Essays in Urban Economics and International Trade

Filippo Boeri

December 2019

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy, London, December 2019

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 41,000 words.

Statement of co-authored work

I confirm that Chapter 2 was jointly co-authored with Marco Di Cataldo and Elisabetta Pietrostefani and I contributed 33% of this work.

Filippo Boeri
December 20 2019

Abstract

This thesis consists of four independent chapters on urban economics and international trade. The first chapter analyses the effect of an improvement of information and communication technologies (ICT) on student performance. By exploiting a vast infrastructural program implemented in Italy, I study how the availability of high-internet speed affects student educational outcomes. On average, I find a positive effect on student performance in numeracy subjects. However, the policy outcome significantly depends on individual socio-economic backgrounds, with disadvantaged students mostly excluded from the benefits. As a result, the program increased the dispersion of student outcomes.

The second chapter (co-authored) analyses a security policy aimed at tackling criminal organisations. The policy, consisting in the confiscation of mafia assets, is meant to eradicate the pervasive presence of the mafia organisations on the Italian territory and harm their business model and earnings. We investigate to what extent this policy is able to regenerate deprived areas by assessing its impact on the value of buildings located in the nearby of confiscated/re-allocated properties. Results show a positive effect of confiscation on housing prices, with the largest gains concentrated in the most deprived areas and where mafia organisations are more deeply rooted.

Third and fourth chapter investigate between and within-industry differences in the spatial distribution of economic activities. The third chapter consists in a cross-country analysis of the economic geography of manufacturing and service plants. I compute continuous agglomeration and co-agglomeration indices using two comprehensive datasets, that cover the whole population of British and French establishments. The estimates are used to analyse in a common framework between- and with-industry variation in agglomeration patterns. The empirical strategy makes it possible to test some of the main predictions suggested by the literature. The fourth chapter, focussing on France only, investigates to what extent exogenous trade shocks are able to reshape the spatial distribution of manufacturing activities. Results unveil a positive effect of import penetration from emerging economies on firm spatial agglomeration.

Acknowledgements

This PhD has been at the same time a challenging and enriching 'solo journey' and a joint experience with extraordinary people.

I owe my utmost gratitude to my supervisors, Olmo Silva and Steve Gibbons. Olmo has guided me through every stage of this process, providing me with constant feedbacks, encouragements and funny jokes. He helped me develop my research interests, identify my strengths and ultimately understand what kind of researcher I wanted to be. I could not think of anything more I could have expected from a supervisor. Steve made me question my approach to research and provided me with countless feedbacks. His comments shaped every chapter of this thesis and the whole way I now address research questions. I know I will not stop asking myself 'what would Steve think about this?' in any future project. I also thank Carlo Altomonte, that helped me discover my research interests and made this whole journey possible.

My work has benefited from discussions with other faculty members at the LSE. I thank Felipe Carrozzi, Gabriel Ahlfeldt, Giulia Faggio, Henry Overman, Riccardo Crescenzi, Simona Iammarino and all other participants of the LSE Economic Geography seminar for the helpful comments. I am also grateful to Chiara Criscuolo and Giuseppe Berlingieri for the opportunity to spend some months at the PBD/STI Division of the OECD. That experience has had a significant impact on my PhD.

I am also grateful to my old teachers, Emanuela that taught me to read and convinced me I could write, and Cini, that told me when I was just a child that one day I would have found a profession to be sweetly obsessed with.

I am deeply grateful to all my PhD colleagues, those that finished before me, and those who started after me, and to all Postdoc fellows in the Department. Thank you for the daily chats, the Friday beers and the intellectual stimuli. I am particularly grateful to Daria, David, Davide, Marco, Patrizia, and to the 'Econ Geo latinos', Alexandra, Eduardo and Elisabetta, without whom I would never have survived this journey. I also thank my coauthors for the intense and fruitful collaboration on my second PhD chapter.

I also thank my family, Antonio, Maddalena and Stefano, that never fully understood what I was doing, but nevertheless supported me during the whole experience. I am grateful to my friends and in particular to Fede, Giulia, Lorenzo, Marco, Martino, Nico, Roger and Vittorio. Without their support, I would not be here.

I gratefully acknowledge the generous financial support from the UK Economic and Social Research Council.

I dedicate this work to Francesca, the reason why I fall asleep every night thinking I don't need anything else and I wake up every morning knowing that I want more.

