the analysis.

Chapter 3

Literature Review

An efficiency of EU regional policy is an attractive topic for many authors. Since the main regional funds, The Structural Funds and the Cohesion Fund use about a third of the whole EU's budget, various research focuses on the impact of the policy on regional performance. Interestingly the results seem not to be consistent among different methods and data specifications. One of the possible reasons for such a disparity is a complicated separation of causal treatment effect and other factors, which causes an endogeneity problem. Besides other things the complexity of the regional policy structure can cause difficulties in revelation of true impacts. Because of many research studies studying the efficiency of regional policy we present only sources, which focus on the last programming periods of structural funding.

One of the first studies processing the programming period 2000-2006 was written by Mohl & Hagen (2011). They apply various panel data methods to estimate impact of structural funding on regional GDP for individual objectives of regional policy and to examine time lag of its effectiveness. They use a two-step GMM estimation system with extensions for correction of spatial autocorrelation. Mohl & Hagen (2011) analyse dataset of NUTS1/NUTS2 regions during ten-year period from 1995 to 2005, which overlaps with two programming periods. The dataset contains information about concrete regional payments and commitments divided into objectives as a percentage of regional GDP. Mohl & Hagen (2011) conclude that the efficiency of regional policy depends on the type of objective for which it is examined. They find positive significant impact only for Objective 1 (about 0,5%), while Objective 2 and

Objective 3 show a negative effect on regional GDP, which seems to be also frequently significant. The impact of regional policy as a whole (Objective 1+2+3) does not present stable significant results. In addition to that, the results suggest that the effect of structural funding occurs with a four years lag. Mohl & Hagen (2011) point out that only Objective 1 has a clear definition of criteria for funding, which can be the explanation for its permanent significance in comparison to other objectives. They also emphasize that Objective 2 and Objective 3 rather affect labour market and their impact on GDP growth is only indirect.

The main inspiration for my thesis was an article written by Becker *et al.* (2010). In their research, they focus on the impact of structural funding under Objective 1 on regional GDP per capita growth and employment growth using a regression discontinuity design approach. Unlike most of the older literature Becker *et al.* (2010) work with a detailed data at level NUTS2 and NUTS3. Specifically, they use 285 NUTS2 and 1213 NUTS3 regions for three programming periods 1989-1993, 1994-1999 and 2000-2006. Since the assignment rule to the Objective 1 funding is not perfect, the fuzzy regression discontinuity design is used for estimation of causal effects of the policy. Based on the suggestions of Lemieux & Imbens (2008) authors estimated model by two stage least square method using a parametric approach with various polynomial specifications and a non-linear first-stage equation. Moreover, they calculate pooled OLS and fixed effect estimates in order to reduce sampling variability and they also adjust the model for spillover and cumulative time effects. The estimation shows a significant positive impact of about 1.6% on annual GDP per capita growth, while there seems not to be any significant effect on the employment growth in the original model specification. After controlling for spillover effects, the impact on employment growth seems to be significant and it has value about 0.9%. Moreover, authors find out that the treatment effect is slightly different for each programming period and it seems to display after at least four years. Finally, Becker *et al.* (2010) conclude that a euro spent on structural funding under Objective 1 causes an increase in GDP by about 1.2 euros and that it is probably related to "a stimulus on the volume and structure investment (e.g. infrastructure) and, eventually, productivity gains but much less so with the creation of new jobs within the same programming period." (Becker *et al.* 2010)

Maynou *et al.* (2016) use extended dataset for period 1990-2010 to study the impact of structural and cohesion funds on performance of regions of Eurozone countries. They use a dynamic panel data model with a spatial adjustment on data at NUTS2 level. Authors also find a positive impact of structural funding on Eurozone GDP per inhabitant growth. Unlike researchers mentioned above, Kyriacou & Roca-Sagalés (2012) analyse the effect of European cohesion policy on income disparities of member states during period 1995-2006 using Feasible General Least Squares estimation. They conclude that the structural funds help reduce regional disparities up to certain level of transfers (about 1.6% of national income), above which the disparities tend to grow. Authors explain such phenomenon by possible existence of moral hazard.

Three years later Becker *et al.* (2013) re-estimated Objective 1 impact using the RDD with heterogeneous local average treatment effects. Basically, they try to reveal different effects of transfers among regions conditionally to the degree of human capital and the government's quality. According to their results only regions with the human capital and the quality of government that is above certain level are able to benefit from the regional transfers. Surprisingly only 30% of the areas seem to be able to use these transfers to boost the regional growth and only 6% of all examined regions show significant impacts on investment while the rest of them present significant effects only on the consumption, which would probably not remain in a long-term period, or not significant results at all. On the other hand, regions with "an absorptive capacity" that is higher than an average level seem to use the subsidy in much more effective way than average regions. (Becker *et al.* 2013)

Rodríguez-Pose & Garcilazo (2015) study the impact of quality of local and regional governments on the regional economic performance and on the efficiency of the EU structural funding. The research is based on a panel data estimation to reveal a relationship between GDP growth per capital and the government quality[1] with specifications for the cohesion policy treatment in period between 1996 and 2007. The authors find out that the government quality is important factor only for regions which receive amount of transfers above a certain threshold. Specifically, for regions with structural funding above 80 euros per capita the government factor seems to have impact on both the economic growth and an efficient use of the funds. In extreme cases (regions with the funding above 120 euros per head) the improvement of the government seems to be the best strategy how to call on the EU regional budget

[1]The government quality is measured by a quality of government index prepared by the Quality of Government Institute at the University of Gothenburg.

efficiently. Few years before Becker *et al.* (2012) studied the impact of regional policy at NUTS3 level using a generalized propensity score method. They try to reveal how the GDP growth changes with various amounts of transfers under Structural Funds and Cohesion Fund objectives. The results show that an optimal level of transfer is about 0.4% of target regional GDP and that a maximal transfer rate is about 1.3%. Only for regions with transfer rate below the maximal threshold it would be profitable to gain an additional subsidy. (Becker *et al.* 2012)

In the spirit of the heterogeneous treatment effect, Doppelhofer *et al.* (2008) estimate factors that have impact on the regional economic growth and convergence and that might be sources of heterogeneity of treatment effect. Authors find positive convergence effects of regions at NUTS2 level. Moreover, they reveal significant differences between economic growths among regions such as regions around capital cities or regions with a significant population share of higher educated workers tend to grow more. There is also diversity between old and new EU member states and the spatial spillovers seem to matter. Gagliardi & Percoco (2017) try to reveal the impact of European cohesion policy in the context of potential spatial heterogeneity in different levels of regional development. They divide regions to categories according to the degrees of urbanization (cities, intermediate and rural areas) and according to distances from the main urban cities (areas close to the urban city and distant areas) and they apply the RDD approach for regions at NUTS3 level in programming period 2000-2006. Authors find a general positive impact of Cohesion policy with a different intensity for development categories. The results show that the impact is the strongest for rural areas close to the cities and such subgroup of regions seem to be the main reason for the general positive effect. For other groups the impact is quite weak or even insignificant. These conclusions are following previous research studies which suggest that the heterogeneity is an important issue in revealing the impact of regional policy in more specific way.

Pellegrini *et al.* (2013) were inspired by their predecessors and they applied the RDD approach on EU-15 regions at NUTS2 level for periods 1994-1999 and 2000-2006. Contrary to Becker *et al.* (2010) they use a sharp regression discontinuity design assuming only regions, which were truly eligible/non-eligible for the Objective 1 treatment based on the stated rule. Moreover, since they consider other sources of funding such as Structural and Cohesion Funds or national financing, they set a maximal level of transfer per capita[2] and the regions with higher level of transfers are excluded from the control group. The authors find positive effects of regional policy on GDP growth in a range of 0.6-0.9 percent per year using parametric and nonparametric approach. Although the research shows seemingly effective results, authors point out that the speed of the regional convergence is quite low. Specifically, it would take about 50-75 years to reach the complete convergence assuming current regional GDP per capita levels.

The effectiveness of European regional policy in connection with technological development and transport infrastructure in the programming period 2000-2006 was studied by Ferrara *et al.* (2017). Authors follow Pellegrini *et al.* (2013) and they use a nonparametric estimation of sharp RDD with similar adjustments. They find positive effects of structural funding on both dependent variables. In the research about the impact of European structural transfers on an outcome and a household income convergence, Checherita-Westphal *et al.* (2009) apply a system of simultaneous equations in order to deal with possible endogeneity problem. They use an EU-19 regional dataset for 1995-2005 period at NUT2 level. Interestingly authors find that while net transfers conduce to the household income convergence, they might become an obstacle to the outcome convergence. On the contrary, Palevičienė & Dumčiuvienė (2015) choose a different path in their research of European structural funds convergence. They focus on socio-economic characteristics of regional performance applying a principal component analysis (PCA) on NUTS2 level data in 2007-2008 to reveal factors influencing the outcome. According to the PCA results Palevičienė & Dumčiuvienė (2015) divide EU member states to four clusters based on the socio-economic factors; employment, high and low economic educational development and active population growth indicator. In other words, authors point out that in spite of a long history of European regional funding there are still large differences in regional development among EU member states.

In similar ideas in mind, Tomova *et al.* (2013) study the impact of structural funding on socio-economic development conditionally on the state of national fiscal and macroeconomic policies. Specifically, they construct an indicator of socio-economic development (SEDI) using several areas of national performances to estimate the impact of the European structural and investment (ESI)

[2]The maximal threshold is equal to 1960 euros per capita.

funds and of the sound fiscal and macroeconomic policy itself. Authors analyse data at national level for period 1980-2010. They find out that the ESI funds have a positive impact on the socio-economic development of member states in general. Moreover, the effect of cohesion policy seems to be strengthened by sound fiscal and economic policy.

When another programming period ended, Becker *et al.* (2016) came with an extended research. In their study they examine basic outcomes such as GDP growth, employment growth and investment rate over the period 1989-2013 and additional variables for the last two periods (growth of total compensation and hours worked, growth of patents, participation rate in training and education and payment relative to commitments). Becker *et al.* (2016) use the RDD approach for estimation of binary treatment effects under Objective 1 criterion at NUTS2 level and a generalized propensity score method for continuous treatment of the European regional policy as a whole at NUTS3 level. The authors find a significant positive evidence of Objective 1 impact on GDP growth during the last two periods which is stronger assuming the complete dataset (4 programming periods). Contrary to the GDP growth outcome the treatment appears to have no effect on employment growth[3] and total investment rate for the whole period, but there seems to be a positive significant effect on employment growth during periods 2000-2006 and 2007-2013. Based on the results of the second approach it can be concluded that "most regions will benefit from a more balanced spending across different spending categories as opposed to concentrating spending on singular categories." Becker *et al.* (2016) Moreover the authors focus on the Objective 1 treatment effect on regions in the UK and they do not find any differences in efficiency of using EU's transfer for an improvement of the regional performance.

In a similar manner, Di Cataldo (2016) studies the Objective 1 impact on performance of the two poorest regions in the UK: Cornwall and South Yorkshire, which voted for the Brexit. He uses a matching, a difference-in-differences and a synthetic control method for evaluation of Brexit impacts on the regional performance. Di Cataldo (2016) concludes that the structural funding policy has a positive temporary effect on economic performance of Cornwall and South Yorkshire and the Brexit may have a negative medium-run impact on the regional economy. Giua (2017) studies the impact of Objective 1 treatment on

[3]While the results show a significant effect on total compensation of employees, there is no impact on wages.

the employment in Italian regions using spatial RDD approach. The author assumes municipalities in treated/non-treated regions which share boundaries and she finds a positive impact of cohesion policy on the employment, which is“concentrated in the economic sectors mostly linked with a territorial development process (construction manufacturing and tourism)”. (Giua 2017)

In 2018 Becker *et al.* (2018) examined the impact of regional policy during Financial and Economic crisis and the impact of receiving/losing the Objective 1 treatment on the regional performance using the fuzzy RDD with a heterogeneous treatment effect specification over period 1989-2013. They find a positive effect on employment growth during the crisis, while there seems not to be any impact on income growth. They also conclude that the Objective 1 treatment has only time-limited impact on regional performance since the “previous growth gains seem to be largely undone once Objective 1 status is lost.” (Becker *et al.* 2018) Similarly to the Becker *et al.* (2018), Bachtrögler (2016) looks at the regional policy during financial and economic crisis by estimating of heterogeneous treatment effects. Moreover, she also focuses on the absorptive capacity of the transfer allocation and extends the current analyses by another hypothesis that not only the education level at regions, but also the employment status might be important for efficient allocation of the funds. Basically “a higher education level in a region may only be able to contribute to using European support effectively when it is employed”. (Bachtrögler 2016) The results suggest that the funding seems to cause an increase in GDP per capita growth in a diminishing way and the higher level of institution quality tends to improve the efficiency of the policy. On the other hand, the impact seems to be weaker during the recent crisis.

In 2016 Pellegrini & Cerqua (2016) prepared a research about impacts of Structural and Cohesion Funds for the European Commission. In their study the authors extends the sharp RDD approach to the case of continuous treatment for estimation of impact of transfer intensity of the regional economic growth. In general, the results show a positive effect of EU funding on the regional growth with a stronger concentration of the effect in new member states (by about 0.7% of annual GDP growth) and a weaker effect during crisis. The authors also point out that for a proper evaluation of the funding impacts for new member states it is necessary to have longer period of time than just one programming period. Finally, Pellegrini & Cerqua (2016) reveal a diminishing impact on the regional growth with increasing intensity of transfers. Another report for the European Commission focusing on the transfer intensity was prepared by Bondonio (2016). Bondonio (2016) uses a method of statistical matching for regions at NUTS2 level for various programming periods including the last one. As the findings suggest, the EU regional funding has a positive impact on the regional performance and the effect is stronger with higher intensity of transfers.

As it was stated earlier, there are many studies focusing on the impacts of EU regional policy on the regional performance. However, there are not many research studies, which study the last programming period 2007-2013 and almost none of them uses the nonparametric estimation of the regression discontinuity design. For this reason, we extend current literature by an analysis of the last programming period using the nonparametric approach, which is believed to be more suitable for the regression discontinuity design. In addition, we use the parametric method for robustness control of the results and we comment on differences and suitability of both methods.

Chapter 4

Data description

For the purposes of estimation, I decided to work with data at NUTS2 level during two last programming periods (2000-2006 and 2007-2013). In each period the eligibility for Objective 1 treatment, which is the policy tool of our main interest, is decided based on regional performance during certain time. In the period between 2000 and 2006 the Objective 1 eligibility was determined based on the average GDP per capita during the period 1994-1996 for original member states (forming the EU15) compared to EU average in the same period and on the average GDP per capita during the period 1997-1999 for new members states (forming EU25) compared to EU average in the same period. For the second period the average GDP per capita between 2000 and 2002 was used. The reason for working only with two periods is the availability of comparable data, because the definitions of variables of interest have changed substantially over time before 2000.

We use data at NUTS2 level to be consistent with the official decision about the Objective 1 treatment. Since there have been significant changes in the European Union composition since 2000, two datasets were created: the first dataset includes both programming periods and the second one contains data from the recent period only. The second dataset is made up of balanced data about regions based on NUTS 2013 specification, while the first one deals with few challenges. In our dataset during period 1994-1996 the regions were defined based on the NUTS 2006 system whereas for the periods 1997-1999 and 2000-2013 the NUTS 2010 and the NUTS 2013 systems, respectively, were used. According to these systems some regional borders changed over time which makes these regions unsuitable for further analysis. Specifically, overseas departments of France Guadeloupe and Mayotte[1], German regions Chemnitz and Leipzig, Italian regions Emilia-Romagna and Marche, Finish regions Pohjois-ja Itä-Suomi, Helsinki-Uusimaa and Etelä-Suomi and British regions Cheshire and Merseyside were excluded for the first period. Although the regions Inner London and Outer London were split to five smaller areas in 2000, we had to use the original composition for both programming periods in order to keep the data comparable. The similar situation occurred in case of Slovenian regions, where the whole area was split to 2 regions Vzhodna Slovenija and Zahodna Slovenija in NUTS 2013 specification. Despite that, we use the whole area to retain consistency with original Council regulations. Moreover, the member states were not obliged to provide information about their regions until 2000 which caused few blank spots in the dataset.

