Università di Firenze, Università di Perugia, INdAM consorziate nel CIAFM

DOTTORATO DI RICERCA
IN MATEMATICA, INFORMATICA, STATISTICA

CURRICULUM IN STATISTICA
CICLO XXXV

Sede amministrativa Università degli Studi di Firenze
Coordinatore Prof. Matteo Focardi

Methods for policy evaluation with panel data

Settore Scientifico Disciplinare SECS-S/01

Dottorando:
Giulio Grossi

Tutore
Prof. Alessandra Mattei

Coordinatore
Prof. Matteo Focardi

Anni 2019/2022

Methods for policy evaluation with panel data

Giulio Grossi

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in

Statistics
DiSIA, University of Florence

Abstract

The focus of this work is present some novel developments in public policy evaluation methods. All four works that make up this dissertation are described under the Rubin causal model (RCM), a framework to estimate causal quantities, under some realistic assumptions and using some of the recently developed estimators. I focused my attention on application to structured data, such as panel data, spatial data and time series. The dissertation is structured as follows: the first chapter introduces the use of the Synthetic Control Method for policy evaluation under a partial interference framework. The second chapter presents the evaluation of conditional cash lotteries, implemented in the US, to foster Covid-19 vaccination. The third chapter presents a novel estimator for causal quantities in presence of spatially correlated treated units. The last chapter deals with a novel application of the principal stratification method to evaluate an active labour market policy. All of these works are research papers, the first is a joint work with Marco Mariani, Alessandra Mattei and Patrizia Lattarulo, the third is a joint work with Alessandra Mattei and Georgia Papadogeorgou, and the fourth is a joint work with Marco Mariani and Alessandra Mattei. Lastly, the second work is on my own.

Acknowledgements

In moments like these, it's easier to realize that we are not alone on an island, but rather every one of our achievements, or successes, comes from our relationship with others. It's as if we are a projection of how others see us, and of the trust they place in us. From this point of view, I have been exceptionally fortunate, having had the opportunity to meet the people that I want to thank here.

I want to thank Giampiero, Patrizia, and Marco, who were the first to take a chance on me in the dark, and laid the foundation of my path as a researcher, especially Marco, for the time he has spent with me.

I want to thank Fabrizia, for her support over these three years and her trust in young researchers, and Georgia, for her enthusiasm and for all that she has imparted to me. I went to the US to work with her, and found a friend. I can never thank Alessandra enough, who teaches me this job daily, I couldn't have imagined a better mentor, you are the researcher I will always aspire to be.

A big hug to those who have shared this journey with me and to my colleagues, Silvia, Carla, Claudio, Maria, Veronica, and Fiammetta, we deserve the best.

A big hug also to all my friends, who over these years have supported me with confidence and who are genuinely happy for me when I talk about research.

I must and always want to remember where I came from, those values that guide me and have brought me this far, to my grandparents, for the sincere and unconditional affection with which I have always been filled. A huge thank you to Claudio and my dad, without whom none of this would have been possible, for the unwavering trust and for being the first people I turn to for advice, you are my compass.

The biggest thank you can only go to Deborah, who more than anyone else has helped me and stood by me on this journey, it's as if this thesis is also her work. Thank you, thank you for being my best friend, my primary support, for being my vision of the future, every small success of mine is dedicated to you.

To my mother, because a mother's dreams never fade.

Contents

Introduction

“Imagine navigating a treacherous road with your eyes blindfolded, that’s what it’s like to implement public policy without data”. This is a statement that was once shared with me by a wise man and it has always stuck with me. As it turns out, this wise person was spot on. The evaluation of public policies is becoming an increasingly crucial topic for all scholars involved in drawing causal claims. The challenges that citizens are facing in the early 21st-century demand informed actions from public operators. Whether it’s health policy, energy policy, infrastructure policy, or action against climate change, these are all topics that should be at the top of the policymakers’ agenda in the coming years. As a result, policymakers are becoming increasingly aware of the fact that their decisions can be evaluated using more rigorous tools than ever before, and that they can leverage these tools to make more informed decisions and ultimately improve the general well-being. The academic community has been paying more attention to the importance of policy evaluation in recent years, as evidenced by the prestigious awards that have been given out. For example, in 2019, Esther Duflo was awarded the Swedish Bank prize for economic sciences for her groundbreaking work on the use of field experiments to study economic development. While in 2021, Joshua Angrist, David Card, and Guido Imbens were also awarded the same prize for their innovative methods for analyzing causal relationships. These awards recognize the crucial role that careful policy evaluation plays in understanding how to best achieve positive social outcomes and make informed decisions.

Policy evaluation examines the effectiveness of government programs in achieving their intended objectives. This includes programs such as financial aid for firms, labour market policies for the unemployed, and legislative interventions in the market. The literature in this field primarily focuses on the post-intervention effects of a given treatment on outcome variables that can be quantitatively measured, often in a longitudinal setting, as is the case in this dissertation.

The methodological framework that contains this work is the potential outcomes framework (Rubin, 1974, Rubin, 1978, Rubin, 1980, Imbens and Rubin, 2015). Using the potential outcomes framework, we focus on the evaluation of an intervention on N units, observed for T times, exposed to some treatment $W_{i,t}$ (e.g.: a law that increases tax on cigarettes) with an outcome variable $Y_{i,t}$ (the value of cigarettes vending). Let $W_{i,t} = 1$ denote the treatment assignment for treated units and $W_{i,t} = 0$ the treatment assignment for controls. In the potential outcome framework, it is assumed that there are no anticipatory effects or dynamic effects and that the Stable Unit Treatment Value Assumption (SUTVA) applies. SUTVA states that there are no spillover effects or hidden versions of the treatment. This means that for each period t and each unit i, there are two potential outcomes - $Y_{it}(W_{i,t} = 1)$ and $Y_{it}(W_{i,t} = 0)$ - which represent the outcome under the treatment and control assignment, respectively. The causal effect of the treatment on each unit is determined by comparing these potential outcomes, typically by looking at their difference. The fundamental problem in causal inference (Rubin, 1974, Holland, 1986) is that it's not possible to observe both potential outcomes for any unit in any period. In principle, before treatments are assigned, it may possible to observe the potential outcomes for all units under both the control and treatment conditions. But once treatments are assigned, we can only observe the outcome for the chosen condition, while the outcome under the other condition is missing, and we call it a counterfactual outcome. It's like a game of "what if" where we can only see the outcome of one scenario, but can only imagine the potential outcome of the other.

The potential outcomes framework has been widely used over the last two decades for policy evaluation, generating significant growth in both methodological and applied literature. The recent rise in the use of causal inference methods in policy evaluation can be attributed to the emergence of the "credibility revolution" in the field (see Angrist and Pischke (2010)). The credibility revolution refers to the growing use of rigorous experimental and quasi-experimental methods to estimate the causal effects of public policies. This movement began in the late 20th century (see for instance: Card and Krueger, 1993, Angrist and Krueger, 1999, Angrist et al., 1996) and has been driven by advances in econometrics, statistics, and experimental design. The availability of such methods enables researchers to infer causality in policy implementation, regardless of whether the study design is a randomized control design experiment or observational study.

Many policy evaluation studies base their causal inference on an unconfoundness assumption. Under unconfoundness, the difference in outcomes between units with the same level of covariates can be attributed to the effect of the treatment (see Rosenbaum et al., 2010, Imbens and Rubin, 2015). In works exploiting unconfoundness, the missing potential outcomes are imputed using the observed values of the control units, which are similar to the treated ones for covariate values, and pre-treatment outcomes values. Among these methods, it is worth reminding methods based on outcome regression, or *horizontal regression* methods (Athey et al., 2021), methods based on the propensity score, used as a weight, in sub-classification or in matching (see Rosenbaum and Rubin, 1983) and Bayesian methods, see Imbens and Rubin (2015), Abadie and Cattaneo (2018) or Athey and Imbens (2017) for recent reviews of such methods.

Such unconfoundness assumption can be less reliable in contexts with longitudinal observation, such as panel data in which a single or few units are treated, and the units are observed for several time periods. Several works have proposed solutions for accounting for possible unobservable confounders and the presence of temporal trends. Some of these works are milestones of the applied research, for example, Difference in Differences (DiD) methods have met a roaring success among applied economics, consider for instance the works of Card and Krueger (1993), Bertrand et al. (2004), or for a more recent development Callaway and Sant'Anna (2021). Based on an assumption of parallel trends across treatment and control group, in DiD the causal effect is estimated as the difference in the pre-post comparison between the treated and control groups.

Another stream of research refers to the Synthetic Control Method (SCM) first introduced by Abadie and Gardeazabal (2003) and Abadie et al. (2010), addressed as "the major innovation in policy evaluation of the last 15 years" by Athey and Imbens (2017). SCM focuses on panel data setting with a single treated unit and many control units. This method proposes to estimate the counterfactual values for the treated unit as a weighted average of the control units. Transparency and intuition behind this method have been the driving force behind its recent popularity and success, stressing its great flexibility. Among the applications fields, I mention the evaluation of economic policy Abadie et al. (2010), health policy Barber and West (2021), analysis of natural disasters Cavallo et al. (2013), the impact of terrorism (Abadie and Gardeazabal, 2003) or organized crime (Pinotti, 2015).

Recent developments in SCM have loosened some assumptions behind it and broadened the application fields. Agarwal et al. (2020), Xu (2017) and Ben-Michael et al. (2021) have proposed methods for unconstrained weights estimation, while Ben-Michael et al. (2022) formalized the use of SCM in staggered adoption context. See Abadie (2021) for a recent taxonomy of SCM. Other recent SCM developments have combined multiple approaches to improve the quality of estimates, as the case of the Synthetic Difference in Differences from Arkhangelsky et al. (2019) or the Penalized Synthetic Control Method from Abadie and L'Hour (2021) that balance across matching estimator and SCM.

Doudchenko and Imbens (2016) and Athey et al. (2021) have proposed to see the imputation of missing potential outcomes as a matrix completion (MCM) problem and solve it by using an unconstrained linear combination of control units, as in the case of the *vertical regression*. Recent discussions and proposals for this class of estimators are presented in Shen et al. (2022) and Arkhangelsky and Imbens (2022).

Other methods exploit the temporal structure of data, drawing causal quantities using the time series literature methods, such as Brodersen et al. (2015), Bojinov and Shephard (2019), Menchetti and Bojinov (2020) and Bojinov et al. (2021).

Some of the most recent approaches to public policy evaluation connect decision theory and causal inference in searching the optimal allocation for a policy in the so-called *policy learning*, as for instance, Viviano (2019), Athey and Wager (2021). The aim of such approaches is in maximising the utility function associated to some treatment in order to exploit the maximum welfare from a policy.

In this work, I aim to add to the ongoing discussion about methods for causal policy evaluation by introducing new approaches in challenging environments. Specifically, I have focused on addressing causal inference estimations with panel data, spatial data, and time series. My contributions include both methodological and applied advancements, specifically addressing complications that can arise when working with observational studies, which are prevalent in policy evaluation research.

The first contribution of this work is to the Synthetic Control Method literature. Using SCM, usually scholars assume SUTVA, Rubin (1980), which rules out the possibility of interference across units Even if this is a quite common assumption in the literature, it could be quite restrictive in many applications. For instance, think about intervention in some specific treatment site that could generate spillovers emanating across space, affecting untreated units. In such contexts, assuming no interference can lead to biased results. In contexts with panel data, there are few works that address explicitly this issue, see for instance the works of Menchetti and Bojinov (2020), Cao and Dowd (2019) and Di Stefano and Mellace (2020). Usually, scholars deal with spillover effects exploiting a partial interference assumption (Sobel, 2006, Hudgens and Halloran, 2008) to rule out spillovers between treated units and control units located far away from treated, or in different clusters of units. See for instance Forastiere et al. (2016), Forastiere et al. (2021a) or Papadogeorgou et al. (2019). The first chapter of this dissertation aims to contribute to this debate by providing estimates for direct and spillover effects arising from an intervention. In particular, it will be presented some novel proposals to estimate both spillover and direct effects of treatment in panel data settings, under a partial interference assumption. We will estimate these quantities using the Penalized SCM from Abadie and L'Hour (2021). Our motivating application is the causal evaluation of the construction of the first line of the Florentine tramway network. In particular, we will estimate and discuss the effects of such infrastructure on commercial vitality, drawing some conclusions that could help urban policymakers.

In economic literature, it is commonly observed that some units adopt a treatment previously adopted by other treated units, a phenomenon known as *policy mimicking*. However, the effectiveness of such behaviour is not guaranteed, and a policy that may be successful for one unit may not be effective for another. The second chapter of the dissertation presents an evaluation of a public policy implemented in the US to promote Covid-19 vaccination through a conditional cash lottery. Following Ohio's lead, 18 other states implemented the policy at different times. The focus of the analysis is on estimating the causal effect of these lotteries on the share of the vaccinated population. Using disaggregated data at the county level, I estimate causal quantities in a staggered adoption setting at different levels of aggregation (county, state, macro-region) by using the Penalized SCM from Abadie and L'Hour (2021)). The goal of this research is to add to the existing literature on health policy by examining how the effects of a treatment vary among counties. Specifically, this study will examine the average treatment effects among clusters of counties that have been grouped based on important socio-demographic characteristics. Additionally, researching the continued impact of a policy after it has ended can provide valuable information for policymakers, enabling them to compare the long-term results in different counties and states. Furthermore, examining the overall impact of a policy across four US regions can help policymakers identify areas that are meeting their expectations or falling short. The analysis in the dissertation aims to provide insight into the policy's effects at different aggregation levels (county, state, and macro-region levels) and periods of treatment. Moreover, I have investigated treatment effect heterogeneity with respect to some key socio-demographic covariates.

Synthetic control methods are often used to evaluate the intervention in spatial areas such as cities, regions or neighbourhoods, with treatment assigned to some specific area, but with possible second-round effects on contiguous units. Synthetic control methods can be used to evaluate the effect that the treatment had in the specific area, but it is often unclear how far the treatment's effect propagates. Common approaches consider separate estimation of treatment effects, disregarding the spatial structure of data and can lead to efficiency loss in spatial settings. In the fourth chapter of this dissertation, I propose to tackle this issue by developing a Bayesian Spatial Matrix Completion Method, that allows the estimation of treatment effects at different distances from the treatment site, accounting for the spatial structure of the data. In particular, I propose to impute the missing potential outcomes for the treated areas as the linear combination of control units, with coefficients that vary smoothly over the distances, following a Gaussian Process specification. Within this framework, I study the effect of the construction of the first line of the Florentine tramway network on the number of stores in the areas surrounding the tramway stops.

In several works from policy evaluation literature, the unit's outcomes are *censored by death*, meaning that they are not defined nor observed for the units who die. In such contexts, the usual approach is identifying the subgroup of units that would be surviving their assignment to the treatment, the so-called "always survivors" by using a principal stratification approach, see for example Frangakis and Rubin (2002), Zhang and Rubin (2003). In the third chapter of this dissertation, I extend the longitudinal principal stratification framework, first proposed by Bia et al. (2020) and propose a framework for the analysis of longitudinal outcomes, in which units can be censored at different times, with the main outcomes defined up to the moment of censoring. Within a bayesian longitudinal principal stratification framework, with units classified according to the longitudinal censoring potential outcomes, I estimate the causal effects for the units that would be alive at time t, irrespective of their treatment assignment. This novel framework has interesting relapses: first of all, it allows the estimation of principal causal effect even in presence of truncation in multiple post-treatment periods. Second, it allows studying the time trend of principal stratum membership and the time trend of the survival average treatment effect, to establish the presence of time patterns and transition probabilities between latent strata. Lastly, we can study differences across the key covariates within the identified longitudinal principal strata. The motivating application for this method is the evaluation of a longitudinal observational policy that aimed to ease access to the credit market for start-ups. The focus of the analysis is on the effect of such policies on the hiring decisions of the firms, to investigate whether assisted credit market access can improve entrepreneurship for vulnerable groups of the population (young citizens and women) by the subsidization of start-up projects and stimulate further job creation.

Finally, in the last part of this work, I will draw some conclusions on these works and propose further research themes related to these topics.

Chapter 1

Direct and spillover effects with the synthetic control method

1.1 Introduction

Synthetic Control Group (SCG) methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015) are an increasingly popular approach used to draw causal inference under the potential outcome framework (e.g., Rubin, 1974) in panel comparative case studies. In these studies, the outcome of interest is observed for a limited number of treated units, often only a single one, and for a number of control units, with respect to a number of periods both prior and after the assignment of the treatment. The SCG method focuses on causal effects for treated units: for each point in time after the assignment of the treatment, a weighted average of the observed potential outcomes of control units is used to reconstruct the potential outcomes under control for treated units. These weighted averages are named synthetic controls. The vector of weights is chosen by minimizing some distance between pre-treatment outcomes and covariates for the treated units and the weighted average of pre-treatment outcomes and covariates for the control units. See Abadie (2021) for a review of the empirical and methodological aspects of SCG methods.

In the last two decades, SCG methods have gained widespread popularity, and there has been a growing number of studies applying them to the investigation of the economic effects on particular locations of a wide range of events or interventions Initially, SCG methods have been used in panel studies where the outcome of interest is observed for a single treated unit (e.g., Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). Recently, they have been generalized to draw causal inference in panel studies where focus is on the average causal effects for multiple treated units (Cavallo et al., 2013; Acemoglu et al., 2016; Gobillon and Magnac, 2016; Kreif et al., 2016; Abadie and L'Hour, 2021). Additional important theoretical and conceptual contributions include the comparison of SCG methods with alternative approaches for program evaluation, the definition of synthetic control units and the development of new estimators (Doudchenko and Imbens, 2016; Xu, 2017; Athey et al., 2021; Bottmer et al., 2021).

In this methodological and applied causal inference literature, SCG methods have been implemented using the potential outcome approach under the Stable Unit Treatment Value Assumption (SUTVA), which rules out the presence of interference and hidden versions of treatments Rubin (1980). The no-interference component of SUTVA, which states that the treatment received by one unit does not affect the outcomes of any other unit, may be arguable in many studies, where the events or interventions of interest may produce their effect not only on the units that are exposed to them (direct effects), but also on other unexposed units (spillover effects). In the presence of interference, both scientists and policy makers may be interested not only in the direct effect of an intervention on the unit(s) where it actually takes place, but also in the effects that the same intervention may have – though in an indirect fashion – on other units not exposed to the intervention. Therefore, disentangling direct and spillover effects becomes the key objective of the analysis. However, the presence of interference entails a violation of the SUTVA, and makes causal inference particularly challenging.

Over the last years, causal inference in the presence of interference has been a fertile area of research. Important theoretical works have dealt with the formal definition of direct and spillover effects and with the development of design and inferential strategies to conduct causal inference under various types of interference mechanisms, in both randomized and observational studies (e.g., Hong and Raudenbush, 2006; Sobel, 2006; Hudgens and Halloran, 2008; Arpino and Mattei, 2016; Forastiere et al., 2021a; Papadogeorgou et al., 2019; Huber and Steinmayr, 2021). Despite such increasing interest, to the best of our knowledge, only the recent works by Cao and Dowd (2019) and Di Stefano and Mellace (2020) deal with the application of synthetic control methods to comparative case studies where the no-interference assumption is not plausible. In particular, Cao and Dowd (2019) introduce – under the assumption that spillover effects are linear in some unknown parameter – estimators for both direct treatment effects and spillover effects. They also investigate their asymptotic properties when the number of pre-treatment periods goes to infinity. Di Stefano and Mellace (2020) introduce a procedure, called "inclusive SCM", under which direct and spillover effects can be estimated using control units potentially affected by spillovers.

Motivated by the evaluation of causal effects of a new light rail line recently built in Florence (Italy) on the commercial vitality of the surrounding area, we propose to contribute to the nascent literature on the use of the SCG approach in a setting with interference. To that end, our work makes both methodological and substantive contributions.

From a methodological perspective, we formally define direct and spillover effects in comparative studies where the outcome of interest is observed for a single treated unit, and a number of control units, for a number of periods before and after the assignment of the treatment. We introduce two types of spillover effects. The first type represents the effect of the treatment on untreated units belonging to treated unit's neighborhood. The second type would flow from untreated units towards the treated unit, in the hypothetical scenario where the untreated units were exposed to the treatment rather than the actual treated unit, representing what would have happened to the treated units if it wasn't treated and instead, the treatment was assigned to some of the neighboring units. In a sense, we can view this type of spillover effect as an "unrealized spillover effect." These causal estimands are defined under a partial interference assumption (**?**), which states that interference takes place between units located near to each other, but not between units that are sufficiently faraway from one another. Under partial interference, we use the penalized SCG estimator recently developed by Abadie and L'Hour (2021) to estimate direct effects and spillover effects of the first type by exploiting information on control units who do not belong to treated unit's neighborhood.

A model-based imputation method is used to estimate the unrealized spillover effects. A bootstrap procedure is used for inference, based on the idea that the set of control units can be reasonably viewed as a sample of control units from a super-population.

From a substantive perspective, we assess the direct effect of a new light rail line built in Florence (Italy) on the retail density of the street where it was built, its spillover on neighboring streets, and the spillover on the treated street that would have emanated from hypothetical, alternative locations of the light rail within the same neighborhood. We measure the retail density of a street using the number of stores every five hundred meters. This kind of application is original with respect to the previous field literature, which has often examined whether the creation of urban rail infrastructure is accompanied by changes in real estate values or gentrification of the area (e.g., Cervero and Landis, 1993; Baum-Snow and Kahn, 2000; Bowes and Ihlanfeldt, 2001; Kahn, 2007; Pagliara and Papa, 2011; Grube-Cavers and Patterson, 2015; Budiakivska and Casolaro, 2018; Delmelle and Nilsson, 2020) and, only more seldom, whether it is accompanied by a higher firm density (Mejia-Dorantes et al., 2012; Pogonyi et al., 2021) or by the settlement of new retailers (Schuetz, 2015; Credit, 2018).

Nevertheless, it is worth noting that not all these empirical studies are fully embedded in an explicit causal framework, and that none of them addresses the issue of spillovers.

The chapter is organized as follows. Section 1.2 describes the application that motivates the methodological development we propose and the available data. Section 1.3 presents the methodology. In Section 1.4, we discuss how the methodology is applied to study the case of the Florentine light rail and present the results of the analysis. Section 1.5 concludes the chapter.

1.2 Motivating application and related data

1.2.1 A new light rail in Florence, Italy

In addition to being a renowned art capital, Florence is also a city with nearly 400,000 residents and the hub of a wide commuting area. Away from the artworks and the pedestrian footpaths packed with store windows in the city center, the thoroughfares of peripheral Florence are often congested with cars. From the early 1900s, the city of Florence developed an extensive public tram network on street running tracks. Such network was dismissed in 1958 in favor of public bus transport. In the following decades, the city of Florence suffered from soaring private motor vehicle transport, which led to congested traffic and undermined both the effectiveness and the attractiveness of public transport. In order to face these issues, the project of a new light rail network has been discussed for a long time, in a climate of doubt about the possibility of raising the necessary funds for the work. Moreover, there has been a strong debate about the appropriateness of this solution compared to others, also in view of the discomfort and discontent that long-lasting construction sites would have created in the areas exposed to the intervention. Nevertheless, a tram network project took shape during the 1990s.

The planned network mostly runs on reserved tracks, thus guaranteeing a more reliable public transport service, especially on long-distance journeys. Once completed, it will develop radially from the city center towards all the main surrounding suburbs.

In the everyday slang of Florentines, the brand new light rail continues to be referred to by the old-fashioned term “tramway.” The first tramway line of the network was constructed between 2006 and 2010. It connects the main railway station, in the city center, with the Southwestern urban area. The most intensive phase of works, when tracks were laid and stations were built, started in 2007. The first line was completed in 2010. It has a total length of 7.6 kilometers, with stops approximately every 400 meters. After the inauguration of this line, some previous long-distance bus services were suppressed, whereas other ones were re-designed as short-distance services to ease the access to the tramway from adjacent areas. The completion of the planned light rail network requires the construction of four additional lines. The construction of two of these lines started in 2014 and was completed in 2018, while the remaining two lines are at a very preliminary stage. The analysis in this work looks at the 2004-2013 period and focuses on the first line of the tramway. In particular, we consider the section of the line that goes along Talenti St. (1.2 kilometers, 3 stops: Talenti, Batoni, and Sansovino), one of the main thoroughfares in the densely inhabited Soutwestern urban neighborhood of Legnaia-Isolotto (Legnaia hereinafter). There are other important thoroughfares and streets in Legnaia, most of which run parallel to Talenti St. but do not host light rail tracks and stations. They are: Pollaiolo St. (about 300 meters far from Talenti St.); Pisana St. (450 meters far); Baccio da Montelupo St. (500 meters far), Scandicci St. (650 meters far); and Magnolie St. (650 meters far). For each of these streets we consider a section of maximum length of 1.2 kilometers, which we select to be geographically the closest to Talenti St.. All these streets fall within 800 meters range from the light rail and its transit stations (corresponding to a walking distance of about 10 minutes), which is considered a reasonable area of impact by the field literature (Guerra et al., 2012). It is worth noting that, unlike previous studies, where streets within a given radius from transit infrastructures are aggregated to form a cluster level unit, we consider each street as a distinct statistical unit.

1.2.2 Conjectures on how light rail could affect the streets' retail activity

Light rail is generally expected to raise accessibility through the improvement of transit times between different points within a urban area (e.g., see Papa and Bertolini, 2015, and the literature review therein). However, citywide accessibility improvements are likely to occur in the presence of an extensive light rail network. This is not the case in our study, where there is only one light rail line, which was mainly conceived to make access to the city center easier from one particular section of urban periphery. A single line like the one subject to our study is expected to yield a rather localized accessibility improvement. At the same time, the light rail may be expected to trigger a process of revitalization of peripheral areas and of the retail sector therein. This may occur once the light rail is in operation thanks to high flows of transit users and renewed site image. However, the previous empirical literature suggests that the boost of the local retail sector, if any, can be small or transitory (Mejia-Dorantes et al., 2012; Schuetz, 2015; Credit, 2018).

Before the light rail inauguration, construction works may temporarily undermine the area's attractiveness and livability. Faced with the light rail construction site in front of their shop windows, incumbent store owners often complain about the risk of lost opportunities owed to poor site image, traffic diversions, very limited street parking, and so forth. For the store owners located on other thoroughfares belonging to the same neighborhood of Talenti St., but with no construction site, the story might go the other way around during the tramway construction, with increased opportunities owed to temporarily higher flows guaranteed by traffic diversions, unchanged image and street parking possibilities, increased relative competitiveness, and so forth.

When the new infrastructure goes into operation in a given site, the prospects of the commercial environment of adjacent sites are hard to envisage. On the one hand, they could also benefit from having the light rail at walking distance, which may increase the footfall for the retailers, constituting a positive spillover effect. On the other hand, they might return to business as usual, or even be crowded out and lose footfall due to the soaring relative attractiveness of the street where stations are located, which may then constitute a negative spillover effect (Credit, 2018; Pogonyi et al., 2021).

The effect of the tramway on the commercial environment of a given shopping site may be heterogeneous depending on the different types of stores. Since stores may belong to a high number of categories, an attractive way to group them into few meaningful classes is to distinguish between purveyors of non-durable goods/frequent-use services (non-durables hereinafter) and purveyors of durable goods/seldom-use services (durables hereinafter). This distinction may help characterize in greater detail the effects of the light rail on a urban neighborhood's retail sector. Indeed, it reflects a difference in the frequency of purchase of the two types of goods and services, which is very high for non-durables and relatively low for durables. It is also correlated with the customers' willingness to travel to purchase each type of goods and services: such willingness is low for non-durables, which are usually purchased in one's vicinity, and high for durables, which may see customers ready to bear some costs to patronize less accessible stores every once in a while (Brown, 1993; Klaesson and Öner, 2014; Larsson and Öner, 2014).

1.2.3 Data

The dataset used to examine the impact of the new ligth rail on the local retail environment includes information on 6 streets in the peripheral urban neighborhood of Legnaia (Talenti St., Pisana St., Pollaiolo St., Baccio da

Montelupo St., Scandicci St., and Magnolie St.) and on 38 further thoroughfares and streets of Florence, clustered in other 10 peripheral neighborhoods that are far from Legnaia. The definition of urban neighborhoods is based on the areas identified by the Real Estate Observatory of the Italian Ministry of Finance. We do not consider any street in the city center, as its commercial environment is completely different from what can be found in the surrounding residential neighborhoods.

Background and outcome variables for each street originate from the Statistical Archive of Active Firms (SAAF, English translation of ASIA, the Italian acronym for "Archivio Statistico delle Imprese Attive"). The SAAF is held by the Italian National Istitute of Statistics (ISTAT). This dataset is available from 1996 onwards. It collects some basic, individual information on all the active local units of firms, including the exact location of the activity and the sector of activity (classified according the Statistical Classification of Economic Activities in the European Community, usually referred to as NACE). We construct background and outcome variables for each street as follows. First, we select firms that are active in the retail sector in the city of Florence. Second, we further select only those stores having their shop windows on the streets involved in the study or that are located within an extremely short distance from such streets (50 meters). Third, in line with the reasoning developed in the previous subsection, we classify each of these stores into a NACE sector of activity in order to elicit the product/service these stores sell, and group them into two categories: purveyors of durable goods (or seldom-use services); and purveyors of non-durable goods (or frequent-use services). For each street and year, we finally construct background and outcome variables aggregating information across stores belonging to the same category. In our application, we focus on the following two outcome variables: number of purveyors of durable goods every 500 meters; number of purveyors of non-durable goods every 500 meters. Figure 5.1 in Appendix 1 shows the observed value of these variables over the time period 1996-2014. The left-hand vertical line marks the start of light rail construction, the right-hand vertical line marks the start of its operation. These descriptive graphs suggest that, in Talenti St., the number of purveyors of non-durable goods (every 500 meters) increases after the tramway goes into operation. On the other hand, the number of stores selling durables on Talenti St. slightly increases during the early phase of construction, but starts to diminish afterwards. On Pollaiolo St., the number of purveyors of non-durables grows during construction and wanes during the operational period. After an initial jump, Pisana St. retains stores selling durables but loses some purveyors of non-durables when the light rail is operational. Also Baccio da Montelupo St. hosts a higher number of outlets during construction, followed by a later loss. On Scandicci St., the number of purveyors is overall stable. Finally, Magnolie St. sees a continuous decrease in the number of stores selling durables, while the decline in the number of purveyors of non-durables begins as the light rail service starts.

1.3 Methodology

1.3.1 Potential outcomes and observed outcomes

We consider a panel data setting with $1+N$ units partitioned into $1+K$ clusters and observed in time periods $t = 1,\ldots,T$. Let $1+N_1$ and N_k be the number of units in cluster 1 and in cluster k, $k \in \{2,\ldots,1+K\}$, respectively: $1+N = (1+N_1)+\sum_{k=2}^{1+K} N_k$; and let $\mathcal{N}_k$ denote the set of numbers indexing units that belong to cluster k, $k = 1,2,\ldots,1+K$. For $k = 1,\ldots,1+K$, let $\mathbf{w}_{kt} = [w_{ki,t}]'_{i\in\mathcal{N}_k}$ be a cluster treatment vector at time t, $t = 1,\ldots,T$. Generally, for $t = 1,\ldots,T$, $\mathbf{w}_{kt} \in \{0,1\}^{\mathbb{I}\{k=1\}+N_k}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. In this work we focus on scenarios where a single unit is exposed to the intervention of interest from a given time period, say T_0+1 with $1 < T_0 < T$, onwards, so that, for $t = 1,\ldots,T_0$, $[\mathbf{w}_{1t},\ldots,\mathbf{w}_{(1+K)t}] = [\mathbf{0}_{1+N_1},\ldots,\mathbf{0}_{N_{1+K}}]$, where $\mathbf{0}_r$ denotes the zero vector in $\mathbb{R}^r$; and for $t = T_0+1,\ldots,T$, $[\mathbf{w}_{1t},\ldots,\mathbf{w}_{(1+K)t}]$ is constant over time and it is a point in $\mathcal{W} = \{[\mathbf{w}_1,\ldots,\mathbf{w}_{(1+K)}]' \in \{0,1\}^{1+\sum_{k=1}^{1+K} N_k}$ with $\mathbf{w}_k = [w_{ik}]_{i\in\mathcal{N}_k}, k = 1,\ldots,1+K : \sum_{k=1}^{1+K}\sum_{i\in\mathcal{N}_k} w_{ik} = 1\}$.

In our motivating study, units are streets of Florence and clusters are naturally defined by urban neighborhoods. Our dataset includes information on $1+N = 1+43 = 44$ streets clustered into $1+K = 1+10 = 11$ urban neighborhoods of Florence, which are observed from 1996 to 2014. Only one of these streets, namely Talenti St., which is in the Legnaia neighborhood, is exposed to the intervention of interest: the construction of a new light rail line. Since construction works started in 2006 and ended in 2010, we have ten pre-treatment, four treatment, and five post-treatment years with $T_0 = 10$ and $T = 19$. In addition to Talenti St., the Legnaia neighborhood, which we refer to as cluster 1, comprises five streets; Pollaiolo St., Pisana St., Scandicci St., Magnolie St., and Baccio da Montelupo St., which we index by $i = 2,3,4,5,6$, with $i \in \mathcal{N}_1$, respectively. The remaining 10 urban neighborhoods, which comprise 38 streets overall, are sufficiently far from Legnaia. See Figure 5.2 in the Appendix 1 for a stylized map.

Under the assumption that there is no hidden versions of treatment (***Consistency Assumption***, **Rubin**, **1980**), let

$$Y_{ik,t}([\mathbf{w}_{k1},\ldots,\mathbf{w}_{kT_0},\mathbf{w}_{k(t_0+1)},\ldots\mathbf{w}_{kT}]_{k=1}^{1+K})$$

denote the potential outcome for unit i in cluster k at time t under treatment assignment $(1+N)\times T$ matrix $[\mathbf{w}_{k1},\ldots,\mathbf{w}_{kT_0},\mathbf{w}_{k(T_0+1)},\ldots\mathbf{w}_{kT}]_{k=1}^{1+K}$, where $\mathbf{w}_{k1}=\cdots=\mathbf{w}_{kT_0}=\mathbf{0}_{\mathbb{I}\{k=1\}+N_k}$ and $\mathbf{w}_{k(T_0+1)}=\cdots=\mathbf{w}_{kT}\equiv\mathbf{w}_k$, with $\mathbf{w}_k$ such that $\sum_{i\in\mathcal{N}_k} w_{ik}\in\{0,1\}$.

We make the assumption of "no-anticipation of the treatment" (e.g. Abadie et al., 2010)

which amount to stating that the intervention has no effect on the outcome before the treatment period, $T_0+1,\ldots,T$:

Assumption 1. *(No anticipation of the treatment). For all $k=1,\ldots,1+K$, $i\in\mathcal{N}_k$, and $t=1,\ldots,T_0$*

$$
\begin{aligned}
&Y_{ik,t}([\mathbf{0}_{\mathbb{I}\{k=1\}+N_k},\ldots,\mathbf{0}_{\mathbb{I}\{k=1\}+N_k},\mathbf{w}_{k(T_0+1)},\ldots\mathbf{w}_{kT}]_{k=1}^{1+K})=\\
&\qquad Y_{ik,t}([\mathbf{0}_{\mathbb{I}\{k=1\}+N_k},\ldots,\mathbf{0}_{\mathbb{I}\{k=1\}+N_k},\mathbf{0}_{\mathbb{I}\{k=1\}+N_k},\ldots\mathbf{0}_{\mathbb{I}\{k=1\}+N_k}]_{k=1}^{1+K})
\end{aligned}
$$

In this study, the no anticipation of the treatment assumption appears to be plausible. In 2000, the city administration announced the construction of the first line of the light rail network, but things soon turned out to be less easy than expected. The first tender for works attracted the interest of no construction companies. The outcome of the second call for tenders, in 2001, was the subject of a legal dispute lasting several years, giving rise to quite a few doubts – in a public opinion that remained divided on the project – as to whether and when a new light rail would ever exist in the city. A third tender followed and the work was awarded to an unexpected consortium of those companies that had fought each other during the previous legal dispute. In light of such a troubled gestation, it is rather difficult to envision what kind of anticipatory behaviors, if any, might have been put in place by private economic agents, especially by the store owners that are the subject of the analysis proposed in the current work.

Under the assumption of no anticipation of the treatment, in our setting where the intervention occurs from time T_0+1 onwards, we can re-write potential outcomes for unit i in cluster k at time t as function of the $(1+N)-$dimensional treatment vector at time t only,

$$
Y_{ik,t}([\mathbf{w}_{k1},\ldots,\mathbf{w}_{kT_0},\mathbf{w}_{k,T_0+1},\ldots\mathbf{w}_{kT}]_{k=1}^{K})=Y_{ik,t}([\mathbf{w}_{1t},\ldots,\mathbf{w}_{(1+K)t}]).
$$

Moreover, because for $k=1,\ldots 1+K$, $\mathbf{w}_{kt}=\mathbf{0}_{\mathbb{I}\{k=1\}+N_k}$ for $t=1,\ldots,T_0$ and $\mathbf{w}_{kt}\equiv\mathbf{w}_k$, with $\mathbf{w}_k$ such that $\sum_{i\in\mathcal{N}_k} w_{ik}\in\{0,1\}$ for $t=T_0+1,\ldots,T$, we can omit the subscript t from the treatment assignment vector. Therefore, for each unit i in cluster k, $k=1,\ldots 1+K$, the observable potential outcomes are $Y_{ik,t}([\mathbf{0}_{1+N_1},\ldots,\mathbf{0}_{N_{1+K}}])$ for $t=1,\ldots,T_0$, and $Y_{ik,t}([\mathbf{w}_1,\ldots,\mathbf{w}_{1+K}])$ with $\mathbf{w}_k$ such that $\sum_{i\in\mathcal{N}_k} w_{ik}\in\{0,1\}$ for $t=T_0+1,\ldots,T$.

Let $[\mathbf{W}_1,\ldots,\mathbf{W}_{1+K}]$ be the treatment vector we observe from time T_0+1 on-wards and let $Y_{ik,t}$ be the observed outcome for unit i in cluster k at time t, $t = 1, \ldots, T_0, T_0 + 1, \ldots, T$. Under consistency and no anticipation of treatment, $Y_{ik,t} = Y_{ik,t}([\mathbf{0}_{1+N_1}, \ldots, \mathbf{0}_{N_{1+K}}])$ for $t = 1, \ldots, T_0$ and $Y_{ik,t} = Y_{ik,t}([\mathbf{W}_1, \ldots, \mathbf{W}_{1+K}])$, for $t = T_0 + 1, \ldots, T$.

When the population can be partitioned into clusters, it is often plausible to invoke the partial interference assumption (Sobel, 2006). Such assumption states that interference may occur within, but not between, groups. Let $\mathbf{w}_k^{(-i)}$ denote a treatment assignment vector for the units other then unit i in cluster k:

$$\mathbf{w}_k^{(-i)} = [w_{1k}, \ldots, w_{(i-1)k}, w_{(i+1)k}, \ldots, w_{\mathbb{I}\{k=1\}+N_k}]'$$

, $k = 1, \ldots, 1 + K$. Then we can formally formulate the partial interference assumption as follows:

Assumption 2. *(Partial Interference). For $t = T_0 + 1, \ldots, T$, for all*

$$[\mathbf{w}_1, \ldots, w_{ki}, \mathbf{w}_k^{(-i)}, \ldots, \mathbf{w}_{1+K}] \quad \textit{and} \quad [\mathbf{w}_1^*, \ldots, w_{ki}^*, \mathbf{w}_k^{*(-i)}, \ldots, \mathbf{w}_{1+K}^*]$$

with $w_{ki} = w_{ki}^*$ and $\mathbf{w}_k^{(-i)} = \mathbf{w}_k^{*(-i)}$,

$$Y_{ik,t}([\mathbf{w}_1, \ldots, w_{ki}, \mathbf{w}_k^{(-i)}, \ldots, \mathbf{w}_{1+K}]) = Y_{ik,t}([\mathbf{w}_1^*, \ldots, w_{ki}^*, \mathbf{w}_k^{*(-i)}, \ldots, \mathbf{w}_{1+K}^*])$$

for all $i \in \mathcal{N}_k$, $k = 1, \ldots, 1 + K$.

Partial interference implies that potential outcomes for unit i in cluster k, $i \in \mathcal{N}_k$, only depend on its own treatment status and on the treatment statuses of the units belonging to the same cluster/neighborhood as unit i, but they do not depend on the treatment statuses of the units belonging to different clusters/neighborhoods. Therefore, partial interference allows us to write $Y_{ik,t}([\mathbf{w}_1, \ldots, \mathbf{w}_k, \ldots, \mathbf{w}_{1+K}]) \equiv Y_{ik,t}([\mathbf{w}_1, \ldots, w_{ki}, \mathbf{w}_k^{(-i)}, \ldots, \mathbf{w}_{1+K}])$ as $Y_{ik,t}(\mathbf{w}_k) \equiv Y_{ik,t}(w_{ki}, \mathbf{w}_k^{(-i)})$ for all $i \in \mathcal{N}_k$, $k = 1, \ldots, 1 + K$, and for $t = T_0 + 1, \ldots, T$.

In our application study, where streets are partitioned into clusters defined by urban neighborhoods, it is rather plausible to assume that interference occurs within streets belonging to the same neighborhood, but not between streets belonging to different, geographically distant, urban neighborhoods. Indeed, we can reasonably expect that customers patronizing stores in a given peripheral area will hardly switch over to other distant, peripheral areas because of a single light rail line connecting only one of these peripheries with the city center, but with none of the other peripheries.

Under partial interference, for $t = T_0 + 1, \ldots, T$, the observed outcome for unit i in cluster k, $k = 1, \ldots, 1+K$, is $Y_{ik,t} = Y_{ik,t}(\mathbf{W}_k) \equiv Y_{ik,t}(W_{ki}, \mathbf{W}_k^{(-i)})$. With no loss of generality, henceforth, we assume that unit 1 in cluster 1 is the single treated unit from time T_0+1 on-wards, so that, $\mathbf{W}_1 = [1, \mathbf{0}_{N_1}]'$ and $\mathbf{W}_k = \mathbf{0}_{N_k}$, for $k = 2, \ldots, 1+K$. For $i \in \mathcal{N}_1$, let $\mathbf{e}^{(i)}_{N_1}$ be a N_1-dimensional vector with all of its entries equal to 0 except the entry corresponding to unit i, which is equal to 1. Let $A \setminus B$ denote the subtraction of sets A and B, A minus B. Therefore, for $t = T_0+1, \ldots, T$, we observe $Y_{11,t} = Y_{11,t}(1, \mathbf{0}_{N_1})$, $Y_{1i,t} = Y_{it}(0, \mathbf{e}^{(1)}_{N_1})$ for all $i \in \mathcal{N}_1 \setminus \{1\}$, and $Y_{ik,t} = Y_{ik,t}(0, \mathbf{0}_{N_k-1})$ for all $i \in \mathcal{N}_k$, $k = 2, \ldots, 1+K$.