Table of Contents

List of Tables

List of Figures

Introduction

This thesis consists of four independent chapters on urban economics and international trade.

The first section of the thesis focuses on the evaluation of spatially targeted policies. Place-based policies, such as placement of infrastructure and targeted economic development programmes, are intended to stimulate local economic growth and tackle spatial inequalities within countries. Other policies, although not explicitly targeting underperforming areas, focus on socio-economic phenomena characterised by significant spatial concentration. In both cases, the assessment of the pursued welfare effects and potential local externalities need to be performed within a common spatial framework. In particular, researchers are expected to isolate the causal effect from spatially correlated confounding factors, to carefully identify the spatial extension of the outcomes and to produce results that are robust to changes in shape and extension of the spatial aggregation. In this section, I analyse a large infrastructural program aimed at reducing the technological divide in Southern Italy as well as a security policy tackling the power exerted by organised crime on deprived territories.

Chapter 1 investigates the impact of high-speed broadband on educational achievements, analysing a large infrastructural program implemented by the Italian Government. The causal effect of internet speed is identified exploiting three key features of the program implementation. First, I focus on designated 'market failure areas', where the Government is the only investor, bypassing potentially confounding effects that could stem from complementarity/substitutability between public and private intervention. Secondly, I exploit the exogeneity of the timing of roll-out of the broadband infrastructure, dictated by technical considerations and other quasi-random factors. Finally, I measure the potential treatment rather than the actual broadband connection to overcome biases related to the - possibly endogenous - decision to connect. To carry out my analysis, I combine a rich panel dataset covering the whole population of Italian students with broadband data measured at a very fine spatial scale. My results exhibit an average positive effect of high-speed internet broadband on educational achievements. However, this masks substantial heterogeneity: while students with better-off backgrounds gain from internet speed, the opposite is found for disadvantaged students.

Chapter 2, co-authored with Marco Di Cataldo and Elisabetta Pietrostefani, focuses on a large-scale policy consisting in the confiscation of properties belonging to individuals convicted for mafia-related crimes, and their re-allocation to a new use. We test the hypothesis that the policy contributes to the regeneration of urban spaces by assessing its impact on the value of buildings in the vicinity of confiscated/re-allocated properties. By exploiting detailed information on the exact location and timing of over 13,000 confiscated and re-allocated properties and over 80,000 geo-localised house sale points, we perform difference-in-differences analyses, investigating the externalities of the policy in the 55 major Italian cities. The results unveil a positive and significant effect of re-allocations of confiscated real estate assets on house prices, that declines with distance from the re-allocation site. The impact is larger in cities with stronger mafia presence and in more deprived neighbourhoods. This suggests that the policy contributes to add value to the territory where it is applied, and favours processes of urban revitalisation. These findings have important implications for the development of deprived urban areas characterised by a strong presence of criminal organisations.

The second part of the thesis investigates the heterogeneity in between- and within-industry spatial agglomeration patterns. Over the past decades, new technologies and rapid decline of trade costs have fostered the globalisation of good and capital markets. Driven by the rising political divide between cities and the countryside, a large body of empirical and theoretical work has unveiled significant effects of trade liberalisations on economic inequality. These often take the form of dramatic increases in spatial disparities. Nevertheless, the use of discrete, non-overlapping and often economically meaningless spatial units prevents researchers from correctly identifying the way exogenous shocks change the internal geography of jobs and economic activities. In this section I study the spatial distribution of economic activities and the way foreign shocks change the location of firms within a country.

Chapter 3 focuses on the spatial distribution patterns of manufacturing and service firms in France and the United Kingdom. By exploiting a large micro-dataset covering the whole population of French and British establishments over the period 2008-2015, I analyse the specific co-agglomeration patterns that characterise various types of plant within each industry. Industry-group location patterns are analysed interpreting agglomeration/coagglomeration estimates within a common framework. By comparing the main within-industry patterns identified in the two macro-sectors, I am able to test some of the most relevant dynamics predicted in the recent literature on agglomeration economies.

Chapter 4, focussing on French manufacturing establishments only, analyses the impact of import competition on the location decisions of plants between and within industries. Results suggest that import penetration slows down the decline of spatial agglomeration that characterises the manufacturing sector in several developed countries. However, behind the aggregate effects, the analysis unveils a significant heterogeneity across different plant types. In general, import competition is found to affect manufacturing industries, reshaping both their aggregate economic geography and the nature of manufacturing activities performed in different regions.

