Essays in Applied Economics

Simone Robbiano

Department of Economics (DIEC)

University of Genoa

Supervisor

Anna Bottasso - University of Genoa
Maurizio Conti - University of Genoa

In partial fulfillment of the requirements for the degree of

PhD Program in Economics

2020

Preface

Underlying this doctoral thesis is the growing importance of the study of the relationship between cause and effect. This relationship can relate to policies, events, actions and processes. The analysis of causal relationships has developed over many years and still remains a central issue today. Indeed, in economics, the measurement of a particular causal effect is often one of the objectives of empirical analyses.

When researchers are unable to conduct a randomized controlled experiment, they must necessarily rely on the observation of non-experimental data, i.e. data from the real world, as is the case with sample surveys for example. A major problem in the use of this data arises from the fact that they are not derived from an experiment specifically designed for the purpose. Therefore, there may be variable factors whose effects are difficult to separate from the specific treatment effect.
The group of statistical techniques which have been developed for the assessment of causal effects from non-experimental data is therefore a very important field of applied econometrics.

In this dissertation, I will face the the relationship between cause and effect and several impact evaluation methods, presenting a collection of three applications of these techniques, adapted from three different works:

i "*Price Matching and Platform Pricing*", presented at the First NetCIEx Workshop (EU Joint Research Centre - Ispra);

ii "*Roads to Innovation: Evidence from Italy*", presented at the First NetCIEx Workshop (EU Joint Research Centre - Ispra);

iii "*Public Funded R&D as a Device for Local Innovation? Evidence from Italian I.I.T. Foundation*".

In the first paper, the effects of Price Matching Guarantees (a PMG refers to the price strategy where different retailers commit themselves to match any lower price offered by competitors on the same item or product category) on U.S. online consumer electronics prices are empirically investigated, by means of a unique dataset developed through sophisticated and computerized scraping procedures.
In such paper, joint use of daily price data, product characteristics derived from User Generated Contents (UGC) and the construction of a control group with a novel approach allow to implement a Difference-in-Difference analysis, aimed to assert the existence of a causal effect of PMG on prices.
The paper finds evidence in favor of price reductions occurring after the PMG policy is repealed. The analysis further investigates if such effect is heterogeneous according to products characteristics, by exploiting UGC (products popularity and quality) and online search visibility measures (Google Search Rank). Estimates suggest that for high quality (visibility) products PMG policies harms competition by keeping prices

high, while for low quality (visibility) products, prices decrease during the policy validity period.

The second article deals on the literature on the economic impact of transport infrastructure, and in particular on the role that road infrastructure can have on innovative regional capacity.
The seminal contribution by Agrawal et al. (2017) is followed to estimate a model of "roads and innovation" where the innovative activity in 1988 is linked to the length of motorways system in 1983, in order to investigate the impact of motorways endowment on the innovative capacity in each Italian NUTS-3 region. The main challenging issue about the estimation of the model arises from the possible endogeneity of highways stock. To deal with this problem, the "historical instrumental variable" approach is followed, by using the length of the ancient Roman roads system dating back to 117 AD as an instrument for the length of current motorways.
Overall, the Instrumental Variable estimates indicates that 1983 highways network has a positive and significant impact on 1988 innovative capacity. Moreover, the analysis find a declining role for highways over time. Furthermore, results suggest a spatial reorganization of economic activity rather than a pure net economic effect.

The third paper concludes this dissertation. In such final analysis, the effects on the regional economy of a prominent Italian place-based policy, the institution of Istituto Italiano di Tecnologia (Istituto Italiano di Tecnologia, IIT, is a scientific research centre established in 2003 that conduct scientific research in the public interest), are investigated by means of a novel identification strategy, the Synthetic Control Method (SCM).
Such identification approach, unlike other counterfactual impact evaluation techniques, is based on the creation of an artificial control unit that not only follows the same pre-treatment trend as the treated unit, but even overlaps the same one. In particular, through the SCM approach, the innovative and economic development (measured by patents per capita, number of local inventors and per capita GDP) of the treated region, namely Genoa, is compared with a set of Italian NUTS-3 control regions with the aim to estimate the causal effect of the location of IIT in 2006.
Estimates show significant effects of IIT presence on local patent activity, highly specialised human capital endowment in research and GDP per capita, suggesting the existence of local spillovers from public research.

Contents

Price Matching and Platform Pricing

This is the result of joint work with

Anna Bottasso*, Paolo Marocco*

*University of Genoa

April 2020

Abstract

In this study we investigate the effects of Price Matching Guarantees (PMG) commercial policies on U.S. online consumer electronics daily prices. By applying a Diff-in-Diff identification strategy we find evidence in favor of price reductions occurring after the PMG policy is repealed.

We further investigate if such effect is heterogeneous according to products characteristics, by exploiting User Generated Contents (products popularity and quality) and online search visibility measures (Google Search Rank). Estimates suggest that for high quality (visibility) products PMG policies harms competition by keeping prices high, while for low quality (visibility) products, prices decrease during the policy validity period.

1 Introduction

Online sales platforms have recently gained increasing importance in both retail and wholesale markets. Such markets are characterized by the supply of personalized services, more convenient delivery schedules and the ability to reach a very high number of consumers. In addition, platforms claim to warrant lower prices with respect to traditional stores through the provision of offers, promotions, down prices and other price discounting policies. Among these options, online platforms often implement Price Matching Guarantees policies (PMG), that is the promise to reimburse price differences when competitors offer a lower prices[1].

PMG policies are surely appealing for customers and can increase consumer confidence and brand fidelity. However, the announcement to tie prices to those of competitors can have anti-competitive effects and sustain high prices, thus harming

[1]For example, a PMG policy, for a generic store, is “If you see a lower price for any good we stock, we will be cheerful to refund you the difference”.

consumers welfare.
Most of theoretical literature agrees on the fact that PMG reduce firms incentive to compete on prices and lower the motivation for consumers to search better sale conditions (Hay, 1981; Salop, 1986). However, in some models, PMG are considered as tools for price discriminating or as real discounting policies (Png and Hirshleifer, 1987; Belton, 1987). Therefore, empirical analyses become particularly relevant in order to understand under what conditions such pricing policies reduce consumer welfare. Indeed, the applied literature analyzing this issue is scant and does not provide conclusive results (Mago and Pate, 2009; Zhuo, 2017).

Our work add on this literature by providing empirical evidence on the effect of platforms Price Matching Guarantees policies on daily consumer electronics prices observed on US online market. We have focused on the NewEgg platform that exclusively sells consumer electronics products and implements PMG policies that turn on and off over time (blinking PMG). Given that our identification strategy is based on a comparison of price levels before and after the policy shutdown, we excluded platforms that never stop offering PMG (like Target).
In particular, we apply a Diff-in-Diff (DiD) approach where we consider as the treated sample the pool of NewEgg products interested by PMG policies. Differently from standard practices in studies adopting a DiD approach, we build the control sample with price data for the same products observed on a different platform, namely Amazon, that never offers PMG to customers. Furthermore, in order to ensure that our counterfactual sample is less likely to be influenced by the PMG policy adopted by NewEgg, we have considered data from the Amazon UK platform, instead of Amazon US. As a matter of fact, price observed on Amazon US might not be completely independent from the policy under scrutiny, because of price tracking practices frequently adopted by platforms.

Estimates provide evidence in favor of an average price reduction of about 3.9% after the interruption of the PMG policy. However, in order to have a more detailed picture of the issue, we investigate if such effect is heterogeneous, depending on products characteristics. In particular, we focus on products features that might affect the outcome of PMG policies and that can be recovered exclusively on online markets. Platform data allow us to obtain information from User Generated Contents (UGC), like product popularity, product quality and online search visibility (Google Search Rank); indeed, we believe that these product characteristics might provide indirect information on consumers heterogeneity. Estimates conducted on specific sub-samples show that when the PMG is interrupted, low quality (low search rank) products experience a price increase of about 3.4%, while for high quality (high visibility) products a price reduction of about 3.7% is observed. These findings are in line with the lack of unambiguous predictions of the theoretical literature.

The anti-competitive effects of PMG observed for high quality (visibility) products has been predicted by theoretical models where such polices make collusion more likely (Hay, 1981; Salop, 1986; Cabral et al., 2018). These findings can be also explained by the theoretical predictions of a class of models, like Corts (1997) and Nalca et al. (2010), where PMG is a tool for discriminating customers according to their sensitiveness to price and products quality. These models also explain our results obtained for products characterized by low visibility (quality). Indeed, the willingness to engage in search activity could indirectly identify those customers whose demand is more rigid, as argued by the search literature (Ellison and Ellison, 2009).

The analysis conducted in this study enriches the literature on the price effects of PMG by using very detailed platform daily price data for a specific market (consumer electronics) where such policies are very common. In particular, the DiD identification strategy adopted is based on the construction of a control group with a novel approach; moreover, products characteristics based on Users Generated Contents (UGC) are employed for the first time in order to study possible heterogeneous effects of such policies.

The rest of the paper is organised as follows. In the next section we briefly discuss related literature and in Section 3 we accurately describe the data extraction process and the variables used in the empirical analysis. Section 4 explains our identification strategy and in Section 5 we discuss estimates results and robustness analysis. Section 6 concludes.

2 Literature Review

The theoretical literature has analysed possible impacts of Price Matching Guarantees (PMG) on different market outcomes, since such commercial policies might affect the behaviour of firms and consumers in different ways.

The most common prediction of the theoretical models is that price matching guarantees hamper competition by keeping prices high and sustaining collusive practices; moreover, some authors suggest that they might be tools for realising price discrimination or signalling cheap prices.

Hay (1981), Salop (1986) and Belton (1987) have first suggested that price matching guarantees can sustain collusion in oligopoly models; they highlight that such clauses might be considered as threats to punishment for firms that lower cartel prices, thus reducing firms incentive to deviate from the agreement. They argue that, if all competitors in the market adopt a PMG policy, none of them has the incentive to lower its price and the latter tends to the monopolistic level. Moreover, they agree on the fact that the adoption of such policies increase the stability of the cartel, as any price cut must be refunded to the consumer, so that the policy generates a credible penalty system.

Several other papers support the pro-collusive argument by extending the basic oligopolistic setting (see among others Doyle (1988), Logan and Lutter (1989) and Baye and Kovenock (1994)), while other authors explore the impact of PMG extending the analysis in dynamic, multi-stage and Hotelling frameworks (see among others Chen (1995), Lu and Wright (2010), Hviid and Shaffer (2010), Pollak et al. (2017), Constantinou and Bernhardt (2018)). Also Cabral et al. (2018) suggest that a PMG can be a collusion enacting practice. In the model two firms alternate over time in setting prices; given that starting a collusion process implies several risks, like for example antitrust penalties, firms include collusion costs in their decisions. The main prediction of the model is that the probability of tacit collusion rises when the policy is in place.

In studies reviewed so far, it is implicitly assumed that customers automatically claim the price guarantee whenever they find a price differential: indeed, this is not always the case, because of lack of information or because there are small costs for the customer to activate a guarantee, the so called “hassle costs”. Hviid and Shaffer (1999) highlight that the presence of hassle costs undermines possible anti-competitive effects of PMG, but do not completely cancel them. Precisely, with symmetric firms PMG are unable to support any price increase in presence of hassle costs. Indeed, each firm will be interested in lowering price levels by an amount that is marginally smaller than these costs, so that buyers are attracted from cheaper firms and do not activate the guarantee. Otherwise, with asymmetric firms, a rise in prices might be supported, but not at the monopolistic level. Moreover, their model can explain why universal adoption is not a realistic assumption of previous studies.

Some other models explore the possibility that sellers use PMG policies as a price discrimination tool. If customers are different in terms of some subjective characteristic, like information on prices and guarantee terms, degree of loyalty to a specific retailer or level of hassle costs in requesting the refund, firms could use the price guarantee to discriminate between different groups of consumers.

Png and Hirshleifer (1987), Belton (1987) and Corts (1997) first suggested duopoly models where firms discriminate between different consumers groups, namely “unsophisticated” customers and “sophisticated” ones. Consumer segmentation and PMG allow firms to set higher prices for unsophisticated consumers, while sophisticated ones benefit from the lowest price guaranteed by the policy. The main intuition from this strand of literature is that price discrimination might at least benefit some customers with actually lower prices[2].

Finally Moorthy and Winter (2006) suggest that PMG might be a credible signal of low prices, if low cost firms adopt the policy and (high cost) competitors can not match the policy. Similarly, Jain and Srivastava (2000) develop a theoretical model that identifies the conditions under which PMG might lead to lower market prices[3]. In the presence

[2]Similar results can be found also in Edlin (1997) and Nalca et al. (2010).

[3]Authors have realized two experiments to analyze the effect of PMG policies on prices consumer

of informed and uninformed consumers (about prices and store characteristics) and of different kind of stores (in terms of size, service quality ans so on), only stores with low prices offer price-matching policies[4].

Despite the theoretical literature is rich and analyses several aspects of price matching policies, the empirical evidence is scant and does not provide conclusive results. Some studies focus on specific markets, like tyre or gasoline, while others analyze retailing prices from supermarkets, grocery stores or online platform markets.

Analysing daily price quotes from the tyre industry advertisements, on 61 US Sunday newspapers observed for three months in 1996, Arbatskaya et al. (2000), through a Feasible Generalised Least Square approach, find weak evidence of anti-competitive effects of PMG and show that an increase in the number of firms implementing the policy leads to a 10% increase in prices[5]. Cabral et al. (2018), focus instead on daily pricing policies adopted by the Shell network of gas stations in Germany in 2015[6]. Leveraging on gas stations localization and consumers demographics as sources of identification, they suggest that PMG can be a collusion enacting policy. Gas station prices have been analyzed also by Byrne and De Roos (2019) for Australia by means of a detailed 15 years time series dataset. Authors argue that the majority of gas stations prices follow a weekly cycle and that dominant firms can use PMG to coordinate market prices and reduce price competition. Similar results can be found in Chilet (2018), who analyses pricing policies of three big retail pharmacy chains in Chile, observed over the period 2006-2008. The author follows an identification strategy based on the estimation of a demand model, in which quantity sold is a function of the differences between own prices with the competitors ones, around the time period where collusive price increases occurred.

Hess and Gerstner (1991) analyse the effect of PMG on prices by collecting weekly data of 114 goods sold in several US supermarkets and grocery stores, from 1984 to 1986. Authors, by means of a time series analysis, provide evidence in favor of higher prices of about 1-2% when the guarantee is introduced. Moorthy and Winter (2006) argue instead that the adoption or non-adoption of the PMG might be interpret as a way to signal the seller service–price profile[7]. Authors analyse data for several product categories from 46 Canadian retailers observed in 2002. They assume the existence of informed and uninformed consumers and show that PMG might be a tool to signal low prices to uninformed consumers. In particular, they find that PMG are mainly adopted by low cost/low service chain stores. Similar results can be found in

perceptions and have shown that consumers did expect lower prices from PMG.

[4]Similar results can be found also in Moorthy and Zhang (2006).

[5]The same authors in Arbatskaya et al. (2006) confirm their results by analysing the same data with a different approach.

[6]See also Wilhelm (2016).

[7]The authors refer to the retailer service-price profile as to any sellers characteristic that might induce customers to choose one seller over another one, like better sales assistance and customer care, a clear Web site, personalised delivery and selling services.

Chung et al. (2016) for three leading hypermarkets in Korea. Finally, Zhuo (2017) focuses on online platforms and collect US price data from online price trackers for 150 products offered on Amazon in 2012. The author observes prices during and after the implementation of PMG policies by two big-box stores (Target and Best Buy) targeted specifically on Amazon prices; by applying DiD and RDD methods, the author suggests that prices increase by about six percentage points during the period of validity of the policy[8]; moreover, the analysis highlights an heterogeneous impact of PMG, with larger price increases for initially lower-priced goods[9].

3 Data

3.1 Data Extraction

In order to study the impact of PMG on prices, we focus on the online consumer electronics market, since it is one of the most widespread sector on online retailing and is often interested by such pricing policies. In particular, electronic products are search goods, whose quality can be evaluated before the purchase: the advent of online markets has made this process much cheaper and faster and is most likely to affect the impact of such policies, whose outcome depends, among other factors, by the level of search and hassle costs. Moreover, electronic goods are barely affected by seasonal effects, so that prices signals are more stable over time and show low price differentials across countries (Gorodnichenko and Talavera, 2017; Stallkamp and Schotter, 2019). These characteristics allow us to improve the identification strategy through the construction of a more refined control group (see the next section).

Among different online retailing platforms we choose to focus on NewEgg, a leading online US retailer of consumer electronics products, that implements blinking PMG on selected items. Given that our identification strategy is based on the comparison of prices before and after a policy shutdown, we do not consider platforms that apply PMG to wide groups of products continuously over time (i.e. Target, among others). In particular, NewEgg communicates the period of validity of the price guarantee by means of a label that appears on the specific product online page; the customer who discovers the PMG badge has 14 calendar days of time to find the same title[10] at a lower price from US competitors belonging to a declared list[11]. PMG policies are

[8]Similar results can be found also in Wu et al. (2015), Haruvy and Leszczyc (2016).

[9]Some other authors analyse the impact of price-beating guarantees, that are less widespread policies with similar terms as price matching ones (in price beating guarantees refund exceeds the price difference). Studies that refer to these policies argue that, with respect to price matching guarantees, they might be serving different purposes in practice and likely be effective in enhancing competition. Experimental literature also focuses on the effect of price matching and price-beating guarantees: however, experimental results lack the complexity of real interactions between sellers and consumers.

[10]With title we refer to a product with the same brand and model number.

[11]NewEgg, after checking the validity of the claim, sends a Customer Care Card to refund the price difference (Source: https://kb.newegg.com/knowledge-base/price-match-guarantee/).

often repeated over time on the same products without any notice, so that consumers looking for deals have to exert an higher effort in the search process.

In order to build the sample we have identified NewEgg products interested by PMG on May 10^{th} 2018 (100 products). For such products we have collected price data until 31^{st} October 2018 (174 days and 9028 observations). We identify as the treatment of interest the interruption of the PMG policy, so that prices observed on the NewEgg platform represent the treated sample. The control sample has been built by recovering price data for the same products observed on NewEgg but sold on the Amazon UK platform, that never offers PMG policies[12]. This reduces the number of observed products, so that the final sample includes 29 products belonging to 19 sub-categories (computer hardware, tablet and computers, mobile phones, printers and scanners, PC accessories, speakers for domotics, screens and audio devices). In the Appendix we provide a detailed list of selected products (Tables A1 and A2).
It is worth noting that, unlike NewEgg, Amazon operates both as a retailer and as a marketplace for third-party sellers who pay fees and royalties to access to Amazon customer base. In such marketplace, Amazon often acts only as a payment intermediary and goods are kept in the third-party sellers inventory. Thus, in order to build a valid control group, we have excluded data on products sold by third-party sellers on the Amazon platform.

The retrieving of sample data has been a challenging task. Given the absence of ready-made and easy-to-use repositories on price data, we have developed an ad-hoc scraping program (in Python language)[13] able to protect the scraping process from unpredictable changes of the page[14] and capable to recover the data without stressing the site, thus limiting the risk of interruptions due to firewalls. In particular, the scraping process has been supported by several alert tools signalling periodical changes of the internal page structure[15].

The process of data collection has required the daily implementation of these main steps:

i Sign up for subscription to Amazon Web Service (AWS) cloud[16], in order to use virtual servers in which to install and launch the program;

ii Accept the norms and terms of use of the platform site, in order to be compliant to the server navigation policies;

iii Launch the daily loop process, in order to navigate among product pages, select

[12]See the next section for a rationale on this choice.

[13]Code available from authors upon request.

[14]A typical problem is to intercept daily changes of web pages not only about prices, but also concerning other dynamic contents, such as the number of customer reviews, the average rating and so forth.

[15]Indeed, platforms frequently change the deep structure of the page, in a not visible way by the human reader but in a way that affects the program code and the scraping process.

[16]AWS is a comprehensive, evolving cloud computing platform provided by Amazon.

the field tags, get the data and save on a server disk. Each scraping session runs about 20 minutes every day.

In addition to products daily prices retrieved on both platforms, we also collect several product characteristics available exclusively on online sales platforms. In particular, we recover some User Generated Content (UGC), like the absolute number of reviews, product rating and product Google search rank.
The absolute number of reviews[17] is a dynamic information which represents a sort of popularity index, since it is proportional to the product market diffusion. We also calculate the relative number of reviews as the ratio between the number of reviews of each product and the amount of reviews received from the most popular product in the same subcategory[18]. This normalized index, that ranges from zero to one, shows the relative popularity of the product with respect to the other items of the same sub-category. We also collect data on product ratings (stars) provided by consumers. We consider the number of stars gained by each product, ranging from zero (low quality) to five (high quality), as a proxy of product quality. Finally, we develop a search index as a proxy of the time spent on search engines to discover the page of a certain product. More precisely, the search index represents the probability to find the product in the first ranked positions of Google results[19].

It is worth noting that, although the products analysed are sold by Amazon and NewEgg in different countries, information on some of the considered UGC (e.g. rating) maintain their consistency. This property is typical of consumer electronics goods that have a standardized nature. Concerning the search index, we adopt a country specific value by launching the Google search engine with specific country settings (UK and US).

3.2 Descriptive Statistics

Our sample consists of 9028 daily price observations (174 days) for 29 products observed on NewEgg and Amazon UK platforms, from 10 May 2018 until 31 October 2018. Table 1 shows the mean and the standard deviations of prices and selected product characteristics for the overall sample and for the treated and control group ones. Prices show a large variability, being the average for the overall sample $240.43 and the standard deviation $283.53. By comparing average values observed over the two platforms, it emerges that both prices and UGC display similar values, thus confirming what has been observed by the previous literature on the low dispersion of consumer electronics prices across countries (Stallkamp and Schotter, 2019); moreover,

[17] In online commerce, product reviews are used by retailing platforms to give consumers an opportunity to comment on products they have purchased, right on the product page.

[18] See Table A1 in Appendix for details.

[19] The ranking position of an item is retrieved launching the Google query composed by the sentence: “the name of product” AND “the name of platform”. The resulting position is then normalized, mapping in the probability range [0,1].

such similarities support our approach for the choice of the control sample. As Table 2 shows, in the case of the treated sample (NewEgg) the average price during the policy validity period (before treatment) is about $18 higher with respect to the post implementation period[20].

Table 1: Summary Statistics. Treated and Control samples.

Variables	Full Sample	Amazon UK	NewEgg
Provider Price ($)	240.43	227.72	253.15
	(283.53)	(262.74)	(302.39)
Product Popularity (0-1)	0.23	0.26	0.20
	(0.27)	(0.30)	(0.23)
Search Rank (0-1)	0.75	0.85	0.64
	(0.30)	(0.17)	(0.36)
Rating (0-5 stars)	4.14	4.14	4.15
	(0.68)	(0.48)	(0.83)

Table 2: Summary Statistics. NewEgg. Pre and Post Treatment.

Variables	Pre Treatment	Post Treatment
Provider Price ($)	226.17	208.41
	(314.57)	(243.15)
Product Popularity (0-1)	0.22	0.20
	(0.25)	(0.23)
Search Rank (0-1)	0.76	0.72
	(0.27)	(0.30)
Rating (0-5 stars)	4.15	4.05
	(0.98)	(1.08)

Notes: The pre-treatment period is the policy implementation period.

Table 3: Summary Statistics. Sub-Samples.

Variables	Low Quality Low Search Rank		High Quality High Search Rank	
	NewEgg	Amazon.uk	NewEgg	Amazon.uk
Provider Price ($)	206.00	95.81	221.10	161.10
	(143.90)	(35.11)	(354.70)	(163.90)
Product Popularity (0-1)	0.01	0.01	0.28	0.29
	(0.01)	(0.01)	(0.25)	(0.33)
Search Rank (0-1)	0.19	0.38	0.90	0.91
	(0.27)	(0.38)	(0.04)	(0.05)
Rating (0-5 stars)	2.93	3.19	4.40	4.45
	(0.71)	(0.32)	(0.49)	(0.20)

Notes: For high quality products we mean those with ratings higher than 4. For high visibility products we mean those with a normalized search index higher than 0,8.

In order to investigate the issue of heterogeneity in the effect of PMG policies on prices, we distinguish products according to products characteristics recovered

[20]We remember that our treatment is the policy shutdown.

from UGC. In particular we classify products depending on their quality and visibility, measured trough UGC as explained in the previous section. Given that quality assessment by consumers is highly correlated to products visibility, in Table 3 we show some descriptive statistics for products classified according to such characteristics[21]. Again, data show that products characteristics stemming from UGC are quite similar across countries/platforms.

As far as the PMG policy is concerned, NewEgg adopts a blinking strategy, so that the policy is applied in a non continuous way, often to the same products. Table 4 shows the total number of days of treatment (absence of PMG) and the average number of treatments occurred in each sample. This latter information suggests that, on average, the policy is applied to each product twice during the sample period (174 days) and such frequency does not seem to be correlated to products quality and visibility. Indeed, since prices are highly correlated to quality, we can reasonably assume that there is not selection into treatment associated to products price or quality (visibility), so that the assumption of random assignment required by the identification strategy seems reasonably fulfilled. On the other side, it seems that, for high quality (visibility) products, the policy implementation period is longer.

Table 4: Summary Statistics. PMG.

Variables	Full Sample	Low Quality	High Quality
		Low Search Rank	High Search Rank
Treatment Duration (days)	38.25	57.38	32.38
	(39.94)	(45.28)	(36.18)
Number of Treatments	2.38	1.80	2.56
	(1.75)	(0.92)	(1.90)

Notes: Treatment duration is the average number of days without PMG. The sample period includes 174 days.

Another important issue is related to the representativeness of our sample. Figure 1 represents the distribution of products by price classes (10). The graph shows that 22 products out of 29 belong to the first two price deciles, with price ranging between 0$ and 240$. This picture closely matches a typical distribution observed in consumer electronics (Coad, 2009), often characterized by a large amount of low cost accessories and few luxury goods. Furthermore, calculating the log-price distribution (Figure 2) and mapping the integer part of this value on the x-axis, we obtain a septile-partition. By plotting the distribution of products by log-price classes we obtain a distribution that resembles the Normal one. Such result is in line with those obtained by Coad (2009).

[21]High quality products are those characterized by a rating higher than 4/5, while high visibility ones are those endowed of a search rank index greater than 0.8.

Figure 1: Products Distribution by Price Classes.

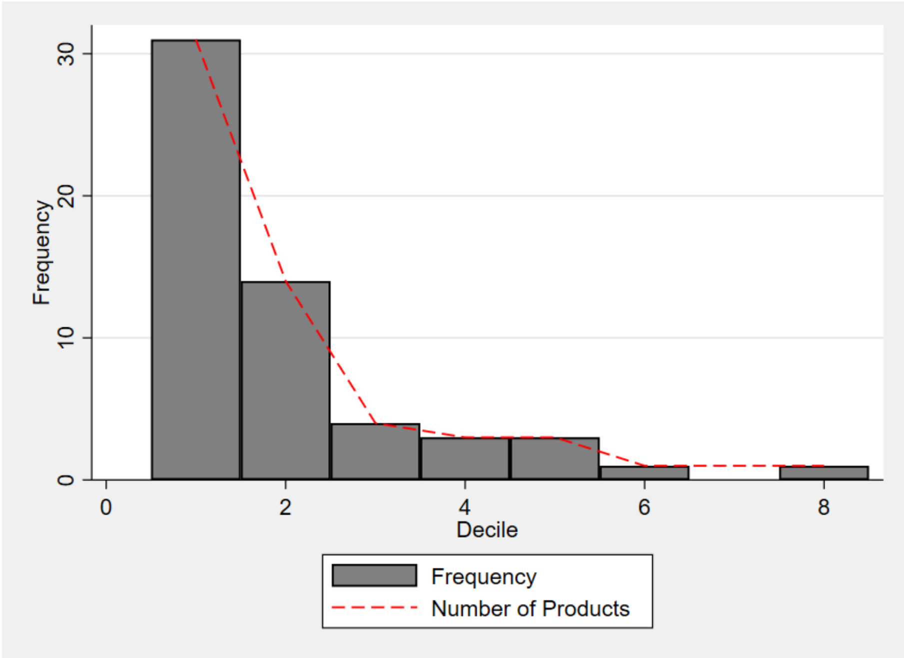
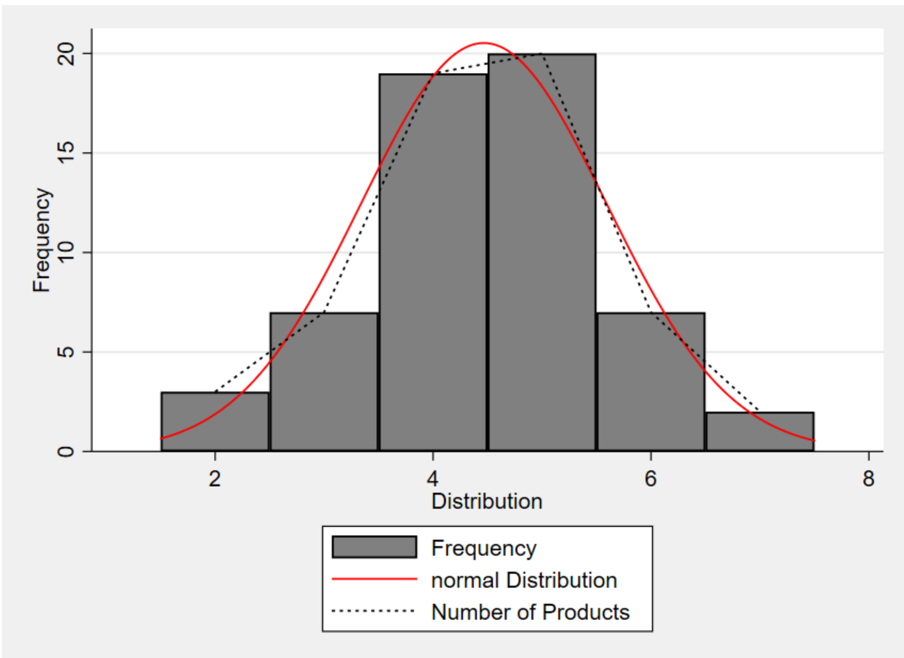


Figure 2: Products Distribution by Log-Price Classes.