Throughout the chapter, we refer to $Y_{11,t}(0, \mathbf{0}_{N_1})$ and $Y_{1i,t}(0, \mathbf{0}_{N_1})$, $i \in \mathcal{N}_1 \setminus \{1\}$, for $t = T_0+1, \ldots, T$ as control potential outcomes for the treated unit and for units who belong to the treated unit's cluster, respectively, and to units who do not belong to the treated unit's cluster as control units.

The observed outcomes at time $t = 1, \ldots, T_0$, $Y_{ik,t}$, $i \in \mathcal{N}_k$, $k = 1, \ldots, 1+K$, are pre-treatment outcomes. In addition to them, we observe a vector of time- and unit-specific covariates, $\mathbf{C}_{ik,t} = [C_{ik,t,1}, \ldots, C_{ik,t,P}]$, $i \in \mathcal{N}_k$, $k = 1, \ldots, 1+K$, $t = 1, \ldots, T_0$, that is, variables that we can reasonably assume to be unaffected by the intervention. Using information on unit-level pre-treatment outcomes and covariates, for each unit i in cluster k, we construct neighborhood-level pre-treatment outcomes, $Y_{\mathcal{N}_{ik},t}$, and neighborhood-level unit×time specific covariates, $\mathbf{C}_{\mathcal{N}_{ik},t} = [C_{\mathcal{N}_{ik},t,1}, \ldots, C_{\mathcal{N}_{ik},t,P}]$, as average of the unit-level pre-treatment outcomes for units belonging to unit i's cluster/neighborhood:

$$Y_{\mathcal{N}_{ik},t} = \frac{1}{\mathbb{I}\{k=1\} + N_k - 1} \sum_{i' \in \mathcal{N}_k \setminus \{i\}} Y_{i'k,t},$$

and

$$C_{\mathcal{N}_{ik},t,p} = \frac{1}{\mathbb{I}\{k=1\} + N_k - 1} \sum_{i' \in \mathcal{N}_k \setminus \{i\}} C_{i'k,t,p}, \quad p = 1, \ldots, P,$$

$k = 1, \ldots, 1+K$, $t = 1, \ldots, T_0$.

1.3.2 Causal estimands

In a setting where only the first unit (Talenti St.) in the first cluster (Legnaia neighborhood) is exposed to the intervention after time point T_0 (with $1 \leq T_0 < T$), and under the assumption of partial interference, we are interested in the following direct and spillover causal effects at time points $t = T_0+1, \ldots, T$.

We define the (individual) direct causal effect of treatment 1 versus treatment 0 for the treated unit/street as

$$\tau_{11,t} = Y_{11,t}(1, \mathbf{0}_{N_1}) - Y_{11,t}(0, \mathbf{0}_{N_1}) \quad t = T_0+1, \ldots, T. \tag{1.1}$$

For all $i \in \mathcal{N}_1 \setminus \{1\}$, let

$$\delta_{1i,t} = Y_{1i,t}(0, \mathbf{e}^{(1)}_{N_1}) - Y_{1i,t}(0, \mathbf{0}_{N_1})$$

be the individual spillover causal effect of treatment 1 versus treatment 0 at time t on unit i belonging to cluster 1, the treated unit's cluster. We define the average spillover causal effect at time t as

$$\delta^{\mathcal{N}_1}_t = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \delta_{1i,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \left[Y_{1i,t}(0, \mathbf{e}^{(1)}_{N_1}) - Y_{it}(0, \mathbf{0}_{N_1}) \right]. \tag{1.2}$$

Finally, we define the unrealized spillover causal effect at time t of unit i in cluster 1, $i \in \mathcal{N}_1$, on the treated unit as

$$\gamma^{(i)}_{11,t} = Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1}) - Y_{11,t}(0, \mathbf{0}_{N_1}). \tag{1.3}$$

The quantity $\gamma^{(i)}_{11,t}$, measures what the spillover effect on unit 1 in cluster 1 could have been in the hypothetical scenario where another unit, say unit i, belonging to the same cluster as the treated unit 1 was exposed to the intervention rather than unit 1. In our application study, $\gamma^{(i)}_{11,t}$ is the effect of the light rail on Talenti St. if the light rail was not located on Talenti St. but on another street belonging to Talenti St.'s urban neighborhood (Legnaia neighborhood). We can interpret $\gamma^{(i)}_{11,t}$ as the spillover that unit 1, namely Talenti St., has not realized precisely because of its exposure to treatment. It recalls the concept of opportunity cost used in public economics for the comparative study of alternative investment plans. It is worth noting that the unrealized spillover effect is defined similarly to the individual spillover effect for the neighboring units, and in fact, it is representing a similar quantity (e.g: what would be my potential outcome if the tramway is located somewhere in my neighborhood but not here?). However, we prefer to treat differently the definition and the estimation of the two spillover effects, as for the unrealized spillover effect we don't observe any potential outcome in the definition of the effect, and thus the inferential procedure could be different.

The difference between the direct effect and the unrealized spillover,

$$\tau_{11,t} - \gamma^{(i)}_{11,t} = Y_{11,t}(1, \mathbf{0}_{N_1}) - Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1}), \qquad i \in \mathcal{N}_1 \setminus \{1\} \tag{1.4}$$

may provide useful insights on whether, among a set of alternatives, the original treatment allocation choice has brought about a gain or a loss for the treated unit. If $\tau_{11,t} > \gamma^{(i)}_{11,t}$, the actual treatment allocation brought about a gain for unit 1 with respect to unit i; if $\tau_{11,t} < \gamma^{(i)}_{11,t}$, then some alternative allocation of the intervention within the cluster would have been preferable for the treated unit; if $\tau_{11,t} = \gamma^{(i)}_{11,t}$, an alternative allocation of the intervention, where unit i rather than unit 1 were exposed to the treatment, would have been equivalent to the actual one for the treated unit.

Here we focus on average unrealized spillover causal effects:

$$\gamma_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \gamma^{(i)}_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1}) - Y_{11,t}(0, \mathbf{0}_{N_1}), \tag{1.5}$$

for $t = T_0 + 1, \ldots, T$.

Two remarks on the causal effects we are interested in are in order. First, it is worth noting that we define direct and spillover effects as comparisons between potential outcomes under alternative cluster treatment vectors. The literature on causal inference under partial interference has generally focused on average direct and spillover effects, defined as comparisons between average potential outcomes under alternative treatment allocation strategies (e.g., Hudgens and Halloran, 2008; Papadogeorgou et al., 2019). Second, we are not interested in assessing causal effects for units/streets belonging to clusters/urban neighborhoods different from the treated unit's cluster (Legnaia), but the availability of information on them is essential for inference, as we will show in the next Sections. We can rewrite the (individual) direct causal effect for the treated unit in Equation (1.1) and the average spillover causal effect in Equation (1.2) at time t, $t = T_0 + 1, \ldots, T$, as function of the observed outcomes:

$$\tau_{11,t} = Y_{11,t} - Y_{11,t}(0, \mathbf{0}_{N_1}) \quad \text{and} \quad \delta^{\mathcal{N}_1}_t = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \left[Y_{1i,t} - Y_{1i,t}(0, \mathbf{0}_{N_1}) \right].$$

These relationships make it clear that we need to estimate $Y_{11,t}(0, \mathbf{0}_{N_1})$ and $Y_{1i,t}(0, \mathbf{0}_{N_1})$ for $i \in \mathcal{N}_1 \setminus \{1\}$ to get an estimate of $\tau_{11,t}$ and $\delta^{\mathcal{N}_1}_t$. The unrealized spillover by the treated unit in Equation (1.5), $\gamma^{(i)}_{11,t}$, depends on two unobserved potential outcomes, $Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1})$, $i \in \mathcal{N}_1 \setminus \{1\}$, and $Y_{1t}(0, \mathbf{0}_{N_1})$, and thus we need to estimate both of them to get an estimate of $\gamma^{(i)}_{11,t}$, and thus, of $\gamma_{11,t}$.

1.3.3 SCG estimators of direct and average spillover effects

Under partial interference (Assumption 4), we creatively exploit information on units within clusters different from the treated unit's cluster to draw inference on direct effects, average spillover effects and unrealized spillover effects using the SCG approach originally proposed by Abadie and Gardeazabal (2003); Abadie et al. (2010), and further developed by

Abadie (2021).

Several exiting SCG approaches exploit the idea of a stable relationship over time between the outcome of the treated units and the outcome of the control units in the absence of intervention (stable patterns across units, e.g., Abadie and Gardeazabal, 2003; Abadie et al., 2010; Doudchenko and Imbens, 2016; Abadie, 2021). Similarly, our the method exploits stable patterns across units belonging to different clusters. Specifically, for each unit i in cluster 1, $i \in \mathcal{N}_1$, we assume that the relationship between the outcome of unit i, $Y_{1i,t}$, and the outcomes of control units, $Y_{ki',t}$, $i' \in \mathcal{N}_k$, $k \neq 1$, is stable over time. This type of stable pattern implies that:

1. the same structural process drives both the outcomes of units in control clusters (clusters of units who do not belong to the treated unit's cluster) as well as the outcomes of the treated unit and its neighbors in absence of treatment

2. the outcomes of control units and their neighbors are not subject to structural shocks during the sample period of the study.

Under these assumptions, building on Abadie et al. (2010), we propose to impute the missing control potential outcomes for the treated unit and the units who belong to the treated unit's cluster as weighted average of outcomes of control units. Formally, for each unit i in cluster 1, $i \in \mathcal{N}_1$,

$$\widehat{Y}_{1i,t}(0, \mathbf{0}_{N_1}) = \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \omega_{ki'}^{(i)} Y_{ki',t} \qquad t = T_0 + 1, \ldots, T,$$

where $\omega_{ki'}^{(i)}$ are weights such that, for each $i \in \mathcal{N}_1$,

$$\omega_{ki'}^{(i)} \geq 0 \quad \text{for all } i' \in \mathcal{N}_k, k = 2, \ldots, 1+K$$

and

$$\sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \omega_{ki'}^{(i)} = 1.$$

For each unit i in cluster 1, $i \in \mathcal{N}_1$, the set of weights

$$\boldsymbol{\omega}^{(i)} = \left[\{\omega_{2i'}^{(i)}\}_{i' \in \mathcal{N}_2}, \ldots, \{\omega_{(1+K)i'}^{(i)}\}_{i' \in \mathcal{N}_{1+K}} \right]'$$

defines the *synthetic control unit* of unit i.

The choice of the weights, $\boldsymbol{\omega}^{(i)}$, is clearly an important step in SCMs. The key idea is to construct synthetic controls that best resemble the characteristics of the units in the treated cluster before the intervention. Unfortunately, the problem of finding a synthetic control that best reproduces the characteristics of a unit may not have a unique solution. We face this challenge using the penalized synthetic control estimator recently developed by Abadie and L'Hour (2021). In our setting, the penalized synthetic control estimator penalizes pairwise discrepancies between the characteristics of units in the treated cluster and the characteristics of the units belonging to untreated clusters that contribute to their synthetic controls.

Let $\mathbf{D}_{ki} = \left[Y_{ki,1}, \ldots, Y_{ki,T_0}, Y_{\mathcal{N}_{ki},1}, \ldots, Y_{\mathcal{N}_{ki},T_0}, \mathbf{C}_{ki,1}, \ldots, \mathbf{C}_{ki,T_0}, \mathbf{C}_{\mathcal{N}_{ki},1}, \ldots, C_{\mathcal{N}_{ki},T_0}\right]'$ be a vector of pre-treatment individual- and neighborhood- level outcomes and covariates for a unit i in cluster k, $i \in \mathcal{N}_k$. For each unit i in the treated cluster 1, and given a positive penalization constant $\lambda^{(i)}$, the penalized synthetic control vector of weights

$$\widehat{\boldsymbol{\omega}}^{(i)} = \left[\{\widehat{\omega}^{(i)}_{2i'}\}_{i' \in \mathcal{N}_2}, \ldots, \{\widehat{\omega}^{(i)}_{(1+K)i'}\}_{i' \in \mathcal{N}_{1+K}}\right]'$$

is chosen by solving the following optimization problem:

$$\arg\min_{\boldsymbol{\omega}^{(i)} \in \boldsymbol{\Omega}} \left\| \mathbf{D}_{1i} - \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \mathbf{D}_{ki'} \omega^{(i)}_{ki'} \right\|^2 + \lambda^{(i)} \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \|\mathbf{D}_{1i} - \mathbf{D}_{ki'}\|^2 \tag{1.6}$$

subject to

$$\omega^{(i)}_{ki'} \geq 0 \quad \forall i' \in \mathcal{N}_k; k = 2, \ldots, 1+K; \qquad \text{and} \qquad \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \omega^{(i)}_{ki'} = 1,$$

where $\|\cdot\|$ is the L^2−norm: $\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}}$ for $\mathbf{v} \in \mathbb{R}^r$ (see Abadie and L'Hour, 2021, for details on the costruction of the weights). It is worth noting that the use of the L^2-norm implies that the same importance is given to all pre-treatment individual- and neighborhood- level outcomes and covariates as predictors of the missing outcome.

Under some regularity conditions, if $\lambda^{(i)}$ is positive, then the optimization problem in Equation (2.8) has a unique solution (see Theorem 1 in Abadie, 2021). The penalization term defines a trade-off between aggregate fit and component-wise fit: the penalized synthetic control estimator becomes the synthetic control estimator originally introduced by Abadie and Gardeazabal (2003); Abadie et al. (2010) as $\lambda^{(i)} \to 0$; and the one-match nearest-neighbor matching with replacement estimator proposed by Abadie and Imbens (2006) as $\lambda^{(i)} \to \infty$.

Given an estimate of the weights, $\widehat{\boldsymbol{\omega}}^{(i)}$ for each unit i in the treated cluster 1, we estimate the direct effects for the treated unit, $\tau_{11,t}$, and the average spillover causal effects $\delta_t^{\mathcal{N}_1}$ $t = T_0 + 1, \ldots, T$, as follows:

$$\widehat{\tau}_{11,t} = Y_{11,t} - \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \widehat{\omega}_{ki'}^{(1)} Y_{ki',t} \tag{1.7}$$

and

$$\widehat{\delta}_t^{\mathcal{N}_1} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \widehat{\delta}_{1i,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \left[Y_{1i,t} - \sum_{k=2}^{1+K} \sum_{i' \in \mathcal{N}_k} \widehat{\omega}_{ki'}^{(i)} Y_{ki',t} \right]. \tag{1.8}$$

In the literature, various approaches have been proposed to quantify uncertainty of SCG estimators, both in the presence of a single treated unit as well as in the presence of multiple treated units. One of the most commonly used approach use falsification tests, also named “placebo studies,” (Abadie et al., 2010, 2015; Ando and Sävje, 2013; Cavallo et al., 2013; Acemoglu et al., 2016; Firpo and Possebom, 2018), but alternative approaches have been recently developed, which include the construction of conditional prediction intervals (Cattaneo et al., 2021), and conformal inference (Ben-Michael et al., 2021).

We opt for a bootstrap based inferential method, which does require neither random assignment of the unit nor random selection of the treatment period, and does not rely on assumptions on the distribution of placebo treatment effects, such as, normality. The use of bootstrap within the SCG methods is not new (e.g., Sills et al., 2015; Xu, 2017). Abadie (2021) discusses the use of bootstrapping in SCG contexts, highlighting that bootstrapping is not always appropriate, since in several contexts we cannot consider the donor pool of control units as a random sample from a super-population, but we must consider it as the entire universe of observable units.

In our study, the donor pool we use to impute $Y_{1i,t}(0, \mathbf{0}_{N_1})$, $i \in N_1$, $t = T_0 + 1, \ldots, T$, consists of streets in urban neighborhoods that do not exhaust the urban neighborhoods of Florence, and thus, we can view it as a sample of urban neighborhoods of Florence. Consequently, the streets belonging to the sampled neighborhoods are a sample of the streets that make up the city. Specifically, we draw inference on the direct and average spillover effects, $\tau_{11,t}$ and $\delta_t^{\mathcal{N}_1}$, using a cluster bootstrap procedure (Davison and Hinkley, 1997), where we sample with repetition control urban neighborhoods: all streets in a sampled neighborhood are included in the bootstrap sample. Bootstrap confidence intervals for the direct and average spillover effects, $\tau_{11,t}$ and $\delta_t^{\mathcal{N}_1}$, are constructed using the bias corrected accelerated bootstrap method (BCa Efron, 1987), which allows for confidence intervals with good coverage properties, even if the distribution of the estimator is skewed. See the Appendix 1 for details on the construction of BCa confidence intervals.

1.3.4 Assessing unrealized spillover effects

Estimating the unrealized spillover effects, $\gamma_{11,t}^{(i)} = Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)}) - Y_{11,t}(0, \mathbf{0}_{N_1})$ $i \in \mathcal{N}_1 \setminus \{1\}$, is particularly challenging because both potential outcomes, $Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)})$ and $Y_{11,t}(0, \mathbf{0}_{N_1})$, are unobserved. Exploiting stable patterns across units' clusters, we can use information on control units outside the treated unit's cluster and their neighbors to construct an estimator for $Y_{11,t}(0, \mathbf{0}_{N_1})$ as described in Section 1.3.3. Unfortunately, the data contain no or little information on the potential outcomes of the form $Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)})$, because they are not observed for any unit in this study. Therefore, in order to construct an estimator for $\gamma_{11,t}^{(i)}$, we need to use an approach that extrapolates information on the potential outcomes $Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)})$ from the observed data.

We deal with this issue using information on units in the treated unit's cluster under a type of unconfoundedness assumption, which requires that for $t = T_0 + 1, \ldots, T$ and for each $i \in \mathcal{N}_1$, potential outcomes of the form $Y_{1i,t}(0, \mathbf{e}_{N_1}^{(j)})$ for $j \neq i \in \mathcal{N}_1$ are independent of W_i conditional on pre-treatment outcomes and covariates. Under this assumption, we propose to use an horizontal regression approach to inference Athey et al. (2021). Contrary to SCM-like estimators, that regress the time series of the treated unit during the pre-treatment period, in *horizontal* regression setting, we regress the cross-section of potential outcomes for the control units at time $t \geq t_0$ on the cross-section of the same units in the pre-treatment periods, in order to impute the missing potential outcomes as a linear combination of pre-treatment periods. Here, we exploit this approach and regress the panel of cross-sections of units receiving the spillover in the post-treatment on its own pre-treatment covariates, in order to estimate the relationship for this type of unit between post-treatment outcomes and pre-treatment outcomes and covariates. Subsequently, we predict on new data - coming from the treated unit - the unobserved potential outcome $Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)})$. Let $\mathbf{D}_{1i}^{*} = \left[Y_{1i,1}, \ldots, Y_{1i,T_0}, \mathbf{C}_{1i,1}, \ldots, \mathbf{C}_{1i,T_0}\right]'$ be a $(P+1) \times T_0-$dimensional vector of pre-treatment individual-level outcomes and covariates for $i \in \mathcal{N}_1$. Specifically, in our analysis we employ the outcome value in the two last periods of pre-treatment (2004 and 2005) and the pre-treatment outcome average. For each unit $i \in \mathcal{N}_1$, let $\bar{\boldsymbol{D}}_{1i}^{*(1)}, \ldots, \bar{\boldsymbol{D}}_{1i}^{*(K)}$, be K linear combinations of the pre-treatment outcomes and covariates. In order to account for a possible post-treatment trend in the outcome, we add a temporal term $(t - T_0)$, representing the distance between the observation period and the beginning of the treatment.

The missing outcomes for the treated unit, $Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1})$, $i \in \mathcal{N}_i \setminus \{1\}$, $t = T_0 + 1, \ldots, T$, are imputed as follows

$$\widehat{Y}_{11,t}(0, \mathbf{e}^{(i)}_{N_1}) = \widehat{\beta}_0 + \sum_{k=1}^{K} \widehat{\boldsymbol{\beta}}_k \bar{\boldsymbol{D}}^{*(k)}_{11} + \widehat{\gamma}(t - T_0),$$

where the regression coefficients are estimated using information on untreated units in the treated unit's cluster:

$$\left(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_K, \widehat{\delta}, \widehat{\gamma}\right) =$$
$$\arg \min_{\beta_0, \widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_K, \delta, \gamma} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \left[Y_{1i,t} - \left(\beta_0 + \sum_{k=1}^{K} \boldsymbol{\beta}_k \bar{\boldsymbol{D}}^{*(k)}_{1i} + \gamma(t - T_0) \right) \right]^2$$

Then, for $t = T_0 + 1, \ldots, T$, the average indirect effects are estimated as

$$\widehat{\gamma}_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \widehat{\gamma}^{(i)}_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \widehat{Y}_{11,t}(0, \mathbf{e}^{(i)}_{N_1}) - \widehat{Y}_{11,t}(0, \mathbf{0}_{|\mathcal{N}_1|}). \tag{1.9}$$

Variance of $\widehat{\gamma}_{11,t}$ is estimated by using the bootstrap variance of $\widehat{Y}_{11,t}(0, \mathbf{0}_{|\mathcal{N}_1|})$ and the robust estimate of the model-based variance of $\widehat{Y}_{11,t}(0, \mathbf{e}^{(i)}_{N_1})$, as:

$$\mathbb{V}(\widehat{\gamma}_{11,t}) = \mathbb{V}(\widehat{Y}_{11,t}(0, \mathbf{0}_{|\mathcal{N}_1|})) + \frac{1}{N_1^2} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \mathbb{V}(\widehat{Y}_{11,t}(0, \mathbf{e}^{(i)}_{N_1})).$$

1.4 Causal effects of a new light rail line on streets' retail density

In this section, we apply the method described in Section 1.3 to estimate the direct, the average spillover and the average unrealized spillover causal effects of a new light rail line on the retail sector density in a number of streets belonging to the same urban neighborhood in peripheral Florence (Italy). Talenti St., where the light rail is located, is subject to direct effects and unrealized spillovers. The nearby streets – namely Pollaiolo St., Pisana St., Baccio da Montelupo St., Scandicci St., and Magnolie St. – may only be subject to spillovers originating from Talenti St.

The streets' retail density is measured using two street-level outcome variables: number of stores selling durable and non-durable goods every 500 meters. We consider stores selling durable and non-durable goods separately, because we believe that effects can be heterogeneous for these two types of stores. Both the outcomes of interest were demeaned for the pre-treatment average outcome.

1.4.1 Penalized synthetic control estimators of direct and spillover effects

We impute the potential outcomes $Y_{1i,t}(0, \mathbf{0}_{N_1})$ for each $i \in \mathcal{N}_1$ and $t > T_0$ applying the penalized synthetic control method. For each street i within the urban neighborhood of Legnaia, $i \in \mathcal{N}_1$, we construct a synthetic street as weighted average of other streets belonging to Florentine urban neighborhoods located sufficiently faraway from Legnaia. From the imputed missing potential outcomes we then estimate the direct, the average spillover and the unrealized spillover causal effects of interest.

In order to estimate the penalized synthetic control weights following the procedure described in Section 1.3.3, we primarily have to select an appropriate value for λ.

In this work we use the leave-one-out cross-validation procedure proposed by Abadie and L'Hour (2021). First, for each post-intervention period $t = T_0 + 1, \ldots, T$, and for each $k = 2, \ldots, K + 1$, we use information on control units belonging to control clusters different from cluster k to derive penalized synthetic control estimators of the potential outcomes under control for units in cluster k under different values of λ. Specifically, for each $i \in \mathcal{N}_k$, $k = 2, \ldots, K + 1$, let $\widehat{Y}_{ki,t}(\lambda)$ denote the penalized synthetic control estimator of $Y_{ik,t}(0, \mathbf{0}_{N_k-1})$ with penalty term λ. For each $t = T_0 + 1, \ldots, T$, and $i \in \mathcal{N}_k$, $k = 2, \ldots, K + 1$, we then calculate

$$Y_{ik,t} - \widehat{Y}_{ki,t}(\lambda) = Y_{ik,t} - \widehat{Y}_{ki,t}(\lambda) = \sum_{k' \neq 1,k} \sum_{i' \in \mathcal{N}_{k'}} w^{(i)}_{k'i'}(\lambda) Y_{k'i',t}.$$

We choose λ to minimize the root mean squared prediction error (RMSPE) for the individual outcomes:

$$\sqrt{\frac{1}{(T - T_0) \sum_{k=2}^{1+K} N_k} \sum_{k=2}^{1+K} \sum_{i \in \mathcal{N}_k} \sum_{t=T_0+1}^{T} \left[Y_{ik,t} - \widehat{Y}_{ki,t}(\lambda) \right]^2}.$$

In order to ensure the uniqueness and sparsity of solution of the optimization problems in Equation (2.8), we focus on values of $\lambda \in (0, 1]$, testing a total of 1000 values. Selected values for λ are reported in Table 5.2 of the Appendix 1.

Once we have selected the penalization term, we move to the calculation of the weights. We estimate weights with the procedure described in 1.3.3, using the covariates and the pre-treatment outcomes scaled with respect to the pre-treatment mean. The estimated weights are reported in Table 5.3

in Appendix 1.

Given a value for λ and the estimated weights, $\boldsymbol{\omega}^{(i)}$, $i \in \mathcal{N}_1$, we estimate direct effects, $\tau_{11,t}$, and average spillover effects, $t = T_0 + 1, \ldots, T = 2006, \ldots, 2014$, using Equations (1.7) and (1.8). The RMSPEs, calculated over the individual- and cluster-level pre-intervention outcomes for each street in Legnaia, $i \in \mathcal{N}_1$, and its synthetic control, respectively, are reported in Table 5.4 of the Appendix 1. We derive 90% confidence intervals for these estimands using the biased corrected accelerated bootstrap procedure described with $B = 1000$ bootstrap replications. It is worth noting that in each bootstrap replication the estimates of the causal effects are derived using the penalized synthetic control method with the penalty term λ derived on the observed data.

1.4.2 Horizontal regression estimators of unrealized indirect effects

Potential outcomes of the form $Y_{11,t}(0, \boldsymbol{e}^{(i)}_{N_1})$ for Talenti St. are imputed using the regression approach described in Section 1.3.4 with $K = 2$ and $\bar{\boldsymbol{D}}^{*(1)}_{1i} = Y_{1i,T_0}$ and $\bar{\boldsymbol{D}}^{*(K=2)}_{1i} = \sum_{s=1}^{T_0} C_{1i,s}/T_0$, where we use as covariate $C_{1i,t}$ the number of purveyors selling non-durable (durable) goods for the outcome variable number of purveyors selling durable (non-durable) goods. We estimate robust model-based standard errors for $\widehat{Y}_{11,t}(0, \boldsymbol{e}^{(i)}_{N_1})$, by using the small sample modification introduced by Imbens and Kolesar (2016), which allow us to account for the small number of cross-section in the estimation.

1.4.3 Results

Estimated direct and average spillover effects

Figure 1.1 shows the estimated direct effect of the new light rail on Talenti St. During the construction phase of the tramway, there is an increase in both the density of stores selling durable and non-durable goods, and the effects are statistically significant. During the operational phase of the tramway, however, the gain of durable goods purveyors fades away, while the effect of the light rail remains positive, and of considerable magnitude, on the density of non-durable goods purveyors. A possible interpretation of these results is that the construction of the tramway initially beckons all types of retailers, who envision that the site will soon offer new commercial opportunities. However, increased demand should translate into higher prices for the available commercial space. Therefore, over a longer time horizon, purveyors of durables, which are goods with a lower frequency of purchase and higher customers' willingness to bear accessibility costs, have

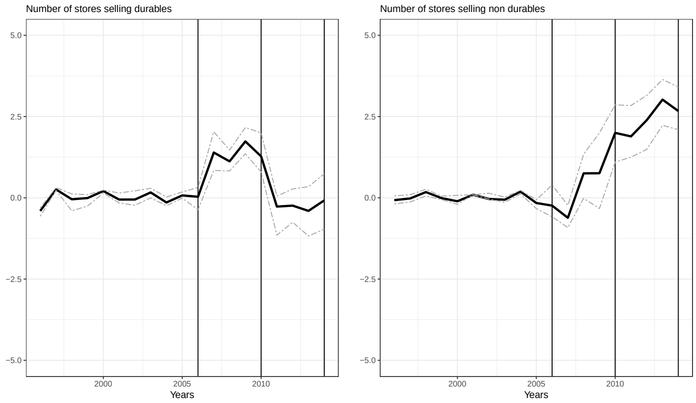


Figure 1.1: Estimated direct effects on Talenti St. (solid) and 90% confidence interval (dashed)

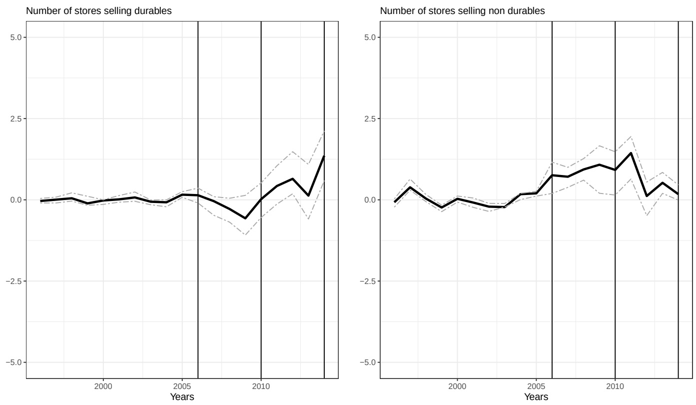


Figure 1.2: Estimated average spillover effects on neighboring streets (solid) and 90% confidence interval (dashed)

less incentive to pay the price required to stay next to the running tramway, because their customer base is not really made up of the occasional crowds of passers-by at stations. In contrast, purveyors of non-durable goods, which have high frequency of purchase in one's vicinity, e.g. cafes, grocery stores, florists, newsagents, depend more on these crowds of passers-by and, therefore, they are willing to pay the higher price required to stay on the site. These results are quite in line with the previous empirical literature, which highlights signs of commercial revitalization close to transit stations located in urban areas (Credit, 2018; Schuetz, 2015).

The average spillover effects on the other streets in the urban neighborhood of Legnaia are shown in Figure 1.2. As long as Talenti St. is undergoing construction works, we estimate slightly negative effects on the density of durable goods retailers in the neighboring streets. Although these effects are not statistically significant, they confirm the idea that the construction of the tramway might have initially raised expectations about Talenti St. to the detriment of other commercial locations nearby. Then, after the light rail goes into service in 2010, the effect on the density of durable-goods purveyors in these alternative locations turns positive but small, as it is less than one store each 500 meters, and statistically negligible for most of the years. Probably, for purveyors that depend little on occasional passers-by, shop windows on these streets are more worth their price than the coveted shop windows on Talenti St. Instead, with respect to stores selling non-durables, we have positive and statistically significant effects on neighboring streets while the light rail is under construction in Talenti St., but such effect tends to fade and lose statistical significance afterwards. A likely interpretation of this result is that, during construction, these alternative streets are expected to offer the opportunity to "steal" some of the customers that used to patronize stores selling non-durables on Talenti St., assuming that these customers would have been willing to flee the construction site to do their daily shopping within walking reach, or obliged to do so due to traffic detours. It is only a short-lived advantage, as Talenti St. later becomes the most lucrative place for non-durable goods purveyors due to the crowds coming and going all day at light rail stations.

In summary, the most noticeable quantitative effects occur in the street where light rail stations are located, as also found by the previous literature, but in the streets close by there is no overt displacement. Rather, our results suggest that the tramway triggered divergent processes of commercial specialization: it strongly encourage the use of commercial spaces near stations by purveyors of non-durables, while it slightly increase the focus of other streets on the retail of durable goods. Highlighting these divergent specialization processes represents, in our view, an original contribution we make to the subject literature.

Estimated unrealized spillover effect

Figure 1.3 shows the unrealized spillover effects on Talenti St., that is, the cost avoided or the benefit forgone by Talenti St. if the tramway had been constructed in some other street belonging to its same urban neighborhood. Although the estimates are surrounded by considerable uncertainty, they suggest that Talenti St. might have have suffered from a minimal negative effect on the density of durable goods retailers during the tramway construction phase, counterbalanced later by an equally minimal positive effect on the same outcome. On the whole, having a tramway somewhere else in the neighbourhood would not have affected the stock of durable goods shops in Talenti St. On the other hand, it is slightly more likely that it would have temporarily affected the stock of non-durable goods purveyors during the construction period, in line with what we estimated to have

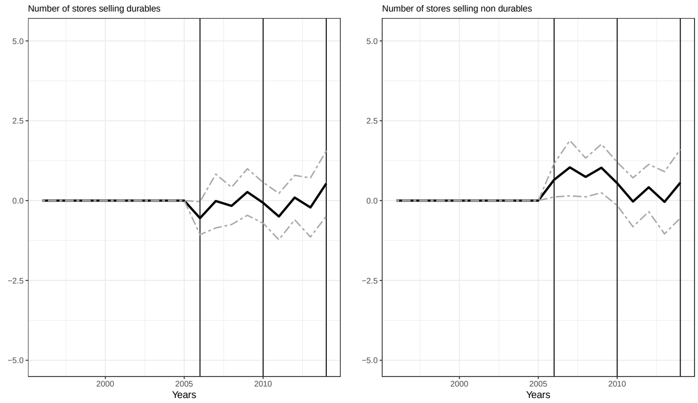


Figure 1.3: Estimated unrealized spillover effect on Talenti St. (solid) and 90% confidence interval (dashed)

happened in the streets that are actually susceptible to spillovers (see 1.2 for comparison). However, the confidence intervals here are quite wide, making it difficult to draw firm causal conclusions.

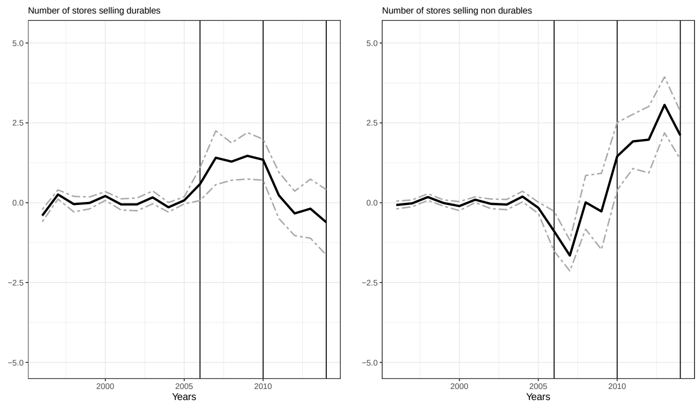


Figure 1.4: Difference between the estimated direct effect and unrealized spillover effect for Talenti St.(solid) and 90% confidence interval (dashed)

Figure 1.4 reports the difference between the direct effect and the unrealized spillover effect on Talenti St., which quantifies – given the choice of locating a light rail in the urban neighborhood of Legnaia – the "net" advantage/disadvantage connected to a situation of immediate proximity to tracks and stations, relative to a situation where the light rail is slightly more distant. From Figure 1.4 we gather that Talenti St. has eventually gained more purveyors of non-durables from being the site of a running tramway instead of being a street only near a running tramway.

1.5 Concluding Remarks

The SCG method has been hailed as "...the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens, 2017) and the ideas initially put forward in Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015) have sparked avenues of methodological research. This chapter has met the challenge of extending the SCG method to settings where the assumption of interference is untenable. This is a nascent stream of research in the SCG literature, which our study contributes to inaugurate, with relevant implications for applied economic and social research.

In this chapter, building on recent methodological works on causal inference with interference in the potential outcomes framework, we have first formally defined unit-level direct effects and average spillover causal effects under a partial interference assumption. We have also introduced a new spillover effect, the "unrealized spillover", which is the spillover that would have taken place on the actually treated unit if another unit had been assigned to the intervention. We believe that these three quantities may be relevant for a comprehensive evaluation of interventions at the meso- and macro-economic level. Then, we have proposed to use the penalized SCG estimator (Abadie and L'Hour, 2021)to estimate direct and average spillover causal effects, capitalizing on the presence of clusters of units where no unit is exposed to the treatment. We have used an horizontal regression approach to estimate unrealized indirect effects.

Our study has been motivated by the evaluation of the direct and unrealized indirect effects of a new light rail line built in Florence, Italy, on the retail environment of the street where it was built, and the spillover effects of the light rail on a number of streets close by. Although we focus on the Florence case study, similar interventions are often planned in other cities, too. Evaluating their direct, indirect and spillover effects may provide precious insight to policy makers, helping them to understand what transformations in the urban landscape are being brought about by creating new transit infrastructure. Our approach is very original also with respect to the field literature, where causal studies are still scarce and scholars usually conduct their analyses by aggregating all streets within a given radius (usually half mile) from the new infrastructure. From such picture, we learn that the light rail has encouraged the emergence of divergent patterns of commercial specialization between the street hosting the stations with the crowds of passers-by, and the streets a little further away from the new light rail.

Our results rely on the the assumption of partial interference, which is plausible in our application study, as it is in many other causal studies (e.g., Papadogeorgou et al., 2019; Huber and Steinmayr, 2021; Forastiere et al., 2021b). Nevertheless, we are aware that some studies might require a more general structure of interference (e.g., Forastiere et al., 2021b,a). Therefore a valuable topic for future research is the extensions of SCM methods to causal studies with general forms of interference.

Chapter 2

Impact heterogeneity of Covid-19 vaccination lotteries in the US

2.1 Introduction

The Covid-19 pandemic has challenged both scholars and policymakers from various points of view. In the first emergency, health management focused on containing the pandemic through non-pharmaceutical interventions (NPI from here on). These interventions were effective and decisive in preventing the contagions, but unsustainable in the long period. In parallel with health emergency management, research has focused on developing vaccines and treatments against Covid-19.

In particular, with the safety and efficacy results of the first vaccines, the organizational plan for the vaccination rollout has begun. It soon became clear that the outcome of the vaccination campaign depended not only on the stocks that each country could secure but also on the attitude of the population towards vaccination and the policies inserted into place to facilitate the campaign.

In various countries, part of the population was eager to get their vaccine shot, to avoid the most dangerous outcomes of Covid-19 and slow down the spread of the virus. Some other people were concerned about the fast development of many effective vaccines and refuse optional or compulsory vaccinations, stating that vaccines are not helpful but dangerous for children and adults. Based on fake news or wrong interpretations of scientific results, these arguments have a particular catch, especially in some echo chambers with similar political orientations and socio-demographic conditions. This disinformation could harm the effectiveness of the vaccination rollout.

With the spread of more transmissible and pathogenic variants than the original strain of Covid-19 (Alpha, Delta, and most recently Omicron), the time factor has become even more critical in limiting the spread of the disease and lowering the most severe outcomes. The focus was on avoiding severe consequences for the susceptible population, the elderly, residents of nursing homes, and essential workers, such as healthcare personnel. In addition, with the broader availability of vaccines, compliance with the vaccination campaign has become a relevant theme of public health policy.

Many governments have promoted several initiatives to entice hesitant to receive the vaccination, including monetary incentives for vaccination, and limitations to public life. For instance, in several European states certification of vaccination was required to travel by plane or train, to go to a restaurant or gym, or even to work. In 2021, Austria, Greece, and Italy governments introduced vaccination requirements, limited to the most at-risk population groups. Other administrations use monetary incentives to foster vaccination: fixed-sum incentives (e.g.:New York City and Pennsylvania) or monetary lotteries.

This chapter focuses on evaluating policies implemented by nineteen US states, which have promoted monetary incentives for vaccination, in the form of lotteries for those vaccinated against Covid-19. The first state to announce this type of policy was Ohio on May 12, 2021, launching the "Vax-a-million" initiative to combat low vaccination levels in the state. Ohio's policy attracted the attention of policymakers in other states, who followed in the subsequent weeks the Ohio example, giving away monetary prizes to vaccinated. On July 21, 2021, in total eighteen states followed Ohio's example. All except one announce the policy by July 1, 2021. Even if, in principle, policymakers design monetary incentives to help the vaccination rollout, in this specific case, the results are not precise a priori. We could expect a positive effect from the monetary incentives, coherent with the literature (Campos-Mercade et al. (2021)). Nevertheless, scepticism towards the safety of fast-developing vaccines, and efficacy doubts, can be enhanced by this kind of public intervention, harming the trust in government (Latkin et al. (2021), Lazarus et al. (2021)). Hesitant citizens may value avoiding the perceived risk connected to the vaccine more than the probability of winning a lottery prize, e.g.: Sprengholz et al. (2021).

It is crucial to analyze the outcome of such policies and what drivers are more tightly related to their impacts. Facing this context, assessing the causal impact of *nudging* toward vaccines is not trivial and it might be heterogeneous with respect to a variety of socio-economic and behavioural factors (Dubé et al. (2015), Savoia et al. (2021), Quinn et al. (2016), Reiter et al. (2020)).

Several papers have investigated the role of incentives in Covid-19 vaccination. Some of them used data from US states (Walkey et al. (2021), Barber and West (2021), AB (2021)), but none of them, to the best of our knowledge, have investigated the county level, addressing in-states differences in vaccination rollout. While the effect of Ohio's program has been studied, there are little to no comparative analyses between US states. There is theoretical ground to suspect that different characteristics correspond to different treatment outcomes.

We contribute to the literature by assessing the impact of conditional cash lotteries in a disaggregated framework in staggered adoption of the policy. We also study the duration of the effect, to exploit whether the treatment impact was temporary or persistent. We also investigate the heterogeneity of treatment effects across the counties. Our goal is to identify the socio-demographic characteristics of counties that performed better or worse. This work contributes to the methodological literature on the synthetic control method by providing estimates of weighted aggregate effects and proposing an inferential procedure for such effects.

We develop the chapter as follows: the relevant literature and the context we wish to evaluate are presented in sections 2.2. We describe data collection in section 2.3, and the causal inference approach in section 2.4. Results are shown and discussed in section 2.5 and section 2.6 concludes.

2.2 Related Literature

Vaccine hesitancy is a known issue in vaccination rollouts, even before the Covid-19 pandemic, as it was observed in vaccine rollout against measles, HPV, and seasonal influenza, see for a review Dubé et al. (2013).

Several health policy interventions over the years have been implemented in order to tackle the concerning trend of reduction of vaccination uptake among children, and in particular, to address directly parental vaccine hesitancy (Gowda and Dempsey (2013), Williams (2014)).

In most previous studies, scholars have posed attention to those socioeconomic drivers that can explain the variety in vaccination uptakes; see Jarrett et al. (2015) for a comprehensive review.

In particular, Robertson et al. (2021), Razai et al. (2021), Willis et al. (2021), Quinn et al. (2016), Reiter et al. (2020), focuses on the relationship between ethnicity and vaccination uptakes, Badr et al. (2021) and Azizi et al. (2017) shed light on the relation between poverty and unemployment and vaccines, interestingly, before and during the Covid-19 pandemic. Bertoncello et al. (2020) find an inverse correlation between the parental level of education and vaccine hesitancy and anti-vaccine sentiment, irrespective of whether children are involved. Malik et al. (2020), Marks (2020) and Joshi et al. (2021) investigated the socio-demographic composition of individuals willing to comply with the US Covid-19 vaccination campaign and find out significant differences in ethnicity, gender, and age groups. Dubé et al. (2014) highlighted another crucial aspect: vaccine hesitancy is not a fully generalized concept but has several different drivers across different countries.