Chapter 1: High-speed Broadband and Educational Achievements

1.1 Introduction

Nowadays, home computers have become an essential tool of modern education in the developed countries. According to the last OECD[1] report on ICT and education, access to home computers is now nearly universal in most OECD countries. However, data still show significant disparities in the access to and quality of home computing. The digital divide is often related to various levels of access to high-speed internet connections and for this reason many countries have invested relevant amounts of public funding in order to upgrade information and communication technologies (ICT) with the aim of increasing the available internet connection speeds. In spite of this, the actual impact of ICT on student performance, as on several other social outcomes, is still debated (Machin, Sandra & Silva, 2007; Barrera-Osorio & Linden, 2009; Checchi, Rettore, & Girardi, 2015; Cristia, Ibarrarán, Cueto, & Santiago, 2017; Faber et al., 2015). Recently, Italy, one of the lowest performers in the Pisa tests among OECD countries, started to gradually reduce this historical gap with other European countries in the diffusion of ICT devices. In 2012, 50% of Italian students browsed the internet for their schoolwork at least once a week at home, 5% more than in 2009 but still below the OECD average of 54%. As such, Italian students are still characterised by a significant digital divide, especially in the poorer Southern regions. In 2012, almost 80% declared that they used the internet outside of school at least once a week in order to find practical information, but this number drops to only 66% for disadvantaged students. Despite efforts to provide schools with ICT devices and laboratories, actual usage is still constrained by the general inadequacy of broadband infrastructures.

This paper sets out to estimate the causal effect of upgrades to the available internet speed on educational achievements. The identification strategy relies on the specific features of a policy implemented by the Italian Government in 2014; the *'National Ultra-Broadband Plan'* (NUBP) is a national plan aimed at ensuring 100% coverage at 30 Mbps and 85% coverage at 100 Mbps by 2020. This study exploits two specific characteristics of the plan. First, the NUBP targets 'market failure areas', where no private operator is interested in investing. This ensures there are no concerns about possible overlaps with the interests of private actors. Second, in order to cover

[1] Organisation for Economic Co-operation and Development

the whole territory in a relatively short period, whilst minimising public spending, the NUBP was implemented progressively in adjacent territories (see Fig. 1.1). As a result, the timing of the implementation can hardly be related to variables associated with educational outcomes. In particular, the main estimation assumption is that, within each province (Nuts-3 level regions), the rollout can be considered exogenous. In addition, I focus on the broadband supply measure, rather than measuring its actual consumption. This allows me to bypass the endogeneity that may characterise the internet usage measures frequently used in the literature. Results suggest a small but significant positive effect of 30 Mbps broadband on educational achievement. When heterogeneity in social background is accounted for, this impact remains positive and significant for the best performers in the previous grade, while it becomes non-significant or negative for disadvantaged students. Overall, students' backgrounds seem to play a relevant role in the heterogeneous policy outcome.

This study contributes to the literature on social outcomes of the internet in two ways. First, it focusses on the impact of high-speed internet broadband on educational outcomes where previous studies have primarily focused on different outcomes, such as employment and productivity (Akerman et al., 2015), electoral outcomes (Falck et al., 2014; Campante et al., 2017), marriage rates (Bellou, 2015) and housing prices (Alhfeldt et al., 2017). A few studies have focused on the relationship between home computer technology and student achievement. For example, Beltran et al. (2006) use a large panel dataset of US students to explore the causal relationship between computer ownership and various educational attainment levels, with a specific focus on high school graduation. Fairlie and Robinson (2013) conducted a large field study involving almost 8,000 students enrolled in grades 6-10 in 15 different middle and high schools in the United States. Fiorini (2010) use data from the Longitudinal Study of Australian Children (LSAC) to analyse the causal relationship between computer usage and children's cognitive and non-cognitive skills. They are able to test individual skills at two moments in time, instrumenting computer usage with their parents' previous computer ownership. Results generally exhibit computer usage to have a positive effect on cognitive skills, whereas the results are mixed for non-cognitive skills. Studies directly relating internet access to educational outcomes are generally descriptive. To my knowledge, the only exception is Faber et al. (2015). Exploiting randomly placed jumps in the available ICT across neighbouring residences, the authors investigate the causal effect of a sensible increase in available internet speeds on educational outcomes in the United Kingdom. The second relevant contribution of this work is that it exploits the features of a large infrastructure policy, bypassing the endogeneity issues that usually affect similar studies.