4 Identification Strategy

We identify the causal effect of price matching guarantees on price levels, by comparing prices before and after the policy shutdown for a sample of products sold by NewEgg (the treatment group), to the prices average change for the same products sold by Amazon UK (the control group). Indeed, and crucially for our identification strategy, PMG implemented by NewEgg only affects products that are sold in US, thereby naturally creating a treatment and a control group; the same products sold by Amazon UK (that never offers price warranties) are less likely to be affected by the policy and well represents a counterfactual sample mimicking what would have happened to prices of treated products in the absence of PMG. This framework provides a quasi-natural experiment that allows us to study the causal impact of PMG on prices through a Diff-in-Diff research design.

This identification approach requires the estimation of the following panel FE model[22]:

$$\log Price_{i,l,t} = \alpha_{i,l} + \gamma\big(T_{i,l,t} * P_{i,t}\big) + X^T_{i,l,t}\beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \tag{1}$$

The dependent variable, $Price_{i,l,t}$ represents the price (natural logarithm) of good i, on platform l, at time t; $T_{i,l,t}$ denotes a binary variable equal to 1 for treated goods; $P_{i,t}$ is a binary variable that is equal to 1 for any day since the policy shutdown and zero otherwise and $\epsilon_{i,l,t}$ is an error term. The model includes a full set of daily time dummies, τ_t, accounting for unobserved time-varying determinants of prices that are common to all goods. Product fixed effects, $\mu_{i,l}$, control for any time invariant unobserved heterogeneity at the product and platform level, that could be correlated with the included regressors and that could also drive prices. Moreover, the presence of individual (product) fixed effects in the Diff-in-Diff research design rises the degree of comparability of treatment and control groups.

We include a set of covariates, $X_{i,l,t}$ in Equation (1), in order to control for products characteristics derived by UGC that might affect the outcome of the PMG policies. The γ coefficient associated to the interaction term $(T_{i,l,t} * P_{i,t})$ represents the DiD estimate of the effect of PMG shutdown on treated products prices and it measures the average price differential between the treated and the control group.

We also explore the issue of heterogeneity in the effect of PMG policies on prices. Indeed, as discussed in the literature review section, most of the predictions of theoretical models on the price effects of PMG policies rely on assumptions related to the presence of heterogeneous consumers. By distinguishing products according to consumers quality assessment, we indirectly assume that consumers are heterogeneous in terms of their preferences towards quality and their availability to pay a

[22]In a Diff-in-Diff context, a classic model would be built like $Y = \alpha + \beta_1(Treated) + \beta_2(Post) + \beta_3(Treated * Post) + \varepsilon$. In all models we exclude $Treated$ and $Post$ terms, since they are multicollinear with time and product fixed effects.

price premium for that. Indeed, for high quality goods the price elasticity of demand is usually assumed to be lower than price elasticity for low profile goods. We further classify products according to their visibility, as measured by the search index described above. We believe that the time spent for finding a product indirectly selects consumers according to their willingness to engage in search activity and that such availability is directly correlated to their price sensitiveness.

Based on the above reasoning, we estimate Equation (1) on different sub-samples built according to product quality and visibility indices. In particular, we analyse separately high (low) quality products, namely products characterized by rating greater (lower) than 4/5, and products characterised by high (low) visibility in terms of Google search rank, namely products whose search index is greater (lower) than 0,8. Moreover, given that products quality and visibility resulted to be highly correlated, we split the sample according to both characteristics. As discussed in the Data Section, such products characteristics do not affect the probability of being treated.

The heterogeneity issue is also investigated with a different approach by estimating a Triple Difference regression (DDD) on the full sample. In particular, we estimate the following model:

$$\log Price_{i,l,t} = \alpha_{i,l} + \varphi\left(T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}\right) + X^T_{i,l,t}\beta + \mu_{i,l} + \tau_t + \omega_{i,l,t} \tag{2}$$

Equation (2) includes an additional component in the interaction term, $HRHV_{i,l,t}$, i.e. a dummy variable assuming value 1 for high quality and high visibility products. The coefficient φ of the triple interaction measures the average treatment effect of PMG on prices for high quality (visibility) products.

All specifications are estimated by OLS and Standard Errors are robustly estimated. Moreover, following Cameron and Miller (2015), we compute bootstrapped standard errors with a cluster structure (at product level) and all results are confirmed. Finally, we conduct an extensive robustness analysis through different falsification and placebo tests (see the Robustness Analysis Section).

5 Empirical Results

5.1 Main Results

In Table 5 and 6 we show DiD estimates; in particular, we first report results obtained by estimating Equation (1) without including control variables (Table 5, column (1)), while Table 6 (column (1)) provides estimates obtained after including all control variables. DiD estimates suggest that the PMG shutdown triggers a significant reduction of price levels of about 3.9%. Rather comfortingly, the inclusion of control variables into model (1) does not significantly affect the result. These findings suggest that, on average, the adoption of PMG has an anti-competitive effect on prices since, after the policy validity period, they show a substantial reduction. These results are consistent with those obtained by Zhuo (2017) on a large sample of products observed on the Amazon platform in 2012. However, we follow a rather different identification strategy. While Zhuo (2017) focuses on price changes observed on the non-adopting platform, before and after the implementation of PMG by competitors, we focus on price changes observed on the adopting platform. Moreover, we innovatively build the control sample with platform price data for the same treated products but observed in another country (UK).

To explore whether product properties affect the impact of PMG on prices, we split the sample according to different classes of product quality and visibility and we re-estimate Equation (1). Columns from (2) to (5) in Tables 5 and 6 show results of this disaggregated analysis. Estimates indicate that a policy repeal produces a price reduction for both low and high quality products; however, the estimated coefficients for the low quality sample are not statistically different from zero, while those for high quality products indicate a statistically significant price reduction of about 2.5%. When we split the sample according to values assumed by the search index, results suggest that, when the PMG is interrupted, products characterised by a low search rank experience a price increase of roughly 2,4%, while for high visibility products prices decreases of about 5,3%. These findings support the hypothesis that, in online consumer electronics market, PMG policies harm competition for high visible products by keeping prices high, while for low visible products, such policies have a pro competitive effect on prices.

Indeed, as highlighted in the data section, quality and visibility are highly correlated in our sample. Hence, we estimate Equation (1) after splitting the sample according to both product properties.
Results shown in column (6) and (7) of Table 5 suggest that the PMG shutdown triggers a reduction of prices for high quality and high visibility products (3,7%), while prices of low quality and low visibility ones raise of about 3,4%. These findings are confirmed when we include control variables into the model (Table 6, columns (6) and (7)) and when we analyse heterogeneous effects of PMG by means of a Triple Difference regression approach, as shown in columns (1) and (2) of Table 7.

Our empirical findings can be explained by the main predictions of theoretical models analysing the impact of PMG on prices and competition.
The anti-competitive effect of PMG observed for high quality (visibility) products has been predicted by theoretical models where such polices make collusion more likely (Hay, 1981; Salop, 1986; Cabral et al., 2018). These findings can be also explained by the theoretical predictions of a class of models, like Corts (1997) and Nalca et al. (2010), where PMG is a tool for discriminating customers according to their sensitiveness to price and products quality. These models also explain our results obtained for products characterized by low visibility (quality).

Table 5: DiD Estimates of the Impact of PMG on Prices.

Products Prices (log)	(1) FULL SAMPLE	(2) L. RATING	(3) H. RATING	(4) L. VISIBILITY	(5) H. VISIBILITY	(6) LR-LV	(7) HR-HV
$T_{i,l,t} * P_{i,t}$	-0.0401***	-0.0064	-0.0250***	0.0242***	-0.0543***	0.0331**	-0.0381***
	(0.00628)	(0.01300)	(0.00688)	(0.00769)	(0.00795)	(0.01370)	(0.00864)
Observations	9,028	2,896	6,132	2,295	6,733	994	4,864
R-squared	0.986	0.985	0.986	0.990	0.984	0.983	0.983
Controls	NO	NO	NO	NO	NO	NO	NO
Product Dummies	YES	YES	YES	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES	YES	YES	YES
F Test	0.000	0.623	0.000	0.002	0.000	0.016	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. High quality products have ratings higher than 4. High visibility products have a normalized search index higher than 0,8. LR-LV are low rating and low search index products, HR-HV are high rating and high search rank products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 6: DiD Estimates of the Impact of PMG on Prices.

Products Prices (log)	(1) FULL SAMPLE	(2) L. RATING	(3) H. RATING	(4) L. VISIBILITY	(5) H. VISIBILITY	(6) LR-LV	(7) HR-HV
$T_{i,l,t} * P_{i,t}$	-0.0424***	-0.0108	-0.0270***	0.0322***	-0.0577***	0.0532***	-0.0398***
	(0.00629)	(0.01510)	(0.00693)	(0.00786)	(0.00799)	(0.01310)	(0.00879)
Observations	9,028	2,896	6,132	2,295	6,733	994	4,864
R-squared	0.986	0.986	0.986	0.992	0.985	0.989	0.983
Controls	YES	YES	YES	YES	YES	YES	YES
Product Dummies	YES	YES	YES	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES	YES	YES	YES
F Test	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. High quality products have ratings higher than 4. High visibility products have a normalized search index higher than 0,8. LR-LV are low rating and low search index products, HR-HV are high rating and high search rank products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7: DDD Estimates of the Impact of PMG on Prices.

Products Prices (log)	(1) DDD	(2) DDD
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}$	-0.0537***	-0.0556***
	(0.00808)	(0.00810)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. $HRHV_{i,l,t}$ is a dummy equal to 1 for high quality and high visibility products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Indeed, most of the predictions of theoretical models on the price effects of PMG policies rely on assumptions related to the presence of heterogeneous consumers. By classifying products on the base of consumers quality assessment, we indirectly assume that consumers are heterogeneous in terms of their preferences towards quality and their availability to pay a price premium for that. Similarly, the time spent for finding a product can indirectly select consumers according to their willingness to engage in search activity and it is reasonable to argue that such availability is directly correlated to price sensitiveness.

5.2 Robustness Analysis

In this section, we discuss empirical results obtained by conducting an in-depth robustness analysis of our results.

The first issue that we tackle is the possibility that the effects of the treatment speed up, stabilize, or mean revert over time. In order to explore this issue, we estimate a specification of Eq. (1) that includes lags à la Autor (2003) and takes on the following form:

$$\log Price_{i,l,t} = \alpha_{i,l} + \sum_{j=0}^{5+} \gamma_j \left(T_{i,l,t} * P_{i,t+j}\right) + X^T_{i,l,t}\beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \quad (3)$$

where $P_{i,t+j}$ assumes the value of 1 in year $t+j$, and 0 otherwise.
Specification (3) allows the PMG repeals to generate different effects over time. In order to lower the number of parameters of the model, we estimate the effect of a PMG shutdown from the implementation day ($j = 0$) until five days later and onward.

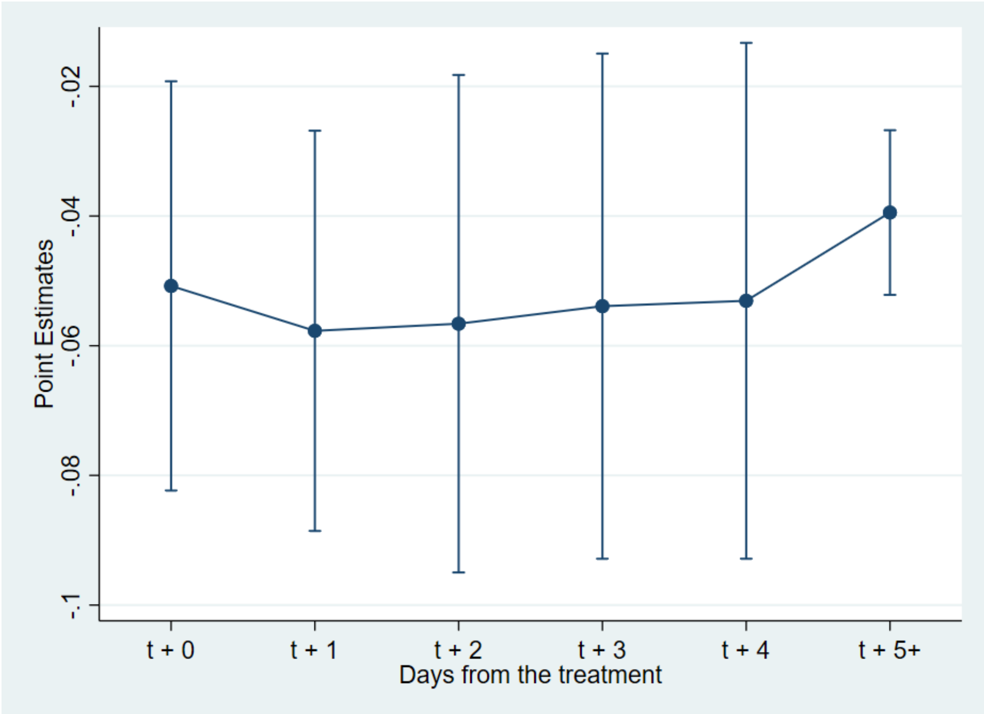
Table 8: DiD Estimates of the Impact of PMG on Prices with Lags à la Autor (2003).

Products Prices (log)	(1) DiD	(2) DiD
$T_{i,l,t} * P_{i,t+0}$	-0.0500***	-0.0508**
	(0.01610)	(0.01610)
$T_{i,l,t} * P_{i,t+1}$	-0.0566***	-0.0577**
	(0.01580)	(0.01570)
$T_{i,l,t} * P_{i,t+2}$	-0.0558***	-0.0566**
	(0.01960)	(0.01960)
$T_{i,l,t} * P_{i,t+3}$	-0.0530***	-0.0539**
	(0.01990)	(0.01990)
$T_{i,l,t} * P_{i,t+4}$	-0.0529***	-0.0531**
	(0.02030)	(0.02030)
$T_{i,l,t} * P_{i,t+5+}$	-0.0368***	-0.0395**
	(0.00646)	(0.00648)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

According to results shown in Table 8, coefficients related to lagged variables are always negative and statistically significant for the full sample. However, point estimates suggest that the impact of the treatment reaches its maximum after one day and starts decreasing afterwards. Figure 3 graphically shows parameter estimates patterns.

Figure 3: DiD Estimates of the Impact of PMG on Prices (Full Sample, with Controls) with Lags à la Autor (2003).



Another important issue in a DiD research design is the presence of pre-treatment common trends for treated and control samples. This assumption is indeed fundamental for the validity of the counterfactual policy evaluation analysis.
In order to explore this issue, we show in Figure 4 point estimates values and the relative confidence intervals of the difference in the level of prices between treated and control products from five days before the treatment to the day of the policy shutdown[23]. Plotted point estimates suggest that price levels for the treated platform do not seem to be significantly different from prices of the control platform before the treatment. This result provides evidence in favor of the validity of parallel trends assumption for our samples.

In order to further analyse this issue, we follow Autor (2003) and we estimate Eq. (3) after including some leads of the treatment interaction variable:

$$\log Price_{i,l,t} = \alpha_{i,l} + \sum_{j=-1}^{-5} \gamma_j \left(T_{i,l,t} * P_{i,t+j}\right) + X_{i,l,t}^{T}\beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \tag{4}$$

[23]In order to obtain these values, we estimate a panel model where we regress average daily price differences between the two samples on lead terms for five days before the treatment. We control for product fixed effects and daily fixed effects.

Figure 4: Price Differentials Between Treated and Control Groups Before the PMG Shutdown.

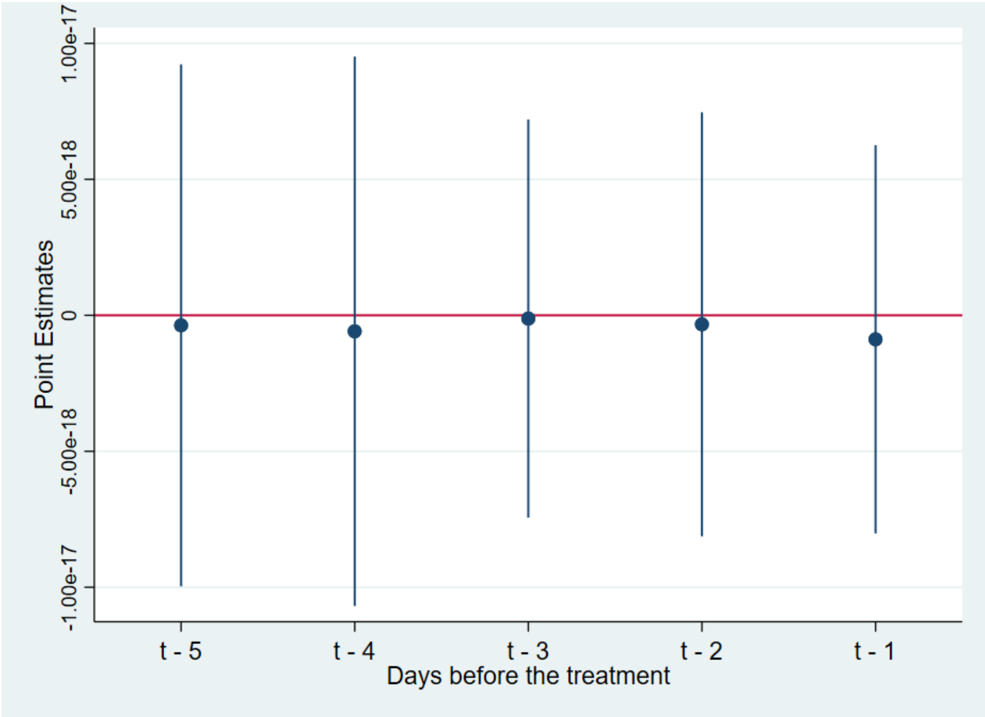


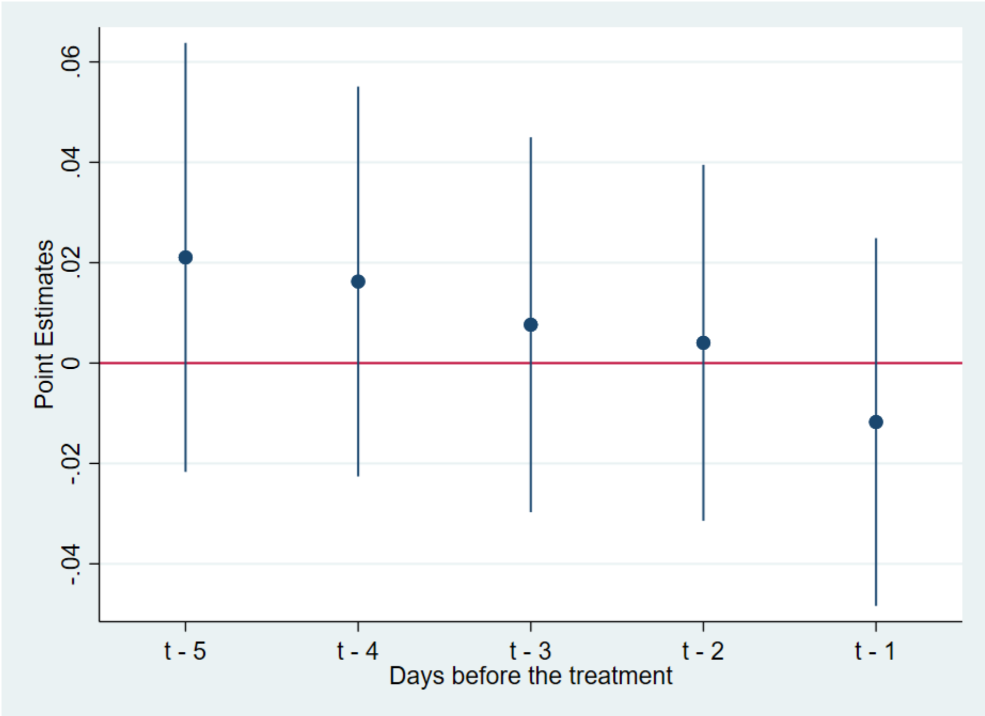
Table 9: DiD Estimates of the Impact of PMG on Prices with Leads à la Autor (2003).

Products Prices (log)	(1) DiD	(2) DiD
$T_{i,l,t} * P_{i,t-1}$	-0.0121	-0.0117
	(0.0187)	(0.0187)
$T_{i,l,t} * P_{i,t-2}$	0.0030	0.0040
	(0.0181)	(0.0181)
$T_{i,l,t} * P_{i,t-3}$	0.0067	0.0076
	(0.0191)	(0.0191)
$T_{i,l,t} * P_{i,t-4}$	0.0155	0.0162
	(0.0199)	(0.0198)
$T_{i,l,t} * P_{i,t-5}$	0.0205	0.0211
	(0.0219)	(0.0218)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.842	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

If leads coefficients turn to be statistically significant, there may be anticipatory effects and a failure in the parallel trend assumption. According to Table 9 and Figure 5, estimated coefficients of the anticipatory effects are not statistically significant,

Figure 5: DiD Estimates of the Impact of PMG on Prices (Full sample, with Controls) with Leads à la Autor (2003).



thus providing further evidence in favor of the existence of a parallel trend between treatment and control sample.

In order to extend our robustness analysis, we implement a complete set of placebo tests. We first estimate our baseline and DDD specifications by introducing artificially timed treatments and artificially treated subjects. Subjects and treatments fake assignments are drawn from two Bernoulli distributions with parameters p (probability of success) derived from the sample distributions of $Treated_{i,l,t}$ and $Post_{i,t}$ respectively. Within this setting, we should not observe any significant effect of PMG repeals on prices. Comfortingly, results reported in Table 10 confirm this prediction. Next, we conduct another falsification test by estimating our models after substituting the dependent variable with a placebo outcome that should not be affected by PMG shutdown. In particular, we generate fake product prices drawn by random distributions resembling sample ones (same mean and variance). Results shown in Table 11 confirm the absence of any impact of PMG repeals on fake outcome.

In order to analyze if our main results are robust to the exclusion of a particular product we estimate the baseline model (1) after dropping one product at a time. Results[24] suggest that this is not the case and confirm all previous findings. In the same spirit, we estimate equation (1) after balancing the panel dataset[25] and all results are confirmed. Finally, it is worth noting that results do not change if we

[24]Results, not reported, are available from the authors upon request.

[25]Precisely, we drop first 34 days in which we observe only some products; results are available upon request.

Table 10: DiD and DDD Estimates of the Impact of *Fake* Implementation Period on Prices for *Fake* Treated/Control Samples.

Products Prices (log)	(1) DiD	(2) DiD	(3) DDD	(4) DDD
$T_{i,l,t} * P_{i,t}(Fake)$	0.0018	0.0019		
	(0.00262)	(0.00261)		
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}(Fake)$			0.0033	0.0033
			(0.00403)	(0.00403)
Observations	9,028	9,028	9,028	9,028
R-squared	0.986	0.986	0.986	0.986
Controls	NO	YES	NO	YES
Product Dummies	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES
F Test	0.502	0.000	0.418	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 11: DiD and DDD Estimates of the Impact of PMG on *Fake* Prices.

$Fake$ Products Prices (log)	(1) DiD	(2) DiD	(3) DDD	(4) DDD
$T_{i,l,t} * P_{i,t}$	-0.0011	-0.0011		
	(0.00154)	(0.00154)		
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}$			-0.0013	-0.0013
			(0.00203)	(0.00204)
Observations	9,028	9,028	9,028	9,028
R-squared	0.999	0.999	0.999	0.999
Controls	NO	YES	NO	YES
Product Dummies	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES
F Test	0.466	0.722	0.516	0.746

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in in parentheses *** p<0.01, ** p<0.05, * p<0.1

compute bootstrapped standard errors at product level.

6 Conclusions

In this work we empirically investigate the effects of Price Matching Guarantees (PMG) commercial policies on U.S. online consumer electronics prices by applying a

Difference-in-Difference research design.

Estimates conducted over a sample of product prices, observed on the NewEgg platform between May and October 2018, provide evidence in favor of an average price reduction of about 3.9% after the interruption of the PMG policy. In order to have a more detailed picture of the issue, we investigate if such effect is heterogeneous across products. In particular, we focus on products features that might affect the outcome of PMG policies and that can be recovered exclusively on online markets. Platform data allow us to obtain information from User Generated Contents (UGC), like product popularity, product quality and online search visibility (Google Search Rank); indeed, we believe that these product characteristics might provide indirect information on consumers heterogeneity. Estimates conducted on specific sub-samples show that when PMG are interrupted, low quality (low search rank) products experience a price increase of about 3.4%, while for high quality (high visibility) products a price reduction of about 3.7% is observed.

These findings are in line with the lack of unambiguous predictions of the theoretical literature and are consistent with models predicting anti-competitive effects of PMG policies and with those interpreting such policies as a price discriminating device. Theoretical models predicting anti-competitive effects of PMG, suggest that such policies might induce higher prices in oligopoly markets (as the online consumer electronics) by sustaining collusion. In particular, online retailing platforms can easily monitor competitors prices trough price-tracking systems and can react faster to price signals, if compared to brick and mortar retailers. This possibility might sustain collusion by decreasing information asymmetries among competitors and reducing detection lags. On the other side, buyers' sensitivity to product quality and the willingness to engage in search activity can indirectly identify those customers whose demand is more rigid, thus allowing price discrimination practices. Indeed, e-commerce allows platforms to easily recover information on buyers, thanks also to Users Generated Contents, thus favoring discrimination policies.

Models that predict anti-competitive effects of PMG on prices are well suited to explain the results for high quality and visible products. The demand of such products is high and stable and consumers are likely to be available to pay a price premium. Such features, together with easily detectable price signals,make collusion more sustainable. Thus, PMG policies might be an invitation to collude that can be quickly and easily captured by competitors. However, it is worth noting that our analysis does not allows us to support such theoretical interpretation of the results since we do not analyse NewEggs competitors' behavior.
Our empirical results are also consistent with theoretical models arguing that PMG act as price discrimination tools. Indeed, such theoretical explanation requires a significant percentage of consumers invoking PMG rights; unfortunately, we do not have data on PMG redemption frequency. However, Moorthy and Winter (2006) find

redemption rates ranging between 5% and 25% on a sample of 46 retailers operating in the United States and in Canada. It is reasonable to assume that online markets redemption rates can be similar to physical ones, thus providing support to the price discrimination interpretation of PMG policies.

References

Arbatskaya, M., Hviid, M. and Shaffer, G. (2000), 'Promises to match or beat the competition: Evidence from retail tire prices', advances in applied microeconomics (advances in applied microeconomics, volume 8)'.

Arbatskaya, M., Hviid, M. and Shaffer, G. (2006), 'On the use of low-price guarantees to discourage price cutting', *International Journal of Industrial Organization* **24**(6), 1139–1156.

Autor, D. H. (2003), 'Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing', *Journal of labor economics* **21**(1), 1–42.

Baye, M. R. and Kovenock, D. (1994), 'How to sell a pickup truck:'beat-or-pay'advertisements as facilitating devices', *International Journal of Industrial Organization* **12**(1), 21–33.

Belton, T. M. (1987), 'A model of duopoly and meeting or beating competition', *International Journal of Industrial Organization* **5**(4), 399–417.

Byrne, D. P. and De Roos, N. (2019), 'Learning to coordinate: A study in retail gasoline', *American Economic Review* **109**(2), 591–619.

Cabral, L., Dürr, N., Schober, D. and Woll, O. (2018), Price matching guarantees and collusion: Theory and evidence from germany, Technical report, Working Paper, New York University.

Cameron, A. C. and Miller, D. L. (2015), 'A practitioner's guide to cluster-robust inference', *Journal of human resources* **50**(2), 317–372.

Chen, Z. (1995), 'How low is a guaranteed-lowest-price?', *Canadian Journal of Economics* pp. 683–701.

Chilet, J. A. (2018), 'Gradually rebuilding a relationship: The emergence of collusion in retail pharmacies in chile'.