For the analysis of the impact of regional policy, the most important variable is the GDP per capita in PPS[2], which is the main indicator in deciding about eligibility of regions for the Objective 1 treatment. The GDP calculation went through several changes caused by improvement of The European System of National and Regional Accounts (ESA). In my datasets the GDP was calculated based on the ESA 79 methodology during period 1994-1996 and based on the ESA 95 methodology during period 1997-1999 (reference periods for the first programming period 2000-2006). The ESA 95 method was also used for calculation of GDP in period 2000-2002 as a reference period for the programming period 2007-2013. To calculate the annual average GDP per capita growth, which is used as main dependent variable in our model, the current ESA 2010 methodology was applied. The differences in calculations, caused by different accounting systems, were made at the lowest level of entities which makes it difficult to recalculate GDP to retain comparability. Luckily, in our analysis we are interested in eligibility of regions to Objective 1 treatment, which is based on the comparison of regional GDP per capita to the EU average during certain period. Moreover, the decision about regional eligibility was made based on the original data (calculated using original ESA method-

[1]The region Mayotte and Extra regions such as ITZZ and NLZZ were excluded from both periods because of absence of information about population.

[2]Purchasing Power Standard (PPS) is an artificial common currency, which is used to eliminate the differences in price levels between countries. In European Comparison Programme a unit of PPS corresponds to the purchasing power of 1 euro in the European Union. Then GDP per capita in PPS is calculated as GDP per capita in national currency divided by the purchasing power parity.

ology), so we do not assume any loss of precision with data constructed this way. Specifically, we use the proportion of regional GDP per capita to the EU average (GDPproportion) as our main independent variable. The information about actual recipients of Objective 1 treatment was collected from the European Commission's documents. The list of regions receiving the Objective 1 support for certain programming period is stated in Council regulations, specifically for period 2000-2006 in Council Regulation 1999/502/EC, published in Official Journal of the European Communities L194 (and for accession countries in Official Journal L 236), and for period 2007-2013 in Council Regulation 2006/595/EC, published in Official Journal L 243.

In more details there are 255 regions at NUTS2 level of 25 EU member states for the period 2000-2006. Based on the official rule and our available dataset, 76 regions are eligible for the Objective 1 support program in the reference years. According to the Council regulations 104 regions received the treatment in the first programming period[3]. In total we miss information about 34 regions. Out of 104 regions which actually received the support 10 regions were parts of the program only transitionally. These regions received treatment in previous programming period but were not classified for the support in the current period assuming the regular rule. Specifically, these regions are Hainaut (BE), Southern and Eastern region of Ireland (IE), Cantabria (ES), Nord-Pas-de-Calais (FR), Corse (FR), Molise (IT), Flevoland (NL), Área Metropolitana de Lisboa (PT), Highlands and Islands of the UK (UK) and Northern Ireland (UK).

As it was stated above, the reference years for the second programming period are 2000, 2001 and 2002. Without loss of generality we use the original ESA 95 calculation of GDP per capita for the reference period. The dataset contains records of 269 regions for programming period 2007-2013. According to available dataset 76 regions were eligible for the Objective 1 treatment in reference years. Out of 94 regions receiving support, 14 regions participated in the program only transitionally. In more details, these regions are Province of Hainaut (BE), Lunburg (DE), Leipzig (DE), Kentriki Makedonia (GR), Dytiki Makedonia (GR), Attiki (GR), Principado de Asturias (ES), Region de Murcia (ES), Ciudad Autonoma de Ceuta (ES), Ciudad Autonoma de Melilla (ES), Basilicata (IT), Burgenland (AT), Highlands and Islands (UK) and Algarve

[3]We miss information about 20 regions which received the support.

(PT). It is also important to note that since we have only formal information about the data, which were used for decision process by EU authority for regional policy, our group of eligible regions differ[4]. Moreover, some values which were not available during the decision-making process were updated later. In Table 4.1 we display total number of regions which received the Objective 1 treatment in each EU country. As it was stated above, the lists of eligible regions and receiving regions differ due to official reasons (for example three Swedish regions receive the support because of low population density) or due to technical reasons caused by unavailability and adjustments of data.

For our analysis we use an annual average GDP per capita growth during each programming period (*GDPGrowth*) as the main dependent variable. We also extend the model by analysing the GDP per capita growth in each year during programming period in the last section. Moreover, we examine the treatment impact on the annual average employment growth (*Employment-growth*) in the last section of our analysis. To properly implement the fuzzy RDD mechanism we use a binary variable *Received/Eligible*, which is equal to 1 when a region received the Objective 1 treatment/was eligible for the Objective 1 treatment and 0 otherwise. For internal validity and sensitivity tests we use various baseline covariates: *Population density* (inhabitants per square kilometer), *Economically active population* (proportion of the population aged 15-64 years), *Agriculture share* (proportion of employees in the agriculture sector), *Industry share* (proportion of employees in the industry sector), *Service share* (proportion of employees in the sector of services), *Patents applications* (number of patents per million inhabitants), *Unemployment rate* (proportion of unemployed people from labour force). To retain comparability, all variables (except for receiving status of regions) were downloaded from the Eurostat database. However, historical data of some variables are available only since 1999. Specifically, in validity testing we had to use the values from 1999 for variables agriculture share, industry share, service share, unemployment rate and economically active population for the first programming period. For the rest of the variables (and for the second period) we use average values during decision making period (1994-1996 and 2000-2002). In Table 4.2 and Table 4.3 we present descriptive statistics for both datasets.

[4]Specifically, there are regions which were officially eligible for the treatment but based on our dataset they are not and vice versa.

Table 4.1: Summary of regions in EU countries

Country	2000-2006		2007-2013	
	NUTS2 reg.	Receiving reg.	NUTS2 reg.	Receiving reg.
Austria	9	1	9	1
Belgium	11	1	11	1
Bulgaria	-	-	6	5
Cyprus	1	0	1	0
Czechia	8	7	8	7
Denmark	5	0	5	0
Estonia	1	1	1	1
Finland	5	1	5	0
France	26	6	25	4
Germany	38	7	38	7
Greece	13	13	13	10
Hungary	7	7	7	6
Ireland	2	2	2	0
Italy	21	7	21	5
Latvia	1	1	1	1
Lithuania	1	1	1	1
Luxembourg	1	0	1	0
Malta	1	1	1	1
Netherlands	12	1	12	0
Poland	16	16	16	16
Portugal	7	7	7	5
Romania	-	-	8	8
Slovakia	4	3	4	3
Slovenia	1	1	1	1
Spain	19	11	19	8
Sweden	8	3	8	0
United Kingdom	37	6	37	3

Table 4.2: Descriptive statistics for period 2007-2013

Variable	*Mean*	*SD*	*Min*	*Max*
GDPproportion	0.935	0.383	0.187	3.178
Eligible	0.289	0.454	0.000	1.000
Received	0.349	0.478	0.000	1.000
GDPgrowth	0.006	0.210	-0.051	0.065
Employmentgrowth	-0.001	0.013	-0.023	0.117
Agriculture share	0.073	0.088	0.001	0.508
Industry share	0.291	0.7078	0.121	0.462
Services share	0.633	0.106	0.248	0.890
Patent applications	104.131	123.542	0.158	806.222
Unemployment rate	0.0849	0.053	0.019	0.260
Econ. act. popul.	0.686	0.062	0.496	0.827
Population density	351.215	853.274	3.300	9137.9

Table 4.3: Descriptive statistics for periods 2000-2006 and 2007-2013

Variable	*Mean*	*SD*	*Min*	*Max*
GDPproportion	0.896	0.347	0.187	3.178
Eligible	0.311	0.463	0.000	1.000
Received	0.378	0.485	0.000	1.000
GDPgrowth	0.022	0.025	-0.051	0.116
Employmentgrowth	0.001	0.010	-0.051	0.116
Agriculture share	0.073	0.088	0.001	0.513
Industry share	0.293	0.072	0.100	0.479
Services share	0.632	0.106	0.248	0.900
Patent applications	88.319	105.051	0.158	806.222
Unemployment rate	0.090	0.053	0.019	0.281
Econ. act. population	0.685	0.060	0.496	0.827
Population density	338.929	799.810	3.300	9137.9

Chapter 5

Methodology

First, we describe theoretical background of the regression discontinuity design, since we assume that the RDD approach is not well-known and for our analysis it is crucial to justify its application. Then we present the applied model and we explain the logic behind it. We begin our analysis with a graphical presentation of the data and we describe the process of sampling the data to clearly display their patterns in the graphs. In next subsections we implement parametric and nonparametric estimation methods, we comment on the results of each method and we explain the differences between them. Since we cannot make any conclusions about the results without checking the validity of the regression discontinuity design, we focus on this problem in the next subsection. In the last subsection, we focus on extensions of our analysis such as inclusion of baseline covariances in the model or analysing the GDP per capita growth in each year of the second programming period. We also estimate the impact of Objective 1 treatment on the employment growth.

5.1 Regression discontinuity design

Regression discontinuity design (RDD) is a quasi-experimental approach frequently used in nonexperimental researches where candidates are chosen according to value of a specific variable (called assignment or forcing variable). In this design a candidate receives a special treatment if the value of his assignment variable exceeds (or does not exceeds) a certain threshold. The regression discontinuity design was introduced by Campbell and Thistlethwaite in 1960 and since 1990s the RDD analyses are widely used for estimating of causal effects of programs in various fields of science. (Lee & Lemieux 2010) The RDD is frequently referred to a description of data generating process rather than a method. As Hahn *et al.* (2001) point out such data can be considered as something in between observational and experimental data. Although the experimental random data are preferable to observational, they are difficult or sometimes even impossible to reach. In those cases, the RDD can be a powerful tool for measuring of the treatment effect because even though the assignment to a treatment depends on a forcing variable and it is not random, the designs are closely related to randomized experiments.

As Lee & Lemieux (2010) describe in their publication, the basic idea of the design is that candidates just above the threshold, who do not receive the treatment, can be considered as a control group to those who belong just below the threshold and do receive the treatment. In this setting the treatment effect can be measured as a difference between mean values of outcomes of treatment and control groups near the cut-off, which may lead to inaccurate results. Another way of finding the treatment effect is to calculate the difference of the outcomes just at the cut-off point. Since it is impossible to observe an individual's outcome with and without treatment at the cut-off at same time, we can only ask what outcome would such hypothetical individual reach if he would/would not receive the treatment. Basically, we rely on approximation using the available data from treatment and control group. For those reasons we measure average treatment effect over a group of candidates rather than for individuals and we rely on local extrapolation. Lee (2008) shows that under few assumptions the RDD mechanism is "as good" as randomized experiment in the neighbourhood of the threshold. In other words, "close to this threshold, all variables determined prior to assignment will be independent of treatment status" and "any variable that is determined prior to the random assignment will have the same distribution in either the treatment or control state. " (Lee 2008) Basically that is the reason why it is possible to use the treatment and control group for the estimation and why the endogeneity should not be a problem in RDD estimation. Lee (2008) provides an explanation that it can be assumed that the assignment variable is partially determined by a random chance. To be able to claim this statement, two main conditions must be satisfied: (1) candidates cannot have precise control over the forcing variable and (2) the density function of assignment variable must evolve continuously. For formal proof we refer interested reader to publication written by Lee (2008).

Other conditions for internal validity are presented later.

In RDD approach the probability of receiving a treatment jumps discontinuously at the threshold. While in a sharp design the probability moves directly from 0 to 1, in fuzzy design the jump can be smaller. In other words, candidates who reach the treatment may not always get it and those who do not reach it may be assigned to it. As a structure of our data suggests, we use the fuzzy design in our analysis.

5.1.1 Fuzzy RDD

The Fuzzy regression discontinuity design (FRDD) is a special type of the design, in which individuals are assigned to the treatment imperfectly. In such design the probability of receiving the treatment changes by less than 1 in the cut-off point while requiring:

$$\lim_{\epsilon \downarrow 0} Pr(D = 1 \mid X = c + \epsilon) \neq \lim_{\epsilon \uparrow 0} Pr(D = 1 \mid X = c + \epsilon) \tag{5.1}$$

where X is an assignment variable, c is a proper threshold, D a dummy variable with value 1 if an individual receives a treatment and 0 otherwise and ϵ is an error term.

In this situation the treatment effect cannot be interpreted as an average treatment effect, since not all eligible candidates receive the treatment and vice versa. Instead, the treatment effect is calculated as "a ratio of the jump in the regression of the outcome on the covariate to the jump in the regression of the treatment indicator on the covariate." (Lemieux & Imbens 2008) The ratio is presented in the following equation:

$$\tau = \frac{\lim_{\epsilon \downarrow 0} E[Y \mid X = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[Y \mid X = c + \epsilon]}{\lim_{\epsilon \downarrow 0} E[D \mid X = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[D \mid X = c + \epsilon]} \tag{5.2}$$

where Y is an outcome and the rest of variables is the same as in previous equation.

Such an estimate is called a local average treatment effect (LATE). The LATE represents the group of individuals who were eligible for a treatment and received it and those who were not supposed to receive the treatment and they did not obtain it, i.e. compliers. Although it would be probably more convenient to estimate a treatment effect that could be applied for all observations, the LATE represents a consistent estimate of average treatment effect. For proper inference of the effect two conditions must be fulfilled: (1) monotonicity (two individuals with the same assignment variable at the threshold cannot belong to different groups) and (2) excludability (assignment variable at the threshold can affect outcome only through the treatment or in other words the density of assignment variable does not jump at the threshold). (Lee & Lemieux 2010) The fuzzy RDD is commonly estimated by two-stage least squares (2SLS) method.

5.2 Model

In this section we present the applied model with various specifications. Since the treatment distribution is imperfect, the fuzzy version of RDD model is used. The fuzzy design requires a two-stage model, which is estimated by the 2SLS method. The model is described below:

$$
\begin{aligned}
&(1) Treatment_{it} = \chi_0 + \delta Eligile_{it} + g(GDPproportion_{it} - Threshold) + v_{it} \\
&(2) Growth_{it} = \beta_0 + \tau Treatment_{it} + f(GDPproportion_{it} - Threshold) + \epsilon_{it}
\end{aligned}
\tag{5.3}
$$

where

Treatment is a dummy variable with value 1 when the treatment was received and 0 otherwise

Eligible is a dummy variable with value 1 when a region was eligible for the program and 0 otherwise

GDPproportion is a continuous variable representing a proportion of regional GDP per capita in PPS to EU average in reference years (1994-96 or 1997-1999 and 2000-2002)

Threshold is a value of a cut-off point, which represents a threshold for participation in the Objective 1 support program. Specifically, it is equal to 0.75 or 75

Growth is a continuous variable representing an annual average GDP per capita growth in PPS during the programming period (2000-2006 and 2007-2013)

ϵ and v are error terms

g(.), f(.) are functional forms of forcing variable

The first stage of the model can be considered as an instrument of the *Treatment* in the second stage of the model. For the purposes of estimation of the first stage, both linear and nonlinear probability models can be used. Since the valid probabilities of the treatment cannot be assured by a linear model, it is advised to use a nonlinear one. On the other hand, Lee & Lemieux (2010) claim that the difference in results while using linear and nonlinear model in the first stage should be minor and they suggest using a linear specification, which is much easier to implement. For this reason, we present both linear and nonlinear model for the first stage of the estimation. The nonlinear model is described more properly in the following equation:

$$P(Treatment_{it} = 1) = f(\chi_0 + \delta Eligile + g(GDPproportion_{it} - Threshold) + v_{it}) \tag{5.4}$$

According to the 2SLS methodology the predicted value of $Treat\hat{m}ent$ from the first stage is implemented in the second stage of estimation. In both stages the *Threshold* is subtracted from the forcing variable *GDPproportion* to move the intercept to the cut-off point for easier interpretation.

