Sapienza
Università di Roma

DOTTORATO DI RICERCA IN
STATISTICA ECONOMICA

CICLO XXXI

Settore Concorsuale:	13/D2
Settore Scientifico Disciplinare:	SECS-S/03

Beyond Treatment Effect Estimation: new developments in policy evaluation

Presentata da: Alessandra Pasquini

Coordinatore Dottorato
Prof. Brunero Liseo

Supervisore
Prof Guido Pellegrini

Esame finale anno 2019

Sapienza
University of Rome

PHD DEGREE IN
ECONOMIC STATISTICS

CYCLE XXXI

Competition Field : 13/D2
Academic Discipline: SECS-S/02

Beyond Treatment Effect Estimation: new developments in policy evaluation

Presented by: Alessandra Pasquini

Ph.D. Director
Prof. Brunero Liseo

Supervisor
Prof. Guido Pellegrini

Final exam year 2019

Abstract

To what extent are hiring incentives targeting a specific group of vulnerable unemployed people more effective with respect to generalized incentives without a definite target? Do targeted policies have negative side effects on the labor market that are too important to accept them? Which are the channels guaranteeing a long-lasting effect of European Structural Funds? Even though there is vast literature on hiring subsidies and European Structural Funds, these questions remained unanswered. To answer them, it is necessary to go beyond the simple treatment effect estimation. We did it following two different paths.

To answer the first two questions we compared the impact of two similar hiring policies, one oriented towards a target group of long-term unemployed people and one generalized, implemented on the Italian labor market. We considered administrative data on job contracts and workers, and applied counterfactual analysis methods. The results show that only the targeted policy has a positive and significant impact, while the effects of the generalized policy on the vulnerable group are negligible. Moreover, we did not detect any indirect negative side effect of the targeted policy on the local labor market.

To answer the third question we introduced a new methodology, namely the Mediation Analysis Synthetic Control (MASC) and applied it to investigate on the impact of European Structural Funds reduction in Abruzzi region. MASC is a generalization of the Synthetic Control Method (SCM) that allows decomposing the total effect of the intervention into its indirect effects and its direct effect when data on only one treated and few control units are available. The results show that the negative impact of European Structural Funds reduction on economic growth is mainly driven by the indirect effect passing through employment reduction. Instead, only a small portion of the negative impact is due to the indirect effect passing through investments reduction.

Acknowledgements

Many people contributed to the development and writing of this dissertation. Many people contributed to my survival through it and through my Phd. I want to thank all my collegues of the XXXI Edition who helped me going through the first year and survive to the never-ending amount of exams we had to do. I want to thank all the guys from room 143 who welcomed me in their room and accepted me without prejudice, almost as if I was a mathematician...rather than leaving me lonely in the statisticians' room. They explained me how to survive to Phd bureaucracy, and how to use the forever-broken printer in the bathroom. They provided me 24h/24h assistance on latex and Simone helped me understanding SCM demonstration. But, most of all, they provided me enjoyable lunches and aperitives and coffee breaks. I want to thank Marco, who introduced me to policy evaluation for the first time and included me in the agreement between INAPP and DSSE from which it came out my first work and the first two chapters of this thesis. I also thank him for keeping me updated on all the institutional "gossips", I love our talks! I want to thank Marusca De Castris and Daniele who gave me my first opportunity to present at a conference and let me get in contact with Giovanni and organise my Visiting abroad. I want to thank Rune, Lina, and all the folks from OBK who immediately welcomed me as part of the family and, together with Beatrice and Jonathan, made my stay in Odense pleasent and enjoyable notwithstanding I was there in the worst season ever. I want to thank all the people from SDU and from VisitINPS program who showed me how academic world works outside Italy and made me realise all my mistakes with respect to it. And in particular Vincenzo and Mircea, who made me notice the biggest mistakes I did with my first work. I want to thank the organisers of IWcee 2017 and LISER. At the first I had the idea for MASC and the second organised a course where I learnt about local randomized experiments in RDD contexts. I want to thank Guglielmo Barone, Francesco David and Guido de Blasio for their paper and their availability to give me the data and the right advices. And particularly Guido, for making me realise which have to be my next steps. I want to thank

Augusto for his advices on my work and his encouragements when I was a bit lost. I want to thank my supervisor for leading me through these three years. I always got out from his office thinking more clearly than when I got in. He always used his experience to give me valuable advices on the choices to make. I want to thank Giovanni, who has been a second supervisor to me, he gave me valuable advices, he lost his time helping me on issues. He showed he really cared. I want to thank all my friends. The last year was very intense. Everyone in his own way, who more who less, you supppoerted me in the harder times, you enjoyed with me my happiest moments, you turned a blind eyes when I was not the best company ever: Livia, Valentina, Alessia, Tullia, Elena, Giulia, the Fatigati, Silvia, Carlos, Ila, Luchino, Andrea, Zari, Giuliano, Judith and Giulietto, Valerio, Valeria and Leo, the stinking Lerci and, obviously, my bestwomen, who basically organised our wedding! I want to thank Francesco who is always there with the right thing to say in the hardest moments. I am happy to have you by my side. Finally, I want to thank my parents. Sometimes I wonder whether all the positive characteristics I have are just them and whether I contributed with something or not. I mean, when we were little, they used to make us change the words of the songs while travelling in the car to stimolate our creativity...It's easy to see how life was a very easy-win situation for me after.

Contents

1. Introduction

The Skinner box is a cage where there are two levers. Pushing the first ones, its floor is electrified. Pushing the second ones, its occupant receives food. In 1948 american psychologist Burrhus Frederic Skinner, inventor of the cage, discovered that a rat placed into it would start soon pushing only the food lever. Somehow, and unconsciously, that rat would learn the cause-effect relation between pushing the lever and getting food. Nonetheless, not all cause-effect relations are so easy to establish that even a rat can understand them. Getting outside of the laboratory, the number of possible causes increase and there can be misleading correlations.

An enlightening example is the history of ergotism's name. This illness was also called "Saint Anthony's fire". This name was originated from the fact that most pilgrims going to Saint Anthony sanctuaries, in Italy, from Northern Europe, would heal. Even though it was interpreted in the past as a holy cause-effect relation, the healing hid a correlation between the location of the sanctuaries and a nutrition free of rye (plant infected by ergotim's fungus).

The philosopher David Hume, in his work "Enquiry Concerning Human Understanding" from 1748, underlined the difficulty of finding a cause-effect relation. He stated: "When we look around us at external objects, and think about the operation of causes, we are never able to discover any power or necessary connection, any quality that ties the effect to the cause and makes it an infallible consequence of it. All we find is that one event does in fact follow the other." To solve this issue, the author proposed the following definitions of cause:
"We may define a 'cause' to be

- an event followed by another, where all events similar to the first are followed by events similar to the second.

Or in other words

- where if the first event hadn't occurred the second wouldn't have occurred either."

Centuries later, two main approaches were developed in statistics to find quantitative proofs of cause-effect relations. Both of them have their roots in these definitions. The first, older, statistical approach follows the first definition of "cause" given by the philosopher. Indeed, in the first approach statisticians try to find the relation between cause and effect looking to the correlation between what is supposed to be the cause and what is supposed to be the effect. Looking for a correlation means checking wheter the event cause is always linked with the event effect. To overcome the risk of interpreting simple correlations as a cause-effect relation, in this approach, researchers try to take into account of all variables that may induce the same effect and being correlated with the cause (in the example of the Saint Anthony's fire a good researcher would have considered nutrition when examining the relation between the illness and the pilgrimage).

The second approach follows the second definition, where causality is described in terms of a counterfactual relation. The idea is that there is a causal effect if the absence of the cause would have implied the absence of the effect. The concept of counterfactual relation is the root of modern causal inference. In this approach, rather than trying to study the correlation between two variables, the statisticians try to determine what would have happened in absence of cause.

Around 200 years after Hume, the concept of counterfactual relation started to be used in scientific research by the polish mathematician Jerzy Neyman. In his work, the mathematician proposed a method for the study of agricoltural field experiments[1] (Neyman 1990). Goal of the study was to compare different crop varieties. In a period when the idea of counterfactual relation was spreading among researchers (see Rubin 1990), Neyman gave the basis for a formal approach to the problem, introducing the concept of potential outcomes. He noticed as, in a field experiment, there are multiple potential yields depending on the type of crop and the plot where it is set. Nonetheless, only some of these yields are observed. Task of the researcher

[1]The last were implemented ever since the twenties.

is to inference on all potential yields using only the observed.

The idea was generalized in Rubin (1974) with the introduction of Rubin's (or Neyman-Rubin) Model. Using modern definitions and notations the idea of the author can be described as follow. There is a group of units of interest, indexed by $i = 1, .., N$. Some of these units are exposed to a treatment (the equivalent of what I have called cause so far) and some of them are not. A dummy variable D_i takes value one if the i unit is treated and zero otherwise. The problem of interest is the impact of the treatment on a specific outcome Y. For each unit i it is possible to define two values of the potential outcome. The first one is the outcome unit i would have when exposed to treatment: Y_i^1. The second is the outcome unit i would have in absence of treatment: Y_i^0. Following the definition of Hume, it is easy to see that the impact of the treatment (or effect of the cause) for unit i is given by:

$$\tau_i = Y_i^1 - Y_i^0 \tag{1.1}$$

Here it comes what P. W. Holland called the "Fundamental Problem of Causal Inference" (Holland 1986): it is impossible to observe both Y_i^1 and Y_i^0. Indeed, a unit is either treated or untreated. More formally, the data will obey to the following observational rule:

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 \tag{1.2}$$

where Y_i is the observed outcome of unit i. Therefore, individual treatment effect is never observed. Nonetheless, in most cases, the purpose of the study is to determine the average effect of the treatment. To reach this goal it would be straightforward to compare the group of treated units with the group of untreated units:

$$\begin{aligned} E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \\ E((Y_i^1 - Y_i^0)D_i + Y_i^0|D_i = 1) - E((Y_i^1 - Y_i^0)D_i + Y_i^0|D_i = 0) = \\ E(Y_i^1 - Y_i^0|D_i = 1) + [E(Y_i^0|D_i = 1) - E(Y_i^0|D_i = 0)] \end{aligned}$$

Nevertheless, as it is clear from the equation, the comparison between the treated units and the untreated units is equal to the sum between the average treatment effect on the group of treated (i.e. on the group of units with $D_i = 1$) and the difference between the potential outcomes of the two groups in absence of treatment (i.e. the term in the square brackets). The second value is commonly called "selection bias" referring to the fact that the two groups' potential outcomes in absence of treatment differ because of a "selection" into treatment of units with determined characteristics. As an example, if we think of a policy targeting long-term unemployed it is easy to see that the group of treated, even in absence of treatment, will be very different from the group of untreated.

The first, intuitive, solution used in that years was to select a group of untreated as similar as possible to the treated group. This method was used even before the introduction of potential outcome concept, and different methods to select the untreated and match between the two groups were proposed (some examples are Cochran 1953 and Cochran and Rubin 1973). Nevertheless, it did not rely on a solid theoretical justification. Therefore, the choice of the characteristics to match the units on, depended mainly on the available covariates. A first step towards a theoretical justification was done by Rubin (1977) who imagined a framework where treatment assignment probability depended on a variable X. He stated that, in that case, the effect of the treatment was given by:

$$\tau = \frac{1}{N}\sum_{x \in X}[E(Y_i|D_i = 1, X = x) - E(Y_i|D_i = 0, X = x)]$$

The basic idea behind this statement is that, once taken into account of the characteristics determining selection into treatment, there is no reason for the other determinants of the outcome to differ between treated and control groups. This intuition was formalized by Rosenbaum and Rubin (1983). They introduced the strong ignorability assumption (also called the unconfoundness or the conditional independence assumption). They demonstrated that, under this assumption, the difference between the conditioned average outcome of the treated and those of the control groups is an unbiased estimation of the average treatment effect (on treated). The conditional independent assumption is the following:

Assumption 1 (Conditional Independence Assumption).

$$(Y^1, Y^0) \perp\!\!\!\perp D|X$$

It requires the potential outcomes to be independent from treatment assignment once conditioning on the proper covariates X. As an example, lets consider a policy targenting long-term unemployed and selecting, for a training course, the group of long-term unemployed living on the west side of a small city. The outcome is the level of wage the individuals have after 10 years. The covariate X satisfying assumption 1 is the time in unemployment. Indeed, once taken into account of the time in unemployment, the additional, non-random, selection "being on the west side of a small city" (small enough that time and energy necessary to go from one side of the city to the other is not determinant in job choice) is independent from future wage. Rosenbaum and Rubin (1983) demonstrated that, if the assumption is satisfied, then:

$$\tau = \frac{1}{N}\sum_{x \in X}[E(Y_i|D_i = 1, X = x) - E(Y_i|D_i = 0, X = x)]$$

is an unbiased estimator of the average treatment effect on treated.

When all the covariates in X necessary to satisfy CIA are observed, we talk of selection on observables (Lechner 2015). Nevertheless, in many policy evaluation frameworks the set of observables is not rich enough for this requirement to be satisfied. In those cases, we talk of selection on unobservables (Lechner 2015) and the approaches based on CIA can't be applied. Many other methods have been proposed to deal with selection on unobservables.

Among others, the Regression Discontinuity Design. This methodology relies on the existence of an observed variable (called "forcing variable") whose value determines selection into treatment with respect to a given threshold. In other words, a variable such that, all units having a value of it higher than a determined threshold are treated. All units having a value of it lower than the threshold are not treated. In such a framework, CIA can be violated if there are unobserved variables influencing both the forcing variable (and, consequently, treatment assignment) and the outcome. The idea is then to take an interval of the forcing variable around the threshold small enough to reasonably assume that treatment assignment is locally randomized (for additional details see 2.3.3.2). This design, can be employed only in frameworks where treatment is defined with respect to a threshold on a forcing variable, and there is no manipulation of the latter.

Regression Discontinuity Design belongs to a set of methods which exploit the availability of cross-sectional data, where many different units are observed in a single time period. Other methods relies on the use of panel data, where the same units are observed in multiple time periods. Some of these methods can be used even when the number of units is particularly low. Therefore, they are useful also when there is selection on observables but the number of units is too low for the estimation to be satisfying. Among them, there is Syntethic Control Method. This methodology has been used for the first time in Abadie and Gardeazabal (2003) and more formally introduced in Abadie, Diamond, et al. (2010) and Abadie, Diamond, et al. (2015). The idea behind SCM method is to use information on pre-intervention period to build a "synthetic control", i.e. a linear combination of control units which mimic what would have happened to the treated unit in post-intervention period in the absence of the intervention. This is done by re-weighting the post-treatment outcomes of control units (whose set is often called donor pool in this framework) by using weights that are chosen to minimize the distance between pre-intervention observable characteristics (including pre-intervention outcomes) of treated and control units (for additional details see 4.1).

All the afore-mentioned methods allow to estimate the average treatment effect, the individual treatment effect or the local average treatment effect. Nonetheless, in many empirical frameworks, it may be policy relevant to have additional information on how the programme works, the mechanism behind it or other of its consequences. As an example, we can imagine to detect a positive effect of a training on long-term unemployed hiring. It could be useful to understand whether (or in which portion) the positive effect is due to an improvement in soft-skills, in technical skills, or in self-confidence. Similarly, it could be useful to understand if the policy had some negative side effects, as a negative impact on short-term unemployed hiring. In this dissertation we show how, starting from two methods that can be used under selection on unobservables (namely, the regression discontinuity design and the synthetic control method), it is possible to go beyond the simple treatment effect estimation following two different paths. The first path is empirical. Depending on the empirical application and the characteristics of the policies under study, it may be possible to extrapolate additional policy relevant information. The second path is methodological. Innovative methods may be developed to estimate additional parameters.

In Part I of the dissertation, the application following the first path is presented. The work starts from Law 407/90, a policy implemented in Italy between 1990 and 2014 and incentivizing firms to create (permanent) employment for long-term unemployed people. In the first chapter, we introduce policy setting and characteristics, some of the existing literature on hiring incentives, the data and empirical strategy we have used to evaluate it, and the results we obtained.

To evaluate the policy, we exploited an administrative micro-database, namely CICO database, provided by the Ministry of Labour and Social Policies. We had access to these data thanks to an agreement between the Dipartimento di Scienze Sociali ed Economiche of Sapienza University of Rome and INAPP (Istituto Nazionale per l'Analisi delle Politiche Pubbliche). The database contains reliable information on contracts stipulated from 2008 to 2016 for a sample of individuals. Following Schünemann et al. (2013), we applied a regression discontinuity design to estimate the policy impact. We used the days of unemployment as a forcing variable and added daily fixed effects to the standard model. In order to select the bandwidth, we applied the method proposed by Cattaneo et al. (2015), modified in order to take into account of daily fixed effects. Bandwidth choice was conditioned by some considerations on the consequences of using a time-varying forcing variable. Even though other authors had to deal with this issue before (as an example, see Schünemann et al. (2013) and Anastasia, Giraldo, et al. (2012)), to the best of our knowledge, no one explicitly considered its implications. Giving general considerations, our hope is to provide guidelines to authors who will have to deal with this issue in the future. From the analysis, it emerged Law 407/90 had a strong, positive and significant effect on LTU hiring. The estimated impact was meaningfully higher than in previous studies using eligibility as treatment. This may be due to the high take-up rate and to country characteristics.

In the second chapter, we went beyond the simple treatment effect estimation. First of all, we investigated on its negative side effects. Two side effects are possible in this framework. In presence of the subsidy, firms may hire subsidized unemployed people instead of unsubsidized ones with similar characteristics (and, consequently, a similar vulnerability level), penalizing the latter. This side effect is called displacement effect. When eligibility is defined by determined conditions (as in targeted policies), the presence of asymmetric information on the government's side may induce agents to cheat in order to appear eligible when they are not. This side effect is called asymmetric information effect. Among the studies estimating the impact of hiring subsidies, few of them took into account of the possible negative side effects of the policies (Schünemann et al. 2013, Blundell et al. 2004, Calmfors et al. 2002, Bucher 2010, Boockmann et al. 2007). To check for the presence of these side effects, we exploited the characteristics of the policy. We compared the differences in hiring before and after policy ending, for values of the forcing variable far from and close to the threshold. We did not detect any displacement or asymmetric information effects.

In spite of the huge amount of international literature on hiring and wage subsidies, targeting a vulnerable category of unemployed people (i.e. long-term-unemployed), little is known about the difference between the latter and generalized subsidies without a definite target. Additional information on the difference between the two type of policies would, nonetheless, be extremely useful. Indeed, it is important to know which of the two policies would be more effective and whether the vulnerable group would be penalized by a switch from the targeted to the untargeted one. With the second part of the second chapter we wish to overcome this lack of information. To do it, we exploited the similarities and the differences between Law 407/90 and a second policy, implemented immediatly after its ending. In this second part of the chapter we start comparing the two policies qualitatively, we introduce some of the literature on this topic, the empirical strategy we used to answer our question and the results we have obtained. The second policy is 2015 *Legge Stabilità* (or Law 190). It consisted, among others, of incentives to create permanent employment without a particular targeted group. It lasted one year and implied the permanent end of Law 407/90. We started checking whether the difference in policy targets was the only relevant ones. Once verified that, we used an interrupted time series analysis, to estimate the impact of the generalized incentives provided by Law 190. In order to parcel out the impact of the generalized incentives from the impact of Law 407/90 ending, we did the estimation on a control group. From the analysis, it emerged Law 190 had no significant effect on hiring of long-term unemployed.

In Part II an application of the second path is presented. This part is composed by two chapter, in the first chapter (the third of this dissertation) we introduce a new methodology allowing to do mediation analysis in frameworks where there is selection unobservables and/or the number of treated units is low. In the second chapter (the fourth of this dissertation), an application for this method is provided.

In the third chapter, the first part is dedicated to the literature and a more detailed introduction to the synthetic control method, while the second part is dedicated to our new methodology, namely the Mediation Analysis Synthetic Control (MASC).

Among the methods dealing with selection on unobservables, Synthetic Control Method (SCM) is particularly popular. Even though this method is very well suited to estimate the total effect of the interventions, in many policy evaluation frameworks it may be highly relevant to have additional information on the mechanism behind this effect. In particular, it may be interesting to investigate on the presence of some intermediate variables (called mediators) that lie on the causal pathway between the treatment and the outcome of interest. Mediation Analysis is a standard approach that allows to investigate on the presence of the mediators and to decompose the total effect into a direct (or net) effect of the treatment on the outcome of interest and some indirect effects, generated through the mediators. A large part of Mediation Analysis literature focus on the identification and estimation of direct and indirect effects under sequential conditional independence (see Pearl 2001, Imai, Keele, and Yamamoto 2010, Imai and Yamamoto 2013, Vansteelandt and VanderWeele 2012, Huber 2014, Huber 2016, Huber et al. 2017), a sort of CIA, extended to take into account of the mediator. Few of them can be used when the sequential conditional independence assumption is violated (see Zheng and Zhou 2017, Frölich and Huber 2017, Deuchert et al. 2018). Nonetheless, the applicability of these methods is limited in empirical analysis. This motivates the introduction of Mediation Analysis Synthetic Control (MASC), a generalization of SCM that allows decomposing the total effect of an intervention into its indirect component, which goes through observed mediators, and its direct component, in frameworks where both the sequential conditional independence assumption and the CIA are violated and/or the number of treated units is low. The main intuition behind this method is that the same idea used in SCM can be applied to identify these parameters. To do it, it is sufficient to add, to the pre-treatment outcome information usually considered during weights choice, pre- and post- treatment mediator information.

In the fourth chapter, we start from Barone et al. (2016) work, where the SCM was applied, and apply the MASC to the same framework. The authors studied the impact of European Structural Funds (SF) reduction on the GDP of Abruzzi region. This analysis was highly relevant to take conclusion on the longevity of European SF effect. Given that this policy's transfers can be used for many different programmes, their impact can travel through different causal channels. In the framework of policy longevity investigation, it is arguably policy relevant to understand what are the channels guaranteeing more longevity and what are those by which transfers have a temporary effect. Consequently, we extended Barone et al. (2016) analysis using investments and employment as mediators to disentagle the portion of the effect due to these particular intermediate variables from the rest. In the first part of the chapter, European Structural Funds and the literature on the topic are introduced. Later on, we describe how MASC was implemented and the results we have obtained. From the analysis it emerged that the end of European SF had a negative and significant impact on indexed GDP per capita (in line with Barone et al. (2016) results). None of this effect was mediated by a reduction in investments while a big portion of it was mediated by a reduction in employment share.

The study presented in the first part of the dissertation was conduced in collaboration with Prof. Marco Centra from INAPP (National Institute for Public Policies Evaluations) and Prof. Guido Pellegrini. The study presented in the second part was conduced in collaboration with Prof. Giovanni Mellace from Southern Denmark University. The whole work was supervisioned by Prof. Guido Pellegrini.

Part I

2. Law 407/90 Evaluation

The aim of this Chapter is twofold. To explore the literature on active labour market policies and to present our evaluation of Law 407/90. We will introduce Law 407/90, the existing literature on hiring incentives, the empirical strategy we've used to evaluate it and the results we have obtained.

2.1. *Long-Term Unemployment and Law 407/90*

The study of active policies whose goal is to reduce long-term unemployment and increase the incidence of permanent contracts, is crucial in today's Europe. Indeed, one of the consequences of 2008 global crises was a steep increase in the rate of LTU[1]. The following austerity policies, worstened the situation from a LTU perspective. Their implementation contributed to slower the recover of labour market (Duell et al. 2016, Junankar 2011). According to Eurostat data, the rate of LTU (calculated with respect to the total labour force) reached 6% in the Euro Area and 7.7% in Italy in 2014[2]. In Italy, according to ISTAT (Italian National Institute of Statistics) data, the incidence of LTU on total unemployment was 57.3% in 2016[3]. The increase of LTU rate was particularly sharp among youths.

A reduction in LTU rate is necessary to guarantee the social inclusion of all individuals. To reach it, the implementation of a proper policy is fundamental. Indeed, it is hard for the LTU rate to lower naturally. LTU condition is characterized by a strong duration dependence. Meaning that, the longer an individual is in the unemployment condition, the harder will be for him/her to exit from it (Mussida 2010, Obermeier and Meier 2016, Heckman and Borjas 1980, Farooq and Kugler 2015, Duell et al. 2016). There are three main causes of the duration dependence (Farooq and Kugler 2015, Brown and Koettl 2012, Brown 2015):

[1]In Italy it happened especially in the southern regions (Anastasia, Giraldo, et al. 2012).
[2]Source: http://ec.europa.eu/eurostat/data/database
[3]Source: ISTAT, Noi Italia 2017

1. The loss of human capital and lack of recent working experience due to the absence from employment. Its consequence is a lower desiderability of the long term unemployed worker.

2. The reduction of connections between the unemployed individual and the labour market. This reason is particularly effective during recession periods since using informal channels to recruit allows employers to reduce recruitment costs.

3. Scarring effect affecting LTU. In real labour market there is imperfect information. Employers tend to exploit the limited tools they have to determine whether a candidate will be productive or not. One of the tools they use is the rejection of the candidate by previous employers. LTU are automatically classified as rejected several times (even though this is not necessary true). Many authors tried to detect and quantify this last effect, parcelling it out from heterogeneities and the loss of human capital. All of them concluded there was a scarring effect due to long-term unemployment (Biewen and Steffes 2010, Oberholzer-Gee 2008, Omori 1997, Kroft et al. 2013, Baert and Verhaest 2014, Ayllòn 2013).

ALMPs lowering the labour cost, such as Law 407/90, can be used to counteract these effects even in the long term. Once a LTU has been hired thanks to the labour cost reduction, he/she will have recent working experience. His/her contacts with the labor market will increase. The hiring employer will have the possibility to screen him/her and overcome imperfect information. Hence, an evaluation of such a policy, in order to determine its effectiveness, can give a big contribution to the fight against LTU.

Law 407/90 was an active labour market policy (ALMP) implemented in Italy from January 1991 until December 2014. The policy was promulgated on December 29th 1990. According to it, any firm hiring, with a permanent contract, individuals who had been either in unemployment status, or suspended from their job, or in Cassa Integrazione (temporary layoff), for at least 24 months, had access to work tax credits for a period of 36 months. In Italy, firms have to pay, for each employee, a rate of his/her wage to the social security service, and a much smaller rate of his/her wage to the institution providing work insurance. The tax credits corresponded to 50% of the total amount of employee taxes, for regular firms. It corresponded to 100% of the same amount for artisans firms and firms located in the Mezzogiorno area. The policy aimed at reducing the rate of long-term unemployment.

A firm, to be eligible, could not have experienced firings or workers' suspending or voluntary resignation or the end of a temporary contract in the last six months. This requirement avoided the temptation, for the employers, to substitute workers of the firm with individuals hired through Law 407/90. In June 28th 2012 rules defining eligibility on firm side were relaxed and conditions in order to be classified as unemployed slightly changed. The law ended on December 31st 2014 according to 2015 Legge Stabilità, promulgated on December 23rd 2014.

This policy was widely exploited on italian labour market. According to INPS' reports (INPS is the National Institute for Social Security) it has been the policy, for the increase of permanent contracts, with the highest number of recipients between 2011 and 2014[4]. In particular, the number of individuals who benefitted by this Law went from a minimum of 295'417 to a maximum of 305'327 among the years 2011-2014. We focused on the impact of this policy only on LTU hirings, not considering the other categories it targeted. According to the data we've used (see section 2.3.1), 28.5% of the LTU hired with a permanent contract in 2014 were benefitting from this policy.

In figure 2.1, it is represented the share of hired unemployed, for the period from 2011 to 2014, with respect to unemployment duration.

[4]Source: Statistiche in breve, Politiche Occupazionali e del Lavoro, http://servizi.inps.it/banchedatistatistiche/menu/index.html

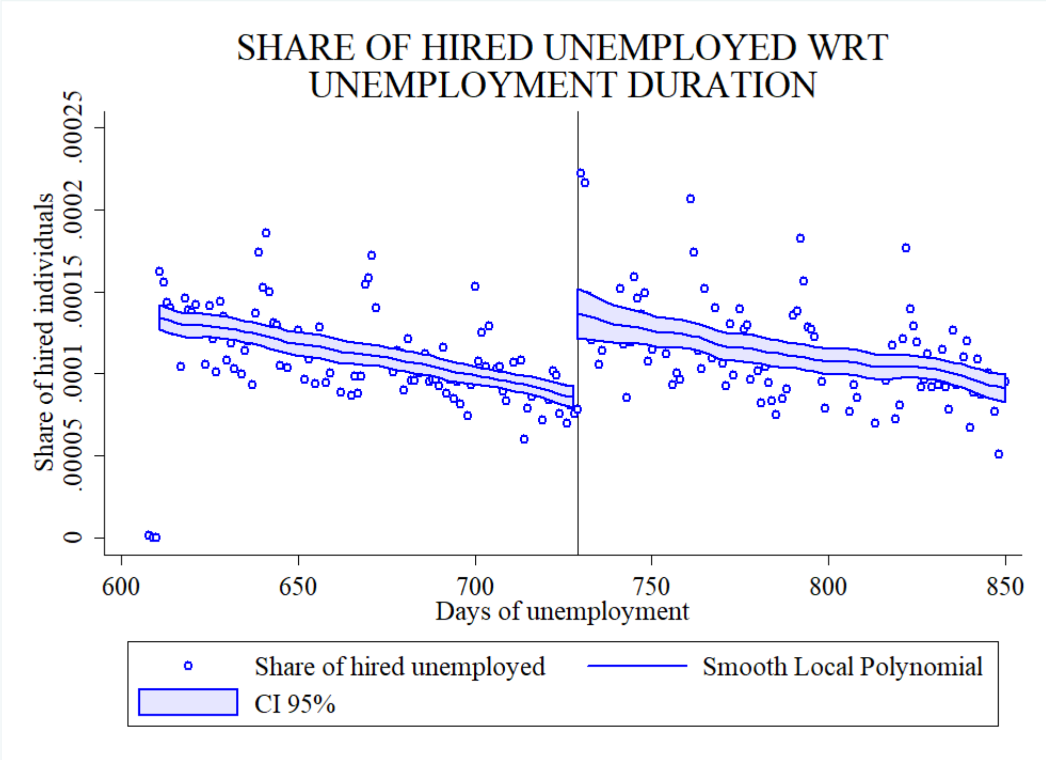


Figure 2.1: The figure represents the share of hired unemployed workers, with respect to unemployment duration, for the period from 2011 to 2014. The variable was based on CICO database.

It is possible to see a clear upward jump at the 729 threshold defining eligibility (evidenced by the black line). This gives a first suggestion on policy effectiveness.

2.2. *Literature Review*

Tax credits can be considered a particular type of wage subsidies given that the basic idea behind both of them is to lower labour cost. Literature on wage subsidies' evaluation is huge. Most of the studies agree on the fact that they have a positive and significant effect (Cockx et al. 1998, Forslund et al. 2004, M. et al. 2005, Bucher 2010, Sianesi 2008, Bernhard et al. 2008, Hamersma 2008, Neubaumer 2010, Jaenichen and Stephan 2011, Anastasia, Giraldo, et al. 2012, Eppel and Mahringer 2013, Farooq and Kugler 2015, Mortensen and Pissarides 2001), few of them concluded they have a null effect (Schünemann et al. 2013[5], Boockmann et al. 20075, Boone and Ours 2004) and, to the best of our knowledge, only one of them concluded they have a negative effect (Kluve et al. 2008). Finally, all the literature reviews suggest they have a positive impact (Martin and Grubb 2001, Martin 2015, Brown 2015, Calmfors et al. 2002). To evaluate them, different methodologies have been used. They can be grouped in the following categories: theoretical models, matching, diff-in-diff (or variations of it), duration models, models relying on instrumental variables approach, regression discontinuity design.

Analysis relying on theoretical models can have bigger or smaller empirical components. As an example, Mortensen and Pissarides (2001) used a model almost completely theoretical. Its only empirical element was model calibration. The authors studied the impact of different hiring subsidies and firing taxes on wages and unemployment, using a search and matching model: Mortensen-Pissarides ones. They used a matching function to describe search and recruiting process, assumed the firm had to face hiring and firing costs and allowed for the presence of idiosyncratic productivity shock causing resource reallocation and job destruction. Hiring costs were assumed to be higher the higher were the required skills. Obviously, hiring subsidies lowered hiring costs and firing taxes increased job destruction costs. From the analysis, it emerged that wage and employment subsidies increase employment, especially among low skilled workers. Job creation and hiring subsidies, instead, decrease unemployment duration but increase its incidence. The authors concluded job creation and hiring subsidies have an ambigous effect on labour market conditions. Nevertheless, in a framework where long-term unemployment has to be reduced, a decrease in unemployment duration should be more meaningful than an increase in unemployment incidence.

Bucher (2010) did a similar use of empirical data still in theoretical model framework. She used real french data to calibrate her model. She studied the impact of

[5] Nevertheless the authors estimated the intention-to-treatment effect rather than the treatment effect.

hiring wage subsidies (under the form of a tax credit), targeting long-term unemployed, in low-skilled labor market. The author used a partial equilibrium model in a continuous time setting allowing for two types of labour force (workers specialized in production and unemployed specialized in job search) and two types of jobs (differening in terms of firm's technology). She modeled the probability to be in unemployment status using a Poisson process. She concluded the program, if implemented on a large scale, would stimulate labor demand of the targeted group, reduce long-term unemployment and increase total welfare. Nevertheless, it would lower the probability of short-term unemployed to find a job.

Neubaumer (2010) went further in the use of empirical data, verifying the results of his theoretical model with a totally empirical matching method. The model described firms' hiring and investment decisions in case of quasi-fixed cost per employee. It allowed for the presence of wage subsidies and training programs. It was based on different assumptions:

- The probability of reaching the intended duration of employment decrease with the increase of unemployment duration.
- Unemployment raise setting-in costs and lower productivity.
- Given a setting-in cost, wage subsidies have an impact only if they are higher than the sum of all the costs of productivity reduction and the costs of firing an employee.
- Training program decrease worker setting-in costs and raise his/her marginal productivity and the probability to reach the intended employment duration.

From the theoretical model the author concluded that firms which need mostly short jobs exploit mainly subsidized workers. Firms which need mostly permanent and highly qualified jobs, instead, choose workers who followed training programs. He concluded in the short run wage subsidies give better results, from an unemployed perspective, while in the long run training programs do. To verify the validity of his conclusions, the author used a matching method. He used German administrative data coming from the German Public Employment Service. The database included all entries, in the evaluated programs, during March 2003, and a random 10% sample of unemployment stock during the same month. He used data about West Germany and individuals aged 25–54 only. As outcome variables, he used the cumulated days spent in regular employment and the share of individuals in regular employment 3.5 years after the program start. To match the units, the author used a propensity score estimated with a probit model. He compared individuals taking a subsidized job, separately, with both the whole group of individuals joining a training program and the subset of this group taking up a job afterwards. The choice to make one of the comparison using only individuals who took a job afterwards is common in this literature. It comes from the fact that, since the subsidy requires the individual to be hired, there may be a selection even once controlled for all eligibility determinants. Indeed, only a restricted group of eligibles, able to find a job under the subsidy, is treated. If selection is not taken into account properly, the comparison between the two programs can be biased. From the analysis, it emerged that there was no significant difference between employment perspectives of wage subsidy beneficiars and those of participants to the training scheme. Wage subsidies beneficiars at first have a better performance, but they're catched up after a while. Nevertheless, wage subsidies were significantly better than training programs in general (hence including in the training program the group of individuals who were not hired after its ending). The author concluded that most of subsidized employment survive after the end of the subsidy. It emerged as well that in medium-long run individuals participating to the programs (both wage subsidies and training programs) have a bigger probability to have a stable job with respect to unemployed who did not participate to any program. Nevertheless, in the short term, the opposite is true, due to a lock-in effect.

The biggest issue about the use of theoretical models in this framework is that often strong assumptions are required to obtain a precise result. Not all of these assumptions are satisfied in real world and they are hard to check.

The empirical method applied by Neubaumer (2010) is one of the most used to study this type of policies. Probably one of the most important work using this method is Sianesi (2008). The author evaluated six different ALMPs implemented in Sweden. Among others, there was a program providing job subsidies and targeting long-term unemployed. She used a database combining two administrative data sources: the program database (Handel) and the unemployment benefits database (Akstat). She included in the analysis only individuals older than 25, who became unemployed for the first time in 1994 and who were recipients of unemployment insurance. Sianesi used a propensity score matching method, estimating the probability to enter in each program with several multinomial models, one for each monthly spell. The matching was based on some observable characteristics as time-varying employment office and local conditions, demographic variables, human capital and labour characteristics, caseworker subjective and time-evolving judgments of client's character, overall prospects and needs of service. Moreover, she included characteristics that could affect individual's past employment history, current employment prospects and self-assessment of the strength of his own chances of re-employment and others variables. She used, as outcome of interest, individual employment and benefit collection probabilities over time. From the analysis, it emerged that entering a job subsidy program, rather than being unemployed, has a positive effect on employment rates soon after the end of the program (35%) and up to five years on (20-25%). From the comparison of different benefit programs, it emerged that job subsidies are the most effective.

A similar empirical strategy has been used by Eppel and Mahringer (2013). They evaluated different ALMPs, including wage subsidies targeting long-term unemployed or individuals at risk. The authors used a database combining two different administrative data sources, the Austrian Social Security Database (ASSD) and the Public Employment Service (PES) database. Their sample included all adults aged 15-54 years old registered as unemployed between January 2003 and December 2006. The authors used a binary logistic regression model to estimate the propensity score. They used four different matching algorithms and two different control groups. The first including unemployed non-participants. The second including non-participants starting non-subsidised employment. It emerged that treated individuals have a higher probability to be in employment after programme ending compared to those who remains unemployed both in the short and in the long term. Nevertheless, there is no significant difference in the short term and there is a slightly significant difference in the long term, when the treated are compared with employed non-participants. The authors concluded taking up a job have a positive impact indipendently from the type of job. The impact was stronger for long-term unemployed.

Jaenichen and Stephan (2011) as well used a matching method. They studied the effect of a wage subsidies program implemented in Germany from 1998 to 2003. They focused on the part of the program targeting hard-to-place workers (i.e. disabled and long-term unemployed). The group of treated included individuals taking a subsidized job during the second quarter of 2002. The control group included, instead, individuals who either didn't participate to the program or participated later in time. The control group was disaggregated to make three different comparisons. One with the unemployed workers. One with the individuals hired during the afore-mentioned period. One with the participants to on-the-job training program. The authors used administrative data coming from the database Integrated Employed Biographies of the German Federal Employment Agency. They used a propensity score nearest-neighbor matching with replacement. Their outcome was whether an individual was in unsubsidized employment at the beginning of a month and whether a person had successfully avoided unemployment at the beginning of a month. To build the propensity score they used as covariates: socio-demographic characteristics, variables on the five-year-history prior to the considered unemployment spell, timing of unemployment entry and informations on regional labour market situation. From the analysis, the authors concluded that three years after the start of the program participants had 25-42% higher probability to be in regular employment than people in the control group who was unemployed three years before. Nevertheless, there was no significant difference between treated and individuals taking an unsubsidized job in the control group.

Bernhard et al. (2008) as well, used a matching model to study the impact of a wage subsidies policy implemented in Germany. They employed the same database exploited by Jaenichen and Stephan (2011) conducing the analysis on individuals who received the UI and were not older than 57 years old. They considered as treated individuals who received the subsidy between February and April 2005. They used as outcomes the percentage of participants in unsubsidized employment, the percentage of non-employed individuals and the percentage of individuals not receiving the UI in subsequent months. As covariates, instead, socio-demographic characteristics, individual's and partner's labour market history during the last five years, household characteristics, local labour market characteristics and some interactions effects. Treated and control units were matched using a propensity score matching, exploiting six different matching algorithms. In line with other results on German wage subsidies programs they concluded there was a positive and significant effect, higher for older and higher skilled individuals and for long-term unemployed.

Kluve et al. (2008) used a methodology that exploited the availability of panel data. The authors, matched the units both according to their socio-demographic characteristics and according to pre-treatment outcomes. The high number of units allowed them to use an exact matching. They calculated treatment effect for each treated unit separately. Subsequently, they averaged the calculated values weighting them according to the frequency, in the sample, of the corresponding unit past labour history. They analysed the impact of a program of subsidized employment and a program of job training implemented in Poland between 1992 and 1996 using data coming from the Polish Labour Force Survey. In contrast with all other analysis, they found a negative effect. The authors concluded that individuals participating to the program may be labeled as less productive.

Hamersma (2008) as well used panel data, to examine the impact of Work Opportunity Tax Credit and Welfare-to-Work Tax Credit on employment outcomes of disadvantaged workers. She applied a propensity score matching combined with a DID model. The author matched with respect to socio-demographic characteristics, Country unemployment rate and months of UI receipt. To avoid the selection problem mentioned before (only unemployed individuals that are able to find a subsidized job can be treated) she defined eligibility for the program as the treatment.