Chung, H. S., Kim, M. et al. (2016), 'Low-price guarantees and pricing behavior: evidence from hypermarkets in korea', *Economics Bulletin* **36**(2), 1223–1229.

Coad, A. (2009), 'On the distribution of product price and quality', *Journal of Evolutionary Economics* **19**(4), 589–604.

Constantinou, E. and Bernhardt, D. (2018), 'The price-matching dilemma', *International Journal of Industrial Organization* **59**, 97–113.

Corts, K. S. (1997), 'On the competitive effects of price-matching policies', *International Journal of Industrial Organization* **15**(3), 283–299.

Doyle, C. (1988), 'Different selling strategies in bertrand oligopoly', *Economics Letters* **28**(4), 387–390.

Edlin, A. S. (1997), 'Do guaranteed-low-price policies guarantee high prices, and can antitrust rise to the challenge?', *Harvard Law Review* pp. 528–575.

Ellison, G. and Ellison, S. F. (2009), 'Search, obfuscation, and price elasticities on the internet', *Econometrica* **77**(2), 427–452.

Gorodnichenko, Y. and Talavera, O. (2017), 'Price setting in online markets: Basic facts, international comparisons, and cross-border integration', *American Economic Review* **107**(1), 249–82.

Haruvy, E. and Leszczyc, P. T. P. (2016), 'Measuring the impact of price guarantees on bidding in consumer online auctions', *Journal of Retailing* **92**(1), 96–108.

Hay, G. A. (1981), 'Oligopoly shared monopoly and antitrust law', *Cornell L. Rev.* **67**, 439.

Hess, J. D. and Gerstner, E. (1991), 'Price-matching policies: An empirical case', *Managerial and Decision Economics* **12**(4), 305–315.

Hviid, M. and Shaffer, G. (1999), 'Hassle costs: the achilles' heel of price-matching guarantees', *Journal of Economics & Management Strategy* **8**(4), 489–521.

Hviid, M. and Shaffer, G. (2010), 'Matching own prices, rivals'prices or both?', *The Journal of Industrial Economics* **58**(3), 479–506.

Jain, S. and Srivastava, J. (2000), 'An experimental and theoretical analysis of price-matching refund policies', *Journal of Marketing Research* **37**(3), 351–362.

Logan, J. W. and Lutter, R. W. (1989), 'Guaranteed lowest prices: do they facilitate collusion?', *Economics Letters* **31**(2), 189–192.

Lu, Y. and Wright, J. (2010), 'Tacit collusion with price-matching punishments', *International Journal of Industrial Organization* **28**(3), 298–306.

Mago, S. D. and Pate, J. G. (2009), 'An experimental examination of competitor-based price matching guarantees', *Journal of Economic Behavior & Organization* **70**(1-2), 342–360.

Moorthy, S. and Winter, R. A. (2006), 'Price-matching guarantees', *The RAND Journal of Economics* **37**(2), 449–465.

Moorthy, S. and Zhang, X. (2006), 'Price matching by vertically differentiated retailers: Theory and evidence', *Journal of Marketing Research* **43**(2), 156–167.

Nalca, A., Boyaci, T. and Ray, S. (2010), 'Competitive price-matching guarantees under imperfect store availability', *Quantitative Marketing and Economics* **8**(3), 275–300.

Png, I. P. and Hirshleifer, D. (1987), 'Price discrimination through offers to match price', *Journal of Business* pp. 365–383.

Pollak, A. et al. (2017), Do price-matching guarantees with markups facilitate tacit collusion? theory and experiment, Technical report.

Salop, S. C. (1986), Practices that (credibly) facilitate oligopoly co-ordination, *in* 'New developments in the analysis of market structure', Springer, pp. 265–294.

Stallkamp, M. and Schotter, A. P. (2019), 'Platforms without borders? the international strategies of digital platform firms', *Global Strategy Journal* .

Wilhelm, S. (2016), Price-matching strategies in the german gasoline retail market, Technical report, Working Paper, Goethe Universität Frankfurt, erhältlich unter: http://ssrn . . . .

Wu, C., Wang, K. and Zhu, T. (2015), 'Can price matching defeat showrooming', *University of California, Haas School of Business, Berkeley* .

Zhuo, R. (2017), 'Do low-price guarantees guarantee low prices? evidence from competition between amazon and big-box stores', *The Journal of Industrial Economics* **65**(4), 719–738.

Appendix

Table A1: Sub-Categories List.

Sub - Categories	# products
CPU Processor	3
Computer Case	2
Mobile Phone	1
Scanner	2
Speaker	2
Motherboard	1
Monitor	3
Headset	1
USB Flash	1
CPU Cooler	1
Speaker for Domotic	1
Tablet	1
Desktop PC	1
Laptop PC	1
Power Supply	1
Printer	2
Memory Card	2
Hard Disk	1
Smart Thing Domotic	2

Table A2: Products List.

Products Titles
AMD Ryzen 5 1500X Processor
Corsair Crystal Series 570X RGB - Tempered Glass; Premium ATX Mid-Tower Case
BlackBerry PRIV (32GB) Verizon Factory Unlocked Phone
Fujitsu fi-7160 Color Duplex Document Scanner
Fujitsu ScanSnap S1300i Instant PDF Multi Sheet-Fed Scanner
Philips BT50B/37 Wireless Portable Bluetooth Speaker
Asus ROG MAXIMUS VIII FORMULA DDR4 ATX Motherboards
ASUS VS247H-P 23.6 Full HD 1920x1080 2ms HDMI DVI VGA Monitor
Samsung Hmd Odyssey Windows Mixed Reality Headset
Samsung 128GB BAR (METAL) USB 3.0 Flash Drive
Corsair CW-9060025-WW Hydro Series Liquid CPU Cooler
Echo Dot (2nd Generation) - Smart speaker with Alexa - Black
ASUS VivoMini Mini PC
Dell XF9PJ Latitude 7490 Notebook
Intel Core i7-8700 Desktop Processor 6 Cores
AMD Ryzen 7 2700X Processor Wraith Prism LED Cooler
Corsair RMx Series RM850 x 80 PLUS Gold Fully Modular ATX Power Supply
ASUS 24-inch Full HD FreeSync Gaming Monitor
Brother Monochrome Laser Printer; Compact All-in One Printer
Team 64GB microSDXC UHS-I/U1 Class 10 Memory Card with Adapter
LG Electronics 21.5 Screen LED-Lit Monitor
HP LaserJet Pro M227fdw All-in-One Wireless Laser Printer
Logitech Z313 Speaker System + Logitech Bluetooth Audio Adapter Bundle
PNY CS900 960GB 2.5 Sata III Internal Solid State Drive (SSD)
Samsung SmartThings ADT Wireless Home Security Starter Kit
Samsung SmartThings Smart Home Hub
Rosewill 2U Server Chassis Server Case (RSV-2600)
Corsair Apple Certified 16GB (2 x 8GB) DDR3 1333 MHz (PC3 10600) Laptop Memory
Acer Iconia One 10 NT.LDPAA.003 10.1-Inch Tablet

Roads to Innovation: Evidence from Italy

This is the result of joint work with

Anna Bottasso*, Maurizio Conti* and Marta Santagata*
*University of Genoa

April 2020

Abstract

In this study we leverage on the ancient Roman roads network as a source of exogenous variation in order to identify the causal effect of the modern highways network on innovation (proxied by patent counts) using Italian NUTS-3 regional data. We find that an increase in the highways stock in a region by 10 per cent causes an increase in the number of patents by about 1 per cent. We also find evidence of important heterogeneous treatment effects associated to the type of patent technology and province population density. Moreover, we find a declining effect over time, possibly associated to the introduction of ICT. Finally, we cannot exclude the existence of important negative spillovers across provinces that might suggest the relevance of reorganization of innovative activity across space.

1 Introduction

The role of transport infrastructure investments in fostering growth has been extensively studied in the economics and regional science literature. As documented by literature surveys, like Ferrari et al. (2019) and meta-analysis, like Melo et al. (2013) and Bom and Ligthart (2014), transport infrastructures have been found to display significant impacts on different economic outcomes. A reduction in transport costs associated to transport infrastructure investments can generate higher productivity of other inputs and lower production costs, can increase trade and competition by enlarging relevant markets and can favor the exploitation of scale economies. Moreover, greater accessibility contributes to raise the market potential of different locations, thus affecting the spatial allocation of human capital and economic activities (agglomeration economies); in particular, transport infrastructures can facilitate knowledge creation and diffusion. Indeed, more recent endogenous growth theory models rests on knowledge spillovers as one of the most important engine of growth (Romer, 1990; Aghion and Howitt, 1990; Acemoglu and Akcigit, 2012). Hence, it becomes crucial to understand if transport infrastructure investments, among other possible policy tools, are able to stimulate innovation and knowledge spillovers that are often constrained by geography (Jaffe et al., 1993). With few exceptions (e.g. Agrawal et al., 2017), this issue has been neglected by previous literature.

This study investigates the impact of road infrastructures (motorways) on the innovative capacity of Italian (NUTS-3) regions . This analysis requires to address a difficult identification issue linked to possible simultaneity between regional technological evolution and transport infrastructure investments. Indeed, such investments are typically not randomly allocated whenever governments tend to build infrastructures in low-income and low-innovation regions, or when high growth driven by local innovation fosters the demand for mobility and therefore the construction of highways. Moreover, there might be omitted factors that drive both infrastructure and innovation. In order to tackle this issue, we follow the historical route instrumental variable approach suggested by the urban and regional economics literature (Redding and Turner, 2015) and pioneered by Duranton and Turner (2012). This approach grounds on the idea that the presence of a transport network built in the past can be a good predictor for successive infrastructure investments[1].

In this study we consider the ancient Roman roads dating back to 117 A.D. as an instrument for the modern motorways endowment of the Italian regions[2]. Following the literature, we argue that the Roman road network is reasonably exogenous, given that Roman roads were built mainly for military purposes; therefore, conditionally on a set of geographic controls, we assume that there are not important local unobservables, that explain both the construction of Roman Roads in certain areas and their (very) long run patterns of growth. Our main result is that the stock of highways has a positive and significant impact on regional innovative activity. Estimates suggest that an increase of 10% in the length of the motorways network leads to a 1% increase in innovation as measured by regional patent fractional count[3]. Our main findings are confirmed when considering patents by region and technological fields as observation units and are robust to the inclusion of a set of geographic and inventor control variables. The impact of highways network is found to decline over time with the diffusion of information and communication technologies. Moreover, we find evidence of important heterogenous treatment effects, as we find that roads favor innovation particularly in fields characterized by a higher level of technological

[1]Studies that have followed this approach include Duranton et al. (2014), Duranton (2015), Agrawal et al. (2017), Baum-Snow et al. (2017) and Martincus et al. (2017).

[2]Such instruments has been firstly adopted by Garcia-López et al. (2015) when studying the impact of highways on the sub-urbanization of Spanish cities and have been successively employed in different studies like Percoco (2015), Holl (2016), Roca and Puga (2017), De Benedictis et al. (2018), Garcia-López (2019).

[3]Similar results are found by Agrawal et al. (2017) for US metropolitan statistical areas.

turnover and in regions where inventors are more scattered over the territory: this is exactly what we should find if we believe that roads foster innovation by making communication easier. Finally, results on possible displacement effects of transport infrastructure investments in one region are not conclusive, but seem to point in favor of the existence of negative spillovers on nearby regions' innovative activity.
This paper is organised as follow. Section 2 describes related literature, Section 3 presents our database and in Section 4 we describe the identification strategy. Results are presented in Section 5 which is followed by the conclusive section.

2 Related literature

This study is related to the literature on the effects of roads infrastructure on regional growth and productivity. The first study that properly addressed this issue is the one by Fernald (1999) on US data. His idea is that some industries rely more on road services so that they should be particularly affected by improvements in the road network. By applying this identification strategy he finds that regional productivity is positively affected by roads investments. Chandra and Thompson (2000) analyze US data at the county level and find that highways have a differential impact across industries and affect the spatial allocation of economic activity. They boost economic activity in the counties that they pass directly, although at the expenses of adjacent counties. Another important study conducted on US Metropolitan Statistical Area (MSA) by Duranton and Turner (2012) finds that an an increase in the stock of highways in the city leads to an increase in employment by about 1.5% after 20 years. Moreover, authors suggest that their result is unlikely to simply reflect the spatial reorganization of economic activity.
A recent paper by Ghani et al. (2016) evaluates the impact on productivity, employment, output and number of establishments of the so called Golden Quadrilateral (GQ) project, a recent major investment program which involved a massive upgrade of the GQ highways network. Authors finds that such investments significantly affected the growth of manufacturing activity. Another interesting study on the impact of the massive investments in highways in China is the one by Xu and Nakajima (2017) who find a positive effect of highways construction on investment and output, with notable differences across types of regions and industries.
Our study contributes to the literature that has analysed the impact of transport infrastructure on innovation. As already mentioned above, this issue has been analysed by very few works. In the seminal work by Agrawal et al. (2017) authors analyze the impact of interstate US highways on regional innovation. They apply an Instrumental Variable (IV) approach to overcome the endogeneity of highway endowment and find that 10% increase in interstate highways leads to a 1.7% increase in regional patenting activity over a five years period. In particular, they suggest that

roads facilitate knowledge creation and diffusion also by favoring knowledge flows within metropolitan statistical areas. Following the spirit of Agrawal et al. (2017), Wang et al. (2018), using an IV approach, examine the impact of road development on innovation in China. In particular, to overcome endogeneity issues about road endowment, they use the mean slope in a city to measure the relative cost of road construction. Authors conclude that a 10% improvement in road density increases the average number of approved patents per company by 0.71%[4].

Recently some authors have investigated the relationship between innovation and rail infrastructure (e.g. Yamasaki, 2017; Dong et al., 2018; Lin, 2017). In particular, Yamasaki (2017) analyses the effect of rail access on the adoption of steam energy in relation to the expansion of the Japanese rail network between the late 1800s and early 1900s. Based on a Difference in Difference (*DiD*) strategy together with an IV approach to overcome the endogeneity of railway construction[5], authors suggest that the growth of rail access from 1888 to 1892 accounts for 67% of the growth of steam energy from 1888 to 1902. Lin (2017) estimates a *DiD* model for a panel of Chinese cities observed over the period 2003-2013 in order to assess the impact of high-speed rail (HSR) on a number of economic outcomes, including patent applications, as a proxy for innovation activities within a city. Among other results, authors suggest that high-speed rail stimulates innovation by favoring greater scientific collaboration between cities and the diffusion of knowledge. Finally, in Dong et al. (2018), the relation between knowledge diffusion and the construction of China's high speed rail is assessed over the period 2006 - 2015. By instrumenting the construction of HSR with the spatial distribution of major military troop deployments in 2005 and Chinese railroad networks in 1962, authors show that in Chinese cities connected to the HSR network, researchers experienced a significant increase in productivity, in terms of quantity and quality of scientific publications.

Our study is the first one that analyze the impact of road infrastructure on regional innovation in Italy. Indeed, empirical evidence on this issue has never been provided other than on US data. In particular, we share the identification strategy based on historical route instrumental variable approach with Agrawal et al. (2017); however our observational units (NUTS 3 regions) are quite different from USA metropolitan statistical areas, from different perspectives, especially from a geographical point of view. Italian regions always share borders and are very heterogeneous in term of population density and economic development. Such characteristics allows us to better analyze the issue of spillover effects of transport investments on nearby regions.

[4] Studies based on firm level data include Li et al. (2017) who suggest that roads have a positive impact on productivity in China by using an approach à la Fernald (1999) and Holl (2016), who evaluates the impact of roads on total factor productivity using Spanish data over a time span characterized by a significant expansion in the highways network. The author finds a negative and statistically significant effect of distance from highways on firm productivity.

[5] Authors construct their instrument by calculating the "cost-minimizing route" between destinations using slope information to account for costs of construction.

3 Data

Our study analyze the relation between roads and innovation, as measured by patent fractional count per capita[6], on 95 Italian provinces (NUTS-3 regions) as defined in 1974[7]. The innovation literature recognizes patents as fundamental instruments of appropriation of the innovative activity; indeed, technologies with greater impact on welfare and economic development are more likely to be patented (Pakes and Griliches, 1980). However, patents measure inventions but do not measure all innovative activity (Smith, 2005) and not all inventions are patented. However, as argued by the innovation literature, patents are an effective measure of local technological capacity.

We recover annual data on patents from the European patent Office (EPO) database (EPO-Patstat) that includes bibliographical and legal status patent data on several countries at NUTS-3 regions level. Patent data refer to patent applications filed directly under the European Patent Convention or to patent applications filed under the Patent Co-operation Treaty and designating the EPO (Euro-PCT). A detailed set of information on applications, like the number of applicants and inventors and their characteristics, the relative technological IPC class of the patent[8] and NACE-2 statistical classification of economic activity are included. We recover patent data for the period 1978 - 2015 and we "regionalise" raw patent information by means of inventor address (NUTS3 codes). Data are finally categorised on the base of technological fields following the WIPO systematic technology classification, based on the codes of the International Patent Classification (IPC). In particular, we identify five patent classes, according to different technological fields, namely Electrical Engineering, Instruments, Chemistry, Mechanical Engineering and the residuals ones. Data are limited to 2015, since the two last years of available data underestimate application counts because of the delays in the publication of patent data (eighteen/twenty-four months since application)[9].

Turning to roads infrastructure regional endowment, we consider the total number of kilometres of motorways in each NUTS-3 region as provided by the Italian Central Institute of Statistics and the Automobile Club of Italy[10]. As regards data on the length of Roman roads, in particular on those defined as major/consular roads, we rely on the Digital Atlas of Roman and Medieval Civilization (DARMC), which provides georeferenced data at regional (NUT-3) level on the road network of the Ro-

[6]The geographical distribution of patent applications is assigned according to the inventor place of residence. If more than one inventor characterizes a patent, the patent application is distributed equally between all of them and consequently between their provinces, eluding thus biased counting.

[7]As the number of Italian provinces has been progressively grown in recent years, we consider only the 1974 local Administrative setting.

[8]WIPO IPC-based technology field classification. Source: WIPO IPC Technology Concordance Table.

[9]See Bronzini and Piselli (2016) for more details.

[10]https://ebiblio.istat.it/SebinaOpac/resource/statistica-degli-incidenti-stradali/IST0010868

man Empire in 117 AD[11]. We then calculate the length of the major Roman roads in each Italian NUTS-3 region. Figure 1 shows the resulting map; in particular, in the Italian peninsula the total length of major roads is almost 10,000 kilometres.

Figure 1: Roman Road Network in Italy: Major Roads.

Source: Authors' elaboration from McCormick et al. (2013)

Following Duranton and Turner (2012) and Agrawal et al. (2017), we include in the analysis a complete set of control variables, like (NUTS-3) regions surface, the difference between maximum and minimum altitude and an index of terrain ruggedness[12]. Table 1 shows the mean and the standard deviations of the main variables used in the study and Figure 2 depicts the territorial distribution of patents in 1988. It is worth noting that, since there are 6 provinces that do not have highways and for which the value of the major Roman roads is different from zero, our sample contains information on 89 provinces[13]. Indeed, we find that, even though some provinces have no motorways, they are nevertheless equipped with road infrastructure similar to them. These infrastructures are however not easily measurable, leading to potentially biased estimates.

[11]The main predecessor of this database is Talbert (2000) which provides maps of the entire Greek and Roman empires, covering the territory of over 75 modern countries.

[12]Authors' elaboration from Nunn and Puga (2012).

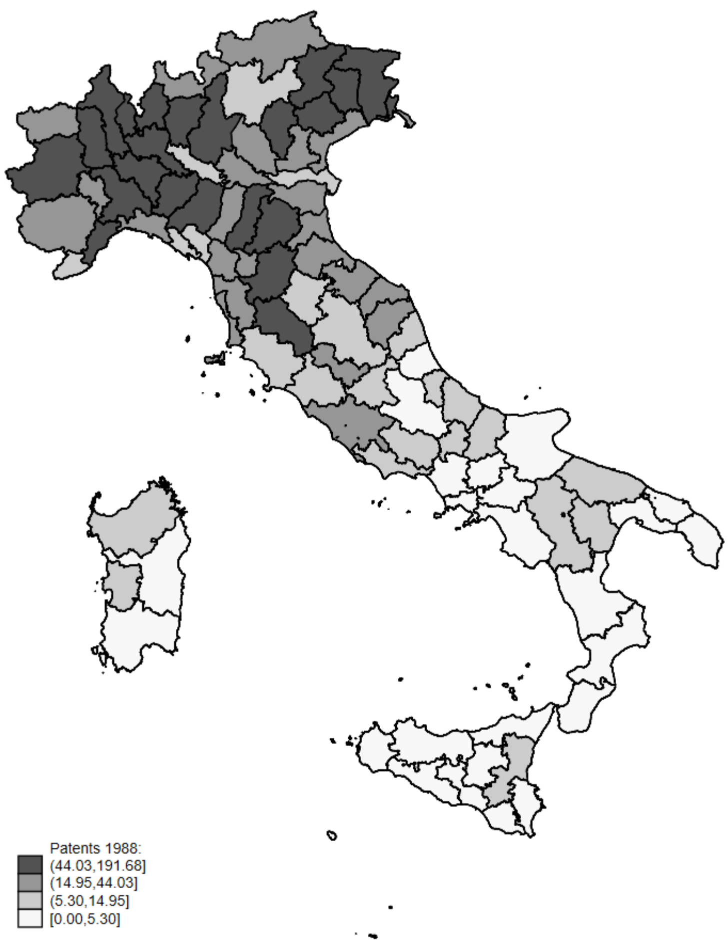
[13]In particular, we drop Brindisi, Matera, Agrigento, Ragusa, Siracusa and Grosseto from the sample.

Table 1: Summary Statistics for Main Variables.

Variables	Obs.	Mean	SD
Patent Fractional Count (per million people) 1983	89	14.62	14.33
Patent Fractional Count (per million people) 1988	89	31.36	35.91
Number of Inventors (per million people) 1988	89	21.01	24.95
Motorways lenght (km)	89	67.30	53.96
Major Roman Roads (km)	89	108.3	98.39
Surface (km^2)	89	2,931	1,594
Range (m)	89	758.8	510.6
Terrain Rugedness Index	89	2.299	1.605

Notes: The final number of observations we use for our computations is 89, since there are 6 provinces that do not have highways and for which the value of the major Roman roads is different from zero. Indeed, analysing these provinces, we find that, even though they have no motorways, they are nevertheless equipped with road infrastructure similar to them. These infrastructures are however not easily measurable, leading to potentially biased estimates. For these reasons, we drop Brindisi, Matera, Agrigento, Ragusa, Siracusa and Grosseto from the sample.

Figure 2: NUTS-3 Patent Fractional Count (per million people) 1988.



4 Identification Strategy

Following Agrawal et al. (2017), we consider an innovation model[14] where the innovative activity in each Italian province at time t is linked to the length of the motorways' system at time $t-j$:

$$\log Innov_{p,t} = \alpha + \beta\left(Motorways_{p,t-j}\right) + \gamma\left(Innov_{p,t-j}\right) + \varphi X_p + \upsilon_p \tag{1}$$

In Eq. (1), $Innov_{p,t}$ refers to the natural logarithm of our innovative measure in province p for patents at time t, while $Motorways_{p,t-j}$ is the length (logarithm) of the motorways' endowment for province p at time $t-j$. Following a common practice in innovation literature, also the lagged dependent variable is included as an input for future knowledge, in order to take into account the cumulative nature of the latter[15]. Moreover, the lagged dependent variable should account for most of the time invariant unobserved heterogeneity at the province level[16]. The econometric model includes a set of inventor and geography control variables, X_p. In particular, inventor controls include a measure for the number of inventors residing in each province, while geography controls include measures for surface, terrain asperity and elevation.

The β coefficient is our parameter of interest and refers to the impact of motorways endowment on local innovative activity: precisely, β measures the rate at which local innovative activity responds to motorways provision. Therefore, we interpret β as the parameter associating the endowment of motorways at time t with innovation growth for the period $t+j$.

Eq. (1) assumes that the analysis is conducted at an aggregate level, i.e. considering the sum of the patents originating in a given province, without variability between sectors. To allow an indication of the relevance of roads for knowledge in a given sector, we also estimate the following equation:

$$\log Innov_{p,f,t} = \alpha + \beta\left(Motorways_{p,t-j}\right) + \gamma\left(Innov_{p,f,t-j}\right) + \varphi X_{p,f} + \upsilon_{p,f} \tag{2}$$

where $Innov_{p,f,t}$ refers to the natural logarithm of our innovative measure in province p for patents in technological field f at time t.

As Drucker (2016) suggests, the effects of innovation activities, as R&D or public research, may spread over space and that spillovers are substantial up to 97 kilometres. For this reason, it is possible that innovation can be also displaced from one province to another. Indeed, the provision of motorways can potentially generate a zero-sum game among provinces. Therefore, in the spirit of Moretti and Wilson

[14]We provide an in depth explanation of the model in Appendix A.

[15]See Aghion and Howitt (1990).

[16]Since the diffusion of knowledge is supported by face-to-face interactions favoured by the presence of highways, one might also expect to find a measure of market potential. Nevertheless, we believe that the presence of the lagged dependent variable can control for this aspect.

(2014) and Agrawal et al. (2017) we extend our main model in Eq. (1) with the inclusion of spatial lag in order to analyze if the impact of roads infrastructure investments generate spillover effects in nearby regions. Specifically, we estimate the following model[17]:

$$\begin{aligned}\log Innov_{p,t} = \alpha + \beta\left(Motorways_{p,t-j}\right) + \gamma\left(Innov_{p,t-j}\right) + \\ +\theta SpatialMotorways_{p,t-j} + \varphi X_p + \upsilon_p\end{aligned} \tag{3}$$

where:

$$SpatialMotorways_{p,t-j} = \sum_{p\neq i}^{P} w_{pi} \log Motorways_{i,t-j} \tag{4}$$

The additional term $SpatialMotorways_{p,t-j}$ represents, for each province p, a weighted average of the motorways stock in other provinces i at time t; it is based on a row normalized matrix of the inverse of the distances multiplied by the levels of innovative capacity between any province p and i, with elements $w_{pi} = (Patent_p * Patent_i)/Distance_{pi}$ such that $\sum_{p\neq i}^{P} w_{pi} = 1$. Notice that to address possible endogenity issues we consider the innovative capacity at the begginning of the sample period, as suggested by Corrado and Fingleton (2012) and Bottasso et al. (2014). Thus, the spatial lag of motorways endowment account for the the possibility that infrastructural endowment in a certain province might have an impact on the performance of nearby provinces. The θ coefficient associated to the spatial lag allows us to detect the nature of possible spillover effects. A negative and significant sign of this parameter would suggest the presence of significant spatial displacement effects on regional innovation generated by road transport infrastructure investments in a particular region.

Estimating the effect of road transport infrastructure investments on regional innovative capacity is a quite challenging task in terms of identification strategy. Indeed, there might be simultaneity between regional technological evolution and transport infrastructure investments. In fact, such investments are typically not randomly allocated whenever governments tend to invest in lagging areas or in low-income and low-innovation regions, or when high growth driven by local innovation fosters the demand for mobility and therefore the construction of highways. Moreover, there might be omitted factors that drive both infrastructure and innovation. Thus, in Eq. (1) and (3) there might be correlation between unobservables, $\upsilon_{p,f}$, and the endowment of motorways in a province. In this case, estimating Eq. (1) and (3) by OLS could generate an underestimation of the causal impact of motorways on innovative performances. To take into account such a concern, we implement one of three approaches usually adopted in the applied literature and described by Redding and Turner (2015), namely, the historical route instrumental variables, the planned

[17]As for Eq. (1), also the Eq. (3) is estimated not only at an aggregate level but also allowing variability between technological fields: $\log Innov_{p,f,t} = \alpha + \beta\left(Motorways_{p,t-j}\right) + \gamma\left(Innov_{p,f,t-j}\right) + \theta SpatialMotorways_{p,t-j} + \varphi X_{p,f} + \upsilon_{p,f}$.

route instrumental variables and the inconsequential units approach.
In this work we rely on the first approach and we use the ancient Roman roads length dating back to 117 AD as an instrument for the current motorways endowment of the Italian provinces. Moreover, for each province p, we build the following instrument for the spatial lag presented in Eq. (4):

$$SpatialRomanRoads_{p,t} = \sum_{p \neq i}^{P} w_{pi} \log RomanRoads_{i,t} \tag{5}$$

which is a weighted average for Roman roads endowment in other NUTS-3 regions[18]. This approach has been pioneered by Duranton and Turner (2011, 2012) that choose the routes of major expeditions of exploration between 1835 and 1850 and the major rail routes in 1898 as instruments for MSA highways endowment. These instruments have also been used in subsequent works, e.g. Duranton et al. (2014), Duranton (2015) and Agrawal et al. (2017).
Other ancient transport network measures have been proposed by the literature as instruments for current roads endowment. Baum-Snow et al. (2017) analyze how urban railroads and highways have influenced urban form in Chinese cities by using the 1962 Chinese transport network as instrument, while Martincus et al. (2017) consider the Inca roads built before 1530 as an instrument for the 2000s Peruvian transport infrastructure.
Among the various instruments used in the literature, one in particular has seen numerous applications, i.e. the ancient Roman road network. One of the first uses of the ancient Roman network as an instrument for the current transport system is represented by Garcia-López et al. (2015). In their work, the authors investigate the effect of highways on the suburbanization of Spanish cities by relying on an IV approach where the instrument is represented by Spain historical roads, namely the old Roman roads and the roads built by the Bourbons in the XVIII century[19]. Subsequent works that have used the roman road network within an IV approach include Percoco (2015), Holl (2016), Roca and Puga (2017), De Benedictis et al. (2018), and Garcia-López (2019).
The validity of the Roman roads network as an instrument for modern roads endowment has been largely discussed by the aforementioned studies, both in terms of relevance and exogeneity. Regarding the first characteristic, i.e. the relevance, as common sense suggests, historical transport networks might be relevant because modern networks are not built in isolation from them (Garcia-López, 2019). This hypothesis has been tested by various studies. In particular a positive correlation between Roman roads and current Spanish highways has been shown by Garcia-López

[18]The weighting matrix has been constructed as the spatial matrix defined in Eq.(4) using Roman roads instead of modern roads.
[19]For a previous application see Garcia-López (2012), where the author limit the analysis to the metropolitan area of Barcelona.

et al. (2015); Garcia-López (2019); Holl (2016); Holl and Mariotti (2018). Also Percoco (2015) and De Benedictis et al. (2018) have found a strong relationship between current and Roman roads network in Italy[20]. We confirm these findings, as suggested by first stage regressions diagnostic. The first-stage F-statistics confirm that weak-identification bias is not a problem, since their values overall exceed thresholds proposed by Stock and Yogo (2005).