The information source plays a role in determining the attitude towards vaccination: e.g.:Featherstone et al. (2019), Engin and Vezzoni (2020) and Mønsted and Lehmann (2022) find out that vaccine conspiracy belief spreads out on social media, especially among those who express conservative political thought. We can find similar results in Covid-19 vaccine rollout analyses about the US and UK, see for example Loomba et al. (2021).

The effects of these different drivers were heterogeneous in the US. In some counties, the Center for Disease Control (CDC) considers the vaccination rollout concerning, especially in the Sunbelt and in the Great Plains, with possible negative effects also on the Covid-19 cases count and on the related deaths.

Interesting literature flourished among those incentives for vaccination and health policy interventions that should direct the general population towards health-policy goals, such as reducing smoking, obesity, and alcohol drinking. Gorin and Schmidt (2015) studies the relationship between the outcome of the policy and the public discussion generated. Persad and Emanuel (2021), Korn et al. (2020), Weisel (2021), and Dotlic et al. (2021) argue about the legitimacy of such kind of monetary interventions in the case of Covid-19, considering the effects in terms of social responsibility. Campos-Mercade et al. (2021), Kim (2021), Jecker (2021) and Taber et al. (2021) discuss results coming from monetary incentives for Covid-19 vaccination, both in form of lotteries and fixed-sum transfers. We note that there is no agreement in the literature either on the legitimacy of monetary incentives for vaccination or on the actual results of such policies, since such incentives may not affect vaccination choices. Governments often used monetary incentives before restricting activities in the absence of a vaccination certificate.

2.3 Data

We collect information on 2925 counties located in 47 US states. We exclude Alaska, Hawaii, and Puerto Rico from the analysis because of their unique characteristics regarding the continental US. We exclude Texas because the primary outcome was not collected at the county level. We also exclude Missouri from the analysis, because of its late lottery announcement, on July 21st,2021. Apart from these five states, we drop counties lacking observations. In particular, we excluded counties that do not report vaccinations at the end of the period, on August 24, 2021, and those with no consistent results for the cumulative number of vaccinations (e.g., decreasing cumulative vaccinations in some time periods).

The primary outcome of our analysis is the share of over 18 citizens vaccinated against Covid-19. The Center for Disease Control (CDC) provides this information on this variable. We chose this measure because it is the most responsive to incentives and it is measured when citizens adhere to the campaign. The percentage of the population fully vaccinated shows a delay between the adherence to the policy and the measurement of the outcome because of time elapsing between the first dose (adherence to the campaign) and the second dose. We focus on the over 18 population because at the beginning of the treatment, the authorization for vaccination for the population over 12 was relatively recent (May 10) and hesitation behaviours were possible, besides the vaccine hesitation measured by the CDC. The primary outcome is the only observed measure over time, from January 1st, 2021, to August 24th, 2021. We focus on this time period because after August 24th the Pfizer-Biontech vaccine received full approval from the Food & Drug Administration (FDA hereinafter). On the basis of this approval, the US government announced compulsory vaccination for military troops. We summarize daily data into weekly data calculating the 7-day moving average of the share of people vaccinated with the first dose. We prefer to work with this outcome rather than daily data because daily calendar effects or temporary delays in vaccination reports could harm the analysis.

Table 2.1 shows some descriptive statistics of the main outcome in three relevant time periods:

- May 12: Announcement of the first lottery in Ohio.
- Jul. 01: Announcement of the last lottery in Michigan.
- Aug. 24: Compulsory vaccination for military staff, end of the observation period

Table 6.2 in Appendix 2 shows the share of first-dose receivers for each state at these three time periods.

Figure 2.1 represents the time series of the share of people vaccinated with the first dose.

We enrich the dataset with information on various socio-demographic, political, economic, and environmental characteristics. In particular, our dataset includes information on the demographic composition, political orientation, and economic indicators (e.g.: percentage of unemployed and median income by county). The US Census Bureau collects and provides socio-demographic data for 2020. The CDC provides data on the percentage of people insured with Medicare. Economic indicators are derived from

	May 12th	July 1st	August 24th
Mean	40.220	45.677	52.282
St.Dev.	13.439	15.985	16.680
5%	19.681	21.055	22.101
50%	41.674	45.228	53.627
95%	57.785	67.690	73.630

Table 2.1: Descriptive statistics for the percentage of first-dose receivers, at county level

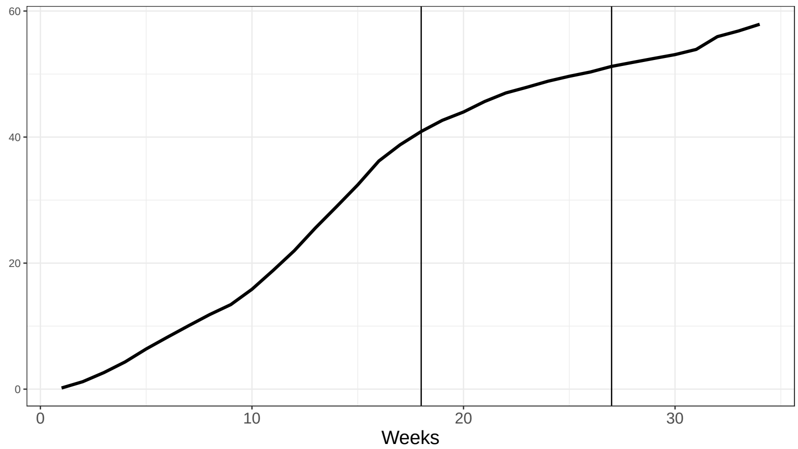


Figure 2.1: % of first-dose receivers on the over 18 population - First vertical line: announcement of first lottery (Ohio), Second vertical line: announcement of last lottery (Michigan)

the work of Kirkegaard (2016), updated to 2020. The New York Times provide data regarding voting in the 2020 presidential election.

We also collect from CDC the total number of Covid-19 related deaths. This dimension could be a relevant effect modifier because the proportion of vaccinated might be higher in counties where happened relevant outbreaks. Table 2.2 shows some descriptive statistics for the variables used in the study.

2.4 Methodology

2.4.1 Notation and Setting

We consider the lottery policy as a causal inference problem in which some counties receive the active treatment (the lottery) starting from some period t_0, and other counties did not receive any monetary stimulus to take part in the vaccination rollout.

Variable	Mean	St.Dev.	5%	50%	95%
Percentage of Hispanic citizens	9.567	13.956	0.960	4.310	40.340
Percentage of Afro-American citizens	8.728	14.006	0.090	2.220	41.448
Percentage of citizens in poverty	14.295	5.617	6.900	13.300	25.060
Percentage of Republican party voters	65.205	15.605	35.288	68.276	85.544
Percentage of high school graduated	34.120	7.201	21.343	34.524	45.397
Percentage of college graduated	22.021	9.426	11.239	19.650	41.261
Unemployment rate	6.704	2.169	3.500	6.500	10.400
Share of treated counties	38.348	48.632	0.000	0.000	100
Covid-19 related deaths/100k citizens	176.597	89.467	55.461	165.292	336.156
Percentage of citizens insured with Medicare	11.890	4.559	5.281	11.392	19.891
Median Age	39.912	4.794	31.700	39.900	47.800

Table 2.2: Descriptive statistics of socio-demographic characteristics of US counties

We consider a panel data setting, in which the total set $\mathbf{\Omega}$ of observed units consists in $|\mathbf{\Omega}| = 2925$ US counties, observed for $t \in T = \{0, \dots, t_0, \dots, t_T\}$, $|T| = 34$ from the January 1st, 2021 to August, 24th, 2021. Let $\mathbf{\Omega}^1$ be the entire set of units in which the lotteries were active at some time t. Let $\mathbf{\Omega}^0$ be the set of units that experiences no form of monetary incentives at any time t.

In total, $|\mathbf{\Omega}^1| = 1134$ counties have experienced a vaccine lottery in the period considered, and $|\mathbf{\Omega}^0| = 1791$ counties have received no kind of monetary incentives for vaccination. Note that $\mathbf{\Omega}^0 \cup \mathbf{\Omega}^1 = \mathbf{\Omega}$. Let $\mathbf{N}^1 = \{1, 2, \dots, n^1\} \subseteq \mathbf{\Omega}^1$ denote a generic subset of treated counties, and similarly let $\mathbf{N}^0 = \{n^1+1, n^1+2, \dots, n^1+n^0\} \subseteq \mathbf{\Omega}^0$ denote a generic subset of control counties. Finally, we denote as $Y_{i,t}$ the primary observed outcome for a generic unit i by time t, which is the percentage of residents who received the first dose of vaccine.

Treatment Uptake

Treatment was allocated at the state level, therefore counties belonging to the same state experienced the same treatment, in terms of duration and prize amount.

We denote the treatment indicator for unit i at time t by $D_{i,t}$

$$D_{i,t} = \begin{cases} 1 & \text{if county } i \text{ is receiving the treatment at time } t \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Note that if $D_{i,t} = 1$, then $D_{i,t'} = 1$, if $t' > t, t \in T$.

Following this specification, we can construct a treatment matrix D:

$$
\mathbf{D} = \begin{bmatrix}
t & 1 & 2 & \dots & n^1 & n^1+1 & n^1+2 & \dots & n^1+n^0 \\
0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
t_0^1 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\
t_0^i & 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
T & 1 & 1 & 1 & 1 & 0 & 0 & \dots & 0
\end{bmatrix}
$$

In the literature, this treatment framework is called *staggered adoption* (Athey and Imbens (2021), Ben-Michael et al. (2022), Callaway and Sant'Anna (2021)). In this scenario, the focus is on evaluating treatment effects started at different times. In particular, we suppose no unit receives the treatment before t_0^1. So, t_0^1 is the time in which the first unit receives the treatment. In our case, t_0^1 is May 12, in which the governor of Ohio announced the "Vax-a-Million" initiative. Treated counties $i \in n^1$ can receive treatment at any time $t_0^i \geq t_0^1$. In particular, we define t_0^i as the period in which unit i receives the treatment. We found it useful to establish a notation for a unit that will or will not receive treatment at any time t: $\mathbf{D}_i = 1$ if $D_{i,t} \neq 0$ for some t, and $\mathbf{D}_i = 0$ if $D_{i,t} = 0 \quad \forall t$.

Nineteen states have announced a vaccine lottery to improve the vaccine rollout (Ohio, Oregon, Washington, California, Nevada, New Mexico, Louisiana, North Carolina, West Virginia, Maine, Kentucky, Michigan, New York, Illinois, Missouri, Arkansas, Colorado, Delaware, Maryland). We report the beginning time of the lotteries and their duration in Figures 2.2 and 2.3, see Table 6.1 in the Appendix 2 for further details.

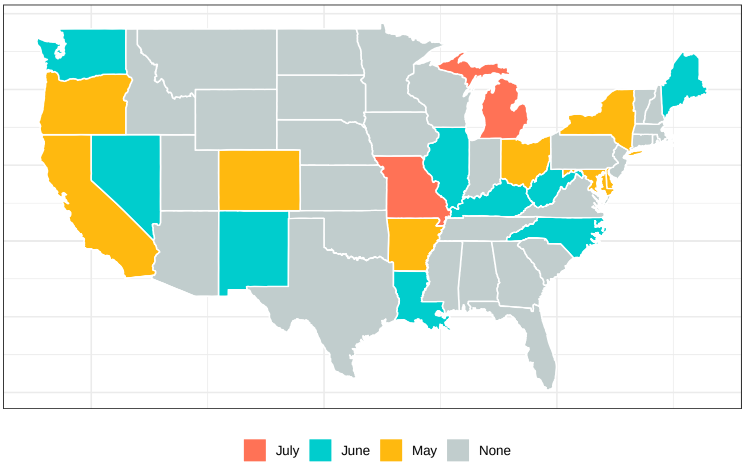


Figure 2.2: Month in which states announced the lottery program

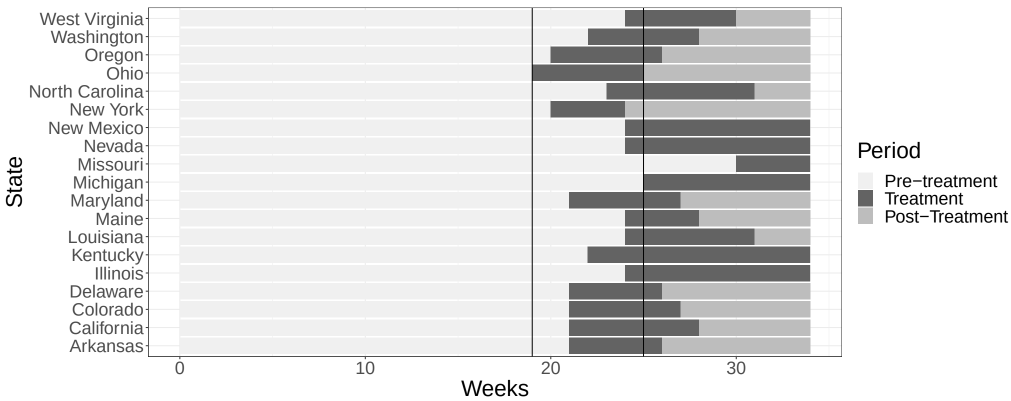


Figure 2.3: Pre-treatment, treatment and post-treatment duration for each state, starting from January 1st. First vertical line: Ohio lottery announcement, Second vertical line: Michigan lottery announcement

Potential Outcomes

We use a potential outcome approach to causal inference (Rubin (1974), Rubin (1978)). For each unit i in each period t we define the following couple of potential outcomes:

$$\begin{cases} Y_{i,t}(1) \equiv Y_{i,t}(D_{i,t} = 1) & \text{as the potential outcome under treatment} \\ Y_{i,t}(0) \equiv Y_{i,t}(D_{i,t} = 0) & \text{as the potential outcome under control} \end{cases} \tag{2.2}$$

This definition implicitly assumes the two following assumptions: the Stable Unit Treatment Value Assumption (SUTVA, Rubin (1980)) and the no-anticipating treatment assumption.

Assumption 3. *SUTVA, Rubin (1980)*

- *No hidden version of the treatment*
- *No interference between units*

Both aspects of SUTVA deserve some more commentary. The no-hidden version of the treatment component of SUTVA may be arguable in situations in which we consider treated counties belonging to different states. In such situations, both the award amount and the probability of winning may differ. We assume these differences are not relevant for determining whether the treatment causes an effect on our outcome of interest. This assumption appears to be credible because it is unlikely that the population applies a quantitative assessment of the cost-benefit ratio associated with vaccination in monetary terms. Consequently, all lotteries are similar for the receiving population. Taber et al. (2021) investigates this aspect, finding no differences in participation choices across twelve different lottery structures, supporting our assumption.

The non-interference assumption states that the potential outcome for any unit does not vary with the treatments assigned to other units. We believe that the non-interference assumption may be valid in our context. For a practical example, the lottery treatment of a county in Ohio should not change the vaccination campaign adherence of an untreated county, such as a county in Wisconsin.

The non-anticipating treatment assumptions require that if a county has not adopted yet the policy, the future adoption has no causal effect on potential outcomes for the current period. In theory, people could have changed their behaviour after the lottery announcement, delaying the vaccine administration to get the lottery ticket. In our study, this behaviour is not plausible because states allow taking part in the lottery even though they had already received the first dose. In addition, the short time between the announcement of the lottery and the start of the program does not allow for noticeable treatment anticipation phenomena. Finally, this policy spreads faster among the US, with most treated states announcing the program within 45 days after the original Ohio governor's announcement. Therefore, we can not expect people delayed joining the vaccination campaign to get a lottery ticket.

Under SUTVA (assumption 3) and no-anticipating treatment, we have $Y_{i,t}(0)$ for $t \in \{0, t_0^1\}$ and $Y_{i,t}(D_{i,t})$ for $t \in \{t_0^1, t_T\}$.

2.4.2 Causal Estimands

We introduce the following individual causal estimand.

$$\Delta_{i,t} = Y_{i,t}(1) - Y_{i,t}(0) \tag{2.3}$$

for each county $i \in \mathbf{N}^1$ receiving the treatment by time t. For treated units $Y_{i,t}(1) = Y_{i,t}$ and thus we compare the observed outcome $Y_{i,t}(1)$ for the treated unit, with his counterfactual outcome $Y_{i,t}(0)$. We think that there are no logistic delays in providing the vaccine shot once the policy was implemented, due to the wide availability of doses and logistic infrastructure at the beginning of the policy.

We are also interested in causal effects for specific states $\mathbf{S}$. We define it as a weighted average of the county effects $\Delta_{i,t}$, multiplied by an appropriate weight η_i. In our case, we chose η_i as the ratio of the population of the i-th county to the total population of the state $\mathbf{S}$. Thus we define the treated set as $\mathbf{N}^1 \equiv \mathbf{S}$.

Formally:

$$\Delta_{\mathbf{S},t} = \sum_{i \in \mathbf{S}} \eta_i \Delta_{i,t} \qquad \eta_i \in (0,1) \tag{2.4}$$

We can define causal effects at the state level (e.g.: pooling counties from California), but also at the supra-state level, by pooling counties from different states (e.g., pooling counties from West Coast). Therefore, the average effect on the supra-state aggregation will result as:

$$\Delta_{\mathbf{N}^1,t} = \sum_{i \in \mathbf{N}^1} \eta_i \Delta_{i,t} \qquad \eta_i \in (0,1) \tag{2.5}$$

In this latter case, the treated set $\mathbf{N}^1$, corresponds to the union of counties belonging to the states that we want to pool together.

Estimation of equation 2.5 at the supra-state level, in presence of staggered adoption, could be a little problematic. Under consistency, we can assume that all the lotteries are considered equal by the population, this assumption is consistent with the results found by Taber et al. (2021). See table 6.1 in the appendix for further information about the lotteries. For

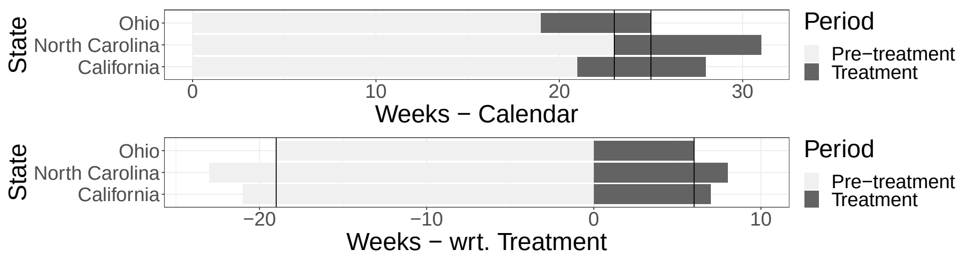


Figure 2.4: Pre-Treatment and treatment period according to calendar weeks criterion and with respect to the treatment, periods within the vertical lines are the overlapping period for this set of states

example, consider three US states: Ohio, California and North Carolina. We observe three different periods for the treatment regime and the post-treatment regime. Table 2.3 reports their duration, in weeks. Naturally, we estimate individual treatment effects in every calendar time t, starting from the first week of 2021. Therefore, week 19 will be the first treatment period for Ohio, week 21 will be the first for California, and week 23 will be the first treatment period for North Carolina. To pool together treatment effects, under staggered adoption, we should sum them up regarding the treatment assignment: to get the pooled effect for the first week after the treatment assignment we should calculate the estimand in equation 2.5 when $t = t_0^i \quad \forall i \in \mathbf{N}^1$. Consequently, to get the effect in the second treatment week we should sum treatment effects when $t = t_0^i + 1$ and so on. Note that $\Delta_{\mathbf{N}^1,t}$ is the weighted average of $\Delta_{i,t}$ for each treated unit $i \in \mathbf{N}^1$. Therefore, it is defined only for time spells $t_0^i - t \geq 0 \quad \forall i \in \mathbf{N}^1$. In our example, we cannot have more than six weeks of treatment, which is the minimum amount of treatment common to the three states.

Figure 2.4 shows the re-alignment of periods regarding calendar criterion and regarding the treatment assignment.

	Pre-treatment	**Treatment**	**Post-treatment**
Ohio	18	6	10
California	20	7	8
North Carolina	22	8	4

Table 2.3: Duration in weeks of pre-treatment, treatment and post-treatment periods, for three example states, starting from the first week of 2021

Last, we define the average effect for the treated set $\mathbf{N}^1$, over the period (t_1, t_2), as

$$\Delta_{\mathbf{N}^1} = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \Delta_{\mathbf{N}^1,t} \tag{2.6}$$

The choice of the period (t_1, t_2) allows us to distinguish between the phase in which the lottery was active and the subsequent phase by averaging $\Delta_{\mathbf{N}^1,t}$, during treatment, or post-treatment, periods. Note that equation 2.6 can be used to estimate both state or supra-state effects over time.

Analysis of the post-treatment helps us understand if the effect of the lottery is temporary or permanent. Indeed, it is likely that the lottery helps to convince latecomers to vaccinate. Thus, lotteries could speed up vaccination but not increase the number of vaccinated patients compared with controls at a more distant endpoint. Conversely, if we observed a permanent increase in the number of vaccinated, we could conclude that lotteries affect those who delayed vaccination and those who had no intention of vaccinating. Assessing the durability of the treatment effects over time may provide useful insights for policymakers.

2.4.3 Penalized SCM

This section explains how we impute missing potential outcomes.

With repeated observations over time, and many units both under treatment and under control, various tools are available to assess the effect of the policy.

We choose to estimate the causal quantity in equation 2.3 with a modification of the Synthetic Control Method, first introduced by Abadie et al. (2010), which is getting growing success in the causal inference community. We estimate causal effects by imputing missing outcomes, $Y_{i,t}(0)$, namely the outcomes that the treated unit i would have been if it had never received the treatment, constructing a weighted mean of control units in the donor pool. This weighted average is called synthetic control. The weights are chosen so that the synthetic control for a treated unit i is very close to the treated unit during the pre-treatment period. This method was extended to allow the estimation of average treatment effects, also in staggered adoption contexts (Dube and Zipperer (2015) Donohue et al. (2019), Ben-Michael et al. (2022)), where the focus is on estimating the treatment effects for each treated unit i and pooling them together.

Among the recent developments of the original estimator (see for example Abadie (2021), Ben-Michael et al. (2021), Doudchenko and Imbens (2016)), we choose to adopt the novel method developed by Abadie and L'Hour (2021), the so-called Penalized Synthetic Control Method (P-SCM).

The penalized control method imputes the missing control outcomes

$Y_{i,t}(0)$ for $i \in \mathbf{N}^1$ as equation 2.7, where the weights are derived solving the following minimization problem equation 2.8. For notation simplicity, let be $i = \{1, 2, \ldots, n^1\}$ the index for treated units and let be $j = \{n^1 + 1, n^1 + 2, \ldots, n^0 + n^1\}$ the index for control units We define the synthetic control for each $i \in \mathbf{N^1}$ as:

$$\widehat{Y}_{i,t}(0) = \sum_{j \in \mathbf{N}^0} \omega_j^{(i)} Y_{j,t} \qquad t \in (t_0^i, \ldots, T) \tag{2.7}$$

For notation simplicity, let be $t_0 = t_0^i$ and $t_T^i = t_T$ for each unit $i \in \mathbf{N}^1$.

Let $\mathbf{X}_i = \left[Y_{i,0}, \ldots, Y_{i,t_0}\right]'$ be a t_0-dimensional vector of pre-treatment outcomes, for each treated unit $i \in \mathbf{N}^1$, and let

$\mathbf{X}_j = (\left[Y_{n^1+1,0}, \ldots, Y_{n^1+1,t_0}\right]', \ldots, \left[Y_{j,0}, \ldots, Y_{j,t_0}\right]', \ldots, \left[Y_{n^1+n^0,0}, \ldots, Y_{n^1+n^0,t_0}\right]')$ be a $(t_0) \times |\mathbf{N}^0|-$dimensional matrix of pre-treatment outcomes, for the set of control units $j \in \mathbf{N}^0$

Given a positive penalization constant $\lambda^{(i)}$, $i \in n^1$, the set of weights

$$\boldsymbol{\omega}^{(i)} = \{\omega_j^{(i)}\}_{j \in \mathbf{N}^0}$$

defines the penalized synthetic control unit for treated unit i. The set of weights $\boldsymbol{\omega}^{(i)}$ is obtained by solving the following minimization problem.

$$\arg \min_{\boldsymbol{\omega}^{(i)} \in \boldsymbol{W}} \left\| \mathbf{X}_i - \sum_{j \in \mathbf{N}^0} \mathbf{X}_j \omega_j^{(i)} \right\|^2 + \lambda^{(i)} \sum_{j \in \mathbf{N}^0} \left\| \mathbf{X}_i - \mathbf{X}_j \right\|^2 \tag{2.8}$$

subject to

$$\omega_j^{(i)} \geq 0 \quad \forall j \in \mathbf{N}^0; \qquad \sum_{j \in \mathbf{N}^0} \omega_j^{(i)} = 1,$$

with $\|n\|$ is the L^2-norm: $\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}}$ for $\mathbf{v} \in \mathbb{R}^r$

We choose the P-SCM estimator because it is specifically designed for estimating average treatment effects with disaggregated treated units. Moreover, it grants us the uniqueness of the weights. Abadie and L'Hour (2021) and Abadie (2021) present the technical details of the estimator and its use. As discussed in Athey et al. (2021), in observational studies where both N and T are large, the choice of which method to use is not straightforward, and P-SCM may allow a data-driven solution to this problem. In fact, when $\lambda^{(i)} \to 0$, the P-SCM collapses into the standard SCM, while when $\lambda^{(i)} \to \infty$ the P-SCM is equivalent to the nearest-neighbour matching estimator. P-SCM enables us to maintain an agnostic behaviour towards the choice between matching methods and SCM.

We chose the tuning parameter $\lambda^{(i)}$ by using the weighted cross-validation approach, proposed by Abadie and L'Hour (2021). We define the function $\mathbf{\Phi}(\lambda)$, exposed in Equation 2.9, which minimizes the overall Root Mean Square Prediction Error (RMSPE) of the donor pool for a given of λ in the treatment period.

$$\Re(\lambda, t_T - t_0) = \sqrt{\frac{1}{(t_T - t_0)} \sum_{t=t_0}^{t_T} \left[Y_{j,t} - \widehat{Y}_{j,t}(\lambda) \right]^2}$$

The tuning parameter λ^* is chosen as follows:

$$\lambda^* = \arg\min_{\lambda \in \Lambda} \mathbf{\Phi}(\lambda) = \sum_{j=1}^{n^0} \Re(\lambda, t_T - t_0) \tag{2.9}$$

Table 6.4 of the Appendix 2 reports values of λ used in the estimation.

Donor pool definition

Abadie (2021) and Abadie et al. (2015) pointed out the necessity of a proper donor pool to estimate synthetic control weights. In particular, they suggest to subset the total control set $\mathbf{\Omega}^0$ in a set that has similar characteristics regarding the treated set. Usually, donor pool selection is driven by the researcher's experience of the characteristics of the treated units (e.g.:Abadie et al. (2015)).

In this work, we search for the proper donor pool for each treated set $\mathbf{N}^1$ by using a data-driven procedure. We choose to select the donor pool by using a matching approach based on propensity scores, first introduced by Rosenbaum and Rubin (1983). Let

$$\pi_i = \pi_i(\mathbf{C}) = Pr(\mathbf{D}_i = 1|\mathbf{C}) \tag{2.10}$$

denote the propensity score, where $\mathbf{C}$ is a set of pre-treatment covariates.

We estimate the propensity score using a logistic regression model on a set of covariates describing socio-economic, political, ethnic, and demographic characteristics (see Table 2.2 for the complete list). Once we have estimated $\pi = \{\pi_j\}_{j \in \mathbf{N}^0}$ for the control units, we select the donor pool by performing a one-to-many matching, with 4 control units for each treated unit. Thus following, $\mathbf{N}^0$ denotes the subset of *matched* control units. Table 6.3 in the Appendix 2 shows two-sample t-tests to evaluate the similarity between covariates of the treated pool and donor pool for each state. Such a procedure for donor pool restriction speeds up weight computations. Moreover, it allows to build the synthetic control of the treated unit from similar control units, without relevant losses in terms of prediction error, see also Abadie et al. (2010).

Inference

We conduct inference using falsification tests, also named "placebo studies," which can be viewed as a type of randomization inference.

A placebo test is constructed by reassigning the treatment to control units in the donor pool and estimating the so-called placebo effects. Then, the estimated treatment effect is compared with the distribution of placebo treatment effects. A treatment effect is considered statistically significant when the magnitude of the effect is large with respect to the distribution of the placebo effects. See Abadie et al. (2010) and Abadie et al. (2015) for details. Firpo and Possebom (2018) extend the original inference to allow for different weights for placebo units, but pointed out that in real applications it is difficult to find these weights.

Building on the aforementioned works and the work by Cavallo et al. (2013), we use an algorithm to conduct inference with SCM and aggregate units.

We focus on assessing the significance of the aggregate $\Delta_{\mathbf{N}^1}$ effects. Therefore, we do not estimate p-values or confidence intervals for individual effects for each i-th county. We neither estimate confidence intervals for each time-specific effect $\Delta_{\mathbf{N}^1,t}$. We define the test statistics as

$$\Theta = \frac{\mathfrak{R}(t_0, t_T)}{\mathfrak{R}(0, t_0)} \tag{2.11}$$

where $\mathfrak{R}(t_1, t_2)$ is the root-mean square prediction error of the estimated $\Delta_{\mathbf{N}^1}$ calculated between t_1 and t_2, as in equation 2.6.

Let $\Delta_{\mathbf{N}^0} = \{\Delta_1, \ldots, \Delta_j, \ldots, \Delta_J\}$ be the set of aggregate placebo effects, created by randomly sampling placebo effects from units in $\mathbf{N}^0$ and pooling them together to find each aggregate placebo effect Δ_j. In our analysis, we choose $J = 1000$. Let $\Theta_{\mathbf{N}^0} = \{\Theta_1, \ldots, \Theta_j, \ldots, \Theta_J\}$ be the set of placebo Θ_j calculated using equation 2.11 on the set of placebo effects $\Delta_{\mathbf{N}^0}$.

We can define the p-value as:

$$\rho_{\mathbf{N}^1} = \frac{1}{J} \sum_{j \in J} \mathbb{I}(\Theta_j > \Theta_{\mathbf{N}^1}) \tag{2.12}$$

In principle, if the effect $\Delta_{\mathbf{N}^1}$ is large relative to the distribution of placebo effects $\Delta_{\mathbf{N}^0}$, there should be very few $\Theta_J > \Theta_{\mathbf{N}^1}$, and we can consider the effect statistically significant. As stated by Abadie et al. (2010), using

$\boldsymbol{\Theta}$ as test statistics instead of Δ allows us to compare treatment units with control units, even in presence of imperfect pre-treatment fit. Algorithm 1 shows the procedure to calculate $\rho_{\mathbf{N}^1}$.

Algorithm 1 SCM inference with many treated units

procedure P-VALUES
Require: $\widehat{\Delta_j}(\mathbf{D}_j = 1)$, $\Theta_{\mathbf{N}^1}$
Ensure: $\rho_{\mathbf{N}^1}$

1. **for** $\mathcal{J} \in (1 : 1000)$ **do**
2. Sample with replacement n^1 units from the donor pool $\mathbf{N}^0$, forming the placebo state j
3. Calculate $\widehat{\Delta_j}$ according to equation 2.5 and 2.6
4. Calculate $\boldsymbol{\Theta}_j$ for each placebo state as in equation 2.11
5. **end for**
6. Calculate the p-value $\rho_{\mathbf{N}^1}$ for the treated state by using Equation

$$\rho_{\mathbf{N}^1} = \frac{1}{J} \sum_{j \in J} \mathbb{I}(\Theta_j > \Theta_{\mathbf{N}^1})$$

7. **end procedure**

2.4.4 Assessing treatment heterogeneity

After the first vaccine lottery announcement in Ohio, several states introduced their own lottery, creating policies mimicking. There is some consensus on the positive results of conditional cash lotteries in Ohio (Barber and West (2021), Acharya and Dhakal (2021)); but very early evaluations from other states (e.g.: Arkansas) head in different directions, leading the policymakers to interrupt the policy.

These results suggest that treatment effects could have been heterogeneous both across states and across counties. We study treatment effect heterogeneity in order to provide helpful insights for future decisions.

We classified treated counties according to some socio-demographic, economic, and cultural characteristics. Table 2.1 shows the dimensions used in this analysis. To classify treated counties, we follow a clustering approach with Gaussian mixture models (see Fraley and Raftery (2002), McLachlan et al. (2019)). We choose the number of clusters that minimize the selection model's BIC for membership in each group of counties. According to the set of covariates $\mathbf{C}$, we found six clusters of counties. Table 2.4 reports the mean values of the covariates in each cluster.

Cluster	1	2	3	4	5	6
Percentage of Hispanic citizens	1.918	9.060	37.379	4.999	3.203	3.794
Percentage of Afro-American citizens	1.004	13.585	2.120	0.869	30.437	5.192
Percentage of citizens in poverty	14.947	13.036	15.252	12.275	23.123	13.869
Percentage of republican party voters	75.117	50.652	53.374	61.721	61.597	66.427
Percentage of High school graduates	42.943	27.466	28.379	32.355	38.953	36.715
Percentage of college graduates	15.336	32.145	21.060	23.542	15.166	19.696
Unemployment rate	8.020	7.260	8.799	7.899	8.404	7.578
Covid-19 related deaths/100k citizens	186.523	136.933	157.225	152.466	262.130	186.280
Percentage of citizens insured with Medicare	11.369	9.001	10.074	14.034	11.675	10.750
Median yearly earnings	24612.784	29142.530	25035.187	25230.088	23198.967	25707.554
Median Age	40.726	36.809	36.742	43.274	38.634	39.975
Share of counties on treated counties set	0.204	0.205	0.126	0.095	0.186	0.184

Table 2.4: Mean value of the covariates in each cluster

As one can see in table 2.4, clusters 3 and 5 have a very different ethnic composition compared to the others. Specifically, in cluster 5, the share of the Afro-American population is the highest among cluster, and cluster 3 represents the one with the highest percentage of Hispanics. Cluster 2 has the richest, and the most educated population on average and the lowest proportion of people voting for the Republican Party. Clusters 1 and 6 are distinguished by their high percentage of votes for Republicans. Cluster 4 has the highest median age, and the highest proportion of citizens ensured with Medicare.

2.5 Results

2.5.1 Causal effects

This section will present the results of our matching+PSCM approach to estimate causal effects.

Figure 6.1 shows county-level average treatment effects (ATE) during the vaccination lottery and after the vaccination lottery. During the treatment period, we find positive effects, in the northwest area (Oregon, Washington, and some areas of California), in the Midwest (Ohio, Kentucky, Illinois, and West Virginia), and on the East Coast (New York, Maryland). In the other areas, the effects of the lotteries are minor, and we find small or adverse effects, in the Southern states. County treatment effects appear to be heterogeneous within the same state and between different states. This heterogeneity may be explained by the different background covariates, the timing of the policy or geographical location.

We now focus on causal effects at state-level, estimated using equation 2.6. Table 2.5 shows treatment effects at state-level, and their p-values, while figure 6.2 shows the time series of the treatment effects at the state level. We can evaluate the goodness of fit from synthetic control estimates using pre-treatment RMSPE. In addition, we use visual inspection of the

	RMSPE	Treatment	Post-treatment
Ohio	0.415	1.23	0.195
p-value		0.371	0.522
New York	0.567	0.748	0.756
p-value		0.835	0.418
Oregon	0.689	3.149	1.79
p-value		0.009	0.463
Delaware	0.710	-1.057	-0.713
p-value		0.521	0.728
Maryland	0.615	0.314	0.638
p-value		0.883	0.852
Arkansas	0.673	-0.662	-0.168
p-value		0.774	0.684
California	0.542	-0.184	0.027
p-value		0.522	0.845
Washington	0.785	2.102	1.431
p-value		0.024	0.804
Kentucky	0.517	0.759	-
p-value		0.413	-
North Carolina	0.439	0.044	0.145
p-value		0.933	0.986
Louisiana	0.647	-0.561	0.309
p-value		0.443	0.822
Nevada	0.452	0.046	-
p-value		0.697	-
Maine	0.742	0.154	0.386
p-value		0.986	0.982
Illinois	0.501	1.019	-
p-value		0.918	-
West Virginia	0.563	2.497	1.615
p-value		0.037	0.132
Michigan	0.526	-0.217	-
p-value		0.714	-

Table 2.5: State-level effects on the share of over-18 vaccinated citizenship

pre-treatment difference between treated and synthetic control as in Abadie et al. (2010). Colorado and New Mexico exhibit poor pre-treatment fit, and thus we do not report results from these two states. These poor pre-treatment fit results may depend on discrepancies in weekly reporting of vaccination, with delays in data entry. All the other states have an RMSPE lower than 1 percentage point, which suggests a reasonable fit. We find statistically significant positive impacts on vaccination rollout for West Virginia, Oregon and Washington, and positive effects, but not statistically significant, for Ohio, New York, Maine, Illinois and Maryland. The effects for the remaining state results are small, or even negative for Delaware, Arkansas, and Louisiana but none of the results is statistically significant.

In the post-treatment period, however, the size of the effects tends to decrease in all states and estimates are not statistically significant for any state. Therefore, there is some evidence that the effect of the policy was temporary, and, after the end of the lotteries, no long-term effect is estimated. Note that we cannot evaluate post-treatment effects from Kentucky, Nevada, New Mexico, Illinois and Michigan because lotteries ended after August 24th, 2021.

We also conduct a macro-region analysis, considering spatial aggregates of states. We focus on four US macro-regions, estimating ATE in each of them:

- East Coast (New York, Maryland, Delaware, Maine)
- Southern (Louisiana, North Carolina, Arkansas)
- Midwest (Ohio, Illinois, Kentucky, Michigan, West Virginia)
- West Coast (California, Oregon, Washington)

In this macro-regions analysis, we evaluate only the effect for the treatment period, as some of the post-treatment effects are missing due to the end of the observation period.

Table 2.6 shows our findings. Results are consistent with those we find at the county level. Midwest states have benefited more from the policy, with positive and significant results. This group comprises states that join the policy either early (Ohio), or fairly late (West Virginia, Illinois, Michigan).

We find some positive results, although not statistically significant, in the East Coast area, and in the West Coast.

It is worth noting that West Coast area comprises Washington and Oregon, in which the policy has positive and statistically significant results, and California, in which small or even negative effects were estimated. Negative, but not statistically significant results come from Southern states of the US, as seen in Louisiana and Arkansas ATE.

	RMSPE	**ATE**	**p-value**
East Coast	0.323	0.536	0.968
Southern	0.294	-0.194	0.997
Midwest	0.181	0.865	0.012
West Coast	0.326	0.581	0.589

Table 2.6: Effects on the share of over-18 vaccinated citizenship for US macro-regions

2.5.2 Staggered adoption results

In addition to causal effects for each state, and for major macro-regions, we assess aggregate effects for states that began vaccine lotteries simultaneously, or within the same short time interval.

This analysis allows us to evaluate whether the early introduction of the policy gives some novelty effect. If so, we expect that the states who adopt the policy at early stages have better outcomes than the states adopting the policy later on. The considerable media buzz coming after the Ohio announcement could, at least partially, explain the heterogeneity of the effects between early and late policy adoption.

We consider the following groups of states:

- Early Bird states: adoption within May 20 (Ohio, New York, Oregon, Delaware)
- Second Echelon states: adoption by the end of May (Maryland, California, Arkansas)
- Third Echelon states: adoption within mid-June (Washington, Kentucky, North Carolina)
- Latecomers states: adoption by the end of June (Louisiana, Nevada, Maine, Illinois, West Virginia, Michigan[1])

We excluded Colorado and New Mexico counties from this analysis because of the unreliability of their pre-treatment fit.

Group	**RMSPE**	**ATE**	**p-value**
Early Bird states	0.285	1.254	0.017
Second Echelon	0.265	-0.017	0.960
Third Echelon	0.240	0.969	0.073
Latecomers	0.188	0.498	0.535

Table 2.7: Effects on the share of over-18 vaccinated citizenship, according to timing announcement of the lottery

[1]We include Michigan, which announced it on July, 01 in the June pool.

Table 2.7 shows the ATE and p-values for the four groups of states, classified according to the timing of policy announcements. Figure 6.3 in Appendix 2 reports the time series of their treatment effects.

For Early Bird states and Third Echelon states the policy seems to have induced an additional 1% of the population vaccinated against Covid-19, and these results are statistically significant. Small and statistically negligible results emerge for Second Echelon states and for Latecomers.

Results suggest that there may have been a novelty effect, especially for Early Bird states. However, even if the June effect results are smaller than those in May, the positive sign of the effect should suggest an intrinsic positive effect from the policy, untied from the media attention. In general, multi-level analysis of individual and pooled treatment effects suggests that a higher level of aggregation for treatment effects grants a better fit for the estimation of missing quantities.

2.5.3 Treatment heterogeneity analysis

In section 2.4.4, we have identified six clusters of treated counties, according to socio-economic characteristics. We now estimate ATE within those six clusters.

Figure 2.5 shows the boxplots for treatment and post-treatment effects in each cluster. We estimate positive effects in mean for clusters 1, 2, and 6, which account for about 65% of the treated units. Specifically, clusters 1 and 6, markedly Republican, showed promising results from the policy, in particular, post-treatment evaluation for cluster 6 remains positive, but with considerable variability. We got interesting results in cluster 2, which was wealthy, democratic and more educated on average, without persistence in the post-treatment. We got negative or no results in clusters 3 and 5 the two clusters with higher ethnic components, these clusters accounting for 20% of the total share. It seems that counties with higher shares of Hispanic and Afro-American citizens and lower wealth have worse policy outcomes. This result is consistent with previous findings in the literature, which found a link between ethnicity and vaccine hesitancy (Quinn et al. (2016), Reiter et al. (2020)). This hesitancy appears to be unaffected by monetary incentives, in contrast to the idea that a financial incentive can alter the choices of the less wealthy population.

Generally speaking, we observe high variability within each cluster. Therefore, the background characteristics may not be enough to explain the heterogeneity of treatment effects.

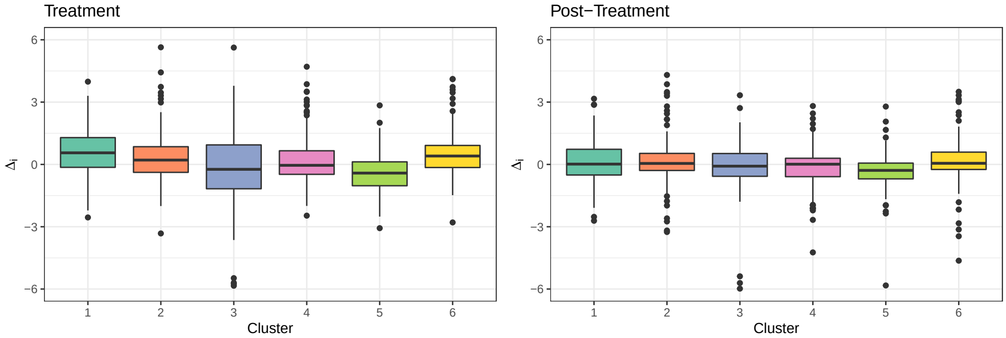


Figure 2.5: Boxplots of county-level effects Δ_i, calculated according equation 2.6, classified by cluster membership

2.6 Conclusions

Policies to incentive vaccination through monetary transfers and lotteries have been rather discussed in the literature, which finds potential benefits in the face of ethical and equity issues, see for example Jecker (2021), Dotlic et al. (2021), Kim (2021), Sprengholz et al. (2021). This study aims to explain the effects of incentives in the treated states, both at different levels of analysis (county, state, macro-region) and at different starting dates, providing an overall picture of policy outcomes. In particular, no previous study has focused on counties and the unique characteristics of each unit. We can see conditional cash lotteries as a viable health policy to incentive people to get a vaccine jab. This is relevant, especially when the vaccination timing is a crucial variable in determining the outcome of the vaccination rollout. However, our results suggest that the outcome of such policies is not unidirectional. In particular, we find different results across counties belonging to the same state and between counties belonging to different states.