The data she used came from three different administrative sources. The control group included all individuals who received 6 to 8 months of receipt, given that 9 months were needed in order to be eligible for the policy. The author concluded the policy has a positive impact in the short run, but it does not in the long run. Nevertheless, she hypothesized this is due to its low participation rate.

The use of matching model in this framework is a valid choice if the data available allows to do it. Indeed, the database has to contain enough informations to satisfy CIA.

Another approach used to determine the impact of a wage subsidies' policy relies on duration dependence analysis. Even though duration dependence wasn't born as a counterfactual method it can be used as one in this framework. Indeed, it can be used to estimate two different hazard functions, one for treated and one for control groups. Duration dependence models allow to obtain the so-called baseline hazard function which describes the probability to exit from unemployment parceled out by all the influence of individuals' characteristics. This means, in this framework, it can be used to obtain the distributions of the outcome for the two groups, cleaned by group's characteristics. To obtain it, the researchers control for observable heterogeneity and integrate out the unobservable ones. The use of duration models give the possibility to better reflect the continuous nature of the problem. Furthermore, it gives an immediate estimation of the probability to reach employment state for different unemployment spell lengths. Cockx et al. (1998) used a duration dependence model to study the impact of three different policies implemented in Belgium on the probability of leaving unemployment. The studied policies were subsidized on-the-job training, classroom training for unemployed and pure wage subsidies to employers. They used a sample period covering the years 1991-1993 and collected data through a survey proposed to a sample of recruiters chosen by the HR manager of the hiring firm. The authors assumed the hazard function to be given by the product between the baseline hazard and the exponential of the linear combination of observable characteristics plus unobserved individual specific effect. To modelize the baseline hazard they used a flexible piece-wise constant specification. They allowed unobservable characteristics to be correlated with the decision to participate to the program. To control for it, the authors used a nested logit specification to model the participation probability as dependent on time-invariant individual characteristics and unobserved individual specific variables. To take into account of sampling probability, instead, they assumed it to be independent from individual characteristics, once conditioned on the hiring firm and the probability to be eligible. Moreover they assumed the probability to work for the hiring firm to have a beta distribution. To model unobservable characteristics they used two different distributions. It emerged that all the policies have a positive a significant effect on job duration.

Carling and Richardson (2004) used a duration dependence model to compare the impact of different ALMPs implemented in Sweden between 1995 and 1997. Among them, there was API, a wage/employment subsidy. The authors used administrative data from the Public Employment Office, reducing the analysis to individuals registered since August 1991, aged between 25 and 54 years old, who became unemployed for the first time between January 1995 and December 1997. Since according to previous studies, there was no self-selection, into different policies, on workers side, they took into account of observable heterogeneity only. The authors used, to modelize the functional form of the hazard function, a piece-wise linear hazard where the spells defining linearity were 30 days long. They concluded that programs implying practice in a firm gave better results with respect to the ones characterized by vocational training.

Notwithstanding all the positive aspects linked with the use of duration dependence models in this framework, they should be employed with caution. Indeed, Heckman and Borjas (1980) and Berg (2001) warned on the strong assumptions needed to use them. In particular, the firsts underlined as strong distributional assumptions about the nature of heterogeneity and the nature of baseline duration dependence were needed to separate the true duration dependence from the spurious ones. The second underlined as the assumption about multiplicative relation between the components of the hazard function in the Mixed Proportional Hazard Model didn't have any justification in economic theory. Heckman and Borjas criticism have been in part overcame thanks to the use of the flexible piecewise modelization for the baseline hazard function and thanks to some theoretical economic justification of the distribution of unobserved heterogeneity. Nevertheless, the issue presented by Berg (2001) is still relevant.

Forslund et al. (2004) studied the impact of a Swedish wage subsidy policy targeting long-term unemployed. The authors applied IV approach exploiting the fact that the sources of the budget used to pay the programme were different across regions and that there were some budget cutbacks in the studied years. They used administrative data coming from the Handel database of the National Labour Market Board. The authors concluded there is a positive treatment effect of participating in the general employment subsidy programme.

Finally, regression discontinuity design has been used by Schünemann et al. (2013) and Anastasia, Giraldo, et al. (2012). Schünemann et al. (2013) studied a policy targeting long-term unemployed. The policy was implemented in Germany from 1989 until 2002 and it consisted in a wage subsidy given to employers hiring an individual belonging to the targeted group. The authors estimated the intention-to-treat effect (see section 2.3.3.1). Hence, they defined eligibles as the treated group. This allowed them to consider as treated all individuals crossing the threshold of unemployment length, defining long-term-unemployment condition. Since multiple programs targeted LTU they combined the RDD with a DID approach exploiting law's ending in 2003. The authors used a random sample of administrative data combining social insurance and program participation records, benefit payment files and job seekers registers. They reduced the analysis to individuals entering in unemployment between April 2000 and December 2002. From the analysis, it emerged the intention-to-treatment effect was close to zero and non-statistically significant. This may be due to the low take-up rate of the policy.

Anastasia, Giraldo, et al. (2012) used a different type of discontinuity. They studied a policy targeting women and under 30 youths and exploited the age threshold. Such policy, was implemented in Italy from 2012 and consisted in tax credits to firms hiring individuals in the targeted group with a permanent contract or turning their contract from temporary to permanent. The study focused on Veneto region and used as outcome the number of contracts turned from temporary to permanent every day. The authors concluded the policy had a positive and significant effect.

As mentioned before, from this literature review, we can conclude most of recent studies detected a positive and significant effect of wage subsidies policies. This result is robust to the use of different methodologies and different data.

2.3. *Data and Empirical Strategy*

2.3.1. Starting Data

Thanks to an agreement between Dipartimento di Scienze Sociali ed Economiche of Sapienza University of Rome and INAPP (Istituto Nazionale per l'Analisi delle Politiche Pubbliche) we had access to all the databases we have used. We mainly used an administrative micro-database, called CICO. It was provided by the Italian Ministry of Labour and Social Policies (Ministero del Lavoro e delle Politiche Sociali). It contains all recorded employment and parasubordinate contracts[6] and some self-employment events (coming from INPS' data) for a random sample of individuals. Each record corresponds to a different contract and reports the worker ID, the firm ID, contract's and job's characteristics and starting and ending dates, and some basic socio-demographic characteristic of the individual, such as age, income and region of residence. The data are collected by the Ministry directly from the employers, who must register the contract and provide all the information. After the collection of records, the last are submitted, by the Ministry, to a validation procedures. The data started to be collected from 2008. Nevertheless, previous observations were re-built by the Ministry to provide additional information. These data have some great advantages. From 2008 on, they are precise and valid from the point of view

[6]The last is a type of contract, present on italian labor market, having some of the characteristics of employment and some of the characteristics of self-employment

of records of contracts' start and end. They report detailed individuals working history in the continuous across nine years. Nevertheless, they have some limits. Given that the data report only employment experiences, there is a lack of informations on individual's status in the period between the end of a contract and the start of the subsequent ones. Hence, we can't determine with certainty whether the individual was actually unemployed in the missing spell. Nevertheless, we took into account of this issue in the analysis (see section 2.3.4 and 2.4.1).

2.3.2. Data Elaboration

Using CICO database records, we built a new database where each unit identifies, rather than a recorded contract, a group of individuals having a determined number of days in a non-CICO-recorded status, at a determined day. To be more precise, in the new database, unit ij identifies the group of individuals with i days in a non-CICO-recorded status[7], at day j. This means unit ij identifies the group having the last registered contract ending i days before day j and the next registered contract starting after, or on, day j. Consequently, each unit starts to be counted from the end of its first contract registered on CICO database. This re-elaboration of the data has a great advantage. Indeed, it allows to exploit the continuus nature of the data, analysing the phenomenon each day of the period under study. At the same time, it maintains a number of observations and variables low enough for the analysis to be computationally feasible. Indeed, without data aggregation, in order to do the analysis across all days, we would have had to register 1825 variables on the employment status of 3'138'373 units. Finally, there is no loss of useful informations due to the aggregation.

Given that the records are not reliable if recorded before 2008, we re-elaborated the data using only records subsequent to December 31st 2007. Since we mostly focused on long-term unemployed this means we have useful informations starting from 2010. Indeed, it is from 2010 that we start having units far enough from the last recorded contract to be possibly considered LTU. Obviously, this re-elaboration

[7]With a non-CICO-recorded status we mean an employment status not recorded in CICO database (from now on, for simplicity, we will call it "non-occupation" since most of the occupational statuses are recorded).

of the data reduces the number of disposable units. Nevertheless, the database is still substantial. Excluding contracts ending before 2008, CICO original database had 11'016'916 observations, corresponding to 3'138'373 different individuals. Once the data were aggregated, and once selected only the units we used in the analysis (for further details on units selection see section 2.3.3), we had 45'291 observations. The variables recorded in the new database, for each unit *ij*, are the share of individuals hired in day *j*, the total number of individuals belonging to that unit and the share of individuals with given socio-demographic characteristics.

In the re-elaboration we excluded from enumeration all individuals starting a contract thanks to other benefits and individuals whose last ended contract started thanks to other benefits. We excluded the second group believing that hiring under a determined benefit may have a different effect on the following hiring with respect to the standard ones. As an example, there could be a stigma linked to benefit's reception.

2.3.3. Empirical Strategy

To identify treatment effect using counterfactual analysis, we have to make three fundamental choices: the outcome we want to study, the definition of the treated group and the method to use. We used as outcome of a unit *ij* the share of individuals belonging to *ij* group that has been hired in day *j*.

2.3.3.1. Eligible vs Subsidised

In our framework, treated group can be defined in two different ways: either as the group of individuals who actually benefitted from the policy, or as the group of individuals who were eligible for it. We followed Schünemann et al. (2013), Anastasia, Giraldo, et al. (2012), Boockmann et al. (2007), Huttunen et al. (2013), Hamersma (2008) and Forslund et al. (2004) and defined it as the group of eligible people. In their work, Schünemann et al. (2013) justified this choice with different motivations. I will now report only the most relevant ones in this framework. First of all, using the actual beneficiary as treated and comparing workers with similair probability to be treated, don't assure the absence of selection bias. This is because the tax credit has to be required by the employer, hence there may be employers selection. Employers who know about the tax credit, and decide to require it, may be different (and consequently have different hiring behaviours) from the employers who don't (i.e. a big firm that have labor consultants will more likely know about the policy than a small ones). Hence, controlling for employees characteristics only, may not take into account of all the selection bias problem. The use of eligibility as treatment solve this problem. Secondly, using eligibility gives the possibility to apply a Regression Discontinuity Design exploiting the eligibility threshold of the policy. As we will explain in the following section, this method is ideal to use in our framework. The last and most important motivation is that, using eligibles as treated group, we estimate the intention-to-treatment effect. The latter is the most policy relevant in this framework. This is because tax credits can't be mandatory. Policies based on them give the possibility to take up them, not the tax credit itself. Hence, policy makers only have control on the intention-to-treatment, they have no power on tax credit use. It is of biggest interest the impact of policies that policy makers can control, and the last have control on eligibility rules only.

2.3.3.2. Methodology: Regression Discontinuity Design

As visible from the literature review, many different methodologies can be used in this framework. Nevertheless, some of them can be used under particular conditions only (see section 2.2). To reduce as much as possible the needed assumptions, we excluded from the available methodologies those based on theoretical and duration dependence models. The available dataset does not allow to use a matching method properly. Indeed, we do not have exact information on the actual status of individuals during non-occupation spells. Hence, we could match a housekeeper with someone who is studying in order to get back on labour market with a higher human capital. Other methods appliable to this framework are the diff-in-diff model and the Regression Discontinuity Design. The last is preferable to the first for two reasons. First of all, if some assumptions are satisfied, when units at the threshold are considered, the results are as reliable as the gold standard of policy evaluation: randomized experiments (Lee 2008). Secondly, it is possible to check the validity of

its assumptions.
The main conditions to apply the regression discontinuity design are the presence of a threshold, on a continuous variable, defining treatment assignment, and the random distribution of individuals around the threshold. In the context of Law 407/90, eligibility (hence our treatment) is defined with respect to the 24 months threshold on a continuous variable reporting unemployment days. Moreover, the assumption of randomness in the distribution of unemployed workers around the threshold and the requirement of absence of sorting are likely to be satisfied (see section 2.3.4).
Hence, we used a regression discontinuity design to estimate the impact of Law 407/90. As mentioned before, Lee (2008) demonstrated that, if units can't manipulate their forcing variable, at the threshold, a regression discontinuity design works similar to a randomized experiment. Nevertheless, when a regression discontinuity design is implemented, there aren't simultaneously treated and untreated units precisely at the threshold, to calculate the average treatment effect on. In past literature, two different approaches were developed in order to extend the area where the average treatment effect could be calculated. In the first approach, a local polynomial regression (either parametric or non-parametric) is used. The idea is that, once controlling for the forcing variable (hence once it is included in the regression model), the notion of local can be extended and a wider bandwidth can be used. Following this approach, it is fundamental the choice of a bandwidth. The last has to guarantee a low MSE and, simultaneously, consistent inference parameters. The units inside the chosen bandwidth can have slightly different characteristics, the presence of the forcing variable (and, sometimes, of other covariates) in the model, allows to take it into account. The second approach has been developed more recently by a new strand of the literature. Its basic idea is to include in the estimation units around the threshold close enough to it, to still consider the model as a local randomized experiment. The literature on this method mostly focused on the conditions necessary to define the regression discontinuity design as a local randomized experiment (Cattaneo et al. 2015, Li et al. 2015). We followed the second approach, and in particular Cattaneo et al. (2015) ones, because of the nature of our forcing variable (see the section 2.3.3.3).

Hence, we used the following linear regression model:

$$y_{ij} = \alpha + \beta D_{ij} + \theta_j + \epsilon_{ij} \tag{2.1}$$

Where the dependent variable is the share (with respect to the number of individuals in the entire group) of individuals of group (ij) hired in day j. Variable D_{ij} is a dummy taking value 1 if individuals in group (ij) are eligibles. The θ_j are the daily fixed effects. The presence of daily fixed effects allow us to consider each daily comparison as independent from the others. Moreover, it allows us to overcome seasonality issues. Contracts' starting and ending dates are characterised by a strong seasonality. Not taking into account of it could bias the estimation of treatment effect. To be more clear, there are months of the year, as December, characterized by a high number of contracts' endings and a low number of hiring. Since individuals become eligible in the month their last contract ended, there is a particularly big group of individuals becoming eligible in a month characterized by low hiring and this can bias the estimation. Controlling for the hiring day we overcome this issue. The error term is ϵ_{ij}. It is easy to see that the effect of Law 407/90 is given by parameter β.

The re-elaboration of data we did (see paragraph 2.3.2) has some consequences on the type of regression we have to use. Since we aggregated the data, a single unit ij can represent both a huge and a low number of individuals. Hence, some values of the outcome may be underepresented with respect to their presence in the real population. When this is the case an OLS estimation gives biased and unconsistent results and WOLS has to be used (Winship and Radbill 1994). To reproduce the distribution of the individuals in the real population we used, as weights, the total number of individuals represented by each unit. This means we weighted unit ij according to the total number of individuals with i days of non-occupation at day j.

When using standard linear regression model, in order to build confidence intervals, and estimate coefficients' standard error, some assumptions are needed. I.e. to build confidence intervals the normality assumption is needed. To estimate the standard error, instead, it is necessary to assume the residuals not to be clustered.

Nevertheless, there are methods to estimate the confidence intervals and the standard error that are robust to the violation of these assumptions, as bootstrap method. To verify whether the assumptions were violated we estimated the confidence intervals in a first regression without weights, both using a standard and using a bootstrap method. The results were similar enough to conclude the previous assumptions were not violated. Nevertheless, the inclusion of weights in a simple homoskedastic OLS estimation, can cause a heteroskedasticy problem (Winship and Radbill 1994). We tested for heteroskedasticity using a Breusch-Pagan test and concluded it was present. Hence, we used a White estimation to determine standard deviation and confidence intervals.

2.3.3.3. Time-Varying Forcing Variable and Bandwidth Choice

As mentioned above, when Regression Discontinuity Design is applied there are two possibilities for bandwidth selection. The first one is to widen the bandwidth and include, among the covariates of the regression model, a linear polynomial of the forcing variable. The second is to select the bandwidth such that, inside it, a randomized experiment is reproduced. The idea behind the first approach is that, through the forcing variable, we can control for all the characteristics determining selection into treatment, and this allows to widen the bandwidth. If the forcing variable is time varying, this approach can't be used. When the forcing variable is not time varying there is a single event where it is defined. The individual characteristics determining this forcing variable are those determining selection into treatment. Therefore, under the assumption that the potential outcomes are continuous at the threshold, controlling for the forcing variable allows to control for all the characteristics determining selection into treatment and affecting the outcome. When the forcing variable has the particularity of changing over time, in different time period, a different value of the forcing variable is observed for the same unit. Consequenlty, all units have, in a determined time period, a forcing variable taking low values, independently from their characteristics. I.e. imagine a firm going bankrupt has to fire two employees. The two have similar competences but one of them is older and he/she is less familiar with technology. Potentially, the older worker will remain unemployed for longer time. Nonetheless, when the firm had just fired them we observe, for the two employees, the same value of the forcing variable. Controlling for that forcing variable we are not able to control for the potential length of their unemployment spells, which determines selection into treatment. This is why, in time-varying forcing variable frameworks, as ours, the first approach for bandwidth selection can't be used.

In this context it may be useful to distinguish between the observable forcing variable and the potential forcing variable. The first is the forcing variable researchers are able to observe directly. The second is a resume of all the characteristics determining selection into treatment. In the example above, the two employees have the same observable forcing variables but the older worker has a bigger potential forcing variable. Obviously, what we are interested in, when applying a regression discontinuity design, is the potential forcing variable. When the variable is time fixed, the observed forcing variable and the potential forcing variable coincides. Hence, it is enough to control for the observed forcing variable to be sure to control for all the characteristics determining selection into treatment. In time-varying forcing variable frameworks, instead, the observed forcing variable takes multiple values and it does not necessary coincide with the potential forcing variable. If we follow the local randomized experiment approach, there is no need to control for the characteristics determining selection into treatment. Indeed, it allows us to choose a bandwidth small enough to be able to assume that all those determinants are equally distributed between treated and control units. Hence, the local randomized experiment is the only possible approach in frameworks where the forcing variable is time-varying. Our opinion is that, even though the literature is full of regression discontinuity designs applied in frameworks with a time-varying forcing variable, this issue was never considered adequately.

Cattaneo et al. (2015) method follows the local randomized experiment approach. Nonetheless, we did not apply their method as it is. Cattaneo et al. (2015) developed a method that allow to do inference even when the number of observations around the threshold is particularly low (as it is often the case in RDD frameworks). This method requires an assumption on the distribution of treatment assignment inside the chosen bandwidth. Nevertheless, in our database, even when the smallest bandwidth is chosen, we have a huge number of observations. Hence, we can avoid to make any assumption on treatment assignment distribution and use standard tests. Moreover, Cattaneo et al. (2015) propose to select the bandwidth according to the results of a balance test on different covariates. In our framework, we want to do the balance test taking into account of the inclusion of daily common effects in the model. Indeed, we compare treated and control groups each day, consequently, we want the two groups to be balanced at daily level. One of the test they used to compare treated and untreated groups and select the bandwidth is the unpaired t-test. This test consists in a comparison between the two means of the two groups. We used a paired t-test rather than a regular one, pairing units belonging to the same day. Using this approach we check whether the difference between the treated and the control groups each day is null on average. Hence, we obtain a treated and a control groups balanced and comparable each single day.

In bandwidth selection process, we used the following covariates: the share of women, of individuals with different educational qualification, of individuals who started their first job at different age class, of foreign citizens, of individuals working in different working sectors or geographical areas during the last contract.

2.3.4. Needed Assumptions

To have a valid result applying the regression discontinuity design, some assumptions have to be satisfied. In the following paragraph we will introduce them and discuss their plausibility.

Since we decided to follow Cattaneo et al. (2015) approach, the model we used required all their assumptions to be satisfied. We adapted them to our framework, in order to take into account of the presence of daily dummies:

Assumption 1 (Local Randomization). *There exists a neighborhood $W_0 = [\underline{\psi}, \overline{\psi}]$ with $\underline{\psi} < \psi_0 < \overline{\psi}$ such that, for all units with $i \in W_0$:*

(i) $F_{\psi_{ij}|\psi_{ij} \in W_0}(\psi|j) = F(\psi|j)$, *and*

(ii) $y_{ij}(\psi|j) = y_{ij}(D_{W_0}|j)$ *for all* ψ.

Where the symbol ψ defines a given value of the forcing variable, ψ_0 is the threshold of interest and ψ_{ij} is the value of the forcing variable of unit (ij) and F is a probability distribution. The first part of the assumption requires the distribution of the forcing variable to be the same for all units inside the interval for a given day. The second part requires the potential outcomes of units inside the interval to depend, each single day, on the forcing variable only through treatment assignment. If this assumption is satisfied, we can ignore the value of the forcing variable inside the interval. From here, the formulation of the regression model only with the treatment dummy (Cattaneo et al. 2015). To be more intuitive, assumption 1 requires the units to be distributed randomly inside the selected bandwidth. Apparently, this assumption is likely to be satisfied. Indeed, being, at a determined day, unemployed from 23 months and 28 days or 24 months, is not under control of the unemployed. There is a random component on the day the previous contract ended. There is another random component in the employers the unemployed got in contact with and the time it happened (to be unemployed since 24 months when hired, the unemployed has to find the right match at the right moment).

Nonetheless, the assumption may be violated in presence of sorting. Sorting may come from different sources. First of all, the unemployed may have been able to control their recorded number of days in unemployment. This is unlikely to be. Indeed, the recorded number of days in unemployment was determined according to the registration of the unemployed worker at an employment agency. Hence, the forcing variable value was completely under control of the employment agency. The unemployed had no control on it. Sorting could come as well from a caseworker who thinks that determined types of unemployed individuals are particularly suitable for this wage subsidy program and it is worth to modify his/her intervention to be sure that they stay in unemployment long enough to become eligible. Nonetheless, this is unlikely to be the case for different reasons. First of all, the value added of the subsidy program is the hiring. Hence, if the unemployed worker can be hired without it, there would be no value added to program participation. Secondly, the number of italian caseworkers is low with respect to unemployment level, suggesting it is unlikely for the LTU to receive a tailored service. I.e. in 2012 there were on average 254.2 individuals per caseworker who made a declaration of immediate availability for work at the employment office[8].

Sorting may derive as well from employers that decide to hire an unemployed person close and under the threshold and wait until the last became eligible to get the subsidy. We checked for this possibility in the following chapter (see paragraph 3.1).

Assumption 2 (Local stable unit treatment value assumption)**.** *For all units with $i \in W_0$: if $D_{ij} = D'_{ij}$ then $y_{ij}(D_{W_0}|j) = y_{ij}(D'_{W_0}|j)$.*

With this assumption, we are requiring nothing more than the SUTVA to be valid locally (inside the interval). We are requiring that locally, how the treatment is enacted is irrelevant to each unit's potential outcome. We are requiring as well that there is not interference among units. Therefore treatment assignment of one unit has no influence on the potential outcome of another unit. This assumption may be violated in presence of negative side effects. This issue will be discussed more in detail in the next chapter (see section 3.1).

Assumption 3 (Zero treatment effect for covariate)**.** *For all units with $i \in W_0$: the covariate $x_{ij}(\psi)$ satisfies $x_{ij}(\psi; j) = x_{ij}(D_{W_0}; j) = x_{ij}(j)$ for all ψ.*

This assumption requires the treatment effect on the covariates to be zero in the interval where assumption 1 holds.

Assumption 4 (Association outside W_0 between covariate and score)**.** *For all units with $i \in \widetilde{W}$ and for all $\psi \in \widetilde{W}$:*

(i) $F_{i|i \in \widetilde{W}}(\psi|j) = F(\psi; x_{ij}(\psi)|j)$, *and*

(ii) *For all couples of units $(ij), (pq)$ with $i \neq p$ and* $\begin{cases} j = q \\ or \\ j \neq q \end{cases}$ *either*

$$x_{ij} > x_{pq} \Rightarrow F(\psi; x_{ij}|Z_j) < F(\psi; x_{pq}|Z_q)$$

[8]Source: "Indagine sui servizi per l'impiego 2013", Cliclavoro

or

$x_{ij} > x_{pq} \Rightarrow F(\psi; x_{ij}|Z_j) > F(\psi; x_{pq}|Z_q)$

where $\widetilde{W} = [\underline{\rho}, \underline{\psi}) \cup (\overline{\psi}, \overline{\rho}]$ for a pair $(\underline{rho}, \overline{rho})$ satisfying $\underline{\rho} < \underline{\psi} < \overline{\psi} < \overline{\rho}$. This assumption requires the score to be correlated with the covariates outside the interval where assumption 1 holds. The last two assumptions justify the use of balance tests between covariates to select the desired interval W_0. Indeed, if there is correlation between the score and the covariates outside the interval and there is no correlation inside the interval the last can be easily identified as the largest interval without correlation. Hence, it corresponds to the largest interval for which the group of treated and the group of untreated units are balanced. Both these assumptions are likely to hold in our framework. Indeed, the selected covariates (such as the geographical area of belonging) are likely to be correlated with unemployment length. On the other hand, the same random components that guarantee us randomization inside W_0 also guarantee uncorrelation between covariates value and treatment assignment. I.e. the previous contract ending five days before or five days later is unlikely to depend on unit's gender.
In addition to the previous assumptions we have to add the following ones:

Assumption 5. *There are no other policies affecting eligibles and non-affecting non-eligibles.*

This assumption maybe violated if other policies use Law 407/90 threshold and we are not able to control for them. Nonetheless, we had the possibility to control for the reception of other benefits. The exclusion of all individuals benefitting from other policies (see section 2.3.2) guarantee us this assumption is not violated.

2.4. *Results*

As mentioned before (see paragraph 2.3.3), to select the bandwidth we used a modified version of Cattaneo et al. (2015) methodology. In table 2.1 the results of the t-test are presented. As Cattaneo et al. (2015) we report different possible windows W_0, on the left column, and the lowest p-value among those obtained from the t-test on each covariate, on the right column. Following Cattaneo et al. (2015), we chose the bandwidth W_0 corresponding to the biggest value of ω having its and all previous p-values higher than 0.15.

ω	p-value
[728:730]	0.6926049
[727:731]	0.6757796
[726:732]	0.7891312
[725:733]	0.7899194
[724:734]	0.7362718
[723:735]	0.6159053
[722:736]	0.5451728
[721:737]	0.5195819
[720:738]	0.4540012
[719:739]	0.4291586
[718:740]	0.3764177
[717:741]	0.2899919
[716:742]	0.2251229
[715:743]	0.1876122
[714:744]	0.149704
[713:745]	0.1151946
[712:746]	0.0905614

Table 2.1: Paired t-tests to select the proper bandwidth: on the left column the bandwidth, on the right column the corresponding lowest p-value

In figure 2.2, it is possible to see a graph of the minimum p-values corresponding to different bandwidth values. The selected bandwidth is highlighted in red.

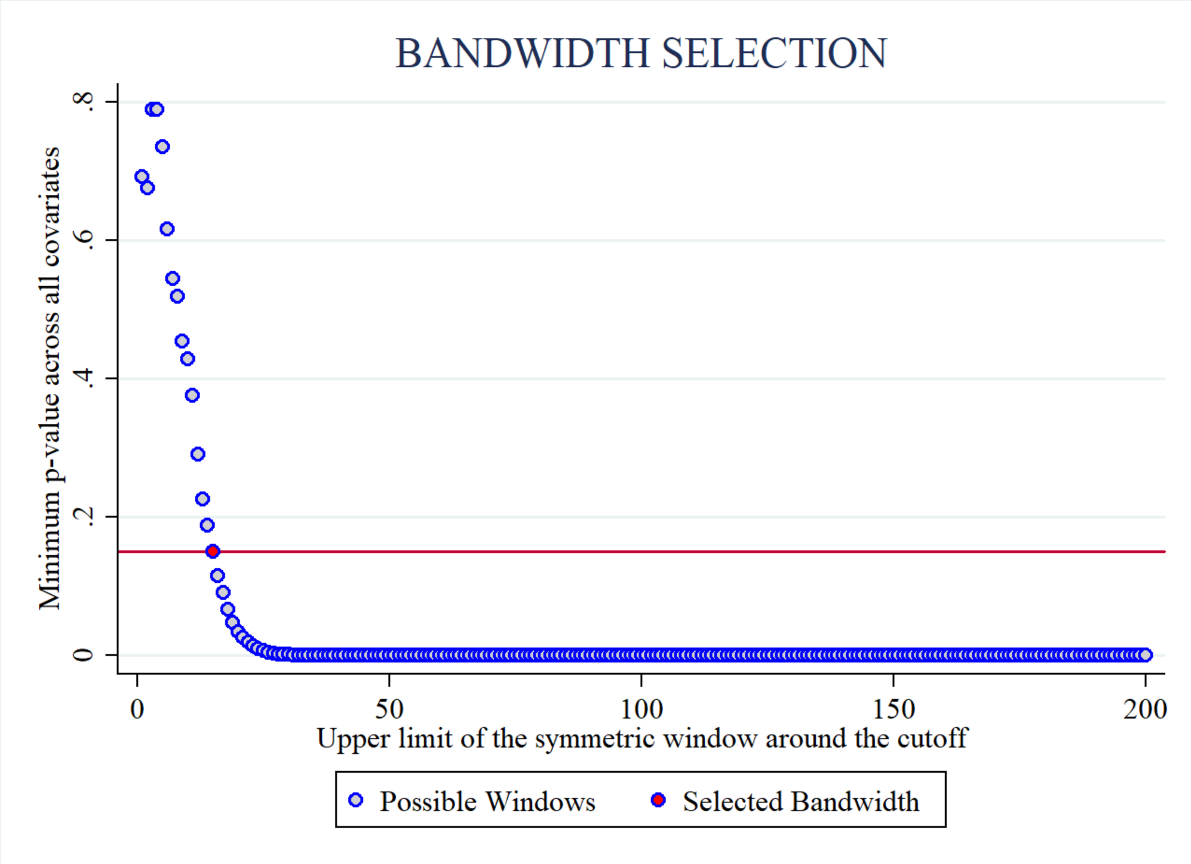


Figure 2.2: The minimum p-values obtained by the paired t-test did on the covariates to select the proper bandwidth is represented with respect to different bandwidth values. The selected bandwidth is highlighted in red. Paired t-tests at a daily level were used.

The selected bandwidth went from 714 to 744 days of unemployment (where 729 is our threshold). Hence, we included all units becoming eligibles in two weeks or less and all units who became eligibles since two weeks or less.

The results of the regression are reported in table 2.2 (obviously, we excluded the daily parameters). As common, we use, as significativity level, a 0.05 ones. The results suggest Law 407/90 had a positive and significant impact. Comparing the estimated impact with the average weighted outcome of the control group, we can conclude the treatment increased the last by 36%. This value is bigger than the results obtained in previous literature. Most of the studies estimating intention-to-treatment effect concluded there wasn't a significant effect (see Boockmann et al.

Table 2.2: ESTIMATION OF THE INTENTION TO TREATMENT EFFECT OF LAW 407/90, REGRESSION WITH DAILY FIXED EFFECTS

VARIABLES	Coefficients
Treat	3.01***
	(0.462)
Constant	9.38***
	(0.282)
Observations	45,291
R-squared	0.094

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

NOTE: We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. The standard errors in the unweighted regression were robust to bootstrap estimation. ITT of Law 407/90 is given by "Treat". For easier reading, coefficients and standard errors were multiplied by 100,000.

2007, Huttunen et al. 2013, Schünemann et al. 2013) and one of them found a positive and significant effect of 10% in the short run (Hamersma 2008). Nevertheless, this difference can be explained by three elements. The first one is that we studied a widely used policy. The second is that it was implemented across a very long period (i.e. 24 years). Hence, it was probably well known by both employers and unemployed individuals. The third is that almost all of the afore-mentioned authors (with the exception of Schünemann et al. 2013) used a methodology which estimated the impact of the policy on the whole population of eligibles. Using a RDD we obtained only local results. This means our estimation describes the effect of the policy only for the group of eligibles included in the bandwidth. This is one of the most affected groups. Indeed, of the hirings under Law 407/90, around 38% involved individuals with 24 to 27 months of unemployment.

In terms of national employment, inside the bandwidth, the policy increase it by

45 individuals per year. If the impact of the policy was the same outside the bandwidth, the policy would have increased the hiring of individuals between 24 and 28 months of unemployment by 1662 per year. The average yearly hiring of individuals in unemployment since 24 to 28 months, hired with a permanent contract, between 2011 and 2014 is 21318.

2.4.1. Model Assumptions and Robustness Checks

We checked whether the results were robust with respect to a seasonal adjustment method different from the inclusion of daily dummies in the model. In particular, we repeated the estimation using data seasonally adjusted with a moving average method. In table 2.3 the result of this estimation (Model 2) is compared with the results obtained using the daily dummies alone (Model 1). The estimated effect do not differ much. We can conclude they are robust to another seasonal adjustment methods.

Table 2.3: ROBUSTNESS CHECK WITH RESPECT TO SEASONAL ADJUSTMENT

VARIABLES	(1) Model 1	(2) Model 2
Treatment	3.01***	2.95***
	(0.462)	(0.466)
Constant	9.38***	9.40***
	(0.282)	(0.291)
Observations	45,291	45,291
R-squared	0.094	0.084

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NOTE: We used a moving average method to seasonal adjust the data. In Model 1 raw data were used. In Model 2 seasonally adjusted data were used. We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. The standard errors in the unweighted regression were robust to bootstrap estimation. ITT of Law 407/90 is given by "Treatment". For easier reading, coefficients and standard errors were multiplied by 100,000.

We also did the robustness checks standardly used when a regression discontinuity design is applied. In particular, in table 2.4, it is represented the sensitivity of the estimated effect with respect to bandwidth choice. The estimation of treatment effect is not very sensitive to bandwidth selection, once assumption 1 is satisfied.

Table 2.4: ROBUSTNESS CHECK OF LAW 407/90 ITT ESTIMATION WITH RESPECT TO BANDWIDTH CHANGES

	(1)	(2)	(3)
VARIABLES	[714:744]	[720:740]	[724:734]
Treat	3.01***	2.43***	2.21***
	(0.462)	(0.575)	(0.822)
Observations	45,291	30,681	16,071
R-squared	0.094	0.124	0.204

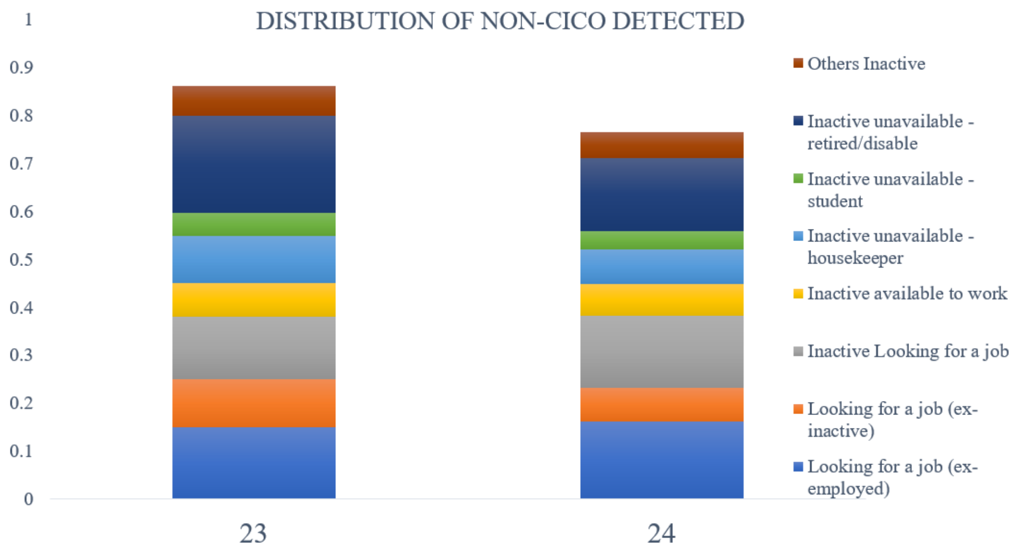
Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

NOTE: We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

Note that the different bandwidths we tested for, are all smaller than the chosen bandwidth. Indeed, given the method we used in bandwidth selection, in a bandwidth bigger than the chosen ones there could be a violation of assumption 1. Hence, treated and control groups may be unbalanced.

Given that in CICO database it is not possible to identify exactly the unemployed individuals we checked as well if treated and non-treated individuals had similar distributions between non-CICO-detected status. To do it, we used RTFL data. The last are survey quarterly data from ISTAT (Italian National Institute of Statistics). It is one of the most important database providing these informations and it is used to build official estimations of labor force status. Data are collected every week interviewing more than 250'000 families living in 1'100 different municipalities (Anastasia, Bertazzon, et al. 2016). We checked whether the rate of individuals in each non-CICO-detected category is equal, each year, between uneligibles with 23 months (from 698 to 728 days) of non-occupation and eligibles with 24 months (from 729 to 760 days) of non-occupation. Given that the more we tight the bandwidth the more the two groups are likely to be equal, this can be considered a conservative check. In the following graph are represented the distributions, among different non-CICO-detected categories, of individuals with 23 and individuals with 24 months of unemployment.



The two distributions are fairly similar.

In table 3.5 the results of an estimation including covariates are presented. The estimated intention-to-treatment effect does not differ from the main estimation.

Table 2.5: ROBUSTNESS CHECK OF LAW 407/90 ITT ESTIMATION WITH RESPECT TO COVARIATES ADDITION

VARIABLES	Coefficients	Std Error
Treatment (β_1)	3***	(0.459)
Women %	8.64	(8.27)
Low Secondary	-5.48	(9.44)
Up Secondary	13.29	(118)
Tertiary no Univ	-109.9*	(64.36)
Tertiary Univ	-4.33	(12.15)
Degree+	-125.2**	(55.14)
20-24	-3.26	(14.16)
25-29	6.86	(12.79)
30-44	11.89	(12.14)
45+	-9.32	(17.01)
Foreigners	-7.47	(12.28)
Industry	16.6	(10.97)
Constructions	31.28***	(7.71)
Services	4.8	(6.52)
NE	-11.04	(10.85)
Center	1.6	(11.48)
South and Islands	7.26	(8.55)

NOTE: We added several covariates. The share of women ("Women"). The share of individuals who completed the following educational stages (where elementary education was used as baseline): lower ("Low Secondary") or upper ("Up Secondary") secondary education, non-university tertiary level ("Tertiary no Univ"), university tertiary level ("Tertiary Univ"), postgraduate education ("Degree+"). The share of individuals in different age ranges (where the age range from 15 to 19 was used as baseline): from 20 to 24, from 25 to 29, from 30 to 44, older than 45. The share of non-italian citizens. The share of individuals working in the industrial ("Industry") or the construction ("Constructions") or the services ("Services") sectors (where the share in agricoltural sector was used as baseline). The share of individuals whose last job was located in the following areas of Italy (where the NW area was used as baseline): North-East ("NE"), center ("Center") or South and Islands ("South and Islands"). We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Obs: 45,291. R-sq. 0.096.

In table 2.6 three different placebo tests are presented. In the first two we use, respectively, 22 and 26 months of unemployment as a threshold (mantaining the same bandwidth of the main estimation). In the third we use Law 407/90 threshold but apply the model to data from 2015.

Table 2.6: PLACEBO TESTS FOR LAW 407/90 ITT ESTIMATION

	(1)	(2)	(3)
THRESHOLD	2011-2014	2011-2014	2015
22 Months	0.00987		
	(0.486)		
26 Months		-0.788	
		(0.504)	
24 Months			-0.960
			(1.28)
Observations	42,369	42,369	11,315
R-squared	0.077	0.109	0.189

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

NOTE: We implemented all weighted regressions, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

All the estimated intention-to-treatment effect are non-significant, suggesting those estimated for Law 407/90 is effectively due to the policy. An additional confirmation comes from the last check. It consists in the study of the correlation between the share of subsidized individuals and the estimated intention-to-treatment effect across different years. In table 2.7, the coefficients of the treatment, the share of subsidized individuals among LTU hired with a permanent contract and the correlation among the two are reported each year.

Table 2.7: LAW 407/90, CORRELATION BETWEEN SHARE OF SUBSIDISED INDIVIDUALS AND ITT ACROSS YEARS

	Correlation	0.79294
Year	Treatment Effect	Subsidized Individuals
2011	4.21e-05	0.398
2012	2.74e-05	0.405
2013	1.46e-05	0.361
2014	3.65e-05	0.41

The share of subsidized individuals and estimated treatment effect present a high correlation. This suggests our regression model effectively detected the impact of the studied policy.