There are two ways of estimating the model: parametric and nonparametric approach. In parametric approach we try to fit proper polynomial function of various orders using the whole dataset. As many references recommend it is common to use the same functional forms in both stages of the estimation for practical reasons. Unlike the parametric approach the nonparametric approach only works with subset of the dataset. The idea behind this is to choose a proper size of a window around the cut-off i.e. a bandwidth, in which proper functional form of the regression can be estimated with greater certainty. Basically, since the main interest of the analysis are values at the cut-off, it would not harm the results if the observation far from the threshold would have no impact on them. For this reason, the nonparametric approach is frequently referred as a local polynomial estimation. Hahn *et al.* (2001) introduce a local linear regression methodology in context of RDD, which is a special case of local polynomial estimation. It is based on choosing bandwidth in which functional form of the regression is approximately linear. According to Lemieux & Imbens (2008) it is recommended to use the same bandwidth in both stages of the estimation, since in the opposite case each stage would be estimated based on different samples of data. The whole process results in local average treatment effect (τ) or LATE, which is the coefficient of the main interest. In other words, it represents an impact of Objective 1 treatment on annual average GDP per capita growth (intercepts) of a hypothetical region with a *GDPproportion* just at the threshold, which received and did not receive the treatment at the same time.

Since it is not straightforward to find a proper functional form of regression, we present polynomial functions of different orders to display treatment effect in various specifications. For the analysis we use polynomial functions of first, second, third and fourth orders for both stages of the regression[1]. In following equations, we present the various specifications of the model using above mentioned polynomials. Since many orders may lead to complex equations we present only the stage of the estimation with polynomial of the first and the second orders to keep the formulas as transparent as possible. We present models with interaction terms, which allow the treatment to affect the intercept and the slope of the regression as well. Moreover, we use *GDPproportion'* for ($GDPproportion_{it} - Threshold$).

$$
\begin{aligned}
Growth_{it} =&\beta_0 + \tau Treatment_{it} + \beta_1 GDPproportion'_{it} + \\
&+ \beta_2 GDPproportion'_{it} Treatment_{it} + \epsilon_{it} \\
Growth_{it} =&\beta_0 + \tau Treatment_{it} + \beta_1 GDPproportion'_{it} + \\
&+ \beta_2 GDPproportion'_{it} Treatment_{it} + \beta_3 GDPproportion'^2 + \\
&+ \beta_4 GDPproportion'^2 Treatment_{it} + \epsilon_{it}
\end{aligned}
\tag{5.5}
$$

As it was stated in the previous section the RDD mechanism can be very close to randomized experiments under some conditions. Lee & Lemieux (2010) point out that in such setting “the assignment to treatment is, by construction, independent of the baseline covariates”. It means that they do not have to be

[1] In case of nonparametric estimation, we use polynomials up to the third order due to practical reasons.

included in the model to get consistent estimates of treatment effect. On the other hand, they can be useful for improvement of precision of estimator by reducing the sampling variability. For those reasons we include them into the regression as an extension of the model. More importantly the covariates are used for validity tests of the research in later subsection.

Chapter 6

Estimation and Results

6.1 Graphical analysis

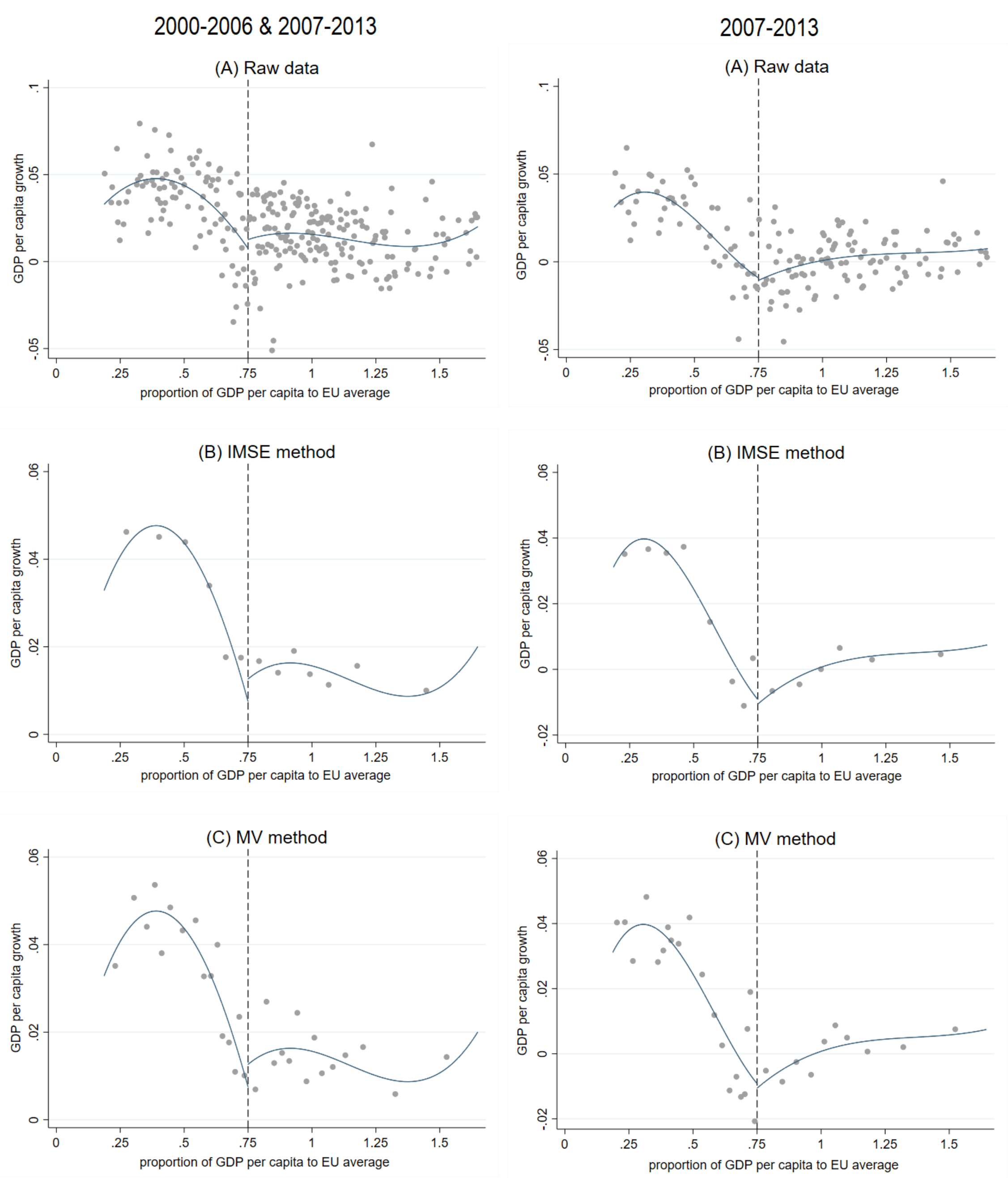
Most of the existing researches concerning the regression discontinuity design agree that the first step of the estimation should be a graphical analysis. As Lee & Lemieux (2010) point out the graphical analysis helps to reveal possible problems in the RDD implementation and to choose proper functional form for the regression model. For the graphical analysis we use a statistical software Stata, specifically a user-written package rdrobust. The process starts with dividing the forcing variable into equally sized or quantile sized intervals (referred to as "bins") to avoid noisy figures. Then the average value of the dependent variable in each bin is calculated and plotted against the midpoint value in the bin. This type of plot can be referred as "local sample mean" and it can be described as "a non-smooth approximation to the unknown regression functions". (Cattaneo *et al.* 2018) Another plot type is based on the fitting a line of proper order polynomial separately at each side of the cut-off. In other words, it is "a smooth approximation to the unknown regression function." (Cattaneo *et al.* 2018) In our analysis we use a combination of both types to retain the original data structure and visualization of the functional form and to observe the local composition of the data. Moreover, we use a quantile spaced specification of bins because such bins contain approximately same number of observations thus the comparability is preserved.

Probably the most challenging problem in this stage of analysis is to choose the width of the bins, since too wide bins may lead to biased results with lower visibility and too narrow bins may lead to imprecise and noisy results. Several tests were designed to detect the proper width. A straightforward method is to choose the width of bins based on visualization while applying various values. Since choosing based on the graphical visualization could be tricky, it is advised to also use a formal method. Calonico *et al.* (2015) present two formal procedures to select number of bins: Integrated Mean Squared Error Method (IMSE) and Mimicking Variance Method (MV). The idea behind the methods is a trade-off between bias and variability within a bin. In other words, with increasing number of bins the bias is lower, but there are also less observations in each bin, which leads to higher variability. The IMSE method minimizes the sum of the expansions of the variance and squared bias by assigning equal weights for both issues. The MV method is based on reaching roughly the same means variability as the true variability of original data. To compare these methods, we present results for both in the Figure 6.1[1]. We also present graphs showing distributions of raw data. Based on the IMSE method in case of two-period dataset, the annual averages of GDP per capita growth are calculated in 6 bins of average length of 9.7% on the left side and in 7 bins of average length of 12.8% on the right side and are plotted against the proportion of the GDP per capita to EU average during reference years. According to the MV method the average length of bins is 3.5% (16 bins) on the left side and 6.4% (17 bins) on the right side. In the second period the data are divided into bins of average length of 7% (8 bins) and 14.9% (6 bins) on the left and right side, respectively based on the IMSE method and into 20 bins of 2.8% length on the left side and 10 bins of 8.9% length on the right side based on the MV method. The solid line represents the 3^{rd} order polynomial function in all graphs. We use a polynomial of lower order to avoid possible overfitting problem, especially in the nonparametric estimation.

According to the Figure 6.1, there seems to be a positive jump at the cut-off in the two-period dataset, which suggests that there are probably differences in GDP per capita growth between treatment and control groups. While assuming the second period separately there does not seem to be any significant differences between the groups. The graphic analysis is rather used for general information about the data distribution than for analysing the treatment effect itself since in the graphs the data are not adjusted for the "fuzzy" nature of

[1]We exclude outliers using box plot analysis in order not to violate the results.

Figure 6.1: GDP per capita growth and GDP proportion



the design[2]. In other words, all regions are treated as perfect compliers, which is an incorrect assumption considering the official reports about eligibility of regions for the Objective 1 treatment. This misspecification will be corrected in the estimation process.

As it was stated in previous section, to use the fuzzy RDD mechanism correctly there must be a discontinuity at the cut-off in probability of receiving the Objective 1 treatment. By definition a size of the discontinuity is equal to 1 for the sharp design and it should be between 0 and 1 in the fuzzy regressions. The probabilities are presented in the Figure 6.2 and Figure 6.3 for two periods dataset and for the 2007-2013 period dataset, respectively. To display the probability in the graph, a nonlinear probability model with a probit function was used.[3] As it can be seen, in both cases there is a visible discontinuity at the cut-off, which means that the validity condition is fulfilled. Moreover, the size of the jump suggests that we apply the fuzzy regression correctly. In simple terms the probability figures represent the first stage of the estimation. Due to complicated calculation of the standard errors, it is common to use the same functional forms in both stages of estimation. In this case the probability values may exceed the limits (0 and 1). For this reason, we also present the graphical analyses of the probability of receiving the treatment using different polynomial in Appendix. In parametric estimation we also present results of a special estimation with a probit first stage designed by Wooldridge (2010).

[2]The discontinuity in the fuzzy estimations tend to be greater than in the sharp ones due to dividing the treatment effect difference by the jump in the regression of the treatment indicator, which should be between 0 and 1.

[3]The fourth order polynomial was used on predicted data by the first stage of the estimation.

Figure 6.2: Probability of receiving treatment to GDP proportion for both periods

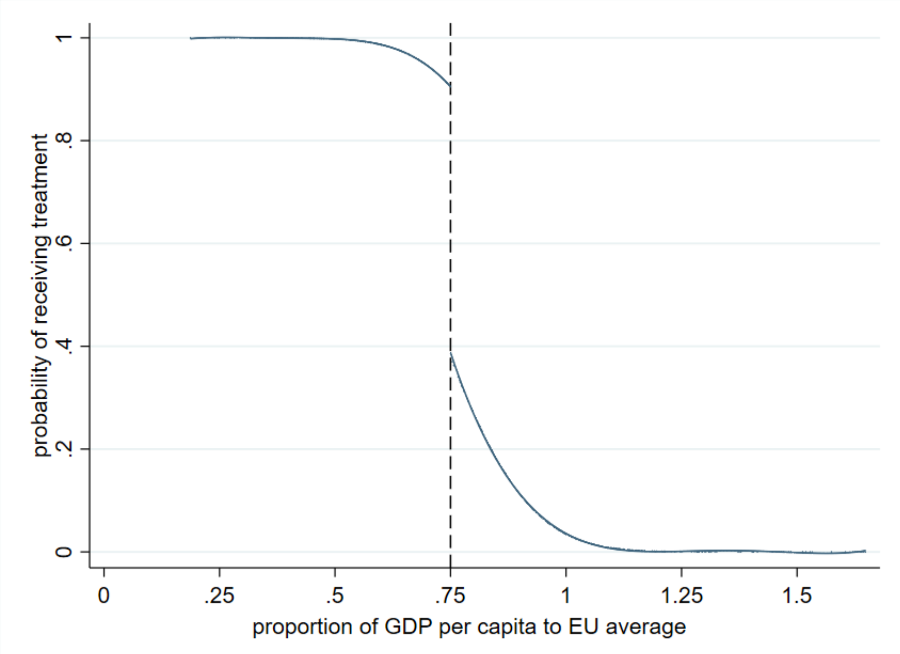
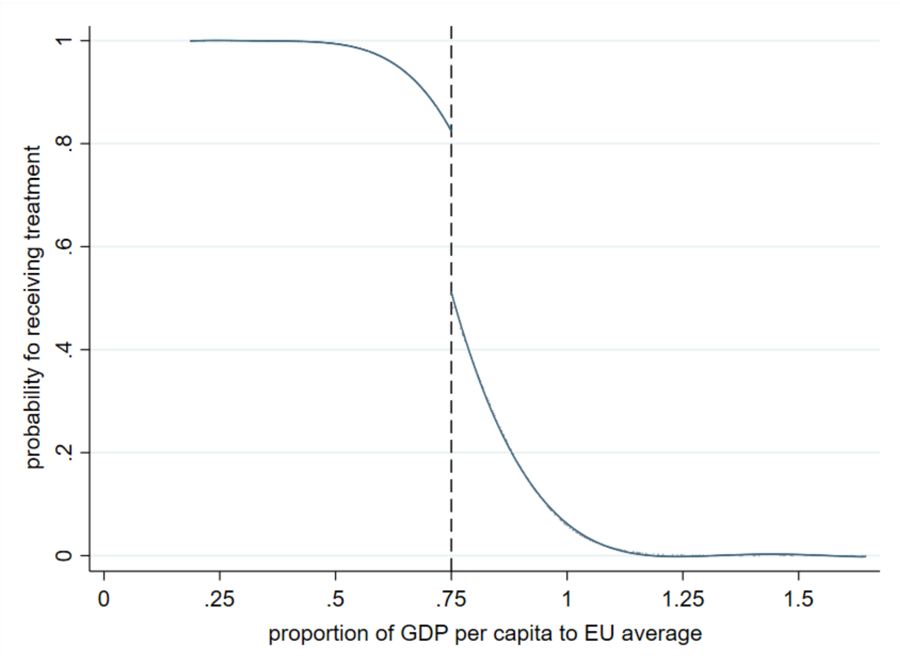


Figure 6.3: Probability of receiving treatment to GDP proportion for period 2007-2013



6.2 Estimation

According to the graphical analysis there seem to be a positive jump in GDP growth at the cut-off for the two-period dataset, which suggests that the treatment may impact the regions in negative way. The situation is different for the second programming period, in which the impact does not seem to be significant. However, as it was explained previously it is not possible to draw any conclusions from presented figures since the observations are treated as in the sharp design. To be able to reveal possible impacts of the Objective 1 treatment we need to use a proper "fuzzy RDD" estimation.