Another condition that our instrument has to satisfy is the exclusion restriction, whereby it should affect regional innovation only through its effect on the current highway endowment; moreover, it should be independent from contemporaneous level of innovation activity at the NUTS-3 regional level. Indeed, the validity of the instrument requires its exogeneity conditional on controls and, according to previous literature, this requirement seems to be satisfied by our chosen instrument. In particular, as argued by Dalgaard et al. (2018), Roman roads are strongly predetermined and, more in general, almost any ancient transport network may be considered as exogenous because of the time that has elapsed since it was built (Duranton and Turner, 2012). Moreover, the literature has identified military reasons as the main purposes of Roman road construction, thus excluding a direct economic reason for their location (e.g. Garcia-López et al., 2015; De Benedictis et al., 2018). However, since geography may have influenced the construction of both Roman roads and modern motorways, the exogeneity of our instruments is also based on the inclusion of a set of geographical controls, as in De Benedictis et al. (2018) and Garcia-López (2019).

5 Empirical Results

5.1 Main Results

Following Agrawal et al. (2017), among other authors, we estimate our model for the year 1988, which precedes the large diffusion of the Information and Communication Technology (ICT) and we use a five year lag both for the motorways stock and the lagged dependent variable. Indeed, the ICT revolution might have boosted or reduced the impact of road infrastructures on innovative activity, depending on the substitutability or the complementarity between personal interactions and ICT in the knowledge production process.

We first estimate Equations (1) and (3) relying on a measure of patent fractional count aggregated at NUTS-3 regional level, without considering the technological field of each patent. In column (1) of Table 2 we show OLS estimates that do not include controls, while in column (2) we report results on estimates that include ge-

[20] Percoco (2015) uses Roman roads as an instrument for road accessibility, as measured by the presence of a motorway exit, while De Benedictis et al. (2018) use Roman roads as an instrument for modern roads endowment.

ographic and inventors controls. Parameter estimates show evidence in favor of a positive correlation between 1983 motorways stock and the patent fractional count in 1988. In column (3) and (4) we present estimates of Eq. (3). The inclusion of the spatial lag, does not affect the positive correlation between highways and innovation; the coefficient of $logMotorways_{p,1983}$ remains positive and significant, while the coefficient of $SpatialMotorways_{p,1983}$ is not significant.

Table 2: Main Results for the Impact of Motorways on the Inovative Activity: OLS Estimation.

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Dependent Variable: log Patents Fractional Count (per capita) 1988				
$logMotorways_{p,1983}$	0.0930**	0.0967**	0.0896**	0.0928**
	(0.0387)	(0.0396)	(0.0388)	(0.0397)
$SpatialMotorways_{p,1983}$			0.410	0.396
			(0.397)	(0.413)
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	89	89	89	89
R-squared	0.777	0.783	0.779	0.784

Notes: All specifications include the lagged dependent variable. Inventor controls include the log of the inventors (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Results in Table 2 show a positive relation between road infrastructure and regional innovative activity; however such results might be biased because of endogeneity issues discussed above. In order to interpret our results in a causal way we need to rely on instrumental variable estimates.

Table 3 presents IV results and shows that the coefficient of $logMotorways_{p,1983}$ is not different from zero in column (1), while it is stable and significant across all other specifications, ranging from 0.10 to 0.13 values.

These results suggest that the endowment of highways has a positive and significant impact on regional innovative activity. In particular, estimates imply that an increase of 10% in the length of the motorways network in 1983 leads to an increase of about 1% in 1988 regional patent fractional count. Therefore, neither the inclusion of control variables nor the spatial lag influence the positive impact of highways. Moreover, the coefficient of the spatial lag variable ($SpatialMotorways_{p,1983}$) is not statistically different from zero, thus suggesting the absence of significant spatial displacement effects.

It is worth noting that, following Agrawal et al. (2017), the analysis focuses mainly on innovative capacity in 1988, but it is only since the early 1990s that Italy has

Table 3: Main Results for the Impact of Motorways on the Inovative Activity: IV Estimation.

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988				
$logMotorways_{p,1983}$	0.0873	0.102*	0.106*	0.130**
	(0.0534)	(0.0582)	(0.0558)	(0.0619)
$SpatialMotorways_{p,1983}$			-0.627	-0.742
			(0.461)	(0.497)
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	89	89	89	89
R-squared	0.777	0.782	0.767	0.771
F-statistic	30.45	24.28	10.95	8.549

Notes: All specifications include the lagged dependent variable. Inventor controls include the log of the inventors (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

increased its percentage share of applications to the EPO by reducing its distance from other industrialized nations (Larédo and Mustar, 2001). Therefore, it is reasonable to think that our estimates may suffer from a substantial underestimation of the innovative activity of small and medium enterprises, which represent the largest share of Italian companies and which are the ones that could benefit most from the expansion of roads.

In order to estimate Eq. (1), we consider one excluded instrument, namely the (log) length of ancient Roman roads in each province, to account for the possible endogeneity of current highways. Moreover, estimation of Eq. (3) (columns 3 and 4) requires an additional excluded instrument, i.e. $SpatialRomanRoads_p$ as defined in Eq. (5). Chosen instruments result to be relevant as the first stage F-statistics in Table 3 suggest. Indeed, the F-statistic ("Kleibergen-Paap rk Wald F statistic") exceeds all critical values for the weak instrument test based on TSLS size (Stock and Yogo, 2005).

It is worth noting that our results are also robust to the adoption of a Limited Information Maximum Likelihood (LIML) estimation method. In addition, the choice to "regionalise" patent information via the inventor address could lead to misleading estimates. In fact, the presence of a specific motorway may increase the patent activity observed in a specific province, p. However, some of the inventors living in that province may work in a neighbouring province, i. So the real impact of the motorway is to allow inventors to live in a given province, p, while working in the neighbouring one, i. By construction, in some cases there may be an increase in innovative activity in the province p when it is attributable to the province i. Therefore, to verify the robustness of our results we replicate the analysis using another criterion to construct the dependent variable, i.e. we use the address of the applicants. The main results are confirmed, as shown in the Table B1 in Appendix B.

To reduce possible unobserved heterogeneity and to reinforce the exclusion restriction assumption, we turn to analyze the relation between the highways stock and patent fractional count in each province p, at time t, in technological field f, according to Eq. (2).

Table 4: Results for the Impact of Motorways on the Innovative Activity by Province and Technological field: IV Estimation.

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 by Province and Tech. Field				
$logMotorways_{p,1983}$	0.0536	0.0759**	0.0879*	0.107**
	(0.0429)	(0.0366)	(0.0513)	(0.0425)
$SpatialMotorways_{p,1983}$			-0.993*	-0.744**
			(0.509)	(0.368)
Field FE	YES	YES	YES	YES
Geography	NO	YES	NO	YES
Inventors	NO	YES	NO	YES
Observations	356	356	356	356
R-squared	0.634	0.690	0.595	0.676
F-statistic	19.26	16.90	8.815	7.138

Notes: All specifications include the lagged dependent variable. Field FE refers to four technological fields classes. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors clustered at the NUTS-3 region level in parentheses *** p<0.01, ** p<0.05, * p<0.1

We consider four different technological classes, constructed by following WIPO systematic technology classification, namely Electrical Engineering, Instruments, Chemistry, Mechanical Engineering[21]. In Table 4 we report results obtained after replicating previous analysis on observation units defined at province and technological field level. Estimated models include also technological field fixed effects and a specific control for the number of inventors in each field[22].

All estimated specifications confirm previous findings and suggest the existence of a positive and significant effect of motorways endowment on innovation activity. The coefficients of $logMotorways_{p,1983}$ are slightly smaller with respect to those reported in Table 3. Moreover, the coefficients of the spatial lag variable are negative and significant, thus suggesting the existence of negative spillovers, whereby one province motorways endowment negatively affects innovation activity in nearby provinces.

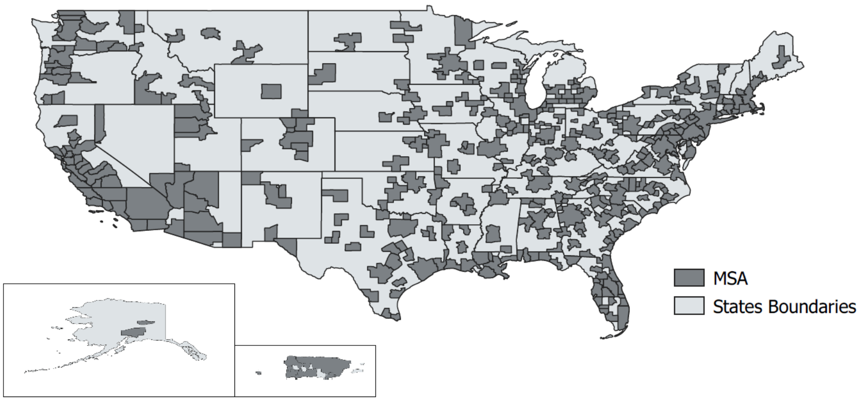
[21]The database is obtained by disaggregating patent fractional count on the basis of the technological class. It is worth noting that not all provinces in 1988 patented in all fields.

[22]As a consequence, inventors control variables include both the total number of inventors in each province and the number of inventors in each province in specific field f.

These results can be interpreted as evidence in favor of spatial reorganization of economic activities associated to road transport investments that generate zero-sum game among provinces. Such evidence on displacement effects is indeed not conclusive, given that our estimates based on observation units not taking into account patent technological field do not provide the same result.

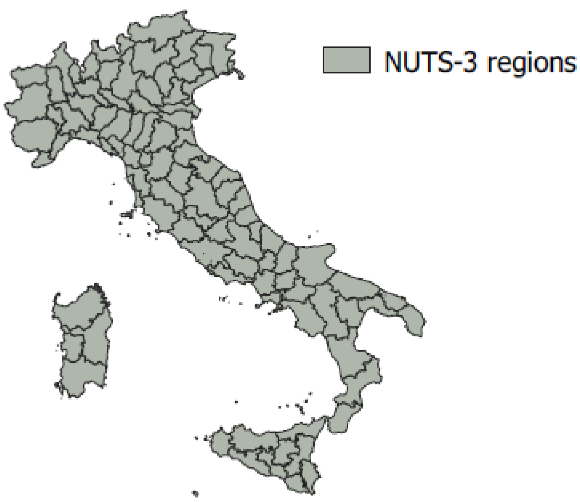
Our (not conclusive) results on the existence of spatial spillovers differ from Agrawal et al. (2017), where no significant displacement effects are detected. A plausible explanation for this discrepancy could lie in the different characteristics of US MSA and Italian NUTS-3 regions.

Figure 3: Metropolitan Statistical Area in the US.



Source: Authors' elaboration from US Census Bureau

Figure 4: NUTS-3 Regions in Italy.



Source: Authors' elaboration from ISTAT

Indeed, as shown in Figure 3, the US MSA are more sparse across the country and not all of them share borders. In this context, the presence of road infrastructures facilitate ideas' circulation within the same MSA, but does not affect the probability that those ideas might be used by someone else in a different MSA. On the other side, the geography of Italian provinces (Figure 4) significantly differs from that of MSA. Indeed, this administrative division covers the whole national territory and each province shares at least one border with another province. In this context, it seems more plausible that local innovation gains from motorways endowment in one region might be offset by losses in nearby regions. Nevertheless, our results have to be interpreted with caution since in this analysis we primarily assess the local impact of highways rather than the national impact of regional motorways. Indeed, as argued by Agrawal et al. (2017), conducing a precise evaluation of the effect of highways at national level is very challenging and would require a general equilibrium type of approach.

5.2 Heterogeneous Effects

In the spirit of Agrawal et al. (2017), we explore possible heterogeneous effects of roads on innovation. We first investigate whether roads differently affect innovation in industries characterized by a faster technological turnover. Since time is needed for additional knowledge to spread sufficiently to be useful for other innovators (Caballero and Jaffe, 1993), motorways should play an even greater role for knowledge diffusion in industries characterized by a faster process of *"creative destruction"*. In these industries the very high speed of technological turnover tends to limit the usefulness of knowledge flows, unless ideas can easily spread; in these contexts, the road network might potentially mitigate this problem by facilitating faster knowledge flows.

In the seminal work by Hall et al. (2001), authors measure the obsolescence of knowledge by technology field, finding that the speed of technological turnover varies across industries. In particular, according to the categories they developed, Computer & Communications and Electrical & Electronics industries show the highest technological turnover rate, while Drugs & Medical, Chemical and Mechanical industries present lower rates. In this spirit, we assume Electrical Engineering to be an "High-Tech Turnover" industry, while Instruments, Chemistry and Mechanical Engineering to be "Low-Tech Turnover" industries. We then split the sample according to this classification and we estimate Eq. (2) for each sub-sample. Results from IV estimation are presented in Table 5.

Estimates reported in column (1) show that a 10% increase in the motorways endowment in 1983 leads to an increase of 1% in patent fractional count in 1988 for industries where technological turnover is high. On the contrary, the coefficient of $logMotorways_{p,1983}$ in column (2) is not different from zero, thus suggesting that 1983 motorways does not affect local innovation levels in Low-Tech Turnover industries. This result is line with Agrawal et al. (2017) obtained don US MSA[23].

[23]As shown in the Appendix A, Agrawal et al. (2017) empirical framework relies on a model in which regional innovative activity is related to the level of highways through the relationship $K_t^* =$

Table 5: Results for the Heterogenous Impact of Motorways on the Innovative Activity at Technological Field Level: IV estimates High/Low Velocity Industries.

	(1)	(2)
Sample	High-Tech Turnover Industries	Low-Tech Turnover Industries
	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 by Tech. Field		
$logMotorways_{p,1983}$	0.104*	0.0604
	(0.0629)	(0.0457)
Field FE		YES
Geography	YES	YES
Inventors	YES	YES
Observations	89	267
R-squared	0.526	0.719
F-statistic	12.96	17.29

All specifications include the lagged dependent variable. High velocity patent field refers only to Electrical Engeneering and Communication technology while low velocity ones include Instruments, Chemistry and Mechanical Engeneering technologies. Field FE are then computed only for low velocity subsample. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patents, inventors and motorways before to taking the log to include observations with value of zero. Robust standard errors clustered at the NUTS-3 region level in parentheses *** p<0.01, ** p<0.05, * p<0.1

We further investigate weather the impact of motorways on innovation is heterogeneous according to different degrees of inventors density. Indeed, geographic proximity favors the development of knowledge flows, learning processes and relations between inventors, which in turn affect innovative activity. In this context, motorways system represents an important tool in facilitating the creation of networks between scientist and organisations.

We construct a measure of inventor density, $(\frac{Inventors_{p,1983}}{Surface_p})$, to capture the degree of inventors dispersion over the regional territory and we split the sample in High-Density or Low-Density regions depending on whether they are above or below the mean inventor density. We expect that the provision of highways would benefits more regions where interactions between inventors require large travelling distances.

Results, reported in Table 6, are in line with Agrawal et al. (2017) and show a zero impact of motorways in High-Density regions, while for Low-Density provinces a 10% increase in motorways endowment leads to a roughly 1.2% increase in patent fractional count.

$A * Motorways_t^{\alpha}$. In such a framework, authors demonstrate that the relation between the coefficient of $logMotorways_{p,t}$, β and the coefficient of the lagged dependent variable, γ, is $\beta = \alpha(1-\gamma)$. Since High-Tech Turnover industries are expected to exhibit lower values of γ, the β coefficient is expected to be larger.

Table 6: Results for the Heterogeneous Impact of Motorways on the Innovative Activity: IV estimates High/Low Density NUTS-3 Regions.

	(1)	(2)
Sample	High Density Regions	Low Density Regions
	IV	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988		
$logMotorways_{p,1983}$	-0.0340	0.117**
	(0.122)	(0.0584)
Geography	YES	YES
Inventors	YES	YES
Observations	18	71
R-squared	0.901	0.686
Geography	YES	YES
Inventors	YES	YES
F-statistic	4.860	27.48

All specifications include the lagged dependent variable. Low density NUTS-3 regions are the ones where the inventor density ($\frac{Inventors_{p,1983}}{Surface_p}$) is below the average value of the whole sample. Respectively, high density NUTS-3 regions show values above the average values. Inventor controls include the log of the inventors (per capita) in field f in each NUTS-3 region in 1983 and the total (log) number of inventors in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, inventor and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.3 ICT and Roads

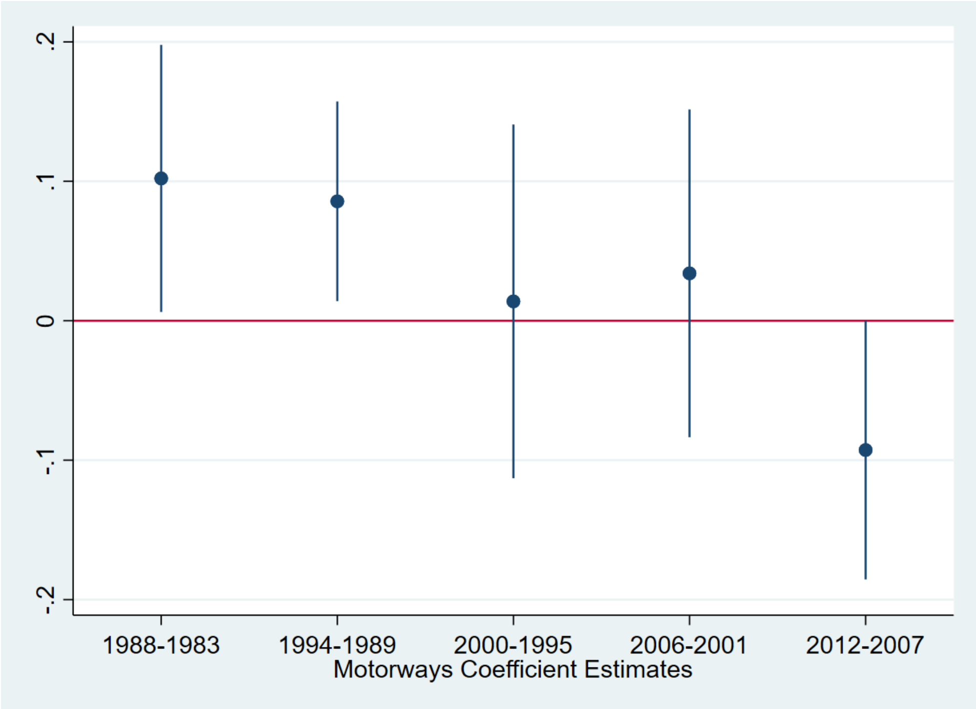
As already mentioned, in this study we focus on the relationship between regional innovative capacity observed in 1988 and the 1983 regional motorways stock. This choice accounts for the fact that, in these years, knowledge flows are supposed not to be affected by the Internet revolution of the 1990s. In fact, our study relies on the idea that roads have a prominent role in favouring knowledge flows, which in turns promote innovative capacity. Since the ICT has brought about revolutionary changes in the way people work, communicate, learn, spend time and interact (Jorgenson and Vu, 2016), it is interesting to investigate whether road infrastructures still have an important role in shaping knowledge flows when new communication technologies have been made available.

We turn to investigate the long term impact of highways, once the ICT revolution occured, by exploring the impact of 1983 regional highways endowment on innovation levels in different subsequent years. We estimate the following model for five different time periods (1989-1983, 1994-1989, 2000-1995, 2006-2001 and 2012-2007):

$$\log Innov_{p,t} = \alpha + \beta\left(Motorways_{p,1983}\right) + \gamma\left(Innov_{p,t-j}\right) + \varphi X_p + \upsilon_p \qquad (6)$$

Figure 5 shows IV estimates for the $Motorways_{p,1983}$ coefficient over five differ-

Figure 5: Effect of 1983 Motorways endowment on level of innovation across time.



Notes: the figure displays IV estimates for the motorways coefficient in five periods: 1988-1983, 1994-1989, 2000-1995, 2006-2001 and 2012-2007. The equation computed for each period is:$\log Innov_{p,t} = \alpha + \beta Motorways_{p,1983} + \gamma Innov_{p,t-5} + \varphi X_p + v_p$. Confidence intervals are reported at 90% level.

ent years. Compared to 1988, in 1994 the coefficient slightly decreases in magnitude, but remains positive and significant. This reflects the fact that the Internet revolution is still in its early phases, especially in Italy. The transition to digital media appears not yet complete and, therefore, the role of road infrastructure remains persistent. On the other hand, ten years after the ICT revolution the role of 1983 motorways have no longer effects on the levels of innovation in 2000. The same result is also evident for the year 2006. It is worth noting that the coefficient of $Motorways_{p,1983}$ for year 2012 is negative. The interpretation of this result, however, is very challenging and should deserve further investigations.

In this context, we interpret such result considering that, in the last twenty years, the focal issue about transport systems is changed from a quantitative perspective to a qualitative one. It is well-known that many Italian provinces face problems of traffic congestion and the presence of highways is not always indicative of adequate mobility. Furthermore, there is increasing discussion about the role of quality of life in the locational choices of activities with a high scientific and technological content. Consequently, given the advent of ICT, which allows instant communications around the world, the choice of innovators on where to work and live is increasingly based on motivations related to sustainability and quality of life (Knox and Mayer, 2013), rather than those purely related to the possibility of using road infrastructure. This trend could open up new opportunities for urban centres that make the combination of high technology and liveability a new model of urban development[24].

6 Conclusions

In this work we assess the impact of motorways endowment on innovative capacity. We explore such relation by estimating a model that links the local innovative capacity, proxied by patent fractional count, at time t with the stock of road infrastructures at time $t-j$. We conduct our analysis on a cross-section of Italian NUTS-3 regions. In particular, our main specification analyzes the relationship between regional innovative capacity in 1988 and regional road stock in 1983. We include in the model the lagged dependent variable and we control for the number of inventors and for a set of geographic charachateristics. We first conduct our analysis with an aggregate measure of patent fractional count in each province, with no distinction on patents technological fields. Secondly, to allow for a finer analysis, we account for technological fields categorisation. The main challenging issue about the estimation of our models arises from the possible endogeneity of highways stock. To deal with this problem, we follow the "historical instrumental variable" approach by using the length of the ancient Roman roads system dating back to 117 AD as an instrument for the length of current motorways. The chosen instrument results to be relevant as our first stage regressions suggest, thus confirming that modern network are not built in isolation from old roads network. Moreover, the validity of such instrument has been largely discussed in the previous literature. Instrumental variables estimates of our baseline specification indicates that 1983 highways network has a positive and significant impact on 1988 innovative capacity, in particular an increase of 10% in the length of motorways corresponds to a 1% increase in patent fractional count.

We further analyse the possible existence of spatial spillovers by introducing in our model a "spatial lag" that accounts for motorways stock in nearby provinces. Results show mixed evidence for potential displacement effects.

We further investigate whether the impact of highways on innovation shows some heterogeneity across provinces and industries. Estimates conducted on different sample splits suggest that motorways endowment has a positive effect on industries where the technological turnover is faster, while no effects are found for industries characterized by a lower technological turnover. Moreover, we find that highways benefit more inventors located in regions where local interaction requires long distances to be travelled.

Finally, we investigate the long term impact of 1983 motorways on 1994, 2000, 2006

[24] However, it is possible that results in the 2000s are biased because of the non-negligible expansion of the highways network that has characterized certain areas of the country in the 1990s: if highways construction has been relatively more important in regions characterized by a low density network in the 1980s, the coefficient of the 1983 highways network might be biased. We plan to test for this in an extension of the current paper.

and 2012 innovative capacity, finding that, due to the ICT revolution of the 1990s, motorways do not seems to display any causal effect on recent innovation outcomes.

References

Acemoglu, D. and Akcigit, U. (2012), 'Intellectual property rights policy, competition and innovation', *Journal of the European Economic Association* **10**(1), 1–42.

Aghion, P. and Howitt, P. (1990), A model of growth through creative destruction, Technical report, National Bureau of Economic Research.

Agrawal, A., Galasso, A. and Oettl, A. (2017), 'Roads and innovation', *Review of Economics and Statistics* **99**(3), 417–434.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A. and Zhang, Q. (2017), 'Roads, railroads, and decentralization of chinese cities', *Review of Economics and Statistics* **99**(3), 435–448.

Bom, P. R. and Ligthart, J. E. (2014), 'What have we learned from three decades of research on the productivity of public capital?', *Journal of economic surveys* **28**(5), 889–916.

Bottasso, A., Conti, M., Ferrari, C. and Tei, A. (2014), 'Ports and regional development: a spatial analysis on a panel of european regions', *Transportation Research Part A: Policy and Practice* **65**, 44–55.

Bronzini, R. and Piselli, P. (2016), 'The impact of r&d subsidies on firm innovation', *Research Policy* **45**(2), 442–457.

Caballero, R. J. and Jaffe, A. B. (1993), 'How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth', *NBER macroeconomics annual* **8**, 15–74.

Chandra, A. and Thompson, E. (2000), 'Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system', *Regional Science and Urban Economics* **30**(4), 457 – 490.

Corrado, L. and Fingleton, B. (2012), 'Where is the economics in spatial econometrics?', *Journal of Regional Science* **52**(2), 210–239.

Dalgaard, C.-J., Kaarsen, N., Olsson, O. and Selaya, P. (2018), Roman Roads to Prosperity: Persistence and Non-Persistence of Public Goods Provision, CEPR Discussion Papers 12745, C.E.P.R.
URL: *https://ideas.repec.org/p/cpr/ceprdp/12745.html*

De Benedictis, L., Licio, V., Pinna, A. M. et al. (2018), The long-term effects of the historical roman road network: trade costs of italian provinces, Working Paper CRENoS 201801, Centre for North South Economic Research, University of Cagliari and Sassari.

Dong, X., Zheng, S. and Kahn, M. E. (2018), The role of transportation speed in facilitating high skilled teamwork, Technical report, National Bureau of Economic Research.

Drucker, J. (2016), 'Reconsidering the regional economic development impacts of higher education institutions in the united states', *Regional Studies* **50**(7), 1185–1202.

Duranton, G. (2015), 'Roads and trade in colombia', *Economics of Transportation* **4**(1-2), 16–36.

Duranton, G., Morrow, P. M. and Turner, M. A. (2014), 'Roads and trade: Evidence from the us', *Review of Economic Studies* **81**(2), 681–724.

Duranton, G. and Turner, M. A. (2011), 'The fundamental law of road congestion: Evidence from us cities', *American Economic Review* **101**(6), 2616–52.

Duranton, G. and Turner, M. A. (2012), 'Urban growth and transportation', *Review of Economic Studies* **79**(4), 1407–1440.

Fernald, J. G. (1999), 'Roads to prosperity? assessing the link between public capital and productivity', *American Economic Review* **89**(3), 619–638.

Ferrari, C., Bottasso, A., Conti, M. and Tei, A. (2019), *Economic Role of Transport Infrastructure: Theory and Models*, Elsevier, Amsterdam.

Garcia-López, M.-A. (2012), 'Urban spatial structure, suburbanization and transportation in barcelona', *Journal of Urban Economics* **72**(2-3), 176–190.

Garcia-López, M.-À. (2019), 'All roads lead to rome... and to sprawl? evidence from european cities', *Regional Science and Urban Economics* **79**, 103467.

Garcia-López, M.-À., Holl, A. and Viladecans-Marsal, E. (2015), 'Suburbanization and highways in spain when the romans and the bourbons still shape its cities', *Journal of Urban Economics* **85**, 52–67.