Washington, Oregon and West Virginia outperformed other states getting positive and significant results. Other states got positive, yet not statistically significant, effects (Ohio, New York and Illinois), while the remaining states get no to negative results.

The analysis of staggered adoption effects shows that causal effects are higher in states that adopted the policy early (Early birds states). Media-buzz effects around the policy can explain these findings, as hypothesized by Gorin and Schmidt (2015). Nevertheless, policies can still produce positive results, even in the case of late adoption, as in the case of West Virginia. Cross-bordering effects are also present. In particular, treated counties in the Midwest of the US seem to have benefited most compared with counties in three other macro-region (East Coast, West Coast and Southern US). The study of the effects after the lottery ending is also unprecedented. This analysis is relevant as it differentiates between counties and states that could have experienced permanent effects from the policy or only temporary ones. This dimension is decisive for the policymaker, adopting this incentive if they should achieve policy goals fast. We observe temporary positive effects in most of the treated counties, especially for the counties in the Midwest and Northwestern areas. However, we got lower effects in the Sunbelt, where monetary incentives did not convince people who are averse to the vaccine.

After initial success, maybe backed up by the media attention, the vaccine lotteries did not significantly affect the vaccination choices of Americans. Small and statistically negligible post-treatment results suggest that lotteries have induced people that would have been vaccinated at some point to anticipate the jab. Nonetheless, even short-term temporary impacts, such as the ones we find for US lotteries, may be meaningful when the health policy goal has to be achieved within a short time.

Chapter 3

SMaC: Spatial Matrix Completion Method

3.1 Introduction

The Synthetic Control Method (SCM hereinafter) is a widespread methodology to estimate causal effects in presence of a single treated unit and many control units, observed over time (Abadie and Gardeazabal, 2003; Abadie et al., 2010; Abadie, 2021). With this method, the impact of an intervention is evaluated as the difference between the observed value of some primary outcome and its counterfactual value, imputed by using a weighted average of control units. The popularity of SCM is growing rapidly, and its field of application range from social science, to ecological studies, to policy evaluation. Athey et al. (2021) defined it as "arguably the most important innovation in the policy evaluation literature in the last 15 years" and nowadays the methods counts over 4,000 citations on Google Scholar.

Evidence of interest in SCM is the flurry of methodological developments. Xu (2017), Amjad et al. (2018) and Ben-Michael et al. (2021) have broadened the original estimator, allowing for unconstrained weights. Another group of works suggests using penalization in the estimation of the weight, such as the recent proposal from Abadie and L'Hour (2021), which penalizes the discrepancies between treated and control units individually. The estimator, called *penalized synthetic control*, results to be an ensemble estimator between the classical SCM and the nearest neighbour matching estimator. A similar proposal comes from Kellogg et al. (2021), trading off between interpolation and extrapolation bias. Arkhangelsky et al. (2019) and Bottmer et al. (2021) introduce double robust estimators, considering the interaction of SCM and Difference-in-Differences (DID) models.

Doudchenko and Imbens (2016) and Athey and Imbens (2021) notice that the SCM is a subclass of a more large set of Matrix Completion Method

(MCM hereinafter) estimators, the so-called *vertical regression* estimators, which impute the missing potential outcomes in post-treatment as a linear combination of a fixed effect and control units. These recent estimators have led to an increase in fitting performances, in spite of the original SCM weights transparency. See Abadie (2021) for an extended review of SCM and MCM.

Recently, the exploration of SCM and MCM alternatives heads toward Bayesian regression models. Menchetti and Bojinov (2020), Kim et al. (2020), Pang et al. (2022) and Pinkney (2021) use Bayesian methods for causal effects estimation, illustrating a simple and effective proposal for inference in SCM-like settings. Finally, recent work from Arbour et al. (2021) investigates the use of multitask Gaussian Processes for weight estimations.

The growing use and acceptance of these methods are leading them to be applied to new research areas and types of data. In particular, many fields where SCM is commonly used study outcomes which are measured in spatial areas such as municipalities, states or regions. Abadie (2021) suggests these as the specific framework of application for SCM-like methods. In such contexts, it is common to see treatment assigned to a single area, and the focus is to estimate the treatment effect on this treated unit. Usually, scholars consider no second-round effects from the treatment, neither in terms of spillovers nor in terms of effect propagation. This is coherent with the no-interference assumption of the SUTVA. Some recent works (Grossi et al., 2020; Cao and Dowd, 2019; Di Stefano and Mellace, 2020) try to estimate causal quantities using SCM in the presence of interference. In particular, Grossi et al. (2020) exploit a partial interference assumption (Forastiere et al., 2021a; Papadogeorgou et al., 2019) to identify the spillover effect of a new tramway line construction.

However, no previous work has addressed spatial treatment effect propagation explicitly within the scope of SCM or MCM. In practice, researchers often evaluate the extent to which treatment effects propagate through space by applying SCM to areas of different sizes around the treated location. In principle, this choice does not harm the unbiasedness of estimation, but it could affect the efficiency of estimators in the presence of spatial (or network) data. Moreover, separate estimation of coefficients can lead to vectors of weights very different for contiguous units, where we expect similar areas to have similar SCM weights. In this work, we propose a Bayesian estimator for missing potential outcomes in presence of spatial correlation among treated units. We exploit a Gaussian process prior to the vertical regression coefficients that take into account spatial correlation, encouraging regression coefficients across similar areas to be similar. We aim to exploit this spatial information to estimate counterfactual quantities that are still unbiased, but have improved properties in terms of mean bias and mean square error of the point estimate with respect to the separated SCM or vertical regression methods. We refer to this method as *Spatial Matrix Completion* or SMaC.

Our motivating application is the impact evaluation arising from the construction of the first line of the Florentine tramway network. In particular, we wish to assess the infrastructural impact on the commercial vitality of the treated neighbourhood, measured as the number of stores located within some distance d from a tramway stop.

The chapter will follow this outline: section 3.2 introduces the related works and the empirical literature and the data related to our application, section 3.3 introduces the causal estimands, and section 3.4 presents the methodology (SMaC). Section 3.5 presents the simulations studies we carried out and section 3.6 presents our study results. Finally section 3.7 concludes and discuss future research.

3.2 The tramway and the city

The impact of investments in transportation infrastructure, such as light rail systems, is a highly debated topic in the field of urban and transport economics (see Cervero and Landis, 1993, Landis et al., 1995, Baum-Snow and Kahn, 2000, Hess and Almeida, 2007, Pan, 2013, Papa and Bertolini, 2015). Naturally, transportation infrastructures are designed to improve the accessibility of the served areas, by reducing the access time, and the congestion charge and improving the general vitality of the area. In fact, it is not rare that transportation infrastructure developments lead to a broader renewal in the served areas, including the revitalization of public spaces, investments in public parks and common areas, as well as an overall improvement in the usability of the neighbouring areas. These effects are expected to be more pronounced when there is a well-connected and extensive network of public transportation, as previously demonstrated in the studies of Mejia-Dorantes et al. (2012) and Credit (2018).

Focusing on the economic outcomes of such interventions, scholars usually analyze the relationship between improvements in urban transport and real estate prices, with a positive correlation between the construction of transport facilities and house selling prices (see for instance Pagliara and Papa, 2011; Yan et al., 2012). Two pieces of work study the causal effects of the construction of the first line of the Florentine tramway. Budiakivska and Casolaro (2018) find an overall positive effect on real estate prices, coherent with the literature. Grossi et al. (2020) focus on the effect on the commercial environment, studying the direct and spillover effects triggered by the intervention.

However, much of this literature has focused on the effects of infrastructure during its operational phase, and less attention has been paid to the impact of the construction period on affected areas. Specifically, decreased accessibility may result in significant negative impacts for areas directly impacted by the construction site, while a shifting of benefits to neighbouring areas may occur. Furthermore, it is also important to consider possible displacement effects, where areas in proximity to the worksite may experience negative impacts while areas farther away may see benefits as a result of increased accessibility and a renewed environment.

It can be reasonably assumed that the impact of the treatment will vary based on proximity to the construction site, with the most significant effects being observed in areas closest to the light rail stops. Thus, during the construction period congestion charge, reduced accessibility, noise pollution and other drawbacks can affect the inner areas leading to a reduction in commercial vitality. On the converse, once the light rail system becomes operational, inner areas may experience increased accessibility and improved surroundings, as previously demonstrated in studies such as Credit (2018) and Pogonyi et al. (2021).

The first Florentine tramway network was built at the very beginning of the XX century, later dismissed in favour of private forms of transport. In the late 80s, arise a debate within the city council about the rebuilding of a tramway network, in separate lines with respect to the car lines. After a long debate, the tramway line we are considering has been built between 2006 and 2010, and since that moment is connecting the central railway station in the city centre with the densely populated municipality of Scandicci, passing by the south-west neighbourhood of Isolotto and Legnaia. In total, the portion of the line we are considering spans 3.2 km with 7 stops.

3.2.1 Data

We collect information on the physical location of businesses in the Florentine neighbourhood. We consider the neighbourhood of Legnaia, in which the tramway has been built and ten additional neighbourhoods located in the peripheral and semiperipheral areas of Florence, most similar to the treated one based on researcher knowledge of the territory. The central city area of Florence is not included in this analysis as it presents distinct characteristics that differentiate it from the peripheral neighbourhoods being studied. We excluded also those peripheral neighbourhoods which have different morphological configurations. In fact, we excluded those hilly neighbourhoods located in the north and south of Florence.

These neighbourhoods are selected based on criteria established by the Real Estate Observatory of the Italian Ministry of Finance.

We also collect information on the tramway stops in Florence. We distinguish between the set of tramway stops in the Legnaia neighbourhoods and the counterfactual sets of tramway stops, located in neighbourhoods

Figure 3.1: Catchment areas around the tramway stops, up to 0.5 km

far away from the tramway site. We treat hypothetical tram stops as being situated in streets similar to those where the tramway already runs. The classification of stores is based on their distance from the closest tram stop. Specifically, buffer zones of 50-meter increments up to 400 meters are created around each tram stop, with no overlap between them. A visual representation of this setup can be seen in the accompanying figure 3.1. Outcome variables for each stop ring originate from the Statistical Archive of Active Firms (SAAF, English translation of ASIA, the Italian acronym for "Archivio Statistico delle Imprese Attive"). The SAAF is held by the Italian National Istitute of Statistics (ISTAT).

This dataset contains information on active businesses in each year of the analysis. It includes the geographic location and economic sector (as classified by the Statistical Classification of Economic Activities in the European Community, or NACE) of each business. This dataset is sourced from the period 1996-2014, and for the purposes of our analysis, we focus on counting the number of active firms in each buffer zone around the tramway stops in each neighbourhood, creating a panel dataset. Our focus is solely on commercial vitality, expressed as the number of stores in a certain area and not on any additional outcome or background variables due to limitations in data availability.

3.3 Causal Framework

3.3.1 Notation

Consider a space $\mathbf{\Omega}$ that can be partitioned into N areas: $\mathbf{\Omega}_i$ in $i \in \mathbf{N} = \{1, \ldots, i, \ldots, N\}$, so that $\bigcup_{i=1}^{N} \mathbf{\Omega}^i = \mathbf{\Omega}$ and $\mathbf{\Omega}^i \cap \mathbf{\Omega}^j = \emptyset$ for each couple $i \neq j \in \{1, \ldots, N\}$. In our study, we consider the natural partition of our sample space into clusters representing the Florentine neighbourhoods.

We observe treatment arising from some specific locations $\boldsymbol{\omega}_1 \in \boldsymbol{\Omega}_1$. We can consider treatment locations as point treatments. Let $\boldsymbol{\omega}_1$ be the set of treatment locations, in our application we consider $\boldsymbol{\omega}_1$ as the tramway stops located in the treated area. Note that all the tramway stops of the line we are considering are located in the treated area and none outside. We also consider sets of locations $\boldsymbol{\omega}_i, i \in \{2, \ldots, N\}$ as sets of points located in neighborhoods far away from the tramway line, in streets similar to the one that receives the treatment. Points in $\boldsymbol{\omega}_i$ are randomly chosen, with a mean distance across points similar to the distance across tramway stops in the treated neighborhood. Note that, for $\boldsymbol{i} \in \mathbf{N}$, the total set of locations $\boldsymbol{\omega} = \bigcup_i \boldsymbol{\omega}_i$ is the set of black dots in figure 3.2.

We define our observation units as the areas around the treatment sites $\boldsymbol{\omega}$. Therefore, for each neighbourhood $\boldsymbol{i}$ we construct a set of buffers areas $\boldsymbol{A}_i = \{A_i^1, \ldots, A_i^h, \ldots A_i^H\}$ around the treatment locations $\boldsymbol{\omega}_i$, using the vector of distances $\mathbf{D} = (d_1, \ldots, d_h, \ldots, d_H)$ representing the distance of the $h-$th area from the treatment site. We sort units and distances such that $d_{h+1} \geq d_h \quad h \in \{1, 2, \ldots, H-1\}$.

In several applied studies scholars consider the treatment arising from one (or multiple) locations, with the effect spreading to neighbouring areas (see for example Zigler et al., 2020). In principle, a treatment can arise effects also to faraway units, but in practice, we can focus on the area within some user-specified distance d. Thus, the choice of the vector of distances $\mathbf{D}$ can be relevant. In our example, we consider distances up to 400 meters, which is a reasonable walking distance to a tramway stop, and

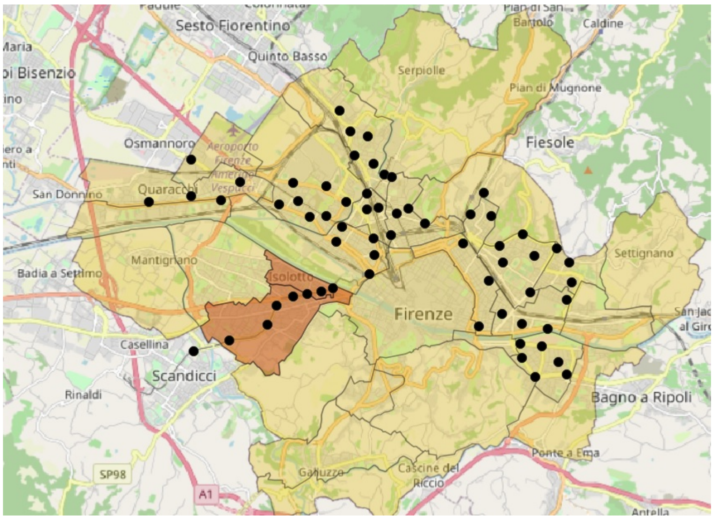


Figure 3.2: Red Area: treatment area $\boldsymbol{\Omega}^1$; Yellow Area: control area $\boldsymbol{\Omega}^0$. Black dots represent actual tramway stops (in red area) or counterfactual stops (in the yellow area).

the area that is more likely to receive the effects of an urban transport facility, in fact Guerra et al. (2012) and Guerra and Cervero (2013) states that the area within the quarter of a mile is the one more likely to receive the effects of new urban infrastructures.

We repeatedly observe units over time, so we consider a panel data setting, with $H \times N$ areas observed for $\boldsymbol{T}^0 = (1, \ldots, t_0 - 1)$ pre-treatment periods, and $\boldsymbol{T}^1 = (t_0, \ldots, T)$ post-treatment periods. All the treated units are treated at $t = t_0$. Let $Y_{i,t}^h$ be our primary outcome, the number of stores in neighbourhood i within distance d_h from the tramway stops in each time period $t \in \boldsymbol{T}$. Let be $\boldsymbol{z} = [z_1^1, \ldots, z_1^h, \ldots, z_1^H, \ldots, z_i^h, \ldots, z_N^H]$, $\quad z_i^h \in \{0, 1\}$ be a neighbourhood-level treatment for each area A_i^h considered. Thus following, units belonging to the same cluster i can be only treated or not-treated together, therefore $\boldsymbol{z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_i, \ldots, \boldsymbol{z}_N]$, with $\boldsymbol{z}_i = [z_i^1, z_i^2, \ldots, z_i^h, \ldots, z_i^H]$. We consider two alternative regimes for $\boldsymbol{z}$:

$$\boldsymbol{z}(1) = [z_1^1 = 1, \ldots, z_1^h = 1, \ldots, z_1^H = 1, \ldots, z_i^h = 0, \ldots, z_N^H = 0]$$

$$\boldsymbol{z}(0) = [z_1^1 = 0, \ldots, z_1^h = 0, \ldots, z_1^H = 1, \ldots, z_i^h = 0, \ldots, z_N^H = 0]$$

$\boldsymbol{z}(1)$ is the scenario in which each area $A_1^h \in \boldsymbol{\Omega}_1$ receives the treatment, and no one outside. Instead, $\boldsymbol{z}(0)$ represents the scenario in which no area results are treated, in our scenario the situation in which the tramway was never built in Florence. We consider that areas $\boldsymbol{A}_1 = \{A_1^1, \ldots, A_1^h, \ldots, A_1^H\} \in \boldsymbol{\Omega}_1$ will receive the treatment starting from the period t_0, and remain treated afterwards. In our application, we consider the treated space as the Legnaia neighbourhood where the tramway stops are located, and t_0 =2006. Units located in other parts of Florence will be considered non-treated units with $\boldsymbol{A}^0 = \{A_2^1, \ldots, A_i^h, \ldots, A_N^H\} \notin \boldsymbol{\Omega}_1$. Areas in A_0 are surrounding the counterfactual tramway stops, which are points chosen at random on streets located faraway from the tramway path. The mean distance between counterfactual tramway stops is mimicking the average distance between stops in the treated area. A sketched map of the treated and non-treated neighbourhoods can be seen in figure 3.2.

We adopt the potential outcome approach to causal inference (Rubin, 1974, Rubin, 1978). Under consistency assumption (Rubin, 1980), for each unit A_i^h in each period t we define the following couple of potential outcomes:

$$\begin{cases} Y_{i,t}^h(1) \equiv Y_{i,t}^h(\boldsymbol{z}(1)) & \text{as the potential outcome under } \boldsymbol{z}(1) \text{ assignment} \\ Y_{i,t}^h(0) \equiv Y_{i,t}^h(\boldsymbol{z}(0)) & \text{as the potential outcome under } \boldsymbol{z}(0) \text{ assignment} \end{cases} \tag{3.1}$$

Under consistency, we consider no hidden version of the treatment, and therefore we define $Y_{i,t}^h(z_{i,t})$ as the potential outcome for areas in neighbourhood $\boldsymbol{i}$, in time $\boldsymbol{t}$ at a distance d_h from the treatment.

In contexts with cluster-level treatment allocation, scholars often invoke a partial interference assumption (Sobel, 2006, Hudgens and Halloran, 2008, Papadogeorgou et al., 2019), which rules that interference may occur, but not within groups. Formally:

Assumption 4. *(Partial Interference). For $t = t_0 + 1, \ldots, T$, for all*

$$[\boldsymbol{z}_1, \ldots, \boldsymbol{z}_i, \ldots, \boldsymbol{z}_N] \quad \textit{and} \quad [\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_i^*, \ldots, \boldsymbol{z}_N^*]$$

with $z_i = z_i^*$,

$$Y_{i,t}^h([\boldsymbol{z}(1), \ldots, \boldsymbol{z}_i, \ldots, \boldsymbol{z}_N]) = Y_{i,t}^h([\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_i^*, \ldots, \boldsymbol{z}_N^*])$$

for all $\boldsymbol{i} \in N, d_h \in \mathbf{D}$.

In our application, this means that the effect of the tramway cannot emanate up to untreated units. This assumption can be considered reasonable, such as control areas are not contiguous to the treated areas nor connected to them by the tramway. Therefore, the potential outcomes for units in the neighbourhood $\boldsymbol{i}$ depend only on the treatment assignment of their own neighbourhood, and thus $Y_{i,t}^h(\boldsymbol{z}_i)$.

We consider that the store location is not influenced by the expectation of future treatment, invoking a *non-anticipating treatment* assumption (Abadie et al., 2010). Under this assumption, the outcome at some time period is not influenced by a treatment applied later. This assumption is crucial in vertical regression contexts, as it allows us to use the relationship between treated and control units in $\boldsymbol{T}^0$ and predict the counterfactual values in $\boldsymbol{T}^1$. It seems to be plausible to consider non-anticipating treatment in our application, as the construction of the first line of the Florentine tramway network experienced a long decisional phase, with many delays that could have discouraged potential early settlers to relocate the activity in the future served area.

Finally, let us define the observed outcomes at time $\boldsymbol{t}$ as:

$$Y_{i,t}^h = \begin{cases} Y_{i,t}^h(\boldsymbol{z}(0)) & \text{if } t < t_0 \\ Y_{i,t}^h(\boldsymbol{z}(1)) & \text{otherwise} \end{cases} \tag{3.2}$$

Before construction starts, we do not observe any unit affected by the treatment, while after t_0, all units in the treated neighbourhood will be impacted by the treatment.

3.3.2 Causal Estimands

In a setting in which treatment arises in some locations $\boldsymbol{\omega}_1$ in the Legnaia neighbourhood, we are interested in evaluating the causal effect on treated areas A_i^d, at times $t \in T^1$. For $i = 1$ and $\forall t \in T^1$ we define the causal effect for the treated units as

$$\Delta_{1,t}^d = Y_{1,t}^h(\boldsymbol{z}(1)) - Y_{1,t}^h(\boldsymbol{z}(0)) \quad \forall t \in T^1, d_h \in \mathbf{D} \tag{3.3}$$

We can define also the intertemporal average :

$$\Delta_1^h = \frac{1}{T - t_0} \sum_{t=t_0}^{T} \Delta_{1,t}^h \tag{3.4}$$

to evaluate the overall effect through the treatment period for an area in the treated cluster, within distance d_h. From the comparison of effects at different distances from the treatment site, we can get precious insights into the transmission of treatment effects through space. In general, we could expect decaying treatment effects up to some boundary of spatial treatment d_∞, in which $\Delta_i^{d_\infty} \to 0$. Studying diffusion effects at different distances from the treatment site can shed light on how the treatment emanates across space. It is possible that, even in presence of spatially correlated units, the treatment effect will differ between inner buffers and outer areas. Such analysis can be particularly useful for those applications related to the spatial dimension of causal effects, such as transport economics, or environmental economics.

3.4 Estimation of causal effects

In order to describe our estimation strategy, we first introduce the SCM estimator and its related developments. Subsequently, we introduce the proposed Bayesian framework for Spatial Matrix Completion. Lastly, we will compare our proposed method with the aforementioned estimators.

3.4.1 Separate vertical regressions

Denote the outcomes for areas in neighbourhood i at distance d_h from treatment as $Y_{i,t}^h$ at time periods $t = \{1, 2, \ldots, T\}$, and $\mathbf{Y}_{i,t} = (Y_{2,t}^1, \ldots, Y_{i,t}^h, \ldots, Y_{N,t}^H)^T$

as the vector of control outcomes at time t. These outcomes can represent both a continuous random variable (for example, the minimum temperature in an area). In our study, $Y^h_{i,t}$ is a count random variable, such as the number of shops within a certain distance.

One might be interested in understanding the effect that treating the specific location ω_i had on the area comprised within a specific distance $d \in \mathbf{D}$ versus not treating it. Remind that for the treated units we observe $Y^h_{1,t} = Y^h_{1,t}(\boldsymbol{z}(1))$ when $t \geq t_0$, so we need to impute the missing quantity $Y^h_{1,t}(\boldsymbol{z}(0))$. In panel data settings, with one o multiple treated units, one could impute the missing outcomes by using a linear combination of the control unit's outcomes, as in SCM-like methods. Specifically, the estimate of the causal effect of the treatment at time period $t \geq T_0$, for treated unit at distance d_h is:

$$\widehat{\Delta}^h_{1,t} = \underbrace{Y^h_{1,t}}_{\text{observed, with treatment}} - \underbrace{\left(\sum_{i=2}^{N} Y^h_{i,t}\beta^h_i\right)}_{\text{SCM-imputed, no treatment}} \tag{3.5}$$

Specifically, one can find $\boldsymbol{\beta}^h_C = (\beta^1_2, \ldots, \beta^h_N)^T \in \mathbb{R}^{N-1}$ such that:

$$\left(\boldsymbol{\beta}^h_C\right) = \underset{\boldsymbol{\beta}^h_C \in \mathbb{R}^{N-1}}{\operatorname{argmin}} \left\{ \sum_{t=1}^{t_0-1} \left(Y^h_{1,t} - (\mathbf{Y}_{i,t})^T \boldsymbol{\beta}^h_C \right)^2 \right\}. \tag{3.6}$$

with $\beta^h_i \geq 0 \; \forall i$ and $\sum_i \beta^h_i = 1$. $\boldsymbol{\beta}^h_C$ corresponds to the solution using the synthetic control method (Abadie et al., 2010), with non-negative constraint on coefficients, that should sum up to one.

The same problem could be faced by removing the constraints on regression coefficients and adding an intercept to the model. As noted by Doudchenko and Imbens (2016) and Ferman and Pinto (2021), this minimization problem is also the solution of a linear regression where the outcome $Y^h_{i,t}$ is regressed on the outcomes of the control units during the same time period, using data only from the first $T_0 - 1$ time periods. Let be $\beta^h_0 \in \mathbb{R}^1$, and $\boldsymbol{\beta}^h = (\beta^h_0, \boldsymbol{\beta}^h_C)$, thus the vector of coefficients

$$\begin{pmatrix} \beta^h_0 \\ \boldsymbol{\beta}^h_i \end{pmatrix} = \underset{\boldsymbol{\beta}^h \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \sum_{t=1}^{t_0-1} \left(Y^h_{1,t} - (1 \;\; \mathbf{Y}_{i,t})^T \boldsymbol{\beta}^h \right)^2 \right\}. \tag{3.7}$$

identifies the solution using a vertical regression method (Doudchenko and Imbens, 2016, Athey et al., 2021).

The synthetic control weights and vertical regression coefficients can be calculated separately for different choices of $d \in [d_1, d_H]$. For example, to find the synthetic control weights at distances $d_1 < d_2 < \cdots < d_H$, one could solve the minimization problem in 3.6 using a constrained optimization procedure, separately for each of these distances. Alternatively, the H different minimization problems could be *stacked*, and one could solve the combined minimization problem

$$\begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_i \\ \vdots \\ \boldsymbol{\beta}_N \end{pmatrix} = \underset{\boldsymbol{\beta}_0,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_i,\ldots,\boldsymbol{\beta}_N \in \mathbb{R}^{NH}}{\operatorname{argmin}} \left\{ \sum_{h=1}^{H} \sum_{t=1}^{T_0-1} \left(Y_{1,t}^h - (1\ \mathbf{Y}_{i,t})^T \boldsymbol{\beta} \right)^2 \right\} \tag{3.8}$$

where $\boldsymbol{\beta}_i = (\beta_i^1, \beta_i^2, \ldots, \beta_i^H)^T$ is the vector of the same parameter β_i^h in the H vertical regression models, which will return the exact same solutions as solving 3.7 separately for each distance. Additionally, we can consider a penalized vertical regression, we refer to it as *pooled ridge* in equation 3.9:

$$\begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_i \\ \vdots \\ \boldsymbol{\beta}_N \end{pmatrix} = \underset{\boldsymbol{\beta}_0,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_i,\ldots,\boldsymbol{\beta}_N \in \mathbb{R}^{N}}{\operatorname{argmin}} \left\{ \sum_{h=1}^{H} \sum_{t=1}^{T_0-1} \left(Y_{1}^h - (1\ \mathbf{Y}_{i,t})^T \boldsymbol{\beta} \right)^2 + \sum_{i=1}^{N} \lambda \boldsymbol{\beta} \mathbb{I} \boldsymbol{\beta} \right\}, \tag{3.9}$$

with λ as a scalar penalization term and $\mathbb{I}$ the identity matrix.

3.4.2 Bayesian approach for Spatial Matrix Completion

In this section, we will introduce the Bayesian framework we will use to impute the missing outcome $Y_{1,t}^d(\boldsymbol{z}(0))$. Building on the vertical regression idea (Abadie and Gardeazabal, 2003, Doudchenko and Imbens (2016), Athey and Imbens (2021)) we will propose a matrix completion algorithm that smooths regression coefficient values through contiguous treated units. Bayesian solutions for vertical regression and SCM-like settings are arising interest, with the recent work of Kim et al. (2020) or Pang et al. (2022) which have proposed Bayesian alternatives to the classical SCM. One of the most salient advantages of Bayesian modelling is its clear and transparent method for evaluating uncertainty in causal effects, which allows for the computation of credibility intervals from the posterior distribution of the parameters. This feature is particularly important in SCM settings in which scholars get a single point from the estimation, and therefore placebo-test-based inference is usually applied. See for a review of such inferential procedures Abadie (2021), Abadie et al. (2010), Cattaneo et al. (2021).

In order to consider the spatial structures of the observed treated units, yet being flexible in the parameter estimation, we follow a Bayesian regression approach, using Gaussian processes as priors for control unit coefficients. For instance, consider the Gaussian process $\mathbf{f}$ as:

$$\mathbf{f} \sim \mathcal{GP}(\mu, \mathcal{K}_{\alpha,\rho})$$

parameterized in the vector of means μ , and covariance function $\mathcal{K}_{\alpha,\rho}$, where α is the amplitude of the process, and ρ) the lengthscale of the process. Let $\mathbf{x} = \{x_1, \ldots, x_p, x_q, \ldots, x_R\}$ be a set of inputs x. A Gaussian process has the appealing property that every finite collection of it has a multivariate normal joint distribution.

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{MVN}\left(\mu(\mathbf{x}), \mathcal{K}_{\alpha,\rho}(\mathbf{x})\right)$$

See Williams and Rasmussen (2006) and Gramacy (2020) for an extended presentation of its properties. A key role in GPs is played by the covariance function $\mathcal{K}_{\alpha,\rho}$, ruling the characteristics of the function we wish to predict, and therefore, our prior beliefs on the data structure. Consider for example, the specification of an exponential smoothing kernel, such as:

$$\mathcal{K}_{\alpha,\rho}(\mathbf{x}) = \alpha \exp\left\{-\frac{(x_p - x_q)^2}{2\rho^2}\right\}$$

We can see that the covariance between the two points x_p and x_q exponentially decays when $(x_p - x_q)$ increases. This representation could be useful to represent longitudinal settings, as a couple of outcomes (Y_p, Y_q) will be less correlated as their temporal distance $(x_p - x_q)$ increases. A similar argument holds for spatial settings where we expect closer points to be in being more correlated, with a covariance matrix ruled by $\mathcal{K}_{\alpha,\rho}(\mathbf{x})$. The relative distance between two points is not the only crucial ingredient in this recipe. In fact, the correlation matrix will depends also on our belief in the total amplitude of the process α, representing the signal variance which determines the average distance of the data-generating function from its mean and on the lengthscale parameter ρ, which represents how smoothed will be our process. Figure 3.3 shows realizations of Gaussian Processes under six different scenarios, using an exponential smoothing kernel. The first row represents GP realizations for different levels of ρ. We can notice

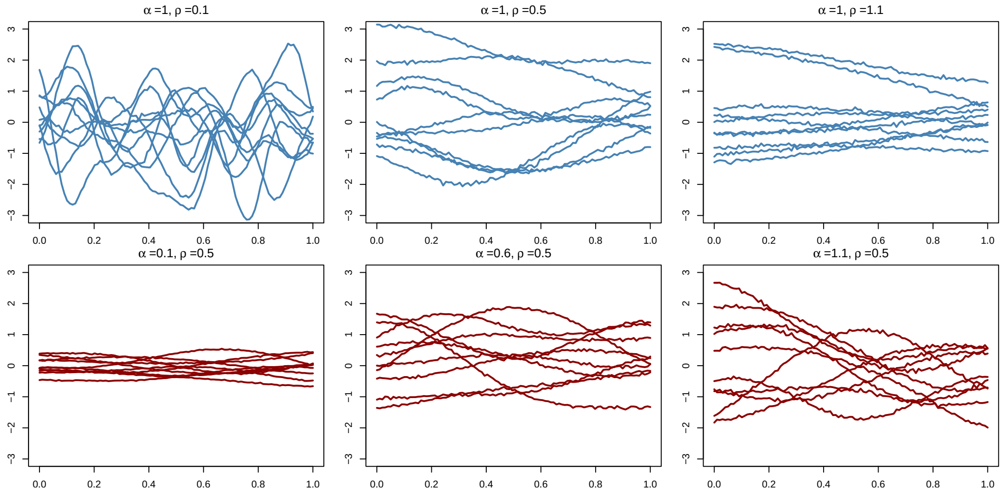


Figure 3.3: Gaussian Process realizations with varying α and ρ

that, at the same level of α, when the lengthscale increases the process will result in smoother realizations. Instead, when ρ is small, even relatively nearby inputs can have a small covariance. In the extreme case of $\rho \to 0$ the kernel matrix will be the identity matrix. The second row instead represents GP realizations when ρ is fixed and α varies. In this case, larger values of α will be associated with larger "width" of the Gaussian process, while with smaller α all the realization will be concentrated around the mean.

By altering the kernel specification, different correlation structures among observations can be achieved, thus providing a prior specification that aligns more closely with our hypotheses. Some examples of correlation structures that can be utilized include the periodic, linear, and Matèrn kernels.

The use of Gaussian processes in machine learning has seen a significant increase in recent years due to their remarkable flexibility and versatility. Researchers have been leveraging the properties of Gaussian processes to estimate causal quantities, as demonstrated in various studies, such as in Alaa and Van Der Schaar (2017), Huang et al. (2019) and Witty et al. (2020).

Building on the multitask Gaussian Process proposal from Bonilla et al. (2007), Arbour et al. (2021) proposes a GP-based estimator for causal quantities in panel data settings. Kanagawa et al. (2018) introduce the similarities between Gaussian Processes and Matrix completion methods, allowing for weighted representations of outcomes.

In our setting, Gaussian processes can be particularly useful, as we could exploit the spatial information in our data for the specification of regression coefficients.

Let $\mathbf{Y_1^d} = (Y_{1,1}^d, Y_{1,2}^d, \ldots, Y_{1,t_0}^d)'$ be the vector of pre-treatment outcomes for the treated area and distances $d \in \{1, \ldots, H\}$. We also define the vector $\mathbf{Y} = (\mathbf{Y_1^1}, \mathbf{Y_1^2}, \ldots, \mathbf{Y_1^H})$ which is of length Ht_0 and includes the pre-treatment outcomes across all distances.

Similarly, let be $\mathbf{Y_i^d} = (Y_{i,1}^d, Y_{i,2}^d, \ldots, Y_{i,t_0}^d)'$ be the vector of pre-treatment outcomes for the control area i, and be $\mathbf{X}_i$ the $t_0H \times H$ block-matrix of control pre-treatment outcomes, such that

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{Y}_i^1 & & & & & \\ & \mathbf{Y}_i^2 & & & & \\ & & \cdots & & & \\ & & & \mathbf{Y}_i^d & & \\ & & & & \cdots & \\ & & & & & \mathbf{Y}_i^H \end{pmatrix}$$

Then let be $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_2, , \ldots, \mathbf{X}_i, \ldots, \mathbf{X}_N]$ be the $t_0H \times H(N)$ matrix of pre-treatment outcomes for the control units plus an intercept matrix $\mathbf{X}_0$. Let $\beta_i = (\beta_i^1, \beta_i^h, \ldots, \beta_i^H)$ be the vector of coefficients for control unit i relative to different treated, and let $\boldsymbol{\beta}$ be the vector of pooled coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta_0}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta_i}, \ldots, \boldsymbol{\beta}_N)$ for each control unit $i \in \{2, \ldots, N\}$ and an intercept. We consider β varying smoothly through space, in particular, the vector of coefficients β^h will be more similar for physically close units. We specify such structure by using a Gaussian process prior for β such that

$$\beta_i(\mathbf{D}) \sim \mathcal{GP}\left(\mathbf{0}, \mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D})\right)$$

with $\mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D})$ as a quadratic exponential smoothing kernel with parameters α_i and ρ_i. Thus, the (p, q) entry of $\mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D})$ is

$$[\mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D})]_{pq} = \alpha_i \exp\left\{-\frac{(d_p - d_q)^2}{2\rho_i^2}\right\}$$

Thus, $\boldsymbol{\beta} \sim \mathcal{MVN}(0, \Sigma)$, where Σ is an appropriate block covariance matrix for the pooled coefficient estimation for multiple treated units. Hence,

$$\Sigma = \begin{pmatrix} \mathcal{K}_{\alpha_0,\rho_0}(\mathbf{D}) & & & & & \\ & \mathcal{K}_{\alpha_2,\rho_2}(\mathbf{D}) & & & & \\ & & \cdots & & & \\ & & & \mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D}) & & \\ & & & & \cdots & \\ & & & & & \mathcal{K}_{\alpha_N,\rho_N}(\mathbf{D}) \end{pmatrix}$$

We define the Bayesian regression model as

$$
\begin{aligned}
\mathbf{Y} &\sim \mathcal{N}(\beta^T \mathbf{X}, \sigma_y \mathbf{I}) && (3.10)\\
\boldsymbol{\beta_i} &\sim \mathcal{GP}(\mathbf{0}, \mathcal{K}_{\alpha_i,\rho_i}(\mathbf{D})) && (3.11)\\
\alpha_i &\sim \Gamma^{-1}(50, 5) && (3.12)\\
\rho_i &\sim \Gamma^{-1}(5, 5) && (3.13)\\
\sigma_y &\sim \Gamma^{-1}(5, 5) && (3.14)\\
& && (3.15)
\end{aligned}
$$

This framework has simple yet powerful relapses. In context with spatially correlated units, Bayesian regression with Gaussian process priors has good frequentist properties, improving inference both in terms of bias and in terms of efficiency. Moreover, from the posterior distribution of $\beta_{\mathbf{i}}$ we can derive the smoothed path of the coefficient for some control unit $\boldsymbol{i}$ across the treated units $d \in \mathbf{D}$. Lastly, we can easily derive credibility intervals for the posterior distribution of the causal effect, retrieving it from the posterior distribution of β_i. Even if we choose to specify weakly informative priors, other approaches are possible: we could also estimate from the data the amount of spatial correlation between treated units, and exploiting an empirical Bayes procedure, provide an estimate for the α_i^{EB}, ρ_i^{EB} and σ_y^{EB} hyperparameters. In fact, the choice of the prior for α_i is heading in this direction, as we wish to avoid the risk of overfitting, we opt for a prior that should induce sparsity in the regressor estimation. Implementation of empirical Bayes methods in the context of the Gaussian process has been also studied by Krivoruchko and Gribov (2019) and Stijnen (1982). Building on Williams and Rasmussen (2006) and van Wieringen (2015), we show in the Appendix that there are no differences, in terms of point estimates, between solving the pooled minimization problem in 3.8 using SMaC and using a generalized ridge regression with penalization matrix $\mathbf{\Phi} = \Sigma^{-1}$ and penalization term $\lambda = \frac{\sigma_y}{\alpha}$.

3.4.3 Estimated SC weights as a function of distance

In fully separated models, coefficients can take very different values also for similar treated areas. Within the context of our study, we implemented separate matrix completion methods for distances $d \in \{50, 100, \ldots, 400\}$. In particular, within a vertical regression method, we estimate control units' coefficients by using the SCM method (equation 3.6), the pooled ridge vertical regression (equation 3.8) and the Bayesian vertical regression with GP priors, or SMaC (equation 3.10). Figure 3.4 shows the SC weight and the pooled ridge coefficient of a randomly chosen control unit in our study as a function of distance from the treatment site d. We notice that the

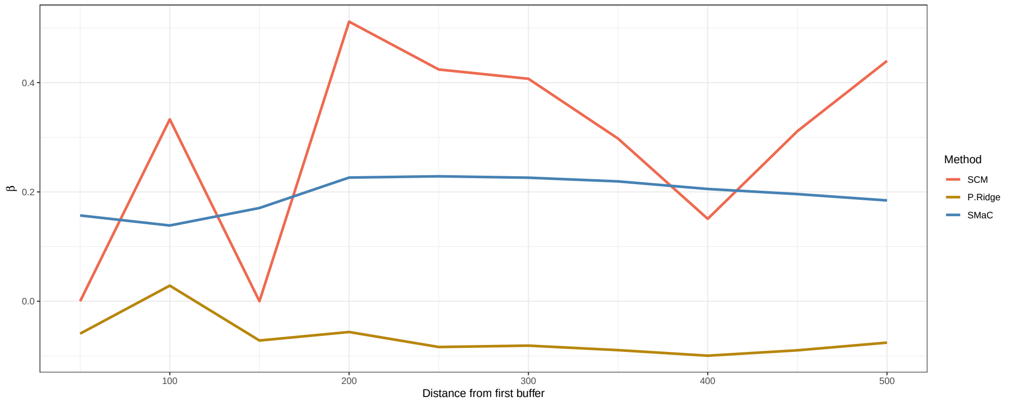


Figure 3.4: Values of an example β_i at increasing distances from the treatment site

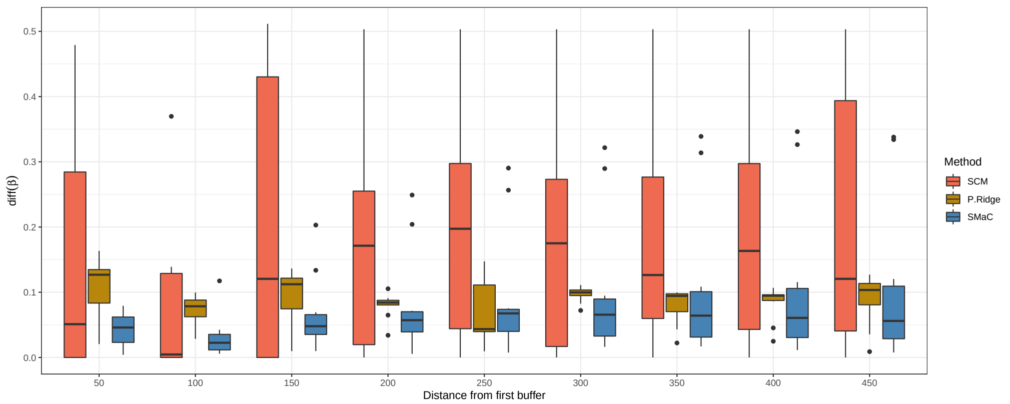


Figure 3.5: Boxplots of the difference $|\beta_i^d - \beta_i^1|$ for $i \in \{2, \dots, N\}$

value of SCM coefficient is varying quite abruptly, even if there is a small increase in distances across units, and one could expect similar values. Pooled ridge regression coefficient instead seems to be all shrunk toward zero. On the converse, SMaC coefficients show similar values at different distances, evidencing a smoothed transition between distance d_h and d_{h+1}

Next, we investigated how different estimated SCM coefficients are for areas of smaller or larger distances from the treated locations. Specifically, if β^1 are the coefficients for $d = 50$, and β^d are the coefficients for any value d, Figure 3.5 reports boxplots of $|\beta^d - \beta^1|$ as a function of $d - 50$. We notice that SCM coefficients at distance d are more different from corresponding coefficients for $d = 50$ as d increases.