In this section the treatment effect is estimated using parametric and nonparametric approaches. The parametric method is based on fitting a proper function on the whole dataset to find a value of the outcome at the cut-off. In the nonparametric approach we try to find a proper subset of data in which proper functional form of the regression can be fitted with greater certainty. Jacob *et al.* (2012) describe the problem of deciding between parametric and nonparametric estimation method as follows. While the parametric approach ensures greater precision because of application of the whole dataset, the nonparametric approach works only with smaller subsample, which may cause trouble to obtain certain level of precision. On the other hand, it is much easier to implement suitable functional form of regression in smaller sample and to reduce possible bias in the results. Generally, it is advised to implement both methods and compare the results to check their robustness.

6.2.1 Parametric approach and results

One way of estimating the treatment effect is the parametric approach. As it was already stated the parametric method works with the whole dataset using a proper functional form of the regression in order to explain the relationship between the GDP per capita growth and the Objective 1 treatment effect. Since there is no formal process of revealing a proper polynomial function it is recommended to apply polynomials of various orders and to test how they fit the data.

As it was stated previously the fuzzy regression is specified by two-stage equations and it is estimated using 2SLS method. The first stage of the estimation is based on a prediction of the treatment effect as a linear or nonlinear probability model using variable *Eligibility* and adjusted variable *GDPproportion* in specific functional form. We present results for both linear and nonlinear probability model in the first stage. Moreover, to be able to calculate standard errors from 2SLS estimation it is recommended to use the same functional forms of the forcing variable in both stages. Since it is not trivial to implement nonlinear probability regression into 2SLS, we consult the theoretical background with the publication written by Wooldridge (2010). Wooldridge (2010) suggests obtaining fitted probabilities from the first stage of the estimation using a probit model and then use the fitted value as one of the instruments. We apply country-clustered standard errors in case of one-period dataset and we include fixed effects for two-period dataset.

In Table 6.1 the results of the parametric estimation for the second period are displayed. For the estimation we use country-clustered standard errors to control for a within-country correlation. Moreover, we performed Pagan-Hall's heteroskedasticity test, designed specifically for the 2SLS estimation.(Baum & Christopher 2006) Under this test we rejected the null hypothesis of homoskedasticity at low significance level for all model specifications (p-value less than 0.01). For this reason, we use robust standard errors. The results are divided into columns based on the functional form of forcing variable in both stages of the estimation. As it can be seen in the Table 6.1 the results do not seem to be statistically significant in any functional form of forcing variable. One of the reasons for the insignificant results is probably a lack of "truly complying" observations due to only one period analysis. Moreover, as it was explained in the previous sections our dataset may not fully correspond to the official values used in the decision-making process because of possible corrections and recalculations. While analysing different specifications of the model the results do not show any stable patterns, since the size and the sign of the treatment effect changes with polynomial orders. Such a divergence of results is not surprising because the approach is highly sensitive to the choice of polynomial order and it is difficult to reveal proper functional form for the whole dataset. Generally, it is believed that the parametric approach relies on higher-order polynomials to properly approximate the functional form of the regression. On the other hand, too high polynomial orders cause overfitting, which leads to unrealistic results. For this reason, we focus on estimation with polynomials up to the 4^{th} order. Although it is advised to present a range of polynomial orders, it is also important to have an idea which polynomial function fits the data best. For this reason, we use Akaike information criterion (AIC) estimator, which balances the bias and the variance of the model. In more details the AIC "increases with both the estimated residual variance as well as with the number of parameters, which moves in opposite direction". (Jacob *et al.* 2012)

The AIC formula is described below:

$$AIC = Nln(\sigma^2) + 2p \tag{6.1}$$

where N is number of observations, p is number of parameters in the model and σ^2 is estimated residual variance.

According to the Table 6.1 the fourth order polynomial function seems to fit the data best. We also use a probit nonlinear model in the first stage of the estimation. The result for such a model is presented in the last column. It is also important to note that the results tend to be higher in the fuzzy RDD specification than in the classic sharp design since due to imperfect compliance the treatment effect is divided by the jump in the regression of the treatment indicator, which should be between 0 and 1.

Table 6.1: Results of parametric estimation for the period 2007-2013

	Linear	2^{nd} order	3^{rd} order	4^{th} order	Probit
Obj 1 effect	-0.001	-0.015	0.020	-0.047	-0.003
	(0.007)	(0.016)	(0.146)	(0.048)	(0.005)
t-statistic	-0.100	-0.960	0.140	-0.970	-0.570
p-value	0.921	0.348	0.890	0.340	0.566
95% CI	(-0.02;0.01)	(-0.05;0.02)	(-0.28;0.32)	(-0.15;0.05)	(-0.01;0.01)
R-squared	0.363	0.362	0.312	0.259	0.369
AIC	-2135.519	-2134.001	-2110.461	-2077.596	-

Next, we analyse the dataset containing both programming periods. Lee & Lemieux (2010) states that unlike in case of more conventional method, it is unnecessary to include fixed effect for the panel data structure in the RDD approach. They suggest using the pooled-cross section estimation with corrected standard errors. On the other hand, they point out that it may help to reduce sampling variability. For this reason, we apply the 2SLS estimation controlling for regional fixed effects (also using robust standard errors). In this approach we are not interested in possible differences between treatment effects across programming periods, we use more periods only for extension of the dataset to design the experiment around the cut-off. The outcome of such an estimation is presented in Table 6.2. Again, we divide results into columns according to the functional form of forcing variable. Based on the results the treatment does not seem to significantly affect the regional GDP growth. Apart of issues described above, we believe that such results are caused because of difficult situation during the global financial crisis, which affects regions in specific ways. Specifically, the regions may not fully draw up funding, which may prevent the treatment effect from affecting the GDP growth. Moreover, the impact of investments in regional employment, infrastructure or technological progress would probably take more time than six years to bear fruits. On the other hand, the estimators of parametric estimation are strongly sensitive to the functional form specification, which may harm the results.

Table 6.2: Results of parametric estimation for period 2000-2006 and 2007-2013

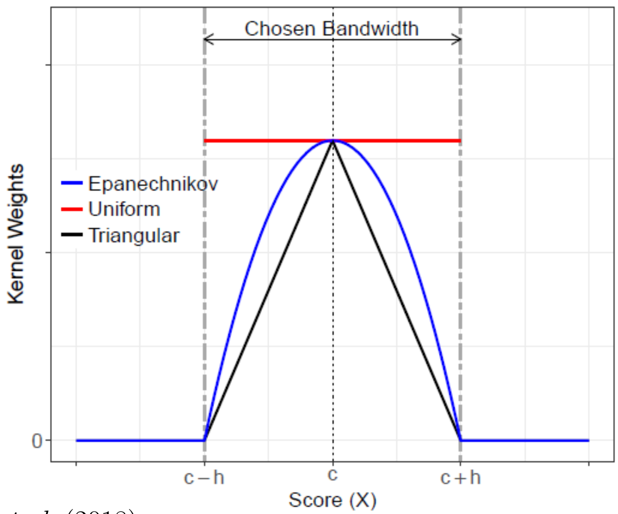
	Linear	2^{nd} order	3^{rd} order	4^{th} order
Obj 1 effect	0.054	0.050	0.003	0.004
	(0.038)	(0.031)	(0.034)	(0.048)
t-statistic	1.308	1.6	0.090	0.07
p-value	0.163	0.109	0.928	0.941
R-squared	0.158	0.160	0.155	0.140
95% CI	(-0.01;0.07)	(-0.01;0.01)	(-0.06;0.07)	(-0.09;0.09)

As it was indicated previously some authors may regard the results of parametric estimation as questionable. Since we are mainly interested in the intercept at the cut-off it might seem odd to work with the whole dataset and to assign equal weights to observations far away and to observations close to the cut-off point. Moreover, Lee & Lemieux (2010) point out that it is complicated to reveal the true functional form of the regression and incorrect specification can lead to biased treatment effects. Assuming the results of graphical analysis, it might seem hard to believe that the true functional form can be approximated by a liner function using complete dataset. Cattaneo *et al.* (2018) add that the parametric global polynomial approximation of functional form "tend to deliver a good approximation overall, but a poor approximation at boundary point", which is the main area of our interest. The parametric approach seems to be generally not ideal for the RDD approach and "starting with Hahn *et al.* (2001), the estimation of RD designs has generally been viewed as a nonparametric estimation problem". (Lee & Lemieux 2010) On the other hand authors also claim that the nonparametric method has its drawbacks and they recommend using both approaches to verify the robustness of the results and of the RDD mechanism itself. To summarize our findings, the results of the parametric estimation should be taken with discretion. Due to difficult revelation of correct functional form of the whole dataset there may be certain bias in the treatment effect. Moreover, allowing observations far from the threshold to impact treatment effect in the same way as observations close to this threshold may also not be intuitive for the RDD mechanism. However, these results may be still useful for general orientation and information about the treatment effect.

6.2.2 Nonparametric approach and results

As it was explained in the previous section, the parametric approach may not be suitable method for the RD analysis. For all reasons, explained in the previous section, modern empirical works mostly rely on the nonparametric estimation method. Unlike the global parametric method, the nonlinear approach is based on finding of a window around the cut-off in which proper function can be fitted. That is why the method is frequently called a local polynomial regression. It is believed that the local polynomial approach is "substantially more robust and less sensitive to boundary and over-fitting problem" (Cattaneo *et al.* 2018) since it is concentrated in the close neighbourhood of the cut-off point, which makes it less sensitive to outliers and misspecifications far from the threshold and it is generally approximated by low-order polynomial. On the other hand, the nonparametric estimation method is largely affected by the choice of the width of the interval around the cut-off i.e. bandwidth. A special case of the nonparametric estimation method is called a local linear regression. It is based on finding proper bandwidth, in which an unknown regression function can be approximated by a linear function. Beside the bandwidth selection the local polynomial regression requires a specification of polynomial order and a kernel function. The kernel function assigns "non-negative weights to each transformed observation based on distance between the observation's score and the cut-off." (Cattaneo *et al.* 2018) Most authors use a triangular kernel function, which assigns symmetric non-zero linear weights only to observations located in the window around the cut-off, because it has convenient properties for the RDD estimation. Figure 6.4 shows weights for different kernel functions. As it can be seen in the Figure 6.4 the Epanechnikov kernel function assigns non-zero quadratic weights to all observations within the bandwidth h. In both cases the maximal weights are set to values at the cut-off and the weights decrease as distance of the observation from the cut-off increases. The simplest function presented in the Figure 6.4 is uniform kernel function[4], which assigns equal non-zero weights to all observations within the bandwidth. According to Cattaneo *et al.* (2018) the choice of kernel function should not cause big difference in results. For our analyses we use a triangular kernel and we test the sensitivity of kernel choice in the last section.

Figure 6.4: Weights of different kernel functions



Source: Cattaneo *et al.* (2018).

The second parameter of choice in the nonparametric method is the local polynomial order. The simplest way how to estimate the treatment effect is to calculate a difference of mean values of outcomes at both side of the cut-off or in other words to apply polynomial function of order zero. As it was stated in the previous section the results of such estimation do not seem to be suitable for the RD design. It is advised to study the impact just at the cut-off point using approximation of the regression function instead. Authors point out that the precision of approximation increases with higher orders of polynomial, but the variability of estimator increases as well. On the other hand, too high orders of polynomial could lead to unreliable results. For those reasons the authors recommend using the local linear regression function and to adjust

[4]A uniform kernel function is frequently called a rectangular kernel function.

the bandwidth in a way that fits the data. The most important specification of the estimation is selection of proper bandwidth. Lee & Lemieux (2010) explain that the bandwidth selection depends on trade-off between precision and bias. Using large bandwidth results in more exact estimates, because the sample contains higher number of observations, which leads to smaller variance of estimated coefficients. However, the results are more likely to be biased since it is difficult to implement correct regression function, especially in case of the local linear regression. Choosing smaller bandwidth fixes the bias in approximation but it leads to higher variance since there are fewer observations in the sample. For those reasons the bandwidth selection is a crucial step in the analysis. As suggested by Cattaneo *et al.* (2018) one of possible methods for selection of optimal bandwidth is minimizing the Mean squared error (MSE), which is defined as a sum of squared bias and variance of the estimator. Authors describe the process of MSE estimation, which is based on "deriving an asymptotic approximation to the MSE of treatment effect, optimizing it with respect to bandwidth h, and using data-driven methods to estimate the unknown quantities in the resulting formula of the optimal h." (Cattaneo *et al.* 2018) For our purposes the main idea of the MSE-optimal bandwidth selection using the trade-off between bias and the variability of estimator is sufficient. Putting these three parameters together, we should be able to estimate local polynomial estimator of treatment effect.

Cattaneo *et al.* (2018) point out that in order to test hypotheses and form confidence interval properly it is necessary to adjust estimates for bias. Basically, since we choose the bandwidth based on the bias-variance trade-off, it would be incorrect to use a conventional OLS method and to ignore the non-parametric form of the estimation (or in other words to act like there is no bias). While using the conventional least-squares estimation (as in parametric approach) we assume that the approximated functional forms are the same as the true functions, which would be correct in very few cases. For this reason, practitioners designed different procedures for conducting a valid inference in cases when the bias seem to be significant. Cattaneo *et al.* (2018) describe few alternatives based on adjustments of bandwidths, estimates or standard errors. In the first procedure the smaller bandwidth than MSE-optimal is used for inference calculations, since in this case the bias is theoretically negligible in large sample approximation. In this procedure there are no empirical rules about the length of the smaller bandwidth and it is generally not suitable for practice use. The second alternative is based on creating confidence intervals around the bias corrected estimates, which lead to more reliable results, however it does not allow to use the same bandwidth for treatment estimate and for bias estimate and it does not include the extra variability from the bias estimation step. The last procedure, which uses the robust bias correction in the inference calculation, incorporate the variability from the bias estimation in the variance of point estimate and it allows to use the same bandwidth for both steps. Moreover, "the robust bias correction approach delivers valid inferences even when the MSE-optimal is used". (Cattaneo *et al.* 2018) According to authors' advice we use the MSE-optimal point estimate and the robust bias corrected (RBC) t-statistic and confidence intervals for valid statistical inference. A formula for 95% confidence intervals is presented below:

$$CI_{rbc} = [(\hat{\tau} - \hat{B}) \pm 1.96\sqrt{V_{bc}}] \tag{6.2}$$

where $\hat{B}$ is estimated bias and V_{bc} is adjusted variance.

Authors also describe the procedure, in which the optimal bandwidth for statistical inference is used. Basically, in this process we use different bandwidths for point estimate and for the inference calculations. The inference bandwidth should minimalize the coverage error, which is "the discrepancy between the empirical coverage of the confidence interval and its nominal level." (Cattaneo *et al.* 2018) We also present the results using this coverage error (CER) bandwidth for statistical inference.[5]

The results for nonparametric estimation for the second programming period are presented in Table 6.3 and for both periods in Table 6.4. The tables are organized into columns based on the functional form of forcing variable in both stages of the estimation. We apply functions with polynomials up to the third order in nonparametric estimation to minimize overfitting problem. The treatment effects and standard errors, which are presented in top part of the table, are estimated using the MSE-optimal bandwidth h displayed in the third row. The bandwidth is the smallest for the linear function, since the interval around the cut-off must be narrow enough so that the true functional form could be approximated by the linear form. As we already discussed ear-

[5]The technical and methodological details of RBC and CER confidence interval and t-statistics calculations are presented in papers written by Calonico *et al.* (2018a) and Calonico *et al.* (2018b).

lier, in such an interval there are less observations, which commonly leads to insignificant results. Moreover, the approximation is more sensitive to overfitting and the estimation tends to be highly influenced by possible outliers. For proper statistical inference we calculated robust bias corrected confidence intervals and t-statistics, presented in the middle part of the tables.[6] Basically, we centred the CI around the treatment effect adjusted by estimated bias and new variance and we also recalculated t-statistics. For the bias estimation we used the same bandwidth as for the point estimation. The last part of tables is dedicated to the statistical inference estimated using coverage error bandwidth for bias estimation. The CER bandwidth should have optimal length for statistical inference and it is wider than the RBC bandwidth. Since the use of the CER optimal bandwidth for the point estimation leads to suboptimal results, we apply the CER bandwidth only for the bias estimation.

