Direct and spillover effects of a new tramway line on the commercial vitality of peripheral streets. A synthetic-control approach

Giulio Grossi[2], Marco Mariani[1], Alessandra Mattei[2], Patrizia Lattarulo[1], and Özge Öner[3]

[1]*IRPET - Istituto Regionale Programmazione Economica Toscana*
[2]*Department of Statistics, Computer Science, Applications, University of Florence*
[3]*Department of Land Economy, Cambridge University*

Abstract

In cities, the creation of public transport infrastructure such as light rails can cause changes on a very detailed spatial scale, with different stories unfolding next to each other within a same urban neighborhood. We study the direct effect of a light rail line built in Florence (Italy) on the retail density of the street where it was built and and its spillover effect on other streets in the treated street's neighborhood. To this aim, we investigate the use of the Synthetic Control Group (SCG) methods in panel comparative case studies where interference between the treated and the untreated units is plausible, an issue still little researched in the SCG methodological literature. We frame our discussion in the potential outcomes approach. Under a partial interference assumption, we formally define relevant direct and spillover causal effects. We also consider the "unrealized" spillover effect on the treated street in the hypothetical scenario that another street in the treated unit's neighborhood had been assigned to the intervention.

We wish to thank Fabrizia Mealli and Georgia Papadogeorgou for the insightful comments

1 Introduction

Synthetic Control Group (SCG) methods (Abadie & 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 & 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 & Magnac 2016, Kreif et al. 2016, Abadie & 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 & 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 & Raudenbush 2006, Sobel 2006, Hudgens & Halloran 2008, Arpino & Mattei 2016, Forastiere et al. 2018, Papadogeorgou et al. 2019, Huber & Steinmayr 2021). Despite such increasing interest, to the best of our knowledge, only the recent works by Cao & Dowd (2019) and Di Stefano & 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 & 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 & 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 paper 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. 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 (Sobel 2006), 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 & 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 & Landis 1993, Baum-Snow & Kahn 2000, Bowes & Ihlanfeldt 2001, Kahn 2007, Pagliara & Papa 2011, Grube-Cavers & Patterson 2015, Budiakivska & Casolaro 2018, Delmelle & 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 paper is organized as follows. Section 2 describes the application that motivates the methodological development we propose and the available data. Section 3 presents the methodology. In Section 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 5 concludes the paper.

2 Motivating application and related data

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 paper 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.

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 & 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 & Öner 2014, Larsson & Öner 2014).

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 6.3.1 in Web Supplementary Material 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.

3 Methodology

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 $t = 1,\dots,T$. Let $1+N_1$ and N_k be the number of units in cluster 1 and in cluster k, $k \in \{2,\dots,1+K\}$, 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, $k = 1,2,\dots,1+K$. For $k = 1,\dots,1+K$, let $\mathbf{w}_{kt} = [w_{ki,t}]'_{i\in\mathcal{N}_k}$ be a cluster treatment vector at time t, $t = 1,\dots,T$. Generally, for $t = 1,\dots,T$, $\mathbf{w}_{kt} \in \{0,1\}^{\mathbb{I}\{k=1\}+N_k}$, where $\mathbb{I}\{\cdot\}$ is the indicator function. In this paper 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 $t = 1,\dots,T_0$, $[\mathbf{w}_{1t},\dots,\mathbf{w}_{(1+K)t}] = [\mathbf{0}_{1+N_1},\dots,\mathbf{0}_{N_{1+K}}]$, where $\mathbf{0}_r$ denotes the zero vector in $\mathbb{R}^r$; and for $t = T_0+1,\dots,T$, $[\mathbf{w}_{1t},\dots,\mathbf{w}_{(1+K)t}]$ is constant over time and it is a point in $\mathcal{W} = \{[\mathbf{w}_1,\dots,\mathbf{w}_{(1+K)}]' \in \{0,1\}^{1+\sum_{k=1}^{1+K} N_k}$ with $\mathbf{w}_k = [w_{ik}]_{i\in\mathcal{N}_k}, k = 1,\dots,1+K : \sum_{k=1}^{1+K}\sum_{i\in\mathcal{N}_k} w_{ik} = 1\}$.

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 6.3.2 in the Web Supplementary Material for a stylized map.