3. Targeted Policies: beyond treatment effect estimation

In the following chapter we will show how we went beyond the simple estimation of treatment effect exploiting laws characteristics and the context it was set in. The chapter starts with an introduction on the possible negative side effects and on how they were investigated in past literature. Later on, we introduce our investigation approach and present the results. The second part of the chapter regards a comparison between targeted and untargeted policies from the perspective of the long-term unemployed workers.

3.1. *Law 407/90 Negative Side Effect*

3.1.1. Which are the possible negative side effects?

Brown and Koettl (2012) and Calmfors et al. (2002) identified several possible negative side effects following targeted ALMPs implementation. In this analysis we focused only on those side effects that can be generated by Law 407/90 (as an example we excluded the locking-in effect which is typical of training programs). Those are the asymmetric information effect, which in our context take the shape of post-poned hiring, and the displacement effect.

3.1.1.1. Post-poned Hiring Effect

Due to asymmetric information, the Government may not have perfect knowledge on what is happening on labor market. This means firms and unemployed can cheat in order to get subsidies even when they're not eligibles. In particular, in our context, an employer who decides to hire an uneligible individual close to eligibility threshold may wait to hire him/her until he/she is eligible in order to get the subsidy. We call this effect post-poned hiring effect.

Literature investigating on post-poned hiring effect is quite contained. One work focusing on this issue is Boockmann et al. (2007). The authors used administrative data coming from the Integrated Employment Biographies database to determine the impact of a policy of wage subsidies implemented in Germany since 1998. To check for the presence of post-poned hiring effect the author verified whether the estimation changed when excluding individuals that were, according to their beliefs, far enough from the threshold not to be affected by it. They concluded there was not post-poned hiring effect. Nevertheless, this can be linked to the fact that, according to their analysis, the policy had no significant effect.

Schünemann et al. (2013) found a similar result. As mentioned before (see 2.2), they studied a policy implemented in Germany consisting in a wage subsidy given to employers hiring unemployed workers from at least 12 months. To check for post-poned hiring effect the authors verified whether there was an unusual behaviour in hiring immediately before the threshold. The authors concluded there was no post-poned hiring effect. Nonetheless, they found no significant impact of the policy either.

3.1.1.2. Displacement Effect

The second possible negative side effect is displacement effect. Non-targeted individuals having similar characteristics to the targeted ones, may be penalised by policy implementation. Indeed, thanks to the policy, they have similar characteristics but a higher labor cost, than targeted individuals. This means the policy lower their relative desirability. In Law 407/90 framework, the implementation of the policy may penalize individuals who are uneligible for it, but close to the threshold. Indeed, workers unemployed from 23 months and workers unemployed from 24 months are similar in terms of lack of recent working experience, scarring effect, etc. Nonetheless, the workers unemployed from 24 months have a lower labor cost. It is likely that an employer prefers to hire an individual with 24 months of unemployment benefitting from the policy. This means that hiring of individuals with, g.e. 23 months of unemployment may decrease when the policy is implemented.

Among the studies testing for the displacement effects of hiring subsidies there is Blundell et al. (2004). The authors evaluated the New Deal for the Young Employed, a policy designed in the UK to move young unemployed back into work. The policy combined job search assistance with subsidized job placement and it was compulsory to have access to unemployment benefits. The program targeted individuals aged 18 to 24 receiving unemployment insurance for at least six months. The authors used as outcome the rate of individuals exiting unemployment within four months since program entrance. They applied a matching difference-in-differences estimator, exploiting the fact that only youths were eligibles and only some areas were treated. They compared treated and untreated areas with similair trends. They found a positive and significant effect on outflows to employment for men. There was an increase of the probability of finding a job of about 5%. Moreover, the authors could not reject the null hypothesis of no displacement and equilibrium wage effect.

Moving towards policies closer to Law 407/90, there is Ham et al. (2011). The authors evaluated the Enterprise Zone Program, a policy providing tax benefits to firms increasing employment in local labor markets of areas lagging behind in terms of economic development. They estimated the impact of the policies using three different difference-in-difference-in-differences estimators. They found a positive and significant effect. To verify the presence of spillover effects, instead, they estimated the impact of being close to ENTZ on the nearest non-ENTZ area using as control the second nearest non-ENTZ. The authors did not find any significant displacement effect. A different result was found in Hanson and Rohlin (2013). The authors used the same method as Ham et al. (2011) to estimate the impact of the federal equivalent of ENTZ programs only and found a significant and positive spillover effect.

In their literature review, Calmfors et al. (2002) included studies on the displacement effect of ALMPs. Among the studies they inspected only one regards general and targeted subsidized employment and it relies on a survey submitted to employers. It was asked to employment officers whether they would prefer to hire a subsidized or an unsubsidized unemployed worker. A strong displacement effect of 84% was found for targeted employment subsidies. Nevertheless, the literature reviewer themselves were sceptical about the results of these studies since qualitative studies have several limits (i.e. the answers may be influenced by the interests of the respondent). They concluded the displacement effect was higher for job creation programs more similar to regular employment.

From this literature review and the ones in paragraph 2.2, it is not possible to take clear conclusions on the existence of negative side effects. Indeed, the studies checking for post-poned hiring effects (Boockmann et al. 2007, Schünemann et al. 2013) analyzed policies having non-significant effects. Hence, it is impossible to determine whether the detection of no post-poned hiring is due to their absence or to the uneffectiveness (or low take-up rate) of the policy. The studies checking for displacement effets, instead, reach contradictory conclusions. Some of them concluded there is no displacement effect (Blundell et al. 2004). Others concluded the opposite (Calmfors et al. 2002, Bucher 2010). Even though they afford the problem of spillovers in the case of hiring tax credits, Ham et al. (2011) and Hanson and Rohlin (2013) results can't be taken into account. Indeed, the two authors investigate on the presence of spillovers effect between regions rather than on the presence of displacement effect between unemployed. In addition, to the best of our knowledge, there are no studies deepening inside the studies of negative side effects for Italian tax credit hiring policies. Our analysis gives therefore an important contribution to the literature on subsidized employment.

Testing for the presence of negative side effects not only gives additional information on the policies consequences, it is also crucial for a proper estimation of policy impact. Indeed, if the control group is affected by displacement or post-poned hiring effects it will not be a good counterfactual for the treated (and SUTVA will be violated). Truly, the control group contains exactly the units most affected by these negative side effects.

To check for negative side effects, approaches exploiting the fact that an area is untreated can't be used in Law 407/90 framework. Even though the fiscal benefit are higher in Mezzogiorno, this area is widely different from the rest of Italy. Hence, a simple comparison is likely to show these differences rather than spillover effects. At the same time, a comparison between Mezzogiorno and the rest of Italy based on a difference-in-differences model can't be used. Indeed, reliable data start from 2010 and the policy began in 1990. Equally, it is not possible to use a DID exploiting policy ending, given that its ending in 2015 coincided with Law 190 start and labor market reaction to the latter could be different in Mezzogiorno area than in the rest of Italy, making the parallel trend assumption hard to believe. Nonetheless, it is possible to follow Schünemann et al. (2013) idea. In the next paragraph we will show how we started from their idea and strengthen it to check for both post-poned hiring and displacement effect.

3.1.2. How we checked for Negative Side Effects

The method we have used to test for the presence of negative side effects, exploit the fact that both of them would imply a reduction in the hiring of uneligible individuals whose value of the forcing variable is close (and obviously under) the threshold. Indeed, individuals close and under the threshold are in competition with the eligibles on the labour market. Consequently, in presence of a displacement effect, their hiring should be lowered by the policy. We can't say the same for individuals far and under the threshold. I.e., it is unlikely that displacement effect affects individuals with 13 months of unemployment. This group is completely different from the group of eligibles. Consequently, there is not competition between them. Similarly, in presence of post-poned hiring effect some of the hiring that should appear close and under the threshold would appear over it. Nonetheless, the same do not happen for hiring far and under the threshold. It is unlikely that an employer would accept to wait 11 months, until the 13 months unemployed individual becomes eligible, to hire him/her, just to get the subsidy. Hence, post-poned hiring effect as well is likely to lower only the rate of hiring of individuals with a value of the forcing variable close to the threshold. The fact that displacement and post-poned hiring effects only affect the units close to the threshold is the only observable element we can use to determine whether the two effects are present (and, eventually, correct the counterfactual estimation for them). To exploit it, we proceeded with two checks. First of all, we did a graphical analysis plotting the average value of the outcome with respect to different values of the forcing variable. The idea is to check whether, getting closer to the threshold, the outcome has an unusual behaviour that can be attributed to the negative side effects of the policy (i.e. it has a sudden decrease). A second check could be to compare the outcomes of groups of units close and far[1] from the threshold. If there are displacement and post-poned hiring effects the outcome of the two groups will be different. Indeed, the units close to the threshold will suffer by the two negative side effects, while it is reasonable to assume, as explained above, that the group of units far from it will not be affected by these effects. Nonetheless, a difference in the outcomes of units far and close to the threshold may be as well a consequence of the different characteristics of the two groups. It would be impossible to determine whether the difference in outcomes is due to the presence of negative side effects or to the different characteristics of the two groups. To solve this issue we assumed that, in absence of negative side effects, the difference between units far from the threshold and units close to the threshold would be constant across time. If this is the case the difference between the two groups after 2014 (i.e. after policy ending) is representative of the same difference before 2014 in absence of displacement and post-poned hiring effects. Indeed, when the policy was not implemented there were no negative side effects. Therefore, in absence of negative side effects, the difference in outcome of the group close to the threshold during policy implementation and after policy ending should be equal to the same difference for the group far from the threshold. For us, it will be enough to compare the two differences far and close to the threshold to verify the presence of post-poned hiring effects. We can think of this analysis as the application of a DID model. To check for the equality between the differences of the two groups we used different statistical tests that allow to compare groups of units. To be more clear, in

[1]When we say units "far" we mean they are far enough to reasonably assume they are not affected by displacement and post-poned hiring effects.

order to determine the presence of the displacement and post-poned hiring effects we followed these steps:

(i) We chose a group of units having values of the forcing variable far enough from the threshold to reasonably assume they're not affected by the two effects[2]. I.e. with a number of months of unemployment going from 12 to 12 and half.

(ii) We selected the group of units used as control in the estimation.

(iii) We calculated weighted outcome average, by month and day, for each year l from 2011 to 2014 separately, and for 2015 and 2016 together, for both the groups:

$$\begin{aligned} AVG_{F,l} &= \frac{1}{(380-365)*365} \sum_{i=365}^{380} \sum_{j\in l} y_{ij} * tot_{ij} \\ AVG_{C,l} &= \frac{1}{(729-714)*365} \sum_{i=714}^{729} \sum_{j\in l} y_{ij} * tot_{ij} \\ AVG_{F,>2015} &= \frac{1}{2*(365)*(380-365)} \sum_{i=365}^{380} \sum_{j\in(2015 \bigcup 2016)} y_{ij} * tot_{ij} \\ AVG_{C,>2015} &= \frac{1}{2*(365)*(729-714)} \sum_{i=714}^{729} \sum_{j\in(2015 \bigcup 2016)} y_{ij} * tot_{ij} \end{aligned} \tag{3.1}$$

Where F define the far from the threshold group and C the close to the threshold ones.

(iv) We calculated the difference between the outcome average of each year l and the outcome average of 2015-2016 for both groups:

$$\begin{aligned} Diff_{F,l} &= AVG_{F,l} - AVG_{F,>2015} \\ Diff_{C,l} &= AVG_{C,l} - AVG_{C,>2015} \end{aligned} \tag{3.2}$$

(v) We led a statistical test to compare the two differences each year.

[2]Unfortunatly, this choice was arbitrary, indeed we do not have any data-driven way to determine whether the group is actually far enough. Nonetheless, repeating the analysis with different groups of units we were able to obtain more robust results.

If the two effects we are testing for are present, we expect to have $Diff_{F,l} >> Diff_{C,l}$ (where the direction of the inequality comes from the fact that the average before 2015 is generally higher than the average after 2014). To compare the two differences we used multiple test whose purpose is to verify whether two populations are equal. In particular, we started with a Student's t-test. The last checks whether the difference between the means of the two populations is equal to a hypothesized value (0 in our framework). This method requires two assumptions: the populations to be normally distributed and the populations' variances to be equal. It is unlikely for the two populations to be normally distributed. Nevertheless, the test is still useful if the sample sizes of the populations are equal, the size is moderate and the distributions have similar shapes. We tested for these hypothesis. In light of the results of hypothesis testing, we used a second test, the Welch t-test, which allows to relax the assumption that the populations' variances are equal. The results reported in the following section refer to the Welch t-tests.

3.1.3. Results

In figure 3.1, is it possible to see a plot of the outcome with respect to the forcing variable. This first graphical analysis suggests there are no post-poned hiring and displacement effects. Indeed, there is no peculiar behaviour of the outcome getting closer to the threshold.

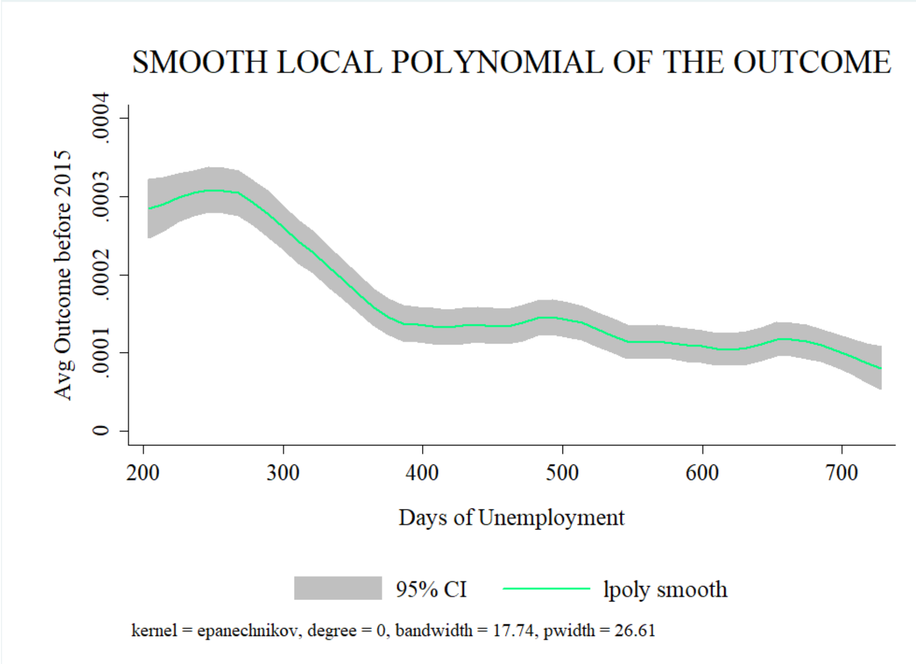


Figure 3.1: The figure shows the outcome with respect to the forcing variable. Data from CICO were used.

Indeed, even though the difference decrease getting closer to the threshold (where the threshold corresponds to a forcing variable value of 729), it does it far from the threshold as well. There is no unusual behaviour getting closer to it.
This conclusion is consistent with the results of the Welch t-test comparing the average difference between two different groups of units far from the threshold and one group of units close to it. Such results are presented in tables 3.1 and 3.2.
From the tests' results it is possible to see that displacement and post-poned hiring effects are not detected in any year.
The fact that these effects aren't present may look surprising. Nevertheless, two elements have to be considered. The first one is that often employers prefer to hire individuals as soon as possible. As an example, from a study by Oberholzer-Gee (2008), it emerged that they often prefer to hire short term unemployed rather than employed individuals, because the first are immediately available to work. Similarly, they may prefer to hire uneligibles close to the threshold immediately, rather than wait for them to become eligibles. In other words, they may prefer a worker imme-

Table 3.1: CHECKING FOR THE PRESENCE OF INDIRECT EFFECTS, COMPARISON WITH THE FIRST GROUP

Year	Diff Mean	Std. Err.	CI 95% Lower	CI 95% Upper
2011	-0.219	5.91	-11.8	11.4
2012	12.88	7.23	-1.32	27.07
2013	-1.17	4-69	-10.39	8.04
2014	-0.0962	4.61	-9.15	8.96

NOTE: Welch t-tests were used to compare the average difference (between the outcome in each year and the outcome after 2015) of individuals with 365 to 380 days of unemployment and those of individuals in the control group. For easier reading, all values were multiplied by 100,000.

Table 3.2: CHECKING FOR THE PRESENCE OF INDIRECT EFFECTS, COMPARISON WITH THE SECOND GROUP

Year	Diff Mean	Std. Err.	CI 95% Lower	CI 95% Upper
2011	-0.364	6.17	-12.47	11.75
2012	11.38	5.8	-0.000496	22.76
2013	8.91	6.04	-2.94	20.76
2014	10.33	5.52	-0.496	21.16

Welch t-tests were used to compare the average difference (between the outcome in each year and the outcome after 2015) of individuals with 545 to 560 days of unemployment and those of individuals in the control group. For easier reading, all values were multiplied by 100,000.

diately available rather than get the subsidy. The second one, is that, in Italy, there are mainly small and medium firms. A small firm probably will not receive a huge amount of applications and curricula at a time. Hence, it is unlikely the firm receives the curriculum of an individual close and under the threshold and the curriculum of an individual close and over the threshold simultaneously. Just as it is unlikely, when it receives a curriculum from an individual close and under the threshold, that the firm will decide not to hire him/her and wait for the curriculum of an eligible individual. Hence, the absence of displacement effect isn't surprising as well. This last result is in line with Blundell et al. (2004) results and in contrast with Bucher (2010) study and Calmfors et al. (2002) literature review. Indeed, according to the second there was a displacement effect damaging short term unemployed. Nevertheless, her analysis was based on a theoretical model which strongly simplified reality. According to the third, the displacement effect of targeted employment subsidies had an average effect of 84%. Nevertheless, the value was based on a single study consisting in surveys to employers. If an employer answer "yes" when someone asked him/her whether she would prefer to hire a subsidized individual with respect to an unsubsidized individual with similar characteristics, this doesn't mean if he/she actually receive the curriculum of the unsubsidized he/she will reject it and wait for another curriculum by an eligible with similar characteristics. The result on the absence of post-poned hiring effect, instead, is consistent with previous literature and, in particular, with Boockmann et al. (2007) and Schünemann et al. (2013) conclusions (even though in their applications policy effect was not significant either).

3.2. *Should Targeted or Untargeted Policies be Preferred?*

Little is known about the difference between subsidies targeting a restricted group of unemployed workers and generalized subsidies without a definite target. To overcome this lack of information we compared the impact of Law 407/90 with the impact of Law 190, a generalized policy, implemented immediately after the first policy ending. In particular, we exploited the fact that the two laws following one another have similar characteristics but only one of them targeted the restricted group of LTU. The study of Law 190 from a LTU perspective can also give an overview on the path Italy is following with respect to ALMPs affecting this category of individuals (given that the last substituted Law 407/90).

3.2.1. Law 407/90 and Law 190: a qualitative comparison

On December 31st 2014, Law 407/90 ended, as a consequence of 2015 Legge Stabilità, promulgated on December 23rd 2014. The generalized incentives provided by the latter, consisted in tax credits to firms hiring unemployed people with a permanent contract or turning a temporary contract into a permanent ones. Tax credits corresponded to 100% of the contributions the firm had to pay to the social security service for each individual. To avoid the temptation, for the employers, to fire and then hire their employees again, firms could not obtain the tax credit hiring individuals who had been fired from a permanent contract less than six months before. The policy, implemented from January 1st 2015, lasted one year[3]. The adoption of this policy raised a huge debate on its effectiveness. Indeed, its implementation was used as a political weapon by all the political fronts. The policy was widely used. According to INPS' data, 1'078'885 were the new permanent contracts stipulated taking advantage from it[4]. In figure 3.2, the share of individuals hired with slightly less than 24 months of unemployment in the period from January 2010 to December 2015 is represented.

[3]It was followed by an equivalent policy with tax credits corresponding to 50% of labour taxes.
[4]Source: Osservatorio Precariato, INPS

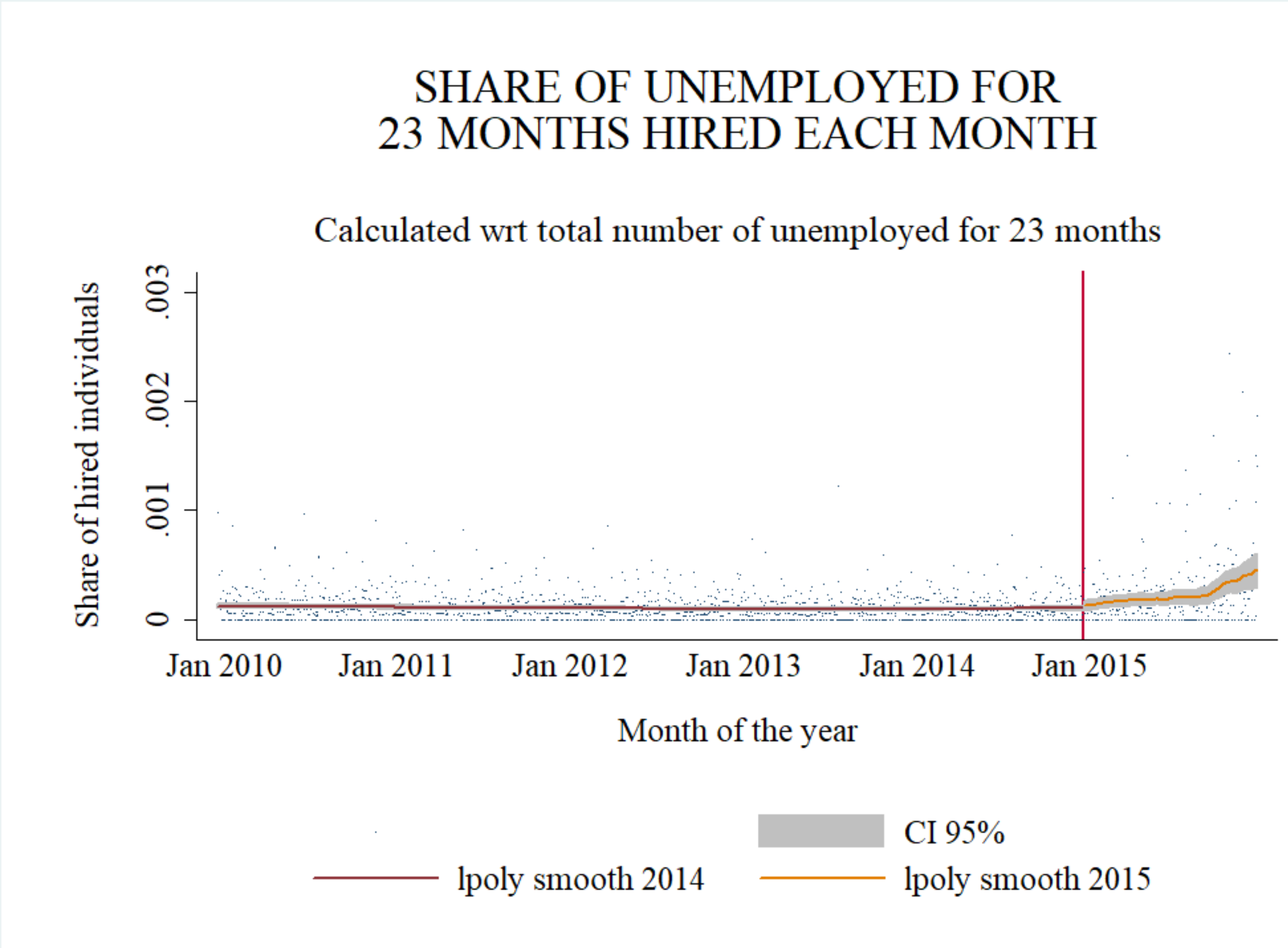


Figure 3.2: The figure represents the share of hired unemployed workers, with respect to time, over the period from January 2010 to December 2015. The variable was based on CICO database. Only individuals with slightly less than 24 months of unemployment were included.

It is possible to see that average hiring in 2015, hence under Law 190 incentives, are higher than in 2014. Nevertheless, year 2015 is characterized by a high variability (as is visible from the wider confidence interval), hence there is not a significant jump from December 2014 to January 2015. Moreover, the biggest increase in hiring is centered in the last months of the year. It is crucial to consider this aspect in the analysis for two different reasons. First of all, the further we get from the threshold, the less reliable the comparison is. Second, it suggests that the biggest effect is due to the fact that the policy was implemented for a limited and short period of time and its duration was communicated in advance. To compare the two types of incentives in a more general way, we estimated the impact of the untargeted incentives excluding the observation about December 2015. Indeed, we are interested in the effect of an untargeted policy, not on the effect of an untargeted policy whose adoption was limited in time. Another important information that can be derived from this graph is that the outcome is linear with respect to time before 2015, while it is quadratic in 2015. We will use this information during the model choice.

The main difference between this policy and Law 407/90 is that the latter targeted LTU, while the former was extended to all unemployed people. This element, and the proximity of the two policies in time, allow us to make general conclusions on the difference between targeted and untargeted incentives from the comparison between their impacts.

Qualitatively, the main difference between targeted and untargeted incentives is that the first lower labor costs in relative terms, while the second lower them only in absolute terms. If targeted incentives are implemented the LTU become economically more convenient than STU. This is not the case with untargeted incentives. The idea behind targeted incentives is that the lower labor costs of the targeted group is enough to counteract its lower desirability. Empirical evidence suggests the hiring behavior of employers in presence and in absence of incentives is heterogeneous. According to it, we can imagine to divide employers into three different groups. In the first group there will be employers for whom the incentive is not enough to counteract the lower desirability of LTU. Their existence is proved by the fact that, under the policy, LTU hiring is still lower than STU. This group of employers, will not react differently to the two different types of incentives. The second group of employers includes those who do not consider LTU as less desirable. The existence of this group is proved by the fact that LTU hiring is higher than zero even in absence of incentives. Like the first group, the employers will not react differently to the two policies. In both cases, the first category of employers will not hire LTU and the second category will hire LTU considering only their lower labor cost. The eventual presence of a third group of employers will make the difference between the two incentive types. There may be a group of employers for whom the LTU have a lower desirability and the relatively lower labor cost under Law 407/90 is enough to counteract it. If this group exists, its hiring will be higher under Law 407/90 than under 2015 Legge Stabilità incentives. Indeed, under the first policy the lower desirability of LTU is counteracted by the relatively lower labor cost. Under the second policy, instead, the labor cost of STU is also lowered, hence in relative terms the labor cost of LTU does not change.

3.2.2. Literature Review

Being the incentives of Law 190 a highly debated policy, there is a significant amount of literature estimating their impact. Among them, Centra and Gualtieri (2016) used a difference-in-differences model to determine the impact of the policy on the incidence of permanent contracts among the total. The authors exploited the fact that, individuals whose previous permanent contract ended less than six months before, were not eligible for the incentive. They added some covariates to the model to make the parallelism assumption more plausible. Moreover, they estimated the outcome of the control group in 2015 using interrupted time series analysis to clean the estimation from a possible displacement effect of the policy. They found a positive and significant effect. Similar results were found by Sestito and Viviano (2016). The authors estimated the impact of the same policy on different outcomes, among which the probability (both for employed and for unemployed workers) to find a permanent job. They used a difference-in-differences model with individual, monthly and annual fixed effects. None of these studies considered 2015 Legge Stabilità from an LTU perspective.

To the best of our knowledge, there are no empirical studies comparing targeted and generalized subsidies. Nevertheless, Brown (2015) did a meta-analysis trying to identify the positive and negative consequences of both types of policies. The author underlined how targeted hiring subsidies have lower deadweight costs[5] with respect to untargeted ones. Nevertheless, he stated that targeted subsidies may have negative consequences at a macroeconomic level, namely, the negative side effects we checked for. Once verified that those effects do not exist (or are negligible) in this context, it remains to check whether the generalized hiring subsidies are more efficient from a long-term unemployed perspective.

3.2.3. Data and Empirical Strategy

We employed the same database used in chapter 2 (see 2.3.1 and 2.3.2). Nonetheless, data were seasonally adjusted using a moving average approach (the results do not change if other seasonal adjustment methods, such as monthly dummies, are applied).

3.2.3.1. Interrupted Time Series Analysis

What we said before about the impossibility to use some of the methods usually applied in wage subsidies impact estimation is still valid (see section 2.3.3). Moreover, the fact that we want to estimate the impact of the policy from a LTU perspective, means we can't use a regression discontinuity design. Indeed, the last has a low external validity. We have two possible choices left. The first is to use a DID model. This arises two issues: the control group (individuals whose last permanent contract ended less than six months before) is significantly different from our target group. Moreover, the control group may be penalized by displacement and post-poned hiring effects. Hence, its value of the outcome could be a bad representation of what would have happened to them in absence of the policy. Following Centra and Gualtieri (2016), we could solve this issue using an interrupted time-series analysis to estimate the value of the outcome for the control group. Nevertheless, the second methodological choice is to directly use an interrupted time-series analysis. The second choice would save us many steps and it would allow us to use a simpler model.

[5]With this term the author identifies the costs of subsidies benefitting individuals who would have been hired indipendently from policy implementation.

Hence, we decided to use an interrupted time-series analysis. We exploited the January 1st 2015 threshold defining policy start. We could not estimate the impact of the policy on the exact same group targeted by Law 407/90. Indeed, we would not be able to distinguish between the effect of Law 407/90 ending and the effect of the generalized incentives start. Hence, we estimated the intention-to-treatment effect on the control group selected according to the modified Cattaneo et al. (2015) method (see 2.3.3) and assumed it not to differ significantly from our target group. We required the treated and control groups to be balanced at a monthly level. Hence, we used the paired t-test pairing units belonging to the same month and year (nevertheless, the results do not change in a significant way if we do the same analysis on the control group used in 2.3.3). The assumption that the two groups do not differ significantly is necessary to compare the intention-to-treatment effects of the two policies. As mentioned before, in order to estimate the impact of untargeted incentives generally, rather than the impact of Law 190 incentives themselves, we excluded observations about December 2015. Indeed, the peak corresponding to those observations is probably due to a characteristic specifically of this policy: its short duration. We want more general results, representing the impact of a generic untargeted policy and being comparable with Law 407/90. As suggested by the graphical analysis (see paragraph 3.2.1), we applied the following regression model:

$$y_k = \alpha + \gamma_1 T_k + \gamma_2 T_k^2 + \gamma_3 P_k + \gamma_4 P_k * T_k + \gamma_5 P_k * T_k^2 + \epsilon_k \tag{3.3}$$

We aggregated the units, with values of the forcing variable inside the selected bandwidth, at a daily level. In the models, T_k is the time variable and P_k is a dummy variable for the period after December 2014. The rest of the notation is as before. The intention-to-treatment effect of the policy is given by coefficient γ_3.

3.2.3.2. Needed Assumptions

The use of models based on time discontinuity requires the following assumption to be satisfied:

Assumption 6. *There is no anticipation effect.*

If policy effect is anticipated with respect to policy implementation starting time, the latter cannot be used as threshold in the discontinuity analysis. Luckily, the policy was announced on December 23rd 2014, few weeks before its implementation starting time. It is therefore unlikely there was an anticipation effect.
Soumerai and Ross-Degnan (2002) suggested another assumption which have to be satisfied when models based on time discontinuity are used:

Assumption 7. *There are no shocks happening at the policy implementation time and having a significant impact on the outcome, other than the policy itself.*

Possible sources of violation for this condition are an exogenous change in overall economic conditions, the implementation, in the same period, of other policies, or a possible change in the composition of the study population. Considering the first source of violation, in our framework, an increase in hiring may follow, as an example, an improvement in overall economic conditions. To verify whether this is the case, we added separately to the regression three exogenous variables that can proxy the economic conditions of the country and observed that the intention-to-treatment effect estimation did not change (suggesting the assumption holds). The first two were the annual final consumptions of non-resident families in Italy both in current and one-year lagged values. The third was the one-year lagged quarterly value of GDP. The use of the lagged value is required for the variable to be exogenous and it is justified by the fact that the improvement in the economic conditions usually have a delayed effect on hiring (Centra and Gualtieri 2016). We used ISTAT data on GDP and consumptions[6].
The second possible source of violation comes from the fact that in 2015 the Jobs Act was implemented. Starting from March 2015 this law brought significant changes to permanent contract rules, relaxing the constraints a permanent contract implied on employers' side. To verify whether this led to an overestimation of Law 190 intention-to-treatment effect, we re-estimated the model excluding observations from March 2015 on.

A third possible source of violation could be a change in the composition of population study. To verify whether that was the case, we included in the regression

[6]Data Sources: http://dati.istat.it/

model as covariates the shares of individuals with different educational qualification and employment sector and observed that the ITT estimation did not change significantly. Another possible source of bias is a change in the share of unemployed inside non-occupied group happening at policy implementation time. To check for it, we included in the regression model, a variable measuring the yearly rate of unemployed among non-CICO-detected individuals and observed that ITT estimation did not change. The variable was built using RTFL data.

The comparison of the two policies is a good representation of the differences between targeted and untargeted incentives only if the following assumption is satisfied:

Assumption 8. *There are no relevant differences between the two policies, other than in their target groups.*

There are three possible sources of violation of this assumption. Law 407/90 gave 100% tax credit only to firms belonging to the Mezzogiorno area or to artisan firms, while the 2015 Legge Stabilità gave it to all enterprises. To verify whether this difference was relevant, we repeated the analysis using only data on the Mezzogiorno area (where both Law 407/90 and Law 190 gave 100% tax credits). Legge Stabilità 2015 incentives covered the contributions due to the social security service only, while Law 407/90 covered both the contributions due to the social service and those due to the institution providing work insurance (the latter are significantly smaller[7]). Given that the rules for the determination of benefit amount are well known and that we had information about employees' wages, we were able to check whether this difference was relevant. In particular, we estimated approximatively[8] the tax credit amount for each individual hired under Law 407/90, under the two policies, and compared the two values. Finally, the awareness of firms about the existence of the policies may have induced a different take-up rate. Nevertheless, the wide use of

[7]The percentage of wage payed to the social security service is, on average, 29.8%. The percentage of wage payed to the insitution providing work insurance is, on average 2.9%

[8]Where it is approximatively because the rate of wage the taxes have to be payed on, depends on the type of work and the sector. Hence, we used an average value.

Law 407/90 and the attention given by the media to 2015 Legge Stabilità suggests both policies were well known by Italian firms.

The results for assumptions checking are presented in section 3.2.4.1.

3.2.4. Results

Table 3.3 reports the results of the baseline estimations. It is possible to see that the intention-to-treatment effect of the incentives is non-significant. The estimated impact is higher than the ones estimated in the placebo tests (see section 3.2.4.1). Nevertheless, the high variability of the outcome in 2015, the fact that the increase in the share of hiring is centered far from the threshold (hence other events influencing the outcome may have occurred biasing the results) and the fact that using other seasonal adjustment methods this difference disappears suggest the policy did not have any relevant effect on the hiring of this particular group. We can conclude the effect of these incentives is significantly smaller than those of Law 407/90. As said before, this difference is mainly attributable to the fact that Law 407/90 incentives are targeted and Law 190 incentives are not. This result suggests the impact of targeted incentives is entirely due to the lower labor cost of the targeted group, which counteract its lower desirability. Consequently, when a generalized policy is implemented, it should be accompanied by additional incentives tailored around the vulnerable groups of unemployed workers in order for them to be affected by it.

To provide additional information on the mechanism behind the policy and on Law 190 incentives impact, we repeated the analysis including observations about December 2015 (the results are reported in appendix .1). The policy had a positive and significant impact. This suggests a significant component of the impact of Law 190 incentives is probably due to the limited duration of the policy itself. In particular, the whole effect of the incentives can be attributed to the December peak in hiring. A policy having a significative impact only when it starts or ends is not new in this literature (Blundell et al. 2004).

Table 3.3: ESTIMATION OF THE INTENTION TO TREATMENT EFFECT OF GENERALIZED INCENTIVES ON THE OUTCOME OF THE LONG-TERM UNEMPLOYED GROUP

VARIABLES	Coefficients
Time	-0.00498**
	(0.00214)
Time^2	0*
	(0)
Treat	739
	(618)
Treat*Time	-0.783
	(0.628)
Treat*Time^2	0.000209
	(0.000160)
Constant	12.2***
	(0.930)
Observations	2160
R-squared	0.075

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NOTE: Generalized incentives ITT on vulnerable group is given by the coefficient of variable "Treat". Observations about December 2015 excluded. We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. For easier reading, coefficients and standard errors were multiplied by 100,000

3.2.4.1. Model Assumptions and Robustness Checks

In the table 3.4, we present the results of the checks for the assumption of no other exogenous shocks happening at policy implementation time and having a relevant impact on the outcome. Model 1 presents the results of the standard model used in the analysis. In model 2, 3 and 4 respectively, lagged GDP, foreign consumption and lagged foreign consumption levels are added as covariates to check for improvements in overall economic conditions. The percentage of unemployed among non-CICO detected units and the share of individuals in different educational level and in different sectors were added, respectively, in model 5 and 7 to check for changements in the composition of population study. In model 6 the observations in correspondence with Jobs Act implementations months were excluded to verify whether the detected effect was due to its implementation.

Table 3.4: ROBUSTNESS CHECK OF GENERALIZED INCENTIVES ITT ESTIMATION WITH RESPECT TO OTHER POSSIBLE EXOGENOUS SHOCKS

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Treat	739	711	735	737	737	-9920	975
	(618)	(618)	(618)	(618)	(618)	(25400)	(634)
Observations	2,160	2,160	2,160	2,160	1,795	1,885	2,160
R-squared	0.075	0.075	0.075	0.075	0.086	0.008	0.107

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

NOTE Model 1: Standard. Model 2: Lagged GDP addition. Model 3: Foreign Consumption addition. Model 4: Lagged Foreign Consumption addition. Model 5: Percentage of unemployed among non-CICO detected addition. Model 6: Exclusion of observations relative to Jobs Act implementation months. Model 7: Educational Level and Sector variable addition. All monthly or daily variables were seasonally adjusted using a moving average method. We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

The estimation is robust to addition of all the covariates in models 2-5 and in model 7. Model 6 estimation of the intention-to-treatment effect, instead, is higher than those obtained excluding December 2015 only (and it is still insignificant). This suggests Jobs Act impact was too small to affect the estimation of Law 190 incentives ITT. This result is in line with Sestito and Viviano (2016) study. From it, it emerged that the effect of Jobs Act changes to permanent contracts rules was negligible with respect to the effect of Law 190. All these results confirm that the assumption of no other relevant exogenous shocks happening at policy implementation time is satisfied.

In table 3.5 we conduce some placebo tests standard in interrupted time series analysis.

Table 3.5: PLACEBO TESTS FOR GENERALIZED INCENTIVES ITT ESTIMATION

VARIABLES	1/1/2014	31/12/2013	1/1/2013	31/12/2012	1/1/2012
Time	-0.00401	-0.00406	-0.00176	-0.00172	-0.00324
	(0.00306)	(0.00306)	(0.00477)	(0.00478)	(0.00944)
Time^2	0	0	-0	-0	0
	(0)	(0)	(0)	(0)	(1.18e-5)
Treat	40.7	32.3	-0.293	0.00329	4.71
	(144)	(143)	(21.6)	(21.5)	(5.80)
Treat*Time	-0.0474	-0.0374	-0.00245	-0.00287	-0.00887
	(0.176)	(0.175)	(0.0300)	(0.0299)	(0.0130)
Treat*Time^2	1.41e-5	1.11e-5	0	0	0
	(5.34e-5)	(5.31e-5)	(1.08e-5)	(1.08e-5)	(1.23e-5)
Constant	12.1***	12.1***	11.7***	11.7***	11.9***
	(1.06)	(1.06)	(1.24)	(1.24)	(1.59)
Observations	1,826	1,826	1,826	1,826	1,826
R-squared	0.006	0.006	0.006	0.006	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NOTE Placebo tests. All monthly or daily variables were seasonally adjusted using a moving average method. We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

With regard to the assumptions necessary for policies comparison, we start presenting the results of the analysis reduced to the group of individuals having the last working experience in a region of the Mezzogiorno Area. In table 3.6 and table 3.7, the results of the estimations of intention-to-treatment effect of, respectively, Law 407/90 and Law 190 in that area are reported together with their values under different robustness checks.