Ghani, E., Goswami, A. G. and Kerr, W. R. (2016), 'Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing', *The Economic Journal* **126**(591), 317–357.

Hall, B. H., Jaffe, A. B. and Trajtenberg, M. (2001), The nber patent citation data file: Lessons, insights and methodological tools, Technical report, National Bureau of Economic Research.

Holl, A. (2016), 'Highways and productivity in manufacturing firms', *Journal of Urban Economics* **93**, 131–151.

Holl, A. and Mariotti, I. (2018), 'Highways and firm performance in the logistics industry', *Journal of Transport Geography* **72**, 139–150.

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *the Quarterly journal of Economics* **108**(3), 577–598.

Jorgenson, D. W. and Vu, K. M. (2016), 'The ict revolution, world economic growth, and policy issues', *Telecommunications Policy* **40**(5), 383–397.

Knox, P. and Mayer, H. (2013), *Small town sustainability: Economic, social, and environmental innovation*, Walter de Gruyter.

Larédo, P. and Mustar, P. (2001), *Research and innovation policies in the new global economy: An international comparative analysis*, Edward Elgar Publishing.

Li, Z., Wu, M. and Chen, B. R. (2017), 'Is road infrastructure investment in china excessive? evidence from productivity of firms', *Regional Science and Urban Economics* **65**, 116–126.

Lin, Y. (2017), 'Travel costs and urban specialization patterns: Evidence from china's high speed railway system', *Journal of Urban Economics* **98**, 98–123.

Martincus, C. V., Carballo, J. and Cusolito, A. (2017), 'Roads, exports and employment: Evidence from a developing country', *Journal of Development Economics* **125**, 21–39.

McCormick, M., Huang, G., Zambotti, G. and Lavash, J. (2013), 'Roman road network (version 2008)', *DARMC Scholarly Data Series, Data Contribution Series* **5**.

Melo, P. C., Graham, D. J. and Brage-Ardao, R. (2013), 'The productivity of transport infrastructure investment: A meta-analysis of empirical evidence', *Regional Science and Urban Economics* **43**(5), 695–706.

Moretti, E. and Wilson, D. J. (2014), 'State incentives for innovation, star scientists and jobs: Evidence from biotech', *Journal of Urban Economics* **79**, 20–38.

Nunn, N. and Puga, D. (2012), 'Ruggedness: The blessing of bad geography in africa', *Review of Economics and Statistics* **94**(1), 20–36.

Pakes, A. and Griliches, Z. (1980), 'Patents and r&d at the firm level: A first report', *Economics letters* **5**(4), 377–381.

Percoco, M. (2015), 'Highways, local economic structure and urban development', *Journal of Economic Geography* **16**(5), 1035–1054.

Redding, S. J. and Turner, M. A. (2015), Transportation costs and the spatial organization of economic activity, *in* 'Handbook of regional and urban economics', Vol. 5, Elsevier, pp. 1339–1398.

Roca, J. D. L. and Puga, D. (2017), 'Learning by working in big cities', *The Review of Economic Studies* **84**(1), 106–142.

Romer, P. M. (1990), 'Endogenous technological change', *Journal of political Economy* **98**(5, Part 2), S71–S102.

Smith, K. (2005), 'Measuring innovation'.

Stock, J. H. and Yogo, M. (2005), 'Testing for weak instruments in linear iv regression, in dwk andrews and jh stock, eds., identification and inference for econometric models: Essays in honor of thomas j. rothenberg. cambridge: Cambridge university press'.

Talbert, R. J. (2000), *Barrington Atlas of the Greek and Roman World: Map-by-map Directory*, Vol. 1, Princeton University Press, Princeton.

Wang, X., Xie, Z., Zhang, X. and Huang, Y. (2018), 'Roads to innovation: Firm-level evidence from people's republic of china (prc)', *China Economic Review* **49**, 154–170.

Xu, H. and Nakajima, K. (2017), 'Highways and industrial development in the peripheral regions of china', *Papers in Regional Science* **96**(2), 325–356.

Yamasaki, J. (2017), 'Railroads, technology adoption, and modern economic development: Evidence from japan'.

Appendix A

In this section we strictly follow Agrawal et al. (2017) in explaining the theoretical innovation model underlying Equations 1 and 3.
The deterministic innovation level in a province, K_t^*, is related to the level of motorways, $Motorways_t$, trough the relation $K_t^* = A * Motorways_t^{\alpha}$.
Authors argue that the adjustment rate depends on how far a region is from the deterministic level of innovation. The innovation adjustment rate is defined as $K_{t+j} =$

${K_t^*}^{1-\gamma} K_t^{\gamma}$, with $0 < \gamma < 1$. Authors show that the level of innovation at time $t+j$ is equal to $K_{t+j} = BR_t^{\beta} K_t^{\gamma}$, where $\beta = \alpha(1-\gamma)$ and $B = A^{1-\gamma}$. Taking the log of this latter equation, Eq. (1) is obtained. The parameter of interest, β, describes the rate at which knowledge creation responds to motorways endowment.

Appendix B

Table B1: Main Results for the Impact of Motorways on the Innovative Activity, robustness to choice of the dependent variable, OLS and IV Estimations.

	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Dependent Variable: log Patents Fractional Count (per capita) 1988 based on Applicants Address				
$logMotorways_{p,1983}$	0.185***	0.136*	0.182***	0.163*
	(0.0499)	(0.0818)	(0.0529)	(0.0850)
$SpatialMotorways_{p,1983}$			0.115	-0.802*
			(0.516)	(0.466)
Geography	YES	YES	YES	YES
Applicants	YES	YES	YES	YES
Observations	89	89	89	89
R-squared	0.642	0.638	0.642	0.626
F-statistic		21.51		10.95

Notes: All specifications include the lagged dependent variable. Applicant controls include the log of the applicants (per capita) in each NUTS-3 region in 1983. Geography controls include surface, terrain ruggedness and elevation. We add one to all patent, applicants and motorways before to taking the log to include observations with value of zero. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Public Funded R&D as a Device for Local Innovation? Evidence from Italian I.I.T. Foundation

Simone Robbiano*

*University of Genoa

April 2020

Abstract

This paper analyzes the effects on the regional economy of a prominent Italian place-based policy, the institution of Istituto Italiano di Tecnologia (IIT) research centre, by means of a novel identification strategy, the Synthetic Control Method (SCM). Such identification approach, unlike other counterfactual impact evaluation techniques, is based on the creation of an artificial control unit that not only follows the same pre-treatment trend as the treated unit, but even overlaps the same one.

In particular, through the SCM approach, the innovative and economic development (measured by patents per capita, number of local inventors and per capita GDP) of the treated region, namely Genoa, is compared with a set of Italian NUTS-3 control regions with the aim to estimate the causal effect of the location of IIT in 2006.

Estimates show significant effects of IIT presence on local patent activity and highly specialised human capital endowment in research. In particular, local patent activity is about 18% higher with respect to control regions. Moreover, in the post-implementation period, Genoa shows on average about 66 more inventors per million inhabitants, with a difference of 34% compared to the control regions. Turning to economic factors, there is then strong evidence that the intervention has also triggered a rise in GDP per capita.

These results suggest that the rises in research funding, development of human capital and formal competence have a significant impact on local economies over a 10 year post-intervention period.

1 Introduction

Economists and policy makers are well aware that each country is characterised by large and persistent geographical differences in economic performances and often try to address them by means of public investments and subsidies targeting lagging areas.
By significantly affecting the geographical displacement of economic activities, high-skilled workers and innovators, hence inducing changes in the industry mix of an area, place-based innovation policies are therefore becoming increasingly important in order to generate competitive advantages, stimulating private sector investments and supporting lagging regions through an innovation-driven economic transformation. Indeed, innovation is a fundamental driver for economic growth and the economic literature has reached a broad consensus on the issue[1].
Seminal studies from Romer (1990) and Aghion and Howitt (1990) assert the importance of the expansion of the knowledge base and illustrate its fundamental role for the economic growth. Similarly, Grossman and Helpman (1993) focus on the diffusion of new knowledge as a necessary condition for a long-term growth in production and wealth. Moreover, since Arrow (1972) seminal work, the literature agrees on the fact that incentives for private R&D investments are lower with respect to the social optimum[2].
Such policies may be formalized as financial support to socially optimal levels of R&D and innovation (including fiscal incentives for R&D), as support for capabilities and skills to generate innovation and as support to several forms of interaction and learning at the local level, including cluster support, an issue that has recently received increasing attention from economists and governments. Some policy makers have also tried to support the economic and innovative development of deprived areas by means of the placement or promotion of new universities and research institutes[3]. Indeed, academic and basic research is extensively believed to favour economic growth and productivity[4] due to a causal chain of effects between academic investments, knowledge spillovers and economic agglomeration. In particular, it is well known that world top-ranked universities are usually situated in economically prospering territories and the vast majority of high-tech clusters are unequivocally connected to prestigious research universities (Carlino et al., 2012); indeed, "Silicon Valley" and

[1]Such policies are frequent in the US (Kline and Moretti, 2014), and a growing effort is also directed by European Union to EU Cohesion Policy, mainly addressed to regional innovation. Generally, policy makers around the world commonly target huge part of public expenditure toward territories suffering poor economic performances.

[2]This because of knowledge spillovers, low appropriability of R&D, constrains in financing R&D projects caused by information asymmetries in financial markets (Harhoff, 2000).

[3]For instance, the US Morrill Act funds several land-grant colleges, agricultural and technical educational institutions usually located in rural counties, whose mission is to ensure higher education in deprived States.

[4]See Adams (1990).

"Route 128" owe their success as primary economic hubs to their closeness to Stanford and MIT (Jaffe, 1989).

Such place-based policies are usually very expensive and difficult to appraise and evaluate, due to their direct and indirect quantity effects. Therefore, it is essential to analyse the variations in the variables of interest caused by the policy related to a "no-intervention" alternative, and to evaluate the social value of these changes.
Although the impact of proximity to a research centre on the accumulation of knowledge, or more generally the agglomeration process and its economic effects, are well known issues for scholars, empirical evidence on the impact of opening new universities and research centres on the economic performances of lagging areas is scant and inconclusive. The extent to which knowledge spillovers from academic and public research institutions affect regional innovation and growth remains unclear and it is thus a significant empirical question.
In the most recent literature, the only existing papers that study the relationship between research centres and regional economic development, implementing reliable methods for causal inference, are Liu (2015) and Bonander et al. (2016). In the first paper, the author highlights a small impact of US universities on local manufacturing output over a period of 10 years, while huge increases in productivity arise when an 80-year period is observed. Instead, Bonander et al. (2016) find no effects of Swedish granting of research universities on local economic performances, on the number of students, patent applications and firm start-ups, while their study highlights robust increases in research competences (namely the number of PhDs and professorships).

By implementing a Synthetic Control Method (SCM) for comparative case studies (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015), this work adds to the above mentioned literature by investigating the causal effect on regional innovative and economic performances of a prominent Italian place-based policy, the institution of Istituto Italiano di Tecnologia (IIT). Established by law in 2003[5] and based in Genoa (Italy, EU) as a result of a political process[6], such scientific research centre promotes excellence in basic and applied research[7] and represents a probably exogenous policy change that allows the estimation of a causal effect of research centres on the regional economy.
The existence of IIT has led to the birth of several high tech start-ups operating in fields such as robotics, energy, health tech and new materials. In this context, in addition to public funds directly addressed to IIT own researches, potentially increases in private capital flows are likely to be attracted, especially in the form of FDI, coming from financial and industrial related sectors[8]. Besides to own basic and applied

[5]The Institute has been active since October 2005.

[6]https://www.repubblica.it/rubriche/la-scuola-siamo-noi/2016/02/29/news/la_fragilita_dell_iit_l_istituto_privato_che_comandera_la_ricerca_italiana-134509491/.

[7]https://www.iit.it/it/istituto/iit.

[8]https://www.iit.it/technology-transfer/start-ups.

research, IIT also promotes scientific activities commissioned by external firms, other public/private research institutions and public administrations, with the aim to transfer the results of its own technology research to the industrial productive fabric[9]. Therefore, IIT is likely to affect local growth and improving productivity, favouring knowledge accumulation, economic agglomeration and knowledge spillovers.
The main goal of the study is thereby to assess whether the institution of central scientific laboratories of IIT in 2006 in the Genoa metropolitan city[10] has had an impact on regional innovation and economic performances in subsequent years. In particular, the regional potential for innovation in 1980-2015 period is firstly analysed, using patents per capita and the number of local inventors as a proxy; subsequently, the effects of the presence of IIT on local GDP per capita are investigated.

The assessment of the effects of the IIT on Genoa economic performances is in any case challenging, since the scale, the location and intrinsic characteristics of such institutions are non-random in spatial terms; moreover, it should be taken into account the presence of possible unobservable characteristics that affect both the location of such new research centre and the potential increases in local research and economic performances. More specifically, aspects such as firms innovation, local productivity, the industrial and skill geographical structure might themselves affect research performances. Moreover, unobservable local shocks and characteristics are likely to create problems in the interpretation of the results of both private-sector activity and public-sector research activity. In particular, Genoa does not necessarily share similar economic characteristics and economic trends with other Italian provinces. Consequently, a comparison between the patterns of the variables of interest for the Genoa metropolitan city and the rest of the Italian provinces is likely to produce biased estimates.
Therefore, to construct a dependable control group, the SCM matching technique is implemented with the aim of achieving an increase in the quality of impact estimation. The SCM builds a synthetic control province for the Genoa metropolitan city. This "synthetic Genoa" captures the development (of the outcome of interest) of the real Genoa in the pre-treatment period through a weighted average of control provinces (such weights are chosen to guarantee the best quality of matching with the treated unit), replicating the outcome path that the treated unit would have experienced in the absence of the treatment[11]. Therefore, when the treated unit and the synthetic

[9]IIT technology transfer contributes to the innovation process also through licensing activities. `https://www.iit.it/technology-transfer/industrial-liaison`.

[10]The metropolitan cities are administrative divisions of Italy, operative since 2015. The metropolitan city, as defined by law, includes a large core city and the smaller surrounding towns that are closely related to it with regard to economic activities and essential public services, as well as to cultural relations and to territorial features. Such administrative division is comparable to a province, the NUTS-3 italian regions.

[11]The SCM approach is based on the creation of an artificial control unit that not only follows the same pre-treatment trend as the treated unit, but even overlaps the same one.

control one are matched on outcome and relevant economic covariates in the pre-treatment period, the divergence in outcome trajectories can be interpreted as the causal impact of the treatment.

Overall results suggest significant effects of IIT presence on the local innovation capacity (estimated gap ranges from 18% to 24% depending on the sample). Instead, considering the potential for innovation, estimates show how the intervention has triggered an increase in research skills: in fact, Genoa shows about 66 more inventors per million inhabitants than the synthetic one, with a relative average difference of 34%. Finally, evidence for a positive effect of the IIT on per capita GDP is also found. These results suggest that increased local funding for research, human capital development and formal skills have a significant impact on Genoa economy in the 10 years following the intervention.
Therefore, this work contributes to the literature by suggesting local spillovers from public research and confirming results of Liu (2015), who highlights how the increase in manufacturing productivity might reflect both a direct impact of university knowledge spillovers and general agglomeration economies. However, such results are contradictory to the paper of Bonander et al. (2016), which does not find significant effects of new granting research universities on the Swedish regional economy. In particular, authors highlight that the intervention has caused a rise in awarded PhDs and professorships, whereas no effects are detected for the number of students, patent applications, and firm start-ups. Therefore, the main findings of the analysis are in line with the lack of unambiguous predictions of the empirical literature.

The rest of the work is structured as follows. The next section focuses briefly on the literature on innovation, research and knowledge spillovers. Section 3 provides historical background and detailed information on the Italian Institute of Technology. Section 4 explains the identification strategy and the empirical methodology, the data sources, the dataset, variable construction and the data statistics. Empirical results are presented in Section 5, including robustness checks and placebo tests. Section 6 concludes.

2 Literature Review

A relevant part of the literature focuses on the role of economic agglomeration and knowledge spillovers in affecting local growth and improving productivity. Moreover, there is a strong debate among policy makers and researchers on the role that universities and research centres have in promoting innovation and regional growth, and consequently on whether or not to attract and motivate increased public and private spending on research (Drucker and Goldstein, 2007; Power and Malmberg, 2008).

Recently, the intersection among endogenous growth models and regional location models has gained increasing attention; the common thread linking these models is represented by the assumption that the production of knowledge tends to be geographically concentrated. Consequently, local innovativeness is likely to generate knowledge spillovers and positive technological externalities; such elements also affect the location of firms and people, thus inducing a dynamic process that is the engine of endogenous growth[12].

Indeed, innovation is primarily affected by new economic knowledge[13] (Audretsch and Feldman, 1996) and universities or research institutes are traditionally players that originate and stimulate the transmission of knowledge: the latter significantly contributes to industrial innovations (Mansfield and Lee, 1996).

Specifically, Goldstein et al. (1995) emphasize the mechanisms through which modern research institutes may potentially influence regional economic development. The authors mainly refer to the creation of knowledge and human capital, the relocation of existing know-how, the support to technological innovation, the potential increase in capital investment, the development of a regional leadership, the raise in knowledge infrastructure production and, finally, the influence on the regional milieu.

Universities and research organisations are clearly central players in the knowledge accumulation process, both in terms of basic research tough the creation of highly skilled workers. The literature generally recognizes how knowledge spillovers and human capital development could attract private sector research, scientists and high-tech firms. Precisely, a face-to-face interaction among research institutions and industries are essential elements for an effective transfer into production of research findings. This knowledge transfer often supports the creation of start-ups and/or high-tech firm branches in the neighbourhood of a research center. Consequently, local human capital benefits from the propensity of high-skilled workers to remain and work in the local area; moreover, new scientists and high-quality workers could be attracted from neighbourhood regions, further raising the level of human capital in the area.

Furthermore, agglomeration of knowledge and geographical proximity affect innovation also through the development of relationships between different organizations (Baptista, 1998; Hervas-Oliver and Albors-Garrigos, 2009), potentially rising the likelihood of joint projects (Guillain and Huriot, 2001). Precisely, geographical contiguity increases the number of interactions, simplifying the development of knowledge flows and learning processes, key elements for the transmission of tacit knowledge[14].

Geographical proximity decreases also uncertainty, search costs (Feldman, 1999) and

[12] Grossman and Helpman (1993), in their seminal work, highlight the agglomeration effects induced by free trade and localized knowledge spillovers.

[13] The growth theory supports the view for which the non-rivalrous nature of new knowledge explains growth in income per capita and the presence of increasing returns to scale (Aghion and Howitt, 2005; Jones, 2005).

[14] See Polanyi (2009), Amin and Wilkinson (1999), Torre and Gilly (2000).

transaction costs that firms suffer for joint projects: as a result, firms construct more stable and longer durable relationships (Bennett et al., 2000; Love and Roper, 2001) and benefit of increasing returns from collaboration (Izushi, 2003; Abramovsky and Simpson, 2011; Agrawal et al., 2017).
Therefore, inter-organizational interactions among public and/or private research institutions and firms are central drivers for innovation processes (Perkmann and Walsh, 2007); this result is also emphasized in the triple helix paradigm[15] (Etzkowitz et al., 1995) for regional innovation systems.
Obviously, geographical proximity is a necessary but not sufficient condition for an effective transmission of knowledge (Lane and Lubatkin, 1998): technological proximity[16] and organizational proximity[17] are thus necessary.

More recently, several papers have dealt with the possibility that location choices and growth could be jointly determined.
Black and Henderson (1999) first propose a model where there is a link between geography and growth. The authors highlight how urbanization generates knowledge spillovers, which are geographically localized, and in turn influence local economic endogenous growth.
Fujita and Thisse (1996, 2002, 2003), combine instead an endogenous growth model with a core-periphery one, with horizontally differentiated products, showing the existence of a core-periphery equilibrium. Precisely, in presence of sufficiently low transaction costs and tradable patents, firms and R&D facilities tend to be clustered in one region.
In an interesting model by Baldwin and Martin (2004), growth affects geography through cumulative causation processes and human, physical and knowledge accumulation; similarly, Minerva and Ottaviano (2009) propose a two region (North-South) model, focusing on the impact of public investments on growth and agglomeration patterns.

While the theoretical literature seems to be in agreement, the empirical literature shows instead a number of conflicting results, possibly in the light of large differences in methodological approaches.
Anselin et al. (1997) argue for the existence of spatial spillovers between US university

[15]The triple helix paradigm of innovation refers to a model where several interactions between academic research, firms and policy makers favour economic and social development. The basic hypothesis consists in the assumption that mutual influences between science, industry and governments are fundamental factors in an innovation process. The model highlights the mutual interactions between the typical functions of entrepreneurship, science (as academic spin-off) and government, which has the role of regulating the process. The relationships between such actors increase within this framework, each actor evolves to adopt some features of the other players, which then results in hybrid institutions and a growth in innovation capacity.

[16]Defined as "the level of overlap of the knowledge bases of two collaborating actors" (Lane and Lubatkin, 1998).

[17]Defined as the set of explicit or implicit routines that refer to organizational structure, organizational culture, performance measurements systems, language (Rallet and Torre, 1999).

research and high technology innovations. Similarly, Monjon and Waelbroeck (2003) find that mainly high-tech firms are advantaged from collaboration projects with French public research centres, emphasizing however the importance of knowledge spillovers for firms that imitate existing technologies or those that are involved in incremental innovation[18].

Darby et al. (2004), Belderbos et al. (2004) and Mohnen et al. (2006) focus on the positive effect of proximity to universities for US and European firms: authors highlight how firms that collaborate with universities show a higher innovation activity and sales increases for new or significantly improved products[19].

In turn, Andersson et al. (2004, 2009), Schubert and Kroll (2016) and Drucker (2016), using Sweden, German and US data respectively, analyse the positive impact of higher education institutions on regional environment, highlighting increases in local productivity, innovation levels, GDP per capita and employment.

Woodward et al. (2006), Abramovsky et al. (2007), Abramovsky and Simpson (2011) investigate instead whether the proximity of a research institute or highly rated university research departments, is related with the location of firms and R&D facilities. The authors find evidence of a slight positive relationship between public research and firm localization, with the geographic allocation of R&D laboratories that is skewed towards places with highly rated, industrially relevant university research departments.

An interesting work is also by Cowan and Zinovyeva (2013), who scrutinize how the creation of new Italian universities has had an impact on regional innovation during 1985–2000 period. The authors show that the location of new universities improved local innovation activity, with effects largely caused by the new high quality scientific research.

Kantor and Whalley (2014) analyse the relationship between knowledge spillovers from universities and labour income in US urban areas, finding evidence that a growth in university expenditure rises non-education sector wages. Moreover, they highlight how those firms that are technologically closer to public research benefit from higher knowledge spillovers, which are relevant in places with a high density of research-intensive institutes. Unlike this last work, Goldstein and Renault (2004), using a quasi-experimental approach, analyse variations in average earnings per job across US MSA[20], finding instead no evidence for a positive relationship between universities and local economic development.

[18]Similarly, Lööf and Broström (2008) show that primarily large manufacturing firms are affected.

[19]Similar results can be found in the paper of Nieto and Santamaría (2007), which focuses on the importance of technological cooperation networks for a higher degree of novelty in product development.

[20]The US Office of Management and Budget (OMB) has defined 384 metropolitan statistical areas (MSAs) for the United States and eight for Puerto Rico. The OMB defines a Metropolitan Statistical Area as one or more adjacent counties or county equivalents that have at least one urban core area of at least 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties.

Finally, Moretti et al. (2019) scrutinise the effects of public funded R&D on privately implemented R&D and the impact on productivity growth, by means of industry-country level data from OECD countries and firm level data from France. Specifically, they find that a rise in public funded R&D causes an increase in privately financed R&D and productivity gains[21]. Similarly, Valero and Van Reenen (2019) show how increases in the number of universities are associated with 0.4% higher future GDP per capita. Furthermore, authors highlight positive spillover effects from universities to neighbouring areas. They finally show that the effect of universities on growth is often mediated through a better supply of highly specialised human capital and higher innovation.

As mentioned above, only Liu (2015) and Bonander et al. (2016) implement reliable methods for causal inference.
Combining a Synthetic Control Method and an event study approach, Liu (2015) analyses the effects of US academic research[22] in the 1890s on several economic outputs, finding evidence of local spillovers from universities and agglomeration economies. Specifically, the author estimates small effects of US universities on local economic performances in 10 years, but a strong productivity growth over an 80-year period.
While Liu (2015) focuses on the effects of an historical intervention, Bonander et al. (2016) analyse the effectiveness of actual (1993-2011) Swedish research universities on the regional economy through several different measures of regional outcome[23]. Using a Synthetic Control Method as identification strategy, authors show how the presence of a research university improves the number of researchers in the region, but they find no effects on technological innovation, regional GDP per capita and compensation of employees during a 13-year follow-up period.

3 Italian Institute of Technology

The Italian Institute of Technology (IIT) is an Italian research centre based on the legislative decree 269/03 (transformed into law No. 326/2003), which has been initially conceived in 2003 for initiative of the Minister of Economy, located in the Italian city of Genoa as a result of a politic process[24].

IIT is supported by government funds with the aim of achieving technological and economic development through qualified basic and applied research. IIT is managed by a foundation that follows the rules of private law, as is the case of the Max Planck

[21] Similar results for private innovation can be found also in Toole (2012) and Azoulay et al. (2019).
[22] Precisely, the institutions of land grant universities in deprived rural States.
[23] Precisely, economic and technological outcomes, as well as intermediate university-related outcomes such as region-specific awarded doctoral degrees and number of professors.
[24] https://www.ilsecoloxix.it/economia/2013/01/18/news/i-baroni-della-ricerca-all-assalto-dell-iit-1.32294420.

Institute in Germany[25].

The institute has been active since October 2005 in Genoa: secondary research laboratories are presents in several national and international territories[26].
The research organisation in Genoa reckon on departments and laboratories that operate in many technological fields such as advanced robotics, drug discovery and development, neuroscience and brain technologies robotics, robotics, brain and cognitive sciences, nanochemistry, nanostructures, nanophysics, pattern analysis and computer vision[27].
In addition, IIT is present in several remote centres, where scientists collaborate with researchers at the university hosting the centre, chasing conjoint scientific aims for the institute and the university[28]. Figure 1 shows the IIT human resources endowment assigned to research.

In particular, research activities follow a specific strategic plan (currently based on 2018-2023 time-period and concerning Robotics, Nanomaterials, Lifetech and Computational Sciences, namely the 4 fundamental research domains[29] on which the activities of the institute are concentrated): this one consists of 16 scientific purposes, divided into 4 research domains (RDs).

i Robotics supports the developing of new hardware or software robotic platforms; in particular, there are 5 priorities, that are Mechatronics, Soft Robotics, Social Cognition and Human Robot Interaction, Biomedical Robotics and Intelligent Companion Robots.

ii Nanomaterials domain focuses with new sustainable and or biodegradable materials, nanocomposites, 2D materials, nanofabrication technologies and nano-devices, and new colloid chemistry approaches. In particular, research activities affect

[25]The choice of a Foundation as type of institutional government is ascribable to a consolidated legislative orientation.

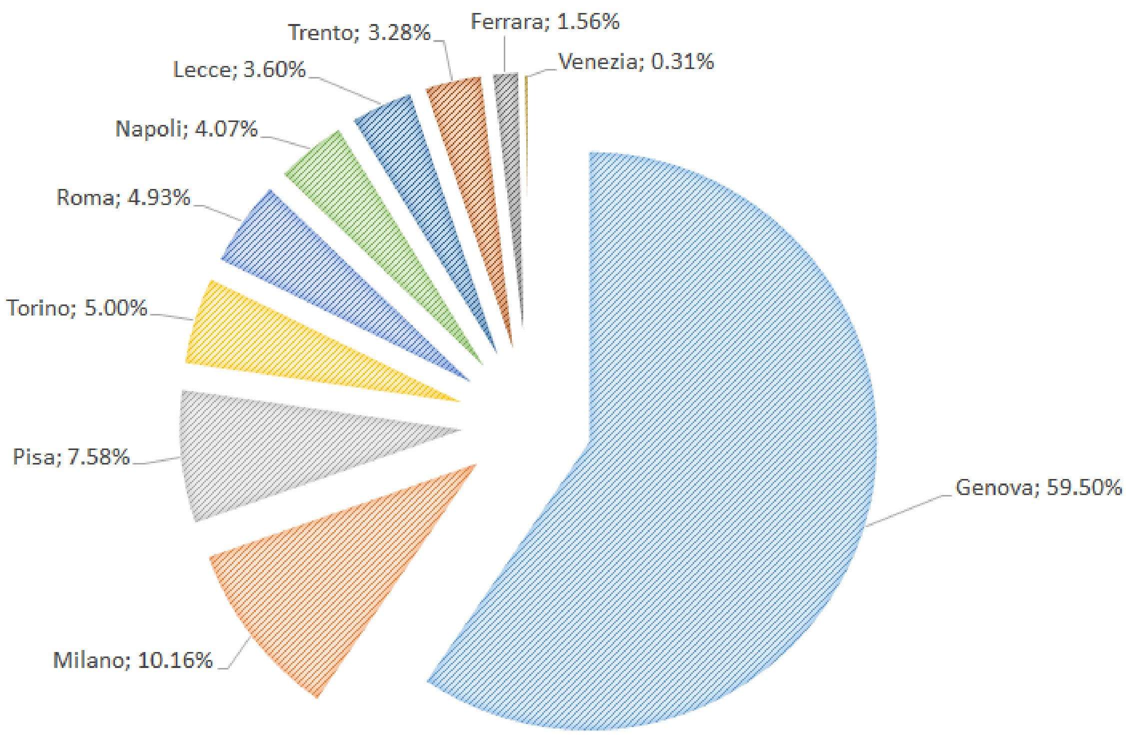
[26]Research take place in Genoa Central Research Laboratories, 11 IIT technological centres across Italy and 2 IIT outstations in US.