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

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

denote the potential outcome for unit i in cluster k at time t under treatment assignment $(1 + N) \times T$ matrix $[\mathbf{w}_{k1}, \dots, \mathbf{w}_{kT_0}, \mathbf{w}_{k(T_0+1)}, \dots \mathbf{w}_{kT}]_{k=1}^{1+K}$, 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, $T_0 + 1, \dots, T$:

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

$$Y_{ki,t}([\mathbf{0}_{\mathbb{I}\{k=1\}+N_k}, \dots, \mathbf{0}_{\mathbb{I}\{k=1\}+N_k}, \mathbf{w}_{k(T_0+1)}, \dots \mathbf{w}_{kT}]_{k=1}^{1+K}) = Y_{ki,t}([\mathbf{0}_{\mathbb{I}\{k=1\}+N_k}, \dots, \mathbf{0}_{\mathbb{I}\{k=1\}+N_k}, \mathbf{0}_{\mathbb{I}\{k=1\}+N_k}, \dots \mathbf{0}_{\mathbb{I}\{k=1\}+N_k}]_{k=1}^{1+K})$$

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 paper.

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,

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

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

Let $[\mathbf{W}_1,\dots,\mathbf{W}_{1+K}]$ 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, $t=1,\dots,T_0,T_0+1,\dots,T$. Under consistency and no anticipation of treatment, $Y_{ik,t}=Y_{ki,t}([\mathbf{0}_{1+N_1},\dots,\mathbf{0}_{N_{1+K}}])$ for $t=1,\dots,T_0$ and $Y_{ik,t}=Y_{ki,t}([\mathbf{W}_1,\dots,\mathbf{W}_{1+K}])$, for $t=T_0+1,\dots,T$.

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:

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

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

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

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

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

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

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

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 $Y_{ki,t}([\mathbf{w}_1, \ldots, \mathbf{w}_k, \ldots, \mathbf{w}_{1+K}]) \equiv Y_{ki,t}([\mathbf{w}_1, \ldots, w_{ki}, \mathbf{w}_k^{(-i)}, \ldots, \mathbf{w}_{1+K}])$ as $Y_{ki,t}(\mathbf{w}_k) \equiv Y_{ki,t}(w_{ki}, \mathbf{w}_k^{(-i)})$ 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 $Y_{ik,t} = Y_{ki,t}(\mathbf{W}_k) \equiv Y_{ki,t}(W_{ki}, \mathbf{W}_k^{(-i)})$. 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 $\mathbf{e}_{N_1}^{(i)}$ 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 $Y_{11,t} = Y_{11,t}(1, \mathbf{0}_{N_1})$, $Y_{1i,t} = Y_{it}(0, \mathbf{e}_{N_1}^{(1)})$ for all $i \in \mathcal{N}_1 \setminus \{1\}$, and $Y_{ki,t} = Y_{ki,t}(0, \mathbf{0}_{N_k-1})$ for all $i \in \mathcal{N}_k$, $k = 2, \ldots, 1+K$.

Throughout the paper, 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 $t = 1, \ldots, T_0$, $Y_{ki,t}$, $i \in \mathcal{N}_k$, $k = 1, \ldots, 1+K$, are pre-treatment outcomes. In addition to them, we observe a vector of time- and unit-specific covariates, $\mathbf{C}_{ki,t} = [C_{ki,t,1}, \ldots, C_{ki,t,P}]$, $i \in \mathcal{N}_k$, $k = 1, \ldots, 1+K$, $t = 1, \ldots, T_0$, 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, $Y_{\mathcal{N}_{ki},t}$, and neighborhood-level unit×time specific covariates, $\mathbf{C}_{\mathcal{N}_{ki},t} = [C_{\mathcal{N}_{ki},t,1}, \ldots, C_{\mathcal{N}_{ki},t,P}]$, as average of the unit-level pre-treatment outcomes for units belonging to unit i's cluster/neighborhood:

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

and

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

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

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 $t = T_0 + 1, \dots, T$.