In a spatial setting, we expect that abrupt variations of regression coefficients such as the ones seen for SCM in 3.4 are unlikely. One would expect that an area does not change much when the distance from the treated locations increases slightly. Therefore, the weighted combination of controls that adequately represent the treated area is expected to be similar for similar distances d. The results in 3.5 support this claim since SMaC coefficients are more similar when the treated areas of interest correspond to areas of similar distance from the treated location. The proposed SMaC estimator explicitly takes advantage of this structure, and returns smoothed transitions across distances (3.4) and smaller differences of coefficients across distances (3.5).

3.5 Simulation Study

We conduct simulation analyses to study the performances of the proposed estimator, in terms of bias and efficiency of the point estimate with respect to the reference estimators for causal effects with panel data. We focus on the comparison with the SCM and the ridge vertical regression method, as in Doudchenko and Imbens (2016). Other estimators were possible (Augmented Synthetic Control from Ben-Michael et al. (2021), OLS, Bayesian Regression, GLS) but we are focusing on the most commonly used estimators in our context.

3.5.1 Design

We simulate the time series outcome $Y_{i,t}^h(0)$, for the areas A_i^h, with H treated units if $i = 1$ and spatially correlated, and $(N-1)\times H$ control units, uncorrelated with the treated and among themselves. We consider three possible data-generating processes (DGP): one generated from the Homogeneous Poission Point process (HPP), one arising from a linear additive model, and the last generated from N Gaussian processes.

- The first data generating process is a Homogeneous Poisson Point process (HPP hereinafter) simulating the location of stores within some area $\mathbf{\Omega}$. As in our empirical application, we split the total space $\mathbf{\Omega}$ into a treated space $\mathbf{\Omega}^1$ and a control space $\mathbf{\Omega}^0$, and create time-fixed and non-overlapping buffers
$\mathbf{A} = \{A_1^1, \dots A_1^H, \dots, A_i^h, \dots, A_N^H\}$ around some random location in $\omega_i \in \mathbf{\Omega}$, at distance $d \in \mathbf{D}$. For each time $t \in T$ we simulate the point locations over a space $\mathbf{\Omega}$, and generate a collection of points as follows:

$$Y_{i,t}^h = \begin{cases} Y_{i,t-1}^h + Bi(Y_{i,t-1}^h, \delta_t) - Bi(Y_{i,t-1}^h, \eta_t) & \text{if } t \neq 1 \\ Y_{i,1}^h = Po(\kappa) \end{cases}$$

where κ is the baseline intensity of the HPP, and $\delta_t = \{\delta_t^1, \delta_t^0\}$ and $\eta_t = \{\eta_t^1, \eta_t^0\}$ are respectively the birth and death probability in each time t for treated and control units.

- The Second DGP assumes that the outcomes follow an additive linear model with normal components. For each unit of observation A_i^h, we consider the following model:

$$Y_{i,t}^h = \begin{cases} Y_{i,t-1}^h + \delta_t^1 - \eta_t^1 + \epsilon_{i,t}^h & \text{if } i = 1 \text{ and } t > 1 \\ Y_{i,t-1}^h + \delta_t^i - \eta_t^i + \epsilon_{i,t}^h & \text{if } i \neq 1 \text{ and } t > 1 \\ Y_{i,1}^h = \mathcal{U}(a, b) & \text{if } t = 1 \end{cases}$$

with δ_t^i and η_t^i normal terms representing the number of arising and closing firms during year t, sampled from a normal distribution, and $\epsilon_{i,t}$ an idiosyncratic error.

- Under the last DGP, that the outcome variable for the treated area Ω^1, $Y_{1,t}^h$ is generated from a Gaussian Process with hyperparameters (α_1, ρ_1), so that treated units are spatially correlated, with smoothing correlation across space. The outcomes for the control units outcomes $Y_{i,t}$ are generated from $(N-1) \times H$ independent Gaussian Processes with hyperparameters (α_i^h, ρ_i^h).

In all three DGP, once $Y_{i,t}$ is simulated, we round up to the unit the outcome. We consider three different lengths for the duration of the observation period. The first length defines a short period with 20 time points, the second length defines a medium period with 50 time points, and the third length defines a longer period with 100 time points. These choices are motivated by the fact that most applications fall within the range of 100 periods. The pre-treatment period is defined setting $t_0 = 3T/4$.

For each temporal scenario, we choose to simulate two alternative scenarios for the number of control units. Let $\#C$ denote the number of control units, we specify a scenario with $\#C = 0.5T$, and another with $\#C = 1.5T$ to test whether the performances increase or decrease across different specifications. In total, we study $3 \times 2 \times 3 = 18$ scenarios. We run simulations over 200 datasets for each scenario. For Bayesian inference we use RStan, simulating posterior distributions by running chains with 6000 iterations, and 3000 warmup iterations.

3.5.2 Simulations Results

We now focus on evaluating the point estimation properties of SMaC, comparing it with simulation results obtained with SCM and the pooled ridge

Table 3.1: Overall bias and MSE - Homogeneous Poisson Process DGP

Time periods	Method	# Controls=0.5*T				# Controls=1.5*T			
		Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2	Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2
20	SCM	6.944	4.628	24.330	14.056	7.173	4.371	25.089	13.326
20	P.Ridge	5.686	3.227	19.979	9.238	5.335	3.041	18.950	8.777
20	SMaC	5.438	3.195	19.291	9.291	5.538	3.145	19.501	9.277
50	SCM	6.504	4.187	23.000	12.715	6.654	4.265	23.491	13.335
50	P.Ridge	4.682	2.557	17.195	7.608	4.568	2.506	16.917	7.742
50	SMaC	4.887	2.829	18.151	8.590	4.790	2.847	17.973	8.863
100	SCM	5.983	3.857	21.996	11.866	5.612	3.626	20.416	11.052
100	P.Ridge	4.596	2.756	17.630	8.612	5.333	3.660	20,127	10.934
100	SMaC	5.054	3.160	20.083	9.774	5.308	3.670	19.860	11.045

Table 3.2: Overall bias and MSE - Additive Process DGP

Time periods	Method	# Controls=0.5*T				# Controls=1.5*T			
		Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2	Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2
20	1	37.901	23.749	144.922	61.646	36.992	25.886	143.059	72.505
20	2	31.877	19.904	136.014	62.374	40.131	31.944	178.187	101.514
20	3	31.145	20.549	116.839	58.428	31.758	22.602	124.188	67.903
50	1	56.134	37.250	208.178	102.544	51.222	34.647	192.251	97.440
50	2	43.186	31.145	194.738	88.078	51.355	38.514	252.797	122.288
50	3	44.630	31.683	173.413	90.570	41.763	29.185	181.384	86.177
100	1	73.489	55.002	271.152	158.010	69.745	53.765	267.098	144.836
100	2	60.296	44.211	265.843	131.965	65.480	45.003	272.742	136.883
100	3	66.163	49.033	251.960	143.905	63.211	43.457	242.994	134.179

vertical regression. In particular, we evaluate the performance of the alternative inferential procedures using the mean intertemporal bias:

$$\theta^1 = \frac{1}{T - t_0} \sum_{h=1}^{H} \sum_{t=t_0}^{T} Y_{1,t}^h(0) - \widehat{Y}_{1,t}^h(0)$$

and the mean intertemporal Mean Square error:

$$\theta^2 = \frac{1}{T - t_0} \sum_{h=1}^{H} \sum_{t=t_0}^{T} (Y_{1,t}^h(0) - \widehat{Y}_{1,t}^h(0))^2$$

Table 3.3: Overall bias and MSE - Gaussian Process DGP

Time periods	Method	# Controls=0.5*T				# Controls=1.5*T			
		Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2	Mean θ_1	st.Dev. θ_1	Mean θ_2	st.Dev. θ_2
20	SCM	1.140	0.552	6.608	1.829	1.189	0.568	6.669	2.036
20	P.Ridge	1.203	0.863	6.817	2.599	1.234	0.987	6.841	3.011
20	SMaC	0.856	0.412	5.971	1.781	0.800	0.386	5.781	1.847
50	SCM	0.892	0.418	6.571	1.194	0.943	0.424	6.925	1.512
50	P.Ridge	0.896	0.753	6.693	1.982	0.976	0.656	7.071	1.990
50	SMaC	0.549	0.247	6.040	1.139	0.619	0.315	6.335	1.507
100	SCM	0.761	0.338	6.691	0.952	0.764	0.361	6.934	1.012
100	P.Ridge	0.724	0.383	6.664	1.084	0.947	0.476	7.234	1.154
100	SMaC	0.455	0.203	6.253	0.852	0.536	0.271	6.550	0.933

The simulation results in table 3.3 suggest that under a Gaussian process DGP our proposed method has better performances than the SCM estimator and the vertical regression estimator with a ridge penalization. In particular, we can see that the mean bias for the treated units is around 30% lower than the other estimators. Similar results are derived also for the mean MSE. We notice also a lower standard deviation, using SMaC, for both θ_1 and θ_2 under all the scenarios considered.

Under a Homogeneous Poisson Process DGP, we notice in table 3.1 that SMaC mean bias and mean MSE are lower when T=20 and when we have $\#C \leq T$. This result sounds reasonable, when the longitudinal information is scarce and cross-sectional information is not that important, SMaC exploits the spatial structure of the treated units to obtain better performances than the separated counterparts. With longer time series, both the SCM and the pooled ridge estimators can exploit the temporal pattern to obtain estimates with a slightly lower bias and MSE with respect to SMaC.

Under the linear additive DGP, in table 3.2 we see that in terms of mean bias, SMaC is performing better than the other estimator when $\#C \leq T$ and T=20. Similar results hold also when we consider $\#C \geq T$, and results are valid for all the time scenarios considered. Boxplots of the summary statistics for θ_1 and θ_2 are reported in figure 7.1, 7.2 and figure 7.3 in the appendix 3.

In general, under the three DGP considered, we notice that SMaC has good performances in terms of bias and MSE, especially when the number of time periods considered is not so large. In fact, working with few time periods and spatially correlated units are common situations in policy evaluation, and this makes particularly attractive the estimator we propose.

3.6 Estimating the effect of the Florentine tramway construction

Figure 3.6 and table 3.4 show the results we obtained applied the proposed methods to the Florentine tramway study. We evaluate the treatment effect arising from the construction of the tramway, for areas surrounding the tramway stops. We consider the treatment effect as the average of the posterior distribution of $\Delta^{d}_{i,t}$, estimated according to equation 3.3. Notice that for each unit the outcome $Y^{h}_{i,t}$ is centered around its own pre-treatment mean and divided by the pre-treatment standard deviation. Table 3.4 shows the posterior mean and the posterior 90% credible intervals based on quantiles of the simulated posterior distribution. Our results show that the tramway has provoked generally an increase in the commercial vitality of the area considered. These results are particularly significant for the areas closer to the tramway stops, as we find significant average treatment effects for the areas within 50 and 100 meters of the treatment sites. The positive, yet non-statistically significant effects are present for the outer areas, from 150 to 400 meters away from the tramway stops.

We also report some goodness of fit measures, such as the mean bias of the pre-treatment effect and the root mean square prediction error. In both cases, we obtain a good fit from the model. These results are confirmed by the visual inspection of pre-treatment trends in figure 3.6, from which we can also see that the credibility intervals in every pre-treatment time period for each treated unit comprise 0.

By inspecting the time trends of the causal effects we can derive additional insights into the effect of construction worksites, and the subsequent effect of the tramway.

Worksites have not extensively damaged the commercial environment of the treated area. We can note a significant and negative effect for the area within 100 meters during the period 2006-2010. That time span was the construction period of the tramway, and thus we could expect worse outcomes for areas close to the construction site. However, the number of stores steadily recovered in 2010, the inauguration year, and the overall effect, even for this particularly affected area, is still positive. For this purpose, it is worth noting that in the closer area to the treatment site, the positive effect is present since the start of the construction period, some retailers anticipate their competitors by locating the shops in the most served areas even before the start of tramway operations. Similar results are present also in the first chapter of the dissertation instead considers a linear buffer of 50 meters around the tramway line.

The effect on the outer bands ($d \geq 200$ meters) is similar to the ones found for inner areas. In particular, we notice that worksites have not affected the commercial environment of the outer areas, while the tramway has improved the accessibility of the area, leading to an increase in the number of shops present. The estimated causal effect has a growing tendency, especially for the outer areas, that exhibit statistically significant effects in the last observational periods.

Except for the effect on the second and third buffers, we cannot notice any relevant effect coming from the construction site on the commercial vitality of the surrounding areas.

Table 3.4: ATE by areas within d distance from the tramway stop, mean value , 90% Credibility intervals, and pre-treatment mean Bias and RMSPE

	ATE	5%	95%	Mean Bias	RMSPE
50	3.059	1.335	4.790	0.000	0.196
100	2.022	0.266	3.804	0.000	0.186
150	5.313	-2.374	13.063	0.000	0.127
200	8.042	-11.423	25.651	0.000	0.080
250	7.951	-14.828	30.905	0.000	0.080
300	10.414	-10.643	29.602	0.000	0.108
350	12.235	-4.837	28.173	0.000	0.150
400	10.263	-5.392	25.516	0.000	0.165

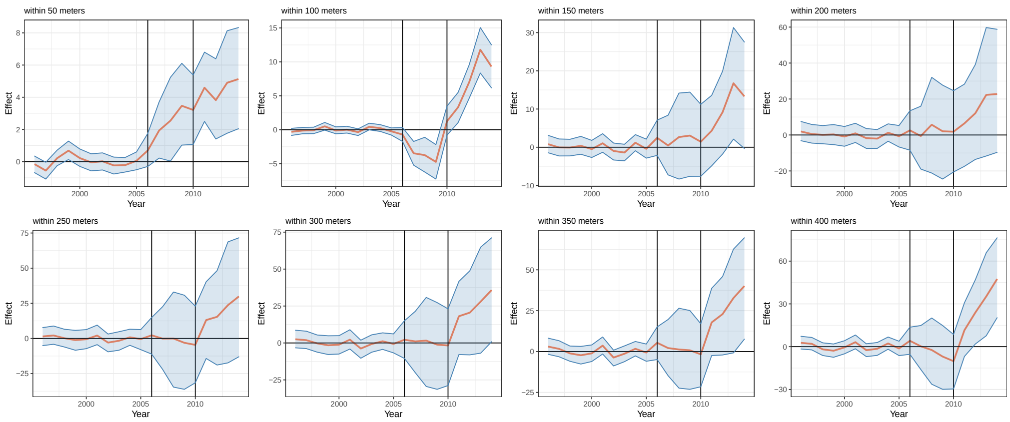


Figure 3.6: Treatment effect for areas within d meters from a tramway stop, Red line: Treatment effect, Blue area: 90% Credibility interval - First vertical line: tramway worksite starts (2006) - Second vertical line: tramway operational (2010)

3.7 Concluding Remarks

In this work, we propose a novel method for causal effects estimation in a spatial setting. Working on the previous literature of SCM (Abadie et al., 2010), vertical regression (Doudchenko and Imbens, 2016) and MCM (Athey and Imbens, 2021), we propose to exploit the properties of a Bayesian regression model, using a Gaussian Process prior for the regression coefficients we use as weights of the control units to impute the missing control potential outcomes for the treated units.

The core idea of the proposed method is the following: in context with spatially correlated units, it is reasonable to assume that the linear combination we use to impute the potential outcomes of a unit in a given location d, will be similar to the linear combination we exploit for imputing the missing control potential outcomes for a unit in the treated area located at a distance $d+\varepsilon$ from the treatment site. In settings with spatially correlated units, we could expect that inferential procedures that account for the spatial dimension could be more efficient than procedures that do not account for it. For this purpose, we simulate three different data-generating processes for spatial settings and evaluate the model performances in terms of bias and MSE of the predicted values. We find out that our SMaC grants at least equal results to the commonly used estimators in these contexts, the SCM and the Ridge estimator for vertical regression, with better results than counterparts when the number of time periods available is small. This could be a particularly relevant feature, as the lack of long time series could be an issue that affects the use of SCM and vertical regression in many policy evaluation problems.

We illustrate the results from this model by investigating the treatment effects on the commercial vitality arising from the construction of the first line of the Florentine tramway. The results from the motivating applications underline the potential of our estimator, allowing us to consider the diffusion of treatment effects around some treatment sites, the tramway stops. We find an overall positive effect for the treated area, and statistically significant for the inner areas, located within 100 of a tramway stop. Outer areas also exhibit positive results, yet non statistically significant, but with a promising trend in the last observation periods.

Worksites have not generally affected the commercial environment, with some negative effects during the construction period for the buffers between 50 and 100 meters from the tramway stop, which steadily recovered once the tramway starts working. Notably, similar results have been found also in the first work of this dissertation. In general, our results suggest that investments in urban light rail systems can be a viable solution that grants improvements in the commercial vitality of served areas.

Further research on the topic is needed. Possible extensions could embrace both methodological and applied extensions. In particular, it could deserve attention to the comparison between SMaC and the generalized ridge estimators, the extensions to multiple sources of treatments, bipartite designs, or even the application of SMaC to networked data.

Chapter 4

Bayesian longitudinal principal stratification

4.1 Introduction

Start-up businesses can contribute significantly to urban regeneration and regional development through the creation of innovative and creative environments, see for instance Román et al. (2013). This is supported by research indicating that start-ups are a key driver of productivity growth, and economic renewal(Haltiwanger et al., 2013, Dumont et al., 2016). Moreover, start-ups are also a viable solution to tackle unemployment.

In recent years, policymakers have turned their attention to promoting self-employment by directly supporting the creation of new businesses. These policies aim to create and maintain a favourable environment for small businesses, including those that have been successful in the past but may be looking to reinvent themselves. By lending their support to promising entrepreneurial endeavours, policymakers aim to decrease unemployment by providing self-employment opportunities for aspiring business owners and potentially, additional hiring opportunities for the unemployed. However, it is not uncommon for start-ups to face challenges and fail during their early stages, e.g., due to economic conditions and the inexperience of the founders.

In such a context, the evaluation of public programs targeting these businesses can be severely harmed by post-treatment complications. Outcomes can be censored by death, in the sense that they are neither observed nor defined for units who die. A first solution for these complications spans from the work by Heckman (1976) and Heckman (1979), who sees this problem as a sample selection situation and provides a structural equation modelling approach to deal with it. Some authors propose Instrumental Variables (IV) approaches (Angrist and Krueger, 1999) or conditional Difference in Differences (DiD), as Heckman et al. (1998). Other recent methods rely on imputation approaches (Leete et al., 2019) or DiD methods (Sant'Anna, 2016).

In this work, we choose to deal with the problem of truncation by death by using the principal stratification approach, first introduced by Frangakis and Rubin (2002), see also Zhang et al. (2008) and Mealli and Mattei (2012). Within the principal stratification approach, the focus is on causal effects on latent sub-populations of units (the principal strata). The use of principal stratification to evaluate causal effects under censoring complications is quite widespread, see for instance Zhang and Rubin (2003), Chiba and VanderWeele (2011), Mattei and Mealli (2011), Mealli and Mattei (2012), Frumento et al. (2012).

We will show the proposed approach to study the causal effects of a policy promoting start-ups on firms' survival and hiring policy. Building on a recent strand of the literature (Bia et al. (2020)), we propose an extended framework for the analysis of longitudinal studies, where units can be censored at different time points, and the main endpoints are observed and well-defined only up to the censoring time.

This work aims to contribute to the existent literature about the evaluation of public-policy programs by expanding to multiple time periods the framework developed by Bia et al. (2020), and possibly opening the strand to a multiple-times principal stratification framework. Moreover, we wish to contribute to the thematic literature about the evaluation of public support in self-employment programs and start-up development, which is a relevant discussion topic in the current literature strand.

The work proceeds as follows: section 4.2 illustrates our motivating application, its main complication and the relevant literature on the applied field. Section 4.3 shows our data sources and some descriptive statistics. Section 4.4 illustrated the methodological environment and the empirical strategy to estimate causal effects. Section 4.5 present the main findings of the work. Section 4.6 concludes.

4.2 The subsidized start-up puzzle

The evaluation of programs supporting start-ups is a debated theme in economic policy literature. The idea behind financial aid for start-ups is articulated into several levels of analysis. At first glance, there is some consensus on the positive effect of start-up businesses in fostering employment. For reference, Kane (2010) and Decker et al. (2014) analyze the US labour market finding out the positive role of start-ups in job creation, moreover, they find that older firms are net job destroyers. In these pieces of study, start-ups work as re-distributors of resources from low-productivity sectors to more profitable ones. Moreover, Aghion et al. (2009) stresses the pressure posed by new-established firms on incumbents, accelerating the natural renewal process of entrepreneurship. Job creation ability of start-ups has been inquired by Kuschel et al. (2018) focuses on the role of women in job creation, finding no relevant differences with a male counterpart. Choi et al. (2020) study the role of Korean tech start-ups in terms of job creation and job quality, finding out that innovative tech projects grant desirable results.

On the other hand, the results we find in the start-up literature are not fully generalized. Several works find dependence between the starting blocks for each firm and its future development. See for example Brown et al. (2019), that shed light on founder characteristics for a successful project.

The financing ability of firms could be a serious issue for new entrepreneurial projects. Freshly established firms are usually disadvantaged in accessing credit because of the little relational capital of the firm owner, and the asymmetric information about the project potential, as pointed out by Peneder (2008). From a financial point of view, start-ups do not usually rely on past liquidity provisions, and they usually lack robust guarantees for their loan application, see for instance Colombo and Grilli (2007) and Nigam et al. (2020). Thus following, the credit market can present significant barriers to the development and growth of new businesses. In these cases, public sector intervention through market corrections can play a crucial role in addressing these deficiencies and promoting an environment that is more conducive to the emergence of fresh-starting projects. This can be done by various means like creating a favourable regulatory framework, providing access to financing, investing in infrastructure and education or directly investing or supporting those projects or ventures. The decision to provide public support to start-up businesses is a nuanced one as start-ups can be fragile and uncertain. Several studies focus on the impact of public policy on the survival of start-ups. Battistin et al. (2001), studies the relationship between public subsidies and the start-ups' survival, Boyer and Blazy (2014) studies the drivers for the economic durability of start-ups, finding a relationship between the outcome and the age, gender, financial support and network of the firm owner. Duhautois et al. (2015) and Alonso-Nuez and Galve-Górriz (2012) shed light on the effect of public support on start-ups. Recently, Mariani et al. (2019) used a causal framework to examine the survival of subsidized firms using data from the same program as the one analyzed in this study. They find ambiguous results: public support helps young men and women to exit from unemployment and create additional jobs in the short term but at the price of investing resources in projects with low potential.

From a policymaker's perspective, there are some additional topics that deserve to be mentioned. Start-up subsidies can be also seen as a viable policy to promote entrepreneurship among the disadvantaged classes of the population, promoting self-employment (Caliendo and Künn, 2014). Lack of experience, self-confidence and contacts could harm even those potential entrepreneurs. Public interventions may help to level the playing field and create a more equitable environment for new ventures to compete and succeed in the marketplace, ultimately fostering a more robust and dynamic economy. On the one hand, the social purpose of these programs is pretty straightforward. First, they can reduce the barriers to getting access to the credit market, reducing the gap between youth and female entrepreneurs and the other firm owners, as pointed out by Caliendo (2016). Second, self-employment can be seen as an active labour policy, with a direct effect on the unemployment level of more vulnerable social groups. Additional relapses can encompass job formation, due to the hiring in freshly established firms and innovation dissemination. On the other hand, there is little consensus on the positive effects of such policies. The most relevant issue regards the personal abilities of the firm owner and the potential of the start-up projects. In fact, funding people that have not shown particular entrepreneurship ability could result in a net loss of public funding, see Shane (2009). Lukeš et al. (2019) find out that start-up incubators should not be financed by public actors, as they are not granting results in terms of job creation and business growth. Interestingly, Caliendo et al. (2020), catch the double nature of start-up subsidies: on the one hand subsidization has a positive impact as an active labour policy in reducing unemployment, but on the other hand, subsidized firms cannot reach the performances of the untreated counterparts. In the short run, this is explained by the different starting abilities, while in the long run is explained by the different development paths. Audretsch et al. (2020) confront 38 different start-up projects finding out that there are few constant results across different environments, and policy outcomes will also depend on the economic environment. Additionally, Koski and Pajarinen (2013) found significant differences between incumbent firms and start-ups and argued that financial support does not appear to be a critical factor in business development. Other scholars find that subsidization could help as an active labour policy where conformal approaches fail to generate employment, and therefore they should be taken into account in the policymaker toolkit as Pfeiffer and Reize (2000), Caliendo and Kritikos (2010), or Duhautois et al. (2015).

Few causal studies focused on the results of such policies. This turns out to be a relevant topic as the positive or negative results found from descriptive analysis often are not confirmed by causal studies Pfeiffer and Reize (2000). Most of the causal studies focusing on the impact of subsidies on start-ups focus on the causal effects on the survival probability of firms at some endpoints. Mealli and Pagni (2001) and Mariani et al. (2019) investigate the effects of this program on the survival probability of Tuscan start-ups, finding longer survival probability for subsidized firms but at the price of supporting low-productivity projects. A recent work from Manaresi et al. (2021) analyze the results of the Italian "Start-up Act" on several business dimensions and found promising results in terms of market failure corrections and improved survival of start-ups.

4.3 Doing business in Tuscany

The data used in this work are referred to the program implemented in Tuscany, called "Fare Impresa" (Doing Business), which has the goal of fostering entrepreneurship between young and female business in the very first period of activity through bank loans assisted by public guarantees.

The eligibility criteria for this public policy program are based on the age of the firms. Both newly established firms (no older than two years, or starting activity within six months) and established firms seeking expansion opportunities (with less than 5 years of activity) are eligible to participate. In addition to these requirements, applicants have to meet the following criteria: they need to be female of any age, or male aged 18-40.

During the period 2011-2015, firms participating in the program were eligible to receive a public-assisted guarantee to help them obtain bank loans to start or grow their businesses. These loans could last up to ten years, with the guarantee covering up to 80% of the requested amount. A regional financial intermediary oversaw the loan request process and ensured that it was conducted correctly. Additionally, once approved for the program, firms were eligible to receive a reduced interest rate on their loans.

In total, 1837 firms received the guarantee backed by public authorities, among these 1563 projects were funded by banks, with 274 credit rejections. We consider a treatment $W_i \in \{0, 1\}$, the granting of a loan from a bank. Thus, we consider as treated ($W_i = 1$ the 1563 units which have received the bank credit to promote their projects, and as control units ($W_i = 0$) those which have not received the loan. Moreover, 185 firms ceased the activity by the end of the observation period. The first chart of Table 4.3 depicts the count of firms that were censored, active, and closing in each time period. The second chart illustrates the number of firms that hired employees in each period. Table 8.1 in the appendix 4 reports the mean values for hiring decisions within the groups defined by the treatment and censoring status.

We obtained information on the main characteristics of the firms in the study from the records of a regional financial intermediary (Fidi Toscana).

Table 4.1: First pane: Proportion of ceased, active and closing firms in each year. - Second pane: Proportion of censored, hiring and no-hiring firms in each year.

Status	**t=1**		**t=2**		**t=3**	
	$W_i = 0$	$W_i = 1$	$W_i = 0$	$W_i = 1$	$W_i = 0$	$W_i = 1$
Ceased	NA	NA	0.193	0.022	0.270	0.071
Active	0.807	0.978	0.730	0.929	0.701	0.877
Closing	0.193	0.022	0.077	0.049	0.029	0.052

Status	**t=1**		**t=2**		**t=3**	
	$W_i = 0$	$W_i = 1$	$W_i = 0$	$W_i = 1$	$W_i = 0$	$W_i = 1$
Censored	0.193	0.022	0.270	0.071	0.299	0.122
Hiring	0.628	0.598	0.591	0.655	0.599	0.707
No-Hiring	0.179	0.380	0.139	0.274	0.102	0.171

These records included information on the start date for the firm's activity, the business sector, the location of the investment, and demographic characteristics of the owner such as age and gender. We also collected data on the end date for each firm's activity from the Chamber of Commerce archives. In addition, we gathered hiring information from the Tuscan Job Information system, which tracks changes in hiring and resignation. This system also provides information on the type of contract and its duration.

Table 4.2 reports the proportion of each covariate used in the study. As all the variables that we considered in the study were binary variables, we are reporting their proportion in the sample as well as the number of firms as a way to present descriptive statistics.

Around three forty of the firms in the study were owned by young individuals (76.2%), while female owners made up 57.1%. The majority of firms were new start-ups (92.1%), and 60.6% were owned by a single individual. Around half of the loans were granted by a local bank, and only 19.8% of the projects were located in urban areas. This suggests that the program was particularly focused on rural areas in Tuscan, where local banks are more common. Additionally, 25.1% of the firms have already hired workers at the beginning of the observation period.

The firms participating in the program represent a range of different industries, with 11.1% of the sample involved in manufacturing, 32.7% engaged in retail, 27.6% in hospitality, and 12.2% in beauty and hairdressing. The remaining portion of the sample represents other types of economic activities. All of the eligible firms participating in the study applied for

Table 4.2: Descriptive statistics on covariates in the study

Covariates	**Mean**			**Num. of firms**		
	$W_i = 0$	$W_i = 1$	Overall	$W_i = 0$	$W_i = 1$	Overall
Young	0.770	0.760	0.765	211	1188	1399
Female	0.540	0.576	0.571	148	901	1049
Has employed	0.226	0.257	0.251	62	402	464
Start-up	0.901	0.925	0.921	247	1445	1692
Sole-ownership	0.657	0.597	0.606	180	933	1113
Manufacturing	0.131	0.107	0.111	36	168	205
Retail	0.296	0.332	0.327	81	519	600
Hospitality	0.226	0.285	0.276	62	445	507
Service to person	0.080	0.130	0.122	22	203	225
Local bank	0.401	0.553	0531	110	865	975
Urban location	0.197	0.198	0.198	54	310	364

a bank loan backed by a public guarantee, with around the 85% which receives bank loans. We do not have data on loan rejections, but we have data on the firm's and owner's characteristics, as well as the local or national business area of the bank. It is possible that loan denials will be related to some specific condition of the firm, such as the young age of the owner, or the type of ownership. It is also possible that loan denials may be connected to the firm owner's relationship capital. Local banks may be more inclined to fund local projects, and thus small business projects can be financed. Additionally, national banks may have stricter requirements for loan approval, which can result in the rejection of applications that do not meet their standards. Covariate imbalance between treated and control units is reported in figure 8.1, and we can notice the similarity between funded and non-funded firms.

4.4 Methodology

4.4.1 Notation and Setting

We consider a set of N firms indexed by $i = 1, \dots, N$, and observed in three post-treatment periods, namely $t \in \{1, 2, 3\}$, post-treatment periods which correspond to years 2012, 2013 and 2014.

For each unit i, we observe the main outcome $Y_{i,t}$ which is a binary indicator for the hiring decision for start-up i in each post-treatment period t, with 0 if no contracts have been formed during the year t, and 1 otherwise. $Y_{i,t}$ is truncated by death, in the sense that is neither observed nor defined for firms that cease their activity before year t. Let $S_{i,t}$ denote the survival status of firm i, with $S_{i,t} = 1$ if firm i does not cease activity during the year t and 0 if it interrupts the activity during the year t. Therefore, $Y_{i,t} = *$ when $S_{i,t} = 0$, where $*$ is a non-real value. Please note that $S_{i,t}$ and $Y_{i,t}$ are collected together, with $S_{i,t} \in \{0,1\}$ and $Y_{i,t} \in \{0,1\} \cup \{*\}$.

Moreover, we observe for each unit i a vector of time-invariant and firm-specific covariates $\boldsymbol{X}_i$. Let $\mathbf{W}$, $\mathbf{Y}_t$, $\mathbf{S}_t$, $t \in (1,2,3)$, be N-dimensional vectors with ith entries equal to W_i, $Y_{i,t}$ and $S_{i,t}$, respectively. Let $\mathbf{X}$ be a $N \times K$ matrix of pre-treatment variables, with the i-th row equal to $\mathbf{X}_i$.

Causal estimands are defined under a potential outcome approach (Rubin, 1974), so we have to specify potential outcomes under treatment and control for each post-treatment variable. In particular, we assume the validity of SUTVA, (Rubin, 1980):

Assumption 5. ***SUTVA***

- *No hidden version of the treatment*
- *No interference between units*

SUTVA implies a single version of the treatment and the absence of interference between units. We can consider the treatment as homogeneous across the different firms, as all the businesses are similar in size, and therefore we could expect similar entrepreneurial projects with similar financed amounts.

We can also safely assume no interference between units, considering the fact that treated businesses are scattered through Tuscany, and thus we could expect low interaction between them. Moreover, it is unlikely that the granting of a loan to firm 1 has a sizeable effect on the hiring decisions for firm 2.

For each start-up i and time t we define the following couple of potential outcomes for the hiring decision:

$$\begin{cases} Y_{i,t}(1) \equiv Y_{i,t}(W_i = 1) & \text{If firm i is assigned to treatment} \\ Y_{i,t}(0) \equiv Y_{i,t}(W_i = 0) & \text{If firm i is assigned to control} \end{cases} \tag{4.1}$$

Similarly, we define the potential outcomes for the firm survival:

$$\begin{cases} S_{i,t}(1) \equiv S_{i,t}(W_i = 1) & \text{If firm i is assigned to treatment} \\ S_{i,t}(0) \equiv S_{i,t}(W_i = 0) & \text{If firm i is assigned to control} \end{cases} \tag{4.2}$$

Let $\mathbf{Y}_t = (Y_{i,t}(0), Y_{i,t}(1))$ and $\mathbf{S}_t = (S_{i,t}(0), S_{i,t}(1))$ be the $N \times 2$ matrices of potential outcomes.

4.4.2 Principal Stratification Approach

The hiring decision for each firm can be seen as a consequence of bank loans granted to the firm, but also from the overall performances of the firm during the first years of activity. In this situation, our treatment evaluation could in be harmed by the non-ignorable missingness of the potential outcome $Y_{i,t}$, due to the censoring by death in the multiple post-treatment periods. To face this problem, we use the principal stratification approach, firstly proposed by Frangakis and Rubin (2002). Applications of principal stratification embrace typical post-treatment complications, as non-compliance (Forastiere et al., 2016), treatment switching (Mattei et al., 2020) or censoring by death (Mattei and Mealli, 2007).

Principal stratification allows us to classify units into latent groups, the principal strata defined by the joint potential outcomes of the intermediate variable, under each of the treatments given in the study, $S_{i,t}(W_i) = (S_{i,t}(0), S_{i,t}(1))$.

In our study, at each time point t, t=1,2,3, we classify units into four latent groups with respect to their survival status under treatment and control, as shown in Table 4.3.

$S_{it}(0)$	$S_{it}(1)$	**Definition**
0	0	Never survivor (NS)
1	0	Defiant survivor (DS)
0	1	Compliant survivor (CS)
1	1	Always survivor (AS)

Table 4.3: Indicators for the membership of unit i at time t

Let G_{it} be the indicator for the principle strata membership of unit i at time t, $G_{it} \in NS, DS, CS, AS$. *Never survivors* are those firms that would cease their activity in time t, irrespective of their treatment assignment. *Always survivors* businesses would continue the activity both under treatment and under control. *Compliant survivors* are businesses that would continue the activity, *if* they are assigned to the treatment, but would have ceased the activity if assigned to control. Finally, *defiant survivors* are the firms that would cease the activity *if* they received the treatment but would have continued the activity if assigned to control.

Considering the longitudinal framework we are dealing with, we should classify our observations according to the repeated membership to each of the four principal strata in table 4.3. In particular, we observe three post-treatment variables, the survival of the businesses in each post-treatment period, and therefore we classify the principal strata membership on the joint value of these variables in each post-treatment time period. Following Bia et al. (2020), we define the longitudinal principal strata membership for each unit i as follows:

$$(S_{i,1}(0), S_{i,1}(1), S_{i,2}(0), S_{i,2}(1), S_{i,3}(0), S_{i,3}(1))$$

Let

$$\mathbf{G}_i = (G_{i,1}, G_{i,2}, G_{i,3})$$

$$\mathbf{G}_i \in (AS, DS, CS, NS) \times (AS, DS, CS, NS) \times (AS, DS, CS, NS)$$

be the indicator for the longitudinal principal strata membership.

Therefore, according to this cross-classification, we have $4 \times 4 \times 4 = 64$ latent strata.

4.4.3 Causal Effects

In this section, we introduce the causal estimand we are interested in. Causal effects with censored data could be challenging, as the potential outcomes we want to estimate are not defined under treatment and under control for all the units. For example, according to the cross-classification we give in table 4.3, we notice that $Y_{i,t}(0) = *$ for compliant survivors. Similarly, $Y_{i,t}(1)$ will be not defined for defiant survivors, which continue activity only under control assignment. Finally, both the potential outcomes for never survivors units will be not defined, and therefore no comparison are possible. Rubin et al. (2006) proposes to focus on the survivor average treatment effects, specifically at each time point we are interested in the survival average causal effect:

$$SATE_t(Y) = \mathbf{E}[(Y_{i,t}(1) - Y_{i,t}(0)|G_{i,t} = AS)] \qquad t \in \{1, 2, 3\} \tag{4.3}$$

In our work, we are also interested in longitudinal SATE:

$$SATE_2(Y) = \mathbf{E}[(Y_{i,2}(1) - Y_{i,2}(0)|G_{i,2} = AS, G_{i,1} = AS)] \tag{4.4}$$

$$SATE_3(Y) = \mathbf{E}[(Y_{i,3}(1) - Y_{i,3}(0)|G_{i,3} = AS, G_{i,2} = AS, G_{i,1} = AS)] \tag{4.5}$$

$SATE_t(Y)$ $t \in 2, 3$ is the average causal effect for firms that would survive under both treatment and control assignment at least up to time t.

These estimands are particularly interesting as we can derive useful insights by observing the evolution of the causal effect over different times. In particular, comparing SATE effect through time can help us in understanding the short and long-term effects of the policy. Specifically, in our application hiring decision can be boosted during the first time by the bank loan concession, while long-term impact will remain more difficult to predict

4.4.4 Assumptions

The identification of causal effects introduced above can be particularly troublesome, as we deal with observational data. Unfortunately, we do not generally observe the principal stratum membership for any unit, as we usually cannot observe both the potential outcomes in the post-treatment. We can only observe the realized outcome for both the intermediate variable and the main outcome under the actual treatment, while the other potential outcomes are missing.

For inference, we need to posit the treatment assignment mechanism. In particular, in this work, we have to deal with observational data describing firms that voluntarily participate in a public policy program. In such a context, we invoke the strong ignorability assumption (Rosenbaum and Rubin, 1983).

Assumption 6. ***Strong Ignorability of treatment assignment***

- *Unconfoundness:* $Pr(W_i|S_{i,1}, S_{i,2}, S_{i,3}, Y_{i,1}, Y_{i,2}, Y_{i,3}, \mathbf{X}_i) = Pr(W_i|\mathbf{X}_i)$
- *Overlap:* $0 < P(W_i = 1|\mathbf{X}_i) < 1$

Strong ignorability consists of two parts. The first part regards unconfoundness assumption, which is not testable in the majority of cases. Even if this is the case, we are not particularly worried about making this assumption, as there are no structural differences between the treated units and the control ones. Both of them were willing to receive the treatment, but some of their applications were rejected for unknown causes. Overlap means that, within the cells defined by the $\mathbf{X}$, there are present both treated and control units. With these two sub-assumptions, within the cells defined by the set of covariates $\mathbf{X}_i$, we can consider the treatment as given at random.

Even under assumption 6 SATE are not identifiable, as different principal strata correspond to the same joint distribution (see Gustafson, 2010, and Ricciardi et al., 2020). In fact, we can observe directly only some groups, classified on the treatment assignment and their survival status at each time. These observed groups are mixtures of several latent strata, see for reference table 4.5.

To disentangle such latent strata, we should first simplify our environment by introducing some realistic yet useful assumptions. Even if, by construction, we define 64 principal strata, it is that some of these latent strata are unfeasible.

For instance, consider the units that will be classified as never surviving at time 1 ($G_{i,1} = NS$). In principle, all of these units could evolve as always survivors, compliant survivors or defiant survivors in the following times. We state that these transitions are unfeasible, as the never surviving firms will be censored at time 1. In particular, if $S_{i,t^*}(0) = 0$ then $S_{i,t}(0) = 0 \quad \forall t > t^*$, and similarly if $S_{i,t^*}(1) = 0$ then $S_{i,t}(1) = 0 \quad \forall t > t^*$.

Therefore, we invoke an assumption on the never surviving firms:

Assumption 7. ***Never survivor firms***

$$\begin{cases} \textit{If } G_{i,1} = NS \textit{ then } G_{i,2} = NS, G_{i,3} = NS \\ \textit{If } G_{i,2} = NS \textit{ then } G_{i,3} = NS \end{cases}$$

We cannot allow a unit to be censored under treatment and under control the first time, and then be active, for instance, under treatment in the following times.

In our study, the existence of defiant survivors firms that would cease the activity if received the bank loan but would survive if did not receive the loan seems to be implausible. In fact, we think that is improbable that a business closes in the following three years after receiving a bank loan, given that they start to refund the loan from the seventh year. Therefore we exclude this subgroup from the study with a *monotonicity* assumption.

Assumption 8. ***Monotonicity***

$$S_{i,t}(1) \geq S_{i,t}(0) \qquad \forall i, t$$

Lastly, we should consider that some transitions between strata are impossible by construction, a unit classified as compliant survival at time t would result censored under control assignment. Therefore it is impossible to classify this unit as always survivor at time $t + h$, because its potential outcome under control would be not definite. We refer to this consideration as *dominance*. Thus, $G_{i,t} = CS$ then $G_{i,t+h} \in \{CS, NS\} \qquad \forall h \in N^+$

Assumptions 7 and 8, imply that the number of possible principal strata reduces to ten. Table 4.4 reports the cross-classification of units according the potential outcome of the survival variable, and the longitudinal principal strata.

When we examine the observed compositions of the subgroups, we find that 95.5% of the treated units fall into the final subgroup in which units are not censored at the end of the study. In decreasing order, we also see units censored at the first time (2.44%), second time (1.46%), and third time (0.56%). In the control group, the composition is more varied, with uncensored units making up 47.8% of the observations. Censored firms become more prevalent as the time periods increase, with 13.2% censored at time 1, 18.9% censored at time 2, and 20.2% censored in the final time period.