[27]IIT also has several joint technology laboratories with companies and public institutes.

[28]The list includes the Centre for Space Human Robotics in collaboration with Polytechnic University of Turin; the Centre for Nano Science and Technology in partnership with Polytechnic University in Milano; the Centre for Genomic Science in collaboration with European School of Molecular Medicine in Milano; the Centre for Neuroscience and Cognitive Systems in association with Trento University, at the headquarters of Rovereto; the Centre for Nanotechnology Innovation in collaboration with Normale University in Pisa; the Centre for Micro-Biorobotics in collaboration with Sant'Anna School of Pisa, in Pontedera; the Centre for Advanced Biomaterials for Health Care in partnership with Naples Federico II University; the Centre for Biomolecular Nanotechnologies in alliance with Lecce University; the Centre for Nano Science in collaboration with Sapienza University in Roma; the Centre for Translational Neurophysiology in collaboration with University of Ferrara; the Center for Cultural Heritage Technology in association with Ca' Foscari University in Venice; the LifeTech laboratories in formal collaborative arrangement between IIT and Harvard University; the Laboratory for Computational and Statistical Learning at the Massachusetts Institute of Technology, Boston.

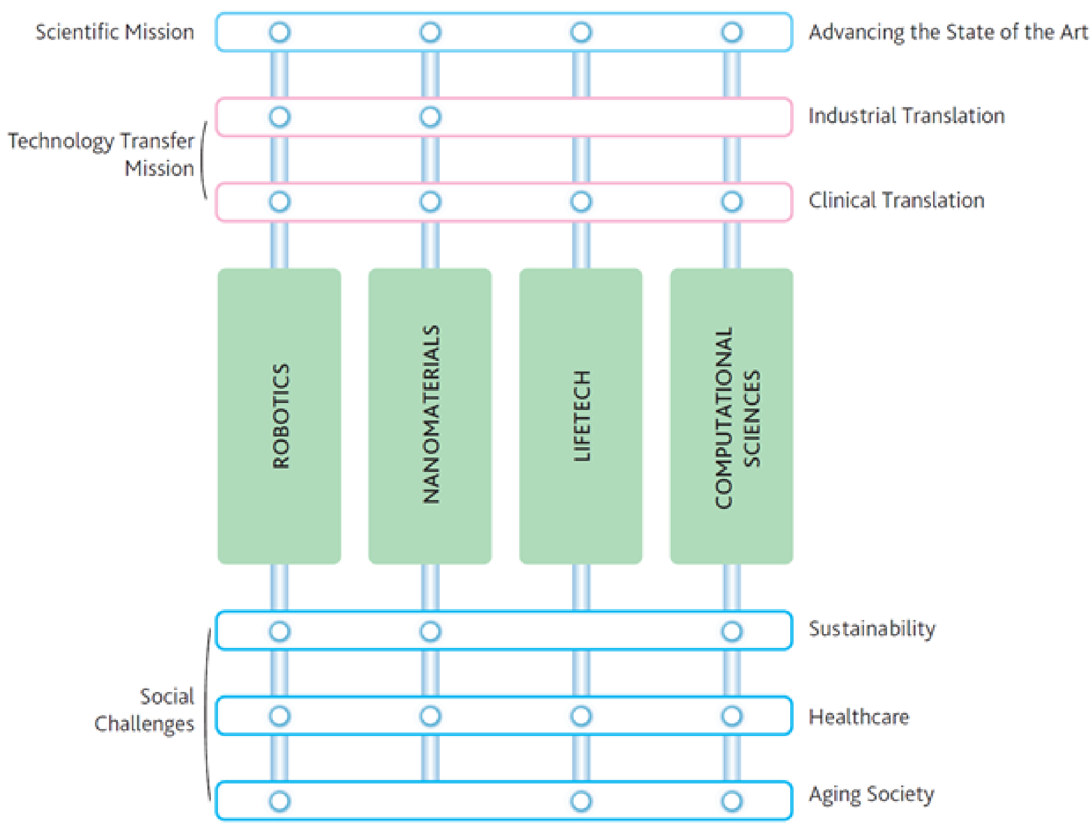
[29]Several Principal Investigators (PIs), working in complete autonomy and independence, develop innumerable strands and research programs in line with the Research Matrix created by the Institute, support each of them.

Figure 1: IIT Human Resources Endowment for Research.



Source: Author elaboration from `https://www.iit.it/people`.

Figure 2: IIT Research Organization.



Source: https://www.iit.it/it/ricerca/domini

Nanomaterials for Sustainability, Nanomaterials for Energy, Nanomaterials for Health and Exploratory Materials Science.

iii Lifetech supports progresses in advanced electrophysiological, computational, genetic, molecular imaging and perturbation tools for dissecting the microscopic neural processes underlying brain functions. This domain is divided in 3 Priorities: Neuroscience and Brain Technologies, RNA Technologies and Technologies for Healthcare.

iv Computational Sciences tends to develop massive simulations of physical systems, repeated numerous times to generate robust statistics, and data mining of vast datasets to identify unexpected patterns. This domain will focus on 4 Priorities: Development HPC Algorithms & Software, Computational Modelling, Machine Learning, Deep Learning & AI and Computer Vision[30].

In general, IIT research results should be a public asset benefiting society: from 2006 to 2018, the Institute activities have generated a flow of approximately 11720 publications in international scientific journals and about over 200 discoveries, which conduct to more than 600 patent applications.
Moreover, the Institute puts in place a set of services to transfer knowledge from research to the marketplace, especially regarding the changing needs of the high-tech market: IIT activities include protection of new inventions through intellectual property rights, without forgetting the strategic licensing of IIT technological and scientific knowledge.
Finally, IIT promotes and supports the origination of innovative start-up companies, the negotiation and definition of settlements with industries to realize R&D and competitive industrial research, the dissemination and training activities for the scientific community[31].

4 Identification Strategy and Data

4.1 The SCM Approach

The empirical researchers in assessing the effects of a treatment on some areas, like cities or regions, since early 1990s implement Diff-in-Diff methods. In particular, Diff-in-Diff methods are usually used when some areas experience a policy change, while others do not; it is worth noting that the choice of which of them receive the treatment is often not necessarily random and economic outcomes are not necessarily the same across areas in the absence of the treatment. Treatment and control groups are observed before and after the policy change. In this context, a fundamental issue

[30]Source: https://www.iit.it/research/domains.
[31]Source: https://www.iit.it/technology-transfer.

for causal inference is to develop a reliable estimate of what the outcomes would have been for treatment areas in the absence of the treatment[32].
The underlying assumption of the Diff-in-Diff approach is that the variation of the outcome over time for the control areas is indicative of what the change for the treatment group would have been in the absence of the treatment: however, those econometric approach requires that variables of interests have fixed relations over time. It is worth noting that the main challenge in assessing the impact of a treatment is thus the choice of a control group that is suitable to be a valid counterfactual for the treatment group. Indeed, if treated areas do not share similar economic characteristics and economic trends in the pre-treatment period with respect to control ones, a comparison between them is likely to produce biased estimates.

In this work, the Synthetic Control Method (SCM), a novel identification strategy, is implemented to estimate the effects of IIT research on the degree of innovativeness at local level. Such approach identifies the location of IIT central laboratories in Genoa in 2006 as a natural experiment. Indeed, the designation of Genoa as IIT headquarter has been affected by many factors, arguably exogenous, other than economic considerations[33]. After controlling for confounding factors with the SCM, the institution of IIT in 2006 thus represents a probably exogenous policy change that allows the identification of the causal effect of public funded research centres on the local economy.

Besides being a useful econometric approach when only one unit experiences the treatment and the other ones do not, the SCM, built by Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), relies on the Diff-in-Diff framework but is more sophisticated and overcomes the use of a single control unit or a simple average of control units. Implementing a weighted average of all controls, such method thus systematically offers comparisons that are more appealing with respect to Diff-in-Diff and other matching techniques. The SCM estimates the impact of a policy by matching the evolution of a variable of interest for a treated unit to the evolution of the same aggregate outcome for a synthetic control group. The counterfactual is indeed a weighted average of control units[34], whose observable features are comparable to those of the treated unit in absence of the treatment. Precisely, the SCM relies on a minimum distance approach, joined with the constraint that the resulting weights are non-negative and sum to one. This method guarantees the best quality of matching with the treated unit and synthetically replicate the outcome path that the treated unit would have experienced in the absence of the treatment. By considering the absence

[32]This involves estimating a (counterfactual) change over time for treatment regions if the policy change has not occurred.

[33]It is worth noting that the location of IIT in Genoa was the result of a political process. See `https://www.ilsecoloxix.it/economia/2013/01/18/news/i-baroni-della-ricerca-all-assalto-dell-iit-1.32294420`.

[34]The list of control units constitute what is called the "donor pool", the set of potential control units out of which the synthetic control unit is constructed.

of further shocks for treated areas in the post-treatment period, when the treated unit and the synthetic control one are matched on outcome and relevant economic covariates in the pre-treatment period the divergence in outcome trajectories can therefore be interpreted as the causal impact of the treatment.

Taking into account that the Genova metropolitan city is the treated unit, the donor pool is constitute by the set of control units that are not exposed to the treatment. Precisely, the donor pool refers to the 94 remaining provinces (Italian NUTS-3 regions). As the number of Italian provinces has been progressively changed in recent years[35], since many new ones are carved out of older ones and several others have been abolished, for the main analysis only the 1974 territorial setting is indeed considered[36]. The treated territory and the donor pool are presented in Figure 3.

To provide a formal description of the SCM method, Abadie et al. (2010) have been strictly followed. Suppose there exists a sample of $J+1$ territories, indexed by j, among which province $j=1$ is the treated one and units $j=2$ to $j=J+1$ represent the donor pool. In this framework, provinces in the sample are observed for time periods $t=1,2,\dots,T$; in particular, a balanced panel dataset is employed, which includes a 26 years pre-intervention period, T_0, as well as a 10 years post-intervention period T_1, with $T=T_0+T_1$.
Let's assume now that $W=(w_2,\dots,w_{J+1})^T$ is a $(J+1)$ vector containing weights, with $0 \leq w_j \leq 1$ for $j=2$ to $j=J$ and $\sum_{j=2}^{J} w_j = 1$. Define then X_1 as the $(k+1)$ vector of pre-intervention characteristics of the treated province and X_0 as a $(k*J)$ matrix containing the values of the same variables for the donor pool.
Finally, consider two potential outcomes: $Y_{j,t}^{I}$ is the outcome of interest in case province j in time t is exposed to the treatment, while $Y_{j,t}^{N}$ is the unobserved outcome for the same province at the same time if the latter would not be exposed to the treatment. The aim of the paper is to detect the post-treatment effect in the Genoa metropolitan city ($j=1$), defined as $\alpha_{1,t} = Y_{j,t}^{I} - Y_{j,t}^{N}$. Since the outcome for the same province at the same time in absence of treatment is unobserved, it is replicated through the Synthetic Control Method.

[35]Italian Provinces in 1974 are 95. In 1992-1995 reorganization, eight provinces have been created: Verbano-Cusio-Ossola, Biella, Lecco, Lodi, Rimini, Prato, Crotone, and Vibo Valentia, while Forlì was renamed as Forlì-Cesena. Four new provinces have been then created in Sardinia in 2001, with effect from 2005: Olbia-Tempio, Ogliastra, Medio Campidano and Carbonia-Iglesias. In the period between 2004 and 2009, three new provinces have been created: Monza and Brianza, Fermo, and Barletta-Andria-Trani, making a total of 110 provinces. In 2016, Sardinia, after a popular referendum, has decreed the abolition of the four provinces established in 2001. In 2017, the autonomous region of Friuli-Venezia Giulia, as part of an administrative reorganisation, has abolished the three provinces of Trieste, Gorizia and Pordenone. In the same Region, during 2018, the last province of Udine has also been abolished.

[36]As the selection of control units is fundamental, since using inappropriate ones may lead to wrong conclusions (Abadie et al., 2015), the provinces of Biella, Verbano-Cusio-Ossola, Lecco, Lodi, Rimini, Prato, Crotone, Vibo-Valentia, Olbia, Ogliastra, Medio-Campidano, Carbonia-Iglesias, Monza-Brianza, Fermo and Barletta-Andria-Trani are then excluded, since they suffer of territorial changes during the observation period.

Figure 3: Italian 1974 Provincial Administrative Divisions. Treated Territory and the Donor Pool.

As aforementioned, the "synthetic Genoa" is created as a weighted average of control provinces from the donor pool, $j = 2, 3, \ldots, J+1$, and characterized by the weighting vector W. Therefore, each variation of W leads to a specific characterization of the possible synthetic control.

With the aim to replicate the path of the outcome variable, and to be as similar as possible in terms of pre-treatment predictors to the treated area, the synthetic control estimator of the impact of the intervention is calculated as the difference between the outcome of the treated unit and its synthetic control unit, $\hat{\alpha}_{1,t} = Y_{1,t} - \sum_{j=2}^{J+1} w_j^* Y_{j,t}$. Precisely, the set of weights W^* is computed so that the "synthetic Genoa" best approximates the real Genoa, exposed to the intervention, with respect to the pre-intervention outcome predictors and a linear combination of pre-intervention outcomes. Optimal weights w_j^* are the ones that minimize $\sum_{m=1}^{k} \vartheta_m \left(X_{1,m} - X_{0,m} W\right)^2$, where ϑ_m reflects the relevance of matching variables in accordance to their outcome predictivity. Indeed, an optimal choice of such element is fundamental to minimize the Mean Squared Prediction Error (MSPE)[37] of the synthetic control estimator.

[37]The Mean Squared Prediction Error is the expected value of the squared difference between the fitted values implied by a predictive function $\widehat{g}$ and the values of a (unobservable) function g. It is an

For the main analysis, the best fitting matching specification[38] has been selected using the model suggested by Ferman et al. (2017): this test chooses the model that minimizes the pre-intervention Root Mean Square Prediction Error (RMSPE)[39]. As a result, the synthetic Genoa, with respect to each outcome of interest, is developed as to be the best reproduction of the most relevant characteristics of the real Genoa prior to the intervention. To this end, a complete set of observed covariates for each province, considered as relevant outcome predictors, and some linear combinations of pre-intervention outcomes, aimed to control for unobserved common factors whose effects vary over time, have been employed.

Subsequently, in order to alleviate the risk of biased estimates due to the presence of secondary IIT scientific sites in other provinces (see Section 3), the main analysis has been replicated with the removal of the latter from the donor pool. In particular, the provinces where the major four IIT centres (in terms of research human resources, see Figure 1) are located, namely Milan, Pisa, Turin and Rome, have been dropped from the sample, as sensitivity test.

Moreover, in order to explore the possibility that results may be solely driven by IIT own patent activity and, thus, to alleviate concerns of misleading estimates[40] (for a detailed description of variables of interest and results, see Sections 4.2 and 5), all patents that refer to the IIT have been identified and dropped from the sample. The SCM technique has then been replicated for all above mentioned analyses using such restricted sample.

Finally, several placebo and falsification tests are proposed. In particular, placebo permutation tests, as suggested by Abadie et al. (2010) are performed. Indeed, by implementing the SCM approach to each possible control unit allows to assess the distribution of the test statistic under the null hypothesis of no treatment effects[41]. Precisely, in every reiteration the assignment of IIT is redistributed among units of the control group. Subsequently, the effect accompanying each placebo is analysed with the aim to build a distribution of estimated impacts for the untreated regions. Then, the specific causal effect could be compared with those estimated for a randomly chosen area.
Sensitivity tests are then proposed.

---

inverse measure of the explanatory power of $\hat{g}$ and can be used in the process of cross-validation of an estimated model.

[38]The matching variables used in this work are a set of pre-intervention province-specific characteristics and pre-intervention outcome variables, which are described in detail in the data section.

[39]Similar to MSPE, the RMSPE is a measure of the quality of a predictor. Researchers can evaluate the goodness of fit by calculating RMSPE between the real and the synthetic region during the pre-treatment period. A poor fit might be caused by many factors, as i.e. weak predictors.

[40]This issue could potentially undermine the evidence of technological and knowledge spillovers arising from the IIT own research.

[41]For details, see Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015).

4.2 Data

In this work, several annual data (1980-2015), which derive from a variety of sources, are used for the 95 Italian provinces (NUTS-3 territories)[42]. The database mainly focuses on innovation measures, but also comprises features of the university system, indicators of industrial performance and economic indicators, observed for the 26 years pre-intervention period (1980-2005), as well as the 10 years post-intervention period (2006-2015).

The main measure of interest refers to the local patent activity. As it is widely known in literature, patents are fundamental instruments of appropriation of the innovative activity and, therefore, they appear as a suitable instrument for the detection of the results deriving from the innovative process. Consequently, patents are undoubtedly an important incentive for innovation but, at the same time, it should be noted that they are also public documents that disseminate information on innovation.

Patent counts are sometimes considered as an inadequate innovation proxy, since they are a measure of invention and not innovation and they do not cover all economic significant innovations (Smith, 2005); however, patent data are readily available, contain considerable details and are useful to develop time series. Moreover, the innovation literature agrees on the fact that technologies with greater impact on welfare and economic development are more likely to be patented (Pakes and Griliches, 1980). Therefore, patent counts appear to be valid indicators of local technological capacity, indicating cognitive resources.

In this work, the regional potential for innovation is measured following the guidelines of OECD Patent Statistics Manual (Zuniga et al., 2009); in fact, different criteria can be selected in order to measure patent activity and, on the basis of this choice, resulting indicators could have different values and different meanings.

Patent data used in this paper are collected from the Patstat database, which is distributed by the European Patent Office (EPO). EPO data specifically refer to patent applications directly filed under the European Patent Convention or to patent applications filed under the Patent Co-Operation Treaty and designating the EPO (Euro-PCT): therefore, the work mainly focuses on EPO patent applications because of their high significance.

The EPO Patstat database is actually a point of reference in the field of patent statistics, allowing sophisticated statistical analyses of bibliographical and legal status patent data. The database includes bibliographical and legal status patent data from several countries.

[42]The NUTS classification (Nomenclature of Territorial Units for Statistics) is a hierarchical system for dividing up the economic territory of the EU: NUTS3 refers to small regions.

In order to extract such information, several queries[43], in SQL language[44], have been created and run through the EPO web-based interface, obtaining annual bulk datasets, from 1980 to 2018, which contain detailed information on applications, applicants, inventors, technological class[45] and statistical classification of economic activity[46] for each filed patent.
Subsequently, to obtain a measure of the provincial potential for innovation, annual row datasets have been joined together and the geographic distribution of patent applications is finally assigned according to the inventor place of residence (NUTS-3 codes). Therefore, if a patent is characterized by more than one inventor, the patent application is distributed equally between all of them and consequently between their provinces (fractional counting), avoiding thus double counting (OECD, 2013).

The result is a balanced panel for 95 Italian provinces in the 1980-2015 period, with 3420 observations for the Patent Fractional Count[47]. Data are limited to 2015 because of the existence of an underestimation for application counts in the last two years of coverage of the database, due to delays in the publication of EPO data (eighteen/twenty-four months since application or priority date; see Zuniga et al. (2009) and Bronzini and Piselli (2016) for more details).

A full set of control variables are also included, with the aim to increase the comparability of the treatment and control groups. Precisely, the dataset contains several features of the university system, indicators of industrial performance and economic indicators collected from the "Urban Data Platform+" repository, a joint initiative of the Joint Research Centre (JRC) and the Directorate General for Regional and Urban Policy (DG REGIO) of the European Commission[48]. Indeed, this platform is a unique data source, aimed to provide information on status and trends of European regions; it is also a fundamental element of the Knowledge Centre for Territorial Policies.
The pre-intervention university features which are considered as controls refer to the number of active academic researchers, departments, universities and student enrolments. Moreover, several measures strictly related to the innovation process, as the number of local inventors and the number of European trade-marks (ETM) registered, are considered. Industrial and economic performances are approximated by Gross Domestic Product (GDP), Gross Value Added (GVA, for industrial sectors),

[43]Query available on request.

[44]Structured Query Language is a domain-specific language used in programming and designed for managing data held in a relational database management system (RDBMS), or for stream processing in a relational data stream management system (RDSMS). It is particularly useful in handling structured data, i.e. data incorporating relations among entities and variables.

[45]IPC framework.

[46]NACE2 framework.

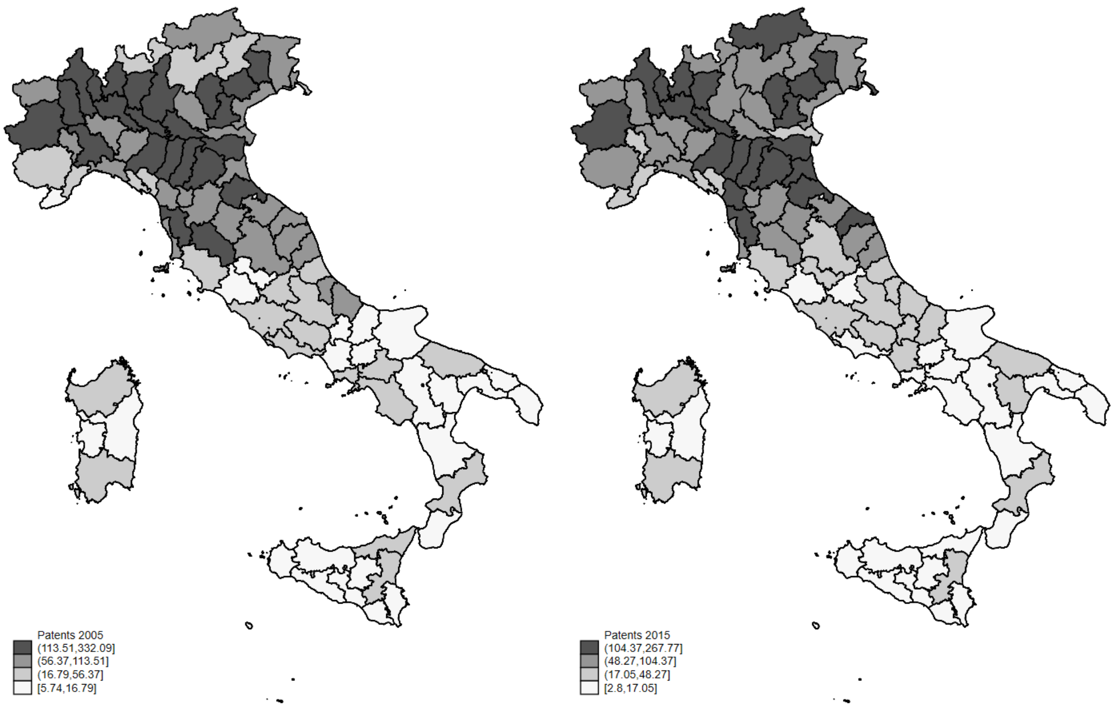
[47]A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero.

[48]Source: `https://urban.jrc.ec.europa.eu/rel2018/#/en/`.

Gross Fixed Capital Formation (GFCF)[49], the number of worked hours (for industrial sectors), the compensation of employees (for industrial sectors) and the number of employed people. Territorial-specific features, as population, surface and working age population are also considered.

Table 1 illustrates the mean values and the standard deviations of the outcome and pre-intervention matching variables computed for the overall sample and for the treated and control territories. Descriptive statistics are then reported for the overall time-period, for the specific implementation year 2006 and for the last observational year 2015. Finally, Figure 4 shows local patent activity for Italian NUTS-3 regions in the last year prior to the intervention, namely 2005, and 2015.

Figure 4: Italian Patent Activity. Patent Fractional Count (million inhabitants).



Notes: the left panel shows patents (fractional count - million inhabitants) in 2005 pre-intervention period. The panel on the right shows the same measure in 2015.

[49]GFCF (Eurostat, 2013), also known as Investments, consists of resident producers acquisitions, less disposals, of fixed assets during a given period plus certain additions to the value of non-produced assets.

Table 1: Summary Statistics for Main Variables.

	(A) Whole Sample					
	mean	sd	2006	sd	2015	sd
Patents (Fractional Count)	32.77	70.69	53.00	97.12	45.85	72.74
Inventors (Number)	56.74	135.90	94.72	190.70	88.19	151.20
European Trade Marks (Number)	31.23	100.60	58.94	123.90	92.65	186.00
GDP (millions)	13610.00	19140.00	16000.00	22340.00	14930.00	22170.00
GVA (millions)	12310.00	17430.00	14420.00	20120.00	13640.00	20140.00
GFCF (millions)	16930.00	12810.00	22010.00	16340.00	15740.00	12400.00
Worked Hours (Number)	2855.00	2002.00	3065.00	2173.00	2854.00	2068.00
Compensations (millions)	30310.00	27600.00	40710.00	32280.00	44530.00	37250.00
Employed People (Number)	231187.00	266968.00	249655.00	292373.00	244767.00	307762.00
Population (Number)	570284.00	588493.00	577414.00	596769.00	600171.00	642943.00
Surface (sq. KM)	2917.00	1555.00	2917.00	1555.00	2917.00	1555.00
Working Age Population (Number)	376573.00	392059.00	378534.00	397547.00	387225.00	420848.00
Univerity Enrolments (Number)	18135.00	34499.00	19136.00	35734.00	-	-
Researchers (Number)	189.10	378.70	-	-	-	-
Universities (Number)	1.08	1.56	1.04	1.59	-	-
University Departments (Number)	5.73	8.21	6.14	9.03	-	-

	(B) Treated Unit					
	mean	sd	2006	sd	2015	sd
Patents (Fractional Count)	53.06	30.07	90.39	0.00	81.33	0.00
Inventors (Number)	100.40	67.95	165.00	0.00	204.00	0.00
European Trade Marks (Number)	36.50	41.34	77.00	0.00	112.00	0.00
GDP (millions)	23600.00	2442.00	26410.00	0.00	24670.00	0.00
GVA (millions)	21690.00	1866.00	23790.00	0.00	22390.00	0.00
GFCF (millions)	7505.00	597.20	7830.00	0.00	6360.00	0.00
Worked Hours (Number)	1189.00	49.63	1193.00	0.00	1097.00	0.00
Compensations (millions)	12910.00	3699.00	16230.00	0.00	17950.00	0.00
Employed People (Number)	371892.00	14727.00	381142.00	0.00	387330.00	0.00
Population (Number)	926585.00	60407.00	876579.00	0.00	861253.00	0.00
Surface (sq. KM)	1806.00	0.00	1806.00	0.00	1806.00	0.00
Working Age Population (Number)	571323.00	46787.00	541225.00	0.00	520119.00	0.00
Univerity Enrolments (Number)	35505.00	2513.00	35110.00	0.00	-	-
Researchers (Number)	503.90	150.00	-	-	-	-
Universities (Number)	1.35	0.49	1.00	0.00	-	-
University Departments (Number)	11.82	0.39	12.00	0.00	-	-

	(C) Donor Pool					
	mean	sd	2006	sd	2015	sd
Patents (Fractional Count)	32.55	70.97	52.60	97.56	45.47	73.04
Inventors (Number)	56.28	136.40	93.97	191.50	86.96	151.60
European Trade Marks (Number)	31.18	101.00	58.74	124.60	92.45	186.90
GDP (millions)	13500.00	19210.00	15880.00	22430.00	14830.00	22270.00
GVA (millions)	12210.00	17500.00	14320.00	20210.00	13550.00	20230.00
GFCF (millions)	17030.00	12840.00	22160.00	16360.00	15840.00	12430.00
Worked Hours (Number)	2872.00	2005.00	3085.00	2176.00	2873.00	2071.00
Compensations (millions)	30490.00	27690.00	40970.00	32350.00	44820.00	37350.00
Employed People (Number)	229641.00	268017.00	248210.00	293662.00	243200.00	309098.00
Population (Number)	566494.00	590429.00	574231.00	599158.00	597394.00	645817.00
Surface (sq. KM)	2929.00	1559.00	2929.00	1559.00	2929.00	1559.00
Working Age Population (Number)	374423.00	393650.00	376804.00	399318.00	385811.00	422878.00
Univerity Enrolments (Number)	17949.00	34634.00	18966.00	35887.00	-	-
Researchers (Number)	185.80	379.00	-	-	-	-
Universities (Number)	1.08	1.57	1.04	1.60	-	-
University Departments (Number)	5.67	8.22	6.07	9.06	-	-

Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero.