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

$$\tau_{11,t} = Y_{11,t}(1, \mathbf{0}_{N_1}) - Y_{11,t}(0, \mathbf{0}_{N_1}) \quad t = T_0 + 1, \dots, T. \tag{3.1}$$

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

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

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

$$\delta_t^{\mathcal{N}_1} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \delta_{1i,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \left[Y_{1i,t}(0, \mathbf{e}_{N_1}^{(1)}) - Y_{it}(0, \mathbf{0}_{N_1}) \right]. \tag{3.2}$$

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

$$\gamma_{11,t}^{(i)} = Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)}) - Y_{11,t}(0, \mathbf{0}_{N_1}). \tag{3.3}$$

The quantity $\gamma_{11,t}^{(i)}$, 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, $\gamma_{11,t}^{(i)}$ 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 $\gamma_{11,t}^{(i)}$ 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.

The difference between the direct effect and the unrealized spillover,

$$\tau_{11,t} - \gamma_{11,t}^{(i)} = Y_{11,t}(1, \mathbf{0}_{N_1}) - Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)}), \qquad i \in \mathcal{N}_1 \setminus \{1\} \tag{3.4}$$

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 $\tau_{11,t} > \gamma_{11,t}^{(i)}$, the actual treatment allocation brought about a gain for unit 1 with respect to unit i; if $\tau_{11,t} < \gamma_{11,t}^{(i)}$, then some alternative allocation of the intervention within the cluster would have been preferable for the treated unit; if $\tau_{11,t} = \gamma_{11,t}^{(i)}$, 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:

$$\gamma_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \gamma_{11,t}^{(i)} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)}) - Y_{11,t}(0, \mathbf{0}_{N_1}), \tag{3.5}$$

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 & 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 re-write the (individual) direct causal effect for the treated unit in Equation (3.1) and the average spillover causal effect in Equation (3.2) at time t, $t = T_0 + 1, \dots, T$, as function of the observed outcomes:

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

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 $\delta_t^{\mathcal{N}_1}$. The unrealized spillover by the treated unit in Equation (3.5), $\gamma_{11,t}^{(i)}$, depends on two unobserved potential outcomes, $Y_{11,t}(0, \mathbf{e}_{N_1}^{(i)})$, $i \in \mathcal{N}_1 \setminus \{1\}$, and $Y_{1t}(0, \mathbf{0}_{N_1})$, and thus we need to estimate both of them to get an estimate of $\gamma_{11,t}^{(i)}$, and thus, of $\gamma_{11,t}$.

3.3 SCG estimators of direct and average spillover effects

Under partial interference (Assumption 2), 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 & Gardeazabal (2003), Abadie et al. (2010), and further developed by Abadie & L'Hour (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 & Gardeazabal 2003, Abadie et al. 2010, Doudchenko & Imbens 2016, Abadie & L'Hour 2021). Similarly, our 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 patterns 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 & 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 $\mathbf{D}_{ki} = [Y_{ki,1}, \ldots, Y_{ki,T_0}, Y_{\mathcal{N}_{ki},1}, \ldots, Y_{\mathcal{N}_{ki},T_0}, \mathbf{C}_{ki,1}, \ldots, \mathbf{C}_{ki,T_0}, \mathbf{C}_{\mathcal{N}_{ki},1}, \ldots, C_{\mathcal{N}_{ki},T_0}]'$ 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

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

is chosen by solving the following optimization problem:

$$\arg\min_{\boldsymbol{\omega}^{(i)}\in\boldsymbol{\Omega}} \left\| \mathbf{D}_{1i} - \sum_{k=2}^{1+K}\sum_{i'\in\mathcal{N}_k} \mathbf{D}_{ki'}\omega_{ki'}^{(i)} \right\|^2 + \lambda^{(i)} \sum_{k=2}^{1+K}\sum_{i'\in\mathcal{N}_k} \|\mathbf{D}_{1i} - \mathbf{D}_{ki'}\|^2 \quad (3.6)$$

subject to

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

where $\|\cdot\|$ is the L^2-norm: $\|\mathbf{v}\| = \sqrt{\mathbf{v}'\mathbf{v}}$ for $\mathbf{v} \in \mathbb{R}^r$ (see Abadie & 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 (3.6) has a unique solution (see Theorem 1 in Abadie & L'Hour 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 & Gardeazabal (2003), Abadie et al. (2010) as $\lambda^{(i)} \to 0$; and the one-match nearest-neighbor matching with replacement estimator proposed by Abadie & 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:

$$\widehat{\tau}_{11,t} = Y_{11,t} - \sum_{k=2}^{1+K}\sum_{i'\in\mathcal{N}_k} \widehat{\omega}_{ki'}^{(1)} Y_{ki',t} \quad (3.7)$$

and

$$\widehat{\delta}_t^{\mathcal{N}_1} = \frac{1}{N_1}\sum_{i\in\mathcal{N}_1\setminus\{1\}} \widehat{\delta}_{1i,t} = \frac{1}{N_1}\sum_{i\in\mathcal{N}_1\setminus\{1\}} \left[Y_{1i,t} - \sum_{k=2}^{1+K}\sum_{i'\in\mathcal{N}_k} \widehat{\omega}_{ki'}^{(i)} Y_{ki',t} \right]. \quad (3.8)$$