Intermediate variable						**Long. Strata**
$S_{i,1}(0)$	$S_{i,1}(1)$	$S_{i,2}(0)$	$S_{i,2}(1)$	$S_{i,3}(0)$	$S_{i,3}(1)$	$\mathbf{G}_i$
0	0	0	0	0	0	NS.NS.NS
0	1	0	1	0	1	CS.CS.CS
0	1	0	1	0	1	CS.CS.NS
0	1	0	0	0	0	CS.NS.NS
1	1	1	1	1	1	AS.AS.AS
1	1	1	1	0	1	AS.AS.CS
1	1	1	1	0	0	AS.AS.NS
1	1	0	1	0	1	AS.CS.CS
1	1	0	1	0	0	AS.CS.NS
1	1	0	0	0	0	AS.NS.NS

Table 4.4: Longitudinal principal strata classification

4.4.5 Bayesian Inference

Our study aims to investigate the causal effects of various factors on contract formation within a sample over time. However, this sample is not homogenous and consists of latent subpopulations, or distinct groups, which may experience different causal effects. To examine these changes and effects over time, we are taking a longitudinal perspective. This presents a challenge because the number of latent subpopulations increases over the course of the study, resulting in more missing data. Under assumptions 1-4, we have reduced from 64 to 10 the number of principal strata. Although this simplification process, principal strata proportions and SATE remain not fully parametrically identifiable. To address these inferential challenges, we propose using a Bayesian approach to causal inference, introducing modelling assumptions.

The model-based Bayesian approach, introduced by Rubin (1978), is a comprehensive and versatile framework for analyzing complex data with missing information. It posits that all unknown quantities - including parameters and missing potential outcomes - can be treated as random variables with a joint posterior distribution given the observed data. In this way, we can make inferences about the causal effects of interest by examining the posterior distributions of the relevant parameters.

The general structure for conducting Bayesian causal inference with principal stratification was first outlined by Imbens and Rubin (1997). It was initially developed to address issues of all-or-none treatment noncompliance but has since been extended and applied in various contexts. Our work builds on previous research in this area, including contributions by Mattei and Mealli (2007), Jin and Rubin (2008), Baccini et al. (2017), Ricciardi et al. (2020) and Bia et al. (2020).

W	$S_{i,1}$	$S_{i,2}$	$S_{i,3}$	$\underline{G_i}$	$P(\{S_{i,t}\}_{t=1}^{3} \mid W_i)$
1	0	0	0	NS.NS.NS	2.44%
1	1	0	0	AS.NS.NS CS.NS.NS	1.46%
1	1	1	0	AS.AS.NS CS.CS.NS AS.CS.NS	0.56%
1	1	1	1	AS.AS.AS CS.CS.CS AS.CS.CS AS.AS.CS	95.54%
0	0	0	0	NS.NS.NS CS.CS.CS CS.NS.NS CS.CS.NS	13.18%
0	1	0	0	AS.CS.CS. AS.CS.NS AS.NS.NS	18.91%
0	1	1	0	AS.AS.CS AS.AS.NS	20.15%
0	1	1	1	AS.AS.AS	47.76%

Table 4.5: Composition and proportion of observed groups

Modelling assumptions

We first introduce an assumption over the hiring decision and future expectations on the survival status:

Assumption 9.

$$P(Y_{i,1}(W_i)|G_{i,1}, G_{i,2}, G_{i,3}, \mathbf{X}_i) = P(Y_{i,1}(W_i)|G_{i,1}, \mathbf{X}_i) \qquad \forall i \in \{1, \ldots, N\}$$

and

$$P(Y_{i,2}(W_i)|G_{i,1}, G_{i,2}, G_{i,3}, \mathbf{X}_i) = P(Y_{i,1}(W_i)|G_{i,1}, G_{i,2}, \mathbf{X}_i) \qquad \forall i \in \{1, \ldots, N\}$$

We state that hiring decisions at time t will not be affected by the expectations over the survival status in the following times. This is a reasonable assumption, as we think that present hiring will be ruled by the present "health status" and needs of the firm.

We now show the model for the joint distribution of the principal strata membership in each time, $\mathbf{G_{i,t}}$ and a couple of potential outcomes $(Y_{i,t}(0), Y_{i,t}(1)) \quad t \in 1, 2, 3$. Under assumption 6 and unit exchangeability (Rubin, 1978), the joint distribution can be expressed as

$$\prod_{i=1}^{N} P(G_{i1}, G_{i2}, G_{i3}, Y_{i1}(0), Y_{i2}(0), Y_{i3}(0), Y_{i1}(1), Y_{i2}(1), Y_{i3}(1)|\mathbf{X}, \boldsymbol{\theta}) \qquad (4.6)$$

Notice that, under units exchangeability, the joint posterior distribution can be viewed as the product of the model for the conditional probability of the longitudinal principal strata membership, given the covariates $Pr(\mathbf{G}_3, \mathbf{G}_2, \mathbf{G}_1|\mathbf{X}_i, \theta)$ and the model for the principal outcome, given the covariates and the longitudinal principal strata membership. By applying the law of total probability, and under the assumptions over the firm's behaviour, we can rewrite the joint probability distribution as:

$$
\begin{aligned}
&\prod_{i=1}^{N} P(G_{i1}, G_{i2}, G_{i3}, Y_{i1}(0), Y_{i2}(0), Y_{i3}(0), Y_{i1}(1), Y_{i2}(1), Y_{i3}(1)|\mathbf{X}, \boldsymbol{\theta}) = \\
&\qquad \prod_{i=1}^{N} \prod_{W_i \in \{0,1\}} P(G_{i1}, G_{i2}, G_{i3}, Y_{i1}(W_i), Y_{i2}(W_i), Y_{i3}(W_i)|\mathbf{X}, \boldsymbol{\theta}) = \\
&\qquad \prod_{i=1}^{N} P(G_{i1}|\mathbf{X}, \boldsymbol{\theta}) \times P(G_{i2}, |G_{i1}, \mathbf{X}, \boldsymbol{\theta}) \times P(G_{i3}, G_{i1}, G_{i2}|\mathbf{X}, \boldsymbol{\theta}) \times \\
&\qquad \prod_{W_i \in \{0,1\}} P(Y_{i1}(W_i)|G_{i1}, \mathbf{X}, \boldsymbol{\theta}) \times P(Y_{i2}(W_i)|G_{i1}, G_{i2}, Y_{i1}, \mathbf{X}, \boldsymbol{\theta}) \\
&\qquad \times P(Y_{i3}(W_i)|G_{i1}, G_{i2}, G_{i3} Y_{i1}, Y_{i2}, \mathbf{X}, \boldsymbol{\theta})
\end{aligned}
\tag{4.7}
$$

Parametric modelling

In subsection 4.4.5, we have specified the modelling assumption and structure of the joint probability distribution of longitudinal principal strata membership, covariates and principal outcome. Now we present the parametric model we use to estimate both the probability of longitudinal principal strata membership and the probability of a hiring for each unit.

Equation 4.8 shows the parametric models we used to estimate the principal stratum membership each year. We employ a multivariate logit model, using the $G_{i,t} = NS$ as a reference level. In the model for principal stratum membership, we use the matrix of pre-treatment covariates $\mathbf{X}_i$ and a fixed effect δ_0 that depends on the principal strata membership in the previous period. Please remember that, by construction, some transitions are not feasible. Thus following we have that $P(G_{i,t} = AS|G_{i,t-1} = CS) = 0$ and $P(G_{i,t} = NS|G_{i,t-1} = NS) = 1$

$$
\begin{cases}
\log\left(\dfrac{P(G_{i,1} = g_{i,1}|\mathbf{X}_i)}{P(G_{i,1} = NS|\mathbf{X}_i)}\right) = \delta_{0,1}^{W_i} + \boldsymbol{\delta_1}\mathbf{X}_i & g_{i,1} \in \{AS, CS\} \\
\log\left(\dfrac{P(G_{i,2} = g_{i,2}|G_{i1}, \mathbf{X}_i)}{P(G_{i,2} = NS|\mathbf{X}_i)}\right) = \delta_{0,2}^{W_i, G_{i,1}} + \boldsymbol{\delta_2}\mathbf{X}_i & g_{i,2} \in \{AS, CS\} \\
\log\left(\dfrac{P(G_{i,3} = g_{i,3}|G_{i1}, G_{i2}\mathbf{X}_i)}{P(G_{i,3} = NS|\mathbf{X}_i)}\right) = \delta_{0,3}^{W_i, G_{i,1}, G_{i,2}} + \boldsymbol{\delta_3}\mathbf{X}_i & g_{i,3} \in \{AS, CS\}
\end{cases}
\tag{4.8}
$$

We consider the potential outcome dependent on past values of Y_i and $\mathbf{G}_i$, and contemporaneous values of $\mathbf{G}_i$ but not on future values. We specify different fixed terms according to the treatment assignment and principal strata membership. To account for the potential temporal dependence of hiring decisions, we included an autoregressive coefficient in our analysis. We do not specify different coefficients for covariates belonging to different latent strata $\mathbf{G}_i$: $\boldsymbol{\beta}_t^{AS} = \boldsymbol{\beta}_t^{CS} = \boldsymbol{\beta}_t^{NS}$, but we consider principal strata and treatment specific fixed effect $\beta_t^{W_i,\mathbf{G}_i}$.

$$\begin{cases} \log\left(\dfrac{P(Y_{i,1}(W_i)=1|G_{i,1},\mathbf{X}_i)}{1-P(Y_{i,1}(W_i)=1|G_{i,1},\mathbf{X}_i)}\right) = \beta_{0,t}^{W_i,G_{i,1}} + \boldsymbol{\beta}_1\mathbf{X}_i \\ \log\left(\dfrac{P(Y_{i,2}(W_i)=1|G_{i,1},G_{i,2},\mathbf{X}_i)}{1-P(Y_{i,t}(W_i)=1|G_{i,1},G_{i,2},\mathbf{X}_i)}\right) = \beta_{0,t}^{W_i,G_{i,1},G_{i,2}} + \lambda_t^{W_i,G_{i,1}}Y_{i,t-1} + \boldsymbol{\beta}_1\mathbf{X}_i \\ \log\left(\dfrac{P(Y_{i,t}(W_i)=1|G_{i,1},G_{i,2},G_{i,3},\mathbf{X}_i)}{1-P(Y_{i,t}(W_i)=1|G_{i,1},G_{i,2},G_{i,3},\mathbf{X}_i)}\right) = \beta_{0,t}^{W_i,G_{i,1},G_{i,2},G_{i,3}} + \lambda_t^{W_i,G_{i,1},G_{i,2}}Y_{i,t-1} + \boldsymbol{\beta}_1\mathbf{X}_i \end{cases} \tag{4.9}$$

We specify proper, yet non-informative prior distributions. It has been argued by Gustafson (2010) that even if a statistical model is only partially identified, by utilizing appropriate prior distributions, we can still obtain meaningful posterior distributions. Therefore, from a Bayesian perspective, there is no fundamental difference between fully and partially identified models.

To obtain posterior distributions for the relevant parameters and missing quantities, we use a statistical algorithm called Hamiltonian Monte Carlo (HMC). The estimations were carried out using the RStan software. We used 2000 iterations with 1000 warm-up iterations, and no pathological behaviour is found in the diagnostic results. Prior specifications, further results and posterior checks are reported in Appendix 4.

4.5 Results

In this section we collect results from our analysis. Valuable insights can be derived both from the principal strata membership and from the longitudinal causal effects.

4.5.1 Principal strata membership

Table 4.7 shows the summary statistics (mean st. deviation, fifth and ninety-fifth quantiles) for the probabilities of longitudinal principal stratum membership. We note that the majority of firms are compliant survivors (58.7 %) which are carrying on their activity because of the bank loan. A sizeable part of start-ups are classified as always survivors (34.5 %) suggesting that around one-third of the units exhibit some level of entrepreneurial ability and their businesses would have survived irrespective of treatment assignment. Few units would have failed anyway. In particular, the 2.4 %,

Table 4.6: Descriptive statistics for strata membership probability in $t \in \{1, 2, 3\}$

	Mean	st.dev	0.025	0.05	0.95	0.975
$P(G_{1,t} = AS)$	0.384	0.129	0.191	0.207	0.631	0.674
$P(G_{1,t} = CS)$	0.592	0.129	0.303	0.343	0.768	0.786
$P(G_{1,t} = NS)$	0.024	0.002	0.021	0.021	0.027	0.028
$P(G_{2,t} = AS)$	0.343	0.128	0.149	0.165	0.583	0.636
$P(G_{2,t} = CS)$	0.603	0.128	0.313	0.365	0.780	0.799
$P(G_{2,t} = NS)$	0.054	0.002	0.050	0.050	0.055	0.055
$P(G_{3,t} = AS)$	0.343	0.130	0.148	0.165	0.587	0.638
$P(G_{3,t} = CS)$	0.604	0.130	0.308	0.360	0.782	0.798
$P(G_{3,t} = NS)$	0.054	0.001	0.052	0.053	0.054	0.054

5.3 % and 5.3% are never survivors in the first, second and third period, respectively. Other strata (AS.CS.NS, AS.AS.CS, CS.NS.NS) seem to be very uncommon, as the posterior mean of belonging to these strata is lower than 1%. Table 4.6 reports the summary statistics for the principal strata membership at each time. It is worth noting that transitions between different categories, such as from always survivors to Compliant Survivors are relatively uncommon. Once a firm is classified in a specific subgroup, it tends to remain in that subgroup.

From the posterior probability of the longitudinal principal strata, it appears that easing credit access is a key factor in the survival of these firms. This can be seen as a positive outcome of the policy, as it achieved its goal of supporting self-employment for young people and women. However, there seems to be a strong connection between the survival of businesses and the receipt of public support, indicating that many firms that receive subsidies may not have strong growth potential. This raises the question of whether this type of support is the most effective use of public funds, and whether it exposes public actors to the financial risk of unpaid loans.

Figure 8.6 in the Appendix 4 reports the boxplots for each covariate across different longitudinal principal strata. There are some differences in the pre-treatment covariates among the different strata. In particular, we can see that firms classified as always survivors in the first two periods differ from compliant survivors in terms of their characteristics. Compliant survivors are more likely to have a young, single owner and to be new establishments in the retail and service sectors. On the other hand, always survivors are often more established businesses, owned by women, and operating in the manufacturing sector, which is relatively rare among compliant survivors. Compliant survivors also tend to rely more on local bank loans, while always survivors may have access to national banks due

Table 4.7: Descriptive statistics for the posterior probability of longitudinal principal strata memberships

	Mean	St.Dev	0.05	0.95
NS.NS.NS	0.024	0.002	0.021	0.027
CS.NS.NS	0.007	0.002	0.003	0.011
AS.NS.NS	0.046	0.003	0.041	0.051
AS.AS.NS	0.016	0.007	0.007	0.031
AS.CS.NS	0.000	0.000	0.000	0.001
CS.CS.NS	0.037	0.007	0.022	0.047
CS.CS.CS	0.587	0.138	0.322	0.778
AS.CS.CS	0.013	0.036	0.000	0.070
AS.AS.CS	0.002	0.003	0.000	0.009
AS.AS.AS	0.345	0.130	0.165	0.587

to their stronger financial standing. In general, there seems to be a distinction between established firms with higher potential and start-up projects that are mainly led by young entrepreneurs and heavily reliant on bank loans in the first period but do not show particularly strong potential.

4.5.2 Principal strata effects

	AS	AS.AS	AS.AS.AS
Mean	0.1079	0.0576	-0.0287
SD	0.0548	0.0680	0.0573
0.05	0.0038	-0.0735	-0.1305
0.95	0.2177	0.1892	0.0802

Table 4.8: Descriptive statistics for SATE effects in each period $t \in \{1, 2, 3\}$

Table 4.8 and figure 4.1 show the summary statistics (mean, st.deviation, fifth and ninety-fifth quantiles) of the posterior distribution of SATE. We can get precious insights from the analysis of such effects.

First, we can observe a positive effect of the treatment in the first period after the bank loan. We can hypothesize that the additional liquidity provided by the bank loan help firms not only to start the activity but also to hire the human resources needed for the firm's operations. Instead, for firms established recently, it is possible that this loan support an enlargement opportunity with an increase in the employed number.

Second, we can observe a temporal trend in the treatment effects. Comparing the SATE causal effects in $t \in \{1, 2, 3\}$ we can see that the estimated effect diminishes with the passage of time. The SATE causal effect is positive and statistically significant in the first period. During the second period, the effect is still positive, but smaller and not statistically significant. In the last period, the effect is very small and not statistically significant.

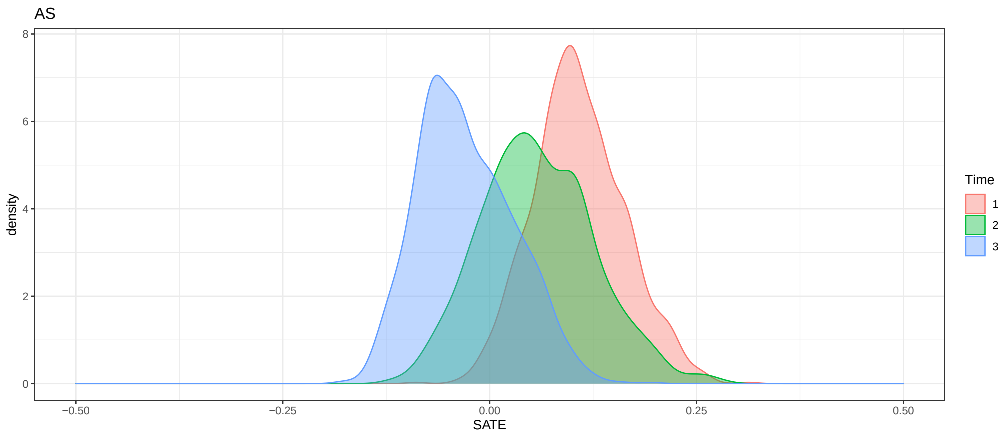


Figure 4.1: SATE posterior probabilities at $t \in \{1, 2, 3\}$

These results can be reasonable: firms hire new workers during the kick-off periods of the financed project, and these projects seem to have not generated enough growth to allow for additional hiring in the subsequent years. This is a negative result because even the more promising projects have not generated enough sales to justify further investment in human resources.

Table 4.9 reports some measures for the goodness of fit (GOF) of the estimations. Working with a binary outcome (hiring yes/no in year t), we have compared the predicted outcomes from our model with the observed outcomes. We also construct a pseudo-R^2 measure as the share of corrected estimates on the total observations. We get 58.6% of correct estimates in the first period, and around 70% in the following periods, which can be considered a reasonable estimate. Table 4.10 instead reports the descriptive statistics for the probability of correcting estimating the hiring or non hiring decision in the three post-treatment period. Interestingly, we can notice that model performance increases in periods 2 and 3. Figure 8.4 and 8.5 in the Appendix 4 reports share of corrected predicted outcomes for the two levels of hiring decision.

Table 4.9: Goodness of fit measures, mean results over the HMC iteration

	Underestimation	Correct	Overestimation	Pseudo-R2
t=1	255.006	1024.197	469.796	0.586
t=2	250.525	1170.957	230.518	0.709
t=3	154.120	1062.444	347.435	0.680

Table 4.10: Descriptive statistics for the correct estimation of hiring decision (first pane) and non-hiring decision (second pane) of the outcome model

	Mean	St.Dev	0.05	0.95
t=1	0.603	0.089	0.471	0.759
t=2	0.463	0.107	0.258	0.623
t=3	0.477	0.128	0.254	0.661

	Mean	St.Dev	0.05	0.95
t=1	0.575	0.082	0.422	0.695
t=2	0.805	0.068	0.691	0.919
t=3	0.726	0.086	0.578	0.858

4.6 Conclusions

This work investigates the impact of a public policy aiming to ease access to the credit market for start-ups. One major challenge we faced is non-ignorable censoring. In particular, it was impossible to evaluate the effect of the policy on the main outcome (hiring decisions made by the funded start-ups) when the outcome was censored by the closure of the business.

To address this issue, we adopted a principal stratification approach based on the proposal by Bia et al. (2020) within the potential outcomes framework. We modified this method to accommodate the longitudinal structure of our data and identified principal strata based on the survival of the firms in each post-treatment time period. This allowed us to identify subpopulations of units that would have survived regardless of their treatment assignment and to estimate the causal effects within these groups.

Our work expands on the literature on using principal stratification to address censoring problems in a longitudinal setting. We identified principal strata for three time periods and provided reasonable assumptions to simplify the number of strata to be estimated. Additionally, the longitudinal perspective allowed us to examine the temporal patterns of the estimated effects.

Exploiting a Bayesian approach for inference, we imputed missing potential outcomes via a data augmentation algorithm and derive the posterior distribution for longitudinal stratum membership and causal effects on hiring decisions.

Policy evaluation of subsidies to start-ups is a debated theme in economic literature, but few studies focus on the second-round effect of such policies and even fewer exploit a proper causal structure. In this work, we explicitly address both of these gaps. We focus on the secondary outcome of this policy, which is the hiring decisions made by funded start-ups, and aim to determine if this active labour policy has the potential to create additional jobs and provide a "double dividend" for the community. Our study adds to the existing literature on public support for start-ups by using a rigorous causal framework to analyze the effect of subsidies on a collateral outcome, namely the employment generated by public intervention.

Our results show that the policy has a positive effect on the stratum of always survivors, which is statistically significant in the first year after the treatment. This may be due to the initial period of the projects. In the subsequent years, no statistically significant effects are observed, suggesting that there is a temporal dependence between the loan period and the hiring decisions of the start-ups. This suggests that the policy is successful in creating new job openings, but once these positions are filled, the start-ups struggle to expand and create additional employment. Therefore, job creation appears to be dependent on the fundraising ability of the firm owner.

We also found that the survival of the majority of firms is heavily dependent on the policy, as about 60% of the firms would have closed without treatment. These start-ups seem to be the most fragile, with less potential for growth. About 35 % of the firms would have remained active in the market regardless of their treatment assignment, suggesting that they have higher potential.

We also characterize firms according to their longitudinal principal strata membership. We found that incumbent firms were more likely to survive regardless of their treatment assignment, which highlights the strength of their entrepreneurial projects. Additionally, we observed that start-up projects were more likely to be undertaken by young firm owners, but these projects were also more dependent on financial subsidies. There were also differences in treatment dependence among economic sectors, with manufacturing firms being less reliant on public aid and restaurants and service firms being more reliant on support.

Our research concurs with the conclusions of Mariani et al. (2019) in suggesting that the provision of public support for facilitating credit market access constitutes a viable strategy for active labour market policy. However, it must be acknowledged that such a strategy entails the utilization of public resources for projects whose potential outcomes may be open to debate.

Overall, our results provide insight into the complementary effects of public support for start-ups. On the one hand, we found that easing access to credit had notable effects on self-employment and job creation, resulting in a double impact on unemployment reduction. On the other hand, the effect of this policy is temporary, and the majority of firms seemed to be dependent on public support, indicating limited potential. These findings can inform policymakers as they consider the use of public subsidies to support start-ups.

Conclusions

In this work, I focused my attention on the methods and applications of policy evaluation for panel data.

Policy evaluation studies often deal with observational data, and thus many complications could arise from real-data environments. Even in relatively simple settings, many challenges could harm the correctness of the study, and thus, I provide some proposals to deal with such complications.

One of these complications is the presence of spillovers between treated and untreated units. In policy evaluation problems with panel data settings, usually scholars rule out the interference across units. In this work, we provide a solution for estimating direct, indirect and spillover effects using the Synthetic Control Method. Our approach is very innovative and poses itself into the emerging literature of policy evaluation in interference settings, which will be more and more important for describing real-world applications in the following years. We exploit such a method in evaluating the causal effect of the construction of the first line of the Florentine tramway, providing useful insights to policymakers interested in understanding better the dynamics of urban infrastructures and retail vitality.

Estimating causal effects with panel data in presence of a multiplicity of treated units and a staggered adoption framework could be challenging. In the second chapter of this work, we evaluate the effects of a lottery policy implemented by several US states to foster the Covid-19 vaccination in the population. Using a disaggregated framework, we study the causal effect of such policy, estimating the causal effect at the county level, state level and macro-region level. Results from our analysis show a wide heterogeneity in the treatment effects across different areas, even within the same state. By studying the timing of policy implementation we derive some useful insights for policymakers. We also study the treatment effect heterogeneity investigating the role of socio-demographic characteristics.

The third chapter of this dissertation proposes a novel estimator for estimating causal effects in panel data framework, with spatially correlated treated units. We call this method SMaC, as Spatial Matrix Completion Method. Usually, in such environments, scholars estimate causal effects by estimating one-by-one the causal effects for each treated unit. Even if this is correct, it does not take into account the underlying correlation structure. We propose to consider it by estimating the causal effect with a Bayesian regression, built on the SCM and vertical regression literature, with Gaussian Processes as priors for the regression coefficients. We show with simulation under several scenarios the properties of our estimator. We apply SMaC to the evaluation of the construction of the first line of the tram in Florence, estimating the causal effect of the tramway on the number of shops comprised within various distances. This method can be particularly effective in helping policymakers to understand how a treatment effect emanates through space.

The last chapter of the dissertation proposes a novel method to deal with non-ignorable censoring in policy evaluation. In panel data studies, with longitudinal outcomes and post-treatment complications, we propose a longitudinal principal stratification approach to evaluate the effects of a policy implemented in Tuscany to stimulate entrepreneurship among young and female citizens. The results of our analysis underline the importance of public support to freshly established firms, especially in easing their access to the credit market. In doing so, we got promising results from the further job creation of treated units stressing the double dividend effect of such a policy.

Further research on these topics is needed. In particular, real-world observations often present underlying structures, as in spatial, or spatiotemporal data and networked data, which policy evaluation should address explicitly. On this basis, the future focus of my research will be on the bridge between econometrics and causal inference in order to develop and disseminate reliable and transparent methods for policy evaluation. Nevertheless, further research will focus on a common approach to deal with interference across units in panel data settings, providing a unified framework to estimate the direct and spillover effects of policies.

Chapter 5

Appendix - Chapter 1

5.1 Bootstrap-accelerated confidence intervals

Let $\theta \in \Theta$ the estimand of interest, where Θ is the parameter space and let $\widehat{\theta}$ be an estimate of θ. Let $\widehat{F}(\cdot)$ denote the bootstrap cumulative distribution function of the estimator of θ. Define $g : [0,1] \longrightarrow \Theta$, such that for each $u \in [0,1]$

$$g(u) = \widehat{F^{-1}}\left(\Phi\left(z_0 + \frac{z_0 + z_u}{a(z_0 + z_u)}\right)\right),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $z_0 = \Phi^{-1}(\widehat{F}(\widehat{\theta}))$, $z_u = \Phi^{-1}(u)$ and a is an acceleration constant. For $\alpha \in (0,1)$, the accelerated bootstrap $(1-\alpha)$ confidence interval is given by $[g(\alpha/2), g(1-\alpha/2)]$. We estimate the acceleration constant, a, as

$$\hat{a} = \frac{\sum_{i=1}^{n} I_i^3}{6(\sum_{i=1}^{n} I_i^2)^{\frac{3}{2}}}$$

where n is the sample size and I_i denotes the influence of data point i on the estimation of θ that we approximate using the finite-sample Jackknife method.

5.2 Tables

Table 5.1: Values for the outcomes of interest for the streets in the treated neighborhood

Number of stores selling durable goods						
	Talenti	Pollaiolo	Pisana	Scandicci	Magnolie	Baccio
1996	**7.983**	13.524	10.007	6.299	10.448	7.092
1997	**9.166**	15.026	11.042	6.693	10.448	7.092
1998	**8.870**	13.899	10.352	5.906	10.448	7.447
1999	**9.758**	13.899	10.697	6.299	11.194	7.801
2000	**11.236**	15.402	11.042	6.693	12.687	8.511
2001	**11.236**	15.402	11.387	7.874	12.687	9.220
2002	**10.053**	13.148	10.697	7.087	11.940	8.511
2003	**10.053**	12.772	10.697	7.874	11.194	9.929
2004	**9.758**	12.772	10.007	6.693	11.940	10.638
2005	**9.758**	12.772	11.042	6.299	12.687	12.057
2006	**9.462**	12.772	11.042	6.693	12.687	12.057
2007	**10.053**	13.148	11.732	5.512	11.194	12.057
2008	**9.758**	11.270	11.732	5.512	9.701	11.348
2009	**9.758**	10.894	11.732	5.118	8.955	11.348
2010	**9.758**	12.397	11.387	5.118	10.448	11.702
2011	**8.575**	12.021	12.077	5.906	10.448	12.057
2012	**7.688**	12.397	11.732	6.299	11.194	10.638
2013	**7.688**	11.645	9.662	6.299	11.194	9.574
2014	**7.096**	12.021	9.662	7.480	11.194	8.865
Number of stores selling non-durable goods						
	Talenti	Pollaiolo	Pisana	Scandicci	Magnolie	Baccio
1996	**6.801**	9.767	10.697	6.299	11.194	8.156
1997	**7.392**	10.518	12.422	7.087	11.940	9.220
1998	**7.392**	10.143	10.007	5.512	11.194	7.801
1999	**7.688**	10.894	10.697	5.906	11.940	8.156
2000	**8.575**	11.270	11.387	6.299	11.940	9.929
2001	**9.758**	11.270	11.732	5.906	12.687	10.993
2002	**7.392**	9.016	10.697	5.118	11.940	9.220
2003	**7.983**	8.640	10.697	5.512	11.194	9.220
2004	**7.688**	8.640	9.662	5.512	12.687	9.929
2005	**7.392**	9.391	11.042	5.906	12.687	10.993
2006	**7.983**	10.143	12.077	5.906	14.925	10.638
2007	**7.688**	10.894	12.077	4.724	14.925	11.348
2008	**7.688**	10.518	12.077	4.724	14.925	10.638
2009	**7.688**	11.645	12.077	4.724	14.179	11.348
2010	**9.166**	12.397	11.387	5.118	14.179	11.702
2011	**9.462**	12.021	12.767	5.512	14.179	12.057
2012	**9.758**	10.518	10.697	5.118	12.687	11.348
2013	**9.462**	8.640	10.697	4.724	11.940	10.284
2014	**8.279**	7.137	8.972	4.331	11.194	9.929

Table 5.2: Penalization terms - A: Number of stores selling durables, B: Number of stores selling non-durables

$\lambda^{(i)}$		λ	
Talenti St.		**Streets in Legnaia neighbourhood**	
A	**B**	**A**	**B**
0.089	0.009	0.089	0.009

Table 5.3: Weights through which the synthetic control values of the outcome variables $Y_{1t}(0, \mathbf{0}_{N_1})$ and $Y_{it}(0, \mathbf{0}_{N_1})$. A: number of stores selling durables; B: number of stores selling non-durables

	$Y_{11,t}(0, \mathbf{0}_{N_1})$		$Y_{1i,t}(0, \mathbf{0}_{N_1})$									
	Talenti		Pollaiolo		Pisana		Scandicci		Magnolie		Baccio da M.	
	A	B	A	B	A	B	A	B	A	B	A	B
Affrico	0	0	0	0	0	0	0	0.2703	0.0520	0.0483	0.2103	0
Alderotti	0	0	0	0	0	0	0	0	0.0719	0	0	0
Aretina	0	0.1335	0	0.1097	0	0.0521	0	0	0	0.0180	0	0
Baracca	0	0	0	0	0	0	0	0	0	0	0	0
Caracciolo	0.115	0	0.1292	0.0542	0.02	0.0834	0.184	0.1140	0	0	0	0
Centostelle	0.0678	0	0	0	0	0	0	0	0	0	0	0
Corsica	0	0	0	0	0	0	0.2531	0	0	0	0	0.2702
DAnnunzio	0.1306	0	0	0	0	0	0	0	0	0	0	0
Datini	0.0454	0.0344	0.1587	0	0	0	0	0	0	0	0	0
DeSantis	0	0	0	0	0	0	0	0	0	0.0233	0	0
Europa	0	0	0.1235	0	0	0	0	0	0	0	0	0
Faentina	0	0.1742	0	0	0.0182	0	0	0	0	0	0	0
Galliano	0	0	0	0.3826	0	0	0	0.5391	0	0.0653	0	0
Giuliani	0.0291	0	0	0	0	0	0	0	0.0449	0	0	0
Guidoni	0	0	0	0	0	0.3818	0	0.0085	0	0.3946	0	0.0376
Maffei	0	0.0649	0	0	0	0	0	0	0	0.0164	0.0319	0.2277
Maragliano	0.0085	0.1879	0.0688	0	0.3725	0	0.0572	0	0	0	0	0
Mariti	0	0	0	0	0	0	0	0	0	0.1349	0.198	0.3335
Masaccio	0	0	0	0	0	0	0.2861	0	0	0	0	0.0668
Mille	0.1372	0	0	0	0	0	0	0	0.4193	0	0	0
Morgagni	0	0.2263	0.1424	0	0.0177	0	0	0.0681	0	0.1089	0	0
Novoli	0	0	0	0	0	0	0	0	0	0	0	0
Panche	0	0	0	0.0109	0.0707	0	0	0	0	0	0	0
Peretola	0	0	0	0	0	0	0	0	0	0	0	0
Piagentina	0	0	0	0	0	0	0	0	0.0003	0	0	0
Pistoiese	0	0	0	0	0	0	0	0	0	0	0	0
PontealleMosse	0	0	0	0	0	0	0	0	0.1295	0	0.1942	0
PontediMezzo	0	0	0	0	0	0.3168	0	0	0	0	0	0
Pratese	0.2016	0	0	0	0	0	0	0	0.1521	0	0	0
Redi	0	0	0	0	0	0	0	0	0.1132	0	0.3657	0.1018
Ripoli	0	0	0	0	0	0	0.1636	0	0	0	0	0
Romito	0	0	0.1554	0.3063	0	0	0	0	0	0	0	0
Rondinella	0	0	0	0	0	0.1389	0	0	0.0072	0	0	0.0217
Tavanti	0	0	0.0353	0.0407	0	0	0	0	0	0	0	0
Toselli	0.2648	0.0407	0	0	0	0	0	0	0	0	0	0
Villamagna	0	0.0357	0	0.0798	0.5009	0	0	0	0	0.1867	0	0
VittorioEmanuele	0	0.1289	0.1867	0	0	0	0.056	0	0	0	0	0
Volta	0	0	0	0	0	0	0	0	0	0	0	0

Table 5.4: RMSPE for the treated street, Talenti St. and the for untreated streets in the treated cluster

	$Y_{1i,t}(0, \mathbf{0}_{N_1})$ Number of stores selling	
Street	durable goods	non-durable goods
Talenti St.	0.1829	0.1112
Pollaiolo St.	0.2148	0.1082
Pisana St.	0.1854	0.2305
Scandicci St.	0.2452	0.2598
Magnolie St.	0.2080	0.3024
Baccio St.	0.2619	0.4416

5.3 Figures

Figure 5.1: Observed values of the number of purveyors of durable (left panel) and non-durable (right panel) goods every 500 meters over the time period 1996-2014 in the treated street (Talenti St.) and in other streets belonging to the same urban neighborhood (Pollaiolo St., Pisana St., Baccio da Montelupo St., Scandicci St., and Magnolie St.)

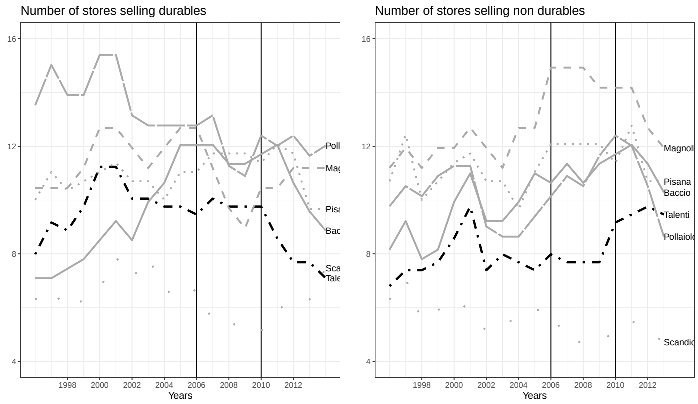
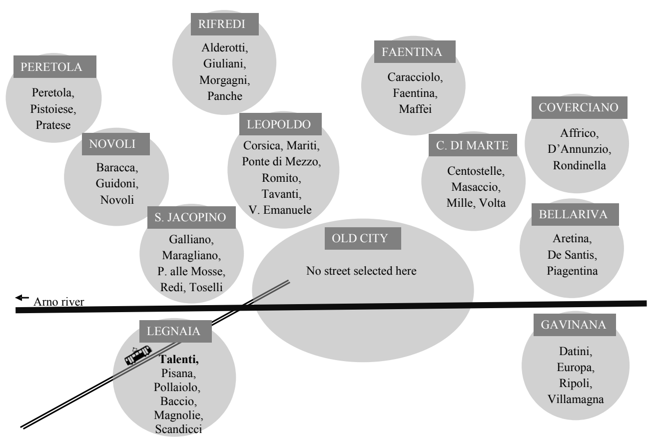


Figure 5.2: Streets involved in the analysis, clustered in their own urban neighborhoods



Chapter 6

Appendix - Chapter 2

6.1 Tables

State	Lottery Start	Lottery End	Tr. Days	Tr. Weeks	Program Cost (in $)
Ohio	13/05/2021	23/06/2021	41	6	5000000
New York	20/05/2021	11/06/2021	22	4	5078340
Oregon	21/05/2021	28/06/2021	38	6	1360000
Delaware	24/05/2021	29/06/2021	36	5	302000
Maryland	25/05/2021	04/07/2021	40	6	2000000
Colorado	25/05/2021	04/07/2021	40	6	5000000
Arkansas	25/05/2021	30/06/2021	36	5	1000000
California	27/05/2021	18/07/2021	52	7	15000000
Washington	03/06/2021	13/07/2021	40	6	2000000
Kentucky	04/06/2021	27/08/2021	84	8	3000000
North Carolina	10/06/2021	04/08/2021	55	8	4000000
Louisiana	17/06/2021	31/07/2021	44	7	1400000
Nevada	17/06/2021	31/08/2021	75	10	5000000
New Mexico	17/06/2021	31/08/2021	75	10	10000000
Maine	17/06/2021	04/07/2021	17	3	896809
Illinois	17/06/2021	27/08/2021	71	10	7000000
West Virginia	20/06/2021	04/08/2021	45	6	2000000
Michigan	01/07/2021	30/08/2021	60	9	4500000
Missouri	21/07/2021	30/08/2021	40	4	9000000

Table 6.1: Relevant information for each lottery programs

	May 12th	July 1st	August 24th
Alabama	30.281	34.506	41.873
Arizona	52.893	62.479	68.471
Arkansas	34.336	38.589	49.084
California	45.829	53.536	59.193
Colorado	19.585	24.051	55.048
Connecticut	67.267	76.883	82.550
Delaware	53.533	62.167	67.400
District of Columbia	57.900	67.700	73.500
Florida	43.426	49.759	58.054
Georgia	14.497	17.199	22.760
Illinois	45.740	52.076	57.314
Indiana	40.218	44.730	51.424
Iowa	49.362	54.032	58.071
Kansas	41.564	45.436	51.099
Kentucky	38.741	44.318	50.206
Louisiana	33.606	37.961	48.013
Maine	60.337	68.744	72.938
Maryland	55.258	64.729	69.871
Massachusetts	48.257	55.471	58.771
Michigan	47.984	51.783	53.338
Minnesota	51.134	55.913	59.277
Mississippi	34.158	37.906	50.096
Missouri	32.781	36.793	43.314
Montana	42.341	47.350	50.822
Nebraska	21.554	21.870	22.123
Nevada	38.065	42.865	47.682
New Hampshire	53.120	66.750	70.680
New Jersey	57.571	67.671	73.700
New Mexico	21.583	22.090	22.090
New York	53.625	62.395	66.354
North Carolina	43.316	48.482	53.916
North Dakota	41.784	45.020	48.438
Ohio	31.113	33.336	34.493
Oklahoma	37.418	41.906	48.047
Oregon	49.874	57.883	62.137
Pennsylvania	49.219	57.064	61.949
Rhode Island	29.840	29.840	78.440
South Carolina	38.641	44.830	50.911
South Dakota	19.858	20.617	21.698
Tennessee	34.167	38.199	45.519
Utah	36.035	39.504	58.773
Vermont	50.500	59.008	61.623
Virginia	24.056	27.471	30.078
Washington	49.329	58.712	63.483
West Virginia	21.873	26.602	29.631
Wisconsin	49.865	55.525	59.699
Wyoming	37.113	40.770	45.591
Mean	40.220	45.677	52.282
St.Dev.	13.439	15.985	16.680
5%	19.681	21.055	22.101
50%	41.674	45.228	53.627
95%	57.785	67.690	73.630

Table 6.2: First panel:% of 18+ citizens vaccinated with first dose per each state, Second panel: % of 18+ citizens vaccinated with first dose in the US

	Hispanic	Black	Poors	Republicans	High School	College	Unemployment	Death/100k	Medicare	Median Earnings	Median Age
Ohio	0.000	0.288	0.373	0.023	0.001	0.038	0.153	0.843	0.816	0.184	0.696
New York	0.760	0.872	0.131	0.869	0.577	0.242	0.296	0.607	0.890	0.204	0.685
Oregon	0.401	0.569	0.171	0.023	0.008	0.088	0.163	0.026	0.365	0.516	0.417
Delaware	0.615	0.653	0.557	0.883	0.824	0.698	0.912	0.325	0.887	0.893	0.783
Maryland	0.151	0.543	0.231	0.179	0.121	0.210	0.261	0.391	0.144	0.032	0.201
Arkansas	0.988	0.555	0.803	0.974	0.748	0.811	0.737	0.935	0.818	0.337	0.995
California	0.000	0.000	0.004	0.000	0.000	0.013	0.000	0.000	0.232	0.005	0.163
Washington	0.785	0.008	0.076	0.017	0.028	0.503	0.001	0.000	0.492	0.582	0.162
Kentucky	0.000	0.093	0.000	0.002	0.085	0.000	0.012	0.941	0.536	0.016	0.530
North Carolina	0.708	0.490	0.479	0.291	0.706	0.754	0.353	0.319	0.792	0.839	0.837
Louisiana	0.095	0.359	0.076	0.692	0.028	0.583	0.134	0.791	0.125	0.560	0.030
Nevada	0.713	0.636	0.601	0.741	0.520	0.535	0.577	0.579	0.725	0.369	0.753
Maine	0.175	0.125	0.316	0.426	0.695	0.264	0.788	0.069	0.194	0.316	0.214
Illinois	0.778	0.793	0.906	0.483	0.685	0.303	0.136	0.465	0.440	0.859	0.710
West Virginia	0.000	0.291	0.045	0.044	0.003	0.104	0.000	0.133	0.874	0.061	0.276
Michigan	0.170	0.843	0.674	0.963	0.813	0.157	0.000	0.399	0.010	0.013	0.010

Table 6.3: P-values for two-sample t-test of equality between treatment and control group for each state

State	λ^*	$\Phi(\lambda)$
Ohio	0	394.175
New York	0	412.276
Oregon	0	390.725
Delaware	0.1	13.385
Maryland	0	364.435
Colorado	0	465.4
Arkansas	0	514.787
California	0	466.172
Washington	0	479.462
Kentucky	0	707.136
North Carolina	0	445.95
Louisiana	0	488.767
Nevada	0	500.973
New Mexico	0	408.573
Maine	0	385.986
Illinois	0	571.956
West Virginia	0	417.775
Michigan	0	522.446

Table 6.4: Cross-validation results of λ^* for each states

6.2 Figures

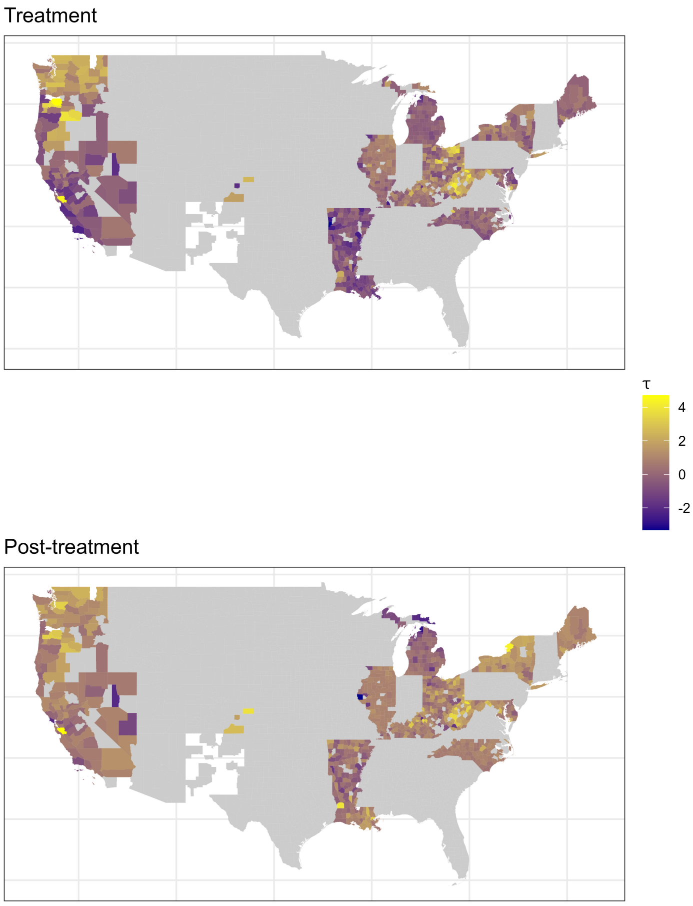


Figure 6.1: Average Treatment Effect Δ_i during treatment and post treatment, represented at county level, estimated according equation 2.6

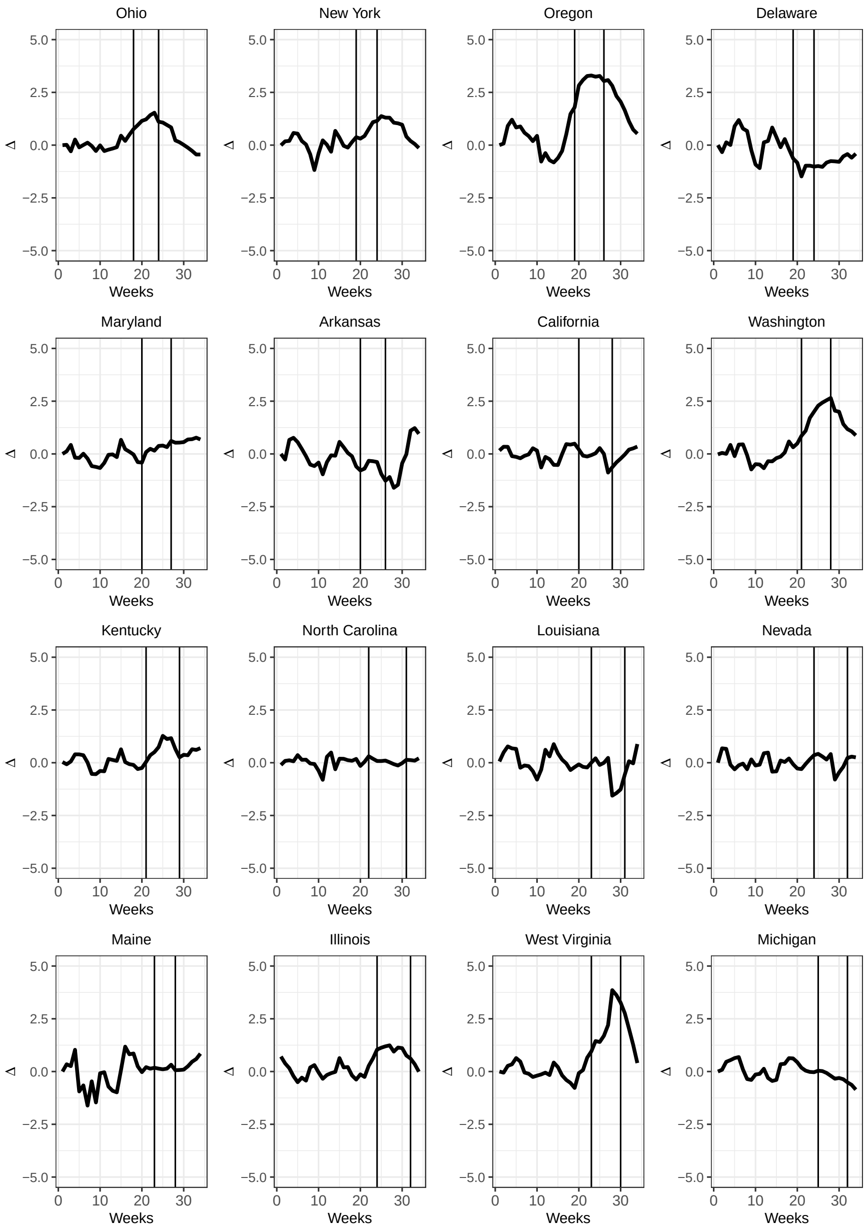


Figure 6.2: Intertemporal effect $\Delta_{\mathbf{S},t}$ - solid line: estimated effect. First vertical line: lottery announcement for the treated state, second vertical line: end of lottery.