The chapter is organised as follows. Section 1.2 provides a general background of the Italian School systems and describes the main features of the NUBP. Section 1.3 presents a simple theoretical framework. Section 1.4 describes the different data sources used and the procedure implemented to define school catchment areas. In Section 1.5 describes the empirical strategy. In section 1.6, 1.7 and 1.8, I present the results and discuss the main policy implications.

1.2 Institutional Background

1.2.1 The National Ultra-Broadband Plan

In 2014, the Italian Government set up the '*National Ultra-Broadband Plan*' (Piano Nazionale Banda Ultra-Larga - NUBP), a massive program aiming to ensure 100% coverage at 30 Mbps and 85% coverage at 100 Mbps by 2020. The plan was developed in accordance with the 'European Broadband Guidelines', which set out how the EU State aid rules apply to public funding for the rollout of broadband networks. The national territory was classified into three different areas according to existing or expected future broadband infrastructure deployment:

1. *White areas*: areas where no provider of broadband services is currently operating and where no such provider is expected to enter the market in the coming three years.
2. *Grey areas*: areas where one (infrastructure-based) provider is already active, but another network is unlikely to be developed in the next three years.
3. *Black areas*: areas where there are or there will be in the next three years at least two basic broadband networks of different operators.

The NUBP is based on four main pillars. First, the State guarantees administrative simplification and a reduction in burdens for all of the target regions. Second, private investments are encouraged in black and grey areas through the creation of tax exemption tools for infrastructure operations. Grey areas also benefit from various measures to facilitate the access to financial resources, the establishment of a guarantee fund and access to credit at subsidised rates. Finally, in white areas (often incorrectly defined as "market failure areas[2]") the Public Sector intervenes directly to realise the infrastructures.

After an initial delay, due to a number of legal disputes, in 2015 the program started to produce positive effects (see Fig. 1.3). The 30 Mbps broadband penetration rate increased between 2015 and 2017, from 25% to almost 50%. Thus far, the implementation speed has widely varied across different regions. Basilicata, Puglia and Sicily, three poor Southern regions, experienced, respectively, a growth from 9% to 58%, from 15% to 80% and from 21% to 67%. The specific characteristics of the broadband infrastructures and the way the policy was implemented are such that these results are generally driven by an increase in high-speed broadband penetration in individual municipalities from 0% to 60%-95%. These significant results were made possible by the availability of EU structural funds (the European Regional Development Fund, European Agricultural Fund for rural development and the Development and Cohesion Fund), which complemented public funds, and the general inadequacy of existing infrastructures, offering an

[2] These areas are defined According to Article 107(3)(c)TFEU (Treaty of Functioning of the European Union). State aid is allowed in order "*facilitate the development of certain economic activities or of certain economic areas, where such aid does not adversely affect trading conditions to an extent contrary to the common interest*". In the Communication from the Commission on EU State Aid Modernisation (SAM). Brussels, 8.5.2012. COM(2012), it is stated that "*State aid policy should focus on facilitating well-designed aid targeted at market failures and objectives of common European interest. State aid measures can, under certain conditions, correct market failures, thereby improving the efficient functioning of markets and enhancing competitiveness. Further, where markets provide efficient outcomes but these are deemed unsatisfactory from a cohesion policy point of view, State aid may be used to obtain a more desirable, equitable market outcome. In particular, a well-targeted State intervention in the broadband field can contribute to reducing the 'digital divide' between areas or regions where affordable and competitive broadband services are on offer and areas where such services are not.*"

average broadband speed below 2 Mbps. On the other hand, richer Northern regions, where many large cities and industrial clusters already had access to 100 Mbps broadband networks, did not experience any relevant changes in access to 30 Mbps broadband.
The policy presents two peculiar features that makes it extremely interesting from a research point of view. First, the focus on white areas allows me to overcome the classical endogenous problems that characterise infrastructural policies. As a matter of fact, in the absence of private intervention, local demand is unlikely to influence the timing of the implementation. Second, extremely detailed spatial data (95,000 local areas) makes it possible to track the progressive implementation of the policy across the treated regions over time.