Table 6.3: Results of nonparametric estimation for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.003	-0.017	-0.019
	(0.094)	(0.805)	(0.793)
MSE bandwidth h	0.185	0.234	0.267
RBC t-statistic	-0.143	-0.264	-0.030
RBC p-value	0.887	0.792	0.976
RBC CI	(-0.248;0.214)	(-0.210;0.160)	(-1.772;1.827)
CER bandwidth b	0.288	0.332	0.340
CER t-statistic	-0.105	-0.046	-0.273
CER p-value	0.916	0.963	0.785
CER CI	(-0.278;0.250)	(-2.370;2.261)	(-0.435;0.329)

In Table 6.3 the treatment effect does not seem to be significant in any polynomial specification. Since the nonparametric method works only with a subset of observations, it requires a lot of observations in cut-off neighbourhood. Although the results of the nonparametric estimation should be more reliable for all reasons explained earlier, we still did not find significant results.

In Table 6.4 we present the results for the both programming periods using clustered standard errors to control for a within-individual correlation.[7] The

[6]To keep the table as clear as possible we do not present the point estimate and standard error for RBC bandwidth estimation nor for the CER bandwidth estimation.

[7]There is no existing process for estimation of the suitable bandwidth for the panel data. For this reason, we do not use the fixed effect method for the estimation.

Table 6.4: Results of nonparametric estimation for periods 2000-2006 and 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.013	-0.013	-0.019
	(0.251)	(0.112)	(0.088)
MSE bandwidth h	0.16	0.219	0.337
RBC t-statistic	-0.057	-0.145	-0.23
RBC p-value	0.955	0.884	0.818
RBC CI	(-0.603;0.569)	(-0.271;0.234)	(-0.215;0.170)
CER bandwidth b	0.288	0.311	0.446
CER t-statistic	0.365	-0.164	-0.101
CER p-value	0.715	0.870	0.920
CER CI	(-13.295;19.385)	(-0.507;0.429)	(-0.319;0.288)

treatment status also does not seem to have statistically significant impact in any polynomial specifications. As we already explained the nonparametric estimation requires high concentration of observations around the cut-off, so we would probably need another programming period to extend the dataset.

As it can be seen in next sections we analyse strength of the impact in each year of the programming period individually to investigate the development of the treatment. It is also important to note that due to correction and redefinition of the GDP we had to make several adjustments of the dataset, which may lead to differences from values that were officially used. We comment on these adjustments in detail in the section Data description. Moreover, the dataset is not well balanced because out of non-complying regions there are more regions that received the treatment and were not eligible than regions that did not receive the help but were eligible. Such a discrepancy makes sense from the point of view of region policy makers, since the policy provides transitional help to regions that were not eligible for the support due to accession of new countries to EU, which caused a decrease in the EU average of GDP per capita. In the opposite case the regions did not receive the treatment in less cases, for example because information about their GDP per capita was not available. In this case it is natural that we also miss this information since we get the data from the EU's official source.

In Table 6.5 and Table 6.6 we present the results of nonparametric estimation using different bandwidths at each side of the threshold. The MSE-optimal

Table 6.5: Results of nonpar. estimation with different bandwidths for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.008	-0.009	0.005
	(0.096)	(0.057)	(0.056)
MSE bandwidth $h-$	0.173	0.239	0.297
MSE bandwidth $h+$	0.189	0.286	0.439
RBC t-statistic	-0.088	-0.101	0.244
RBC p-value	0.930	0.920	0.807
RBC CI	(-0.234;0.214)	(-0.139;0.126)	(-0.111;0.142)

Table 6.6: Results of nonpar. estimation with different bandwidths for periods 2000-06 and 2007-13

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.024	-0.023	0.019
	(0.225)	(0.128)	(0.090)
MSE bandwidth $h-$	0.211	0.186	0.316
MSE bandwidth $h+$	0.155	0.234	0.332
RBC t-statistic	-0.280	-0.149	0.202
RBC p-value	0.779	0.881	0.840
RBC CI	(-0.598;0.449)	(-0.313;0.268)	(-0.217;0.177)

bandwidth for the left side of the cut-off is showed in the third row (h-) and for the right side in the fourth row (h+). For statistical inference we use the RBC procedure with the same bandwidths for bias estimation. As it can be seen the treatment effect also does not seem to be statistically significant. To sum up the treatment does not seem to affect GDP growth significantly. Apart from possible explanations presented previously, the insignificant results are probably caused by lack of observations around the threshold. The results also seem to be sensitive to the bandwidth choice and due to lack of observations around the cut-off the results tend to be highly influenced by each observation. We analyse this sensitivity in the final section.

It was already stated at the beginning of this section that it is very difficult to choose between the parametric and nonparametric estimation processes. While some authors claim that the RDD problems should be estimated by the nonparametric approach, others do not use this process at all (Becker *et al.* 2016), (Becker *et al.* 2018) or just for robustness checks (Becker *et al.* 2010).

We would also like to present here the opinion of Lee & Lemieux (2010). They point out that incorrect specification of parameters in both approaches leads to biased results and it is impossible to decide which misspecification is more serious in a finite sample without having information about the true function. They recommend using both methods with more specifications since “results that are stable across alternative and equally plausible specifications are generally viewed as more reliable than those that are sensitive to minor changes in specification. RD is no exception in this regard.”(Lee & Lemieux 2010) According to this theory we implemented both methods to reveal a possible effect of the Objective 1 treatment and to deal with problems related to the estimation of impacts of certain policy in general. In both specifications the impact of Objective 1 treatment does not seem to be statistically significant. In the next section we focus on model extensions to try to analyse characteristic of the RDD method and the regional policy itself.

6.3 Validity tests

It was already stated in the previous section that the RDD mechanism is regarded as a powerful tool for revealing the treatment effect. The main advantage of this approach is its close connection to the randomized experiments. While in most approaches it is very difficult to get an unbiased estimate of treatment effect due to the endogeneity problem, the RDD method is designed to use the similarity to randomized experiment to calculate a valid estimate. The endogeneity problem (or simultaneity problem) arises when an outcome is affected by both an explanatory variable directly and by error term through the explanatory variable. Basically, there is a correlation between the explanatory variable and the error term, which causes biased results of estimation. As it was already explained the main idea of the RDD method is that individuals just above the threshold are very similar to individuals who belong just below the threshold and can be included in the control group as opposed to the treatment group, which receive the treatment. In extreme case we can consider both individuals (groups) to be the same in all characteristics except the treatment.[8] Theoretically we assume a situation when an individual located just at the

[8]For this reason, one way how to estimate the treatment effect is to compare mean values of outcomes of treatment and control groups in a certain interval around the cut-off point. However, such estimation would be probably inaccurate, and it is advised to use approximation to the values just at the threshold as we presented in the previous section.

cut-off receives the treatment and does not receive the treatment at the same time and we are analysing the difference between these two situations using the approximation of the function. In other words, we design an experiment from selected observations in the neighbourhood of the threshold in order to estimate the local treatment effect. To be able to employ the method properly, certain assumptions must be fulfilled. As in the case of a randomized experiment, the RD design is highly sensitive to self-selection of the individuals and misspecification in definition of treatment and control group. The self-selection problem may arise when an individual knows the rule that assign treatment and he may try to manipulate its running variable to receive the treatment. The assigning process may also be manipulated by decision makers if they try to privilege some individuals. In order to reveal such rule violation, one has to know and fully understand the whole assigning process, which is frequently unfeasible. For those reasons the RD design relies on empirical validation methods.

As in case of a randomized experiment, it is important to have transparent information about the sampling process to reveal possible manipulation. (Lee & Lemieux 2010) define three types of such manipulation based on the way how can a candidate impact the assigning variable: complete control, precise control and imprecise control. They explain that while it is unlikely for candidates to have complete or precise control over the selection process, in real-life situations they may impact the assigning variable imprecisely. They point out that such imprecise control over the rating variable should not violate the RDD analysis. Authors show that "the behavioural assumption that individuals do not precisely manipulate assigning variable around the threshold has the prediction that treatment is locally randomized" (Lee & Lemieux 2010) and that the RDD approach is valid. To be able to examine the nature of individual's control over the rating variable we use a test based on the density of the rating variable (proportion of GDP per capita). In Figure 6.5 and Figure 6.6 we display the density of the assigning variable for the second and both programming periods, respectively. It is obvious that most observations are situated around the EU average. More importantly, the number of observations in treatment and control group around the cut-off seems to be similar, which suggest that there should be no manipulation in sampling process.

Since the graphical analysis of the assigning variable is rather indicative, it is advised to use proper statistical test to exclude possible manipulation of selecting process. (Cattaneo *et al.* 2018) describe a statistical test, in which a

Figure 6.5: The histogram of assigning variable for period 2007-2013

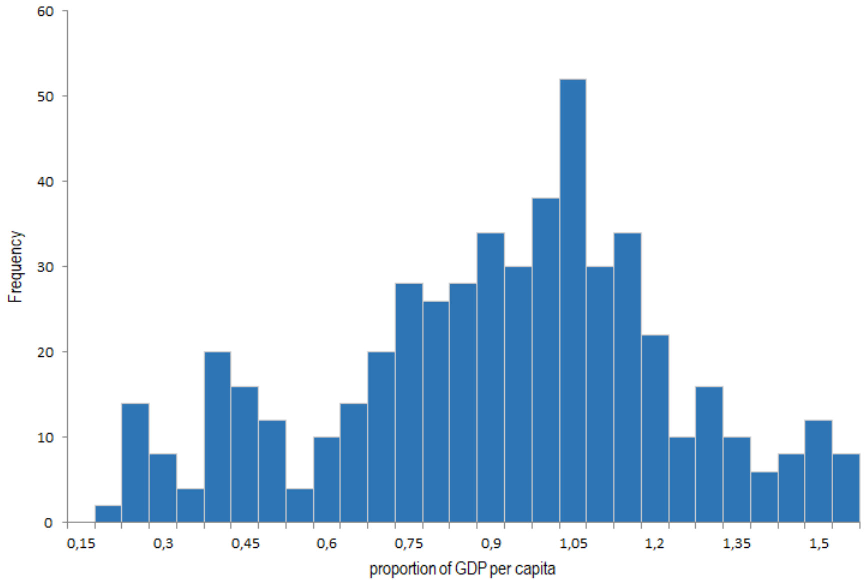
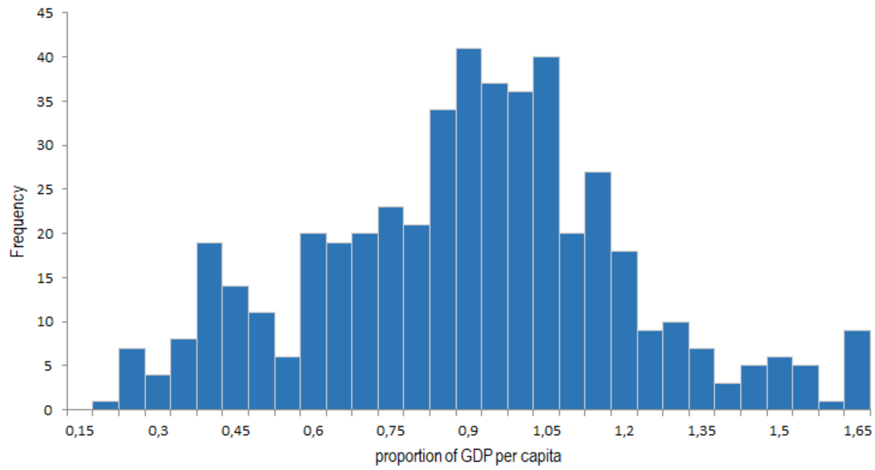


Figure 6.6: The histogram of assigning variable for periods 2000-2006 and 2007-2013



null hypothesis is that the density of the assigning variable is continuous at the cut-off. They explain that the test is based on "local polynomial estimation of the density of observations near the cut-off separately for observation above and below the cut-off". (Cattaneo *et al.* 2018) We apply this test using the statistical package *rdrobust* in Stata. The results of the test confirm our previous analysis since we fail to reject the continuity hypothesis with p-value 0.3019 for the both programming periods. We also fail to reject the null hypothesis for the second programming period with p-value 0.6392. To sum up our findings, we can use the RDD approach without worries about possible manipulation of the assigning variable.

In previous section we presented the results of our estimation. To be able to draw conclusions about the treatment impact we must put the validity of the RDD to the test, since there is still a chance that the difference in outcomes for treatment and control group is caused by another effect than by the treatment and that these groups are not that similar as we thought. Another test for internal validity is based on examination of the baseline covariates in the neighbourhood of the cut-off. The main idea is to analyse the relationship between specific baseline covariate and assigning variable. The test is based on examining whether the covariates are balanced on both side of the cut-off and whether the local randomization is not ruled out. Since the covariates are predetermined prior to the assigning process, they should not be affected by the treatment status. In other words, we try to graphically and empirically investigate whether there are significant differences between treatment and control groups (on either sides of the threshold) in other characteristics than treatment status prior to the selection process, which would violate the RDD validity assumption.

The results of graphical analysis are presented in the Figure 6.7 and Figure 6.8 for period 2007-2013 and for both periods 2000-2006 and 2007-2013, respectively. For all cases the 3^{rd} order polynomial functions are used. As it can be seen the graphical analysis did not reveal any visible discontinuities for most cases, except for population density and unemployment rate for the second period 2007-2013 and for unemployment rate for periods 2000-2006 and 2007-2013. Since we want to display functional characteristics of the dataset, we use the whole dataset for the graphical analysis. However, to be able to draw conclusions about the internal validity of RD design, we must also perform an empirical test around the cut-off. Moreover, it is necessary to investigate the suspicious covariates from the graphical analysis. For this purpose, we apply nonparametric estimation using the MSE-optimal bandwidth for point estimate and standard errors calculations and the robust bias corrected bandwidth for statistical inference. We present the results of suspicious covariates in Table 6.7 and Table 6.8 for each dataset and the results for other covariates are presented in Appendix. As it can be seen we did not find any statistical evidence that there are discontinuities at the threshold since we fail to reject the hypothesis that the RDD mechanism is appropriate for our research.[9]

Table 6.7: Internal validity test of population density and unemployment rate for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Pop. density	1769.3	545.48	-482.21
	(8172.2)	(1489.4)	(1107.1)
RBC t-statistic	-0.4153	0.1295	-0.1695
RBC p-value	0.678	0.897	0.865
RBC CI	(-22820.1;14840.8)	(-2934.2;3349.36)	(-2742.69;2306.05)
Unempl. rate	-0.421	-0.207	-0.509
	(1.126)	(0.296)	(1.435)
RBC t-statistic	0.049	-0.325	-0.265
RBC p-value	0.961	0.745	0.791
RBC CI	(-2.406;2.531)	(-0.779;0.558)	(-3.354;2.556)

Table 6.8: Internal validity test of unemployment rate for both periods 2000-2006 and 2007-2013

	Linear	2^{nd} order	3^{rd} order
Unemployment rate	0.101	0.223	0.711
	(1.350)	(0.404)	(1.461)
RBC t-statistic	-0.458	0.645	0.621
RBC p-value	0.647	0.619	0.534
RBC CI	(-3.629;2.253)	(-0.574;1.137)	(-1.982;3.821)

[9] As we already stated previously there may be problems with lack of observations or with corrections and adjustments of the dataset. For these reasons the validity tests may be different when extending the dataset for another period.