Table 3.6: LAW 407/90 ITT ESTIMATION AND ROBUSTNESS CHECK FOR THE *MEZZOGIORNO* AREA

	(1)	(2)	(3)	(4)	(5)	(6)
THRES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
24	4.72***	4.30***	2.73***	1.99		
	(0.901)	(0.912)	(1.02)	(1.55)		
22					−0.464	
					(0.810)	
26						−0.900
						(0.891)
Obs	33,603	33,580	27,390	12,782	43,378	40,595
R-sq	0.068	0.071	0.093	0.192	0.059	0.078

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

NOTE Model 1: Standard. Model 2: Covariates addition. Model 3: Different bandwidth [722:736]. Model 4: Different bandwidth [726:732]. Model 5 and 6: Placebo Tests. We implemented all weighted regressions, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

Table 3.7: GENERALIZED INCENTIVES ITT ON VULNERABLE GROUP, ESTIMATION AND ROBUSTNESS CHECKS FOR THE *MEZZOGIORNO* AREA

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Treat	991	929	988	993	248	
	(687)	(688)	(688)	(688)	(628)	
Constant	13.7***	-714	-123	17.9***	6.86***	1.33***
	(1.17)	(576)	(301)	(6.30)	(2.36)	(1.32)
Observations	2,160	2,160	2,160	1,795	2,160	1,826
R-squared	0.076	0.077	0.076	0.088	0.090	0.003

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

NOTE Model 1: Standard. Model 2: Lagged National GDP addition. Model 3: Foreign Consumption of National products addition. Model 4: Percentage, at national level, of unemployed individuals among non-CICO detected addition. Model 5: Educational Level and Sector variables addition. Model 6: Placebo test with January 1st 2014 as threshold. All monthly or daily variables were seasonally adjusted using a moving average method. We implemented all weighted regressions, using the total number of individuals corresponding to each unit as weights. For easier reading, all coefficients and standard errors were multiplied by 100,000.

Assumptions and robustness checks suggest the models are valid. Both the intention-to-treatment effects are slightly higher in the Mezzogiorno area than in Italy as a whole. Nevertheless, the Mezzogiorno area estimations present a high volatility, hence the coefficients are not significantly different from those of our baseline estimations. Moreover, the results in terms of juxtaposition between targeted and untargeted policies do not change.

The comparison between different benefit amounts under the two policies is reported in table 3.8.

Table 3.8: COMPARISON BETWEEN ESTIMATED AVERAGES OF SUBSIDIES AMOUNTS UNDER THE TWO POLICIES

Year	Avg 407	Avg 190	Diff /Avg 190
2010	7023	5726	0.227
2011	5816	5479	0.061
2012	5846	5502	0.063
2013	5722	5378	0.064
2014	6821	6366	0.071

As expected, the difference between the average amount of subsidies given under each policy is negligible in most years. Indeed, the difference is lower than 7.2% of the smallest average almost all years, with the exception of 2010. Consequently, observations relative to that year have been excluded from the estimation of Law 407/90 intention-to-treatment effect.

Results of table 3.6, 3.7 and 3.8, suggest the assumption of no other relevant differences between the two policies is satisfied.

Part II

4. Identify more, observe less: Mediation Analysis Synthetic Control

The aim of this Chapter is twofold. To explore the literature on synthetic control method and mediation analysis and to introduce a new methodology allowing to do mediation analysis when there is selection on unobservables or the number of treated units is low.

4.1. *Synthetic Control Method*

Synthetic Control Method introduction has been defined by Athey and Imbens (2017) as "arguably the most important innovation in the policy evaluation literature in the last 15 years". This methodology can be used in frameworks were panel data are available and the intervention is implemented from a determined time period T. The idea behind SCM is to use information on pre-intervention period to build a synthetic value of the potential outcome of the treated unit in absence of treatment (Y^0_{1t} where 1 is the treated unit) in post-intervention period. This is done by re-weighting the post-treatment outcomes of control units (whose set is often called donor pool in this framework) by using weights that are chosen to minimize the distance between pre-intervention observable characteristics (including pre-intervention outcomes) of treated and control units. The weights to use are constrained such that Y^0_{it} lies in the convex hull of non-treated outcomes. Namely, it can be written as a linear combination of the latter. To introduce the method more formally, and justify the use of a linear combination of control units to build a synthetic control, we will report the same motivating example used in Abadie, Diamond, et al. (2010). Assume to be interested in the effect of an intervention D, implemented from time T, on an outcome Y. To see this, let D_{it} be a binary indicator which is equal to one if unit i was exposed to the intervention at time t. Assume that we observe J units ordered such that units 1 to n are treated while units $n+1$ to J are not. Without loss of generality, we will first consider the first treated, unit 1 (in

SCM the parameters of interest are normally defined with respect to a single unit). The effect of interest for unit 1 is the total effect α_{1t} in the post-intervention period ($t \geq T$), defined as

$$\alpha_{1t} = Y^1_{1t} - Y^0_{1t},$$

where at time $t > T$ only the outcome Y^1_{1t} is observed. Hence, to identify the total effect α_{1t}, the outcome Y^0_{1t} has to be estimated.

Assume the outcome of unit i at time t in absence of treatment can be modelized as:

$$Y^0_{it} = \zeta_t + \eta_t X_i + \lambda_t \mu_i + \epsilon_{it} \tag{4.1}$$

where ζ_t is an unknown common factor with constant factor loadings across units, X_i is a $(r \times 1)$ vector of observed covariates, η_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, ϵ_{it} are unobserved transitory shocks.

Consider a $(1 \times (J - n))$ vector of positive and adding up to 1 weights $L = (l_{n+1}, ..., l_J)$. Different values are possible for L and to each of them it is associated a potential synthetic control whose outcome is given by:

$$\hat{Y}^0_{1t} = \sum_{i=n+1}^{J} l_i Y^0_{it} = \zeta_t + \eta_t \sum_{i=n+1}^{J} l_i X_i + \lambda_t \sum_{i=n+1}^{J} l_i \mu_i + \sum_{i=n+1}^{J} l_i \epsilon_{it}. \tag{4.2}$$

Now suppose it exists a vector $L^* = (l^*_{n+1}, ..., l^*_J)$ such that $\forall\ t = 1, ..., T - 1$, it satisfies

$$\sum_{j=n+1}^{J} l^*_j Y_{jt} = Y_{1t},$$

$$\sum_{j=n+1}^{J} l^*_j X_j = X_1.$$

Abadie, Diamond, et al. (2010) demonstrated that, under these and other proper assumptions (we refer to that work for more details, even though the same assumptions are used in appendix .2 for our demonstration), the average of the outcome of the synthetic control is given by Y^0_{1t} plus a bias which goes to zero as the number of pre-treatment periods goes to infinity. More formally:

$$E(\hat{Y}^{0,1}_{1t'}) = Y^{0,1}_{1t'} + o(T).$$

Consequently, estimating the total effect as $\hat{\alpha}_{1t} = Y_{1t} - \hat{Y}^0_{1t}$ is justified by the fact that

$$\lim_{T\to\infty} E(\hat{\alpha}_{1t}) = \alpha_{1t} \ \forall \ t \geq T \tag{4.3}$$

This justifies choosing the weights that minimize the distance between the observable characteristics of the treated and the one of the control units in pre-treatment period. Notice that the assumption on the model underlying potential outcome in (4.1) has been introduced to simplify the afore-mentioned demonstration and it is more stringent than needed. For the method to give valid results the following assumption (defined for the first time in this dissertation) would be enough:

Assumption 1. [Common Time-varying Shocks Assumption (CTSA)]
All relevant time-varying shocks influencing the outcome are common between treated unit and donor pool.

If CTSA is satisfied, the outcomes of the units in the donor pool are determined by the same observables and unobservables of the outcome of the treated unit. The unobservables are given by the difference between the outcome and the observables. Hence, imposing constraints on the last two in pre-treatment period, indirectly constrains the former. Under CTSA, the equivalence of observable and unobservable variables in pre-treatment period guarantees the equivalence of those variables in post-treatment period as well. Indeed, if the assumption is satisfied, an eventual relevant shock would be common to the two units. This assumption is "contained" in model specification (4.1). Indeed, in this specification, the outcome depends on some individuals characteristics constant in time, some characteristics varying in time but common to all units and random errors which change both in time and among units but are irrelevant and have zero mean. This is somehow the same concept behind matching on pre-treatment outcomes (Lechner 2015).

This identification strategy was widely used in empirical literature after its implementation. Moreover, it was widely studied in theoretical literature. Several extensions have been proposed. Kreif et al. (2016) proposed a method to apply SCM to frameworks with multiple treated units. Doudchenko and Imbens (2017) proposed a method that allows the weights not to sum to one and to take negative values. They propose to select weigths using a LASSO with an elastic-net type penalty to substitute these two assumptions with the requirement of weights as smaller as possible, and with a small number of them which is non-zero. This method allows to obtain the equality condition, and consequently the synthetic control, in a wider number of cases. Nonetheless, it should be used with caution. Indeed, looking for the synthetic control outside the convex hull of the donor pool allows for the use of less similar control units, increasing the probability that CTSA is violated.

Looking to extensions that go further away from standard SCM, Xu (2017) introduced the generalized synthetic control method. This method can be used in frameworks with multiple treated units and it provides uncertainty estimates valid under frequentist conception. Simulations showed this method to be more efficient than SCM (when its assumptions are satisfied). Nonetheless it relies on stronger functional form assumptions. Indeed, while the modelization of the outcome can be relaxed in SCM application, the same can't be done with the generalized synthetic control method.

To extend SCM, Athey, Bayati, et al. (2017) proposed a general methodology, which approximates methods based on unconfoundedness if there are few time periods and many units, and approximates methods based on synthetic control approach if there are many time periods and few units. The method was developed starting from interactive fixed effects literature.

A method very similar to the SCM has been proposed in Hsiao et al. (2012). It relies on model (4.2) and it can be used in the same frameworks where SCM is applied. As in SCM a synthetic control is built to mimic treated unit potential outcome in absence of treatment. The outcome of the treated unit is regressed on the outcomes of the controls for all pre-treatment period to determine the weights to use. Using a standard regression the risk of extrapolating results and creating a conterfactual units even when the treated unit is far from the convex hull of the donor pool is higher (Abadie, Diamond, et al. 2015). Gardeazabal and Vega-Bayo (2017) compared this method with the SCM and showed that the last is more robust to changes in the donor pool and it performs better when the number of pre-treatment period is high enough. From their simulations it also emerged that the SCM is severely biased when the model constraints are not properly satisfied. The same results was found in Gobillon and Magnac (2016), where the authors compared the SCM with interactive fixed effects models and concluded the first performed very well only under the satisfaction of model constraints.

Although synthetic control method (and its extensions) is very well suited for estimating the total effect of an intervention, it does not take into account the fact that it might exist some intermediate variables (namely, the mediators) that lie in the causal path between the treatment and the outcome of interest, such that the knowledge of the total effect might not be enough for policy conclusions. In the presence of one or more mediators, the total effect can be generally decomposed into a direct or net effect of the treatment on the outcome of interest and some indirect effects, generated through the mediators. Policy conclusions that ignore the presence of such intermediate outcomes, might be misleading. For example, one may find a zero or even negative total effect of the intervention, even though its direct effect is positive. Moreover, it is often important to quantify the indirect effects to better target the intervention. Consider the huge decrease in tobacco consumption after the introduction of California's anti-tobacco law, Proposition 99, estimated in Abadie, Diamond, et al. 2010. Proposition 99 not only increased tobacco price but also introduced several anti-tobacco informational campaigns. It would be extremely relevant to know how much of the decrease in tobacco consumption triggered by Proposition 99 is due to the increase in prices and how much of it is due to investments in informational campaigns. Both effects can be thought as indirect effects of Proposition 99. In paragraph 4.3 we will show that the idea behind SCM can be applied to identify the direct effect and the indirect effect. The approach we will present to extend SCM to do mediation analysis can be easily applied to extend the methodologies developed in Athey, Bayati, et al. (2017), Xu (2017), Kreif et al. (2016) and Doudchenko and Imbens (2017).

4.2. *Investigating the mechanism behind the effect: Mediation Analysis*

In mediation analysis, there is a variable (namely, the mediator) mediating the impact of a treatment on an outcome. The idea is that the treatment has an impact on the mediator which, changing, has an impact on the outcome. The mediator can be seen as an intermediate outcome. The choice of the two outcomes depends on the case under study. Their classification as intermediate or final should be based on researchers' knowledge and beliefs and on the literature (for and example see the next chapter). The path treatment-mediator-outcome is not the only causal pathway that exist between the treatment and the outcome. Goal of mediation analysis is to distinguish between the effect of the pathway of interest and those of the others. To make this distinction, the total effect is split into the amount of effect going through the mediator (indirect effect) and the residual effect (direct effect).

To the best of our knowledge, the concept of mediation was introduced for the first time in scientific literature in 1928. The psychologist Robert S. Woodworth developed the concept of stimulus-organism-response (SOR). More or less half a century later, the need for a scientific method to investigate on SOR mechanism was satisfied by Smith (1982). He proposed an experiment to investigate, in psychological frameworks, on the organism between the causal relation stimulus-response. Some years later, Baron and Kenny (1986) proposed to do mediation analysis applying a structural equation model. In particular, they proposed to do three different regressions. In the first, regressing the mediator on the treatment (we are adapting authors' terminology to our framework of interest). In the second, the outcome on the treatment. In the third, the outcome on both the treatment and the mediator. The authors claimed that if treatment effect is effectively mediated by the variable of interest then:

- The treatment would affect the mediator in the first equation.

- The treatment would affect outcome in the second equation.
- The mediator would affect the outcome in the third equation.
- The effect of the treatment on the outcome would be smaller in the third equation than in the second.

The direct effect should be given by treatment's coefficient in the third equation while the indirect effect should be given by the product between the coefficient of the treatment in the first equation and the coefficient of the mediator in the third. Even though it is very simple to use, this method has been widely criticized (Imai, Keele, and Yamamoto 2011, Pearl 2014, Imai, Keele, and Tingley 2010). The critics are mostly on the strong parametric and distributional assumptions needed and the ambiguity of direct and indirect effects interpretation in this framework.

From psychology, literature in other disciplines started to realize how useful mediation analysis could be. When other disciplines got interested in mediation analysis, the methodology improved significantly. To the best of our knowledge, the first to use Rubin's model was Holland (1988). In mediation analysis the main challenge using a counterfactual approach is to estimate the potential outcomes under treatment and mediator combinations that are never observed for any unit. To be more clear, we will now introduce formally the mediation analysis in counterfactual framework. We will use the notation needed in the following sections rather than authors' notation, to make the following sections easier to understand. For the same reason, we will add the time component even though none of the following authors used it.

Define treatment as D, the outcome of interest as Y and the mediator of interest as M. Then, the total effect of the intervention can be decomposed into an indirect effect which goes through M and the direct effect, which goes through other causal pathways. For each unit i, we can define the potential mediator at time t as follow:

$$M_{it}(d) \text{ for } d \in \{0, 1\}.$$

$M_{it}(d)$ is the value that the mediator of unit i would take, at time t, if D_{it} was set to d. Assuming that there are no anticipation effects on the mediator, in pre-intervention period, and that the standard Stable Unit Treatment Value Assumption (SUTVA) holds, the observed and the potential mediators are related through the following observation rule:

$$M_{it} = M_{it}(0)(1 - D_{it}) + M_{it}(1)D_{it}.$$

Note that M_{it} is always equal to $M_{it}(0)$ for both treated and control units in pre-intervention ($t < T$) period and we can only observe one of the two potential mediators in each period. Similarly, for each unit i at time t, we define the potential outcome as follow:

$$Y_{it}(d, M_{it}(d')) \equiv Y_{it}^{d,d'} \text{for } d, d' \in \{0, 1\}.$$

$Y_{it}^{d,d'}$ is the value that the outcome of unit i would take at time t if we set $D_{it} = d$ and $M_{it} = M_{it}(d')$. The potential outcome is a function of both the treatment and the potential mediator. Under SUTVA, and assuming no anticipation effects in pre-intervention period, the observed and the potential outcomes are related by the following observation rule:

$$Y_{it} = Y_{it}^{0,0}(1 - D_{it}) + Y_{it}^{1,1}D_{it}.$$

The effects of interest in mediation analysis are the total effect α_t, the direct effect $\theta_t(d)$, and the indirect effect $\delta_t(d)$ in the post-intervention period ($t \geq T$) that are defined as

$$\begin{aligned}
\alpha_t &= E(Y_{it}^{1,1} - Y_{it}^{0,0}), \\
\theta_t(d) &= E(Y_{it}^{1,d} - Y_{it}^{0,d}), \\
\delta_t(d) &= E(Y_{it}^{d,1} - Y_{it}^{d,0}), \quad t \geq T.
\end{aligned}$$

It is possible to see that the direct effect is defined keeping fixed the treatment status defining mediator value and changing the actual treatment status. The indirect effect, instead, is defined keeping fixed the actual treatment status and changing the treatment status defining the mediator. Notice that the indirect effect is written as a function of the treatment status defining the mediator to simplify notation. Nonetheless, its value depends on the potential value of the mediator depending on that treatment status. It is important to underline this aspect given that while the treatment is typically binary, the potential mediator can take more than two values. Differently from the standard setting only $Y_{it}^{0,0}$ and $Y_{it}^{1,1}$ are observed (and still only one of them for each unit in each period), while $Y_{it}^{0,1}$, $Y_{it}^{1,0}$ are never observed. This makes the identification of the direct and indirect effect more challenging.

Identification of the direct and indirect effects is particularly hard also because there may be selection bias on the mediator as well. Suppose we decide to estimate the indirect effect comparing a group of treated with another group of treated having mediator's value equals to the $M(0)$ of the first group (hence equals to the potential mediator of the first group in absence of treatment). The second group may have a different mediator under treatment because of some characteristics correlated with the outcome. In that case the estimated indirect effect would be biased because of selection on the mediator. Indeed, the comparison between groups with different values of the mediator, incorporates both the indirect effect and the different characteristics that are the cause of mediator differences. Mediator selection bias was noticed for the first time in Robins and Greenland (1992). The authors underlined as, treatment randomization was not enough to identify correctly all the parameters of interest in mediation analysis. The first to formalize this intuition, defining a set of assumptions needed to estimate the direct and indirect effects in observational frameworks was Pearl (2001). The author defined the direct effect as treatment effect holding all possible intermediates fixed. He said that the average direct effect can be identified under the following assumptions:

$$
\begin{aligned}
Y^{d,d'} &\perp\!\!\!\perp M(d'')|X, \\
P(Y^{d,d'} = y|X = x) &\text{ is identifiable,} \\
P(M(d'') = m|X = x) &\text{ is identifiable}
\end{aligned}
$$

for all values of d, d' and d''. Where the notation is as before. He showed as, under these conditions, the average direct effect under a value m of the mediator, is given by the difference between the average conditioned outcome for treated and the average conditioned outcome for the untreated times the conditional probability of the mediator being equal to m averaged with respect to all covariates values. He also showed that total effect can be written as the sum of the direct and the indirect effects, with treatment set to proper values. In particular, using today's notation, he showed that:

$$\begin{aligned}\alpha_t &= E(Y_{it}^{1,1} - Y_{it}^{0,0}),\\ &= E(Y_{it}^{1,1} - Y_{it}^{1,0}) + E(Y_{it}^{1,0} - Y_{it}^{0,0}),\\ &= \delta_t(1) + \theta_t(0),\end{aligned}$$

and

$$\begin{aligned}\alpha_t &= E(Y_{it}^{1,1} - Y_{it}^{0,0}),\\ &= E(Y_{it}^{1,1} - Y_{it}^{0,1}) + E(Y_{it}^{0,1} - Y_{it}^{0,0}),\\ &= \theta_t(1) + \delta_t(0).\end{aligned}$$

Thus, if α_t is identified, the identification of $\theta_t(1)$ and $\delta_t(1)$ automatically implies the identification of $\theta_t(0) = \alpha_t - \delta_t(1)$ and $\delta_t(0) = \alpha_t - \theta_t(1)$. An additional contribution to the literature was the introduction of the difference between natural and controlled direct effects. The former is the direct effect for a value of the mediator equal to the actual potential mediator value. The latter, instead, is the direct effect under a chosen, generic, value of the mediator m.

Thus far, the literature focused on two sets of interconnections among the variables. The first set is represented in figure 4.1, where the causal effect is depicted as an arrow going from the cause to the effect (notice that the direct effect is given by the blue arrow while the indirect effect is given by the green arrow).

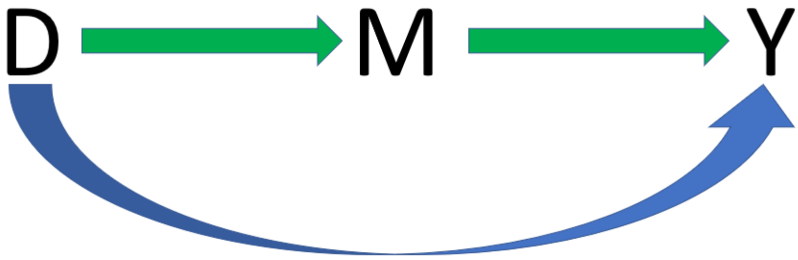


Figure 4.1: First set of interconnections: No Covariates, no Post-Treatment Confounders. M is the mediator. Y is the outcome. D is treatment status. Direct effect is given by the blue arrow. Indirect effect is given by the green arrow.

This structure of relations is unlikely to be found in non-experimental studies. Indeed, it relies on the assumption that there are no covariates correlated with the mediator and/or with treatment assignment and influencing outcome value. This is mostly the case of randomized experiments where there is the possibility to control both for treatment and for mediator values. The second set of interconnections, more common in non-experimental studies, can be represented graphically as in figure 4.2.

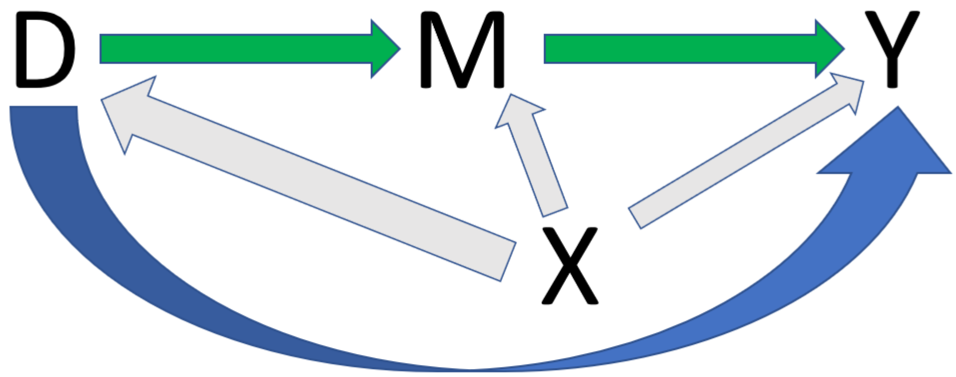


Figure 4.2: Second set of interconnections: Covariates, no Post-Treatment Confounders. M is the mediator. Y is the outcome. D is treatment status. X are the covariates. Direct effect is given by the blue arrow. Indirect effect is given by the green arrow.

Where X represents covariates influencing the outcome and correlated with either treatment assignment or mediator values or both. In this framework the identification of the parameters of interest is more challenging. Vanderweele and Vansteelandt (2009) noticed as this set of interconnections as well can be restrictive. In particular, they warned on the possible presence of post-treatment confounders, i.e. covariates influenced by the treatment and influencing the mediator and the outcome. Interconnections allowing for the presence of post-treatment confounders can be represented as in figure 4.3.

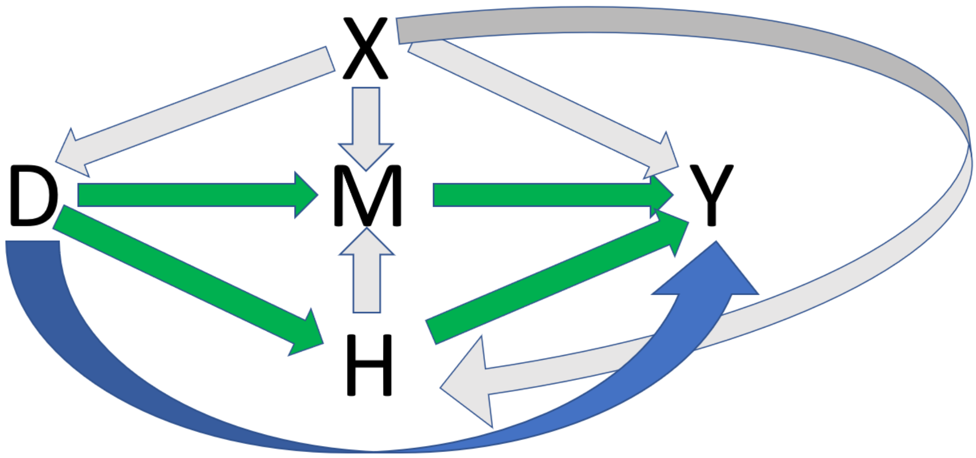


Figure 4.3: Third set of interconnections: Covariates and Post-Treatment Confounders. M is the mediator. *H* is the post-treatment confounder. Y is the outcome. D is treatment status. X are the covariates. Direct effect is given by the blue arrow. Indirect effect is given by the green arrow.

Where *H* represents post-treatment confounders. In presence of post-treatment confounders, the independence assumption can't be reached conditioning only on pre-treatment covariates influencing the outcome and the mediator. If their presence is not properly taken into account, the estimation will be biased. Different methods have been proposed to deal with the three different frameworks. In this literature review we will focus on observational methods only, therefore on the second and the third frameworks only. Among the methods proposed, few of them showed how to deal with the framework depicted in figure 4.3. All those methods exploited the fact that post-treatment confounders can be considered as other mediators.

To deal with the framework depicted in image 4.2, Imai, Keele, and Yamamoto (2010) proposed a "competitor" to the set of assumptions introduced by Pearl (2001). The authors modified the sequential ignorability assumption (a version of the CIA used with time-varying treatments) to use it for mediation analysis. They required the following assumption:

Assumption 2 (Sequential Ignorability (or Sequential Conditional Independence)).

$$\{Y_{it}^{d',m}, M_{it}(d)\} \perp\!\!\!\perp D_{it}|X_{it} = x,$$
$$Y_{it}^{d',m} \perp\!\!\!\perp M_{it}(d)|D_{it} = d, X_{it} = x$$

for $d, d' = 0, 1$, *and all* $x \in X$ *where it is also assumed that* $0 < Pr(D_{it}|X_{it} = x)$ *and* $0 < Pr(M_{it} = m|D_{it} = d, X_{it} = x)$ *for* $d = 0, 1$ *and all* $x \in X$ *and* $m \in M$.

The authors proved that under this assumption average natural direct and indirect effects can be identified non-parametrically as follow:

$$\bar{\delta}_t(d) = \int\int E(Y_{it}|M_{it} = m, D_{it} = d, X_{it} = x)\{dF_{M_{it}|D_{it}=1,X_{it}=x}(m) - dF_{M_{it}|D_{it}=0,X_{it}=x}(m)\}dF_{X_{it}}(x),$$
$$\bar{\theta}_t(d) = \int\int \{E(Y_{it}|M_{it} = m, D_{it} = 1, X_{it} = x) - E(Y_{it}|M_{it} = m, D_{it} = 0, X_{it} = x)\}dF_{M_{it}|D_{it}=d,X_{it}=x}(m)dF_{X_{it}}(x),$$

where $F_Z(.)$ and $F_{Z|Q}(.)$ represent the distribution function of a random variable Z and the conditional distribution function of Z given Q.

In mainstream mediation analysis, most of the approaches relies on SCIA[1].

Another identification strategy has been proposed in Huber et al. (2017). The authors showed how to identify the direct and the indirect effects on treated, under SCIA and common support assumption. Their identification strategy uses a radius propensity score matching with bias adjustment. The propensity score is built conditioning on the covariates and the mediator to identify the average direct effect on treated. While it is built conditioning on covariates only to estimate the average total effect on treated. The average indirect effect on treated is obtained by the difference

[1]SCIA is very similair to the assumptions set used in Pearl (2001). Nonetheless, the former is less restrictive (Imai, Keele, and Yamamoto 2010, Shpitser and VanderWeele 2011). Hence, in the following literature review we will focus on identification strategies that rely on SCIA even though others relying on Pearl's set exist.

between the second and the first.

Huber (2014) proposed a similar identification strategy relying on the same set of assumptions, based on inverse probability weighting. He proved that, under the two assumptions, the average direct effect is identified by:

$$\theta_t(d) = E\left[\left(\frac{Y_{it} * D_{it}}{Pr(D_{it} = 1|M_{it}, X_{it})}\right) * \frac{Pr(D_{it} = d|M_{it}, X_{it})}{Pr(D_{it} = d|X_{it})}\right]$$

Where $Pr(D_{it} = 1|M_{it}, X_{it})$ is a propensity score built on both M and the pre-treatment covariates and $Pr(D_{it} = 1|X_{it})$ is a propensity score matching built on pre-treatment covariates only. The idea is that weighting the observed outcomes with respect to propensity scores, the direct effect can be identified. The author proved that, under the same assumptions, the average indirect effect can be identified by:

$$\delta_t(d) = E\left[\frac{Y_{it} * \mathbb{I}\{D_{it} = d\}}{Pr(D_{it} = d|M_{it}, X_{it})} * \left(\frac{Pr(D_{it} = 1|M_{it}, X_{it})}{Pr(D_{it} = 1|X_{it})} - \frac{1 - Pr(D_{it} = 1|M_{it}, X_{it})}{1 - Pr(D_{it} = 1|X_{it})}\right)\right]$$

where $\mathbb{I}$ is the indicator function taking value 1 if its argument is true. These estimators depend on a propensity score which can be estimated either parametrically or non-parametrically. No other functional form assumptions are required. While this first identification strategy does not allow for the presence of post-treatment confounders the author proposed another method to make the identification in the framework depicted in figure 4.3. He considered each post-treatment confounder as an additional mediator. He defined (using a different notation), for each unit i, a second potential mediator $H_{it}(d)$ with the same characteristics of $M_{it}(d)$. He re-wrote the outcome and the first mediator as functions of the post-treatment confounders. Hence, they became $M_{it}(d, H_{it}(d'))$ and $Y_{it}(d, H_{it}(d'''), M_{it}(d', H_{it}(d''))) \equiv Y_{it}^{(d,d''',(d',d''))}$ where $d, d', d'', d''' = 0, 1$. The author defined, with respect to this new framework, the total indirect effect as (we use our notation):

$$\delta_t^{TOT}(d) = E\left[Y_{it}^{d,d,(1,1)} - Y_{it}^{d,d,(0,0)}\right]$$

The total indirect effect includes the impact of a variation of the mediator due a variation of both first and second mediator treatment assignment. This value has to be distinguished from the partial indirect effect, which includes the effect of a mediator variation due to a variation in its treatment assignment while keeping fixed that of the second mediator. More formally, the partial indirect effect is defined by the author as:

$$\delta_t^{PAR}(d) = E\left[Y_{it}^{d,d,(1,d)} - Y_{it}^{d,d,(0,d)}\right]$$

In accordance with the author, I believe in most empirical frameworks the total indirect effect is more relevant.

Huber defined the direct effect as the impact of treatment assignment variation keeping fixed all the mediators. This definition is not common to all authors dealing with post-treatment confounders. As an example, in Imai and Yamamoto (2013) the second covariate is allowed to vary. Huber direct effect is defined as:

$$\theta_t(d)^{Hub} = E\left[Y_{it}^{1,d,(d,d)} - Y_{it}^{0,d,(d,d)}\right]$$

The author showed how to identify the average direct effect and the partial indirect effect under the assumptions of conditional independence of the treatment, conditional independence of the mediator and common support (we refer to the author for a complete and formal representation of the assumptions). These set of assumptions can be considered a version of the SCIA adapted to the presence of post-treatment confounders. To estimate the total indirect effect, instead, he did an additional assumption on the functional form of average potential outcome-mediator relation. This identification strategy model was extended in Hsu et al. (2017), where the authors proposed a completely non-parametric method.

The challenge of direct and indirect effects estimation in presence of post-treatment confounders, was taken up as well in Imai and Yamamoto (2013). As Huber, the authors wrote the first mediator and the outcome as functions of the second mediator. They defined the indirect effect as Huber's total indirect effect while their definition of the direct effect was:

$$\theta_t(d)^{Ima} = E\left[Y_{it}^{1,1,(d,d)} - Y_{it}^{0,0,(d,d)}\right].$$

The idea behind this definition is that the direct effect is defined as direct only with respect to the mediator of interest, i.e. the first mediator. This is consistent with the definition of direct effect normally used in mediation analysis. Moreover, this definition allows to write the total effect as a sum between the direct and the total indirect effects (as the authors showed). The authors showed how to identify these effects under a modified version of the sequential conditional ignorability assumption. The new assumption is very similair to Huber's. In order to estimate the indirect effect non-parametrically, they made the assumption of no interaction between treatment and mediator for any unit. This means the direct effect has to be constant with respect to mediator value. Nonetheless, they propose a semi-parametric model to relax this assumption.

Most of the models proposed to do mediation analysis relies on SCIA, or on similar assumptions, requiring conditional unconfoundedness (i.e. Pearl's set of assumptions). Nonetheless, there are many frameworks where SCIA does not hold. To deal with this possibility, different approaches have been suggested in past literature. Many of them consist in sensitivity tests to verify how biased the estimation can become in presence of a SCIA violation (see, for example Imai, Keele, and Yamamoto 2010, Imai, Keele, and Yamamoto 2011, Vanderweele and Chiba 2014). Unless the analysis is robust enough to a SCIA violation this solution is not very satisfying. Gallop et al. (2009) and Small (2012) proposed two different methodologies that allows to make point identification under SCIA violation. Nonetheless, both of them require the treatment to be randomized. Another method has been proposed in Zheng and Zhou (2017) for survival models frameworks. The method is based on a softer version of the sequential ignorability assumption. Nonetheless, it can only be applied in survival analysis frameworks and it requires strong functional form assumptions and the CIA to hold.

To the best of our knowledge, there are only two non-experimental methods for mediation analysis that can be used in case of SCIA (or similar assumptions) violation and allows a point identification of the parameters of interest. The first approach relies on the use of Instrumental Variables. Instrumental Variables approach has been extended to do mediation analysis in Frölich and Huber (2017). This approach is used when treatment and mediator are still endogenous after conditioning on observables. The method is based on the use of two different instruments. A first instrument (which has to be binary) to solve treatment endogeneity and a second (which can be either discrete or continuous) to solve mediator endogeneity.

The two authors imagined a structural model given by a system of non-separable non-parametric equations where the outcome is a function of the treatment and some observable (X) and unobservable (U) covariates. The mediator is a function of the treatment, the second instrument and observable (X) and unobservable (V) covariates. Finally, treatment assignment depends on a function of the first instrument and observable (X) and unobservable (W) covariates. In this framework, there is endogeneity if there is correlation among the groups of unobservable variables U, V, W. As in standard IVs framework, the total, direct and indirect natural effects are estimated with respect to the group of treatment compliers. The authors show how to estimate these parameters under two main assumptions:

- The two instruments have to be independent from unobservables U, V conditioning on observables X and group of belonging.
- The first instrument has to be independent from unobservables U, V and the group of belonging conditioning on the second instrument and the observables.

For the identification of some of the parameters of interest, an additional assumption is required: the independence between the two instruments given the observables. As in standard IVs frameworks, the identification also requires the assumption of absence of defiers and, in cases where the mediator and the second instrument are continuous, the weak monotonicty assumption on the mediator. The intuitive idea behind their identification method, is that the first instrument can be used to control for treatment assignment while the second instrument is simultaneously used to counteract the impact of the first on the mediator. Using this approach, it is possible to change treatment assignment while mediator value is kept fixed. The authors showed as well how to use this approach to identify the controlled direct effect. Moreover, they allowed for the presence of post-treatment confounders when it does not violate the assumptions on the instruments. Even though this approach is ideal in cases of SCIA violation it can be applied only in a restricted amount of frameworks. Indeed, it is often hard, and sometimes impossible, to find proper instruments satisfying the independence assumptions.

The second approach which does not rely on SCIA has been proposed in Deuchert et al. (2018) and it is an extension of a DID model to do mediation analysis. The model has been proposed for frameworks where treatment is randomized (or the CIA is plausible) and the mediator is a binary variable. The authors divided the population of interest in groups of compliers, always takers, never takers and defiers, defined with respect to the combination of treatment and mediator values as in IVs approach. The intuition behind this method is that, imposing common trend assumptions between two different groups, the outcome of the first can be used as a counterfactual for the others. This approach allows to identify the direct effect for the always and never takers and the direct and indirect effect for the compliers under the afore-mentioned assumptions, the monotonicity of the mediator in the treatment and some effect homogeneity restrictions. This method does not allow to identify the parameters of interest when the CIA is violated.

4.3. *Mediation Analysis Synthetic Control*

The method we propose in this section, allows to estimate the natural direct and indirect and controlled direct effects in frameworks where the number of treated units is low and/or SCIA and CIA are violated and no valid instruments can be found. Moreover, it can be easily extended to be used in the framework depicted in figure 4.3.

4.3.1. The General Method

In this paragraph we will introduce the Mediation Analysis Synthetic Control (MASC), a generalization of SCM that allows to do mediation analysis, i.e. to decompose the total effect of an intervention into its indirect component, which goes through observed mediators, and its direct component. Similarly to SCM, MASC creates synthetic values of the non-observed potential outcomes of the treated units using a linear combination of the donor pool. As we will discuss in more details, MASC re-weights control unit post-intervention outcomes by choosing weights that minimize the distance between treated and the linear combination of the controls in pre-intervention observable characteristics (including pre-intervention values of the outcome and the mediator) as well as in post-intervention values of the mediator. This allows us to mimic what would have happened to the treated in absence of the intervention if her mediator were set to her potential mediator under treatment. Following Abadie, Diamond, et al. 2010 we illustrate MASC with a simple dynamic factor model with interactive fixed-effects and show that both the direct and the indirect effects estimators are unbiased as the number of pre-intervention periods goes to infinity. Before, we have to better introduce the framework of interest.

As in standard synthetic control method, assume we are interested in the effect of an intervention D, implemented from time T, on an outcome Y. Consider the same notation and the definitions introduced in section 4.2. Notice that, in our framework the mediator M does not have to be binary but for each individual we define at most 2 potential values for the mediator at each time $M_{it}(1)$ and $M_{it}(0)$ (we assume no anticipation effects in the pre-intervention period such that if $t < T$ then $M_{it} = M_{it}(0)$ for any unit). This implies that for each individual there exists at most 4 potential outcomes at each time (under no anticipation effects in the pre-intervention period such that if $t < T$ then $Y_{it} = Y_{it}^{0,0}$ for any unit).

Following the synthetic control literature, we will define the parameters of interest with respect to a single treated unit. If more than one unit is exposed to the intervention (see Gobillon and Magnac 2016, Adhikari 2015) our method can be be easily extended to decompose the Average Treatment Effect on the Treated (Vansteelandt and VanderWeele 2012, Huber et al. 2017).

Once again, assume that we observe J units ordered such that units 1 to n are treated while units $n+1$ to J are not. Without loss of generality, we will first consider the first treated, unit 1. The effects of interest for unit 1 are the total effect α_{1t}, the direct effect $\theta_{1t}(d)$, and the indirect effect $\delta_{1t}(d)$ in the post-intervention period $(t \geq T)$ that are defined as

$$\begin{aligned}\alpha_{1t} &= Y_{1t}^{1,1} - Y_{1t}^{0,0} \\ \theta_{1t}(d) &= Y_{1t}^{1,d} - Y_{1t}^{0,d}, \\ \delta_{1t}(d) &= Y_{1t}^{d,1} - Y_{1t}^{d,0}, \quad t \geq T.\end{aligned}$$

As we have seen we can write α_{1t} as the sum of the direct and the indirect effect. Hence, if it is identified, identifying $\theta_{1t}(1)$ and $\delta_{1t}(1)$ automatically implies identification of $\theta_{1t}(0) = \alpha_{1t} - \delta_{1t}(1)$ and $\delta_{it}(0) = \alpha_{1t} - \theta_{1t}(1)$. For this reason, we will focus only on those two parameters here-after. Intuitively, the total effect α_{1t} can be identified with the synthetic control method as described in Abadie, Diamond, et al. 2010 and briefly summarized before. We will now show that a similar idea can be applied to identify the parameters of interest in mediation analysis framework. To this end we need to show that $Y_{1t}^{1,1}$, $Y_{1t}^{0,0}$, $Y_{1t}^{1,0}$ and $Y_{1t}^{0,1}$ are identified in post-intervention period. First notice that $Y_{1t}^{1,1}$ is observed in post-intervention period and $Y_{1t}^{0,0}$ can be estimated using standard SCM. Our main challenge is the identification of $Y_{1t}^{1,0}$ and $Y_{1t}^{0,1}$, both of which are never observed for any individual at any time and cannot be estimated through a standard SCM.

For $Y_{1t}^{0,1}$, we propose to re-weight the control unit post-intervention outcomes by choosing weights that minimize the distance between treated and control pre-intervention observable characteristics as well as post-intervention values of the mediator. The intuition is that, since $M_{1t} = M_{1t}(1)$ in post-intervention period, minimizing the distance between treated and controls with respect to post-treatment values of the mediator as well, when the weights are chosen, will mimic what would have happened to the treated in absence of the intervention if her mediator was set to her potential mediator under treatment $M_{1t}(1)$. Similar to SCM, MASC only works if the CTSA is satisfied both with respect to the outcome and with respect to the mediator.