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 & Sävje 2013, Cavallo et al. 2013, Acemoglu et al. 2016, Firpo & 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 & 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 Web Supplementary Material for details on the construction of BCa confidence intervals.

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 3.3. Unfortunately, the data contain no or little information on the potential outcomes of the form $Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1})$, because they are not observed for any unit in this study. Therefore, in order to construct an estimator for $\gamma^{(i)}_{11,t}$, we need to use an approach that extrapolates information on the potential outcomes $Y_{11,t}(0, \mathbf{e}^{(i)}_{N_1})$ from the observed data.

We deal with this issue using information on units in the treated unit 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 $Y_{1i,t}(0, \mathbf{e}^{(j)}_{N_1})$ 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). Let $\mathbf{D}^*_{1i} = [Y_{1i,1}, \ldots, Y_{1i,T_0}, \mathbf{C}_{1i,1}, \ldots, \mathbf{C}_{1i,T_0}]'$ be a $(P + 1) \times T_0-$dimensional vector of pre-treatment individual-level outcomes and covariates for $i \in \mathcal{N}_1$. For each unit $i \in \mathcal{N}_1$, let $\bar{\boldsymbol{D}}^{*(1)}_{1i}, \ldots, \bar{\boldsymbol{D}}^{*(K)}_{1i}$, be K linear combinations of the pre-treatment outcomes and covariates. Moreover let Δ_{1i} and $(t - T_0)$ be the distance between the centroid of unit i and centroid of the treated unit; and $(t - T_0)$ the duration of 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

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

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

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

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

$$\widehat{\gamma}_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \widehat{\gamma}^{(i)}_{11,t} = \frac{1}{N_1} \sum_{i \in \mathcal{N}_1 \setminus \{1\}} \widehat{Y}_{11,t}(0, \mathbf{e}^{(i)}_{N_1}) - \widehat{Y}_{11,t}(0, \mathbf{0}_{|\mathcal{N}_1|}). \quad (3.9)$$

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:

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

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

In this section, we apply the method described in Section 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.

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 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 & 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 $Y_{ki,t}(0, \mathbf{0}_{N_k-1})$ 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

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

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

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

In order to ensure the uniqueness and sparsity of solution of the optimization problems in Equation (3.6), we focus on values of $\lambda \in (0, 1]$, testing a total of 1000 values. Selected values for λ are reported in Table 6.2.2 of the Web Supplementary Material.

Once we have selected the penalization term, we move to the calculation of the weights. We estimate weights with the procedure described in 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 6.2.3 in Web Supplementary Material.

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 (3.7) and (3.8). 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 6.2.4 of the Web Supplementary Material. 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.

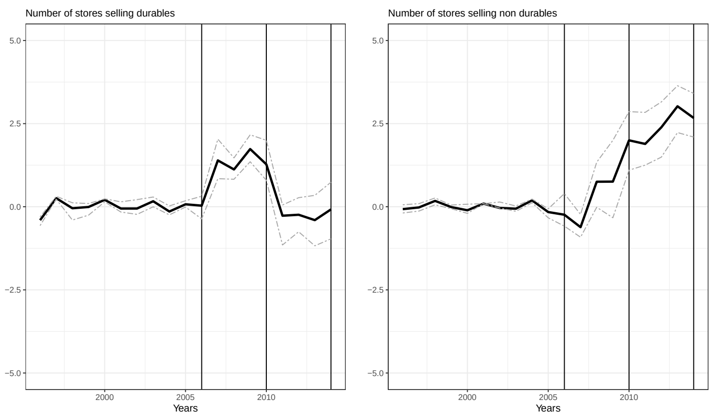

Figure 4.3.1: Estimated direct effects on Talenti St. (solid) and 90% confidence interval (dashed)