Figure 6.7: Internal validity test of covariates for period 2007-2013

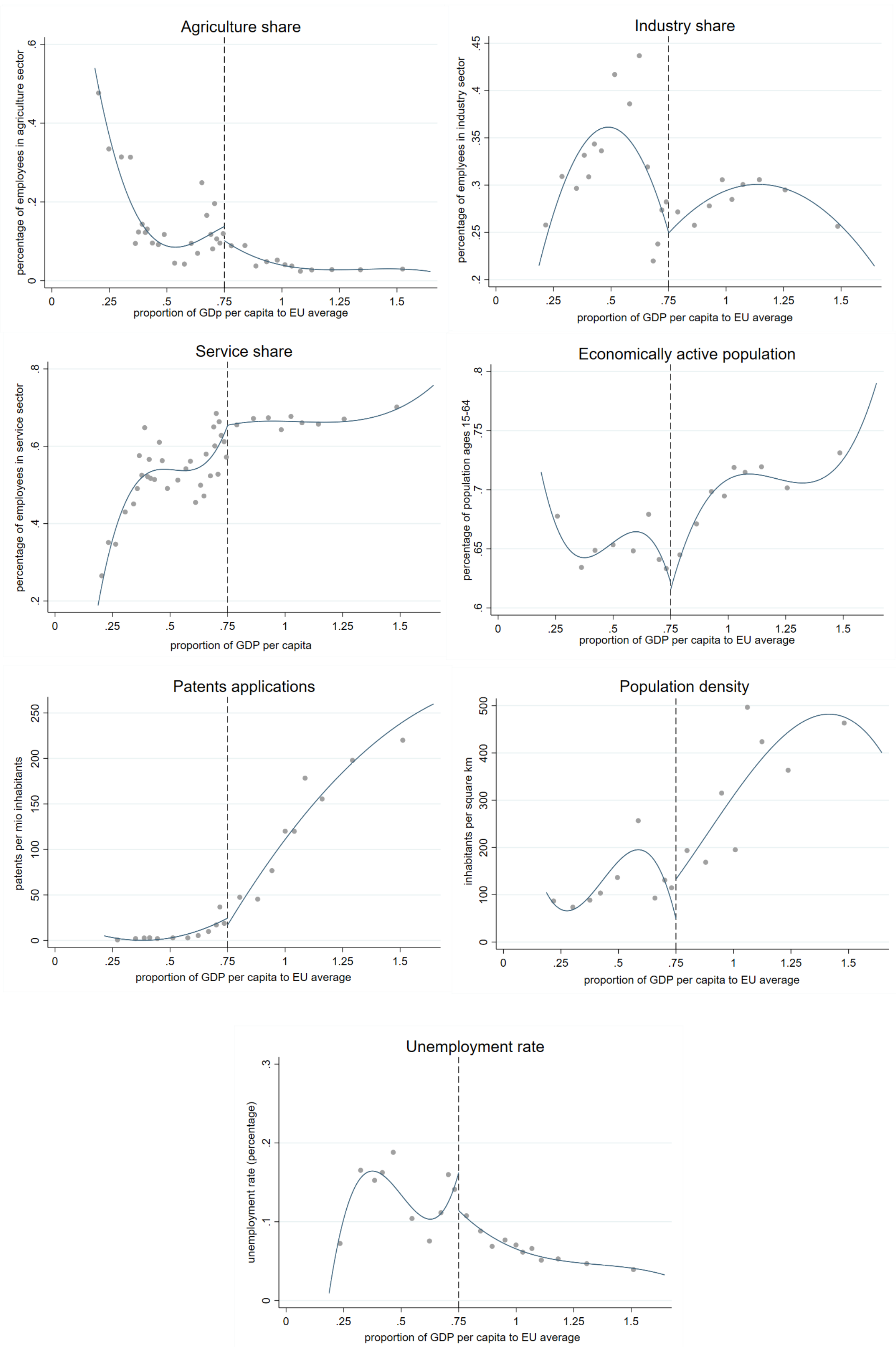
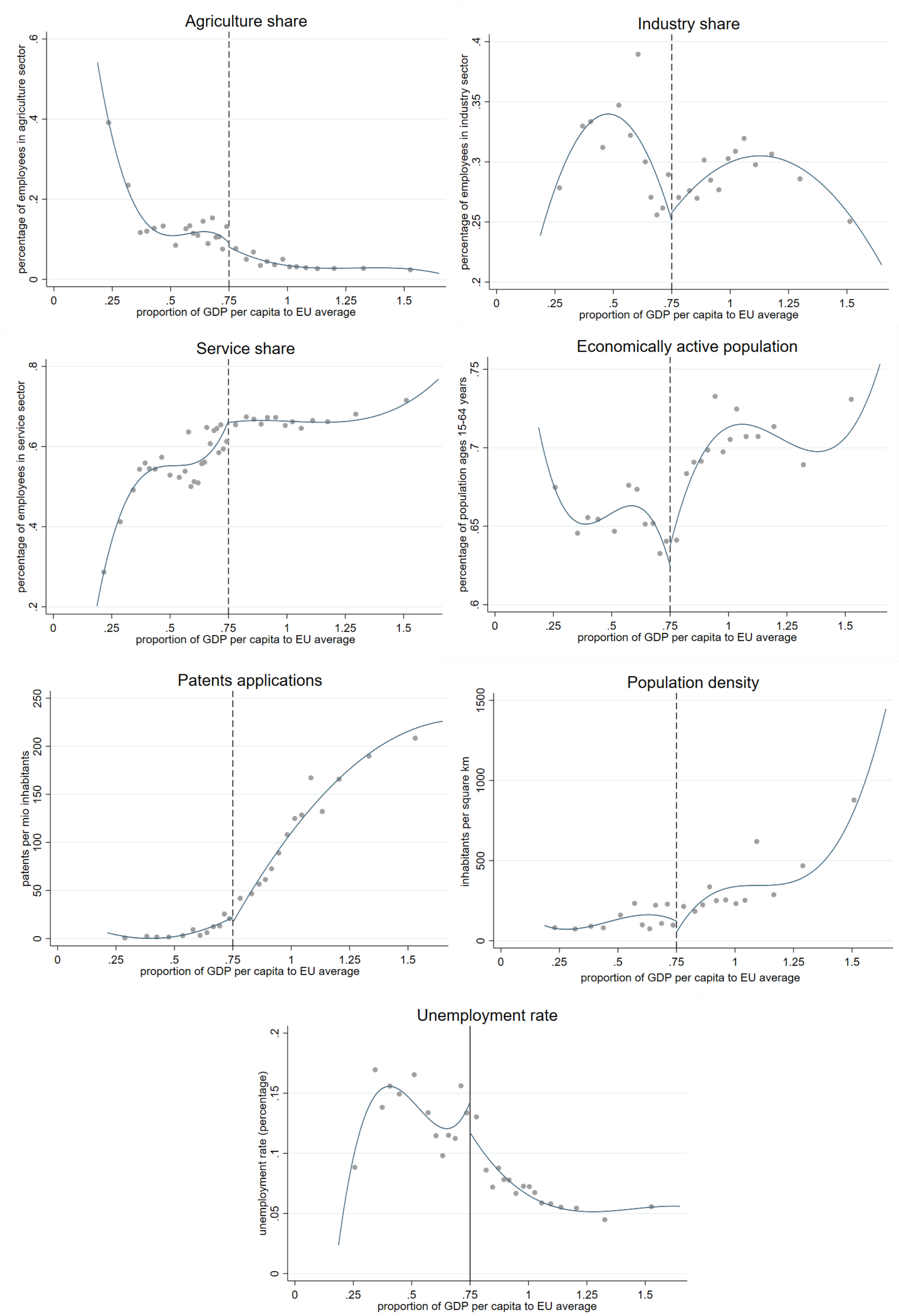


Figure 6.8: Internal validity test of covariates for both periods 2000-2006 and 2007-2013



6.4 Model extensions

6.4.1 Estimation including covariates

As it was explained in previous sections, under certain conditions the RDD mechanism is very close to the randomized experiment. In such setting the treatment effect should be independent of covariates and it is not necessary to include them in the model to get consistent estimates of treatment effect. On the other hand, they could be useful for checking the RDD validity and for improving the precision of the results. In Table 6.9 we present the results of nonparametric estimation including covariates, which were used for internal validity tests. Inclusion of covariates did not cause any significant changes in treatment effects and it led to narrower confidence intervals, which justifies the use of the RDD method. The impact of the treatment still does not seem to be statistically significant.

Table 6.9: Results of nonpar. estimation incl. covariates for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.006	-0.015	-0.009
	(0.068)	(0.044)	(0.047)
MSE bandwidth h	0.173	0.190	0.247
RBC t-statistic	-0.109	-0.330	-0.114
RBC p-value	0.913	0.741	0.910
RBC CI	(-0.163;0.146)	(-0.114;0.081)	(-0.106;0.094)

6.4.2 Objective 1 treatment impact during programming periods

In this subsection we analyse an accumulation of treatment effect over the programming periods. In other words, we want to investigate whether the impact of Objective 1 evolves over time. For our previous analyses we used average annual GDP per capita growth as a dependent variable to deal with a problem of missing data. Moreover, it seems right to apply such specification to better reflect the GDP per capita evolution. Similarly, we use average annual GDP per capita growth up to certain year to estimate the intensity of the treatment during the programming periods. In Table 6.10 and Table 6.11 we present results of such estimation for period 2007-2013 and for both periods

2000-2006 and 2007-2013 together, respectively. To keep result presentation as transparent as possible, we present only the treatment effects and standard errors using the MSE-optimal bandwidths with a sign of $***$ for statistical significance at 1% level, $**$ for significance at 5% level and $*$ for significance at 10% level, calculated based on the robust-bias corrected statistical inference. We apply functions with polynomials up to the third order in nonparametric estimation. To check the sensitivity of the results we use both methods of the same MSE-optimal bandwidths and different MSE-optimal bandwidths on each side of the threshold. According to these analyses the results seem to be consistent in both specifications. For purpose of transparency of the text, we present only one of the methods for each dataset. Specifically, we use the same bandwidths on each side of the cut-off for the second period 2007-2013 (Table 6.10) and different bandwidths on each side for both programming period together (Table 6.11).

Table 6.10: Development of treatment effect during period 2007-2013 (the same bandwidths on both sides)

	Linear	2^{nd} order	3^{rd} order
Objective 1 - 1^{st} year	0.033	-0.089	0.004
	(0.103)	(0.402)	(0.219)
Objective 1 - 2^{nd} year	0.005	-0.038	-0.039
	(0.060)	(0.142)	(0.128)
Objective 1 - 3^{rd} year	-0.025	-0.036	-0.050
	(0.120)	(0.090)	(0.418)
Objective 1 - 4^{th} year	-0.026	-0.040	-0.014
	(0.138)	(0.097)	(0.430)
Objective 1 - 5^{th} year	-0.003	-0.017	-0.019
	(0.094)	(0.080)	(0.793)

To sum up the results of this analysis, the impact does not seem to be statistically significant in any year. The results are consistent among different datasets and methods. Generally, using more extended dataset (for example by another programming period) may lead to different results, since the observations around the threshold are limited. Moreover, there are several issues with the data, described in previous sections. As it was already noted, we used the average values of GDP per capita as dependent variables in our analyses also due to problem of missing data. In order to reflect possible side effects of this decision, we also estimate the treatment effect using the cumulative GDP per capita growth with the first year of programming period as a base year. The

Table 6.11: Development of treatment effect during programming periods (different bandwidths on each side)

	Linear	2^{nd} order	3^{rd} order
Objective 1 - 1^{st} year	0.003	-0.088	-0.098
	(0.151)	(0.253)	(0.187)
Objective 1 - 2^{nd} year	0.012	-0.068	-0.042
	(0.224)	(0.209)	(0.129)
Objective 1 - 3^{rd} year	-0.003	-0.061	-0.042
	(0.120)	(0.177)	(0.111)
Objective 1 - 4^{th} year	-0.010	-0.034	-0.029
	(0.195)	(0.134)	(0.095)
Objective 1 - 5^{th} year	-0.025	-0.036	-0.035
	(0.189)	(0.126)	(0.097)

results of such estimation are presented in Table A.4 in Appendix for period 2007-2013, we omit analysis of both periods together due to missing data. The results seem to be consistent with previous analyses.

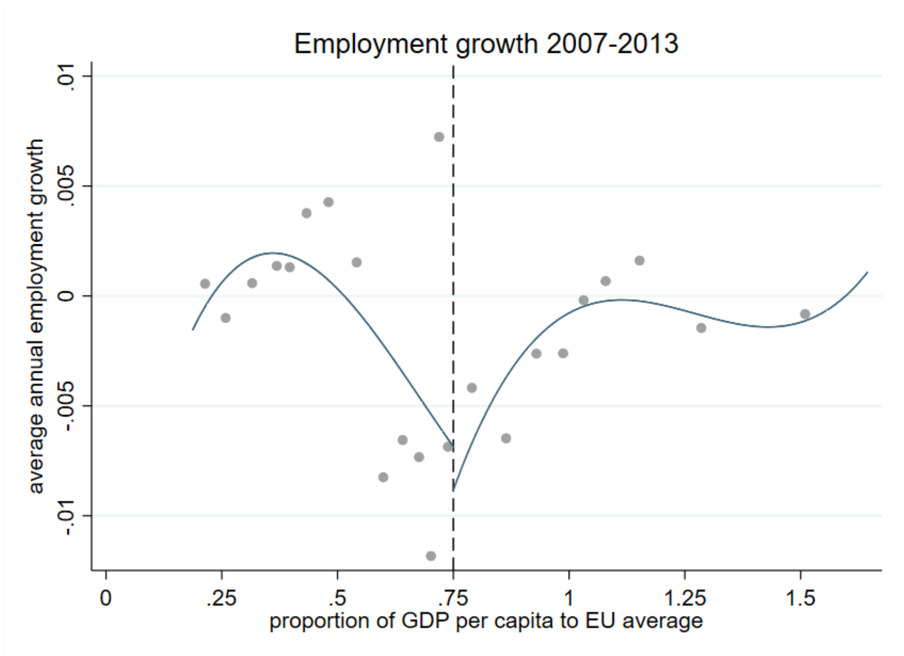
6.4.3 Objective 1 treatment impact on employment growth

Many projects, financed based on the Objective 1 criterium, focus also on supporting full and fair employment[10], which contributes to the economic performance of regions. For this reason, we also analyze how the Objective 1 treatment affects the employment growth in regions. For our analysis we use an average annual employment growth as a dependent variable, since it describes the employment development better and it deals with the problem of incomplete dataset. Firstly, we display the data graphically in Figure 6.9 and Figure 6.10 for period 2007-2013 and periods 2000-2006 and 2007-2013 together, respectively. For both analyses, we use the Mimicking Variance Method (MV) and polynomial functions of third order. In Figure 6.9 the annual average employment growths for the second programming period are calculated in 14 bins of average length of 4% on the left side and in 9 bins of average length of 9.9% on right side. In Figure 6.10 the average employment growths for both programming periods are based on calculations in 13 bins of 4.3% length on the left side and in 13 bins of 6.9% length on the right side. In both cases, the figures display a small negative jump at the threshold, which suggests that

[10]Although there are other objectives that are specialized on employment, the Objective 1 (with the largest budget) also affects this area in direct or indirect way.

the treatment should have a negative impact on the regional employment. As we noted previously, it is not possible to analyze the treatment effect itself from the graphical analyses, since the figures are not adjusted for the complying/noncomplying data structure. On the other hand, the figures help us with a general description of data structure.