1.2.2 The Italian School System

The Italian education system consists of 4 stages: nursery school (children between 3 and 6 years of age), primary education (children between 6 and 11), first grade (lower) secondary school (between 11 and 14 years of age) and second grade (upper) secondary school (from 14 to 19 years of age). Once these stages have been successfully completed, students can access the higher education offered by universities, institutes for Higher Education in Art and Music as well as Higher Technical Institutes. Education is compulsory for ten years, between the ages of 6 and 16. As a result, all students are expected to gain at least a 'Licenza media' (lower secondary school diploma). Over the last decade, the country experienced a reasonable decrease in the number of high school early dropouts. In 2014, only 1.6% (mostly first-generation foreigners) of the population in the 16-19 year-old cohorts did not hold a lower secondary school diploma.
Primary and lower secondary school together form the first cycle of education, lasting 8 years. According to the new ministerial guidelines, the general aim of lower-secondary education is '*the harmonious and comprehensive development of the individual, according to the principles of the Italian Constitution and European cultural tradition, to be achieved through the promotion of knowledge, respect for individual diversity and the active involvement of students and their families*' (Framework for Key Competences for Lifelong Learning set up by the European Parliament and the Council of the European Union through the Recommendation). The subjects taught in this stage are: Italian, English, a second foreign language, mathematics, science, technology, geography, history, music, art, sports science and Catholic religious education (optional). Schools are expected to provide 30 hours of teaching per week (990 hours per year), allocated according to a common timetable (see Table A1. 1 in the Appendix).

School Councils can offer to some or all classes an 'Extended timetable' (from 36 to 40 hours per week). In this case, the mandatory education goals remain the same, but students are expected to allocate less time to at-home study. At the end of the three-year program, students need to pass a uniform national examination in order to obtain a diploma and to access the following stage. The examination consists of a national written test set by INVALSI (also used by the Institute as a national assessment for grade 8) and four written tests set by a mixed (internal-external) committee. The subjects covered in the tests are Italian language, mathematics, science, ICT and two foreign languages.

In contrast to the following two cycles (upper secondary school and tertiary education), the primary and lower-secondary schools are characterised by a very low, if not entirely absent, degree of autonomy (Ichino & Tabellini, 2014). First, individual schools have almost no autonomy in the design of the education programs. The national Ministry designs the course contents, defines the number of hours to allocate to each subject and authorises a limited number of textbooks for each field. Single institutions are only allowed to use a limited budget to set up laboratories and extra-program activities, such as optional courses. Teachers are allowed to choose among a certain number of authorised textbooks for each subject and can design their classes based on the national program, but need to report each semester to the Ministry. Teaching methods and contents must be consistent with each school's educational offer plan, which in turn must be consistent with the educational goals established at the national level. Second, the homogeneity in the service is also guaranteed by a rigid financial system. Between 97% and 100% of the school budget depends on transfers from the Central Government. Every three years, the Ministry allocates resources based on specific criteria (i.e. number of students, number of disabled students, specific needs, etc.). The uniformity of the system is also guaranteed by the human resource management, which is primarily conducted at the national level. Teachers apply to province-level lists and are assigned a ranking. Vacancies in each school are covered on the basis of teachers' preferences and rankings, with little to no involvement of the school directors. Salaries and career development are defined by national agreements.

Another relevant feature of the lower secondary education system is the limited competition among schools. Classes can have between 15 (down to 10 in remote areas) and 26 students. Classes that exceed this number are split using the limited extra budget allocated by the Central Government. When the number of students applying exceeds the available places, schools are allowed to select entrants according to various criteria, but are expected to take into account distance as the main criterion. All of these characteristics together enforce a high degree of homogeneity among different schools. Even though national data still show disparities in teaching standards among different regions, mostly based on the quality of buildings and teachers' self-selection, there is strong evidence of a generally uniform service quality within provinces (Nuts 3 areas).

1.3 Theoretical framework

This section presents a basic model to guide the empirical analysis. I study the effect of changes in access to high-speed internet on learning outcomes using a simple production function. Following Faber et al. (2015) I distinguish two main mechanisms:

1. ICT improvement can change the productivity associated with a given amount of time spent studying (MOOC effect).
2. High-speed internet can affect the supply of time spent studying relative to leisure activities (online-gaming effect).