4.2 Horizontal regression estimators of unrealized indirect effects

Potential outcomes of the form $Y_{11,t}(0, \boldsymbol{e}_{N_1}^{(i)})$ for Talenti St. are imputed using the regression approach described in Section 3.4 with $K = 2$ and $\bar{\boldsymbol{D}}_{1i}^{*(1)} = Y_{1i,T_0}$ and $\bar{\boldsymbol{D}}_{1i}^{*(K=2)} = \sum_{s=1}^{T_0} C_{1i,s}/T_0$, 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 $\widehat{Y}_{11,t}(0, \boldsymbol{e}_{N_1}^{(i)})$, by using the small sample modification introduced by Imbens & Kolesar (2016), which allow us to account for the small number of cross-section in the estimation.

4.3 Results

4.3.1 Estimated direct and average spillover effects

Figure 4.3.1 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

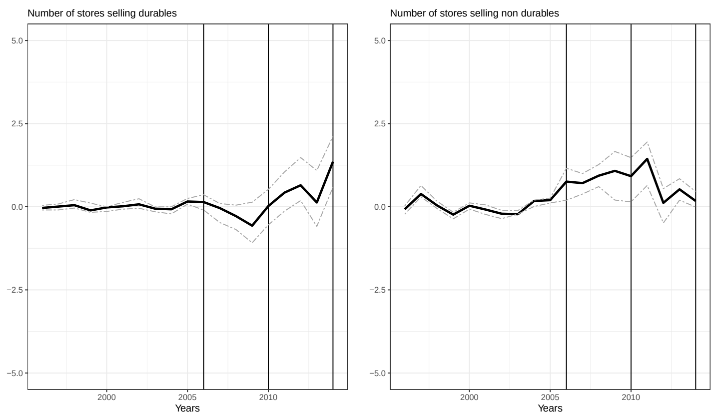

Figure 4.3.2: Estimated average spillover effects on neighboring streets (solid) and 90% confidence interval (dashed)

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 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 4.3.2. 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.

4.3.2 Estimated unrealized spillover effect

Figure 4.3.3 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

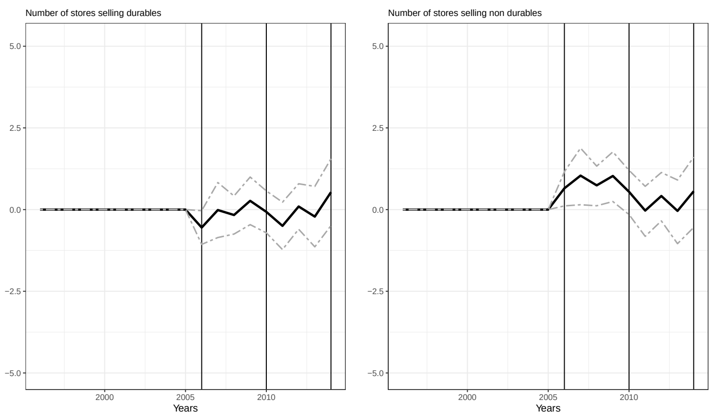

Figure 4.3.3: Estimated unrealized spillover effect on Talenti St. (solid) and 90% confidence interval (dashed)

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 happened in the streets that are actually susceptible to spillovers (see 4.3.2 for comparison). However, the confidence intervals here are quite wide, making it difficult to draw firm causal conclusions.

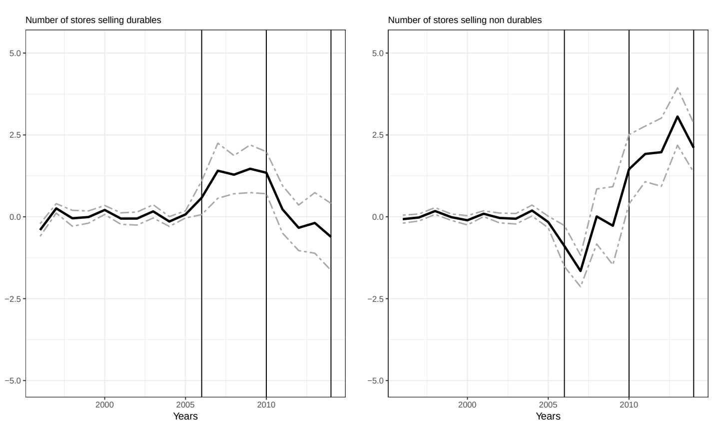

Figure 4.3.4: Difference between the estimated direct effect and unrealized spillover effect for Talenti St.(solid) and 90% confidence interval (dashed)

Figure 4.3.4 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 4.3.4 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.