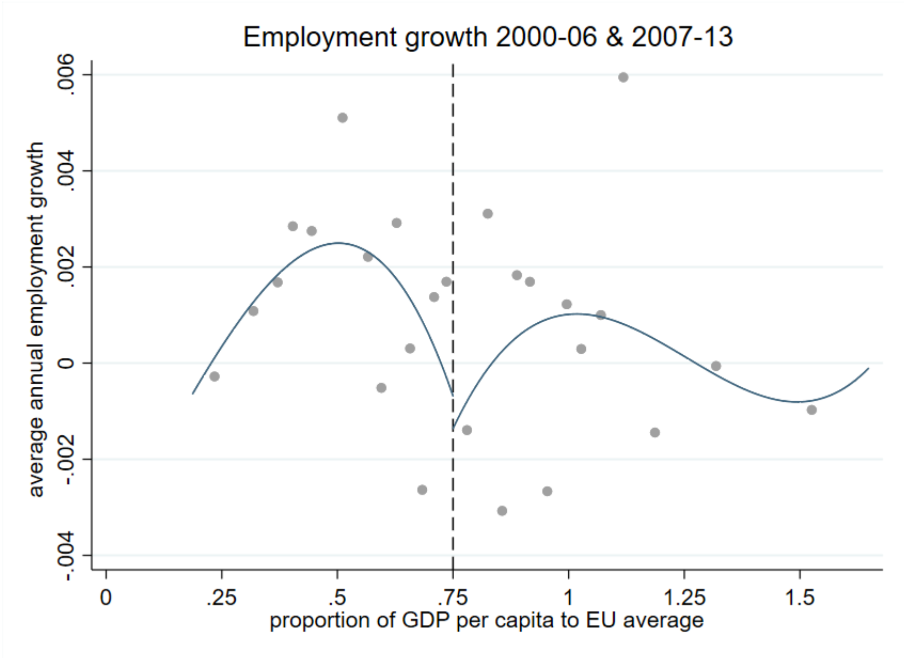
Figure 6.9: Employment growth and GDP proportion for period 2007-13



To be able to make conclusion about the Objective 1 impact on the employment growth in regions, it is necessary to perform a proper analysis. As in the previous cases we use a nonparametric "fuzzy" estimation[11] with polynomials up to the third order to reveal a treatment effect. In Table 6.12 and Table 6.13 we show the results of such estimation for period 2007-2013 and for both periods 2000-2006 and 2007-2013, respectively. The tables are organized into columns based on the polynomial order of forcing variable. For the treatment and standard errors estimation, we use the MSE-optimal bandwidths h, which can differ on each side of the threshold. The concrete values of the bandwidths are presented in the third rows of both tables. The last parts of the tables describe the statistical inference of the treatment effect using the robust bias correction approach. To sum up our findings, although the figures suggests a negative treatment effect on the employment growth, we cannot confirm these

[11]For the analysis of both periods together, we use the clustered standard errors to control for a within-individual correlation.

Figure 6.10: Employment growth and GDP proportion for period 2000-06 and 2007-13



findings using nonparametric estimation, since the impact does not seem to be statistically significant assuming any functional form of forcing variable.

Table 6.12: Treatment impact on employment growth for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.001	-0.003	0.014
	(0.045)	(0.026)	(0.036)
MSE bandwidth $h-/h+$	0.208/0.185	0.298/0.283	0.253/0.374
RBC t-statistic	-0.063	-0.202	0.429
RBC p-value	0.949	0.840	0.668
RBC CI	(-0.110;0.103)	(-0.066;0.054)	(-0.063;0.099)

Table 6.13: Treatment impact on employment growth for period 2000-06 and 2007-13

	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	0.004	-0.012	0.018
	(0.032)	(0.054)	(0.056)
MSE bandwidth $h-/h+$	0.141/0.207	0.242/0.243	0.237/0.331
RBC t-statistic	0.205	-0.084	0.323
RBC p-value	0.838	0.933	0.747
RBC CI	(-0.070;0.086)	(-0.127;0.116)	(-0.099;0.139)

Chapter 7

Treatment effect and comparison

In this section, we summarize our findings and we compare them with existing literature. Moreover, we perform tests to examine the sensitivity of results to the choice of kernel function and to different bandwidths. Specifically, besides the MSE-optimal bandwidth we also use two times and three times wider bandwidth.

7.1 Sensitivity tests

Although the nonparametric estimation tends to be more suitable for estimation of the treatment effect at the threshold, it relies on the choice of parameters such as bandwidth and kernel function. For this reason, it is convenient to test sensitivity of results on changes of these parameters. As it was stated in previous sections, the kernel function assigns non-negative weights to each observation based on its distance from the cut-off point. Following common practice, we use a triangular kernel function, which is based on the linear function. In this section we also apply an Epanechnikov function, which is more complex since it assigns quadratic weights to observations within the bandwidth. The results of treatment estimation for the last programming period using triangular and Epanechnikov kernel functions are presented in Table 7.1 and Table 7.2.[1] Based on the results we can confirm that the values did not change significantly, and the choice of kernel function does not seem to play

[1]The estimation results using triangular function correspond to the values presented in section Nonparametric approach and results.

a crucial role in the estimation, since the RBC confidence interval remained almost the same.

Table 7.1: Sensitivity test of different kernel functions - triangular

Triangular	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.008	-0.009	0.005
	(0.096)	(0.057)	(0.056)
MSE bandwidth $h-$	0.173	0.239	0.297
MSE bandwidth $h+$	0.189	0.286	0.439
RBC t-statistic	0.088	0.101	0.244
RBC p-value	0.930	0.920	0.807
RBC CI	(-0.234;0.214)	(-0.139;0.126)	(-0.110;0.142)

Table 7.2: Sensitivity test of different kernel functions - Epanechnikov

Epanechnikov	Linear	2^{nd} order	3^{rd} order
Objective 1 effect	-0.017	-0.012	0.001
	(0.088)	(0.051)	(0.050)
MSE bandwidth $h-$	0.156	0.217	0.299
MSE bandwidth $h+$	0.186	0.276	0.437
RBC t-statistic	-0.071	-0.161	0.190
RBC p-value	0.944	0.872	0.849
RBC CI	(-0.209;0.0195)	(-0.127;0.108)	(-0.1;0.122)

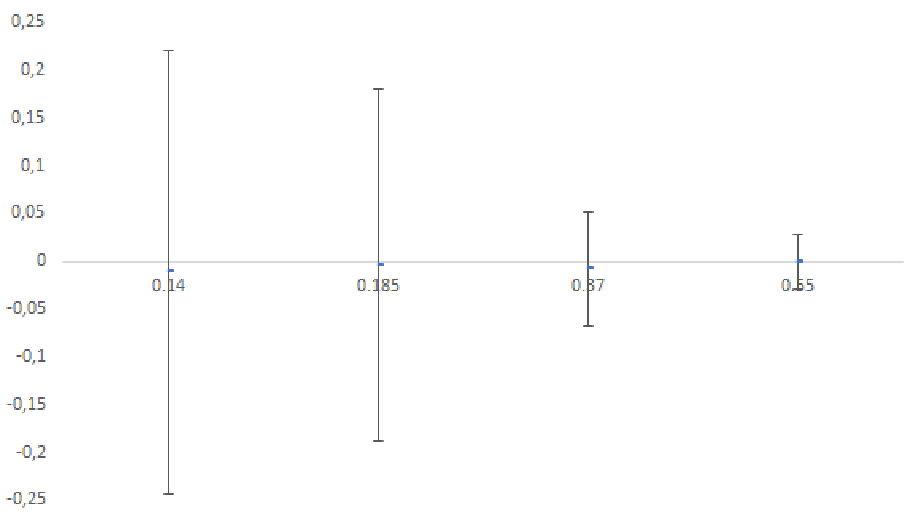
The way of assigning weights to observations in the intervals around the threshold does not seem to be game-changing parameter. On the other hand, the choice of interval may affect the results fundamentally. To analyse the sensitivity of estimated treatment effect to such changes we compare results of nonparametric estimation using various bandwidths. We present results of such estimation in Table 7.3 and we also display coefficients and confidence intervals in Figure A.2. The Table 7.3 is divided into columns based on the selected bandwidths: 0.14 (h_{CER}), 0.185 (h_{MSE}), 0.37 ($2h_{MSE}$) and 0.555 ($3h_{MSE}$). We also present results for conventional and robust-bias corrected method, which is used for proper statistical inference. According to Cattaneo *et al.* (2018) using different than optimal bandwidths affect the results in straightforward way, since with larger bandwidth the variance of local polynomial estimator decreases, which causes a decrease in confidence intervals. On the other hand, it causes an increase in bias of the point estimator, which leads to displacement.

As it can be seen the results using the CER-optimal bandwidth are consistent with results using MSE-optimal bandwidth, since the treatment effect estimates are similar. The development of confidence interval according to changes of bandwidths supports the theory described above. Generally, the results of sensitivity analysis are consistent with our previous findings.

Table 7.3: Sensitivity of bandwidth selection

	0.14(CER)	0.185(MSE)	0.37	0.55
Obj 1 effect	-0.011	-0.003	-0.008	-0.000
	(0.119)	(0.094)	(0.030)	(0.015)
Conv. t-stat	-0.093	-0.037	-0.259	-0.024
Conv. p-value	0.926	0.971	0.795	0.981
Conv. 95% CI	(-0.244;0.222)	(-0.188;0.181)	(-0.067;0.052)	(-0.029;0.028)
Obj 1 effect	-0.014	-0.017	-0.025	-0.009
	(0.135)	(0.118)	(0.049)	(0.022)
Conv. t-stat	-0.105	-0.143	-0.506	-0.395
Conv. p-value	0.916	0.887	0.613	0.693
Conv. 95% CI	(-0.278;0.250)	(-0.248;0.214)	(-0.120;0.071)	(-0.051;0.034)

Figure 7.1: Sensitivity of bandwidth selection



7.2 Treatment effect and comparison

In this part of the thesis we summarize our findings, which were commented on continuously in the section Estimation, and we compare the results with other research studies. Many authors tried to evaluate the impact of the regional policy of the European Union in past years, but the results do not seem to be stable. Although some authors find generally positive impact of EU regional policy, the intensity of the effect does not seem to be clear.[2] The divergence of results was probably caused by complex system of structural funding and organisation of regional policy, which makes it hard to prepare proper analysis, and by difficult separation of causal treatment effects from other factors.

During past years authors used various methods on various programming periods with different level of efficiency. Some researchers count on more traditional approaches like standard panel data methods with model extensions (Mohl & Hagen (2011), Rodríguez-Pose & Garcilazo (2015), Maynou *et al.* (2016)). Other authors prefer methods specifically designed for evaluation of causal effects such as matching on generalized propensity scores (Bondonio (2016), Becker *et al.* (2012), Becker *et al.* (2016)) or regression discontinuity design. Over the past years the RDD method became popular, since it has convenient properties for treatment effect estimation. While some authors used a conventional sharp form (Pellegrini *et al.* (2013), Ferrara *et al.* (2017), Pellegrini & Cerqua (2016)) excluding non-complying regions of EU15 member states, other researchers decided to use all available regions and to implement a fuzzy RDD mechanism (Becker *et al.* (2010), Gagliardi & Percoco (2017), Becker *et al.* (2013), Doppelhofer *et al.* (2008), Becker *et al.* (2016), Bachtrögler (2016), Becker *et al.* (2018)). Out of them only few authors studied the last programming period 2007-2013 (Becker *et al.* (2016), Bachtrögler (2016), Becker *et al.* (2018)) and to the best of our knowledge none of them used a nonparametric estimation method as the main tool for treatment effect analysis, although many theoreticians believe that this method is more suitable for the problem of this nature. On the other hand, the nonparametric estimation method is designed only for the homogeneous treatment effect analysis without using any panel data specifications.

In our analysis we focus on the evaluation of Objective 1 treatment, since it covers the largest part of the budget of regional policy and its eligibility rule is precisely defined. We use the nonparametric estimation on regional data at NUTS2 level (unlike Gagliardi & Percoco (2017), who work with data at

[2]These studies probably catch more public attention, than literature presenting insignificant results. For this reason, we mostly describe research studies presenting positive impact. Moreover, for transparency reasons we describe only studies, that analyse the last programming periods. There seems to be much more discrepancy in older studies.

NUTS3 level), assuming various model settings and we also provide a parametric estimation as a certain robustness check. Based on our analysis, we did not find significant effect of the Objective 1 treatment on the regional GDP per capita growth assuming two last programming periods 2000-2006 and 2007-2013 or the period 2007-2013 separately. Such finding seems to be robust across both parametric and nonparametric methods and various model settings. Like other authors we study the Objective 1 impact on regional performance from different perspective, so we also analyse the effects on employment growth during the last programming periods. According to our results there does not seem to be any significant impact on the employment growth.

As it was indicated previously it is difficult to compare results of different methods and programming periods, since there does not seem to be general consensus. On the other hand, it seems more than reasonable to comment differences between research studies, which implement the same estimation method. Unlike Becker *et al.* (2016) and Becker *et al.* (2018), who found a general positive impact of the Objective 1 treatment on the regional GDP growth using parametric approach with controlling for regional heterogeneity, we analyse the homogeneous treatment effect around the threshold, i.e. local average treatment effect. We use this approach to be able to fully implement the nonparametric estimation. Our findings can be supported by the results presented by Bachtrögler (2016), who found insignificant homogeneous treatment effect using parametric approach (and using nonparametric method as robustness check). She also analysed the heterogeneous treatment effect and she found a weak diminishing effect of regional policy on regional performance. Generally, although many theoreticians prefer the nonparametric estimation of RDD problems, it may lead to loss of efficiency. Moreover, because of unavailability of complete and comparable data for years before 2000, we had to work with the last two programming periods only, which prevented us from analysing the policy impacts for periods up to 2006 only like Becker *et al.* (2010), who found a positive homogeneous treatment effect on GDP per capita growth. Based on their findings the Objective 1 treatment effect on employment growth does not seem to be statistically significant. Also, we had access to Eurostat database in current form, which went through several adjustments and changes of variable definitions, that might cause differences in results. On the other hand, it seemed right to work with "our own" dataset, that was built by working with regional data "one by one" than to use prepared dataset. We also believe that the results of last programming period were affected by the global financial crisis, so the regions may not fully draw up funding, which may prevent the treatment effect from affecting the regional performance.

To sum up, based on our findings the Objective 1 treatment does not seem to significantly affect the performance of regions, which received the funding. The results are robust across parametric and nonparametric estimation methods assuming various functional forms of forcing variable and bandwidths. The impact also does not vary during the programming period. On the other hand, according to results of validity tests the regression discontinuity design mechanism seems to be appropriate for this problem, since there is not any violation of assumptions about density of assigning variable or impacts of baseline covariates. Although we did not find significant impact on regional performance, it does not mean that the importance of common regional policy should be questioned. Even though, it is very difficult to evaluate the efficiency of the whole regional policy, it participates in many local projects, which help people to improve quality of their lives and to build better future society.

Chapter 8

Conclusion

Over the past years, the European Union invested larger amounts with each new programming period to boost growth of disadvantaged regions and to improve the business and living conditions of the society. Currently, the EU plans to allocate about EUR 461.1 billion, which represents almost a half of the EU's budget. Such an extensive investment plan assumes that the treatment has a positive impact on the regional performance and that it helps to reduce economic and social disparities between regions.

This thesis examines how the Objective 1 treatment affects the GDP per capita and employment growth during two last programming periods 2000-2006 and 2007-2013 using mostly nonparametric estimation method of the regression discontinuity design. The aim of this thesis is to contribute to existing literature since the current research studies do not provide conclusive results. Moreover, although many authors tried to evaluate the impacts of the EU's regional policy, almost none of them used the nonparametric approach of the RD design, which is believed to be more suitable for this kind of problem.

First, we applied the parametric estimation method, which uses the whole dataset to fit proper polynomial function of the proportion of GDP per capita to EU average as a forcing variable. As some theoreticians point out it might seem odd to use the whole dataset for the local approximation and to let observations far away from the threshold to affect the local estimation in the same way as observations almost at the threshold. For this reason, we mostly focus on the nonparametric estimation method that is based on finding a bandwidth around the cut-off, in which a proper function can be fitted, and on choosing proper kernel function that assigns weights to observations according to their distance from the threshold.

According to our estimation results we did not find significant impact of the Objective 1 treatment on the GDP per capita growth nor employment growth. These findings are robust to various model specifications and estimation methods. To be able to confirm the validity of the results, we performed several sensitivity and validity tests and we concluded that the regression discontinuity design approach is appropriate for this kind of analysis.

The results are subject to limitation of the data quality, since we had to face to many problems while preparing the datasets such as redefinition of variables of our main interest, changes in definition of borders of regions at NUTS2 level or correction and adjustments of the data in Eurostat's database. These issues forced us to make few simplifying assumptions and to exclude some observations from our dataset. Moreover, although the RD design is considered to be one of the best methods for treatment effect estimation, it does not account for the intensity of the support and whether the regions fully draw up the funding, which might affect the Objective 1 treatment efficiency.