Finding a "synthetic" value of $Y_{1t}^{1,0}$ is more challenging and requires more than 1 treated unit. First, we need to estimate what value the mediator of unit 1 would have taken in the absence of the intervention ($M_{1t}(0)$). This could be done with a standard SCM, using the mediator as an outcome. Second, we propose to treat the remaining treated as controls in a SCM where we use also the distance between the first step estimate of $M_{1t}(0)$ and the other treated mediators in post-intervention period, in computing the weights.

As mentioned before, in the spirit of Abadie, Diamond, et al. 2010, to further illustrate our approach, we will introduce a factor model in which we assume that potential mediators of unit i are given by

$$M_{it}(0) = \gamma_t + \beta_t Z_i + \vartheta_t \varrho_i + \nu_{it}, \tag{4.4}$$

$$M_{it}(1) = \gamma_t + \beta_t Z_i + \vartheta_t \varrho_i + \psi_t D_{it} + \nu_{it}, \tag{4.5}$$

where γ_t is an unknown common factor with constant factor loadings across units. Z_i is a $(p\times 1)$ vector of observed covariates, β_t is a $(1\times p)$ vector of unknown parameters, ϑ_t is a $(1\times v)$ vector of unobserved common factors, ϱ_i is an $(v\times 1)$ vector of unknown factor loadings, ψ_{it} is an unknown parameter describing the impact of the treatment on the mediator, and ν_{it} are unobserved transitory shocks.

Similarly, we assume that the four potential outcomes are given by

$$Y_{it}^{0,0} = \zeta_t + \eta_t X_i + \lambda_t \mu_i + \varphi_t(0) M_{it}(0) + \epsilon_{it}, \tag{4.6}$$

$$Y_{it}^{0,1} = \zeta_t + \eta_t X_i + \lambda_t \mu_i + \varphi_t(0) M_{it}(1) + \epsilon_{it}, \tag{4.7}$$

$$Y_{it}^{1,0} = \zeta_t + \eta_t X_i + \lambda_t \mu_i + \varphi_t(1) M_{it}(0) + \rho_t(M_{it}(0)) D_{it} + \epsilon_{it}, \tag{4.8}$$

$$Y_{it}^{1,1} = \zeta_t + \eta_t X_i + \lambda_t \mu_i + \varphi_t(1) M_{it}(1) + \rho_t(M_{it}(1)) D_{it} + \epsilon_{it}, \tag{4.9}$$

where ζ_t is an unknown common factor with constant factor loadings across units, X_i is a $(r \times 1)$ vector of observed covariates which includes all the variables included in Z_i but might also include other observable variables which affects the treatment and the outcome but not the mediator, η_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, ϵ_{it} are unobserved transitory shocks, and $\varphi_{it}(d)$ and

$\rho_{it}(M_{it}(d))$ capture the impact, on the potential outcomes, of the potential mediator and the treatment, respectively. In this model the total, direct and indirect effects of unit 1 are then given by

$$\begin{aligned}
\alpha_{1t} &= \varphi_t(1)M_{it}(1) - \varphi_t(0)M_{it}(0) + \rho_t(M_{it}(1)),\\
\theta_{1t}(1) &= \rho_t(M_{1t}(1)) + (\varphi_t(1) - \varphi_t(0))M_{1t}(1),\\
\theta_{1t}(0) &= \rho_t(M_{1t}(0)) + (\varphi_t(1) - \varphi_t(0))M_{1t}(0),\\
\delta_{1t}(1) &= \rho_t(M_{1t}(1)) - \rho_t(M_{1t}(0)) + \varphi_t(1)(M_{1t}(1) - M_{1t}(0)),\\
\delta_{1t}(0) &= \varphi_t(0)(M_{1t}(1) - M_{1t}(0)).
\end{aligned}$$

As mentioned above, for the total effect we can just use the standard SCM. In particular, we assume that there exists a $(1 \times (J - n))$ vector of positive and adding up to 1 weights $L^* = (l^*_{n+1}, ..., l^*_J)$ such that in the post intervention period

$$Y^{0,0}_{1t} = \sum_{i=n+1}^{J} l^*_i Y_{it}.$$

As in Abadie, Diamond, et al. 2015 we assume that $\forall\ t = 1, ..., T - 1$, L^* also satisfies

$$\begin{aligned}
\sum_{j=n+1}^{J} l^*_j Y_{jt} &= Y_{1t},\\
\sum_{j=n+1}^{J} l^*_j M_{jt} &= M_{1t},\\
\sum_{j=n+1}^{J} l^*_j X_j &= X_1.
\end{aligned}$$

This justifies choosing the weights that minimize the distance between the observable characteristics of the treated and the one of the control units in pre-treatment period. More formally, let $\Omega^{\alpha}_1 = (X_1, Y_{11}, \ldots, Y_{1,T-1}, M_{11}, \ldots, M_{1,T-1})$ be a $((2(T-1)+r)\times 1)$ vector, $\omega^{\alpha}_{0i} = (X_i, Y_{i1}, \ldots, Y_{i,T-1}, M_{i1}, \ldots, M_{i,T-1})$ be a $(1\times(2(T-1)+r))$ vector, and $\Omega^{\alpha}_0 = (\omega^{\alpha}_{0,n+1}, \ldots, \omega^{\alpha}_{0J})'$, then

$$\begin{aligned}
L^* &= \min_{l_{n+1},\ldots,l_J} ||\Omega^{\alpha}_1 - L\Omega^{\alpha}_0||\\
&\quad s.t.\quad l_{n+1} \leq 0, ..., l_J \leq 0, \sum_{i=n+1}^{J} l_i = 1,
\end{aligned}$$

where $||\Omega_1^\alpha - L\Omega_0^\alpha|| = \sqrt{\left(\Omega_1^\alpha - L\Omega_0^\alpha\right)' \left(\Omega_1^\alpha - L\Omega_0^\alpha\right)}$. It is also possible to give more weight to specific observable characteristics, by using the alternative distance $||\Omega_1^\alpha - L\Omega_0^\alpha||_V = \sqrt{\left(\Omega_1^\alpha - L\Omega_0^\alpha\right)' V \left(\Omega_1^\alpha - L\Omega_0^\alpha\right)}$ (we refer to section 4.3.2 for the procedure to choose V).

The first step of MASC is the estimation of $Y_{1t}^{0,1}$. This requires additional constraints. Indeed, we want to construct a "synthetic" unit which is identical to the treated, not affected by the intervention, and, at the same time, has the same value of the mediator as the treated unit. Similar to standard SCM, we want to find a $(1 \times (J-n))$ vector of positive and adding up to 1 weights $W_t^* = (w_{n+1,t}^*, ..., w_{Jt}^*)$ such that in post-intervention periods

$$Y_{1t}^{0,1} = \sum_{i=n+1}^{J} w_{it}^* Y_{it}.$$

Notice that, in our simple factor model, $Y_{1t}^{0,1}$ depends on the value that M takes at time t only[2]. Also notice that the weights need to be calculated at each post-intervention period in this model. Let $t' \geq T$ be the time at which we want to estimate the direct effect, similar to Abadie, Diamond, et al. 2010, we assume that $W_{t'}^*$ exists and it also satisfies $\forall\ t = 1, ..., T-1$

$$\begin{aligned} \sum_{j=n+1}^{J} w_{jt'}^* Y_{jt} &= Y_{1t}, \\ \sum_{j=n+1}^{J} w_{jt'}^* X_j &= X_1, \end{aligned}$$

and $\forall\ t = 1, ..., T-1, t'$,

$$\sum_{j=n+1}^{J} w_{jt'}^* M_{jt} = M_{1t}.$$

[2]It is easy to let $Y_{1t}^{0,1}$ depend on all the values that the mediator takes between T and t. This is done by replacing $\Omega_1^{\theta_{t'}(1)}$ and $\omega_{0i}^{\theta_{t'}(1)}$ defined further below with $\Omega_1^{\theta_{t'}(1)} = (X_1, Y_{11}, \ldots, Y_{1,T-1}, M_{11}, \ldots, M_{1,T-1}, M_{1,T}, \ldots, M_{1,t'})$ and $\omega_{0i}^{\theta_{t'}(1)} = (X_i, Y_{i1}, \ldots, Y_{i1,T-1}, M_{i1}, \ldots, M_{i,T-1}, M_{i,T}, \ldots, M_{i,t'})$, respectively.

The vector of weights $W^*_{t'}$ is estimated in a similar way as L^*. The only difference is that we now need to include the post-treatment mediator in the distance. More formally, let $\Omega_1^{\theta_{t'}(1)} = (X_1, Y_{11}, \ldots, Y_{1,T-1}, M_{11}, \ldots, M_{1,T-1}, M_{1,t'})$, $\omega_{0i}^{\theta_{t'}(1)} = (X_i, Y_{i1}, \ldots, Y_{i,T-1}, M_{i1}, \ldots, M_{i,T-1}, M_{i,t'})$, and $\Omega_0^{\theta_{t'}(1)} = (\omega_{n+1}^{\theta_{t'}(1)}, \ldots, \omega_J^{\theta_{t'}(1)})'$, then

$$\begin{aligned} W^*_{t'} &= \min_{w_{n+1,t'},\ldots,w_{Jt'}} ||\Omega_1^{\theta_{t'}(1)} - W_{t'}\Omega_0^{\theta_{t'}(1)}||_V \\ & s.t. \quad w_{n+1,t'} \leq 0, ..., w_{Jt'} \leq 0, \sum_{i=n+1}^{J} w_{it'} = 1, \end{aligned}$$

where $||\Omega_1^{\theta_{t'}(1)} - W_{t'}\Omega_0^{\theta_{t'}(1)}||_V = \sqrt{\left(\Omega_1^{\theta_{t'}(1)} - W_{t'}\Omega_0^{\theta_{t'}(1)}\right)' V \left(\Omega_1^{\theta_{t'}(1)} - W_{t'}\Omega_0^{\theta_{t'}(1)}\right)}$.

Let $\hat{Y}^{0,1}_{1t'} = \sum_{i=n+1}^{J} w^*_{it'} Y_{it'}$, as we show in the appendix, if $W^*_{t'}$ exists, under standard conditions

$$E(\hat{Y}^{0,1}_{1t'}) = Y^{0,1}_{1t'} + o(T)$$

Then, we can estimate the direct effect $\theta_{1t'}(1)$ and the indirect effect $\delta_{it'}(0)$ as

$$\hat{\theta}_{1t'}(1) = Y_{1t'} - \hat{Y}^{0,1}_{1t'}, \qquad \hat{\delta}_{1t'}(0) = \hat{\alpha}_{1t'} - \hat{\theta}_{1t'}(1),$$

respectively.

If the number of treated is big enough, we can also create a "synthetic" $Y^{1,0}_{it'}$. This is done in two steps. In a first step, we estimate $M_{1t'}(0)$ by $\hat{M}_{1t'}(0) = \sum_{i=n+1}^{J} k^*_{it'} M_{it'}$ with $K^*_{t'} = (k^*_{n+1,t'}, \ldots, k^*_{Jt'})$ chosen with a standard SCM. Note that those weights need to be calculated for each t'. In a second step, we need to find a vector of positive and adding up to 1 weights $Q^*_{t'} = (q^*_{2t'}, ..., q^*_{nt'})$, such that $Y^{1,0}_{it'} = \sum_{i=2}^{n} q^*_{it'} Y_{it'}$. $Q^*_{t'}$ is estimated with a SCM but using only the other treated units. More specifically, let $\Omega_1^{\delta_{t'}(1)} = (X_1, Y_{11}, \ldots, Y_{1,T-1}, M_{11}, \ldots, M_{1,T-1}, \hat{M}_{1t'}(0))$, $\omega_{0i}^{\theta_{t'}(1)} = (X_i, Y_{11}, \ldots, Y_{i,T-1}, M_{11}, \ldots, M_{i,T-1}, M_{i,t'})$, and $\Omega_0^{\theta_{t'}(1)} = (\omega_2^{\theta_{t'}(1)}, \ldots, \omega_n^{\theta_{t'}(1)})'$, then

$$\begin{aligned} Q^*_{t'} &= \min_{q_{n+1,t'},\ldots,q_{Jt'}} ||\Omega_1^{\theta_{t'}(1)} - Q_{t'}\Omega_0^{\theta_{t'}(1)}||_V \\ & s.t. \quad q_{n+1,t'} \leq 0, ..., q_{Jt'} \leq 0, \sum_{i=n+1}^{J} q_{it'} = 1, \end{aligned}$$

Let $\hat{Y}^{1,0}_{1t'} = \sum_{i=2}^{n} q^*_{it'} Y_{it'}$, similar as before, we assume that $Q^*_{t'}$ exists and satisfies $\forall\ t = 1, \ldots, T-1$

$$
\begin{aligned}
\sum_{j=2}^{n} q^*_{jt'} Y_{jt} &= Y_{1t}, \\
\sum_{j=2}^{n} q^*_{jt'} X_j &= X_1,
\end{aligned}
$$

$$\sum_{j=2}^{n} q^*_{jt'} M_{jt} = M_{1t},$$

$\forall\ t = 1, \ldots, T-1$

and

$$\sum_{j=2}^{n} q^*_{jt'} M_{jt'} = \hat{M}_{1t'}(0).$$

Under extra standard conditions and assuming that $\rho_{t'}(\cdot)$ is a linear function, as we show in the appendix

$$E(\hat{Y}^{1,0}_{1t'}) = Y^{1,0}_{1t'} + o(T).$$

The latter assumption can admittedly be restrictive in many applications. However, it is substantially weaker than assuming a constant $\rho_{t'}$. Then, we can estimate the indirect effect $\delta_{it'}(1)$ and the direct effect $\theta_{1t'}(0)$ as

$$\hat{\delta}_{1t'}(1) = Y_{1t'} - \hat{Y}^{1,0}_{1t'}, \qquad \hat{\theta}_{1t'}(0) = \hat{\alpha}_{it'} - \hat{\delta}_{1t'}(1),$$

respectively. Intuitively, $Q^*_{t'}$ exists under the same assumption discussed above. However, if the number of treated units is too small $\hat{Y}^{1,0}_{1t'}$ may be a very poor approximation of $Y^{1,0}_{1t'}$. Hence, in those settings is only possible to estimate $\delta_{it'}(0)$ and $\theta_{it'}(1)$.

If the number of treated units is big enough, the approach used in MASC can be easily extended to estimate the controlled direct effect. In particular, if the parameter of interest is the direct effect $\theta_{1t'}(m) = Y^{1,m}_{1t'} - Y^{0,m}_{1t'}$, MASC has to be applied twice (where in this framework the value m represents the chosen value of the potential mediator instead than the treatment status determining the potential mediator). In a first step, using untreated units as the donor pool, to estimate $Y^{0,m}_{1t'}$. In a second step, using treated units as the donor pool, to estimate $Y_{1t'}^{1,m}$. In both steps, it has to be imposed a post-treatment constraint forcing the mediator of the synthetic unit to be equal to m.

Furthermore, MASC can be easily extended to identify direct and indirect effects in presence of post-treatment confounders, i.e. in the setting depicted in figure 4.3. To give an intuition, making some assumption on post-treatment confounder structure, constraints in post-treatment period can be used to obtain a synthetic unit with the required mediator and post-treatment confounder values. I.e. to estimate $Y_{1t}^{0,0,(1,1)}$ (and identify $\theta_{1t}^{Ima}(d)$) two constraints in post-treatment period should be used. A first one requiring the synthetic unit mediator to be equal to the treated unit mediator. A second one requiring the synthetic unit post-treatment confounder to be equal to those of the treated unit in absence of treatment.

4.3.2. Implementation

To implement MASC it can be used the same implementation method proposed in Abadie, Diamond, et al. (2010) and in Abadie, Diamond, et al. (2015). In particular, the weights to use to build the synthetic control can be selected minimizing the following quantity[3]: $||\Omega_1 - W_{t'}\Omega_0||_V = \sqrt{(\Omega_1 - W_{t'}\Omega_0)' \, V \, (\Omega_1 - W_{t'}\Omega_0)}$. In SCM implementation, even though the estimation is valid under any value of V, an optimal choice of this matrix can reduce the mean squared error of the estimator (Abadie, Diamond, et al. 2010). Abadie, Diamond, et al. (2015) proposed to search for the optimal matrix V using a cross-validation method, employing a subset of pre-treatment data to choose the matrix V that minimizes the root mean squared prediction error (RMSPE) in it. This approach was criticized in Klößner et al. (2018). The authors underlined as using a cross-validation technique the matrix V is not uniquely identified and this can be a problem in the placebo tests and in the leave-one-out tests used to make inference when SCM is applied (see section 4.3.3). To solve this problem, M. Becker et al. (2018) proposed to impose an additional constraint during weights

[3]Where with this representation we are not referring to the estimation of a particular parameter and what we will say in this section is valid generally for the estimation of both $Y_{1t}^{0,0}$, $Y_{1t}^{1,0}$ and $Y_{1t}^{0,1}$ unless specified otherwise.

choice, such that they are uniquely identified. The constraint implies that no covariates can be accidentally irrelevant.

In MASC framework, the choice of matrix V has to follow some additional considerations. The first consideration is that, in the estimations requiring post-treatment constraint (i.e. $Y_{1t}^{1,0}$ and $Y_{1t}^{0,1}$), there are only one (or few) constraint(s) on post- and several constraints on pre-intervention period. If this is not taken into account properly in V choice, the post-treatment constraint is unlikely to be satisfied. Hence, we suggest to choose V such that an equal weight is given to pre- and post- intervention information. The second consideration is that optimizing weights V with respect to outcome's RMSPE does not guarantee the equalities on pre-treatment mediator values to be satisfied. This is true especially if the mediator has a low correlation with the outcome or with its determinants. In MASC the equality between pre-treatment mediator of the treated and of the syntetic units is important for a proper estimation of all the parameters of interest. The simplest way to solve this issue is to repeat the estimation trying different weights partition between outcome and mediator information and to choose the partition giving the most satisfying results in terms of both pre-treatment outcome and mediator. A more complicated (but more rigorous) solution is to do the same using an optimization function. The latter can either minimize both the distances or minimize those in pre-treatment mediator values under the constraint that the distance in pre-treatment outcomes is smaller than a fixed value. Given all these considerations, and the fact that V choice does not have an impact on estimation validity, we suggest to follow Gobillon and Magnac (2016) and use fixed V. The weights should be divided equally between pre- and post- treatment information and, among pre-treatment information, between outcome and mediator information such that pre-treatment constraints are satisfied for both of them. As an alternative, once the weights are divided between pre- and post- treatment information and between outcome and mediator pre-treatment information the method proposed in M. Becker et al. (2018) can be easily applied, separately, to the subset of weights concerning pre-treatment outcome, pre-treatment mediator and post-treatment mediator information.

A strand of the literature on SCM implementation focused on covariates selection, i.e. the choice of pre-treatment outcomes functions and of the observables. In particular, Botosaru and Ferman (2017) studied the relevance of covariates constraints in SCM. They demonstrated that, if the assumption of non-collinearity and non-irrelevance of the unobservables is extended to the observables, the constraint on pre-treatment outcomes is enough for the estimator bias to go to zero when the number of pre-treatment periods goes to infinity. Moreover, they demonstrated that, under these assumptions, the equality on pre-treament outcomes guarantees the equality on observables and unobservables. Their demonstration is very simple, they simply treated the observables as the unobservables in standard SCM. Hence, their conclusions can be easily extended to MASC framework. The authors showed as well that the first result is still valid when the two assumptions are relaxed. Kaul et al. (2018) have a completely different opinion. The authors show as the use of all available pre-treatment outcomes is equivalent to the exclusion of all covariates. They claim that the exclusion of all covariates is in contrast with the idea behind SCM, even though it still gives consistent results.

MASC implementation requires an additional choice: the time periods to include in post-treatment constraints. Indeed, as mentioned above, if we are estimating the effects at time t' we can either impose post-treatment constraints on all time periods between T and $t'-\phi$ (where ϕ value should be chosen according to the authors beliefs on the timing of mediator effect) or on $t'-\phi$ only. The difference between these two approaches is that in the first case we ask the treated and the synthetic units to share the same mediator trend in the whole post-treatment period. In the second we ask them to share a single value of the mediator in a single time period. To the end of the demonstration of estimators validity this choice does not change much. Nonetheless, it can make a difference in terms of assumptions plausibility. Indeed, the constraint only on period t' is more likely to be satisfied. On the other hand, under the constraint on the whole trend the synthetic unit will be more similar to the treated ones, hence it will be a more valid counterfactual.

An additional recommendation, during MASC implementation, is to check graphically the overlap between treated and synthetic units for both pre-treatment outcome and pre-treatment and post-treatment mediator values (rather than pre-treatment outcome only as in SCM).

4.3.3. Inference

Inference can be carried over in a similar manner as in standard synthetic control method. For example, one can run similar placebo tests as the one suggested in Abadie, Diamond, et al. 2015, estimating the effects (in our case also the direct and indirect) of the intervention either before its implementation or for units not exposed to it. Abadie, Diamond, et al. 2015 criticize the former type of placebo tests (often called in-time placebos) arguing that there may be other shocks in the past affecting treated and control units differently. They suggest to use the ratio between post- and pre- intervention Root Mean Square Prediction Error (RMSPE) as test statistics during inferential procedure. For unit i and synthetic $\hat{Y}^{d,d'}$ the pre-intervention RMSPE can be defined as

$$RMSPE_i^{\hat{Y}^{d,d'},pre} = \frac{\sum_{t=1}^{T-1}(Y_{i,t} - \hat{Y}_{i,t}^{d,d'})^2}{T-1}.$$

The post RMSPE is defined similarly

$$RMSPE_i^{\hat{Y}^{d,d'},post} = \frac{\sum_{t=T}^{t'}(Y_{i,t'} - \hat{Y}_{i,t'}^{d,d'})^2}{t'-T-1}.$$

The test-statistic can then be defined as

$$Test_i^{Y^{d,d'}} = \frac{RMSPE_i^{\hat{Y}^{d,d'},post}}{RMSPE_i^{\hat{Y}^{d,d'},pre}}.$$

The choice to use the ratio between $RMSPE_i^{\hat{Y}^{d,d'},post}$ and $RMSPE_i^{\hat{Y}^{d,d'},pre}$ rather than the first measure alone is crucial to have a correct inference. Indeed, Ferman and Pinto (2017) showed as using $RMSPE_i^{\hat{Y}^{d,d'},post}$ only the test statistics may have different marginal asymptotic distributions under different permutations. Moreover, they showed as the test statistics based on the ratio performs better and is less sensitive to violations of the assumptions. They proposed as well another test statistics which is robust to serial correlation in temporary shocks.

Firpo and Possebom 2017 generalized the procedure proposed in Abadie, Diamond, et al. (2015) such that it could be used for other test statistics and for any sharp null hypothesis. The authors also showed that the method proposed in Abadie, Diamond, et al. (2015) was more powerful than standard test statistics in presence of a single treated unit. They also extended their method to frameworks with multiple outcomes and multiple interventions in different regions. Chernozhukov et al. (2018) proposed an inferential method which can be applied more generally to any framework where a counterfactual outcome is generated. They use a test statistics based on residuals and provides as well a formal justification of the inference procedure proposed in Abadie, Diamond, et al. (2015). This methodology can be easily extended to be used in our framework as well.

Yet another inference procedure is described in Gobillon and Magnac (2016). This procedure is based on different steps. First of all, the outcome of the treated units is reduced by treatment effect. In our framework, the mediator of treated units should be reduced as well, using the estimation of the potential mediator in absence of treatment. Later on, 10'000 samples without replacement of a number of units equal to the number of treated units have to be selected from the group containing all units. For each of the 10'000 samples the selected units should be used as treated units and the rest of the group as control, to apply the synthetic control method (the MASC in our framework). Finally, the estimated values should be used as distributions to make inference on the estimated effects. In MASC, this method would consist in the following steps:

1. Substitute Y_{it} with $Y'_{it} = Y_{it} - \hat{\alpha}_t$ for $i = 1, ..., n$ and $t \geq T$. Where α_t is given by the average among all the total effect estimated.

2. Substitute M_{it} with $M'_{it} = M_{it} - E(M_{it} - \hat{M}_{it}(0))$ for $i = 1, ..., n$ and $t \geq T$.

3. Iterate 10'000 times:

  - Select *n* units.
  - Apply MASC on selected unit.
  - Calculate the average total, direct and indirect effects

4. Use the calculated effects to determine the distribution of the real effects and do inference.

We refer to Gobillon and Magnac (2016) for more details.

Note that this inference procedure, just as Abadie, Diamond, et al. 2015 ones, is based on the strong assumption that the disturbances across units are exchangeable. Indeed, the basic ideas behind these methods is that the noise of the placebos can be used to approximate the noises of the treated units.

In SCM (and, consequently, in MASC) framework there is a second source of uncertainty. Unfortunately, the choice of the control units (donor pool) can dramatically affect the results. To solve this issue Abadie, Diamond, et al. (2015) suggest to make a sensitivity test excluding one by one each of the units in the donor pool (if the donor pool is particularly big another possibility is to select a sample with replacement from the donor pool). If the estimated effects do not change much, the results are not sensitive to the chosen donor pool and can be considered robust.

5. Put the Dream Back Together

In this chapter, MASC is applied to investigate on the mechanism behind the temporariness of European Structural Funds effect. The chapter starts with an overview on European Structural Funds, followed by some information on the allocation of transfers in our area and period of interest and some literature on their effectiveness. The aim of the literature review is twofold. We want both to show the state of the literature on the estimation and analysis of SF impact and to provide measures of it for the years and regions of interest. Therefore, we focused on the literature including at least one of the two programming periods between 1988-2000 in the time of interest and Abruzzi region in the sample. After this overview, details on MASC implementation and its results will be provided.

5.1. *An Overview*

5.1.1. European Structural Funds

European Structural Funds were introduced for the first time in 1975. Their goal was (and it is) to reduce the wide regional disparities among European regions. To use the words of the Commissioner for Regional Policy in 1971, Albert Borschette, the convergence among european regions was essential because "The monetary union is not possible if the current disparities and differences present among the community continues to exist". In 1988, there was a fundamental reform, the funds took the form they have today and the reduction of economic and social disparities became one of the three highest priorities for the Union. As a consequence, today, Structural and Cohesion Funds account together for one-third of Community policies budget. The Structural Funds include the European Regional Development Fund, the European Social Fund, the European Agricultural Guidance and Guarantee Fund and the Financial Instrument for Fisheries Guidance. ERDF and ESF are the most relevant in terms of budget. The first accounted for 45% of total structural funds during the period 1989-1999 (which, as we will see, is our period of interest). Its goal is to streghten the economic potential of the beneficiary regions through support to productive investments (especially those involving innovation or fostering sustainability), infrastructural projects (i.e. for transport, telecommunications and energy) and co-financing operational programmes. It also provides assistance to large projects implementation and during preparatory studies. The ESF, instead, accounted for 30% of total structural funds during the same period. Its goal is to fight long-term unemployment and improve the employability of youths and other vulnerable categories of unemployed workers and to promote adaptation to industrial change (i.e., contributing to human capital development). The main target of SF are the regions lagging behind. In particular, most of the Funds are used in regions having a GDP per capita in PPP terms lower than 75% of EU average. The so-called Objective 1 regions. Most of the Countries belonging to this group are largely agricoltural, with a low R&D level and a high level of unemployment (Cappelen et al. 2003). In the programming period 1988-1993, 70% of the programme budget was used for Objective 1 regions. In 1994-1999 programming period 68%. The goal of Objective 1 transfers is not just to induce economic development in the short term. Innovative investments, unemployment reduction and improvements in trasportation, telecommunication and energetic infrastructures should bring the lagging behind regions on a faster self-sustaining growing path Barone et al. (2016). Programmes subsidized by the funds have to be co-financed by the beneficiary Country or region.

5.1.2. European Structural Funds in Mezzogiorno

Italy has always been characterized by strong internal disparities. Immediatly after unification, the northern area of the Country was characterized by an environment more favorable for industrial development. Indeed, it was closer to European markets and water resources. On the opposite, Mezzogiorno area (i.e. the Southern area of Italy) was only specialized in industries that would have had low relative growth rates in the following decades and lagged behind in terms of infrastructures and human and social capital (Iuzzolino et al. 2011). Over most of the Country history different policies were implemented as an attempt to reduce those

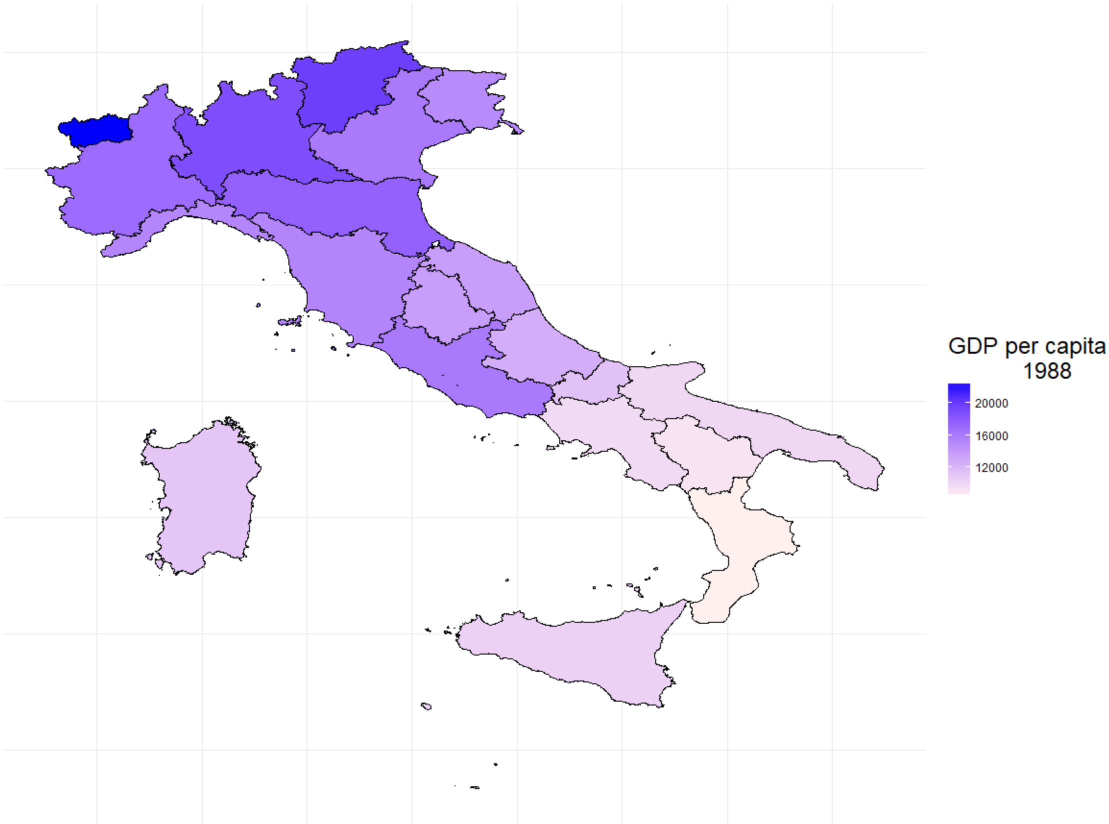


Figure 5.1: Different color intensity represents different GDP per capita levels in 1988. Source: Barone et al. (2016) database.

disparities. Nonetheless, as it is visible from figure 5.1, where different color intensities represent different GDP per capita levels in 1988, when the "modern" European Structural Funds started, disparities were still very high. Consequently, in 1988, as visible from figure 5.2, all regions of Mezzogiorno area belonged to Objective 1. During the period 1989-1999 the total transfers received in Mezzogiorno were equivalent to 3% of the GDP of the area. According to the European Commission report, they had a strong role in sustaining investment levels in the area and reducing unemployment. A big portion of the funds was used to extend the natural gas distribution network, reaching 75% of the population of southern regions. The transfers were used as well to improve communications and water supply in the area, to do training courses, boost tourism and transfer innovative production methods to

Figure 5.2: Regions belonging to Objective 1 group during the programming periods 1989–1993 and 1994–1999 are evidenced in red.

local farms. Nonetheless, serious delays during the whole implementation period reduced the effectiveness of Structural Funds transfers (Commission 1997).

In Abruzzi region, as it can be seen from table 5.1, most of the funds were used for agricolture and tourism sectors. In the latter, they were especially used for restoration of touristic attractions or touristic services (Abruzzo 2001).

Categories of use	Funds Amount	Share of the Total
Telecommunications	129311	0.12
Industry, artisan and service sectors	106752	0.099
Tourism	261468	0.243
Agricolture	373552	0.347
Infrastrucures in support of business activities	116322	0.108
Human capital promotion	86248	0.080
Technical assistance, monitoring	3866	0.004
Total	1077519	

Table 5.1: European SF transfers allocation in Abruzzi over the programming period 1994-1997 with respect to different mode of use.

In 1997, Abruzzi GDP per capita crossed the threshold of 75% of European average and the region exited from Objective 1. It was the only EU region which exited the program without a phasing out support. Due to delays in programmes implementation the transfers were effectively reduced only starting from 2000. In 2000-2006 cycle Abruzzi was an objective 2 region. The transfers, and the co-financing public resources were halved (Barone et al. 2016).

5.1.3. Literature Review

As soon as the European Structural Funds were introduced, a vibrant debate on their effectiveness started. When data on the firsts programming periods became available, several empirical analysis were conducted in support of this debate. To the best of our knowledge, the first empirical analysis was Boldrin and Canova (2001). To verify the effectiveness of European SF the authors followed beta- and sigma-convergence approaches and regressed the variation in the logarithm of GDP per capita on its initial value. They repeated the regression separately before and during policy implementation and compared the results. Given that there was no difference in the coefficients between the two periods, they concluded the funds were not effective to foster convergence among regions. Nonetheless, as the authors themselves underlined, convergence level before SF implementation may not be a good representation of the potential convergence in absence of the policy in the following years. Indeed, if in absence of SF the convergence had slowed down, a constant convergence would reveal a positive effect of the funds.

The subsequent studies, based on beta- and sigma-convergence or on other theoretical growth models, took into account of this possibility, including the SF as a covariate in the regression. Moreover, they included additional variables that could influence the outcome. Among them, Cappelen et al. (2003) used a panel data model with fixed effect to study the effect of SF at a regional level, over the period 1989-1997. They chose the covariates and the structure of the regression according to a theoretical model, based on the idea that development is mainly driven by technological innovation set in a proper environment. They wrote productivity as a multiplicative function of the level of knowledge coming from outside the region, those born inside it, the capacity of the region to exploit benefits of knowledge and a constant. They assumed diffusion of external knowledge follow a logistic curve. In the regression model, they included measures of physical infrastructure, population density, industrial structure and long-term unemployment. To solve the problem of possible correlation between structural funds and some of the covariates, they exploited the fact that structural funds increased strongly from 1980-1988 programming period to 1989-1997 programming period. In particular, they included a dummy variable indicating the first programming period, interacted with the other covariates. They concluded structural funds have a positive and significant impact on economic growth of European regions, contributig to reach major equality in productivity and income. Moreover, they said that the increase in the funds amount in 1988 had a positive impact on the effectiveness of the policies. The effects are stronger in more developed environments.

In line with this result another Country-level evaluation showed that SF effect was higher in Countries with high-quality institutions (Ederveen et al. 2006). The authors based their regression on the theoretical convergence model introduced by Mankiw et al. (1992). They used data on 13 EU Countries and 5-years observations over the period from 1960-1965 to 1990-1995. They focused only on European Regional Development Fund. Interacting SF variable with some proxies of institutions quality the authors found a positive and significant impact of SF in countries with high-quality institutions and a negative and significant impact in countries with low-quality institutions.

Puigcerver-Penalver (2007) conduced her empirical analysis on data at a regional level. Focusing on Objective 1 regions over the period 1989-1999, she applied a panel data model with fixed effects. She regressed the GDP per capita in PPP terms growth on different measures of SF amount and private and public expenditures. As the previous authors, she based her regression on a theoretical model. In particular, she proposed a hybrid model where technological growth is a consequence of both exogenous and endogenous forces. The author modeled production as a Cobb-Douglas aggregate production function with constant returns to scale. She combined the macroeconomic level with a minimization problem at household level. From the empirical analysis, it emerged that SF had a positive and significant impact on rates of growth of Objective 1 regions during the whole period under analysis.

As Puigcerver-Penalver (2007), Esposti and Bussoletti (2008) used panel data at a regional level, on 200 european regions, and focused on Objective 1 transfers over the period 1989-2000. The authors described convergence process with a beta-convergence model, where average income per worker growth, conditional on initial income, depended on total factor productivity growth rate, speed of convergence, initial total factor productivity, production (modeled as a Cobb-Douglas production function), initial investment rate, initial population and capital depreciation rate. The authors focused on the long-run effect of the policy on supply side, claiming that SF should boost investments increasing in this way total factor productivity and, consequently, it should induce growth in the labor market. A possible issue employing this regression model was the endogeneity of the lagged value of the outcome with respect to the other covariates. Hence, they employed, for the estimation, a Generalized Method of Moments based on first differences and one based on system of equations (using the difference between past and present outcome as an instrument). They concluded there was convergence and the impact of SF was negligible. Nevertheless, both these results were not robust to alternative specifications and estimators (i.e. the impact of SF was positive under some specifications).

Other authors used a GMM regression. Beugelsdijk and Eijffinger (2005) used it to regress the GDP growth on its lagged values and its initial level, a measure of the structural funds received and some additional covariates. The authors wanted to investigate whether the selection-into-treatment rule of SF induced moral hazard in the most corrupt Countries. I.e. they wondered whether the most corrupt governments decided not to use the funds properly to avoid threshold crossing or use them for policies they would have done anyway (substitution effect). To corroborate their hypothesis, they interacted, in the regression, SF levels with a corruption index. They found a positive and significant effect of SF on GDP growth and concluded it was not affected by corruption level of the Countries. With regard to endogeneity they concluded there was no serious autocorrelation in the outcome.

The works summarized so far have two main limits. The first one is that most of them do not take into account of a possible correlation between the variable measuring SF and the other covariates (with the exception of Cappelen et al. (2003)). The second is that none of them take into account of selection bias. Even though some of them include other outcome determinants as covariates, they are not chosen following the counterfactual approach and nothing guarantees CIA is satisfied. In particular, the accuracy of the estimation strongly depends on the right specification of the theoretical model the analysis relied on. Dall'Erba and Fang (2017), in their meta-analysis, underlined as more attention should be provided to approaches different from the neoclassical beta-convergence model and looked favorably to a second strand of the literature based on counterfactual models.

One study in-between the first and the second strands is Dall'erba and Le Gallo (2008), where the authors applied a beta-convergence model and took into account of SF endogeneity. They used data at a regional level over the period 1989-1999 and GDP growth as an outcome. The authors took into account of possible spillover effects, determining the spatial correlation among regions in absence of SF and using it inside the final model. Moreover, they took into account of spatial heterogeneity dividing the regions into a peripheric and a core groups through quantitative analysis and including dummies for the two groups (interacted with all variables of interest) in the final model. Finally, they took into account of endogeneity using several instruments. To control for endogeneity of the spatial lag outcome, they used the spatial lag of all explanatory variables. To control for the endogeneity of share of agricolture and long-term unemployment they used two quasi-instrument. They classified the regions in three different categories according to each variable and used the new factor variables as instruments. Finally, to control for the endogeneity of SF they used four different instruments: the distance by road to Bruxelles, travel time from the most populated town of the region to Bruxelles, a quasi-instrument built as before with respect to SF level and its spatial lag. Their Hausman tests suggested that only the share of agricolture and the SF were endogenous. The regression results, instead, suggested that SF had no impact on the outcome.

To the best of our knowledge, the first study relying only on counterfactual methods is Hagen and Mohl (2008). The authors estimated SF effect on economic growth rate applying a generalized propensity score matching on a sample of 122 European regions. They measured treatment as the actual regional SF payment (rather than the initial commitment) and excluded all observations with zero payment. They restricted the estimation of the dose-response function on a range of treatment up to the 75% quantile given that there was a small number of observations in the upper tail of the distribution. According to the authors, covariates choice was mostly driven by data availability. They included pre-treatment values of the outcome variable, the ratio between GDP per capita in PPP terms and EU average, population density, employment structure and unemployment rate, the ratio between long-term unemployment and total unemployment, lagged and squared unemployment level and country dummies. From their analysis, it emerged SF effect increased as their amount increased. Nonetheless, the confidece interval at 95% increased together with the estimated effect, making it always non-significant. The peculiar divergence in confidence interval upper and lower bounds suggest the non-significance may be due to the low amount of observations in correspondence of higher SF amounts. Hence, we believe these results should be accompanied by more reliable analysis to take conclusions on SF effectiveness.

Luckily for us, this is not the unique estimation based on counterfactual analysis. S. O. Becker et al. (2010) used a fuzzy regression discontinuity design to estimate the impact of Objective 1 transfers on average annual growth of GDP per capita in PPP terms and employment. They exploited the 75% threshold on GDP per capita defining Objective 1 regions (see 5.1.1). After the estimation, the authors did several sensitivity checks, controlling, among the others, for spillover effects and different programming periods. The estimation was robust to all of them. The analysis showed that eligibility for Objective 1 regions had a positive and significant effect on GDP growth and a non-significant effect on employment growth. They concluded transfers had an immediate impact on investments while it took them longer time to affect employment.