5 Results

5.1 IIT and Innovation

In the analysis a synthetic control unit is developed, the so-called “synthetic Genoa”, which best replicates real values of predictor variables for the main outcome, i.e. patents in fractional count, in the 1980-2005 pre-intervention period.
For the Genoese territory, the impact of the creation of IIT on the above-mentioned innovation proxy is analysed as the difference in the number of patents between the treated province and its synthetic version in the period after the 2006 intervention.
As aforementioned, the best fitting matching specification has been chosen with the aim to minimize the pre-intervention RMSPE (see Section 4.1). In particular, the pre-treatment control variables that minimize the RMSPE refer to the overall mean and several lags of the outcome variable (ten lags, from 1996 until 2006), the number of inventors in the province (overall mean and 10 lags) and the overall mean of GDP, GVA, worked hours, number of university departments in the area.

Results are reported in Figure 5 and Table 2. Tables A1 and A2, in Appendix, provide instead the predictor balance and provincial weights.
In particular, Table A1 shows a fundamental characteristic of SCM approach. Unlike other matching estimators, the SCM forces scholars to prove the similarity among areas exposed to the treatment and their synthetic counterparts, that is, the weighted average of units in the donor pool. Consequently, the SCM prevents the estimation of “extreme counterfactuals”, that are those that fall far outside the convex hull of the data (King and Zeng, 2006).

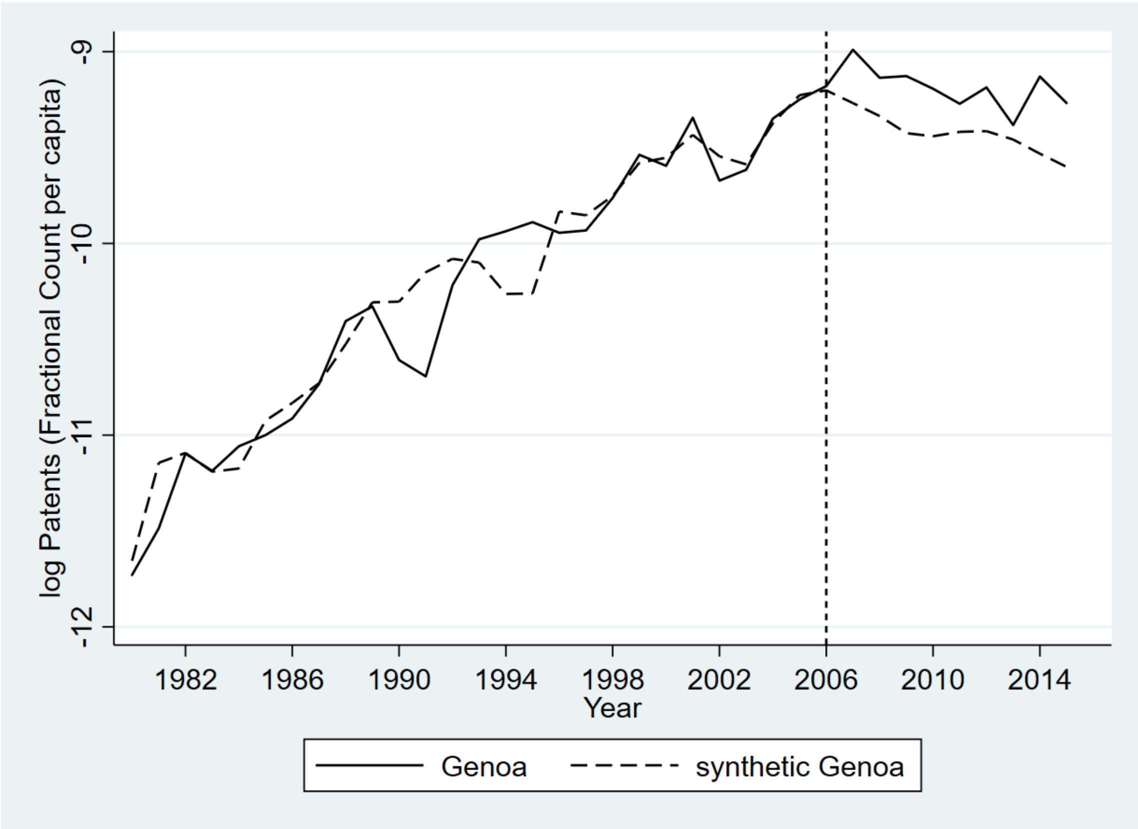
Table 2: Effect of the location of IIT in 2006: Effect Estimates.

Year	log Patents Treated (FC per capita)	log Patents Synthetic (FC per capita)	Patents - Treated (FC million inhabitants)	Patents - Synthetic (FC million inhabitants)	Absolute Effect	Relative Effect
2007	-8.98959350	-9.26806820	124.70	94.39	30.31	27.67%
2008	-9.13605310	-9.33557950	107.71	88.23	19.48	19.89%
2009	-9.12699600	-9.42512120	108.69	80.67	28.02	29.59%
2010	-9.19335080	-9.44103240	101.71	79.40	22.32	24.64%
2011	-9.27167320	-9.41797110	94.05	81.25	12.80	14.60%
2012	-9.18631170	-9.41441910	102.43	81.54	20.89	22.71%
2013	-9.38275430	-9.45811070	84.16	78.05	6.11	7.53%
2014	-9.12955860	-9.53175040	108.41	72.51	35.90	39.69%
2015	-9.26758770	-9.60039020	94.44	67.70	26.73	32.98%

Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. The absolute effect is the total difference between treated and synthetic control units, while the relative effect is the absolute effect divided by the mean between real outcome and synthetic control estimate.

Graphically (Fig. 5), the synthetic control closely matches the outcome of interest in the pre-intervention period, except for a small period (1990-1994) not in proximity to the intervention. Indeed, predictor balance in Table A1, in 1980-2006 period, shows reasonably comparable results for the Genoa metropolitan city and its synthetic version for pre-treatment control variables. Consequently, the treated unit and its synthetic equivalent are very likely to be similar in the period prior the implementation of IIT.

Figure 5: Effect of the location of IIT in 2006: Innovative Capacity.



Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero.

Finally, Table A2 contains critical weights on donor pool units and indicates that patent activity trend in Genoa prior to the implementation of IIT is best reproduced by a combination of 15 Italian provinces.

Figure 5 shows the effect of IIT research on patent activity as the absolute difference between per capita patent applications (fractional count) in Genoa and in its synthetic counterpart after the 2006 implementation. Immediately after 2006, the two lines remarkably diverge, thus suggesting a large positive effect of IIT own research on per capita patent applications.

Table 2 is aimed at depicting the magnitude of the impact of the IIT on Genoese innovation levels for the whole post-treatment period (2007-2015). The absolute effect is the total difference between treated and synthetic control units, while the relative effect is the absolute effect divided by the mean between real outcome and synthetic control estimate. The joint analysis of Figure 5 and Table 2 allows to highlight that, for the treated territory, there is strong evidence for an increase in the number of patent filed. After the implementation of IIT in 2006, Genoa patent activity experiences a faster growth with respect to its synthetic counterpart. This trend is constant until 2007, while from 2008 to the end of the sample period the trend is reversed, although it remains at higher levels than “synthetic Genoa”. The causal impact of the IIT implementation ranges from a minimum of 6.11 (7.53%) to a maximum of 35.90 (39.69%) more patents every million inhabitants. On average, the gaining of IIT impacts on Genoa innovation levels by about 22.5 patents for million inhabitants (24.37%) during the entire post-intervention period.

However, this result can be biased by two opposing issues, leading to an underestimation or overestimation respectively, namely the influence of 2008 financial crisis on patenting activity and the possibility that Genoa patenting trend may be specifically driven, after 2006, just from IIT patents.

Indeed, several studies illustrate how the effects of financial crisis on the economic process constraint patent applications[50]; it is also well known the bad circle where the lack of financial resources origins underinvestment, causing lower levels of innovation and, consequently, a slowdown in patenting activity. While the paradigm in which short-term shocks in patent activity may or may not have an effect on the long-term trend of innovation is generally unclear in the literature, economic conditions are likely to influence patenting activity. In particular, in a first phase 2008 financial crisis has affected Genoese local economy less harshly with respect to other similar areas, due to several structural factors that tend to moderate the sensitivity of Genoa (and Liguria) to fluctuations of economic cycle. Such factors refer to the higher level of tertiarisation, to the low level of openness to international trade, the relevant proportion of household income from pensions and public salaries. However, during the following years and with the extension of crisis effects from the financial side to the real economy, the local economy has also been largely hit, leading to significant contractions in consumption, investments and employment levels, a decrease in disposable income for households, a decrease in bank lending and a substantial growth in impaired loans (Bankitalia, 2016)[51].

For all these reasons, it seems even more fundamental to control enough for the economic structure of the area: nevertheless, it is necessary to taking into account the possibility of an underestimation in coefficient estimates. However, it is worth noting that, as aforementioned and clearly displayed in Figure 5, during this period, Genoa suffers from a slowdown in patent activity, like its synthetic counterpart, but still remains at a higher levels.

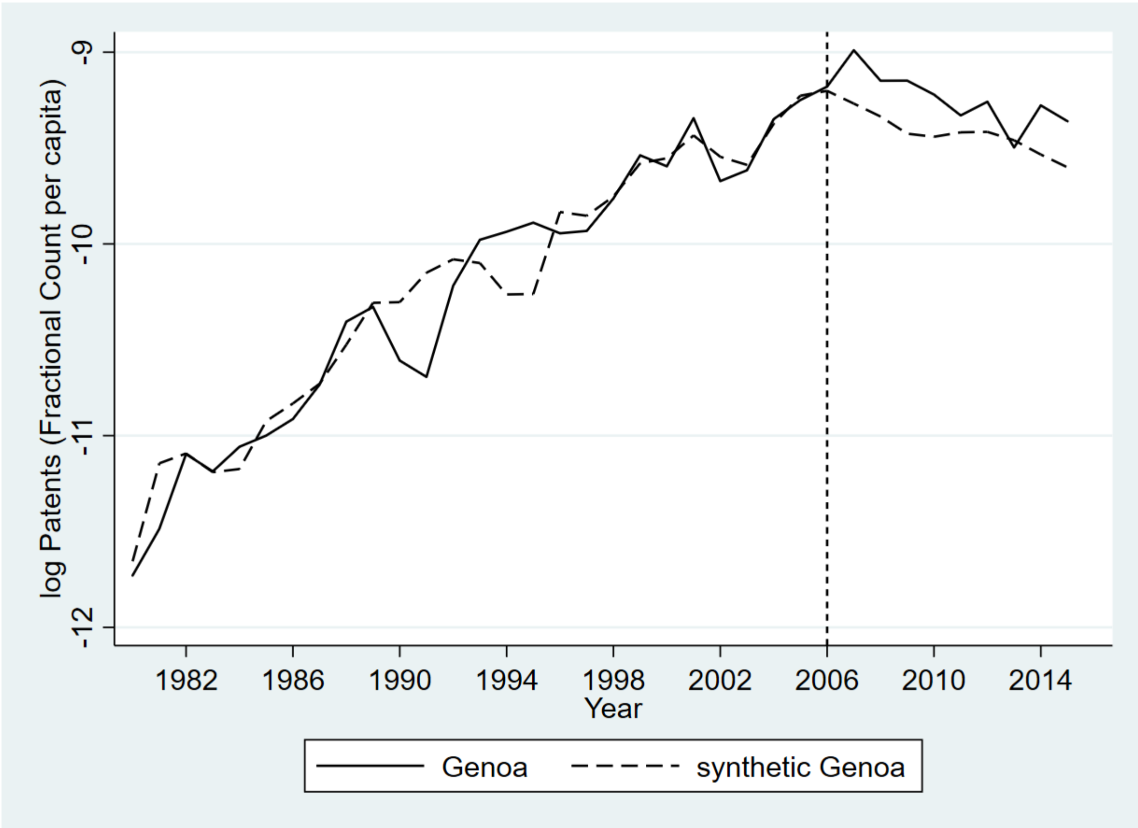
Another important issue is that one might argue that the results are potentially driven by IIT own patent activity, with actually little spillovers. Indeed, IIT implements an international model of public research that specifically focuses on the development of technologies for the market. IIT own research, both basic, "curiosity-driven" and applied, has thus led to file a large number of patents in different study areas defined by 4 research domains (See Section 3).

[50]For details, see Benoliel and Gishboliner (2014).

[51]`www.bancaditalia.it/chi-siamo/beni-immobili/edifici-storici/palazzo-de-gaetani/centenario_palazzo_de_gaetani_genova.pdf`.

This issue could potentially undermine the evidence of technological and knowledge spillovers arising from the IIT own research. Hence, all patents that refer to the IIT have been identified and dropped from the sample. The SCM approach has therefore been replicated on such restricted sample to refine the analysis and with the aim to scrutinize such potential bias.
Results are listed below in Figure 6 and Table 3 (Tables A3 and A4, in Appendix, usually show predictors balance and province weights).

Figure 6: Effect of the location of IIT in 2006: Innovative Capacity - Restricted Sample.



Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. Restricted sample: all patents that refer to the IIT have been identified and dropped from the analysis.

Table 3: Effect of the location of IIT in 2006: Effect Estimates - Restricted Sample.

Year	log Patents Treated (FC per capita)	log Patents Synthetic (FC per capita)	Patents - Treated (FC million inhabitants)	Patents - Synthetic (FC million inhabitants)	Absolute Effect	Relative Effect
2007	-8.98959350	-9.26806820	124.70	94.39	30.31	27.67%
2008	-9.14891720	-9.33557950	106.33	88.23	18.11	18.61%
2009	-9.14834120	-9.42512120	106.40	80.67	25.72	27.50%
2010	-9.22104840	-9.44103240	98.93	79.40	19.54	21.91%
2011	-9.32976530	-9.41801580	88.74	81.25	7.50	8.82%
2012	-9.25833320	-9.41551550	95.31	81.45	13.86	15.69%
2013	-9.49707130	-9.45951160	75.07	77.94	-2.87	-3.76%
2014	-9.27738090	-9.53373180	93.52	72.37	21.15	25.50%
2015	-9.36031060	-9.60071590	86.07	67.68	18.39	23.93%

Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. The absolute effect is the total difference between the treated and the synthetic control units, while the relative effect is the absolute effect divided by the mean between real outcome and synthetic control estimate. Restricted sample: all patents that refer to the IIT have been identified and dropped from the analysis.

Regarding the quality of fit in the pre-intervention period, nothing appears to have significantly changed. Indeed, predictor balance (Table A3, in Appendix) remains reasonably similar in the treated region and the synthetic versions for all pre-treatment control variables.

The results in Figure 6 highlight the persistence of a suggestive evidence for a positive causal effect of the location of IIT in 2006 for the Genoa area. Although estimates change in magnitude, it should be noted that the analysis on the restricted sample provides a picture that is roughly comparable with those in the main analysis. In the treated area, it is then possible to observe clear effects of IIT on patenting activity: these effects are quite similar in trends to those previously analysed. In particular, after the creation of IIT Genoa tends to have 16.86 more additional patents every millions inhabitants (18.43% higher, in average with respect to the synthetic Genoa). Specifically, effect estimates range from 7.50 (8.82%) to 30.31 (27.67%) more patents every million inhabitants. This finding seems to confirm the effectiveness of IIT public research in fostering local growth and development of innovation.

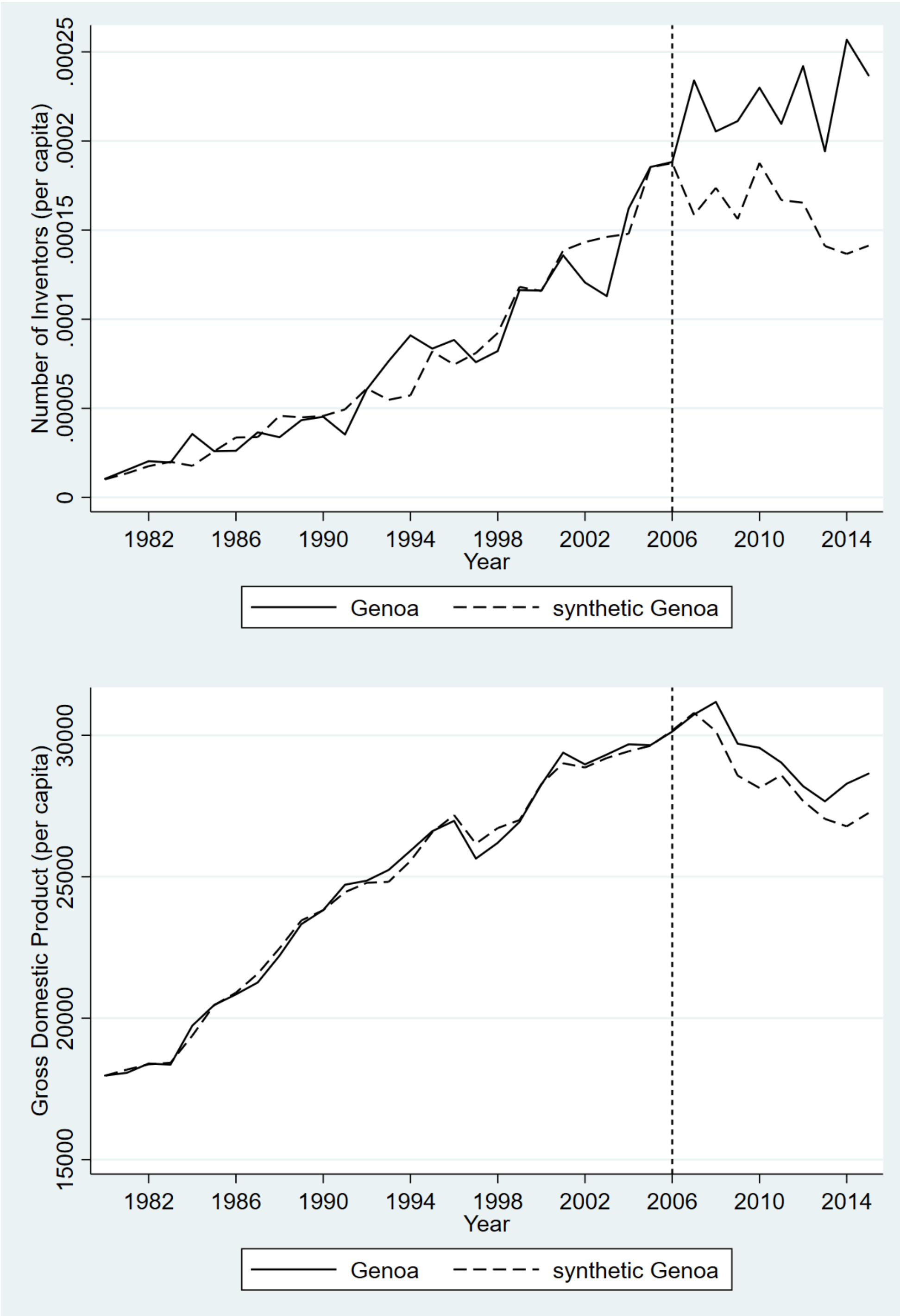
Overall results seem to be aligned to Moretti et al. (2019), suggesting that public funded R&D "crowds in" rather than "crowds out" firm innovation and patent activity. Moreover, the results in Cowan and Zinovyeva (2013) are also confirmed, suggesting that the location of new research centres improves local innovation activity, with effects largely caused by the high quality scientific research.

It is also worth noting that estimates could be seen as contradictory with respect to results in Bonander et al. (2016), since authors find no impact of new granting research universities on the Swedish regional economy. In particular, authors highlight that the intervention has caused a rise in awarded PhDs and professorships, whereas no effects are detected for the number of students, patent applications, and firm start-ups. A possible explanation could be that IIT is a public research centre, not a university. Specifically, despite universities are central players for knowledge accumulation, in terms of basic research and creation of highly skilled workers, IIT specifically conduct basic and applied scientific research in the public interest, for the purpose of pure technological development. Moreover, a IIT primary goal is to transfer own technology research results to the productive fabric. For these reasons, the increase in patent activity shown by Genoa in the post-implementation period, therefore, appears plausible, and the presence of knowledge spillovers is more likely than what would happen with a classic university.

5.2 The Impact on Research Competences and Economic Outcomes

In this section, following Bonander et al. (2016), the possibility that the creation of IIT in 2006 may have also affected the endowment of highly specialised human capital in research and primary economic measures, which might have an effect on the local economy during the follow-up period, is analysed. Specifically, the analysis is focused on the number of local inventors and per capita GDP.

Figure 7: Effect of the location of IIT in 2006: Research Competences and Economic Measures.



Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. In the top panel results for research competences are shown, while results for per capita GDP are shown in the bottom panel.

Figure 7 plots the trends in the number of inventors per capita and GDP per capita during the period of analysis. In the top panel results for research competences are

Table 4: Effect of the location of IIT in 2006: Research Competences and Economic Measures - Effect Estimates.

Year	Inventors - Treated (million inhabitants)	Inventors - Synthetic (million inhabitants)	Absolute Difference	Relative Difference
2007	234.05	158.56	75.49	38.46%
2008	205.37	173.70	31.67	16.71%
2009	211.18	156.23	54.95	29.91%
2010	229.95	187.67	42.28	20.25%
2011	209.63	166.85	42.78	22.73%
2012	242.07	165.38	76.69	37.64%
2013	194.22	141.07	53.15	31.70%
2014	256.87	136.65	120.22	61.10%
2015	236.86	141.27	95.59	50.56%
Year	GDP - Treated (per capita)	GDP Synthetic (per capita)	Absolute Difference	Relative Difference
2007	30 735.35	30 792.13	-56.78	-0.18%
2008	31 179.37	30 147.47	1031.90	3.37%
2009	29 703.49	28 576.60	1126.90	3.87%
2010	29 557.11	28 139.01	1418.10	4.92%
2011	29 035.90	28 593.49	442.41	1.54%
2012	28 200.95	27 677.46	523.49	1.87%
2013	27 666.15	27 045.79	620.36	2.27%
2014	28 290.43	26 781.69	1508.74	5.48%
2015	28 645.48	27 252.68	1392.80	4.98%

Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. The absolute effect is the total difference between the treated and the synthetic control units, while the relative effect is the absolute effect divided by the mean between real outcome and synthetic control estimate.

shown, while results for per capita GDP are depicted in the bottom panel.
It should then be noticed that, as explained in previous sections, the synthetic Genoa has been developed as a convex combination of provinces in the donor pool that closely imitate the treated unit in terms of pre-intervention values of outcome predictors. The results are displayed in Table 4.

First, the analysis looks to a fundamental intermediate entrepreneurial outcome referring to research competences which likely affects innovation and local technological development, namely the number of inventors that operate in a certain area.
As shown in Figure 7, the "synthetic Genoa" closely matches the real one in almost all pre-intervention period[52]: in the treated area, an increase in the number of inventors is clearly highlighted. Indeed, synthetic Genoa shows about 66 fewer inventors every million inhabitants then the real one. In particular, relative differences range from 16.71% to 61.10%. In addition, the number of inventors displays an increasing trend for Genoa, while this is not true for the synthetic corresponding area. These findings

[52] Province weights and predictor balance are available upon request.

seem to confirm the idea that the realization of IIT in 2006 has caused an increase in highly specialised human resources for research, in line with main predictions of the literature on knowledge spillovers and human capital development.

Subsequently, the analysis also focuses on a primary economic measure as the Gross Domestic Product, with the aim to explore the link between IIT presence and local economic growth. Indeed, GDP per capita is a significant measure of average living standards and economic well-being and is one of the core indicators of economic performance (OECD, 2013).

The synthetic control sensibly replicates the real Genoa in the whole pre-treatment period, as the bottom panel in Figure 7 and the predictor balance confirm[53]. Trusting in the goodness of fit among the synthetic control and Genoa during the period prior the intervention, estimates show evidence for a small but meaningful lagged impact of the IIT presence on local GDP per capita. Obviously, the implementation of IIT took some time before producing its effects on GDP, and the delay in the impact thus appears plausible.

Overall economic results seem to be aligned to Liu ones (2015), who highlights a 7% effect on manufacturing per worker from land-grant universities in the US in the 1860s, and to results in Valero and Van Reenen (2019), which show how increases in university presence are positively associated with faster subsequent economic growth. Moreover, public funded R&D may also affect growth by not only stimulating innovation in their area, but also increasing economic performances in neighbouring areas.

These results agree with the idea that research institutes are central actors in the knowledge-based economy, key drivers of innovation and major agents of economic growth.

5.3 Placebo and Falsification Tests

In this section, in order to assess the robustness of the results, several placebo and falsification tests are proposed. First, placebo permutation tests, as suggested by Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), are implemented; as second step, a conventional set of sensitivity checks, commonly employed in SCM applied literature, are then proposed.

The evaluation of SCM ability to replicate the evolution of a counterfactual Genoa without IIT requires a placebo study; indeed, applying the SCM to each possible control unit allows to assess the distribution of the test statistic under the null hypothesis of no treatment effects[54].

In particular, in every reiteration the assignment of IIT is redistributed among units of the control group, and such iterative method provides us with a distribution of

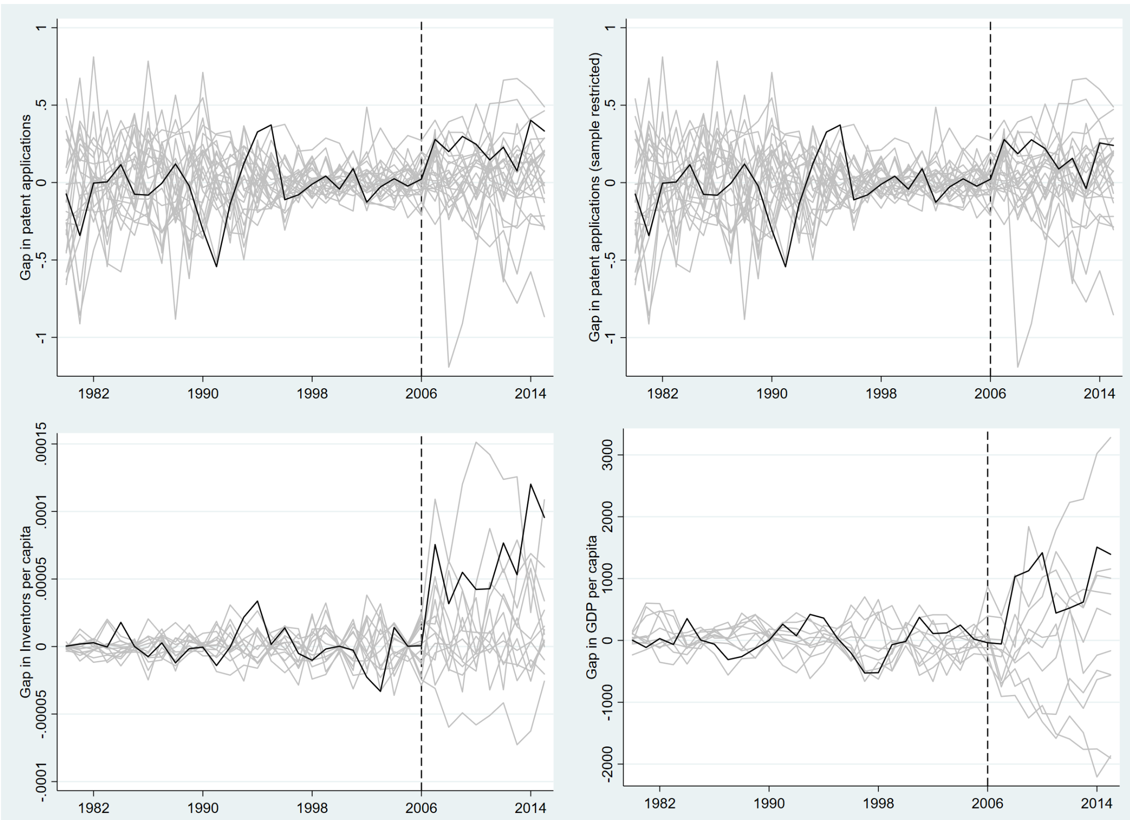
[53]Province weights and predictor balance are available upon request.

[54]For details, see Abadie et al. (2010).

estimated gaps for the provinces where no intervention took place. Therefore, the specific causal effect on the real Genoa is compared in magnitude with those estimated for an area randomly chosen.
Previous findings would obviously be undermined if a similar or even greater effect is achieved when the treated provinces are randomly chosen (where the intervention has not taken place)[55].

Results are presented in Figure 8. The grey lines show outcome differences related with each iteration of the placebo test. Specifically, the grey lines represent the gap for each measure of outcome between each control province in the donor pool and its own synthetic counterpart. The superimposed black line finally indicates the real output differences estimated for Genoa. Following Abadie and Gardeazabal (2003), placebo effects are not included in the pool for inference if the match quality of that control is significantly dissimilar to the match quality of the real treated unit.

Figure 8: Effect of the location of IIT in 2006: Placebo Plot Test.



Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. In restricted sample, all IIT patents are dropped.