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 & Imbens 2017) and the ideas initially put forward in Abadie & Gardeazabal (2003), Abadie et al. (2010, 2015) have sparked avenues of methodological research. This paper 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 paper, 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 & 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 & Steinmayr 2021, Forastiere et al. 2021). Nevertheless, we are aware that some studies might require a more general structure of interference (e.g., Forastiere et al. 2018, 2021). Therefore a valuable topic for future research is the extensions of SCM methods to causal studies with general forms of interference.

References

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

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

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

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

Abadie, A. & 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. & Mitton, T. (2016), 'The value of connections in turbulent times: Evidence from the united states', *Journal of Financial Economics* **121**(2), 368–391.

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

Arpino, B. & 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. & Khosravi, K. (2021), 'Matrix completion methods for causal panel data models', *Journal of the American Statistical Association* **116**(536), 1716–1730.

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

Baum-Snow, N. & 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. & Rothstein, J. (2021), 'The augmented synthetic control method', *Journal of the American Statistical Association* **116**(536), 1789–1803.

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

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

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

Budiakivska, V. & 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**.

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

Cattaneo, M. D., Feng, Y. & 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. & Pantano, J. (2013), 'Catastrophic natural disasters and economic growth', *Review of Economics and Statistics* **95**(5), 1549–1561.

Cervero, R. & 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.

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. & Hinkley, D. V. (1997), *Bootstrap methods and their application*, number 1, Cambridge university press.

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

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

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

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

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

Forastiere, L., Airoldi, E. M. & Mealli, F. (2021), '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., Mealli, F., Wu, A. & Airoldi, E. (2018), 'Estimating causal effects under interference using bayesian generalized propensity scores', *arXiv preprint arXiv:1807.11038* .

Gobillon, L. & Magnac, T. (2016), 'Regional policy evaluation: Interactive fixed effects and synthetic controls', *Review of Economics and Statistics* **98**(3), 535–551.

Grube-Cavers, A. & 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., Cervero, R. & Tischler, D. (2012), 'Half-mile circle: Does it best represent transit station catchments?', *Transportation Research Record* **2276**(1), 101–109.

Hong, G. & 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.

Huber, M. & 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. & Halloran, M. E. (2008), 'Toward causal inference with interference', *Journal of the American Statistical Association* **103**(482), 832–842.

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

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.

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

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

Larsson, J. P. & 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.

Mejia-Dorantes, L., Paez, A. & 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.

Pagliara, F. & 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.

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

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

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

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

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

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

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.

Xu, Y. (2017), 'Generalized synthetic control method: Causal inference with interactive fixed effects models', *Political Analysis* **25**(1), 57–76.

6 Web Supplementary material

6.1 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.

6.2 Tables

Table 6.2.1: 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 6.2.2: 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 6.2.3: 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 6.2.4: 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

6.3 Figures

Figure 6.3.1: 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.)

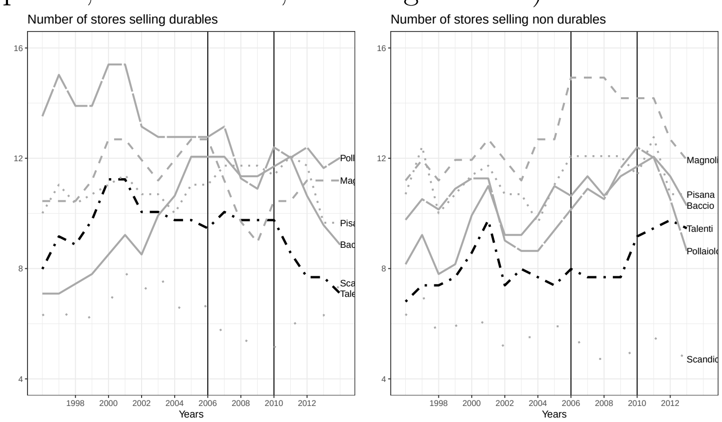

Figure 6.3.2: Streets involved in the analysis, clustered in their own urban neighborhoods