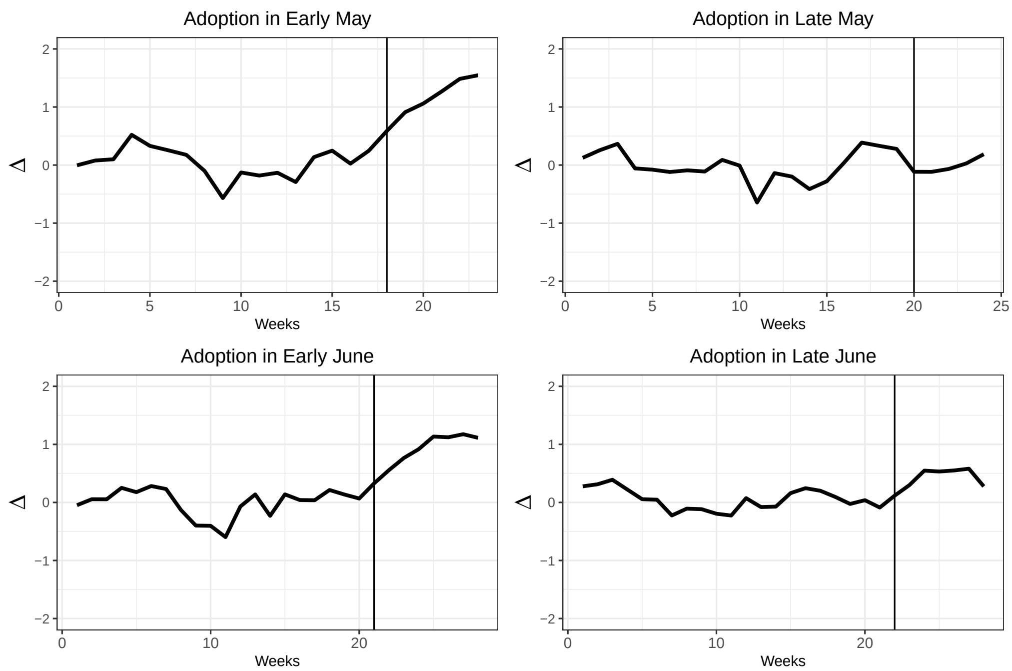


Figure 6.3: Treatment effect pooled for staggered adoption of the treatment

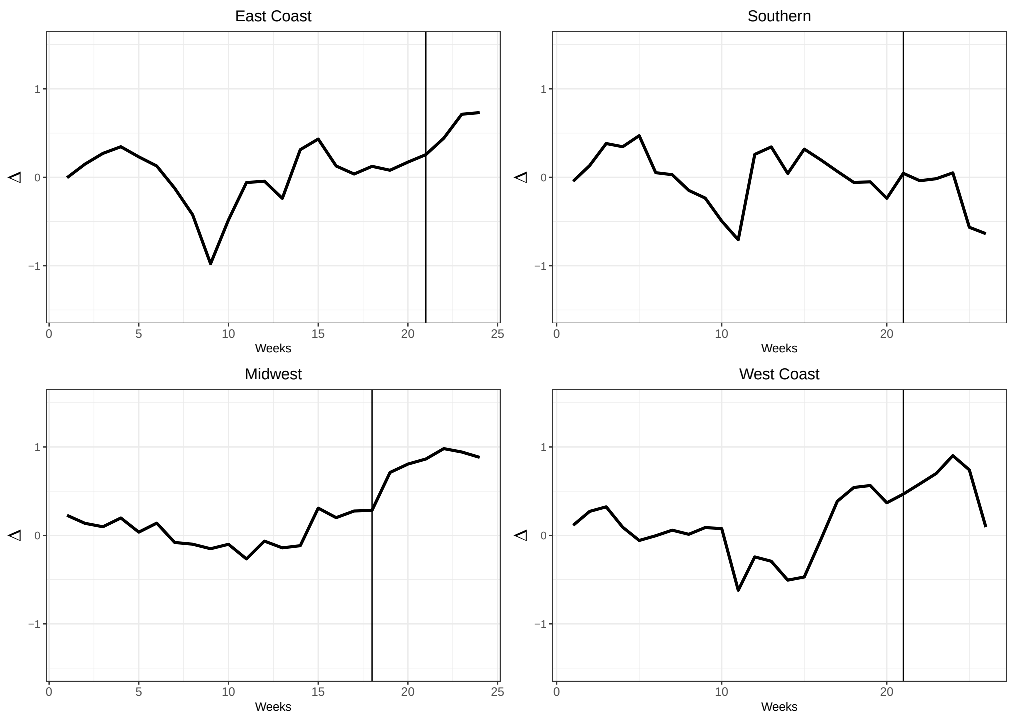


Figure 6.4: Treatment effect pooled for US macro regions

Chapter 7

Appendix - Chapter 3

7.1 Spatially-penalized vertical regression

We return to the stacked SCM in 3.8, and we consider a penalized version of the form

$$\begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_i \\ \vdots \\ \boldsymbol{\beta}^H \end{pmatrix} = \underset{\boldsymbol{\beta}_0, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_N \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \sum_{h=1}^{H} \sum_{t=1}^{T_0 - 1} \left(Y_{1,t}^h - (1 \ \mathbf{Y}_{i,t}^T)^T \boldsymbol{\beta}_i \right)^2 + \sum_{i=1}^{N} \lambda_i \boldsymbol{\beta}_i \boldsymbol{\Phi}_i \boldsymbol{\beta}_i \right\}, \quad (7.1)$$

where $\boldsymbol{\beta}_i = (\beta_i^1, \beta_i^2, \ldots, \beta_i^H)^T$ is the vector of the same parameter β_i in the H vertical regression models, $\boldsymbol{\Phi}_i$ is $(H) \times (H)$ positive definite distance matrices with 1s on the diagonal, and $\lambda_i > 0$ are scalars. For $\lambda_i = 0$ for all i, this minimization problem is equivalent to the one in 3.8. Also, for $\boldsymbol{\Phi}_i$ equal to the identity matrix, this minimization problem is equivalent to using separate ridge regressions to estimate synthetic control weights.

We consider $\boldsymbol{\Phi}_i$ matrices that are not diagonal. Following Tibshirani et al. (2005), we could specify the penalization matrix $\boldsymbol{\Phi} = diag[k_i(1)_{i=2}^N]$ where $k_i(1)$ is the 1-d penalization matrix

$$k_i(1) = \begin{pmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \cdots & & \\ & & & & 1 & -1 \end{pmatrix}.$$

Fused estimators (Tibshirani et al., 2005, Irie, 2019) can represent a viable solution for a smoothed estimation. This class of estimators, commonly applied to spatial structures and image detection problems, penalizes the distance between coefficients, allowing them to vary smoothly. The majority of previous applications in literature follows an l^1-penalty, specifying a lasso penalization, while few work has been previously publicized for l^2-penalty terms as in ridge estimators, e.g. Goeman (2008) or van Wieringen (2015). Recently, Obakrim et al. (2022) use the generalized ridge to predict outcomes when the covariates exhibit a spatial structure. Note that with this penalization matrix, all control unit coefficients are equally penalized. Other orders of penalization are possible, see Tibshirani and Taylor (2011) and van Wieringen (2015) for further examples.

7.2 Generalized Ridge and Gaussian process

In this section we show the similarities across these two estimators. In the two previous sections we have presented Spatial matrix completion methods, under a frequentist approach with a Generalized ridge regression and in a Bayesian framework, using a Gaussian process prior for the coefficients. It is worth noting that, under some assumptions and specifications, these two models are equivalent in term of point estimates, which is the focus of this work.

let be

$$p(y|\mathbf{X},\boldsymbol{\beta}) = \prod_{i=1}^{N} p(y_i|\mathbf{X}_i,\boldsymbol{\beta}) =$$
$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{(y_i - \mathbf{X}_i^T\boldsymbol{\beta})^2}{2\sigma_y}\right)$$
$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_y^n} \exp\left(-\frac{(y_i - \mathbf{X}_i^T\boldsymbol{\beta})^2}{2\sigma_y}\right) \sim \mathcal{N}(\mathbf{X}_i^T\boldsymbol{\beta}, \sigma_y \mathbf{I})$$

Let also be $\boldsymbol{\beta} \sim \mathcal{GP}(\mathbf{0}, \alpha\Sigma_\theta)$, the posterior distribution of $\boldsymbol{\beta}$ will result as $p(\boldsymbol{\beta}|y,\mathbf{X}) = \dfrac{p(y|\mathbf{X},\boldsymbol{\beta})p(\boldsymbol{\beta})}{p(y|\mathbf{X})}$. $p(y|\mathbf{X})$ is a constant term so, $p(\boldsymbol{\beta}|y,\mathbf{X}) \propto p(y|\mathbf{X},\boldsymbol{\beta})p(\boldsymbol{\beta})$. Thus,

$$p(\boldsymbol{\beta}|\mathbf{Y},\mathbf{X}) \propto \exp\left(-\frac{1}{2\sigma_y^2}(\mathbf{Y} - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}_i^T\boldsymbol{\beta})\right) \exp\left(-\frac{1}{2\alpha}(\boldsymbol{\beta} - 0)^T\Sigma^{-1}(\boldsymbol{\beta} - 0)\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma_{\mathbf{Y}}^2}(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta}) - \frac{1}{2\alpha}(\boldsymbol{\beta} - 0)^T\Sigma^{-1}(\boldsymbol{\beta} - 0)\right)$$
$$\propto \exp\left(-(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta}) - \frac{\sigma_{\mathbf{Y}}}{\alpha}\boldsymbol{\beta}^T\alpha\Sigma^{-1}\boldsymbol{\beta}\right)$$

Thus, we can maximise it with $\lambda = \frac{\sigma_{\mathbf{Y}}}{\alpha}$ and $\boldsymbol{\Phi} = \Sigma^{-1}$, as it corresponds to solving equation 7.1.

$$\underset{\boldsymbol{\beta}_0,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_N\in\mathbb{R}^{N\times H}}{\operatorname{argmax}} \exp\left(-(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta}) - \lambda\boldsymbol{\beta}^T\boldsymbol{\Phi}\boldsymbol{\beta}\right)$$
$$\underset{\boldsymbol{\beta}\in\mathbb{R}^{N\times H}}{\operatorname{argmax}}\left(-(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta}) - \lambda\boldsymbol{\beta}^T\boldsymbol{\Phi}\boldsymbol{\beta}\right)$$
$$\underset{\boldsymbol{\beta}\in\mathbb{R}^{N\times H}}{\operatorname{argmin}}(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta})^T(\mathbf{Y}_i - \mathbf{X}_i^T\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\Phi}\boldsymbol{\beta}$$
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} + \lambda\Phi)\mathbf{X}^T\mathbf{Y}$$

which is the solution to the maximisation problem in 7.1 and the solution for generalized ridge with Tikhonov matrix $\boldsymbol{\Phi}$.

7.3 Figures

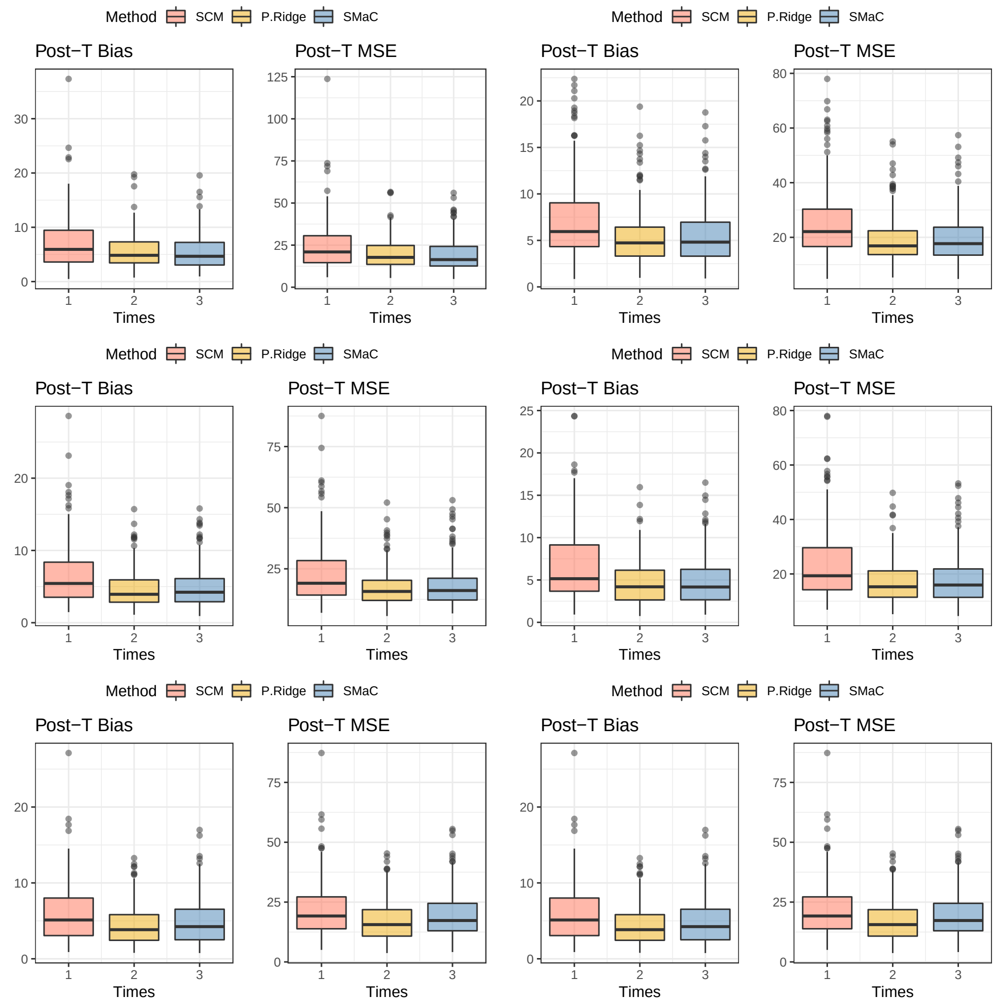


Figure 7.1: Post-Treatment bias and MSE for the selected estimators - HPP DGP. first row: T=20, second row: T=50, third row: T=100, first and second column: C=T*0.5, third and last column: C=T*1.5

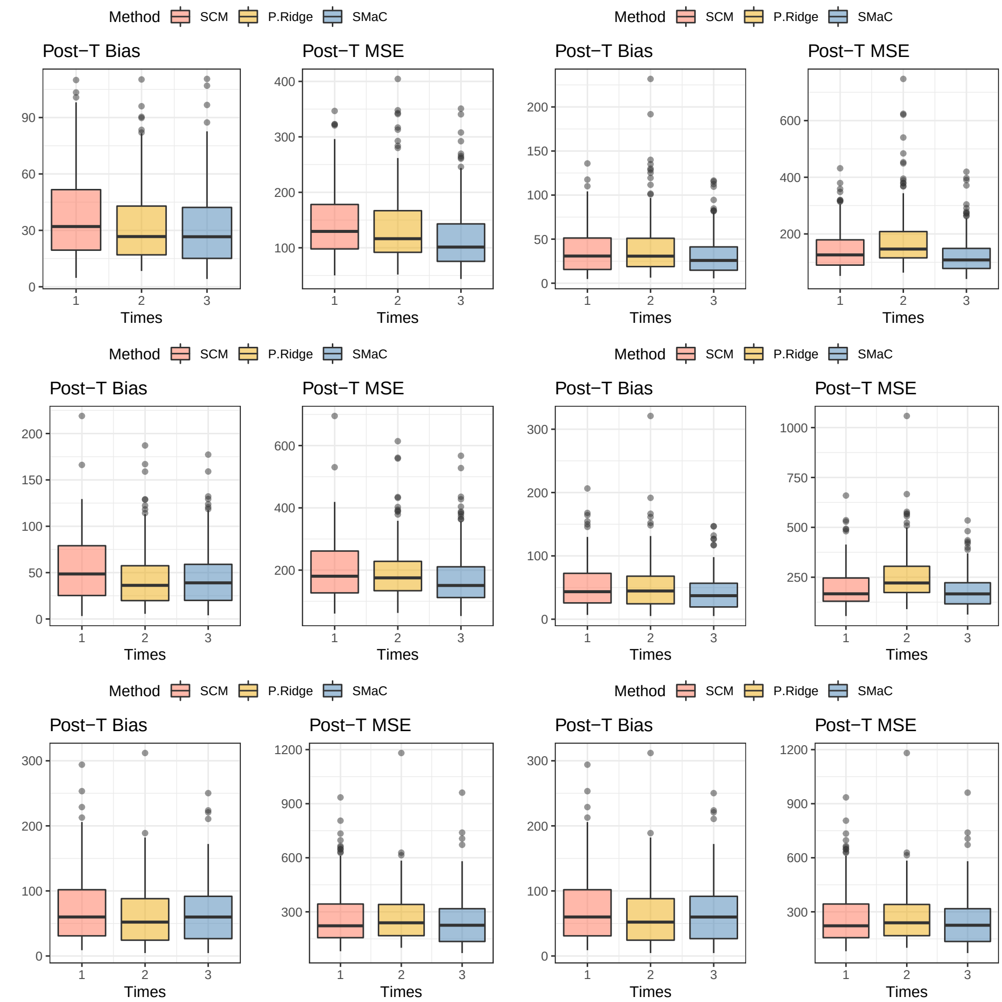


Figure 7.2: Post-Treatment bias and MSE for the selected estimators - Additive Linear DGP. first row: T=20, second row: T=50, third row: T=100, first and second column: C=T*0.5, third and last column: C=T*1.5

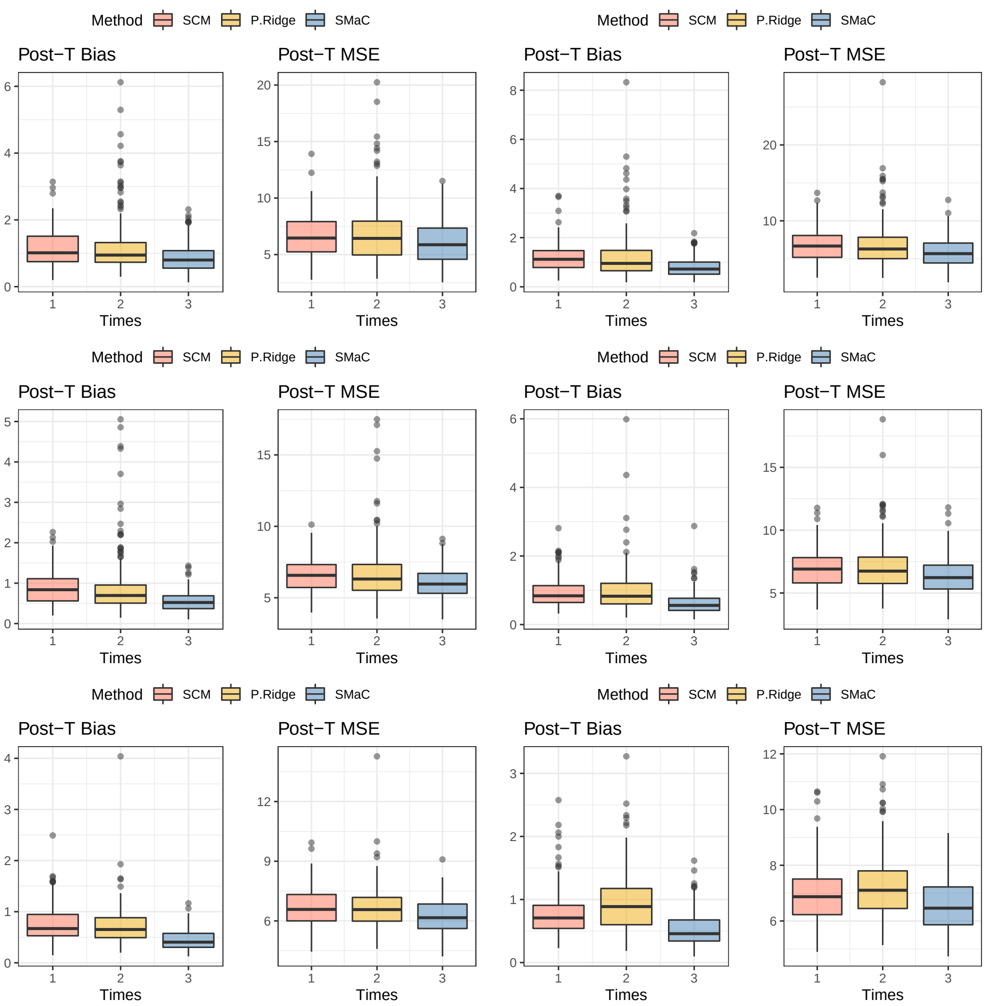


Figure 7.3: Post-Treatment bias and MSE for the selected estimators - Gaussian Process DGP. first row: T=20, second row: T:50, third row: T=100, first and second column: C=T*0.5, third and last column: C=T*1.5

Chapter 8

Appendix - Chapter 4

8.1 Priors specification

In this subsection of the appendix we report the prior specifications we used in our analysis. We assume prior distributions are a-prior independent.

Let $\mathbf{0}_{\boldsymbol{k}}$ be a k-dimensional vector of zeros. Similarly, let be $\mathbf{I_k}$ be a identity matrix with dimensions $k \times k$.

- Priors for the strata membership model

$$
\begin{aligned}
&\delta_{0,1} = \delta_{0,2} = \delta_{0,3} \sim \mathcal{N}(0,2)\\
&\boldsymbol{\delta_1} = \boldsymbol{\delta_2} = \boldsymbol{\delta_3} \sim \mathcal{N}_k(\mathbf{0}_{\boldsymbol{k}}, \sigma \mathbf{I_k})\\
&\gamma_1^{AS} = \gamma_1^{CS} \sim \mathcal{N}(0,2)\\
&\gamma_2^{(AS|G_{i,1}=AS)} = \gamma_2^{(CS|G_{i,1}=AS)} \sim \mathcal{N}(0,2)\\
&\gamma_2^{(CS|G_{i,1}=CS)} \sim \mathcal{N}(0,2)\\
&\gamma_3^{(AS|G_{i,1}=G_{i,2}=AS)} = \gamma_3^{(CS|G_{i,1}=G_{i,2}=AS)} \sim \mathcal{N}(0,2)\\
&\gamma_3^{(CS|G_{i,1}=AS,G_{i,2}=CS)} \sim \mathcal{N}(0,2)\\
&\gamma_3^{(CS|G_{i,1}=CS,G_{i,2}=CS)} \sim \mathcal{N}(0,2)
\end{aligned}
$$

- Priors for the strata membership model

$$
\begin{aligned}
&\beta_{0,1} = \beta_{0,2} = \beta_{0,3} \sim \mathcal{N}(0,2)\\
&\boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \boldsymbol{\beta_3} \sim \mathcal{N}_k(\mathbf{0}_{\boldsymbol{k}}, \sigma \mathbf{I_k})
\end{aligned}
$$

$$\lambda_2^{(W_i=1,G_{i,1}=AS)} = \lambda_2^{(W_i=0,G_{i,1}=AS)} \sim \mathcal{N}(0,2)$$
$$\lambda_2^{(W_i=1,G_{i,1}=CS)} \sim \mathcal{N}(0,2)$$
$$\lambda_3^{(W_i=1,G_{i,1}=G_{i,2}=AS)} = \lambda_3^{(W_i=0,G_{i,1}=G_{i,2}=AS)} \sim \mathcal{N}(0,2)$$
$$\lambda_3^{(W_i=1,G_{i,1}=AS,G_{i,2}=CS)} \sim \mathcal{N}(0,2)$$
$$\lambda_3^{(W_i=1,G_{i,1}=CS,G_{i,2}=CS)} \sim \mathcal{N}(0,2)$$
$$\alpha_1^{W_i=1,G_{i,1}=AS} = \alpha_1^{W_i=0,G_{i,1}=AS} \sim \mathcal{N}(0,2)$$
$$\alpha_1^{W_i=1,G_{i,1}=CS} = \alpha_1^{W_i=0,G_{i,1}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_1^{W_i=1,G_{i,1}=NS} = \alpha_1^{W_i=0,G_{i,1}=NS} \sim \mathcal{N}(0,2)$$
$$\alpha_2^{W_i=1,G_{i,1}=AS,G_{i,2}=AS} = \alpha_2^{W_i=0,G_{i,1}=AS,G_{i,2}=AS} \sim \mathcal{N}(0,2)$$
$$\alpha_2^{W_i=1,G_{i,1}=AS,G_{i,2}=CS} = \alpha_2^{W_i=0,G_{i,1}=AS,G_{i,2}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_2^{W_i=1,G_{i,1}=AS,G_{i,2}=NS} = \alpha_2^{W_i=0,G_{i,1}=AS,G_{i,2}=NS} \sim \mathcal{N}(0,2)$$
$$\alpha_2^{W_i=1,G_{i,1}=CS,G_{i,2}=CS} = \alpha_2^{W_i=0,G_{i,1}=CS,G_{i,2}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_2^{W_i=1,G_{i,1}=CS,G_{i,2}=NS} = \alpha_2^{W_i=0,G_{i,1}=CS,G_{i,2}=NS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=AS} = \alpha_3^{W_i=0,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=AS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=CS} = \alpha_3^{W_i=0,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=NS} = \alpha_3^{W_i=0,G_{i,1}=AS,G_{i,2}=AS,G_{i,3}=NS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=AS,G_{i,2}=CS,G_{i,3}=CS} = \alpha_3^{W_i=0,G_{i,1}=AS,G_{i,2}=CS,G_{i,3}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=CS,G_{i,2}=CS,G_{i,3}=CS} = \alpha_3^{W_i=0,G_{i,1}=CS,G_{i,2}=CS,G_{i,3}=CS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=AS,G_{i,2}=CS,G_{i,3}=NS} = \alpha_3^{W_i=0,G_{i,1}=AS,G_{i,2}=CS,G_{i,3}=NS} \sim \mathcal{N}(0,2)$$
$$\alpha_3^{W_i=1,G_{i,1}=CS,G_{i,2}=CS,G_{i,3}=NS} = \alpha_3^{W_i=0,G_{i,1}=CS,G_{i,2}=CS,G_{i,3}=NS} \sim \mathcal{N}(0,2)$$

8.2 Tables

W	**S1**	**S2**	**S3**	**Y1**	**Y2**	**Y3**
0	0	*	*	*	*	*
1	0	*	*	0.257	*	*
0	1	0	*	0.047	*	*
1	1	0	*	0.303	*	*
0	1	1	0	0.125	0.125	*
1	1	1	0	0.309	0.186	*
0	1	1	1	0.244	0.193	0.146
1	1	1	1	0.398	0.302	0.194

Table 8.1: Observed outcomes

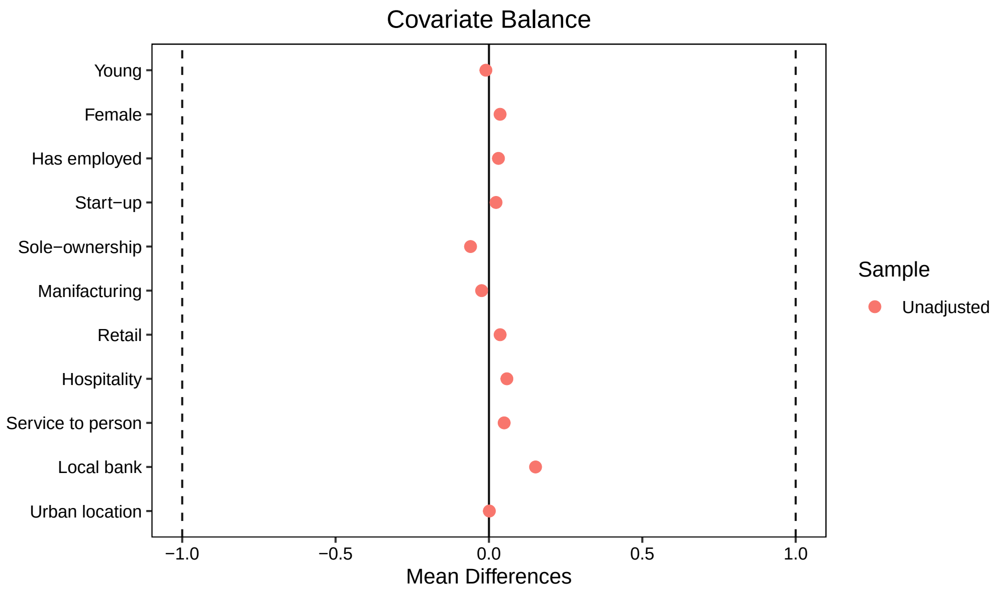


Figure 8.1: Covariate Balance

8.3 Figures

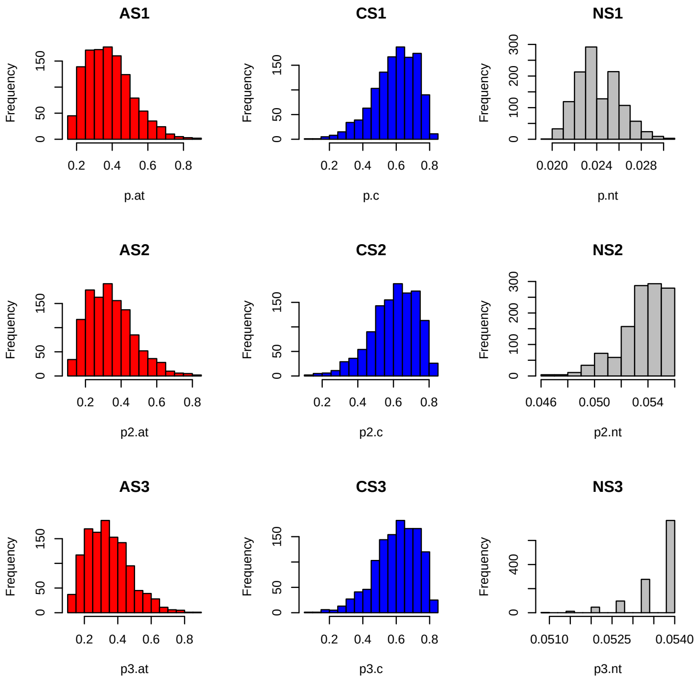


Figure 8.2: Posterior probability for principal strata membership in each post-treatment period

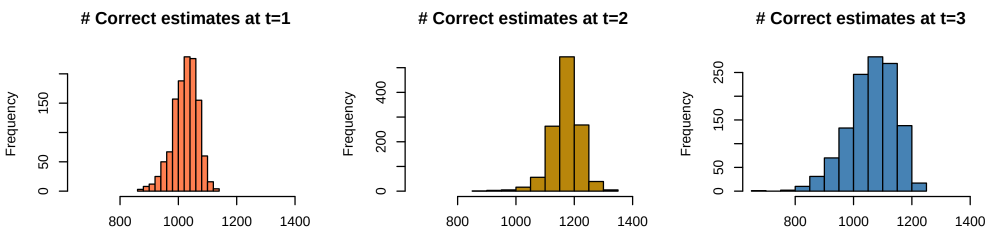


Figure 8.3: Number of corrected estimated outcomes for hiring decision in each period $t \in 1, 2, 3$ over the HMC iterations