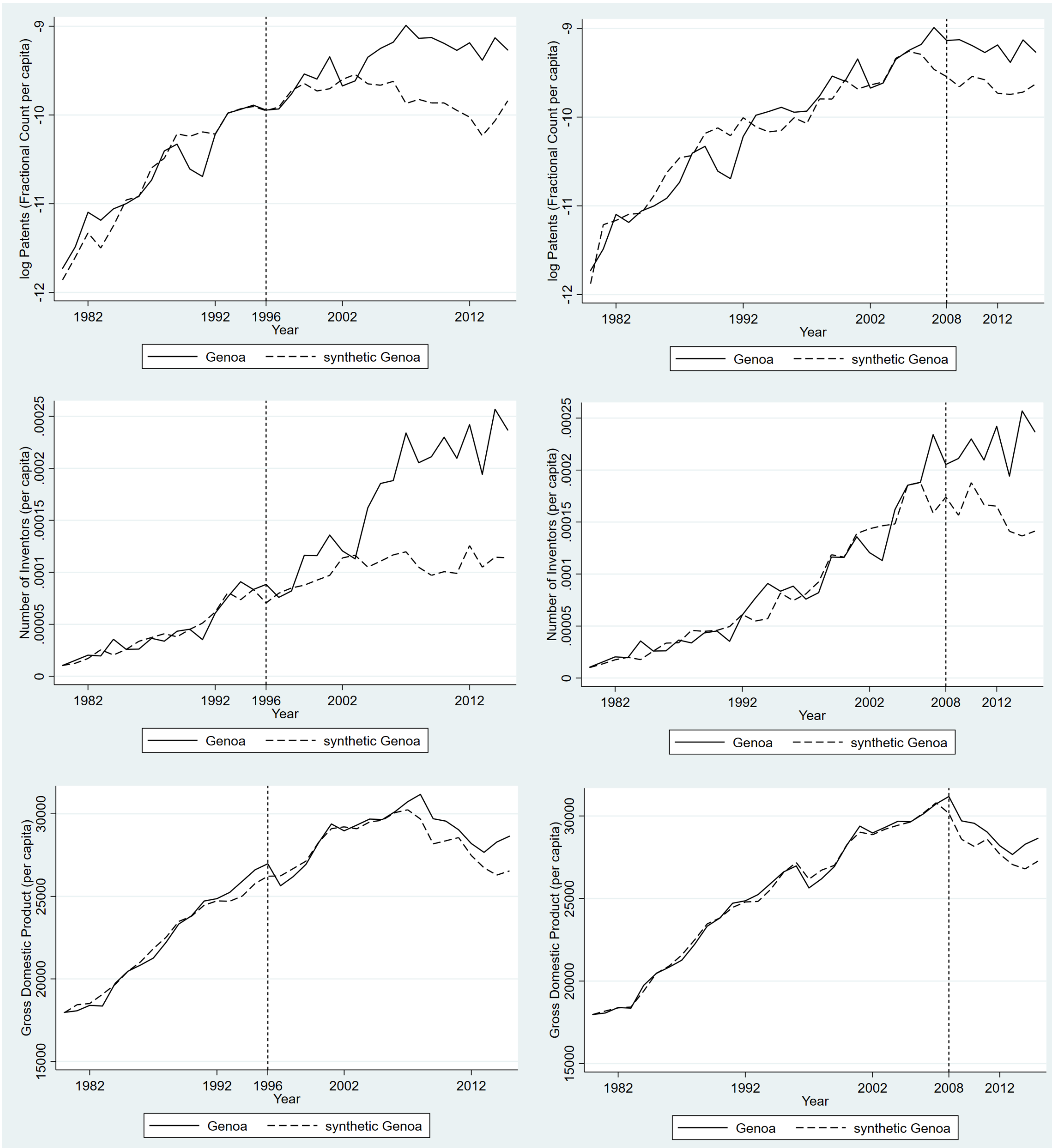
As Figure 8 clearly highlights, the estimated outcome difference for Genoa during the 2006–2015 post-implementation period is usually abnormally large with respect

[55]More generally, this inferential tool scrutinises whether or not the estimated impact of the IIT implementation is large with respect to the effects distribution for the provinces not exposed to the intervention. Under the null hypothesis of no intervention effect, the estimated impact of the intervention is then not expected to be abnormal with respect to the distribution of the placebo effects.

to the distribution of placebo gaps for almost all variables in the entire post-treatment period. These results likely confirm the robustness of paper findings.

With the aim of further investigate the robustness of the analysis, the main analysis is performed implementing a fake treatment year, in the spirit of common falsification tests: precisely, such falsification test assumes the realization of IIT in 1996 and 2008. Indeed, in 1996, Genoese economy should not be affected by the IIT, while in 2008 one would expect to find anticipatory effects. Thus, in this context, any discovered impact of IIT should be suspicious, casting some doubts on the effects found in previous analysis.

Figure 9: Effect of the location of IIT in $fake$ years: Falsification Test.



Notes: A one has been added to all patent and inventor count variables before taking the log to include observations with values of zero. In the left panel fake implementation year 1996 is presented. In the right panel fake implementation year 2008 is presented.

In particular, Figure 9 presents results for the fake implementation year 1996 in the left panel, while those for the fake implementation year 2008 are displayed in the right panel. Reassuringly, no direct effects of the IIT $fake$ implementation on any outcome variable are detected, thus corroborating the validity of the research design.

Finally, it is verified whether the main results are sensitive to the exclusion of some provinces. Furthermore, due to the presence of secondary IIT scientific sites in other provinces, to eliminate the risk of biased estimates the analysis has been replicated after removing the latter from the donor pool. Reassuringly, results[56] show trends that are in line to that found in the main analysis, staying virtually unaffected regardless of which and how many predictor variables (and provinces) are included.

6 Conclusions

In this work causal effects on regional innovative and economic performance of a prominent Italian place-based policy, the institution of Italian Institute of Technology (IIT), are analysed by means of a novel identification strategy, the Synthetic Control Method (SCM). The regional potential for innovation in 1980-2015 period is first analysed, using patents per capita as a proxy; subsequently, the effects are also analysed on several economic and entrepreneurial measures, as the number of local inventors and per capita GDP.

The analysis finds significant effects of IIT presence on the local patent activity (estimated average gap ranges from 18% to 24% depending on the sample).
Turning to intermediate entrepreneurial factors as research skills, there is strong evidence that the intervention has triggered a rise in highly specialised human resources for innovation and research: indeed, Genoa shows, in average, about 66 more inventors every million inhabitants then the synthetic one, with an average relative difference of 34%.
Finally, GDP per capita is also positively affected by the location of the IIT in 2006.
These results suggest that the rises in public research funding, development of human capital and formal competences have a significant impact on local economies over the 10-year post-intervention period.

Main findings of the paper seem to be aligned to Liu ones (2015), who highlights a 7% effect on manufacturing per worker from land-grant universities in the US in the 1860s. Improvements in local innovation activity, by means of high quality scientific research due to the location of new research centres, are confirmed also in Cowan and Zinovyeva (2013), Valero and Van Reenen (2019) and Moretti et al. (2019).
It is also worth noting that overall results could be seen as contradictory with respect to Bonander et al. (2016). In particular, authors find no impact of new granting research universities on the Swedish regional economy: precisely, they highlight that

[56]Results available by the author upon request.

the intervention caused a rise in awarded PhDs and professorships, whereas no effects are detected for the number of students, patent applications, and firm start-ups.
One possibility could be that, unlike Bonander et al. (2016), this work refers to a public research centre which conducts basic and applied scientific research for the purpose of pure technological development. It should be noticed that one of IIT primary goals is also to transfer the results of its own technology research to the productive fabric, by means of joint laboratories and industrial liasons. Moreover, IIT supports the innovation process also by transferring intellectual property to both established firms and start-ups, hence converting new technologies into concrete goods and services. Consequently, the presence of knowledge spillovers is more likely than what would happen with a classic university and, therefore, the increase in patent activity shown by Genoa appears plausible.
In addition, the effects on research competences and economic growth seem to be reasonable: as shown in Valero and Van Reenen (2019), increases in university presence are positively associated with faster subsequent economic growth. Indeed, authors highlight how the link among GDP per capita and public research might be not merely driven by the direct expenditures of research centres, their staff and students, but it is in fact mediated through an increased supply of highly specialised human capital and higher innovation.

Therefore, it is possible to conclude that public funded R&D is likely to deliver a stream of private and social returns, in terms of innovation and also economic growth, by attracting high quality researchers, PhD students and star scientists, those that larger benefit productivity and that uniquely have positive long-lasting effects on local knowledge spillovers, as suggested also in Waldinger (2016). In addition, the IIT technological transfer process to the productive fabric can certainly enhance knowledge accumulation and, therefore, knowledge spillovers.

On the other side, there are several points of reflection on which it is worth dwelling on in view of future research developments. Indeed, as Drucker (2016) suggests, benefits of research may not be location specific. In particular, author finds that spillovers are substantial up to 97 kilometres, reflecting considerable influence across space. For this reason, it is worth noting that innovation can be also displaced from one province to another.
It should be noticed that this analysis focuses on local effects of IIT own research and does not consider the possibility of spatial reorganization of economic activity. For this reason, a further research line should take into account this possibility, maybe including spatial lags in the spirit of Moretti and Wilson (2014).

References

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.

Abramovsky, L., Harrison, R. and Simpson, H. (2007), 'University research and the location of business r&d', *The Economic Journal* **117**(519), C114–C141.

Abramovsky, L. and Simpson, H. (2011), 'Geographic proximity and firm–university innovation linkages: evidence from great britain', *Journal of economic geography* **11**(6), 949–977.

Adams, J. D. (1990), 'Fundamental stocks of knowledge and productivity growth', *Journal of political economy* **98**(4), 673–702.

Aghion, P. and Howitt, P. (1990), A model of growth through creative destruction, Technical report, National Bureau of Economic Research.

Aghion, P. and Howitt, P. (2005), 'Growth with quality-improving innovations: an integrated framework', *Handbook of economic growth* **1**, 67–110.

Agrawal, A., Galasso, A. and Oettl, A. (2017), 'Roads and innovation', *Review of Economics and Statistics* **99**(3), 417–434.

Amin, A. and Wilkinson, F. (1999), 'Learning, proximity and industrial performance: an introduction', *Cambridge Journal of Economics* **23**(2), 121–125.

Andersson, R., Quigley, J. M. and Wilhelmson, M. (2004), 'University decentralization as regional policy: the swedish experiment', *Journal of Economic Geography* **4**(4), 371–388.

Andersson, R., Quigley, J. M. and Wilhelmsson, M. (2009), 'Urbanization, productivity, and innovation: Evidence from investment in higher education', *Journal of Urban Economics* **66**(1), 2–15.

Anselin, L., Varga, A. and Acs, Z. (1997), 'Local geographic spillovers between university research and high technology innovations', *Journal of urban economics* **42**(3), 422–448.

Arrow, K. J. (1972), Economic welfare and the allocation of resources for invention, *in* 'Readings in industrial economics', Springer, pp. 219–236.

Audretsch, D. B. and Feldman, M. P. (1996), 'R&d spillovers and the geography of innovation and production', *The American economic review* **86**(3), 630–640.

Azoulay, P., Graff Zivin, J. S., Li, D. and Sampat, B. N. (2019), 'Public r&d investments and private-sector patenting: evidence from nih funding rules', *The Review of economic studies* **86**(1), 117–152.

Baldwin, R. E. and Martin, P. (2004), Agglomeration and regional growth, *in* 'Handbook of regional and urban economics', Vol. 4, Elsevier, pp. 2671–2711.

Bankitalia (2016), *100 Genova. Centenario di Palazzo De Gaetani Sede di Genova della Banca d'Italia*, Banca d'Italia.

Baptista, R. (1998), 'Clusters, innovation and growth: a survey of the literature', *The dynamics of industrial clusters: international comparisons in computing and biotechnology* pp. 13–51.

Belderbos, R., Carree, M. and Lokshin, B. (2004), 'Cooperative r&d and firm performance', *Research policy* **33**(10), 1477–1492.

Bennett, R. J., Bratton, W. A. and Robson, P. J. A. (2000), 'Business advice: the influence of distance', *Regional Studies* **34**(9), 813–828.

Benoliel, D. and Gishboliner, M. (2014), 'The effect of economic crises on patenting activity across countries', *Chi.-Kent J. Intell. Prop.* **14**, 316.

Black, D. and Henderson, V. (1999), 'A theory of urban growth', *Journal of political economy* **107**(2), 252–284.

Bonander, C., Jakobsson, N., Podestà, F. and Svensson, M. (2016), 'Universities as engines for regional growth? using the synthetic control method to analyze the effects of research universities', *Regional Science and Urban Economics* **60**, 198–207.

Bronzini, R. and Piselli, P. (2016), 'The impact of r&d subsidies on firm innovation', *Research Policy* **45**(2), 442–457.

Carlino, G. A., Carr, J., Hunt, R. M., Smith, T. E. et al. (2012), The agglomeration of r & d labs, Technical report, Federal Reserve Bank of Philadelphia.

Cowan, R. and Zinovyeva, N. (2013), 'University effects on regional innovation', *Research Policy* **42**(3), 788–800.

Darby, M. R., Zucker, L. G. and Wang, A. (2004), 'Joint ventures, universities, and success in the advanced technology program', *Contemporary Economic Policy* **22**(2), 145–161.

Drucker, J. (2016), 'Reconsidering the regional economic development impacts of higher education institutions in the united states', *Regional Studies* **50**(7), 1185–1202.

Drucker, J. and Goldstein, H. (2007), 'Assessing the regional economic development impacts of universities: A review of current approaches', *International regional science review* **30**(1), 20–46.

Etzkowitz, H., Leydesdorff, L. A. et al. (1995), 'Universities and the global knowledge economy: A triple helix of university-industry-government relations'.

Eurostat, E. (2013), 'European system of accounts esa 2010', *Official Journal of the European Union* **174**, 56.

Feldman, M. P. (1999), 'The new economics of innovation, spillovers and agglomeration: Areview of empirical studies', *Economics of innovation and new technology* **8**(1-2), 5–25.

Ferman, B., Pinto, C. and Possebom, V. (2017), 'Cherry picking with synthetic controls'.

Fujita, M. and Thisse, J.-F. (1996), 'Economics of agglomeration', *Journal of the Japanese and international economies* **10**(4), 339–378.

Fujita, M. and Thisse, J.-F. (2002), 'Agglomeration and market interaction', *CEPR Discussion Paper* (3362).

Fujita, M. and Thisse, J.-F. (2003), 'Does geographical agglomeration foster economic growth? and who gains and loses from it?', *The Japanese Economic Review* **54**(2), 121–145.

Goldstein, H. A., Maier, G. and Luger, M. (1995), 'The university as an instrument for economic and business development: Us and european comparisons', *Emerging patterns of social demand and university reform: Through a glass darkly* pp. 105–133.

Goldstein, H. and Renault, C. (2004), 'Contributions of universities to regional economic development: A quasi-experimental approach', *Regional studies* **38**(7), 733–746.

Grossman, G. M. and Helpman, E. (1993), *Innovation and growth in the global economy*, MIT press.

Guillain, R. and Huriot, J.-M. (2001), 'The local dimension of information spillovers: a critical review of empirical evidence in the case of innovation*.(dialogue)', *Canadian Journal of Regional Science* **24**(2), 313.

Harhoff, D. (2000), Are there financing constraints for r&d and investment in german manufacturing firms?, *in* 'The economics and econometrics of innovation', Springer, pp. 399–434.

Hervas-Oliver, J.-L. and Albors-Garrigos, J. (2009), 'The role of the firm's internal and relational capabilities in clusters: when distance and embeddedness are not enough to explain innovation', *Journal of Economic Geography* **9**(2), 263–283.

Izushi, H. (2003), 'Impact of the length of relationships upon the use of research institutes by smes', *Research policy* **32**(5), 771–788.

Jaffe, A. B. (1989), 'Real effects of academic research', *The American economic review* pp. 957–970.

Jones, C. I. (2005), Growth and ideas, *in* 'Handbook of economic growth', Vol. 1, Elsevier, pp. 1063–1111.

Kantor, S. and Whalley, A. (2014), 'Knowledge spillovers from research universities: evidence from endowment value shocks', *Review of Economics and Statistics* **96**(1), 171–188.

King, G. and Zeng, L. (2006), 'The dangers of extreme counterfactuals', *Political Analysis* **14**(2), 131–159.

Kline, P. and Moretti, E. (2014), 'Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority', *The Quarterly Journal of Economics* **129**(1), 275–331.

Lane, P. J. and Lubatkin, M. (1998), 'Relative absorptive capacity and interorganizational learning', *Strategic management journal* **19**(5), 461–477.

Liu, S. (2015), 'Spillovers from universities: Evidence from the land-grant program', *Journal of Urban Economics* **87**, 25–41.

Lööf, H. and Broström, A. (2008), 'Does knowledge diffusion between university and industry increase innovativeness?', *The Journal of Technology Transfer* **33**(1), 73–90.

Love, J. H. and Roper, S. (2001), 'Outsourcing in the innovation process: locational and strategic determinants', *Papers in Regional Science* **80**(3), 317–336.

Mansfield, E. and Lee, J.-Y. (1996), 'The modern university: contributor to industrial innovation and recipient of industrial r&d support', *Research policy* **25**(7), 1047–1058.

Minerva, G. A. and Ottaviano, G. I. (2009), 'Endogenous growth theories: Agglomeration benefits and transportation costs', *Handbook of Regional Growth and Development Theories. Edward Elgar, Cheltenham* .

Mohnen, P., Mairesse, J. and Dagenais, M. (2006), 'Innovativity: A comparison across seven european countries', *Economics of Innovation and New Technology* **15**(4-5), 391–413.

Monjon, S. and Waelbroeck, P. (2003), 'Assessing spillovers from universities to firms: evidence from french firm-level data', *International Journal of Industrial Organization* **21**(9), 1255–1270.

Moretti, E., Steinwender, C. and Van Reenen, J. (2019), The intellectual spoils of war? defense r&d, productivity and international spillovers, Technical report, National Bureau of Economic Research.

Moretti, E. and Wilson, D. J. (2014), 'State incentives for innovation, star scientists and jobs: Evidence from biotech', *Journal of Urban Economics* **79**, 20–38.

Nieto, M. J. and Santamaría, L. (2007), 'The importance of diverse collaborative networks for the novelty of product innovation', *Technovation* **27**(6-7), 367–377.

OECD (2013), *OECD Compendium of Productivity Indicators 2013*, OECD Publishing.

Pakes, A. and Griliches, Z. (1980), 'Patents and r&d at the firm level: A first report', *Economics letters* **5**(4), 377–381.

Perkmann, M. and Walsh, K. (2007), 'University–industry relationships and open innovation: Towards a research agenda', *International journal of management reviews* **9.4**, 259–280.

Polanyi, M. (2009), *The tacit dimension*, University of Chicago press.

Power, D. and Malmberg, A. (2008), 'The contribution of universities to innovation and economic development: in what sense a regional problem?', *Cambridge journal of regions, economy and society* **1**(2), 233–245.

Rallet, A. and Torre, A. (1999), 'Is geographical proximity necessary in the innovation networks in the era of global economy?', *GeoJournal* **49**(4), 373–380.

Romer, P. M. (1990), 'Endogenous technological change', *Journal of political Economy* **98**(5, Part 2), S71–S102.

Schubert, T. and Kroll, H. (2016), 'Universities' effects on regional gdp and unemployment: The case of g ermany', *Papers in Regional Science* **95**(3), 467–489.

Smith, K. (2005), 'Measuring innovation'.

Toole, A. A. (2012), 'The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry', *Research Policy* **41**(1), 1–12.

Torre, A. and Gilly, J. (2000), 'Debates and surveys: On the analytical dimension of proximity dynamics" regional studies; apr 2000; 34, 2', *Academic Research Library pg* **169**.

Valero, A. and Van Reenen, J. (2019), 'The economic impact of universities: Evidence from across the globe', *Economics of Education Review* **68**, 53–67.

Waldinger, F. (2016), 'Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge', *Review of Economics and Statistics* **98**(5), 811–831.

Woodward, D., Figueiredo, O. and Guimaraes, P. (2006), 'Beyond the silicon valley: University r&d and high-technology location', *Journal of Urban Economics* **60**(1), 15–32.

Zuniga, P., Guellec, D., Dernis, H., Khan, M., Okazaki, T. and Webb, C. (2009), 'Oecd patent statistics manual', *Francia: OECD Publications* .

Appendix

Table A1: Effect of the location of IIT in 2006: Predictor Balance.

Predictor Balance	Treated	Synthetic
log Patents (per capita 1980-2006 mean)	-10.238651	-10.210187
log Patents (per capita 1996)	-9.9445887	-9.8336593
log Patents (per capita 1997)	-9.9322319	-9.8531022
log Patents (per capita 1998)	-9.7626934	-9.7520115
log Patents (per capita 1999)	-9.5378551	-9.5794194
log Patents (per capita 2000)	-9.5951233	-9.5531842
log Patents (per capita 2001)	-9.3440561	-9.4338958
log Patents (per capita 2002)	-9.6724463	-9.5455846
log Patents (per capita 2003)	-9.6158409	-9.5876474
log Patents (per capita 2004)	-9.3491268	-9.3737893
log Patents (per capita 2005)	-9.2493191	-9.2267484
log Patents (per capita 2006)	-9.1796532	-9.2017387
log Inventors (per capita 1980-2006 mean)	-9.7599277	-9.7600755
log Inventors (per capita 1996)	-9.3341398	-9.4217893
log Inventors (per capita 1997)	-9.486268	-9.5345111
log Inventors (per capita 1998)	-9.4077177	-9.3215714
log Inventors (per capita 1999)	-9.0593538	-9.034706
log Inventors (per capita 2000)	-9.0616417	-9.0058939
log Inventors (per capita 2001)	-8.9040632	-8.8539103
log Inventors (per capita 2002)	-9.0228987	-8.9674115
log Inventors (per capita 2003)	-9.0885267	-8.9841819
log Inventors (per capita 2004)	-8.7276134	-8.7726373
log Inventors (per capita 2005)	-8.5928745	-8.6203939
log Inventors (per capita 2006)	-8.577837	-8.5529802
log GDP (per capita 1980-2006 mean)	10.095055	10.08823
log GVA (per capita 1980-2006 mean)	8.10848	8.2304448
Worked Hours (per capita 1980-2006 mean)	.00127079	.00839626
University Departments (per capita 1980-2006 mean)	.00001314	.0000135

Table A2: Effect of the location of IIT in 2006: Province Weights.

Province	Weight	Province	Weight	Province	Weight
Torino	0	Ragusa	0	Lecce	0
Vercelli	.14	Siracusa	0	Foggia	.042
Novara	0	Sassari	0	Bari	0
Cuneo	0	Nuoro	0	Potenza	.034
Asti	0	Cagliari	0	Matera	0
Alessandria	0	Oristano	0	Cosenza	0
Aosta	.032	Bolzano	0	Catanzaro	0
Imperia	0	Trento	0	Reggio di Calabria	0
Savona	0	Verona	0	Trapani	0
La Spezia	0	Vicenza	0	Palermo	.005
Varese	0	Belluno	0	Messina	0
Como	.1	Treviso	0	Agrigento	0
Sondrio	0	Venezia	0	Caltanissetta	0
Bergamo	0	Padova	.104	Enna	0
Brescia	0	Rovigo	0	Catania	0
Pavia	0	Pordenone	0	Pisa	0
Cremona	0	Udine	0	Arezzo	0
Mantova	0	Gorizia	0	Siena	.048
Milano	.022	Trieste	0	Grosseto	0
L'Aquila	0	Piacenza	0	Perugia	0
Teramo	0	Parma	0	Terni	0
Pescara	.072	Reggio nell'Emilia	0	Pesaro e Urbino	0
Chieti	0	Modena	.068	Ancona	0
Isernia	0	Bologna	0	Macerata	0
Campobasso	0	Ferrara	.217	Ascoli Piceno	0
Caserta	.038	Ravenna	0	Viterbo	0
Benevento	0	Forlì-Cesena	0	Rieti	0
Napoli	.032	Massa-Carrara	0	Roma	0
Avellino	.037	Lucca	0	Latina	0
Salerno	0	Pistoia	0	Frosinone	0
Taranto	0	Firenze	0		
Brindisi	.007	Livorno	0		

Table A3: Effect of the location of IIT in 2006: Predictor Balance - Restricted Sample.

Predictor Balance	Treated	Synthetic
log Patents (per capita 1980-2006 mean)	-10.238651	-10.210187
log Patents (per capita 1996)	-9.9445887	-9.8336593
log Patents (per capita 1997)	-9.9322319	-9.8531022
log Patents (per capita 1998)	-9.7626934	-9.7520115
log Patents (per capita 1999)	-9.5378551	-9.5794194
log Patents (per capita 2000)	-9.5951233	-9.5531842
log Patents (per capita 2001)	-9.3440561	-9.4338958
log Patents (per capita 2002)	-9.6724463	-9.5455846
log Patents (per capita 2003)	-9.6158409	-9.5876474
log Patents (per capita 2004)	-9.3491268	-9.3737893
log Patents (per capita 2005)	-9.2493191	-9.2267484
log Patents (per capita 2006)	-9.1796532	-9.2017387
log Inventors (per capita 1980-2006 mean)	-9.7599277	-9.7600755
log Inventors (per capita 1996)	-9.3341398	-9.4217893
log Inventors (per capita 1997)	-9.486268	-9.5345111
log Inventors (per capita 1998)	-9.4077177	-9.3215714
log Inventors (per capita 1999)	-9.0593538	-9.034706
log Inventors (per capita 2000)	-9.0616417	-9.0058939
log Inventors (per capita 2001)	-8.9040632	-8.8539103
log Inventors (per capita 2002)	-9.0228987	-8.9674115
log Inventors (per capita 2003)	-9.0885267	-8.9841819
log Inventors (per capita 2004)	-8.7276134	-8.7726373
log Inventors (per capita 2005)	-8.5928745	-8.6203939
log Inventors (per capita 2006)	-8.577837	-8.5529802
log GDP (per capita 1980-2006 mean)	10.095055	10.08823
log GVA (per capita 1980-2006 mean)	8.10848	8.2304448
Worked Hours (per capita 1980-2006 mean)	0.00127079	0.00839626
University Departments (per capita 1980-2006 mean)	0.00001314	0.00001357

Table A4: Effect of the location of IIT in 2006: Province Weights - Restricted Sample.

Province	Weight	Province	Weight	Province	Weight
Torino	0	Ragusa	0	Lecce	0
Vercelli	.14	Siracusa	0	Foggia	.042
Novara	0	Sassari	0	Bari	0
Cuneo	0	Nuoro	0	Potenza	.034
Asti	0	Cagliari	0	Matera	0
Alessandria	0	Oristano	0	Cosenza	0
Aosta	.032	Bolzano	0	Catanzaro	0
Imperia	0	Trento	0	Reggio di Calabria	0
Savona	0	Verona	0	Trapani	0
La Spezia	0	Vicenza	0	Palermo	.005
Varese	0	Belluno	0	Messina	0
Como	.1	Treviso	0	Agrigento	0
Sondrio	0	Venezia	0	Caltanissetta	0
Bergamo	0	Padova	.104	Enna	0
Brescia	0	Rovigo	0	Catania	0
Pavia	0	Pordenone	0	Pisa	0
Cremona	0	Udine	0	Arezzo	0
Mantova	0	Gorizia	0	Siena	.048
Milano	.022	Trieste	0	Grosseto	0
L'Aquila	0	Piacenza	0	Perugia	0
Teramo	0	Parma	0	Terni	0
Pescara	.072	Reggio nell'Emilia	0	Pesaro e Urbino	0
Chieti	0	Modena	.068	Ancona	0
Isernia	0	Bologna	0	Macerata	0
Campobasso	0	Ferrara	.217	Ascoli Piceno	0
Caserta	.038	Ravenna	0	Viterbo	0
Benevento	0	Forlì-Cesena	0	Rieti	0
Napoli	.032	Massa-Carrara	0	Roma	0
Avellino	.037	Lucca	0	Latina	0
Salerno	0	Pistoia	0	Frosinone	0
Taranto	0	Firenze	0		
Brindisi	.007	Livorno	0		

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Prof. Anna Bottasso and Prof. Maurizio Conti for the continuous support of my Ph.D study and related researches, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisors, I would like to thank my Ph.D thesis external referees, Prof. Fabio Montobbio (Università Cattolica del Sacro Cuore) and Prof. Davide Vannoni (Università di Torino) for their helpful comments and their invaluable advices and suggestions.

I thank the rest of my thesis collaborators: Dr. Paolo Marocco and Dr. Marta Santagata, for the sleepless nights and days we were working together before deadlines, for their insightful comments and encouragements which incented me to widen my research from various perspectives. Without their invaluable support, this thesis would not have been possible.

I thank all my colleagues in the Ph.D office for the stimulating discussions and for all the fun we have had in the last three years.

Last but not the least, I would like to thank my family: my parents and my friends for supporting me spiritually throughout writing this thesis and my my life in general.