Even though we did not find convincing results about impacts of regional policy at regional level, it does not mean that the treatment does not impact growth locally. Future research might investigate the treatment effects in smaller areas or municipalities to evaluate the efficiency of supported projects. Moreover, the result inconsistency in existing literature makes space for deeper analysis of the roots of this variation.

Bibliography

BACHTRÖGLER, J. (2016): "On the effectiveness of eu structural funds during the great recession: Estimates from a heterogeneous local average treatment effects framework." .

BAUM, C. F. & F. CHRISTOPHER (2006): *An introduction to modern econometrics using Stata*. Stata press.

BECKER, S. O., P. H. EGGER, & M. VON EHRLICH (2016): "Effects of eu regional policy: 1989-2013." *Warwick Economics Research Paper Series* **1118**.

BECKER, S. O., P. H. EGGER, & M. VON EHRLICH (2018): "Effects of eu regional policy: 1989-2013." *Regional Science and Urban Economics* **69**: pp. 143–152.

BECKER, S. O., P. H. EGGER, & M. VON EHRLICH (2010): "Going nuts: The effect of eu structural funds on regional performance." *Journal of Public Economics* **94(9-10)**: pp. 578–590.

BECKER, S. O., P. H. EGGER, & M. VON EHRLICH (2012): "Too much of a good thing? on the growth effects of the eu's regional policy." *European Economic Review* **56(4)**: pp. 648–668.

BECKER, S. O., P. H. EGGER, & M. VON EHRLICH (2013): "Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects." *American Economic Journal: Economic Policy* **5(4)**: pp. 29–77.

BONDONIO, D. (2016): "The impact of varying per capita intensities of eu funds on regional growth: Estimating dose response treatment effects through statistical matching." .

CALONICO, S., M. D. CATTANEO, & M. H. FARRELL (2018a): "Coverage error optimal confidence intervals." *arXiv preprint arXiv:1808.01398* .

Calonico, S., M. D. Cattaneo, & M. H. Farrell (2018b): "Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs." *arXiv preprint arXiv:1809.00236* .

Calonico, S., M. D. Cattaneo, & R. Titiunik (2015): "Optimal data-driven regression discontinuity plots." *Journal of the American Statistical Association* **110(512)**: pp. 1753–1769.

Cattaneo, M. D., N. Idrobo, & R. Titiunik (2018): "A practical introduction to regression discontinuity designs: Volume i." *Cambridge Elements* .

Checherita-Westphal, C. D., C. Nickel, & P. Rother (2009): "The role of fiscal transfers for regional economic convergence in europe." .

Council of the European Union (2006a): "Council decision of 6 october 2006 on community strategic guidelines on cohesion 2006/702/ec." *Technical Report 11*, The European Union.

Council of the European Union (2006b): "Council regulation of 11 july 2006 laying down general provisions on the european regional development fund, the european social fund and the cohesion fund and repealing regulation no 1260/1999 1083/2006/ec." *Technical Report 25*, The European Union.

Di Cataldo, M. (2016): "Gaining and losing eu objective 1 funds: Regional development in britain and the prospect of brexit." *LEQS Paper* **(120)**.

Doppelhofer, G., J. Crespo-Cuaresma, & M. Feldkircher (2008): "The determinants of economic growth in european regions." .

European Commission (1999): "The structural funds and their coordination with the cohesion fund: Guidelines for programmes in the period 2000-06." *Publications Office* `https://ec.europa.eu/regional_policy/en/information/legislation/guidance/2000-2006`.

European Commission (2005a): "Guidelines for programmes in 2000-2006." *Publications Office: Summaries of EU legislation* `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:g24202&from=EN`.

European Commission (2005b): "Objective 1." *Publications Office: Summaries of EU legislation* `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:g24203&from=EN`.

European Commission (2005c): "Objective 2." *Publications Office: Summaries of EU legislation* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:g24206&from=EN.

European Commission (2005d): "Objective 3." *Publications Office: Summaries of EU legislation* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:g24207&from=EN.

European Commission (2008): "Working for regions eu regional policy 2007-2013." *Luxembourg: Publications Office of the European Union* .

European Commission (2014a): "European social fund 2007-2013: investing in people." *Publications Office* **MEMO 08(302)**. http://europa.eu/rapid/press-release_MEMO-08-302_en.htm?locale=en.

European Commission (2014b): "The european union explained: Regional policy." *Luxembourg: Publications Office of the European Union* .

European Commission (2018): "Summary report of the programme annual implementation reports covering implementation in 2014-2017." *Publications Office* **COM 2018**.

European Commission (2019): "Shaping the digital single market." https://ec.europa.eu/digital-single-market/en/policies/shaping-digital-single-market.

European Commission (n.d.): "Projects." http://ec.europa.eu/esf/main.jsp?catId=46&langId=en&list=0.

European Statistical Office (2017): "Smarter, greener, more inclusive? indicators to support the europe 2020 strategy." *Luxembourg: Publications Office of the European Union* .

European Union (2013): "Social investment: Commission urges member states to focus on growth and social cohesion." https://ec.europa.eu/social/main.jsp?langId=en&catId=1044&newsId=1807&furtherNews=yes.

European Union (2018): "Regions in the european union nomenclature of territorial units for statistics - nuts 2016/eu-28." *Luxembourg: Publications Office of the European Union* .

European Union (n.d.): "Themes." `https://ec.europa.eu/regional_policy/en/policy/themes`.

Ferrara, A. R., P. McCann, G. Pellegrini, D. Stelder, & F. Terribile (2017): "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)**: pp. 817–841.

Gagliardi, L. & M. Percoco (2017): "The impact of european cohesion policy in urban and rural regions." *Regional Studies* **51(6)**: pp. 857–868.

Giua, 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)**: pp. 109–131.

Hahn, J., P. Todd, & W. Van der Klaauw (2001): "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* **69(1)**: pp. 201–209.

Jacob, R., P. Zhu, M.-A. Somers, & H. Bloom (2012): "A practical guide to regression discontinuity." *MDRC* .

Kyriacou, A. P. & O. Roca-Sagalés (2012): "The impact of eu structural funds on regional disparities within member states." *Environment and Planning C: Government and Policy* **30(2)**: pp. 267–281.

Lee, D. S. (2008): "Randomized experiments from non-random selection in us house elections." *Journal of Econometrics* **142(2)**: pp. 675–697.

Lee, D. S. & T. Lemieux (2010): "Regression discontinuity designs in economics." *Journal of economic literature* **48(2)**: pp. 281–355.

Lemieux, T. & G. Imbens (2008): "The regression discontinuity designâ€"theory and applications." .

Maynou, L., M. Saez, A. Kyriacou, & J. Bacaria (2016): "The impact of structural and cohesion funds on eurozone convergence, 1990–2010." *Regional Studies* **50(7)**: pp. 1127–1139.

Mohl, P. & T. Hagen (2011): "Do eu structural funds promote regional employment? evidence from dynamic panel data." .

Palevičienė, A. & D. Dumčiuvienė (2015): "Socio-economic diversity of european regions: Finding the impact for regional performance." *Procedia economics and finance* **23**: pp. 1096–1101.

Pellegrini, G. & A. Cerqua (2016): "Measuring the impact of structural and cohesion funds using the regression discontinuity design in eu-27, in the period 1994-2011." *Luxembourg: Publications Office of the European Union* .

Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso, & F. Busillo (2013): "Measuring the effects of european regional policy on economic growth: A regression discontinuity approach." *Papers in Regional Science* **92(1)**: pp. 217–233.

Rodríguez-Pose, A. & E. Garcilazo (2015): "Quality of government and the returns of investment: Examining the impact of cohesion expenditure in european regions." *Regional Studies* **49(8)**: pp. 1274–1290.

Tomova, M., A. Rezessy, A. Lenkowski, E. Maincent *et al.* (2013): "Eu governance and eu funds-testing the effectiveness of eu funds in a sound macroeconomic framework." *Technical report*, Directorate General Economic and Financial Affairs (DG ECFIN), European â€¦.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data.* MIT press.

Appendix A

Appendix A

Table A.1: The thematic objectives and their funding over period 2014-2020

	Objective	ERDF	ESF	CF
1	Strengthening research, technological development and innovation	**X**	X	
2	Enhancing access to, and use and quality of, information and communication technologies	**X**	X	
3	Enhancing the competitiveness of SMEs	**X**	X	
4	Supporting the shift towards a low-carbon economy	**X**	X	X
5	Promoting climate change adaptation, risk prevention and management	X		X
6	Preserving and protecting the environment and promoting resource efficiency	X		X
7	Promoting sustainable transport and improving network infrastructures	X		X
8	Promoting sustainable and quality employment and supporting labour mobility	X	**X**	
9	Promoting social inclusion, combating poverty and any discrimination	X	**X**	
10	Investing in education, training and lifelong learning	X	**X**	
11	Improving the efficiency of public administration	X	**X**	X

Table A.2: Internal validity test of covariates for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Econ. active popul.	0.1431	0.0785	-0.93
	(0.429)	(22.15)	(2.682)
RBC t-statistic	0.124	0.396	-0.08
RBC p-value	0.902	0.692	0.936
RBC CI	(-0.887;1.006)	(-0.4;0.602)	(-5.746;5.3)
Agr. share	-0.306	-0.117	-0.127
	(0.580)	(0.267)	(0.924)
RBC t-statistic	0.209	-0.318	-0.504
RBC p-value	0.834	0.75	0.614
RBC CI	(-1.221;1.513)	(-0.685;0.494)	(-24.296;14.361)
Industry share	0.283	0.108	0.088
	(0.537)	(0.277)	(0.818)
RBC t-statistic	0.074	0.230	0.07
RBC p-value	0.941	0.818	0.944
RBC CI	(-1.172;1.264)	(-0.551;0.697)	(-1.649;1.177)
Service share	0.014	-0.003	-0.661
	(0.427)	(0.25)	(1.32)
RBC t-statistic	-0.148	0.083	-0.375
RBC p-value	0.883	0.934	0.708
RBC CI	(-1.024;0.880)	(-0.533;0.58)	(-3.064;2.08)
Patents appl.	281.01	217.26	125.4
	(446.49)	(223.83)	(720.62)
RBC t-statistic	0.362	0.881	0.48
RBC p-value	0.717	0.378	0.631
RBC CI	(-827.733;1202.63)	(-270.37;711.84)	(-1142.72;1883.56)

Table A.3: Internal validity test of covariates for both periods 2000-2006 and 2007-2013

	Linear	2^{nd} order	3^{rd} order
Econ. active popul.	-0.541	-0.084	-0.07
	(4.555)	(0.3239)	(0.2437)
RBC t-statistic	0.322	-0.3124	-0.289
RBC p-value	0.747	0.755	0.772
RBC CI	(-8.134;11.337)	(-0.868;0.63)	(-0.608;0.452)
Agr. share	0.915	-0.173	-0.107
	(4.154)	(0.382)	(0.217)
RBC t-statistic	0.925	-0.07	-0.471
RBC p-value	0.355	0.944	0.638
RBC CI	(-6.044;16.856)	(-0.863;0.803)	(-0.568;0.348)
Industry share	-0.233	-0.151	-0.158
	(1.875)	(0.391)	(0.280)
RBC t-statistic	0.145	-0.506	-0.612
RBC p-value	0.885	0.613	0.541
RBC CI	(-3.783;4.386)	(-1.096;0.646)	(-0.764;0.401)
Service share	0.371	0.386	0.334
	(1.042)	(0.746)	(0.472)
RBC t-statistic	0.05	0.612	0.732
RBC p-value	0.96	0.54	0.464
RBC CI	(-2.182;2.296)	(-1.097;2.092)	(-0.615;1.348)
Patents appl.	233.13	89.66	71.785
	(894.67)	(173.27)	(157.44)
RBC t-statistic	-0.071	0.25	0.393
RBC p-value	0.944	0.803	0.694
RBC CI	(-1964.7;1827.85)	(-330.014;426.39)	(-265.8;399.512)
Pop. density	-614.28	1287.1	-232.04
	(2029.1)	(6477.5)	(1152.7)
RBC t-statistic	-0.64	0.087	-0.528
RBC p-value	0.522	0.931	0.598
RBC CI	(-5900.13;2995.38)	(-12735.7;13916.5)	(-3338.75;1922.43)

Table A.4: Development of treatment effect using cumulative GDP growth for period 2007-2013

	Linear	2^{nd} order	3^{rd} order
Objective 1 - $1^{st}year$	0.048 (0.117)	0.008 (0.071)	-0.037 (0.142)
Objective 1 - $2^{nd}year$	0.009 (0.114)	0.002 (0.071)	0.012 (0.01)
Objective 1 - $3^{rd}year$	-0.06 (0.258)	-0.034 (0.150)	0.048 (0.171)
Objective 1 - $4^{th}year$	-0.145 (0.476)	-0.095 (0.265)	0.087 (0.289)
Objective 1 - $5^{th}year$	-0.186 (0.603)	-0.117 (0.325)	-0.116 (0.366)

Figure A.1: The probability of receiving the treatment for both periods (the 3^{rd} order polynomial)

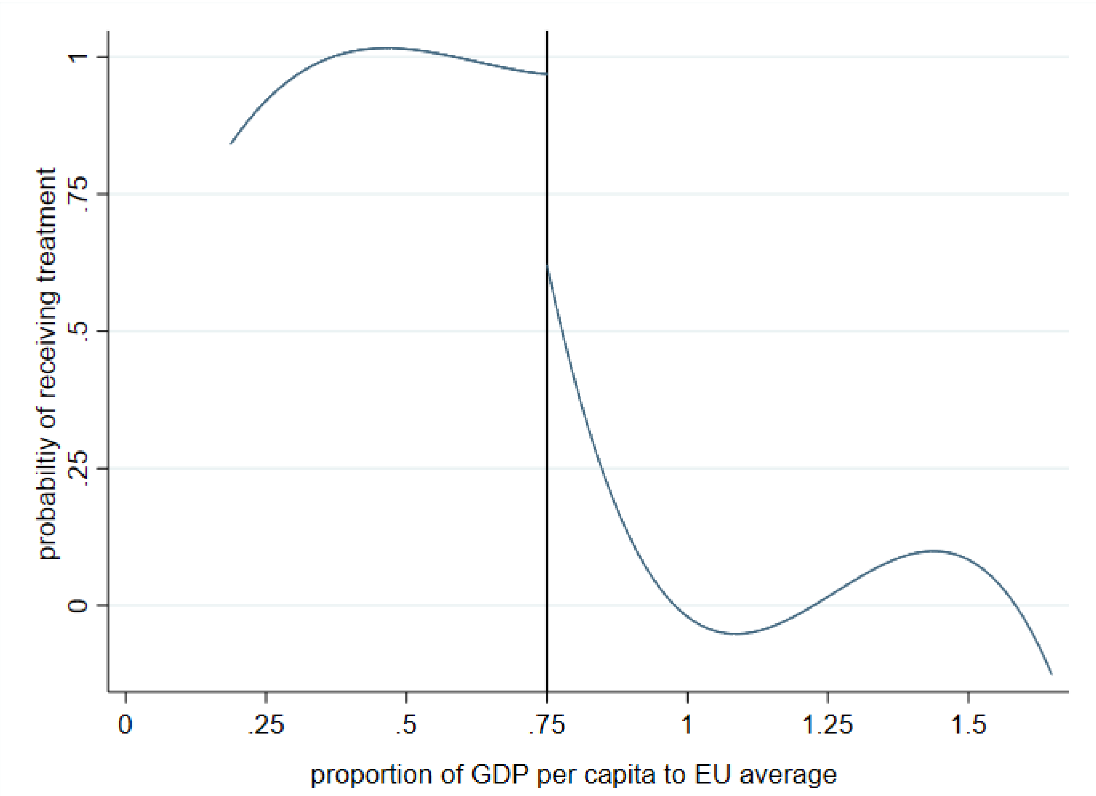


Figure A.2: The probability of receiving the treatment for period 2007-2013 (the 2^{nd} order polynomial)