This model simply extends the one proposed by Faber et al. (2015) by taking into account any potential background-biased effect of ICT access on students' productivity. In this framework, students with a better family background are expected to maximise productivity gains, whereas disadvantaged students may be less likely to offset the negative 'gaming effect'. This assumption is built on the rich - although mostly qualitative - literature investigating the factors related to the relationship between student performance and their access to ICT[3]. Consistently, student *i*'s knowledge production function is given by:

$$H_{i2} = A_{i2} L_{i2}^{\alpha} H_{i1} \tag{1}$$

where H_{it} is the educational achievement at the entrance to (t=1) and exit from (t=2) a given school cycle, A_{it} is an individual learning productivity shifter, L_i is the time spent studying and $\alpha > 0$ is the elasticity of learning outcomes with respect to time spent studying. I assume both productivity and individual labour supply to be functions of student-specific characteristics (λ_i^A and λ_i^L), and school characteristics (μ_s^A and μ_s^L).
Broadband access to high-speed internet (S) affects both the student productivity shifter and the time spent studying:

$$A_{i2} = S^{\delta(B)} \lambda_i^A \mu_S^A e^{\varepsilon^A} \tag{2}$$

$$L_{i2} = S^{\eta} \lambda_i^L \mu_S^L e^{\varepsilon^L} \tag{3}$$

Following a basic labour supply equation, η captures a relative price effect, since S affects the relative attractiveness of studying compared to leisure activities, online or offline. On the other hand, δ, depending on students' background (B), captures the effect on individual productivity. Substituting (2), (3) into (1) and taking logs, I obtain the following estimation equation:

$$H_{i2} = [\alpha\eta + \delta(B)] lnS + \lambda_i + \mu_s + H_{i1} + \varepsilon \tag{4}$$

Where $\mu_s = \ln\mu_s^A + \alpha\ln\mu_s^L$, $\quad \lambda_i = \ln\lambda_i^A + \alpha\ln\lambda_i^L$, $\quad \varepsilon = \varepsilon^A + \alpha\varepsilon^L + \gamma\varepsilon^P$

[3] Several studies suggest that the perception of their own ICT abilities (Aesaert and Van Braak, 2007), their actual competencies (Tounder et al., 2010; Aesaert et al., 2015) as well as their profile of ICT use (Scherer et al., 2010) might be influenced by family background.

The main hypothesis tested is that the interplay between η and δ will determine different effects depending on B, a 'learning multiplier' linked to household characteristics (parents' education, occupational status, etc...). To simplify the empirical analysis, I rewrite the previous equation as:

$$H_{i2} = \beta(B)lnS + \lambda_i + \mu_s + H_{i1} + \varepsilon \quad (5)$$

where $\beta = \alpha\eta + \delta(B)$.

1.4 Data

1.4.1 Main Sources

In this study, I create a unique dataset, linking microdata on student achievements to spatial data relating to internet broadband coverage in Southern Italy. In this way, I build up a comprehensive pupil-level dataset, enriched with regional, town and school-level data.

Data come from four main sources:

a. Schooling outcomes

The Italian National Institute for the Evaluation of the Education System (INVALSI) is a research institute with the status of a legal entity governed by public law. It is responsible for the annual assessment of the competencies of Italian students in both reading and mathematics. Tests are taken at a number of given grades (2, 5, 6, 8 and 10) and at a national level. Historically, the test was optional and suffered from a significant spatial bias in response rate. Since 2012, the test has become compulsory and recently has reached a 97% coverage at secondary school level (2016/2017). Every year, the Institute publishes anonymised microdata on student performance[4]. Since 2008, individual marks have been matched with a rich set of individual information, allowing for control of personal, family and school characteristics. I have information on the test results for the whole population of students in the 2013/2014 2016/17 school years at grades 5 and 8 and have retrieved from the dataset individual information for two large samples of Italian students: 560,000 primary school students (grade 5) who took the test in 2013/2014 and 548,000 lower secondary school students (grade 8) who took the test in 2016/17. Invalsi dataset also provides a rich set of variables concerning student characteristics. In addition to the student-level information, I have access to the level of education (according to the ISCED scale) of both parents, as well as to their occupational status, recorded in the Socio-Economic Index of occupational

[4] As highlighted by a vast literature (see Bertoni, 2013 for a review) a common problem with test-based accountability systems in education is that they generate incentives for teachers, students and school administrators to manipulate test outcomes. For this reason, the dataset reports both raw and cheating-adjusted results, obtained by means of a probabilistic algorithm that takes into account personal information (beyond that available to the public), score fluctuations and suspicious patterns. Moreover, the algorithm takes into account relevant discrepancies with the 5% classes that, for every test, have been assigned an external examiner.