Some other authors applied different versions of the regression discontinuity design. Pellegrini et al. (2013) used a sharp RDD to study the impact of Objective 1 eligibility on real GDP per capita growth over the period 1995-2006. The authors used a sample of 190 regions and applied different non-parametric estimation procedures. They concluded eligibility had a positive and significant effect on the outcome and this result was robust to the different procedures they used. Giua (2017), instead, applied a spatial regression discontinuity design using regional border between municipalities belonging to Objective 1 and non-Objective 1 regions as a threshold. She used the total employment variation between 1991 and 2001 as outcome (the results were nonetheless robust to the choice of 1981-2001 variation, suggesting the effect of the policy was not yet visible in 1991) and repeated the analysis including dummies for different economic sectors. She concluded the policy had a positive and significant impact on employment growth, especially in the sectors most likely to be influenced (manifacturing, construnction, retail and tourism). Cerqua and Pellegrini (2018) started applying a standard RDD to estimate the impact of structural and cohesion funds on economic growth over the period 1994-2006. Later on, they extended this method to estimate the local average treatment level effect. In particular, to estimate the LATLE using an RDD, they exploited the fact that also non-Objective 1 regions received some funds. They defined a first treatment as being highly treated (Objective 1 regions) or low treated (non-Objective 1 regions). Later on, they defined a second measure of treatment, as the distance of one region from the average reception of transfers in its group (i.e. for an Objective 1 region the second treatment is measured as the distance of its reception of funds from Objective 1 average). They identified the LATLE under the assumption that treated and untreated regions with the same value of the second treatment, were equal in the unobservables dimensions relevant for the outcome. To make this assumption more plausible, they controlled for some covariates. The authors measured funds reception as the amount of EU payments, by operational program, per year, at regional level. They used different specification tests as robustness checks. Among the others, they used an IV approach to test for transfers level endogeneity and checked for spatial error correlation. The estimation was robust to all checks. The authors concluded the funds had a positive and significant effect on regional growth. They found a concave conditional intensity-growth function, with the maximum effect of the policy for 305-340 euros of per capita transfers.

Another study, investigating the variability of the effect with respect to different transfers level, is S. O. Becker et al. (2012). As Hagen and Mohl (2008), the authors used a generalized propensity score matching to determine the dose-response function for European SF. They used data at NUTS 3 level. It emerged that treatment effect was positive and significant and the optimal intensity level of treatment was 0.4% of GDP.

It is worth to mention a third strand of the literature on Structural Funds. The main difference between the last and the previous ones has been explained very well by its pioneers, using the words of a World Bank lead economist: " A treatment is an instance of treating someone, say medically. A cure ends a problem. Sometimes, the treatment is a cure. Other times it just keeps the problem under control without curing it: if you remove the treatment, the problem comes back" (Ozler 2014). While the first two strands of the literature focus on the impact of the treatment this third strand of the literature investigates whether the treatment is the cure (Barone et al. 2016). Knowledge about the longevity of SF effect is fundamental to assess the validity of this tool and to understand whether and how it should be improved. Indeed, the main goal of Objective 1 European Structural Funds is to reduce the gap between the poor and the rich areas permanently, activating a self-sustaining faster growing path in the poorer regions (Esposti and Bussoletti 2008, Barone et al. 2016). To the best of our knowledge, the pioneers of this strand of literature are Barone et al. (2016). To investigate whether the funds were the cure, the authors exploited the fact that Abruzzi exited the SF Objective 1 program in 1997 without a phasing out support. Consequently its SF, and the national public resources associated to them, were more than halved in 2000. The authors, applied a Synthetic Control Method using Abruzzi as the treated region and defining the treatment as the end of transfers reception. They included the other regions belonging to the Mezzogiorno area (namely Molise, Campania, Basilicata, Calabria, Sicily, Sardinia, Puglia) in the donor pool. They used the period 1980-2000 as pre-treatment period and calculated the effects on the post-treatment period 2001-2008 (following years were excluded because Abruzzi was hitten by a earthquake in 2009). To make inference they conduced in-time and in-space placebo tests (the latter using all others italian regions). The authors observed that there was, on average, a 5.5% statistically significant drop in indexed GDP per capita (for further information on the outcome and the control variables they have used see 5.2) due to the lost of Objective 1 support. Comparing this result with those obtained in previous literature, the authors concluded most of the effect of the transfers was temporary. Barone et al. (2016) did as well a series of robustness check, i.e. they excluded the neighbouring regions from the donor pool to check for spatial spillovers. The results were robust to them.

To the best of our knowledge, there is only one other work belonging to this strand of the literature, S. O. Becker et al. (2018). The authors started estimating the impact of Objective 1 treatment on average annual growth of GDP per capita in PPP terms, average annual employment growth and total and public investment intensity (as percentages of GDP). Later on, they estimated the impact of loosing an Objective 1 classification on a series of regions. To do it, they exploited the fact that the inclusion of eastern european regions lowered significantly average GDP level, and, consequently, policy eligibility threshold. In both estimations, the authors applied a fuzzy RDD over the period 1989-2013. The authors found a significant and positive effect of the policy on GDP per capita growth. They found a non-significant effect on employment growth and general investments. Finally, they found a positive and significant effect on public investment. From the two last results the authors concluded the public capital stock crowded out some of the private investments. With respect to the lost of Objective 1 classification, the authors started comparing regions exiting from the treated group with those never-entered and found a positive and significant effect of "temporary treatment" of 2.1-2.6% of GDP. Later on, they compared the first group with regions treated the whole time and found a negative and significant effect on GDP of exiting from treated group of 1.7%. They found no significant impact on employment and suggested it is due to the slow response of labour market. They concluded the effect of SF transfers is not permanent and it seems to vanish when transfers are stopped.

To summarize, there are three strands of literature on European Structural Funds. The first one relies on theoretical models of growth. Among the works belonging to this strand, some concluded European Structural Funds have a positive and significant effect on GDP growth (Cappelen et al. 2003, Puigcerver-Penalver 2007, Beugelsdijk and Eijffinger 2005), some of them concluded their effect was negligible (Boldrin and Canova 2001, Esposti and Bussoletti 2008, Freitas et al. 2003) and one of them concluded it depended on Country's institutions (Ederveen et al. 2006). One of them investigated on policy impact on employment and concluded it was negligible (Boldrin and Canova 2001). Finally, none of them investigated on policy impact on investments. As mentioned before, the results of this first strand should be taken with caution given that they do not take into account of possible selection bias and endogeneity issues. To the second strand belong analysis estimating the impact of SF on growth through a counterfactual approach. Most of them identified a positive and significant effect (S. O. Becker et al. 2010, Pellegrini et al. 2013, Giua 2017, Cerqua and Pellegrini 2018, S. O. Becker et al. 2012). Only two of them concluded policy effect was negligible (Dall'erba and Le Gallo 2008, Hagen and Mohl 2008). Nonetheless, the reliability of Dall'erba and Le Gallo (2008) results (independently from how convincing their instruments are) is weakened by the inclusion of investments among the covariates. Given that SF should foster investments, the last is a bad control. The fact that Hagen and Mohl (2008) result is in contrast with the rest of the second strand of the literature may be due to the fact that the authors conduced the analysis only on units with a range of treatment up to the 75% quantile, excluding the most treated areas. Hence, they may have estimated a lower bound of the actual effect.

Few works focused on the estimation of treatment effect on unemployment. One of them found a positive and significant effect (Giua 2017). Two of them found no significant impact (S. O. Becker et al. 2010, S. O. Becker et al. 2018). This difference may be due to the fact that the second two works focused on a post-treatment period too short to observe the slow response of labour market. Only one of the second strand papers estimated treatment effect on investments-to-GDP ratio and public investments-to-GDP ratio and found no significant effect on the first and a positive and significant effect on the second (S. O. Becker et al. 2018).

The third strand of the literature includes studies investigating on the longevity of Objective 1 transfers effect. Both the works belonging to it concluded European Structural Funds have a temporary effect (Barone et al. 2016, S. O. Becker et al. 2018). The first found a negative effect significantly higher than the second. This may be due to the fact that it focused on a region which did not receive a phasing out support, while the second included many units which benefitted from a more gradual exit.

To the best of our knowledge, none of the works based on a counterfactual approach investigated on the causal channels between Objective 1 belonging and economic growth. As we have seen (see 5.1.1 and 5.1.2), Objective 1 transfers can be used in different frameworks, i.e. they can be used to boost investments or to reduce unemployment level. Consequently, their impact can travel through different mediators. Focusing on the third strand of the literature, it is of high interest to understand which are the causal channels guaranteeing more longevity and which are the channels through which the transfers have only a temporary impact. This information would be highly policy relevant. As an example, if the component of transfers effect passing through investments is permanent while the component passing through employment is temporary, it could be useful to center SF efforts on the first or to rethink the second. The contribution of this chapter goes in this direction.

5.2. *MASC Implementation and Inference*

The exit of Abruzzi from the group of Objective 1 regions, and the subsequent reduction in structural funds transfers, caused a reduction of its GDP per capita, indexed at 1995, of 5.5% in the following eight years (Barone et al. 2016). Our goal is to understand to what extent the sudden reduction in SF transfers had an impact through a reduction of investments and to what extent it had impact through a reduction in employment. To reach it, we applied MASC to the case of Abruzzi region and used investments and employment as mediators. The choice of the mediators and the outcome is straightforward. Indeed, the policy itself is built to increase the

GDP in the lagging behind regions acting on investments and employment.

Using investments and employment as mediators we need to assume that the values of these variables one year is not influenced by the value of GDP in subsequent years. This assumption may be violated if firms base their production choices on expected values of GDP (as according to Keynesian theories) and the last coincide with the actual future values of GDP. Nonetheless, firms will probably base their expectations on the whole italian economy rather than on those of Abruzzi alone. There is no reason for the italian economy to drop drammatically in 2001. Moreover, the end of the policy affected directly the level of available resources for investments and employment. Compared with a drop in the available resources the impact of negative expectations is likely to be negligible.

Following Barone et al. (2016), we used as an outcome an index of GDP per-capita in real terms, set equal to 100 in 1995. This choice allows to build a synthetic unit for Abruzzi, notwithstanding the fact that its GDP is higher than those of any other region of Mezzogiorno area. Meanwhile, it still holds all the needed information on regions growth. Consistently with this choice of the outcome, we selected measures of the two mediators relevant for GDP trend of growth. For employment level, we used an index of the share of employed population, set equal to 100 in 1995. For investments, instead, we used simply investments level. While it is the trend of employment, rather than its level, that influence GDP trend we can't say the same for investments. Indeed, the level of investments directly influence GDP trend. This is why we indexed only the first mediator. In addition, we used the same covariates employed in Barone et al. (2016): the initial level of GDP per-capita, the past level of GDP per capita growth, a measure of human capital (the share of graduates among the population), population density, a measure of trade openness (export over GDP), the sectorial composition of value added (measured as the share of agricolture, industry and market services). Following the authors, we imposed the costraints on GDP per capita growth averaged for the 10 years before the intervention and on all the other covariates (including pre-treatment outcomes) averaged 3 years before the intervention. As for the pre-treatment constraints on the mediators, we imposed them on all pre-treatment levels of investments from 1981 until 1999 and on all pre-treatment values of the indexed share of employment from 1980 to 1999 (with the exclusion of the observation relative to 1995 since it takes value 100 for all units). The inclusion of all pre-treatment mediators values allowed us not ot include observables determining the mediator in the estimation of $Y^{0,0}$ (Botosaru and Ferman 2017).With respect to post-treatment constraints, for the estimation of the direct effect, we imposed them on all years from the last pre-treatment year until the last year before the year of interest. I.e. for the year of interest t' the post-treatment constraint were imposed on mediator value from 2000 until $t' - 1$.

Investments are unlikely to have a post-treatment confounder. Indeed, most of the transfers were used directly for investments or in the fight against unemployment. It is hardly believable that employment has an effect on investments. It is instead very likely the opposite. The reduction of investments may have an impact on employment level. Hence, investments may be a post-treatment confounder for employment. Unfortunately, we were not able to calculate the direct effect for employment in presence of post-treatment confounders as suggested in section 4.3.1. There was not enough variability in the data to satisfy the required post-treatment constraints on employment and on investments simultaneously. Nonetheless, as we will see, the indirect effect passing through investments was null while those passing through employment was not. This ensure us that none of the indirect effect through employment is due to investments. Indeed, if they were post-treatment confounders they would have had an impact on the outcome through employment and their estimated indirect effect would have not been null.

To determine the weigths to attribute to each constraint we used the cross-validation method proposed in Abadie, Diamond, et al. (2015). Nonetheless, we changed it to select the weigths that minimize the MSPE of both the outcome and the mediators in pre-treatment periods. For the estimation of the direct effects, instead, half of the weight was divided between pre-treatment constraints (following the total effect division) and half of it was divided equally between post-treatment constraints.

Data were kindly provided by Barone, David and De Blasio. We therefore refer to their work for details on their origin (Barone et al. 2016).

With regard to inference, the low number of control units do not allow us to use in-space placebo. Indeed, if the number of units to use for in-space placebo is too low it is likely to have under-rejection. Indeed, a single outlier is enough for the p-value to be too high to reject the null hypothesis. We partially solved this issue using in-time placebo. Even though the number of cases to use to build the distribution was still low, observations in-time should present lower variability since they refer to the same unit. Hence, the probability to have an outlier or a peculiar behaviour is lower. Following Gobillon and Magnac (2016), in post-treatment period we subtracted the estimated treatment impact from the outcome and from the mediators. Later on, we applied the MASC multiple times using all the years over the period 1990-1994 and 1996-1999 as intervention thresholds. We derived p-values from the estimated values as the rate of placebo estimations with a value of the treatment effect higher than those of the treated.

All the estimations were conduced using, on the software R, the package “Synth” (Abadie, Diamond, et al. 2011).

5.3. *Results*

5.3.1. Total Effect

In table 5.2 the average outcome, mediators and covariates over pre-treatment period for the treated unit, the synthetic unit and the Mezzogiorno area (with the exclusion of Abruzzi) are presented. It is possible to see that the synthetic unit is more similar to Abruzzi than the Mezzogiorno area average with respect to all the most relevant variables (i.e. the outcome and the mediators) and some of the less relevant ones.

Table 5.2: COMPARING THE PRE-TREATMENT AVERAGES OF TREATED AND SYNTHETIC UNITS AND MEZZOGIORNO AREA

Variable	Abruzzi	Synthetic Control	Mezzogiorno Average
Indexed GDP per capita	92.43	92.07	93.93
Investments	3626.6	3611.92	6591.65
Indexed Share of Employed	1.0102	1.0198	1.0413
GDP growth	0.0202	0.019	0.0174
Share of Graduates	0.0403	0.0359	0.0338
Population Density	116.01	105.88	165.66
Export over GDP	0.1053	0.0316	0.0494
VA share of Agricolture	0.0492	0.0602	0.0589
VA share of Industry	0.2914	0.2233	0.2375
VA share of Market Services	0.4379	0.4238	0.4229

NOTE: Average variables over the period 1980-2000. Mezzogiorno area include Molise, Puglia, Calabria, Basilicata, Sardinia, Sicily, Campania. Population density is defined for Km^2.

The weights among the donor pool were distributed in the following way: 0.2935 to Molise, 0.000001 to Campania, 0.000002 to Puglia, 0.000009 to Basilicata, 0.5169 to Calabria, 0.0000004 to Sicily and 0.1896 to Sardinia. In figures 5.3, 5.4 and 5.5 the comparison between the treated and the synthetic unit is presented with respect to the outcome and the two mediators.

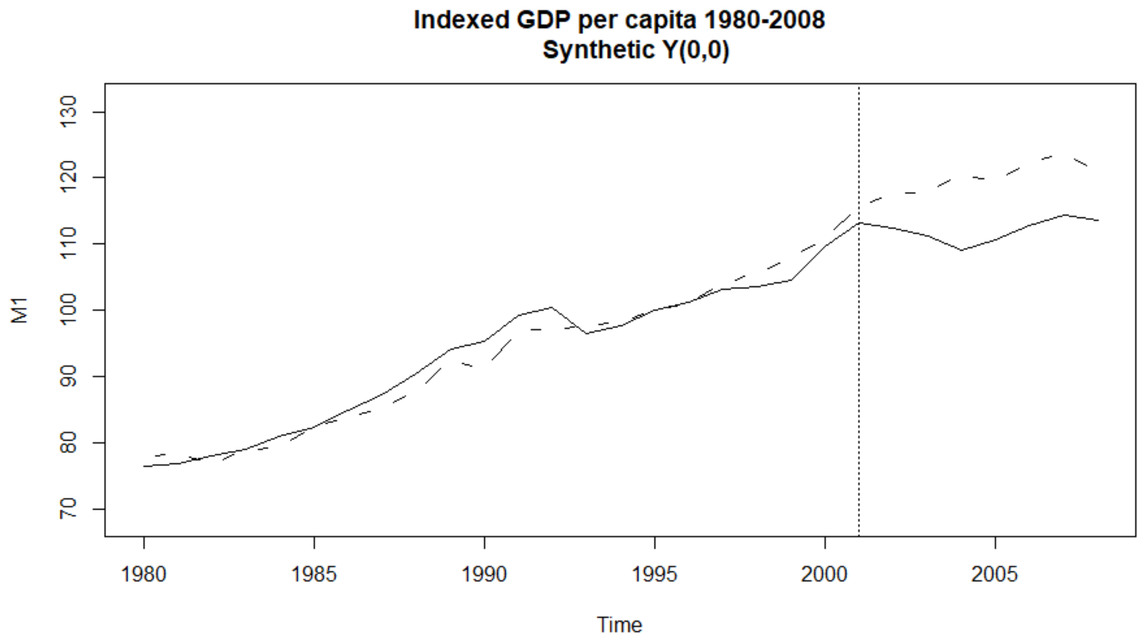


Figure 5.3: In the figure, the outcome of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period only, as it is required to estimate the total effect.

With respect to the outcome, it is possible to see that there is a good pre-treatment fit between the treated and the synthetic unit. After 2001 the two outcomes diverge, suggesting a negative effect of the treatment (in line with Barone et al. (2016) results).

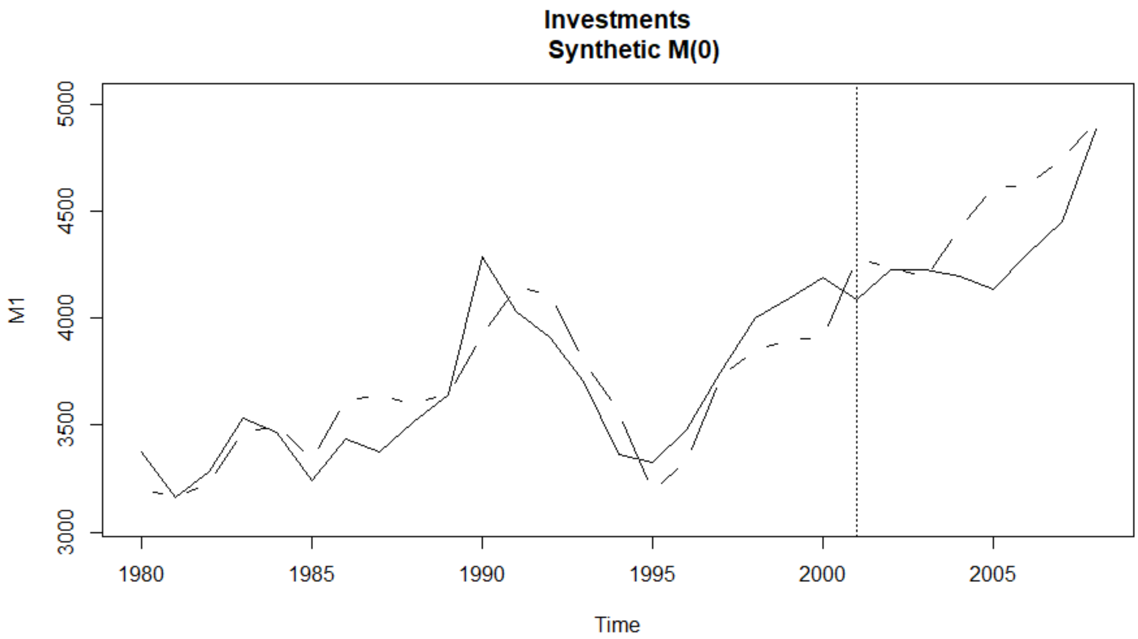


Figure 5.4: In the figure the investments level of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period only, as it is required to estimate the total effect.

The pre-treatment fit between treated and synthetic unit in terms of the two mediators is slightly worst, nonetheless it is still satisfying. The lines of the treated and the synthetic units diverge after the intervention for both mediators. The divergence is nonetheless stronger with respect to the indexed share of employment while it is more moderate for investments. In particular, the lines for the last converge in 2008. The evident divergence in terms of indexed share of employment starts with a little delay.

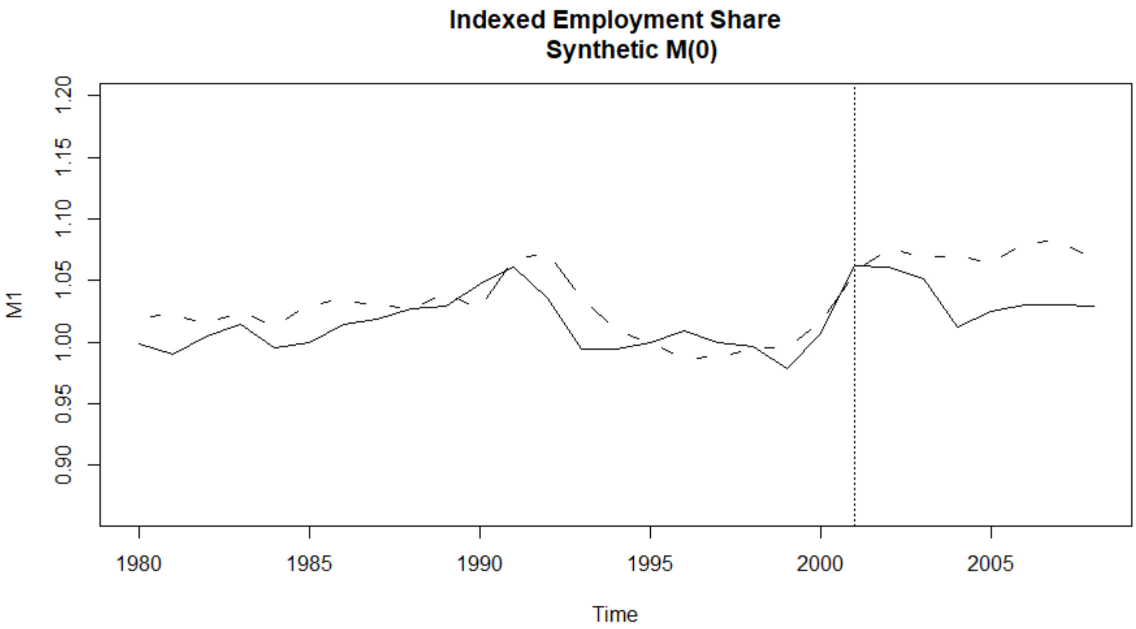


Figure 5.5: In the figure the indexed ratio of employment of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period only, as it is required to estimate the total effect.

This suggest the intervention reduced strongly the employment level and midly investments level. The quantitative results are represented in table 5.3.

Table 5.3: ESTIMATION OF THE TOTAL EFFECT OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Effect on		
	Outcome	Inv	Index Empl Share
2001	-2.42	-187.90	-0.001
2002	-5.06	-16.09	-0.018
2003	-6.79	38.11	-0.027
2004	-11.32	-23.58	-0.076
2005	-9.11	-476.60	-0.060
2006	-9.52	-319.66	-0.072
2007	-9.30	-294.97	-0.072
2008	-7.19	-33	-0.060

Outcome: Indexed GDP per capita
Pre-Treatment Average of the Outcome: 92.43

NOTE: Effect of the intervention on the outcome and the two mediators for each post-treatment year.

5.3.2. Direct and Indirect Effects

5.3.2.1. Investments

In table 5.4 the average outcome and investments over pre-treatment period and the average mediator over post-treatment period for the treated unit, the synthetic unit and the Mezzogiorno area (with the exclusion of Abruzzi) are presented (the weigths used to build these values for the synthetic unit are those obtained from the estimation of last post-treatment year effect, where all post-treament constraints are used). It is possible to see that the synthetic unit is more similar to the treated than the Mezzogiorno area average with respect to all the values. Moreover, the synthetic unit approximate very well the treated unit in both pre-treatment outcome and mediator and post-treatment mediator values.

Table 5.4: COMPARING THE PRE- AND POST-TREATMENT AVERAGES OF TREATED AND SYNTHETIC UNITS AND MEZZOGIORNO AREA

Variable	Abruzzi	Synthetic Control	Mezzogiorno Average
Indexed GDP per capita	92.43	92.23	93.93
Inv pre-treat	3626.6	3503.46	6591.65
Inv post-treat	4311.65	4332.89	7150.55

NOTE: Average variables over the period 1980-2000 and 2001-2008. Mezzogiorno area include Molise, Puglia, Calabria, Basilicata, Sardinia, Sicily, Campania.

The weigths given to the constraints were divided equally between pre- and post-treatment period. In the estimation of last post-treatment year the weights among the donor pool were distributed in the following way: 0.3303 to Molise, 0.000594 to Campania, 0.00474 to Puglia, 0.0000543 to Basilicata, 0.4435 to Calabria, 0.0004825 to Sicily and 0.2203 to Sardinia. Table 5.4 is confirmed by figure 5.6 and 5.7, were outcome and investments of the treated unit are compared with those of the synthetic. Pre-treatment fit is fairly good for both variables, just as the post-treatment fit for the mediator.

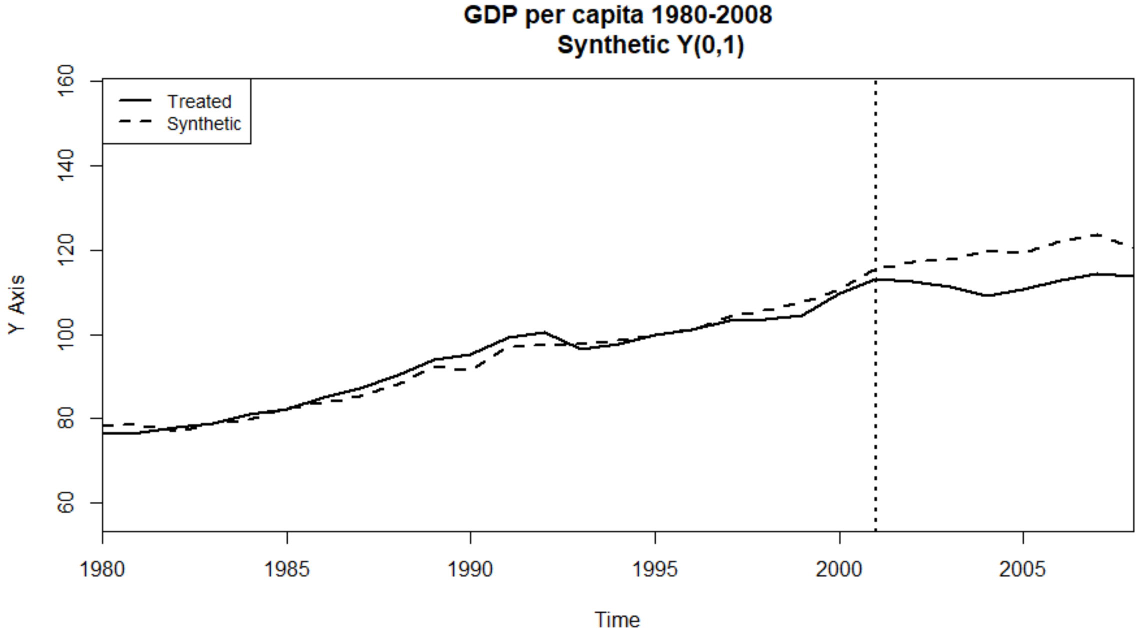


Figure 5.6: In the figure the outcome of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period and post-treatment constraints on investments. As it is required to estimate the direct and indirect effects using investments as mediator.

For the outcome the two lines diverge after intervention similarly to how they used to in the total effect estimation, suggesting a small portion of the effect on the outcome pass through a variation in investments.

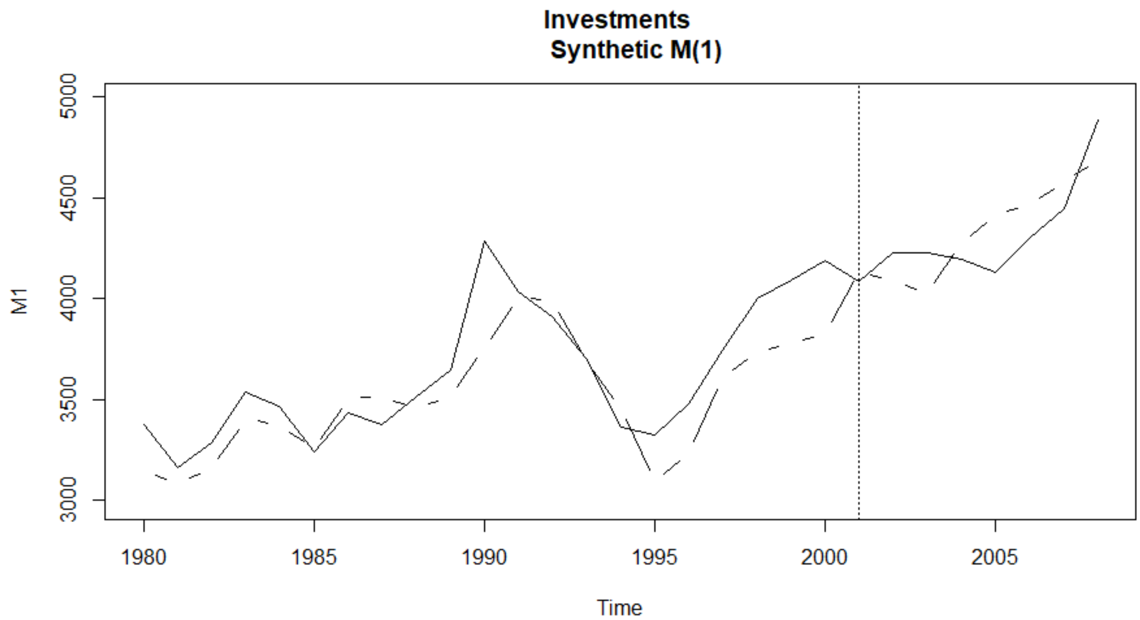


Figure 5.7: In the figure investments of the treated unit are compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period and post-treatment constraints on investments. As it is required to estimate the direct and indirect effects using investments as mediator.

In figure 5.8 the total effect, the direct effect when mediator's treatment is set to 1 and the indirect effect when treatment is set to 0 are represented. The indirect effect mediated by the investments is almost null. There are two possible explanations to this result: European SF may have a low impact on the outcome through investments or their effect through investments may be durable.

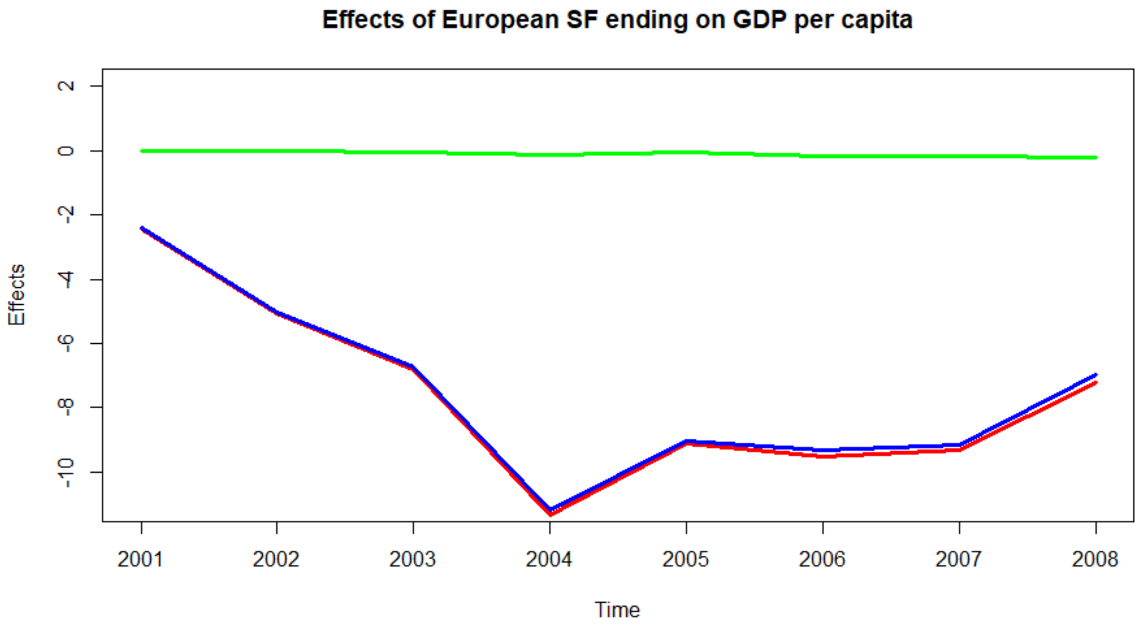


Figure 5.8: In the figure the total, direct and indirect effects are presented. Investment was used as mediator.

To choose with certainty between the two possible explainations we would need more information. Indeed, this result should be compared with mediation analysis on European SF impact. Unfortunately, to the best of our knowledge, such an analysis does not exist. We can nonetheless advance some hypothesis based on the literature (see 5.1.3) and the use of transfers in Abruzzi region (see 5.1.2). Past literature based on counterfactual analysis suggests European Structural Funds have no impact on Investments because they only induce a substitution of private investments with public investments (S. O. Becker et al. 2018). This result would support the first of our explainations. Nonetheless, this result relies on an analysis based on a huge group of regions and the effects may be highly heterogeneous. Hence the results for Abruzzi may be different. Moreover, transfers in Abruzzi region were mostly used for restorations and touristic infrastructures which are mostly public expenses. Hence, it is unlikely that in this context public investments substituted the private ones. Moreover, the reduction, although small, in investments level after intervention, suggests European SF had a positive impact on them and the positive impact of infrastructures (Di Giacinto et al. 2011) and in general investments on GDP growth is by now commonly accepted as a matter of fact. Hence, the first explaination is not satisfying. The fact that Structural Funds effect through investments is durable instead, is likely to be true. Indeed, when the investments end it is still possible to benefit from infrastructures and restorations, at least in a time period of 8 years (before other restorations are needed). In table 5.5 more detailed informations on the portion of effect mediated by investments is presented.

Table 5.5: ESTIMATION OF THE TOTAL, DIRECT AND INDIRECT EFFECTS OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Total	Direct	Direct in %	Indirect	Indirect in %
2001	-2.42	-2.42	99.76	-0.0057	0.24
2002	-5.06	-5.03	99.34	-0.0334	0.66
2003	-6.79	-6.73	99.17	-0.0564	0.83
2004	-11.32	-11.19	98.89	-0.1261	1.11
2005	-9.11	-9.03	99.16	-0.0765	0.84
2006	-9.52	-9.32	97.97	-0.1932	2.03
2007	-9.30	-9.13	98.26	-0.1622	1.74
2008	-7.19	-6.96	96.80	-0.2303	3.20

NOTE: Total, direct and indirect effects. The last two are expressed as well as a share of the total. Mediator: Investments.

5.3.2.2. Indexed Employment Share

In table 5.6 the average outcome and indexed employment share in pre-treatment and the same mediator in post-treatment periods are presented (again synthetic values are used exploiting the weigths from the last estimation where all post-treatment weights were used). The averages both in pre- and post-treatment periods between the treated and the synthetic units are fairly similar. Again, the weights for the constraints were equally divided between pre- and post-treatment constraints. In the last estimation, the weights were distributed in the following way among the donor pool: 0.0008 to Molise, 0.0044 to Campania, 0.9718 to Puglia, 0.000098 to Basilicata, 0.0198 to Calabria, 0.0009 to Sicily and 0.0013 to Sardinia.

Table 5.6: COMPARING THE PRE- AND POST-TREATMENT AVERAGES OF TREATED AND SYNTHETIC UNITS AND MEZZOGIORNO AREA

Variable	Abruzzi	Synthetic Control	Mezzogiorno Average
GDP per capita	92.43	94.27	93.93
Employment Share pre-treat	1.010	1.036	1.0401
Employment Share post-treat	1.0379	1.0570	1.0794

NOTE: All variables are indexed with respect to 1995. Average variables over the period 1980-2000 and 2001-2008. Mezzogiorno area include Molise, Puglia, Calabria, Basilicata, Sardinia, Sicily, Campania.

In figure 5.9 and 5.10 the outcome and the mediator lines of the treated and of the synthetic units are compared. From the first graph it is possible to see that pre-treatment fit for the outcome is slightly worst than before and the divergence between the two lines is significantly lower than when the total effect was estimated. This suggests the indirect effect mediated by employment share is significantly higher than those mediated by investments.

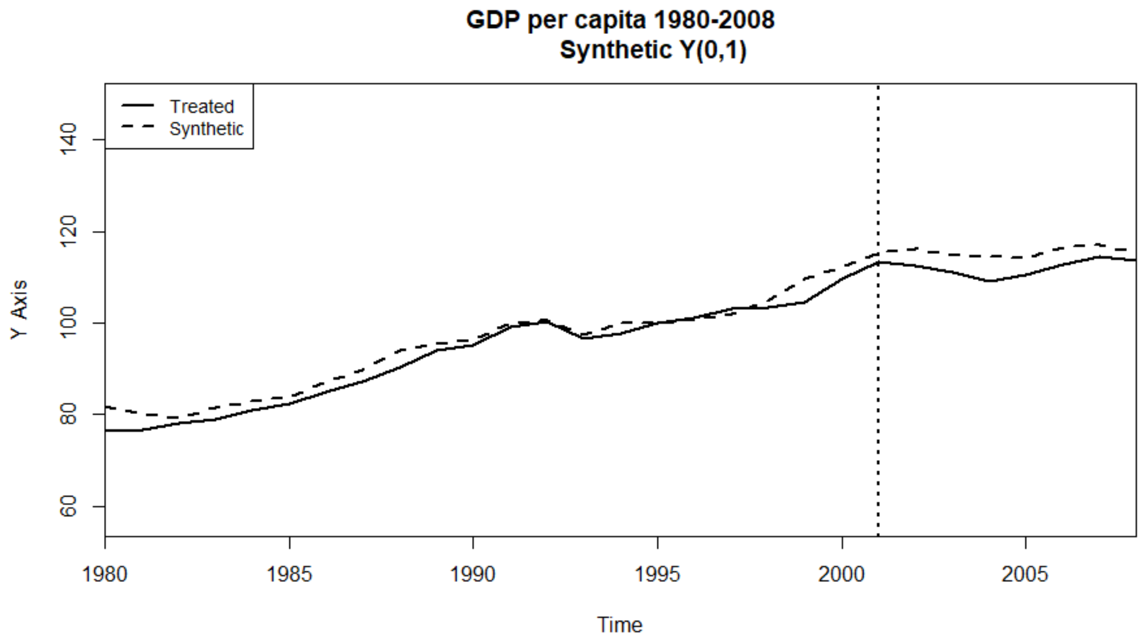


Figure 5.9: In the figure the outcome of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period and post-treatment constraints on indexed employment share. As it is required to estimate the direct and indirect effects using indexed employment share as mediator.

With regard to employment share, it is possible to see that we could not obtain perfect post-treatment fit. Indeed, synthetic unit post-treatment mediator values are slightly higher. This means the estimated direct effect is closer to the total ones than how it is in reality. Hence, the estimated indirect effect can be seen as a lower bound of the real one. Unfortunately, a better post-treatment fit was not possible because we would have had to pay the price of a worst pre-treatment fit for both the outcome and the mediator.

Figure 5.10: In the figure the indexed employment share of the treated unit is compared with those of the synthetic unit across years. The synthetic unit was built using constraints on pre-treatment period and post-treatment constraints on indexed employment share. As it is required to estimate the direct and indirect effects using indexed employment share as mediator.

The high indirect effect mediated by employment share is confirmed by figure 5.11 were it is also possible to see that the indirect effect mediated by employment arrives with a little delay after the intervention. This is probably due to the slow reaction of labour market (S. O. Becker et al. 2010, S. O. Becker et al. 2018).