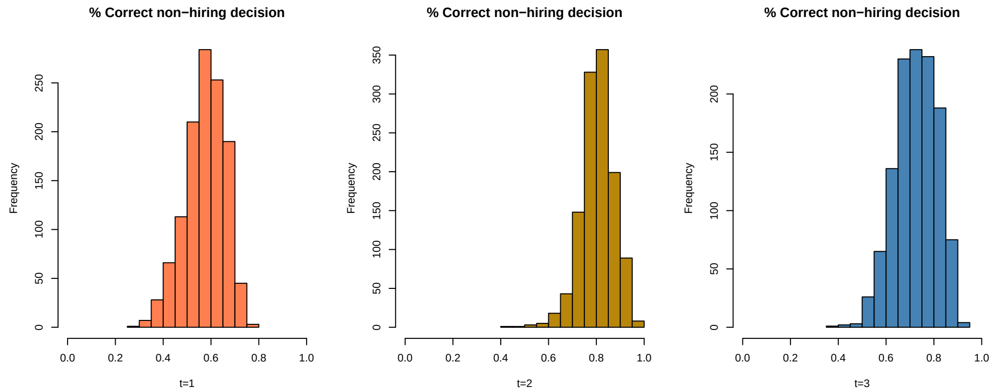


Figure 8.4: % of correct predictions for non-hiring decisions

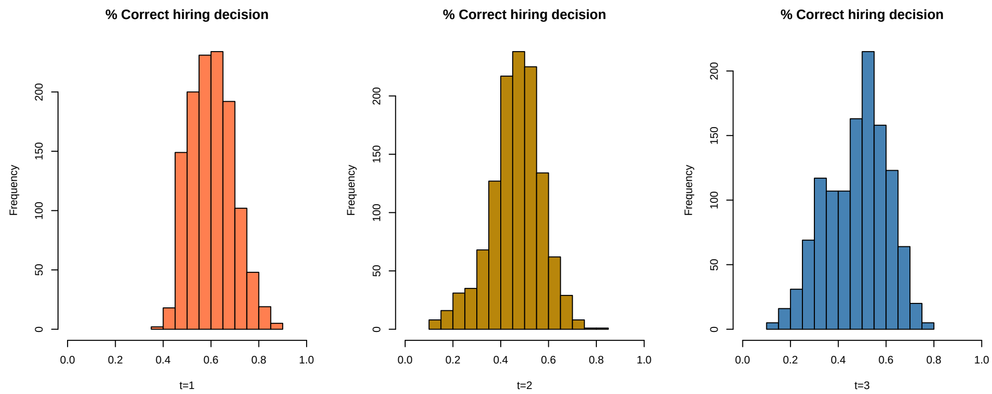


Figure 8.5: % of correct predictions for hiring decisions

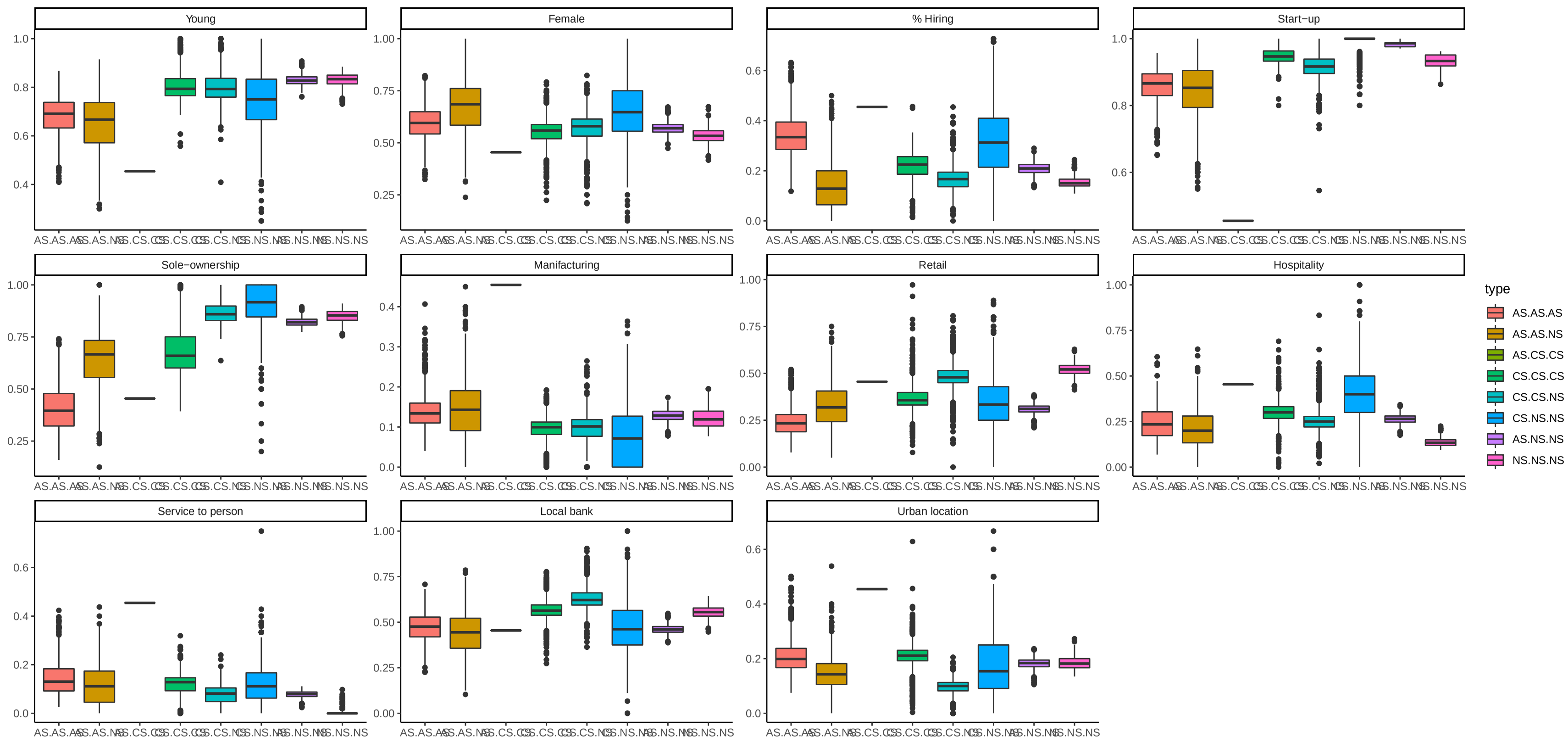


Figure 8.6: Boxplots of covariates across the longitudinal strata

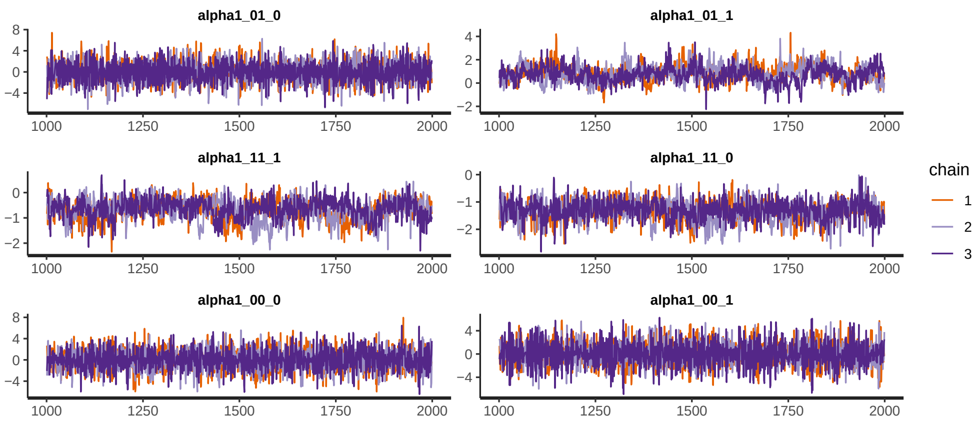


Figure 8.7: HMC traceplots for $\alpha_1^{W_i,G_i}$

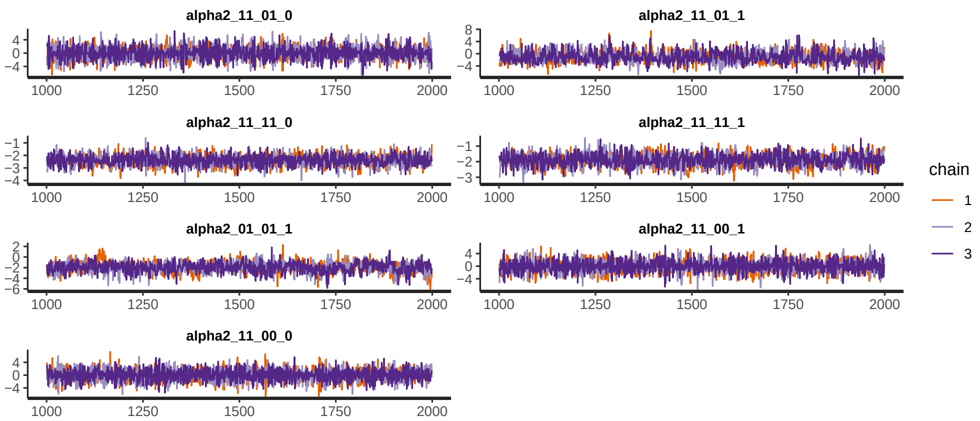


Figure 8.8: HMC traceplots for $\alpha_2^{W_i,G_i}$

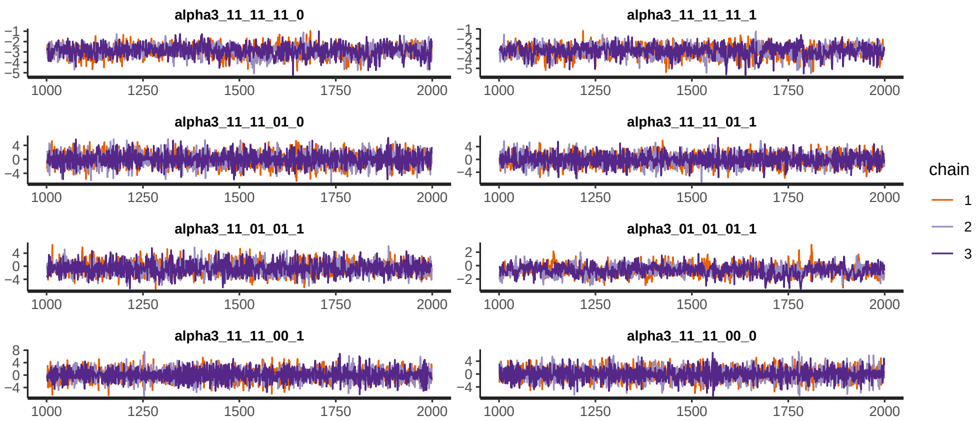


Figure 8.9: HMC traceplots for $\alpha_3^{W_i,G_i}$

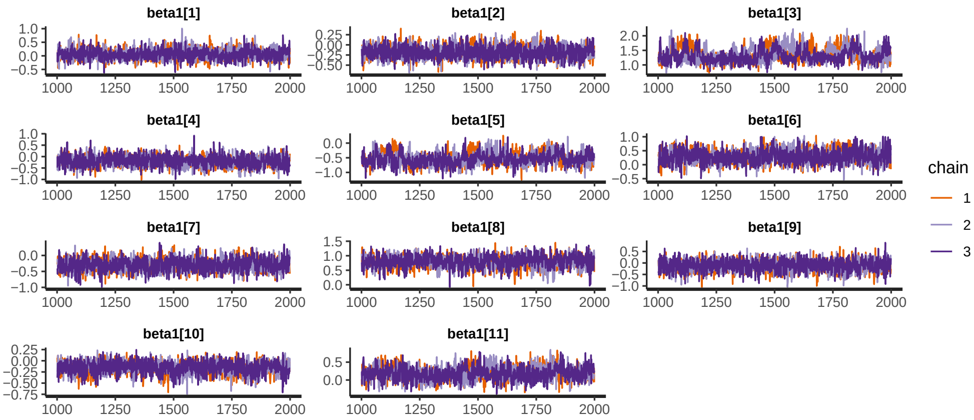


Figure 8.10: HMC traceplots for β_1

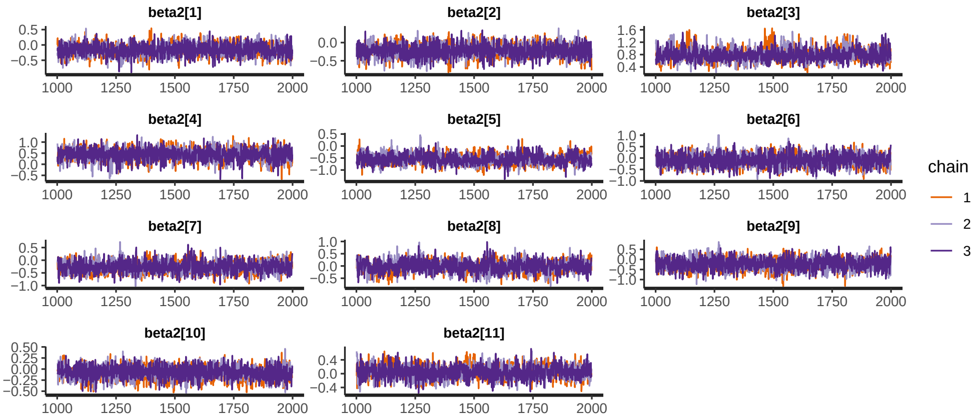


Figure 8.11: HMC traceplots for β_2

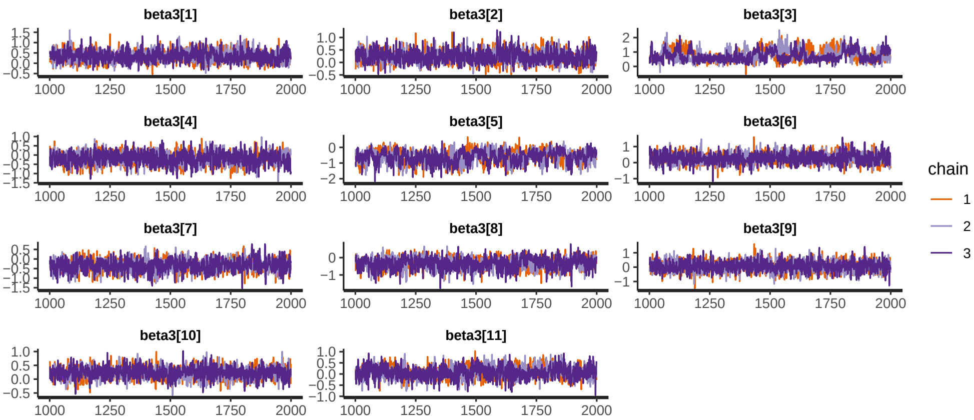


Figure 8.12: HMC traceplots for β_3

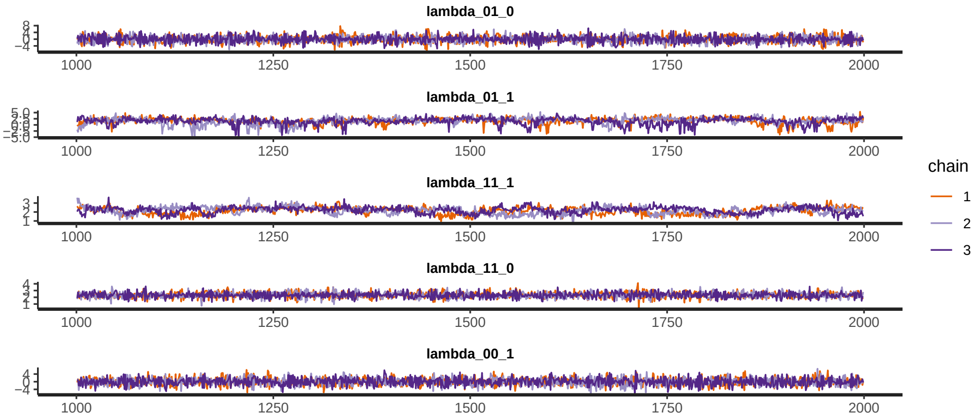


Figure 8.13: HMC traceplots for $\lambda_2^{W_i,G_{i,t-1}}$

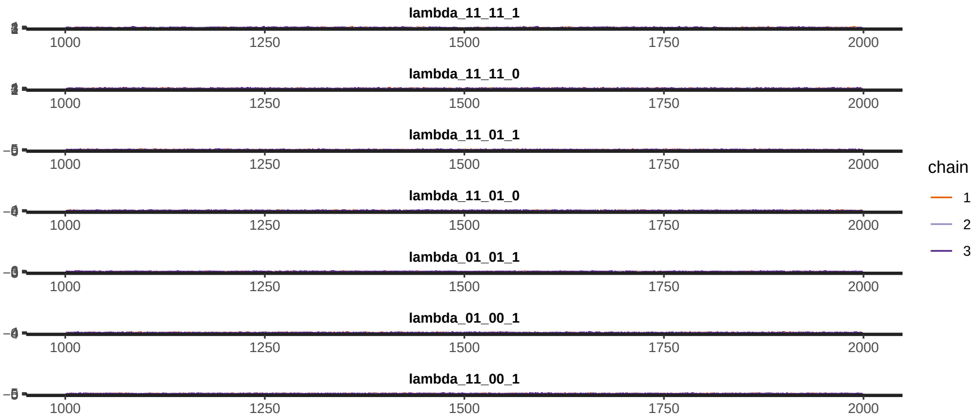


Figure 8.14: HMC traceplots for $\lambda_3^{W_i,G_{i,t-1}}$

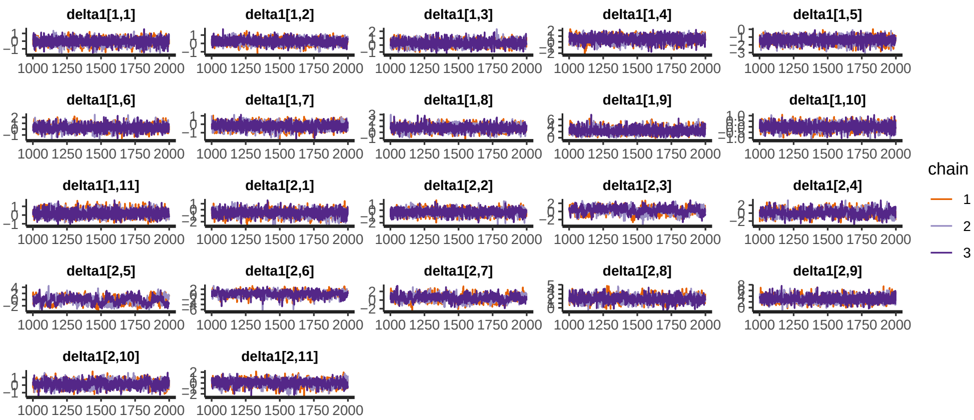


Figure 8.15: HMC traceplots for δ

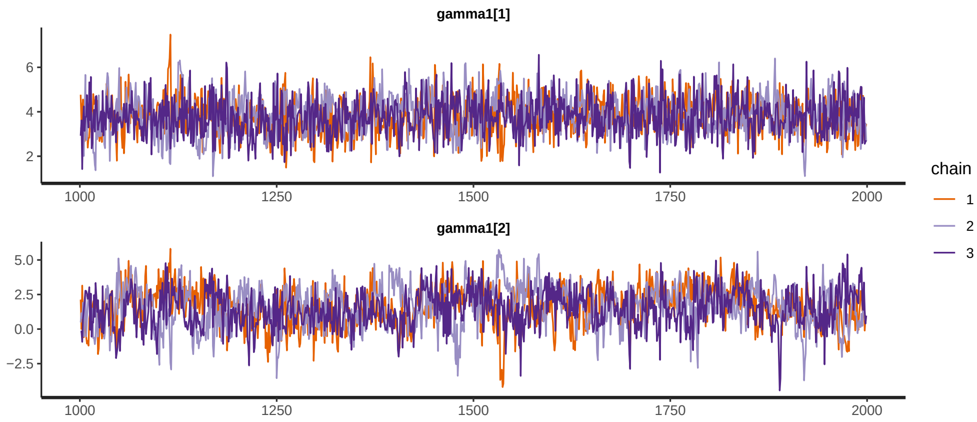


Figure 8.16: HMC traceplots for γ_1

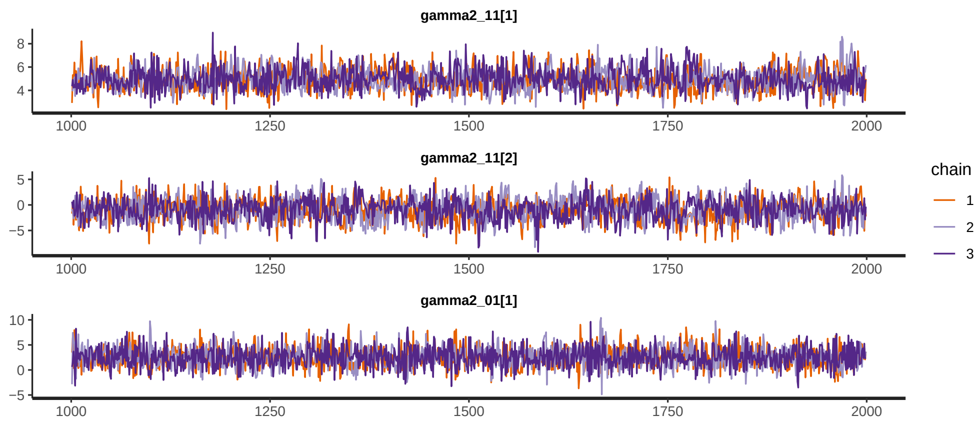


Figure 8.17: HMC traceplots for γ_2

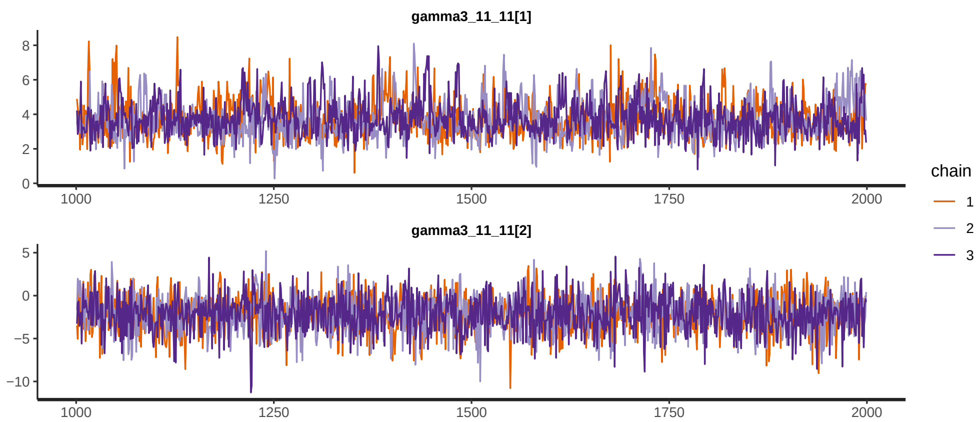


Figure 8.18: HMC traceplots for $\gamma_3^{G_{i,2}=AS}$

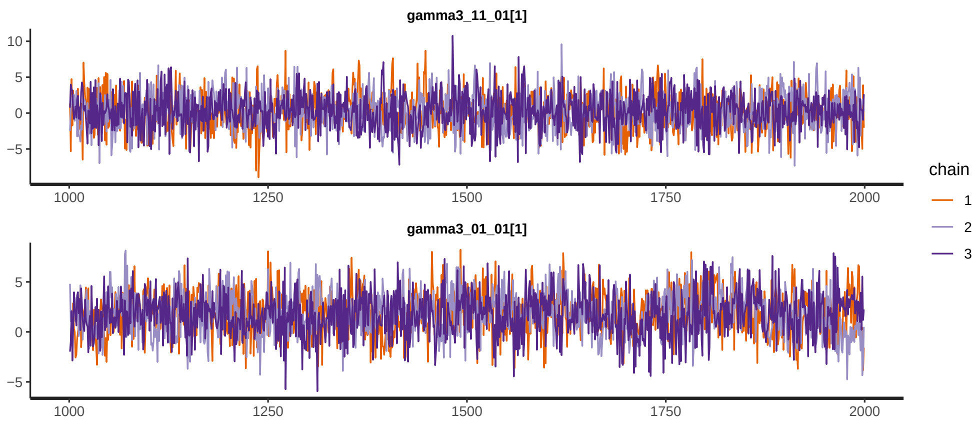


Figure 8.19: HMC traceplots for $\gamma_3^{G_{i,2}=CS}$

Bibliography

AB, N. K. S. (2021). Impact of vax-a-million lottery on covid-19 vaccination rates in ohio. *The American Journal of Medicine*.

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10(1).

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *econometrica*, 74(1):235–267.

Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121(2):368–391.

Acharya, B. and Dhakal, C. (2021). Implementation of state vaccine incentive lottery programs and uptake of covid-19 vaccinations in the united states. *JAMA Network Open*, 4(12):e2138238–e2138238.

Agarwal, A., Shah, D., and Shen, D. (2020). Synthetic interventions. *arXiv preprint arXiv:2006.07691*.

Aghion, P., Blundell, R., Griffith, R., Howitt, P., and Prantl, S. (2009). The effects of entry on incumbent innovation and productivity. *The Review of Economics and Statistics*, 91(1):20–32.

Alaa, A. M. and Van Der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task gaussian processes. *Advances in neural information processing systems*, 30.

Alonso-Nuez, M. J. and Galve-Górriz, C. (2012). The impact of public programs on the survival and profits of startups: Evidence from a region of spain. *Journal of Developmental Entrepreneurship*, 17(02):1250010.

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852.

Ando, M. and Sävje, F. (2013). Hypothesis testing with the synthetic control method. In *European Economic Association and Econometric Society Meeting*, pages 1–35. Unpublished working paper Gothenburg Sweden.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, volume 3, pages 1277–1366. Elsevier.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.

Arbour, D., Ben-Michael, E., Feller, A., Franks, A., and Raphael, S. (2021). Using multitask gaussian processes to estimate the effect of a targeted effort to remove firearms. *arXiv preprint arXiv:2110.07006*.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2019). Synthetic difference in differences. Technical report, National Bureau of Economic Research.

Arkhangelsky, D. and Imbens, G. W. (2022). Doubly robust identification for causal panel data models. *The Econometrics Journal*, 25(3):649–674.

Arpino, B. and Mattei, A. (2016). Assessing the causal effects of financial aids to firms in tuscany allowing for interference. *The Annals of Applied Statistics*, 10(3):1170–1194.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32.

Athey, S. and Imbens, G. W. (2021). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89(1):133–161.

Audretsch, D., Colombelli, A., Grilli, L., Minola, T., and Rasmussen, E. (2020). Innovative start-ups and policy initiatives. *Research Policy*, 49(10):104027.

Azizi, F. S. M., Kew, Y., and Moy, F. M. (2017). Vaccine hesitancy among parents in a multi-ethnic country, malaysia. *Vaccine*, 35(22):2955–2961.

Baccini, M., Mattei, A., Mealli, F., Bertazzi, P. A., and Carugno, M. (2017). Assessing the short term impact of air pollution on mortality: a matching approach. *Environmental Health*, 16(1):1–12.

Badr, H., Zhang, X., Oluyomi, A., Woodard, L. D., Adepoju, O. E., Raza, S. A., and Amos, C. I. (2021). Overcoming covid-19 vaccine hesitancy: Insights from an online population-based survey in the united states. *Vaccines*, 9(10):1100.

Barber, A. and West, J. (2021). Conditional cash lotteries increase covid-19 vaccination rates. *Available at SSRN*.

Battistin, E., Gavosto, A., and Rettore, E. (2001). Why do subsidised firms survive longer? an evaluation of a program promoting youth entrepreneurship in italy. In *Econometric evaluation of labour market policies*, pages 153–181. Springer.

Baum-Snow, N. and Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*, 77(2):241–263.

Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803.

Ben-Michael, E., Feller, A., Rothstein, J., et al. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B*, 84(2):351–381.

Bertoncello, C., Ferro, A., Fonzo, M., Zanovello, S., Napoletano, G., Russo, F., Baldo, V., and Cocchio, S. (2020). Socioeconomic determinants in vaccine hesitancy and vaccine refusal in italy. *Vaccines*, 8(2):276.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Bia, M., Mattei, A., and Mercatanti, A. (2020). Assessing causal effects in a longitudinal observational study with "truncated" outcomes due to unemployment and nonignorable missing data. *Journal of Business & Economic Statistics*, pages 1–34.

Bojinov, I., Rambachan, A., and Shephard, N. (2021). Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics*, 12(4):1171–1196.

Bojinov, I. and Shephard, N. (2019). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*, 114(528):1665–1682.

Bonilla, E. V., Chai, K., and Williams, C. (2007). Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20.

Bottmer, L., Imbens, G., Spiess, J., and Warnick, M. (2021). A design-based perspective on synthetic control methods. *arXiv preprint arXiv:2101.09398*.

Bowes, D. R. and Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of urban Economics*, 50(1):1–25.

Boyer, T. and Blazy, R. (2014). Born to be alive? the survival of innovative and non-innovative french micro-start-ups. *Small Business Economics*, 42(4):669–683.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, pages 247–274.

Brown, J. D., Earle, J. S., Kim, M. J., and Lee, K. M. (2019). Start-ups, job creation, and founder characteristics. *Industrial and Corporate Change*, 28(6):1637–1672.

Brown, S. (1993). Retail location theory: evolution and evaluation. *International Review of Retail, Distribution and Consumer Research*, 3(2):185–229.

Budiakivska, V. and Casolaro, L. (2018). Please in my back yard: the private and public benefits of a new tram line in florence. *Bank of Italy Temi di Discussione (Working Paper) No*, 1161.

Caliendo, M. (2016). Start-up subsidies for the unemployed: Opportunities and limitations. *IZA World of Labor*, (200).

Caliendo, M. and Kritikos, A. S. (2010). Start-ups by the unemployed: characteristics, survival and direct employment effects. *Small Business Economics*, 35(1):71–92.

Caliendo, M. and Künn, S. (2014). Regional effect heterogeneity of start-up subsidies for the unemployed. *Regional Studies*, 48(6):1108–1134.

Caliendo, M., Künn, S., and Weissenberger, M. (2020). Catching up or lagging behind? the long-term business and innovation potential of subsidized start-ups out of unemployment. *Research Policy*, 49(10):104053.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., and Wengström, E. (2021). Monetary incentives increase covid-19 vaccinations. *Science*, 374(6569):879–882.

Cao, J. and Dowd, C. (2019). Estimation and inference for synthetic control methods with spillover effects. *arXiv preprint arXiv:1902.07343*.

Card, D. and Krueger, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania.

Cattaneo, M. D., Feng, Y., and Titiunik, R. (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536):1865–1880.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95(5):1549–1561.

Cervero, R. and Landis, J. (1993). Assessing the impacts of urban rail transit on local real estate markets using quasi-experimental comparisons. *Transportation Research Part A: Policy and Practice*, 27(1):13–22.

Chiba, Y. and VanderWeele, T. J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *American journal of epidemiology*, 173(7):745–751.

Choi, D. S., Sung, C. S., and Park, J. Y. (2020). How does technology startups increase innovative performance? the study of technology startups on innovation focusing on employment change in korea. *Sustainability*, 12(2):551.

Colombo, M. G. and Grilli, L. (2007). Funding gaps? access to bank loans by high-tech start-ups. *Small Business Economics*, 29(1):25–46.

Credit, K. (2018). Transit-oriented economic development: The impact of light rail on new business starts in the phoenix, az region, usa. *Urban Studies*, 55(13):2838–2862.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Number 1. Cambridge university press.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

Delmelle, E. and Nilsson, I. (2020). New rail transit stations and the out-migration of low-income residents. *Urban Studies*, 57(1):134–151.

Di Stefano, R. and Mellace, G. (2020). The inclusive synthetic control method. *Discussion Papers on Business and Economics, University of Southern Denmark*, 14.

Donohue, J. J., Aneja, A., and Weber, K. D. (2019). Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, 16(2):198–247.

Dotlic, J., Stojkovic, V. J., Cummins, P., Milic, M., and Gazibara, T. (2021). Enhancing covid-19 vaccination coverage using financial incentives: arguments to help health providers counterbalance erroneous claims. *Epidemiology and Health*, 43.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research.

Dube, A. and Zipperer, B. (2015). Pooling multiple case studies using synthetic controls: An application to minimum wage policies. *Available at SSRN 2589786*.

Dubé, E., Gagnon, D., Nickels, E., Jeram, S., and Schuster, M. (2014). Mapping vaccine hesitancy—country-specific characteristics of a global phenomenon. *Vaccine*, 32(49):6649–6654.

Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., and Bettinger, J. A. (2013). Vaccine hesitancy: an overview. *Human vaccines & immunotherapeutics*, 9(8):1763–1773.

Dubé, E., Vivion, M., and MacDonald, N. E. (2015). Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert review of vaccines*, 14(1):99–117.

Duhautois, R., Redor, D., and Desiage, L. (2015). Long term effect of public subsidies on start-up survival and economic performance: An empirical study with french data. *Revue d'économie industrielle*, (149):11–41.

Dumont, M., Rayp, G., Verschelde, M., and Merlevede, B. (2016). The contribution of start-ups and young firms to industry-level efficiency growth. *Applied Economics*, 48(59):5786–5801.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397):171–185.

Engin, C. and Vezzoni, C. (2020). Who's skeptical of vaccines? prevalence and determinants of anti-vaccination attitudes in italy. *Population Review*, 59(2).

Featherstone, J. D., Bell, R. A., and Ruiz, J. B. (2019). Relationship of people's sources of health information and political ideology with acceptance of conspiratorial beliefs about vaccines. *Vaccine*, 37(23):2993–2997.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).

Forastiere, L., Airoldi, E. M., and Mealli, F. (2021a). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918.

Forastiere, L., Lattarulo, P., Mariani, M., Mealli, F., and Razzolini, L. (2021b). Exploring encouragement, treatment, and spillover effects using principal stratification, with application to a field experiment on teens' museum attendance. *Journal of Business & Economic Statistics*, 39(1):244–258.

Forastiere, L., Mealli, F., and VanderWeele, T. J. (2016). Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using bayesian principal stratification. *Journal of the American Statistical Association*, 111(514):510–525.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Frumento, P., Mealli, F., Pacini, B., and Rubin, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *Journal of the American Statistical Association*, 107(498):450–466.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Goeman, J. J. (2008). Autocorrelated logistic ridge regression for prediction based on proteomics spectra. *Statistical Applications in Genetics and Molecular Biology*, 7(2).

Gorin, M. and Schmidt, H. (2015). 'i did it for the money': incentives, rationalizations and health. *Public Health Ethics*, 8(1):34–41.

Gowda, C. and Dempsey, A. F. (2013). The rise (and fall?) of parental vaccine hesitancy. *Human vaccines & immunotherapeutics*, 9(8):1755–1762.

Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. Chapman and Hall/CRC.

Grossi, G., Lattarulo, P., Mariani, M., Mattei, A., and Öner, Ö. (2020). Synthetic control group methods in the presence of interference: The direct and spillover effects of light rail on neighborhood retail activity. *arXiv preprint arXiv:2004.05027*.

Grube-Cavers, A. and Patterson, Z. (2015). Urban rapid rail transit and gentrification in canadian urban centres: A survival analysis approach. *Urban Studies*, 52(1):178–194.

Guerra, E. and Cervero, R. (2013). Is a half-mile circle the right standard for tods?

Guerra, E., Cervero, R., and Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation Research Record*, 2276(1):101–109.

Gustafson, P. (2010). Bayesian inference for partially identified models. *The international journal of biostatistics*, 6(2).

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pages 475–492. NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161.

Heckman, J. J., Ichimura, H., Smith, J. A., and Todd, P. E. (1998). Characterizing selection bias using experimental data.

Hess, D. B. and Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in buffalo, new york. *Urban studies*, 44(5-6):1041–1068.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

Hong, G. and Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101(475):901–910.

Huang, B., Chen, C., and Liu, J. (2019). Gpmatch: A bayesian doubly robust approach to causal inference with gaussian process covariance function as a matching tool. *arXiv preprint arXiv:1901.10359*.

Huber, M. and Steinmayr, A. (2021). A framework for separating individual-level treatment effects from spillover effects. *Journal of Business & Economic Statistics*, 39(2):422–436.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.

Imbens, G. W. and Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4):701–712.

Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The annals of statistics*, pages 305–327.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Irie, K. (2019). Bayesian dynamic fused lasso. *arXiv preprint arXiv:1905.12275*.

Jarrett, C., Wilson, R., O'Leary, M., Eckersberger, E., Larson, H. J., et al. (2015). Strategies for addressing vaccine hesitancy–a systematic review. *Vaccine*, 33(34):4180–4190.

Jecker, N. S. (2021). Cash incentives, ethics, and covid-19 vaccination. *Science*, 374(6569):819–820.

Jin, H. and Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, 103(481):101–111.

Joshi, A., Kaur, M., Kaur, R., Grover, A., Nash, D., and El-Mohandes, A. (2021). Predictors of covid-19 vaccine acceptance, intention, and hesitancy: a scoping review. *Frontiers in Public Health*, 9.

Kahn, M. E. (2007). Gentrification trends in new transit-oriented communities: Evidence from 14 cities that expanded and built rail transit systems. *Real Estate Economics*, 35(2):155–182.

Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*.

Kane, T. J. (2010). The importance of startups in job creation and job destruction. *Available at SSRN 1646934*.

Kellogg, M., Mogstad, M., Pouliot, G. A., and Torgovitsky, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, 116(536):1804–1816.

Kim, H. B. (2021). Financial incentives for covid-19 vaccination. *Epidemiology and Health*, 43.

Kim, S., Lee, C., and Gupta, S. (2020). Bayesian synthetic control methods. *Journal of Marketing Research*, 57(5):831–852.

Kirkegaard, E. O. (2016). Inequality across us counties: an s factor analysis. *Open Quantitative Sociology & Political Science*.

Klaesson, J. and Öner, Ö. (2014). Market reach for retail services. *Review of Regional Studies*, 44(2):153–176.

Korn, L., Böhm, R., Meier, N. W., and Betsch, C. (2020). Vaccination as a social contract. *Proceedings of the National Academy of Sciences*, 117(26):14890–14899.

Koski, H. and Pajarinen, M. (2013). The role of business subsidies in job creation of start-ups, gazelles and incumbents. *Small Business Economics*, 41(1):195–214.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics*, 25(12):1514–1528.

Krivoruchko, K. and Gribov, A. (2019). Evaluation of empirical bayesian kriging. *Spatial Statistics*, 32:100368.

Kuschel, K., Labra, J.-P., and Diaz, G. (2018). Women-led startups and their contribution to job creation. In *Technology Entrepreneurship*, pages 139–156. Springer.

Landis, J., Guhathakurta, S., Huang, W., Zhang, M., and Fukuji, B. (1995). Rail transit investments, real estate values, and land use change: A comparative analysis of five california rail transit systems.

Larsson, J. P. and Öner, Ö. (2014). Location and co-location in retail: a probabilistic approach using geo-coded data for metropolitan retail markets. *The Annals of Regional Science*, 52(2):385–408.

Latkin, C., Dayton, L. A., Yi, G., Konstantopoulos, A., Park, J., Maulsby, C., and Kong, X. (2021). Covid-19 vaccine intentions in the united states, a social-ecological framework. *Vaccine*, 39(16):2288–2294.

Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., and El-Mohandes, A. (2021). A global survey of potential acceptance of a covid-19 vaccine. *Nature medicine*, 27(2):225–228.

Leete, O. E., Kallus, N., Hudgens, M. G., Napravnik, S., and Kosorok, M. R. (2019). Balanced policy evaluation and learning for right censored data. *arXiv preprint arXiv:1911.05728.*

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., and Larson, H. J. (2021). Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature human behaviour*, 5(3):337–348.

Lukeš, M., Longo, M. C., and Zouhar, J. (2019). Do business incubators really enhance entrepreneurial growth? evidence from a large sample of innovative italian start-ups. *Technovation*, 82:25–34.

Malik, A. A., McFadden, S. M., Elharake, J., and Omer, S. B. (2020). Determinants of covid-19 vaccine acceptance in the us. *EClinicalMedicine*, 26:100495.

Manaresi, F., Menon, C., and Santoleri, P. (2021). Supporting innovative entrepreneurship: an evaluation of the italian "start-up act". *Industrial and Corporate Change*, 30(6):1591–1614.

Mariani, M., Mattei, A., Storchi, L., and Vignoli, D. (2019). The ambiguous effects of public assistance to youth and female start-ups between job creation and entrepreneurship enhancement. *Scienze Regionali*, 18(2):237–260.

Marks, J. H. (2020). Lessons from corporate influence in the opioid epidemic: toward a norm of separation. *Journal of Bioethical Inquiry*, 17(2):173–189.

Mattei, A. and Mealli, F. (2007). Application of the principal stratification approach to the faenza randomized experiment on breast self-examination. *Biometrics*, 63(2):437–446.

Mattei, A. and Mealli, F. (2011). Augmented designs to assess principal strata direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):729–752.

Mattei, A., Mealli, F., and Ding, P. (2020). Assessing causal effects in the presence of treatment switching through principal stratification. *arXiv preprint arXiv:2002.11989.*

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6:355–378.

Mealli, F. and Mattei, A. (2012). A refreshing account of principal stratification. *The international journal of biostatistics*, 8(1).

Mealli, F. and Pagni, R. (2001). Analisi e valutazione delle politiche per le nuove imprese. *Il caso della LR Toscana*, (27/93).

Mejia-Dorantes, L., Paez, A., and Vassallo, J. M. (2012). Transportation infrastructure impacts on firm location: the effect of a new metro line in the suburbs of madrid. *Journal of Transport Geography*, 22:236–250.

Menchetti, F. and Bojinov, I. (2020). Estimating causal effects in the presence of partial interference using multivariate bayesian structural time series models. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (21-048).

Mønsted, B. and Lehmann, S. (2022). Characterizing polarization in online vaccine discourse—a large-scale study. *PloS one*, 17(2):e0263746.

Nigam, N., Mbarek, S., and Boughanmi, A. (2020). Impact of intellectual capital on the financing of startups with new business models. *Journal of Knowledge Management*.

Obakrim, S., Ailliot, P., Monbet, V., and Raillard, N. (2022). Em algorithm for generalized ridge regression with spatial covariates. *arXiv preprint arXiv:2208.04754*.

Pagliara, F. and Papa, E. (2011). Urban rail systems investments: an analysis of the impacts on property values and residents' location. *Journal of Transport Geography*, 19(2):200–211.

Pan, Q. (2013). The impacts of an urban light rail system on residential property values: a case study of the houston metrorail transit line. *Transportation Planning and Technology*, 36(2):145–169.

Pang, X., Liu, L., and Xu, Y. (2022). A bayesian alternative to synthetic control for comparative case studies. *Political Analysis*, 30(2):269–288.

Papa, E. and Bertolini, L. (2015). Accessibility and transit-oriented development in european metropolitan areas. *Journal of Transport Geography*, 47:70–83.

Papadogeorgou, G., Mealli, F., and Zigler, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics*, 75(3):778–787.

Peneder, M. (2008). The problem of private under-investment in innovation: A policy mind map. *Technovation*, 28(8):518–530.

Persad, G. and Emanuel, E. J. (2021). Ethical considerations of offering benefits to covid-19 vaccine recipients. *JAMA*, 326(3):221–222.

Pfeiffer, F. and Reize, F. (2000). From unemployment to self-employment–public promotion and selectivity. *International Journal of Sociology*, 30(3):71–99.

Pinkney, S. (2021). An improved and extended bayesian synthetic control. *arXiv preprint arXiv:2103.16244*.

Pinotti, P. (2015). The economic costs of organised crime: Evidence from southern italy. *The Economic Journal*, 125(586):F203–F232.

Pogonyi, C. G., Graham, D. J., and Carbo, J. M. (2021). Metros, agglomeration and displacement. evidence from london. *Regional Science and Urban Economics*, 90:103681.

Quinn, S., Jamison, A., Musa, D., Hilyard, K., and Freimuth, V. (2016). Exploring the continuum of vaccine hesitancy between african american and white adults: results of a qualitative study. *PLoS currents*, 8.

Razai, M. S., Osama, T., McKechnie, D. G., and Majeed, A. (2021). Covid-19 vaccine hesitancy among ethnic minority groups.

Reiter, P. L., Pennell, M. L., and Katz, M. L. (2020). Acceptability of a covid-19 vaccine among adults in the united states: How many people would get vaccinated? *Vaccine*, 38(42):6500–6507.

Ricciardi, F., Mattei, A., and Mealli, F. (2020). Bayesian inference for sequential treatments under latent sequential ignorability. *Journal of the American Statistical Association*, 115(531):1498–1517.

Robertson, E., Reeve, K. S., Niedzwiedz, C. L., Moore, J., Blake, M., Green, M., Katikireddi, S. V., and Benzeval, M. J. (2021). Predictors of covid-19 vaccine hesitancy in the uk household longitudinal study. *Brain, behavior, and immunity*, 94:41–50.

Román, C., Congregado, E., and Millán, J. M. (2013). Start-up incentives: Entrepreneurship policy or active labour market programme? *Journal of Business Venturing*, 28(1):151–175.

Rosenbaum, P. R., Rosenbaum, P., and Briskman (2010). *Design of observational studies*, volume 10. Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593.

Rubin, D. B. et al. (2006). Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Statistical Science*, 21(3):299–309.

Sant'Anna, P. H. (2016). Program evaluation with right-censored data. *arXiv preprint arXiv:1604.02642*.

Savoia, E., Piltch-Loeb, R., Goldberg, B., Miller-Idriss, C., Hughes, B., Montrond, A., Kayyem, J., and Testa, M. A. (2021). Predictors of covid-19 vaccine hesitancy: socio-demographics, co-morbidity, and past experience of racial discrimination. *Vaccines*, 9(7):767.

Schuetz, J. (2015). Do rail transit stations encourage neighbourhood retail activity? *Urban Studies*, 52(14):2699–2723.

Shane, S. (2009). Why encouraging more people to become entrepreneurs is bad public policy. *Small business economics*, 33(2):141–149.

Shen, D., Ding, P., Sekhon, J., and Yu, B. (2022). Same root different leaves: Time series and cross-sectional methods in panel data. *arXiv preprint arXiv:2207.14481*.

Sills, E. O., Herrera, D., Kirkpatrick, A. J., Brandão Jr, A., Dickson, R., Hall, S., Pattanayak, S., Shoch, D., Vedoveto, M., Young, L., et al. (2015). Estimating the impacts of local policy innovation: the synthetic control method applied to tropical deforestation. *PloS one*, 10(7):e0132590.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407.

Sprengholz, P., Eitze, S., Felgendreff, L., Korn, L., and Betsch, C. (2021). Money is not everything: experimental evidence that payments do not increase willingness to be vaccinated against covid-19. *Journal of Medical Ethics*, 47(8):547–548.

Stijnen, T. (1982). Empirical bayes rules and gaussian processes. *Journal of Statistical Planning and Inference*, 6(4):363–372.

Taber, J. M., Thompson, C. A., Sidney, P. G., O'Brien, A., and Updegraff, J. (2021). Promoting vaccination with lottery incentives.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *The annals of statistics*, 39(3):1335–1371.

van Wieringen, W. N. (2015). Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.

Viviano, D. (2019). Policy targeting under network interference. *arXiv preprint arXiv:1906.10258*.

Walkey, A. J., Law, A., and Bosch, N. A. (2021). Lottery-based incentive in ohio and covid-19 vaccination rates. *JAMA*.

Weisel, O. (2021). Vaccination as a social contract: The case of covid-19 and us political partisanship. *Proceedings of the National Academy of Sciences*, 118(13).

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Williams, S. E. (2014). What are the factors that contribute to parental vaccine-hesitancy and what can we do about it? *Human vaccines & immunotherapeutics*, 10(9):2584–2596.

Willis, D. E., Andersen, J. A., Bryant-Moore, K., Selig, J. P., Long, C. R., Felix, H. C., Curran, G. M., and McElfish, P. A. (2021). Covid-19 vaccine hesitancy: Race/ethnicity, trust, and fear. *Clinical and translational science*, 14(6):2200–2207.

Witty, S., Takatsu, K., Jensen, D., and Mansinghka, V. (2020). Causal inference using gaussian processes with structured latent confounders. In *International Conference on Machine Learning*, pages 10313–10323. PMLR.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

Yan, S., Delmelle, E., and Duncan, M. (2012). The impact of a new light rail system on single-family property values in charlotte, north carolina. *Journal of Transport and Land Use*, 5(2):60–67.

Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by “death”. *Journal of Educational and Behavioral Statistics*, 28(4):353–368.

Zhang, J. L., Rubin, D. B., and Mealli, F. (2008). Evaluating the effects of job training programs on wages through principal stratification. In *Modelling and Evaluating Treatment Effects in Econometrics*. Emerald Group Publishing Limited.

Zigler, C., Forastiere, L., and Mealli, F. (2020). Bipartite interference and air pollution transport: Estimating health effects of power plant interventions. *arXiv preprint arXiv:2012.04831*.