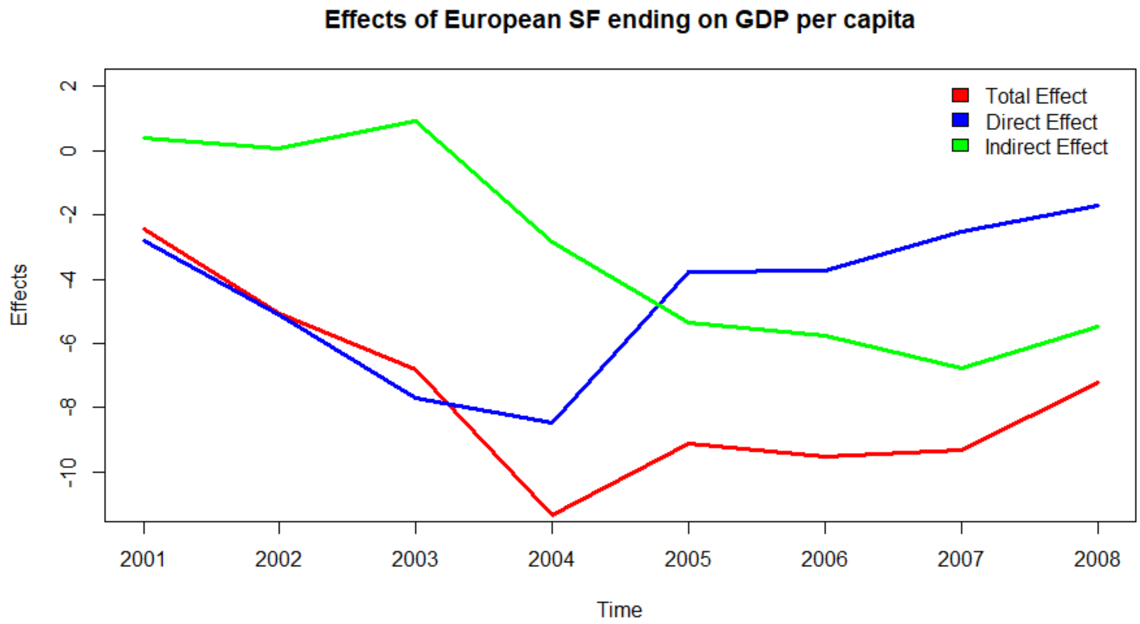


Figure 5.11: In the figure the total, direct and indirect effects are presented. Indexed employment share was used as mediator.

As mentioned before, the fact that the indirect effect through investment has no significant impact allow us to attribute this impact entirely to the employment mediation channel. The indirect effect when treatment is set to 0 is reported in table 5.7 together with the total and the direct effects estimated and the direct and indirect effects expressed as a portion of the total (obviously, we expect the negative indirect effect to be non-significant).

Table 5.7: ESTIMATION OF THE TOTAL AND DIRECT EFFECTS OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Total	Direct	Direct in %	Indirect	Indirect in %
2001	-2.42	-2.80	115.45	0.37	-15.45
2002	-5.06	-5.13	101.28	0.07	-1.28
2003	-6.79	-7.70	113.42	0.91	-13.42
2004	-11.32	-8.46	74.77	-2.85	25.23
2005	-9.11	-3.77	41.42	-5.34	58.58
2006	-9.52	-3.75	39.42	-5.76	60.58
2007	-9.30	-2.53	27.20	-6.77	72.80
2008	-7.19	-1.73	24.08	-5.46	75.92

NOTE: Total effect, direct and indirect effects. The last two are expressed as well as a share of the total. The possible range of the share of indirect effect is built according to the indirect effect of the treatment mediated by investments. Mediator: Indexed Employment Share.

We can conclude the portion of effect due to European SF ending mediated by employment is significantly higher than those mediated by investments. This suggests the transfers used for investments have a higher longevity than those used to boost employment.

5.3.3. Inference and Robustness Checks

5.3.3.1. Inference

As mentioned before, we conduced in-time placebo. In table 5.8 the estimated effect for Abruzzi region over the years 2001-2008 is presented together with the p-values derived from the placebo test.

Table 5.8: TOTAL EFFECT OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Total Effect	p-value
2001	-2.42	0.1
2002	-5.06	0
2003	-6.79	0
2004	-11.32	0
2005	-9.11	0
2006	-9.52	0.1
2007	-9.30	0.1
2008	-7.19	0.4

NOTE: P-values were built using in-time placebo.

It is possible to see that the total effect started to be significant, at a 95% level, one year after treatment and it is significant until 2005.
In table 5.9, the direct effect and the indirect effect mediated by investments are presented together with their p-values derived from in-time placebo.

Table 5.9: DIRECT AND INDIRECT EFFECTS OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Direct Effect	p-value	Indirect Effect	p-value
2001	-2.42	0.1	-0.006	0.9
2002	-5.03	0	-0.033	0.3
2003	-6.73	0	-0.056	0.3
2004	-11.19	0	-0.126	0
2005	-9.03	0	-0.076	0.1
2006	-9.32	0.1	-0.193	0.1
2007	-9.13	0.1	-0.162	0
2008	-6.96	0.4	-0.23	0.4

NOTE: P-values were built using in-time placebo. Mediator: investments.

As the total effect, the direct effect is significant only from 2002 until 2005. The indirect effect, instead, is non-significant for almost all years under study (and its significance in some years is temporary). Hence, we can conclude the indirect effect passing through investments is null.

Different results were found using the indexed share of employed population as mediator. The direct effect non-mediated by employment share was non-significant in the first year after the intervention and since 2005. On the other hand, the indirect effect was strong and significant only since 2004 and non-significant in 2008. The results of the in-time placebo for these two effects are presented in table 5.10.

Table 5.10: DIRECT AND INDIRECT EFFECTS OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION

Year	Direct Effect	p-value	Indirect Effect	p-value
2001	-2.8	0.1	0.37	0.4
2002	-5.13	0	0.07	0.9
2003	-7.7	0	0.91	0.3
2004	-8.46	0	-2.85	0
2005	-3.77	0.4	-5.34	0
2006	-3.75	0.6	-5.76	0
2007	-2.53	0.6	-6.77	0
2008	-1.73	0.9	-5.46	0.1

NOTE: P-values were built using in-time placebo. Mediator: indexed share of employed.

5.3.3.2. Robustness Check: Spillover Effects

The studied intervention may have had spillover effects in the neighbouring regions. As an example, the reduction of employment in the treated region may have induced the population to migrate in the neighbourings, affecting their employment share. Equally, the reduction in investments in the treated region may have induced firms to move in the neighbouring ones. If one of the neighbouring regions is used in the donor pool the presence of spillover effects may under- or over-estimate the real treatment effect. In our empirical application, the only region in the donor pool having its border in common with Abruzzi is Molise. Hence, we tried to exclude Molise from the donor pool, to verify whether the estimated effects changed, suggesting the presence of spillover effects. In figure 5.12 the effects estimated including and excluding Molise from the donor pool are presented. In the first two graphs we compare the effects when investments is used as mediator.

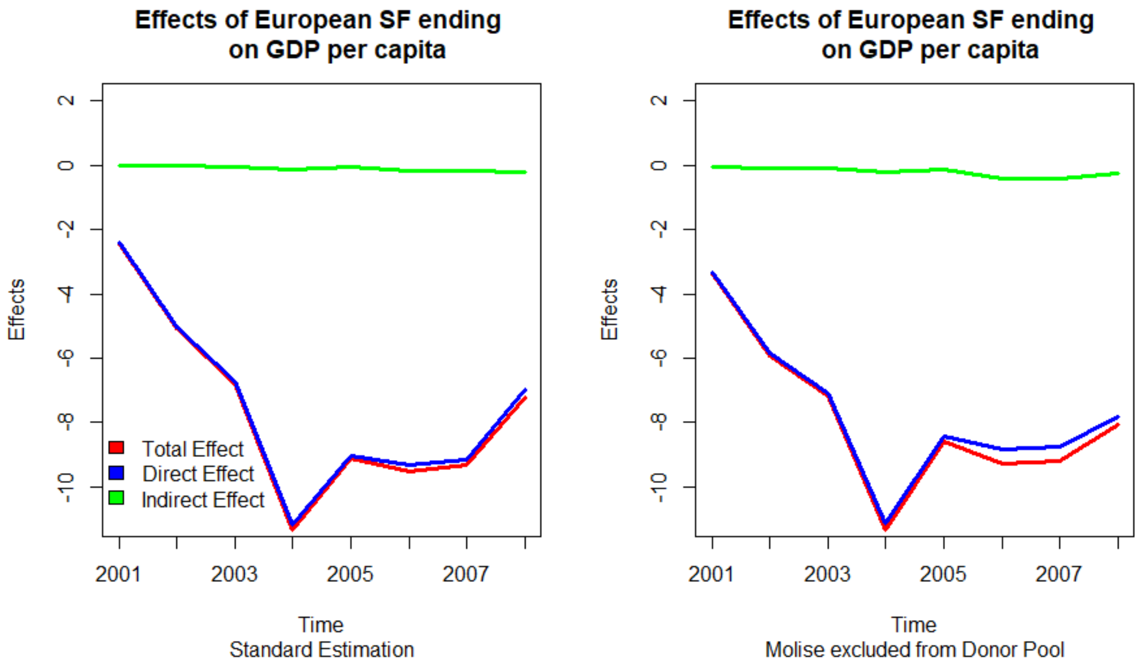


Figure 5.12: In the figure on the left the effects are estimated using the whole donor pool. In the figure on the right the effects are estimated excluding Molise from the donor pool. Mediator: Investments.

It is possible to see that the estimated effects slightly change when Molise region is excluded. Nevertheless, the indirect effect is still negligible with respect to the direct and total ones. As visible from figure 5.13, the direct and indirect effect when employment share is used as a mediator are even more robust to Molise exclusion.

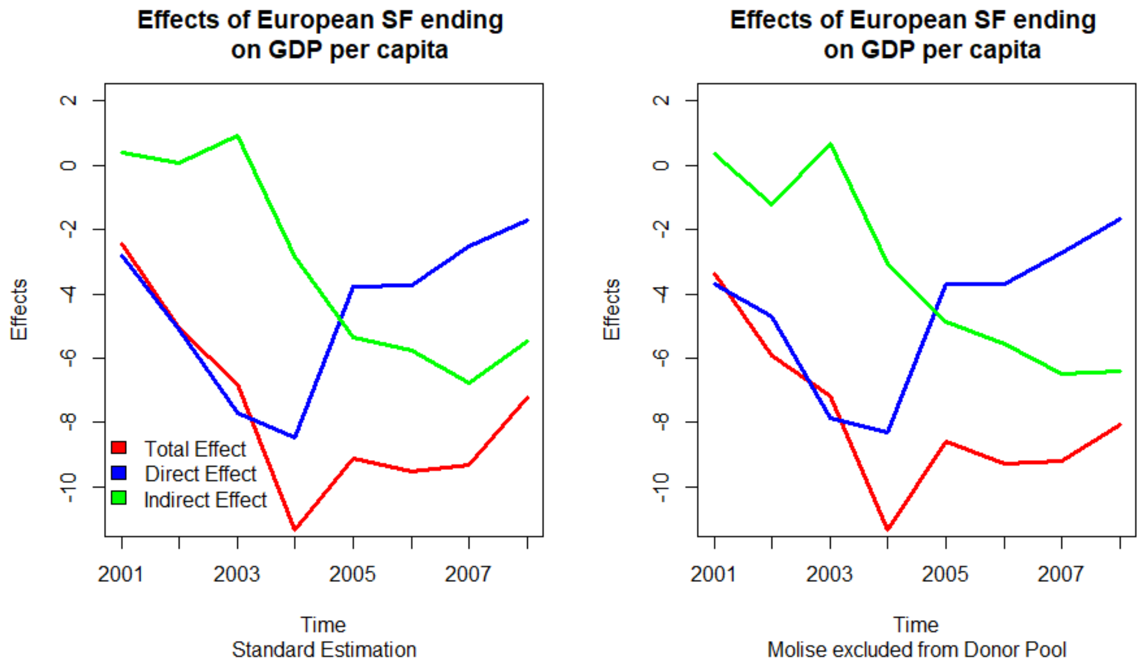


Figure 5.13: In the figure on the left the effects are estimated using the whole donor pool. In the figure on the right the effects are estimated excluding Molise from the donor pool. Mediator: Indexed employment share.

We can therefore conclude there are no spillover effects, or their aggregate effect is null.

5.3.3.3. Robustness Check: Changements in the Donor Pool

When Synthetic Control Method is implemented, there is the risk that a region in the donor pool having a peculiar behaviour drags the synthetic unit towards values of the outcome that are different from those treated unit would have had in absence of treatment. The same can happen when the MASC is implemented. To verify this is not the case, we repeated the estimations several times, excluding one by one all the units in the donor pool. The results for total effect estimation are presented in figure 5.14.

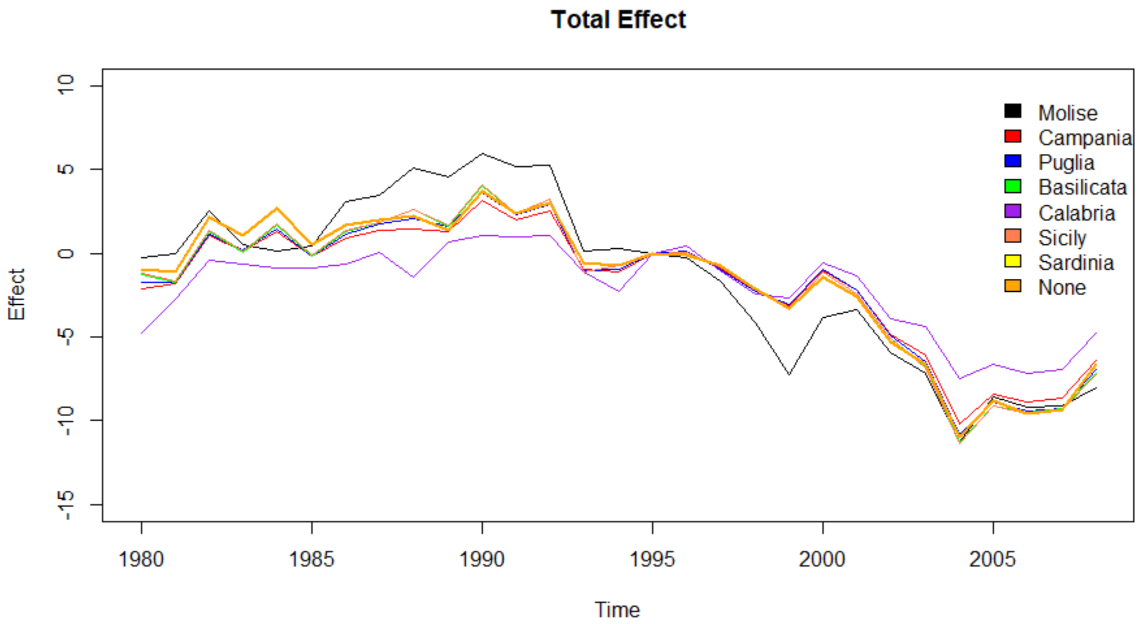


Figure 5.14: Each line represents the total effect estimated excluding the corresponding region from the donor pool.

It is possible to see that the estimation is fairly robust to the exclusion of all the units in the donor pool with the exception of Puglia and Calabria. Nonetheless, for those regions, the pre-treatment fit is very bad as well. Moreover, even though the magnitude of the effect changes for the esclusion of those regions, the effect of the intervention remains negative.

In figure 5.15 and 5.16 the same results for the direct effect with respect to the two mediators are presented (we do not present the same graph for the indirect effects given that if the total and the direct effects are robust they will be robust by difference).

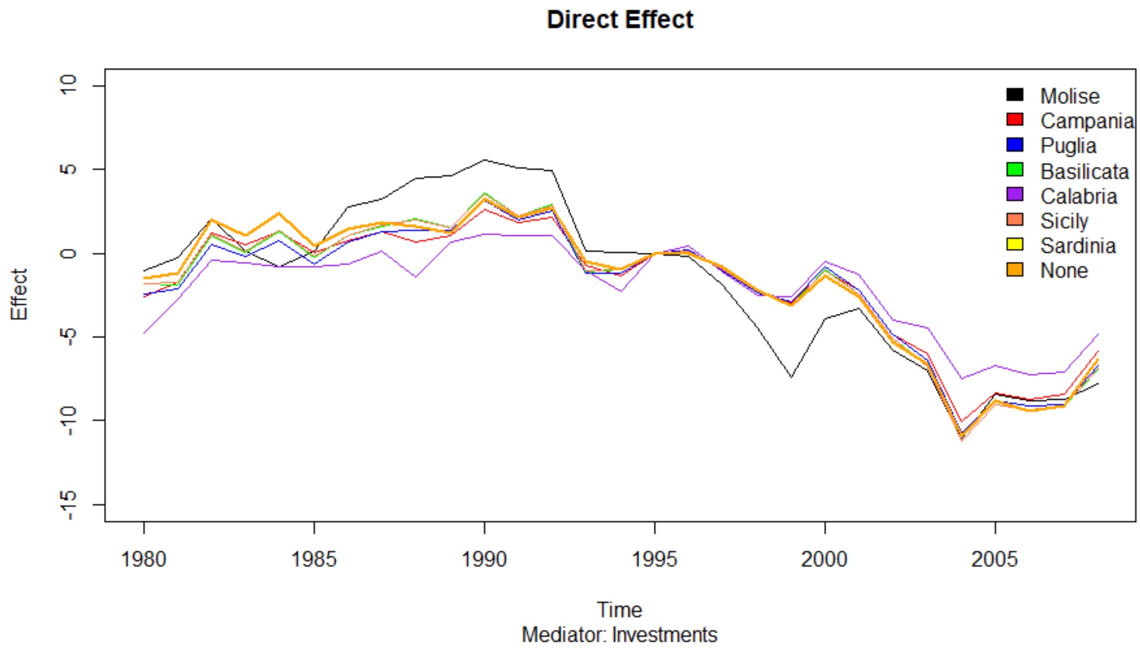


Figure 5.15: Each line represents the direct effect estimated excluding the corresponding region from the donor pool. Mediator: Investments.

It is possible to see that again, when investments level is used as mediator, the estimation is robust for the exclusion of all regions in the donor pool with the exception of Puglia and Calabria for which the pre-treatment fit is nonetheless really bad. When the indexed share of unemployment is used, the estimation is robust to exclusion of all regions but Puglia. Nonetheless, as before, this region has a bad pre-treatment fit as well. The estimation results when Calabria is excluded from the donor pool as well are slighlty different. Nonetheless, the difference is only in the magnitude, while the trend of the effect does not change.

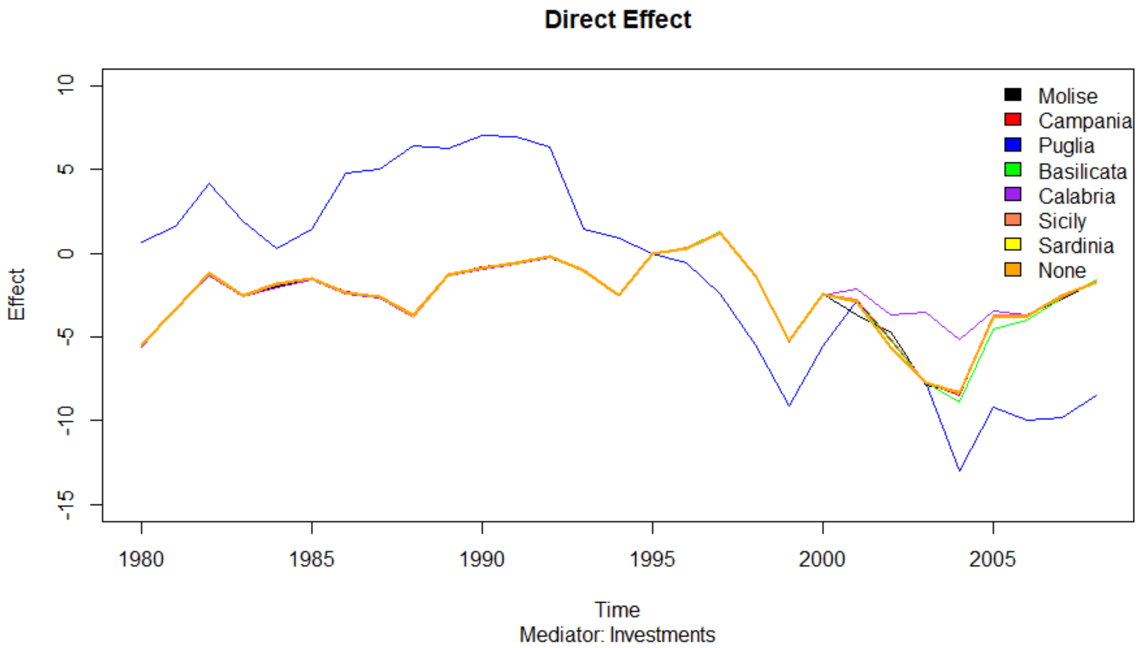


Figure 5.16: Each line represents the direct effect estimated excluding the corresponding region from the donor pool. Mediator: Indexed employment share.

5.3.3.4. Robustness Check: Different Mediator Lags

As mentioned before, we imposed the constraints on the mediators lagged by one year. I.e. we imposed them until 1999 in pre-treatment period and until $t' - 1$ in post-treatment period when estimating the effect at time t'. This allows the mediator to have a delayed effect on the outcome. Nonetheless, we do not know exactly the amount of time necessary for the mediator to have an impact on the outcome. The one-year lag was an arbitrary choice led by common sense. Hence, as a further robustness check, we repeated the estimation increasing the number of lags, to check whether the results changed. In table 5.11 the estimated effects for one-, two- and three-years lag are presented.

The estimation of the total effect is fairly robust to variations in the number of lags on the mediator. This, and the fact that the direct effects get closer to the total ones when the number of lags in the mediators constraints are augmented, suggests that the reduction of post-treatment constraints (due to the higher number of lags) do

Table 5.11: ESTIMATION OF THE TOTAL AND DIRECT EFFECTS OF EUROPEAN STRUCTURAL FUNDS ENDING IN ABRUZZI REGION WRT DIFFERENT MEDIATORS LAGS

Year	Total Effect			Direct Effect Investments			Direct Effect Employment		
	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag	1-y lag	2-y lag	3-y lag
2001	-2.42	-2.18	-2.23	-2.42	-2.18	-2.24	-2.8	-2.83	-2.54
2002	-5.06	-4.84	-4.91	-5.03	-4.82	-4.88	-5.13	-5.68	-5.66
2003	-6.79	-6.16	-6.23	-6.73	-6.12	-6.18	-7.7	-7.43	-7.6
2004	-11.32	-10.3	-10.4	-11.19	-10.24	-10.26	-8.46	-12.75	-12.28
2005	-9.11	-8.44	-8.5	-9.03	-8.36	-8.4	-3.77	-6.49	-9.89
2006	-9.52	-8.95	-9.05	-9.32	-8.9	-8.88	-3.75	-4.95	-6.54
2007	-9.3	-8.72	-8.82	-9.13	-8.66	-8.64	-2.53	-3.15	-3.82
2008	-7.19	-6.48	-6.52	-6.96	-6.13	-6.21	-1.73	-1.77	-2.03

NOTE: Total and direct effects estimated with 1, 2 and 3 years lags in the mediators.

not allow to distinguish properly between the direct and the total effect. Hence, our choice to use a one-year lag seems correct.

6. Conclusions

In this dissertation we showed that, in causal inference frameworks, it is possible to go beyond the estimation of the total effect of a policy following two different paths: exploiting the characteristics of the policy and the context it is set in, or introducing new methodologies that can be applied to other empirical investigations. We followed the first path to go beyond the estimation of Law 407/90 effect. We started showing that the policy had a significant and strong intention-to-treatment effect on eligible people with approximately 24 months of unemployment. Indeed, its implementation increased their likelihood of being hired by 36%. Later on, we exploited policy characteristics to show that it did not present negative side effects such as displacement effect and post-poned hiring effect. This is probably due to the fact that employers prefer to hire the chosen workers immediately and to Italian firms' characteristics.

Finally we exploited the context the policy was set in to show that a generalization of the incentives to all unemployed would strongly penalize the vulnerable group of LTU. Indeed, the intention-to-treatment effect of Law 190, parcel out of the component due to the limited implementation period of the policy, is insignificant. This, and Centra and Gualtieri (2016) and Sestito and Viviano (2016) results, suggest that the generalization of the policy re-allocates the benefits in favor of others, more "desirable", groups of unemployed people. Therefore, in order to be effective, with respect to vulnerable groups of unemployed workers, a policy based on hiring subsidies should lower their relative labor costs rather than their absolute ones. This would avoid benefit redistribution.

We furthermore provided evidence that thanks to the peak in hiring of LTU in December 2015, probably due to the limited duration of the policy, Law 190 incentives had a positive and significant effect on LTU hiring.

We followed the second path introducing a new methodology called Mediation Analysis Synthetic Control (MASC). This method combines Synthetic Control

Method (Abadie and Gardeazabal 2003, Abadie, Diamond, et al. 2010, Abadie, Diamond, et al. 2015) with Mediation Analysis approach allowing to investigate on direct and indirect effects in frameworks with selection on unobservables and a low number of treated and control units. This method is very intuitive and easy to implement (i.e. public available SCM algorithms can be employed) and it can be used even in presence of post-treatment confounders. Even though introduced for the procedure presented in Abadie, Diamond, et al. (2010) and Abadie, Diamond, et al. (2015), it can be easily extended to new approaches such as Athey, Bayati, et al. (2017), Xu (2017), Kreif et al. (2016) and Doudchenko and Imbens (2017). On the other end, the estimation of some of the parameters requires the presence of multiple treated and additional restrictions on direct effect functional form. Moreover, a low variability in the data may not allow to estimate all the parameters properly.

We applied MASC to the case of European SF ending in Abruzzi region, to verify to what extent the impact of this intervention on indexed GDP per capita passed through a variation of investments or through a variation of indexed employment share. We showed that most of the effect passed through a reduction of employment share, even though with a small delay after intervention. No effect instead, went through investments reduction. This may be due to the nature of the investments implemented in Abruzzi region. The results suggest more effort should be committed to increase the longevity of European Structural Funds impact on employment and employment impact on GDP per capita growth. Moreover, it suggests that, to reach a major longevity of SF effect using today tools, transfers should be concentrated on investments rather than employment.

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

— (2011). "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies". In: *Journal of Statistical Software* 42.13, pp. 1–17.

— (2015). "Comparative Politics and the Synthetic Control Method". In: *American Journal of Political Science* 59.2, pp. 495–510.

Abadie, A. and J. Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country". In: *The American Economic Review* 93.1, pp. 113–132.

Abruzzo (2001). "Documento Unico di Programmazione. Obiettivo 2 (2000-2006)". In: *Regione*.

Adhikari, B. (2015). "When Does Introducing a Value-Added Tax Increase Economic Efficiency? Evidence from the Synthetic Control Method". In: *Tulane University Working Paper* 1524.

Anastasia, B., L. Bertazzon, M. Gambuzza, and M. Rasera (2016). "GRAMMATICA DELLE COMUNICAZIONI OBBLIGATORIE /4: GUIDA AI CONFRONTI CON LE ALTRE FONTI STATISTICHE SUL MERCATO DEL LAVORO". In: *Veneto Lavoro: Osservatorio & Ricerca* 206.

Anastasia, B., A. Giraldo, and A. Paggiaro (2012). "L'Effetto degli Incentivi alle Assunzioni e alle Trasformazioni. Prime Evidenze per il Veneto". In: *POLITICA ECONOMICA* XXVIII.2.

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2017). "Matrix Completion Methods for Causal Panel Data Models". In: *https://arxiv.org/abs/1710.10251*.

Athey, S. and G. W. Imbens (2017). "The State of Applied Econometrics - Causality and Policy Evaluation". In: *Journal of Economic Perspectives* 31.2, pp. 3–32.

Ayllòn, S. (2013). "Unemployment persistence: not only stigma but discouragement too". In: *Applied Economics Letters* 20.1, pp. 67–71.

Baert, S. and D. Verhaest (2014). “Unemployment or Overeducation: Which is a Worse Signal to Employers?” In: *IZA Discussion Paper* 8312.

Baron, R. M. and D. A. Kenny (1986). “The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations”. In: *Journal of Personality and Social Psychology* 51.6, pp. 1173–1182.

Barone, G., F. David, and G. de Blasio (2016). “Boulevard of broken dreams. The end of EU funding (1997: Abruzzi, Italy)”. In: *Regional Science and Urban Economics* 60, pp. 31–38.

Becker, M., S. Klobner, and G. Pfeifer (2018). “Cross-Validating Synthetic Controls”. In: *Economics Bulletin* 38.1, pp. 603–609.

Becker, S. O., P. H. Egger, and M. von Ehrlich (2010). “Going NUTS: The effect of EU Structural Funds on regional performance”. In: *Journal of Public Economics* 94.9-10, pp. 578–590.

— (2012). “Too much of a good thing? On the growth effects of the EU’s regional policy”. In: *European Economic Review* 56.4, pp. 648–668.

— (2018). “Effects of EU Regional Policy: 1989-2013”. In: *Regional Science and Urban Economics* 69, pp. 143–152.

Berg, van den (2001). “Duration Models: Specification, Identification, and Multiple Durations”. In: *Handbook of Econometrics* 5.5.

Bernhard, S., H. Gartner, and G. Stephan (2008). “Wage subsidies for needy jobseekers and their effect on individual labour market outcomes after the German reforms”. In: *IAB Discussion Paper* 21.

Beugelsdijk, M. and S. C. W. Eijffinger (2005). “The Effectiveness of Structural Policy in the European Union: An Empirical Analysis for the EU-15 in 1995-2001”. In: *Journal of Common Market Studies* 43.1, pp. 37–51.

Biewen, M. and S. Steffes (2010). “Unemployment Persistence: Is There Evidence for Stigma Effects?” In: *Economics Letters* 106.3, pp. 188–190.

Blundell, R., M. C. Dias, C. Meghir, and J. Van Reenen (2004). “Evaluating the Employment Impact of a Mandatory Job Search Program”. In: *Journal of the European Economic Association* 2.4, pp. 569–606.

Boldrin, M. and F. Canova (2001). “Inequality and Convergence in Europe’s Regions: Reconsidering European Regional Policies”. In: *Economic Policy* 16.32, pp. 206–253.

Boockmann, B., A. Ammermueller, T. Zwick, and M. F. Maier (2007). “Do hiring subsidies reduce unemployment among the elderly? Evidence from two natural experiments”. In: *ZEW Discussion Papers* 001.

Boone, J. and J. C. van Ours (2004). “Effective active labor market policies”. In: *IZA Discussion Paper* 1335.

Botosaru, I. and B. Ferman (2017). “On the Role of Covariates in the Synthetic Control Method”. In: *MPRA Working Paper* 80796.

Brown, A. J. G. (2015). “Can Hiring Subsidies Benefit the Unemployed?” In: *IZA world of Labor* 163.

Brown, A. J. G. and J. Koettl (2012). “Active Labor Market Programs: Employment Gain or Fiscal Drain?” In: *IZA Journal of Labor Economics* 4.1, pp. 1–36.

Bucher, Anne (2010). “Impact of hiring subsidies targeted at the long-term unemployed on the low-skilled labor market: The French experience”. In: *Economic Modelling* 27.2, pp. 553–565.

Calmfors, L., A. Forslund, and M. Hemström (2002). “Does active labour market policy work? Lessons from the Swedish experiences”. In: *CESifo Working Paper* 675.

Cappelen, A., F. Castellacci, J. Fagerberg, and B. Verspagen (2003). “The Impact of EU Regional Support on Growth and Convergence in the European Union”. In: *Journal of Common Market Studies* 41, pp. 621–644.

Carling, K. and K. Richardson (2004). “The relative efficiency of labor market programs: Swedish experience from the 1990s”. In: *Elsevier Labour Economics* 11.3, pp. 335–354.

Cattaneo, M. D., B. R. Frandsen, and R. Titiunik (2015). “Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate”. In: *Journal of Causal Inference* 3.1, pp. 1–24.

Centra, M. and V. Gualtieri (2016). “Incentivi al lavoro permanente e contratto a tutele crescenti: una stima dell’impatto sulle nuove assunzioni nel 2015”. In: *Paper*

per la IX Conferenza ESPAnet Italia "Modelli di welfare e modelli di capitalismo. Le sfide per lo sviluppo socio-economico in Italia e in Europa".

Cerqua, A. and G. Pellegrini (2018). “Are we spending too much to grow? The case of Structural Funds”. In: *Journal of Regional Science* 58.3, pp. 535–563.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2018). “An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls”. In: *https://arxiv.org/abs/171*

Cochran, W. G. (1953). “Matching in Analytical Studies”. In: *American Journal of Public Health* 43, pp. 684–691.

Cochran, W. G. and D. B. Rubin (1973). “Controlling bias in observational studies: a review”. In: *Sankhya: The Indian Journal of Statisics, Series A* 35.4, pp. 417–446.

Cockx, B., B. Van der Linden, and A. Karaa (1998). “Active Labour Market Policies and Job Tenure”. In: *Oxford Economic Papers* 50.4, pp. 685–708.

Commission, European (1997). “Regional Development Studies - The Impact of Structural Policies on Economic and Social Cohesion in the Union 1989-99”. In: *Office for Official Publications of the European Commission, Luxembourg*.

Dall’Erba, S. and F. Fang (2017). “Meta-analysis of the impact of European Union Structural Funds on regional growth”. In: *Regional Studies* 51.6, pp. 822–832.

Dall’erba, S. and J. Le Gallo (2008). “Regional Convergence and the Impact of Structural Funds over 1989-1999: a Spatial Econometric Analysis”. In: *Papers in Regional Science* 87.2, pp. 219–244.

Deuchert, E., M. Huber, and M. Schelker (2018). “Direct and Indirect Effects Based on Difference-in-Differences With an Application to Political Preferences Following the Vietnam Draft Lottery”. In: *Journal of Business & Economic Statistics* 0.0.

Di Giacinto, V., G. Micucci, and P. Montanaro (2011). “L’impatto macroeconomico delle infrastrutture: una rassegna della letteratura ed un’analisi empirica per l’Italia”. In: *Le infrastrutture in Italia: dotazione, programmazione, realizzazione, Workshops and Conferences* 7.

Doudchenko, N. and G. W. Imbens (2017). “Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis”. In: *https://arxiv.org/abs/1610.07748*.

Duell, N., L. Thurau, and T. Vetter (2016). "Long-term Unemployment in the EU: Trends and Policies". In: *Bertelsmann Stiftung, Economix Research & Consulting*.

Ederveen, S., H. L. F. de Groot, and R. Nahuis (2006). "Fertile Soil for Structural Funds? A Panel Data Analysis of the Conditional Effectiveness of European Cohesion Policy". In: *Kyklos* 59.1, pp. 17–42.

Eppel, R. and H. Mahringer (2013). "Do wage subsidies work in boosting economic inclusion? Evidence on effect heterogeneity in Austria". In: *WIFO Working Papers* 456.

Esposti, R. and S. Bussoletti (2008). "Impact of Objective 1 Funds on Regional Growth Convergence in the European Union: A Panel-data Approach". In: *Regional Studies* 42.2, pp. 159–173.

Farooq, A. and A. D. Kugler (2015). "What factors contributed to changes in employment during and after Great Recession?" In: *IZA Journal of Labor Policy* 4.1, pp. 1–28.

Ferman, B. and C. Pinto (2017). "Placebo Tests for Synthetic Controls". In: *MPRA Working Paper* 78079.

Firpo, S. and V. Possebom (2017). "Synthetic control method: Inference, sensitivity analysis and confidence sets". In: *Working Paper*.

Forslund, A., P. Johansson, and L. Lindqvist (2004). "Employment Subsidies - A Fast Lane From Unemployment to Work?" In: *IFAU Working Paper* 18.

Freitas, M. L. de, F. Pereira, and F. Torres (2003). "Convergence ammong EU regions, 1990-2001. Quality of national institutions and "Objective 1" status". In: *Intereconomics* 38.5, pp. 270–275.

Frölich, M. and M. Huber (2017). "Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables". In: *Journal of the Royal Statistical Society Series B* 79.5, pp. 1645–1666.

Gallop, R., D. Small, J. Y. Lin, M. R. Elliot, M. Joffe, and T. Ten Have (2009). "Mediation Analysis with Principal Stratification". In: *Statistics in Medicine* 28.7, pp. 1108–1130.

Gardeazabal, J. and A. Vega-Bayo (2017). "An Empirical Comparison Between the Synthetic Control Method and HSIAO et al.'s Panel Data Approach to Program Evaluation". In: *Journal of Applied Econometrics* 32.5, pp. 983–1002.

Giua, M. (2017). "Spatial discontinuity for the impact assessment of the EU regional policy. The case of Italian objective 1 regions". In: *Journal of Regional Science* 57.1, pp. 109–131.

Gobillon, L. and T. Magnac (2016). "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls". In: *Reviews of Economics and Statistics* 98.3, pp. 535–551.

Hagen, T. and P. Mohl (2008). "Which is the right dose of EU Cohesion Policy for economic growth?" In: *ZEW-Centre for European Economic Research Discussion Paper* 08-104.

Ham, J. C., C. Swenson, I. Ayşe, and H. Song (2011). "Government programs can improve local labor markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Community". In: *Journal of Public Economics* 95.7, pp. 779–797.

Hamersma, S. (2008). "The Effects of an Employer Subsidy on Employment Outcomes: A Study of the Work Opportunity and Welfare-to-Work Tax Credits Abstract". In: *Journal of Policy Analysis and Management* 27.3, pp. 498–520.

Hanson, A. and S. Rohlin (2013). "Do Spatially Targeted Redevelopment Programs Spillover?" In: *Regional Science and Urban Economics* 43.1, pp. 86–100.

Heckman, J. J. and G. Borjas (1980). "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence". In: *Economica* 47.187, pp. 247–283.

Holland, P. W. (1986). "Statistics and Causal Inference". In: *Journal of the American Statistical Association* 81.396, pp. 945–960.

— (1988). "Causal inference, path analysis, and recursive structural equation models". In: *Sociology an Methodological* 18, pp. 449–484.

Hsiao, C., H. Steve Ching, and S. Ki Wan (2012). "A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China". In: *Journal of Applied Econometrics* 27.5, pp. 705–740.

Hsu, Y. C., M. Huber, and T. C. Lai (2017). "Nonparametric estimation of natural direct and indirect effects based on inverse probability weighting". In: *FSES Working Papers* 482.

Huber, M. (2014). "Identifying causal mechanisms in experiments (primarily) based on inverse probability weighting". In: *Journal of Applied Econometrics* 29.6, pp. 920–943.

— (2016). "Disentangling policy effects into causal channels". In: *IZA World of Labor*.

Huber, M., M. Lechner, and G. Mellace (2017). "Why Do Tougher Caseworkers Increase Employment? The Role of Program Assignment as a Causal Mechanism". In: *Review of Economics and Statistics* 99.1, pp. 180–183.

Huttunen, K., J. Pirttilä, and R. Uusitalo (2013). "The Employment Effects of Low-Wage Subsidies". In: *Journal of Public Economics* 97.C, pp. 49–60.

Imai, K., L. Keele, and D. Tingley (2010). "A general approach to causal mediation analysis". In: *Psychological Methods* 15.4, pp. 309–334.

Imai, K., L. Keele, and T. Yamamoto (2010). "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects". In: *Statistical Science* 25.1, pp. 51–71.

— (2011). "Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies". In: *American Political Science Review* 105.4, pp. 765–789.

Imai, K. and T. Yamamoto (2013). "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments". In: *Political Analysis* 21, pp. 141–171.

Iuzzolino, G., G. Pellegrini, and G. Viesti (2011). "Convergence among Italian Regions, 1861-2011". In: *Economic Histoy Working Papers, Banca d'Italia* 22.

Jaenichen, U. and G. Stephan (2011). "The Effectiveness of Targeted Wage Subsidies for Hard-to-place Workers". In: *Applied Economics* 43.10, pp. 1209–1225.

Junankar, P. N. (2011). “The Global Economic Crisis: Long-Term Unemployment in the OECD”. In: *IZA Discussion Paper* 6057.

Kaul, A., S. Klobner, G. Pfeifer, and M. Schieler (2018). “Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates”. In: *MPRA Working Paper* 83790.

Klößner, S., A. Kaul, G. Pfeifer, and M. Schieler (2018). “Comparative politics and the Synthetic Control Method revisited: a note on Abadie et al (2015)”. In: *Swiss Journal of Economics and Statistics* 154.11.

Kluve, J., H. Lehmann, and C. M. Schmidt (2008). “Disentangling Treatment Effects of Active Labor Market Policies: The Role of Labor Force Status Sequences”. In: *Elsevier Labour Economics* 15.6, pp. 1270–1295.

Kreif, N., R. Grieve, D. Hangartner, A. J. Turner, S. Nikolova, and M. Sutton (2016). “EXAMINATION OF THE SYNTHETIC CONTROL METHOD FOR EVALUATING HEALTH POLICIES WITH MULTIPLE TREATED UNITS”. In: *Health Economics* 25.12, pp. 1514–1528.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013). “Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment”. In: *Quarterly Journal of Economics* 128.3, pp. 1123–1167.

Lechner, M. (2015). “Treatment Effects and Panel Data”. In: *The Oxford Handbook of Panel Data*.

Lee, D. S. (2008). “Randomized experiments from non-random selection on U.S. House elections”. In: *Journal of Econometrics* 142.2, pp. 675–697.

Li, F., A. Mattei, and F. Mealli (2015). “Evaluating The Causal Effect of University Grants on Student Dropout: Evidence from a Regression Discontinuity using Principal Stratification”. In: *The Annals of Applied Statistics* 9.4, pp. 1906–1931.

M., Gerfin, M. Lechner, and H. Steiger (2005). “Does subsidised temporary employment get the unemployed back to work?” In: *Labour Economics* 12.6, pp. 807–835.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). “A Contribution to the Empirics of Economic Growth”. In: *Quarterly Journal of Economics* 107.2, pp. 407–437.

Martin, J. P. (2015). "Activation and active labour market policies in OECD countries: stylized facts and evidence of their effectiveness". In: *IZA Journal of Labor Policy* 4.4.

Martin, J. P. and D. Grubb (2001). "What works and for whom: A review of OECD countries' experiences with active labour market policies". In: *Swedish Economic Policy Review* 8.2, pp. 9–56.

Mortensen, D. T. and C. Pissarides (2001). "Taxes, subsidies and equilibrium labor market outcomes". In: *CEPR Discussion Paper* 2989.

Mussida, C. (2010). "Is Long-Term Unemployment Unaffected by Flexible Labour Market Legislation?" In: *Rivista internazionale di Scienze Sociali* 118.1, pp. 77–105.

Neubaumer, R. (2010). "Can Training Programs or Rather Wage Subsidies Bring the Unemployed Back to Work? A Theoretical and Empirical Investigation for Germany". In: *IZA Discussion Papers* 4864.

Neyman, J. (1990). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9". In: *Statistical Science, 1990. Translated and edited by Dabrowska, D. M. and Speed, T. P. originally apperead in Roczniki Nauk Rolniczych Tom X (1923) 1-51 Annals of Agricultural Sciences* 5.4, pp. 465–472.

Oberholzer-Gee, F. (2008). "Nonemployment stigma as rational herding: a field experiment". In: *Journal of Economic Behavior & Organization* 65.1, pp. 30–40.

Obermeier, T. and M. Meier (2016). "Duration Dependence, Dynamic Selection and the Optimal Timing of Unemployment Benefits". In: *Annual Conference (Augsburg): Demographic Change* 145503.

Omori, Y. (1997). "Stigma effects of nonemployment". In: *Economic Inquiry* 35.2, pp. 394–416.

Ozler, B. (2014). "Confusing a Treatment for a Cure". In: *The World Bank Blog*.

Pearl, J. (2001). "Direct and Indirect Effects". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann*, pp. 411–420.

— (2014). "Interpretation and identification of causal mediation". In: *Psychological Methods* 19.4, pp. 459–481.

Pellegrini, G., F. Terribile, O. Tarola, T. Muccigrosso, and F. Busillo (2013). "Measuring the Effects of European Regional Policy on Economic Growth: A Regression Discontinuity Approach". In: *Papers in Regional Science* 92.1, pp. 217–233.

Puigcerver-Penalver, M. C. (2007). "The impact of structural funds policy on European regions' growth. A theoretical and empirical approach." In: *The European Journal of Comparative Economics* 4.2, pp. 179–208.

Robins, J. M. and S. Greenland (1992). "Identifiability and exchangeability for direct and indirect effects". In: *Epidemiology* 3.2, pp. 143–155.

Rosenbaum, P. R. and D. B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". In: *Biometrika* 70.1, pp. 41–55.

Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66.5, pp. 688–701.

— (1977). "Assignment of Treatment Group on the Basis of a Covariate". In: *Journal of Educational Statistics* 2.1, pp. 1–26.

— (1990). "[On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies". In: *Statistical Science* 5.4, pp. 472–480.

Schünemann, B., M. Lechner, and C. Wunsch (2013). "Do Long-Term Unemployed Workers Benefit from Targeted Wage Subsidies?" In: *German Economic Review* 16.1, pp. 43–64.

Sestito, P. and E. Viviano (2016). "Hiring Incentives and/or firing cost reduction? Evaluating the umpact of the 2015 policies on the Italian labour market". In: *Bank of Italy Occasional Papers* 325.

Shpitser, I. and T. J. VanderWeele (2011). "A complete graphical criterion for the adjustment formula in mediation analysis". In: *International Journal of Biostatistics* 7.1.

Sianesi, B. (2008). "Differential effects of active labour market programs for the unemployed". In: *Labour Economics* 15.3, pp. 370–399.

Small, D. S. (2012). “Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables”. In: *https://arxiv.org/abs/1109.1070*.

Smith (1982). “Beliefs, Attribution and Evaluations: Nonhierarchical Models of Mediation in Social Cognition”. In: *Journal of Personality and Social Psychology* 43.2, pp. 248–259.

Soumerai, S. and D. Ross-Degnan (2002). “Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research”. In: *Journal of Clinical Pharmacy and Therapeutics* 27, pp. 299–309.

Vanderweele, T. J. and Y. Chiba (2014). “Sensitivity analysis for direct and indirect effects in presence of exposure-induced mediator-outcome confounders”. In: *Epidemiology, Biostatistics, and Public Health* 11.2.

Vanderweele, T. J. and S. Vansteelandt (2009). “Conceptual issues concerning mediation interventions and composition”. In: *STATISTICS AND ITS INTERFACE* 2, pp. 457–468.

Vansteelandt, S. and T. J. VanderWeele (2012). “Natural Direct and Indirect Effects on the Exposed: Effect Decomposition under Weaker Assumptions”. In: *Biometrics* 68.4, pp. 1019–1027.

Winship, C. and L. Radbill (1994). “Sampling Weights and Regression Analysis”. In: *SOCIOLOGICAL METHODS and RESEARCH* 23, pp. 230–257.

Xu, Y. (2017). “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models”. In: *Political Analysis* 25.1, pp. 57–76.

Zheng, C. and X. H. Zhou (2017). “Causal mediation analysis on failure time outcome without sequential ignorability”. In: *Lifetima Data Analysis* 23.4, pp. 533–559.

.1. *Generalized Policy ITT Estimation Including Data on December 2015*

In table E1 we present the results of the estimation of generalized incentives ITT when observations about December 2015 are included.

.2. *Derivation of "Synthetic"* Y_{1t}^{01}

To easy the notation the subscript t is dropped from the weights. Following Abadie, Diamond, et al. 2010, consider a generic vector of weights $W = (w_{n+1}, ..., w_J)'$ such that $w_j \geq 0$ for all $j = n+1, ..., J$ and $w_{n+1} + ... + w_J = 1$. With these weights (and considering the factor model introduced in the text) the synthetic value of Y_{1t}^{01} is given by

$$\sum_{j=n+1}^{J} w_j Y_{jt} = \zeta_t + \eta_t \sum_{j=n+1}^{J} w_j X_j + \lambda_t \sum_{j=n+1}^{J} w_j \mu_j + \varphi_t(0) \sum_{j=n+1}^{J} w_j M_{jt}(0) + \sum_{j=n+1}^{J} w_j \epsilon_{jt}.$$

The difference between the real potential outcome and the synthetic one is then

$$\begin{aligned} Y_{1t}^{0,1} - \sum_{j=n+1}^{J} w_j Y_{jt} &= \eta_t \left(X_1 - \sum_{j=n+1}^{J} w_j X_j \right) + \lambda_t \left(\mu_1 - \sum_{j=n+1}^{J} w_j \mu_j \right) \\ &+ \varphi_t(0) \left(M_{1t}(I\{t \geq T\}) - \sum_{j=n+1}^{J} w_j M_{jt}(0) \right) \\ &+ \sum_{j=n+1}^{J} w_j (\epsilon_{1t} - \epsilon_{jt}). \end{aligned} \tag{1}$$

Let Y_i^P be the $((T-1)\times 1)$ vector with tth element equal to Y_{it}, ϵ_i^P the $((T-1)\times 1)$ vector with tth element equal to ϵ_{it}, η^P the $((T-1)\times r)$ matrix with tth row equal to η_t and λ^P the $((T-1)\times F)$ matrix with tth row equal to λ_t. Moreover, let $\varphi^P(0)$ be the $((T-1)\times 1)$ vector with tth element equal to $\varphi_t(0)$ and $M_i^P(0)$ the $((T-1)\times 1)$ vector with tth element equal to $M_{it}(0)$. We can now write

$$\begin{aligned} Y_1^P - \sum_{j=n+1}^{J} w_j Y_j^P &= \eta^P \left(X_1 - \sum_{j=n+1}^{J} w_j X_j \right) + \lambda^P \left(\mu_1 - \sum_{j=n+1}^{J} w_j \mu_j \right) \\ &+ \varphi^P(0) \left(M_{1t}^P(0) - \sum_{j=n+1}^{J} w_j M_{jt}^P(0) \right) + \left(\epsilon_1^P - \sum_{j=n+1}^{J} w_j \epsilon_j^P \right). \end{aligned}$$

Table E1: ESTIMATION OF THE INTENTION TO TREATMENT EFFECT OF GENERALIZED INCENTIVES WITH THE INCLUSION OF DECEMBER 2015 OBSERVATIONS

VARIABLES	Coefficients
Time	-0.00498**
	(0.00214)
Time^2	0*
	(0)
Treat	1430
	(618)
Treat*Time	-1.49
	(0.687)
Treat*Time^2	0.000387
	(0.000174)
Constant	12.2***
	(0.930)
Observations	2191
R-squared	0.112

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

NOTE: Generalized incentives ITT on vulnerable group is given by the coefficient of variable "Treat". Observations about December 2015 included. We implemented a weighted regression, using the total number of individuals corresponding to each unit as weights. All monthly or daily variables were seasonally adjusted using a moving average method. For easier reading, all coefficients and standard errors were multiplied by 100,000.

Note that we have $M^P_{1t}(0)$ as $t < T$. It is easy to see that:

$$\begin{aligned}
\lambda^P\left(\mu_1 - \sum_{j=n+1}^{J} w_j\mu_j\right) &= Y^P_1 - \sum_{j=n+1}^{J} w_j Y^P_j - \eta^P\left(X_1 - \sum_{j=n+1}^{J} w_j X_j\right) \\
&- \varphi^P(0)\left(M^P_{1t}(0) - \sum_{j=n+1}^{J} w_j M^P_{jt}(0)\right) \\
&- \left(\epsilon^P_1 - \sum_{j=n+1}^{J} w_j \epsilon^P_j\right) \qquad (2)
\end{aligned}$$

Similar to Abadie, Diamond, et al. 2010 assume that

Assumption 3. $\sum_{t=1}^{T-1} \lambda_t' \lambda_t$ *is non-singular.*

Assumption 3 is equivalent to assume no perfect-collinearity among unobserved common factors and implies that $(\lambda^{P\prime}\lambda^P)^{-1}$ exists. We can then multiply both sides of 2 by $(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}$ to get

$$
\begin{aligned}
\mu_1 - \sum_{j=n+1}^{J} w_j \mu_j &= (\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime} \left\{ Y_1^P - \sum_{j=n+1}^{J} w_j Y_j^P - \eta^P \left(X_1 - \sum_{j=n+1}^{J} w_j X_j \right) \right. \\
&- \left. \varphi^P(0) \left(M_{1t}^P(0) - \sum_{j=n+1}^{J} w_j M_{jt}^P(0) \right) - \left(\epsilon_1^P - \sum_{j=n+1}^{J} w_j \epsilon_j^P \right) \right\}.
\end{aligned}
$$

Substituting in 1 and considering a generic post-intervention period $t' \geq T$, we have

$$
\begin{aligned}
Y_{1t'}^{0,1} - \sum_{j=n+1}^{J} w_j Y_{jt'} &= \lambda_{t'}(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime} \left(Y_1^P - \sum_{j=n+1}^{J} w_j Y_j^P \right) \\
&+ \left(\eta_{t'} - \lambda_{t'}(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\eta^P \right) \left(X_1 - \sum_{j=n+1}^{J} w_j X_j \right) \\
&- \lambda_{t'}(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime} \left[\varphi^P(0)(M_1^P(0) - \sum_{j=n+1}^{J} w_j M_j^P(0)) \right] \\
&+ \varphi_{t'}(0) \left(M_{1t'}(1) - \sum_{j=n+1}^{J} w_j M_{jt'}(0) \right) \\
&- \lambda_{t'}(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime} \left(\epsilon_1^P - \sum_{j=n+1}^{J} w_j \epsilon_j^P \right) + \sum_{j=n+1}^{J} w_j (\epsilon_{1t'} - \epsilon_{jt'}).
\end{aligned}
$$

If we now assume, as we did in the main text, that there exists a set of positive and summing up to 1 weights W^* that satisfies, $\forall\ t = 1, ..., T-1$

$$
\begin{aligned}
\sum_{j=n+1}^{J} w_j^* Y_{jt} &= Y_{1t}, \\
\sum_{j=n+1}^{J} w_j^* X_j &= X_1,
\end{aligned}
$$

and $\forall\ t = 1, \ldots, T-1, t'$, also satisfies

$$\sum_{j=n+1}^{J} w_j^* M_{jt} = M_{1t},$$

replacing in the post-intervention period, the generic weights with W^*, we get

$$Y_{1t'}^{0,1} - \sum_{j=n+1}^{J} w_j^* Y_{jt'} \quad = \quad -\lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}\left(\epsilon_1^P - \sum_{j=n+1}^{J} w_j^* \epsilon_j^P\right) + \sum_{j=n+1}^{J} w_j^*(\epsilon_{1t'} - \epsilon_{jt'}).$$

From here, the proof is identical to the one in Abadie, Diamond, et al. 2010. We can write

$$Y_{1t'}^{0,1} - \sum_{j=n+1}^{J} w_j^* Y_{jt'} = R_{1t'} + R_{2t'} + R_{3t'}$$

where

$$R_{1t'} = \lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}\sum_{j=n+1}^{J} w_j^* \epsilon_j^P \tag{3}$$

$$R_{2t'} = -\lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}\epsilon_1^P \tag{4}$$

$$R_{3t'} = \sum_{j=n+1}^{J} w_j^*(\epsilon_{jt'} - \epsilon_{1t'}) \tag{5}$$

Following Abadie, Diamond, et al. 2010, we impose the following assumptions

Assumption 4. *$\epsilon_{it} \perp \epsilon_{jt}\ \forall i \neq j$ with $i, j = 1, \ldots, J$.*

Assumption 5. *$\epsilon_{it} \perp \epsilon_{it''}\ \forall t \neq t''$ with $t, t'' = 1, \ldots, t'$.*

Assumption 6. *$E(\epsilon_{it}|X_i, \mu_i, M_{it}(I\{t \geq T\})) = E(\epsilon_{it}) = 0$ for $i \in \{1, n+1, \ldots, J\}$ and for $t = 1, \ldots, t'$*

Taking the expected value on both sides of 4 we get

$$\begin{aligned} E(R_{2t'}) &= E(-\lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}\epsilon_1^P) \\ &= -\lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}E(\epsilon_1^P) \\ &= 0 \end{aligned}$$

where the second equality follows from the fact that $-\lambda_{t'}(\lambda^{P'}\lambda^{P})^{-1}\lambda^{P'}$ is non-stochastic and the third equality follows from assumption 6. Taking the expectation on both sides of 5

$$\begin{aligned} E(R_{3t'}) &= E\left(\sum_{j=n+1}^{J} w_j^*(\epsilon_{jt'} - \epsilon_{1t'})\right) = \sum_{j=n+1}^{J}\left[E(w_j^*\epsilon_{jt'}) - E(w_j^*\epsilon_{1t'})\right] \\ &= \sum_{j=n+1}^{J}\left[E(w_j^*)E(\epsilon_{jt'}) - E(w_j^*)E(\epsilon_{1t'})\right] = 0 \end{aligned}$$

where the third equality follows from the fact that weights $W^* = w_{n+1}^*, \ldots, w_J^*$ are determined using constraints on covariates, pre-treatment period outcomes and the mediator which under assumptions 4, 5 and 6 are independent from the error terms at time $t' \geq T$. The fourth equality follows from assumption 6. The remaining 3 can be rewritten as:

$$R_{1t'} = \sum_{j=n+1}^{J} w_j^* \sum_{s=1}^{T-1} \lambda_{t'}(\sum_{h=1}^{T-1} \lambda_h' \lambda_h)^{-1} \lambda_s' \epsilon_{js} \tag{6}$$

As in Abadie, Diamond, et al. 2010, we further assume that

Assumption 7. *Let $\varsigma(M)$ be the smallest eigenvalue of*

$$\frac{1}{M}\sum_{t=T-M+1}^{T-1} \lambda_t' \lambda_t,$$

$\varsigma(M) \geq \underline{\varsigma} > 0$ for each positive integer M.

Assumption 8.

$$\exists\ \underline{\lambda} \mathit{s.t.} |\lambda_{tf}| \leq \underline{\lambda}\ \forall t\text{=1,...,}t'\text{ and } f\text{=1,...,}F.$$

Assumption 7 guarantees that the matrix $\sum_{t=1}^{T} \lambda_t' \lambda_t$ and, consequently, its inverse, are symmetric and positive definite. Thus, for the Cauchy-Schwarz inequality, we have that

$$\begin{aligned} \left(\lambda_t \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_s'\right)^2 &= |\langle \lambda_t, A\lambda_s' \rangle|^2 \leq ||A\lambda_t||^2 ||A\lambda_s||^2 \\ &= \left(\lambda_t \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_t'\right)\left(\lambda_s \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_s'\right) \end{aligned} \tag{7}$$

Where $A = \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1}$. Since A is a symmetric matrix $B = (T-1)A$ is symmetric as well. Thus, it can be decomposed as $B = GOG^{-1}$. Where G is orthogonal and $G^{-1} = G'$ and O is a diagonal matrix with the eigenvalues of B as elements. Thus,

$$\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_t = \frac{1}{T-1}(\lambda_t B \lambda'_t) = \frac{1}{T-1}(\lambda_t GOG' \lambda'_t)$$

Defining $b_t = \lambda_t G$ we have

$$\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_t = \frac{1}{T-1}(b_t O b'_t) = \frac{1}{T-1}\left(b_{t1}^2 \frac{1}{\varsigma_1} + \ldots + b_{tF}^2 \frac{1}{\varsigma_F}\right)$$

where ς_i are the eigenvalues of matrix B. From assumption 7, imposing $M = T-1$, we'll have that $\frac{1}{\varsigma_i} \le \frac{1}{\underline{\varsigma}}$ for $i = 1, \ldots, F$. Indeed the eigenvalues of the inverse of a matrix are given by the inverse of the matrix eigenvalues, and B is the inverse of the matrix in assumption 7. Consequently:

$$\begin{aligned}\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_t &= \frac{1}{T-1}\sum_{f=1}^{F} \frac{b_{tf}^2}{\varsigma_f} \le \frac{1}{(T-1)\underline{\varsigma}}\sum_{f=1}^{F} b_{tf}^2 \\ &= \frac{1}{(T-1)\underline{\varsigma}}||b_t||^2 = \frac{1}{(T-1)\underline{\varsigma}}||\lambda_t G||^2\end{aligned}$$

As we noticed before, G is an orthogonal and thus isometric matrix, hence $||\lambda_t G|| = ||\lambda_t||$. Consequently,

$$\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_t \le \frac{1}{(T-1)\underline{\varsigma}}||\lambda_t||^2 = \frac{\sum_{f=1}^{F} \lambda_{tf}^2}{(T-1)\underline{\varsigma}} \le \frac{\sum_{f=1}^{F} \underline{\lambda}^2}{(T-1)\underline{\varsigma}} = \frac{F\underline{\lambda}^2}{(T-1)\underline{\varsigma}}$$

where the last inequality follows from assumption 8. Applying the same idea to the second part of 7 we get

$$\begin{aligned}\left(\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_s\right)^2 &\le \left(\lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_t\right)\left(\lambda_s \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h\right)^{-1} \lambda'_s\right) \\ &\le \left(\frac{F\underline{\lambda}^2}{(T-1)\underline{\varsigma}}\right)^2 \end{aligned} \tag{8}$$

Following Abadie, Diamond, et al. 2010 we define

$$\overline{\epsilon_j^L} = \sum_{s=1}^{T-1} \lambda_T (\sum_{h=1}^{T-1} \lambda'_h \lambda_h)^{-1} \lambda'_s \epsilon_{js} \tag{9}$$

for $j = n + 1, ..., J$. Assume that

Assumption 9. *The p^{th} moment of $|\epsilon_{jt}|$ for some even p exists for $j = 2, ..., J$ and $t = 1, ..., T - 1$*

Using Hölder's Inequality and taking into account that $0 \leq w_j^* \leq 1$ for $j = n + 1, ..., J$ we have that:

$$\begin{aligned}
\sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L}| &= \sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L} * 1| \leq \left(\sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L}|^p \right)^{1/p} \left(\sum_{j=n+1}^{J} w_j^* |1|^q \right)^{1/q} \\
&= \left(\sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L}|^p \right)^{1/p} \left(\sum_{j=n+1}^{J} w_j^* \right)^{1/q} = \left(\sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L}|^p \right)^{1/p} \leq \left(\sum_{j=n+1}^{J} |\overline{\epsilon_j^L}|^p \right)^{(1/p)}
\end{aligned}$$

where the last equality follow from $w_{n+1}^* + ... + w_J^* = 1$ and the last inequality follows from the condition that $w_{n+1}^* \leq 1, ..., w_J^* \leq 1$. Applying Hölder's Inequality again we get

$$E\left[\sum_{j=n+1}^{J} w_j^* |\overline{\epsilon_j^L}| \right] \leq \left(E\left[\sum_{j=n+1}^{J} |\overline{\epsilon_j^L}|^p \right] \right)^{1/p} \tag{10}$$

Applying Rosenthal's Inequality we have

$$\begin{aligned}
E\left[|\overline{\epsilon_j^L}|^p \right] &= E\left[\left| \sum_{s=1}^{T-1} \lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h \right)^{-1} \lambda'_s \epsilon_{js} \right| \right] \\
&\leq C(p) \max \left(\sum_{s=1}^{T-1} E\left[\left| \lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h \right)^{-1} \lambda'_s \epsilon_{js} \right|^p \right] \right. \\
&\quad , \left. \left(\sum_{s=1}^{T-1} E\left[\left| \lambda_t \left(\sum_{h=1}^{T-1} \lambda'_h \lambda_h \right)^{-1} \lambda'_s \epsilon_{js} \right|^2 \right] \right)^{p/2} \right)
\end{aligned}$$

where $C(p)$ is the pth moment of -1 plus a Poisson random variable with mean 1 (see Abadie, Diamond, et al. 2010). Consider the two elements of $max(.)$. For the first element, we have

$$
\begin{aligned}
\sum_{s=1}^{T-1} E\left[\left|\lambda_t \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_s' \epsilon_{js}\right|^p\right] &= \sum_{s=1}^{T-1} E\left[\left(\lambda_t \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_s'\right)^{2*(p/2)} |\epsilon_{js}|^p\right] \\
&\leq \sum_{s=1}^{T-1} E\left[\left(\frac{F\underline{\lambda}^2}{(T-1)\,\underline{\varsigma}}\right)^{2*(p/2)} |\epsilon_{js}|^p\right] \\
&= \left(\frac{F\underline{\lambda}^2}{\underline{\varsigma}}\right)^p \frac{1}{(T-1)^p} \sum_{s=1}^{T-1} E\left(|\epsilon_{js}|^p\right)
\end{aligned}
$$

where the first equality follows from the distributivity of the power and the inequality follows from 8. For the second element in *max* (.), we have

$$
\begin{aligned}
\left(\sum_{s=1}^{T-1} E\left[\left|\lambda_t \left(\sum_{h=1}^{T-1} \lambda_h' \lambda_h\right)^{-1} \lambda_s' \epsilon_{js}\right|^2\right]\right)^{p/2} &\leq \left[\sum_{s=1}^{T-1} E\left(\left(\frac{F\overline{\lambda}^2}{(T-1)\,\underline{\varsigma}}\right)^2 \epsilon_{js}^2\right)\right]^{p/2} \\
&= \left(\frac{F\overline{\lambda}^2}{\underline{\varsigma}}\right)^p \left[\sum_{s=1}^{T-1} \frac{1}{(T-1)^2} E\left(\epsilon_{js}^2\right)\right]^{p/2}
\end{aligned}
$$

where the first inequality follows from 8. Putting all these results together have

$$
E\left[|\overline{\epsilon_j^L}|^p\right] \leq C(p) \left(\frac{F\overline{\lambda}^2}{\underline{\varsigma}}\right)^p \max\left(\frac{1}{(T-1)^p} \sum_{s=1}^{T-1} E\left(|\epsilon_{js}|^p\right), \left[\sum_{s=1}^{T-1} \frac{1}{(T-1)^2} E\left(\epsilon_{js}^2\right)\right]^{p/2}\right)
$$

As Abadie, Diamond, et al. 2010, we define $\sigma_{js}^2 = E|\epsilon_{js}|^2$, $\sigma_j^2 = (1/(T-1)\sum_{s=1}^{T-1} \sigma_{js}^2)$, $\overline{\sigma^2} = max_{j=n+1,\dots,J}\sigma_j^2$ and $\overline{\sigma} = \sqrt{\overline{\sigma^2}}$. Similarly, we define $\tau_{p,jt} = E|\epsilon_{jt}|^p$, $\tau_{p,j} = \frac{1}{(T-1)}\sum_{t=1}^{T-1} \tau_{p,jt}$, and $\overline{\tau_p} = max_{j=n+1,\dots,J}\tau_{p,j}$. We can write the first element of *max*(.) as

$$
\frac{1}{(T-1)^p} \sum_{s=1}^{T-1} E(|\epsilon_{js}|^p) = \frac{1}{(T-1)^{p-1}} \frac{1}{(T-1)} \sum_{t=1}^{T-1} \tau_{pjt} = \frac{1}{(T-1)^{p-1}} \tau_{pj}
$$

Similarly, the second element can be written as

$$
\left[\sum_{s=1}^{T-1} \frac{1}{(T-1)^2} E(\epsilon_{js}^2)\right]^{p/2} = \left(\frac{1}{T-1}\frac{1}{T-1}\sum_{s=1}^{T-1} \sigma_{js}^2\right)^{p/2} = \left(\frac{1}{T-1}\sigma_j^2\right)^{p/2}
$$

Thus , defining $\varpi = C(p)(\frac{F\bar{\lambda}^2}{\underline{\varsigma}})^p$, we have

$$\begin{aligned}
E\left[|\overline{\epsilon_j^L}|^p\right] &\leq \varpi \max\left(\frac{1}{(T-1)^{p-1}}\tau_{pj}, \left(\frac{1}{T-1}\sigma_j^2\right)^{p/2}\right) \\
\sum_{j=n+1}^{J} E\left[|\overline{\epsilon_j^L}|^p\right] &= E\left[\sum_{j=n+1}^{J} |\overline{\epsilon_j^L}|^p\right] \\
&\leq \varpi \max\left(\frac{1}{(T-1)^{p-1}}\sum_{j=n+1}^{J}\tau_{pj}, \sum_{j=n+1}^{J}\left(\frac{1}{T-1}\sigma_j^2\right)^{p/2}\right) \\
&= \varpi \max\left(\frac{J-n-1}{(T-1)^{p-1}}\frac{1}{J-n-1}\sum_{j=n+1}^{J}\tau_{pj}, \frac{1}{(T-1)^{p/2}}\sum_{j=n+1}^{J}\sigma_j^{2*p/2}\right) \\
\left(E\left[\sum_{j=n+1}^{J} |\overline{\epsilon_j^L}|^p\right]\right)^{1/p} &\leq \varpi^{1/p} \max\left(\frac{\left(\frac{J-n-1}{(T-1)^{p-1}}\right)^{1/p}}{(J-n-1)^{1/p}}\left(\sum_{j=n+1}^{J}\tau_{pj}\right)^{1/p}, \frac{\left(\sum_{j=n+1}^{J}\sigma_j^{2*p/2}\right)^{1/p}}{(T-1)^{(p/2)*(1/p)}}\right) \\
&= \varpi^{1/p} \max\left(\left(\frac{J-n-1}{(T-1)^{p-1}}\right)^{1/p}\overline{\tau}_p^{1/p}, \frac{1}{(T-1)^{1/2}}\left(\sum_{j=n+1}^{J}\overline{\sigma}^{2*(p/2)}\right)^{1/p}\right)
\end{aligned}$$

where the last equality follows from $\frac{1}{J-n-1}\sum_{j=n+1}^{J}\tau_{pj} = E(\tau_{pj}) \leq max_j(\tau_{pj}) = \overline{\tau_p}$. Thus,

$$\begin{aligned}
\left(E\left[\sum_{j=n+1}^{J} |\overline{\epsilon_j^L}|^p\right]\right)^{1/p} &\leq \varpi^{1/p} max\left(\frac{(J-n-1)^{1/p}\,\overline{\tau}_p^{1/p}}{(T-1)^{1-1/p}}, \frac{(J-n-1)\,\overline{\sigma}^{2*(p/2)}}{(T-1)^{1/2}}\right)^{1/p} \\
&= \varpi^{1/p}(J-n-1)^{1/p}\, max\left(\frac{\overline{\tau}_{1/p}}{(T-1)^{1-\frac{1}{p}}}, \frac{\sqrt{\overline{\sigma}^2}}{(T-1)^{1/2}}\right) \qquad (11)
\end{aligned}$$

this implies

$$
\begin{aligned}
E\left[|R_{1t'}|\right] &= E\left[\left|\sum_{j=n+1}^{J} w_j^* \epsilon_j^L\right|\right] \\
&\leq E\left[\sum_{j=n+1}^{J} w_j^* |\epsilon_j^L|\right] \\
&\leq \left(E\left[\sum_{j=n+1}^{J} |\epsilon_j^L|^p\right]\right)^{1/p} \\
&\leq \varpi^{1/p}(J-n-1)^{1/p} max\left(\frac{\overline{\tau}_p^{1/p}}{(T-1)^{1-\frac{1}{p}}}, \frac{\overline{\sigma}}{(T-1)^{1/2}}\right)
\end{aligned}
$$

where, in the second equation, the first equality follows from 4 and 9, the first inequality follows from the triangular inequality, the second follows from 10 and the third from 11. It follows that

$$
E|R_{1t'}| \leq C(p)^{1/p}\frac{\overline{\lambda^2}F}{\underline{\varsigma}}(J-n-1)^{1/p}\max\left\{\frac{\overline{\tau_p^{1/p}}}{(T-1)^{1-1/p}}, \frac{\overline{\sigma}}{(T-1)^{1/2}}\right\}.
$$

Thus, the difference between the expected value of $Y_{1t}^{0,1}$ and its synthetic counterpart can be bounded by something that goes to zero when the number of pre-intervention periods goes to infinity, namely

$$
E\left(Y_{1t'}^{0,1} - \sum_{j=n+1}^{J} w_j^* Y_{jt'}\right) = E(R_{1t'}) = o(T).
$$

.3. *Extra assumptions on the mediator needed for* Y_{1t}^{10}

To create a synthetic Y_{1t}^{10} we need to impose the standard SCM assumptions on the mediator which are:

Assumption 10. $\sum_{t=1}^{T-1} \vartheta_t' \vartheta_t$ *is non-singular.*

Assumption 11. $v_{it} \perp v_{jt}$ $\forall i \neq j$ *with* $i, j \in \{1, n+1, ..., J\}$.

Assumption 12. $v_{it} \perp v_{it''}$ $\forall t \neq t''$ *with* $t, t'' = 1, ..., t'$.

Assumption 13. $E(v_{it}|\{Z_i, \varrho_i\}_{i\in\{1,n+1,...,J\}}) = E(v_{it}) = 0$ *for* $i \in \{1, n+1, ..., J\}$ *and for* $t = 1, ..., t'$

Assumption 14. *$\kappa(M) \geq \underline{\kappa} > 0$ for each positive integer M, where $\kappa(M)$ is the smallest eigenvalue of*

$$\frac{1}{M}\sum_{t=T-M+1}^{T-1} \vartheta_t'\vartheta_t. \tag{12}$$

Assumption 15.

$$\exists\ \underline{\vartheta} s.t. |\vartheta_{tv}| \leq \underline{\vartheta}\ \forall t=1,...,t' \text{ and } v=1,...,V. \tag{13}$$

Assumption 16. *$\exists$ a p^{th} moment of $|v_{jt}|$ for some even p and for $j = n+1,\ldots,J$ and $t = 1,\ldots,t'$*

.4. *Derivation of "Synthetic"* Y_{1t}^{10}

As for Y_{1t}^{01} we drop the subscript t from the weight and we write

$$\begin{aligned}\sum_{j=2}^{n} q_j Y_{jt} = \zeta_t \quad &+ \quad \eta_t \sum_{j=2}^{n} q_j X_j + \lambda_t \sum_{j=2}^{n} q_j \mu_j + \varphi_t\left(I\{t \geq T\}\right) \sum_{j=2}^{n} q_j M_{jt}\left(I\{t \geq T\}\right) \\ &+ \quad \sum_{j=2}^{n} q_j \rho_t\left(M_{jt}\left(I\{t \geq T\}\right)\right) I\{t \geq T\} + \sum_{j=2}^{n} q_j \epsilon_{jt}.\end{aligned}$$

Thus,

$$\begin{aligned}Y_{1t}^{1,0} - \sum_{j=2}^{n} q_j Y_{jt} &= \eta_t\left(X_1 - \sum_{j=2}^{n} q_j X_j\right) + \lambda_t\left(\mu_1 - \sum_{j=2}^{n} q_j \mu_j\right) \\ &+ \varphi_t\left(I\{t \geq T\}\right)\left(M_{1t}(0) - \sum_{j=2}^{n} q_j M_{jt}(I\{t \geq T\})\right) \\ &+ \left(\rho_t\left(M_{1t}(0)\right) - \sum_{j=2}^{n} q_j \rho_t\left(M_{jt}\left(I\{t \geq T\}\right)\right)\right) I\{t \geq T\} \\ &+ \sum_{j=2}^{n} q_j\left(\epsilon_{1t} - \epsilon_{jt}\right)\end{aligned}$$

Using the same notation as before in the pre-intervention period we have

$$\begin{aligned}Y_1^P - \sum_{j=2}^{n} q_j Y_j^P &= \eta^P\left(X_1 - \sum_{j=2}^{n} q_j X_j\right) + \lambda^P\left(\mu_1 - \sum_{j=2}^{n} q_j \mu_j\right) \\ &+ \varphi^P(0)\left(M_1^P(0) - \sum_{j=2}^{n} q_j M_j^P(0)\right) + \left(\epsilon_1^P - \sum_{j=2}^{n} q_j \epsilon_j^P\right)\end{aligned}$$

Thus

$$\begin{aligned}
\lambda^P\left(\mu_1-\sum_{j=2}^n q_j\mu_j\right) &= Y_1^P-\sum_{j=2}^n q_jY_j^P-\eta^P\left(X_1-\sum_{j=2}^n q_jX_j\right)\\
&- \varphi^P(0)\left(M_1^P(0)-\sum_{j=2}^n q_jM_j^P(0)\right)-\left(\epsilon_1^P-\sum_{j=2}^n q_j\epsilon_j^P\right)
\end{aligned}$$

Multiplying both sides by $(\lambda^{P'}\lambda^P)^{-1}\lambda^{P'}$ we get

$$\begin{aligned}
\mu_1-\sum_{j=2}^n q_j\mu_j &= \left(\lambda^{P'}\lambda^P\right)^{-1}\lambda^{P'}\left\{\left(Y_1^P-\sum_{j=2}^n q_jY_j^P\right)-\eta^P\left(X_1-\sum_{j=2}^n q_jX_j\right)\right.\\
&- \left.\varphi^P(0)\left(M_1^P(0)-\sum_{j=2}^n q_jM_j^P(0)\right)-\left(\epsilon_1^P-\sum_{j=2}^n q_j\epsilon_j^P\right)\right\}.
\end{aligned}$$

Substituting in 14 and considering a generic post-intervention period t', we have

$$\begin{aligned}
Y_{1t'}^{1,0}-\sum_{j=2}^n q_jY_{jt'} &= \left(\lambda^{P'}\lambda^P\right)^{-1}\lambda^{P'}\left(Y_1^P-\sum_{j=2}^n q_jY_j^P\right)\\
&+ \left(\eta_{t'}-\left(\lambda^{P'}\lambda^P\right)^{-1}\lambda^{P'}\eta^P\right)\left(X_1-\sum_{j=2}^n q_jX_j\right)\\
&- \left(\lambda^{P'}\lambda^P\right)^{-1}\lambda^{P'}\varphi^P(0)\left(M_1^P(0)-\sum_{j=2}^n q_jM_j^P(0)\right)\\
&+ \varphi_{t'}(1)\left(M_{1t'}(0)-\sum_{j=2}^n q_jM_{jt'}(1)\right)\\
&+ \left(\rho_{t'}\left(M_{1t'}(0)\right)-\sum_{j=2}^n q_j\rho_{t'}\left(M_{jt}(1)\right)\right)\\
&- \left(\lambda^{P'}\lambda^P\right)^{-1}\lambda^{P'}\left(\epsilon_1^P-\sum_{j=2}^n q_j\epsilon_j^P\right)+\sum_{j=2}^n q_j\left(\epsilon_{1t'}-\epsilon_{jt'}\right)
\end{aligned}$$

Assume, as we did in the main text, that there exists weights $q_2^*, ..., q_n^*$ that satisfy

$\forall t = 1, ..., T-1$

$$\begin{aligned}\sum_{j=2}^{n} q_j^* Y_{jt} &= Y_{1t},\\ \sum_{j=2}^{n} q_j^* X_j &= X_1,\\ \sum_{j=2}^{n} q_j^* M_{jt} &= M_{1t},\end{aligned}$$

and it also satisfies

$$\sum_{j=2}^{n} q_j^* M_{jt'} = \hat{M}_{1t'}(0).$$

Substituting the generic weights with $q_2^*, ..., q_n^*$ in the post-intervention period t', we get

$$\begin{aligned}Y_{1t'}^{1,0} - \sum_{j=2}^{n} q_j^* Y_{jt'} &= \left(\rho_{t'}\left(M_{1t'}(0)\right) - \sum_{j=2}^{n} q_j^* \rho_{t'}\left(M_{jt'}(1)\right)\right)\\ &- \left(\lambda^{P'}\lambda^{P}\right)^{-1}\lambda^{P'}\left(\epsilon_1^P - \sum_{j=2}^{n} q_j^* \epsilon_j^P\right) + \sum_{j=2}^{n} q_j^*\left(\epsilon_{1t'} - \epsilon_{jt'}\right)\end{aligned}$$

Note that, as by assumption $\sum_{j=2}^{n} q_j^* M_{t'} = \hat{M}_{1t'}(0)$ and $\hat{M}_{1t'}(0)$ is estimated using a standard SCM

$$E\left(\varphi_{t'}(1)\left(M_{1t'}(0) - \sum_{j=2}^{n} q_j^* M_{jt'}(1)\right)\right) = o(T).$$

As we mention in the main text, for identification we have to impose an extra assumption, namely

Assumption 17. $\rho_{t'}(.)$ *is a linear function*

Under assumption 17 we have

$$\begin{aligned}E\left[\left(\rho_{t'}\left(M_{1t'}(0)\right) - \sum_{j=2}^{n} q_j^* \rho_{t'}\left(M_{jt'}(1)\right)\right)\right] &= E\left[\left(\rho_{t'}\left(M_{1t'}(0)\right) - \rho_{t'}\left(\sum_{j=2}^{n} q_j^* M_{jt'}(1)\right)\right)\right],\\ &= E\left[\left(\rho_{t'}\left(M_{1t'}(0)\right) - \rho_{t'}\left(\hat{M}_{1t'}(0)\right)\right)\right],\\ &= \rho_{t'}\left(M_{1t'}(0)\right) - \rho_{t'}\left(E(\hat{M}_{1t'}(0))\right) = o(T).\end{aligned}$$

Thus,

$$Y^{1,0}_{1t'} - \sum_{j=2}^{n} q^*_j Y_{jt'} \quad = \quad -\left(\lambda^{P'}\lambda^{P}\right)^{-1}\lambda^{P'}\left(\epsilon^{P}_1 - \sum_{j=2}^{n} q^*_j \epsilon^{P}_j\right) + \sum_{j=2}^{n} q^*_j \left(\epsilon_{1t'} - \epsilon_{jt'}\right).$$

This, with an analogous as the one above therefore omitted proof, can be shown to imply

$$E(Y^{1,0}_{1t'} - \sum_{j=2}^{n} q^*_j Y_{jt'}) \quad = \quad o(T).$$