status (ISEI) and an index based on individual-level economic, social and cultural information (ESCS)[5].

b. Ministry of education

In 2011, the Ministry of Education, in accordance with the Community Guideline on public access to information held by public authorities, started to publish data on each State-recognized school, for any grade. Since 2012, all schools provide, on a yearly basis, information concerning the number of students enrolled in each grade by gender and nationality, number of classes, number of teachers, the school basic budget and a self-evaluation document, to be sent to the Ministry at the end of the academic year. Moreover, 80% of schools provide information on staff (age, type of contract, level of training), environment (desktops, ICT technologies, Wi-Fi coverage) and other relevant features. For institutes providing different stages of education, data are gathered for each stage. Moreover, each school provides the full address of each building (*plesso*) belonging to the school.

c. 2011 Italian socio-economic Census

The Italian Census is a large survey conducted every ten years by the National Institute for Statistics (Istat). The survey is divided in three main sections. The first section, the Agricultural General Census, provides complete information relating to the structure of the agricultural system on a national, regional and local level. The Industry and Service Census focuses instead on the production system, providing the most detailed source of information available. Both censuses are used to develop statistical strategies to conduct any sample-based surveys during the following decade. The third and most relevant survey is the Population and Housing census, which covers the whole population residing in the country at the census date. The primary objective of this survey is to update and review personal data, calculate the legal population level and gather information on the number and structural characteristics of houses and other buildings. Since the 1991 census, the collected micro-data have been linked to a complete digital database in ArcInfo format at a scale of 1:25.000, integrating remote sensing images, IGMI maps and technical maps at regional level with information relating to the municipality. This advanced methodology allowed the Istat to produce detailed geocoded data on the Italian territory, which is divided into 402,000 areas. On average, each section hosts 142 people and for each one, the Istat releases information concerning the number of people living in each division, by gender and age class. Furthermore, the dataset can be matched to information on wage, occupational status and other social features on the basis of a 5-10% population sample living in each division.

[5] The ESCS is a composite score built by the indicators parental education (ISCED), parental occupation (ISEI), and home possessions (HOMEPOS) via principal component analysis. The index was developed by the OECD Pisa (Programme for International Student Assessment) and is often used in international standardised tests. Invalsi builds the HOMEPOS component using information about the availability of a ‘quiet place to study’, a private room, a desk, encyclopaedias, a personal computer and access to the internet.

d. Infratel area-level data

The Infratel dataset contains information gathered through the monitoring process of the Italian Ultra-Broadband Plan. In 2014, a consultation was carried out in order to collect information on the availability of broadband telecommunication infrastructures and on the private investment plans for the following three years. Infratel divided the entire Italian territory into 94,645 areas and separately assessed each one. Once it had obtained data on internet speed and had assigned each area to one of the four clusters, according to the presence/interest of private actors, Infratel, on behalf of the Ministry of Local Development, started to update the dataset every three months. Data are based on information provided by the private sector at the beginning of the process together with updated data from the ministry relating to the implementation process. As a result, the dataset is able to provide a comprehensive representation only for white areas, where only the public sector operates. Specifically, the dataset provides historical data on:

- Population coverage to at least 100 Mbps.
- Population coverage to at least 30 Mbps.
- Public Administration coverage to at least 100 Mbps.
- Businesses coverage to at least 30 Mbps.
- Businesses coverage to at least 100 Mbps.

Population coverage is measured as the percentage of houses (flats) that have access to 30/100 Mbps internet speed. The dataset can be linked to a spatial dataset providing geocoded boundaries for each section. In this way, the information gathered can be easily matched with other spatial datasets.

e. Final Dataset

In the final dataset, the census areas are matched with the Infratel areas. In this way, I can create a new micro-aggregated spatial unit, corresponding to the Infratel areas, reporting weighted broadband coverage measures, where the weight is the relative share of students living in the area (see section 1.4.2 for a complete explanation of the full procedure). Information relating to individual students is matched with school data provided by the Ministry of Education and, via the school coordinates, to the data associated with the territory in which each school is located. Thus, for each student, I have the results in the national exams, a rich set of individual and family characteristics, various information about the school attended and the town in which they live and the weighted broadband coverage measured in the proximity to the school. A complete description of the variables used in the analysis is reported in Table A1. 2.

1.4.2 Broadband Measure and Catchment Areas

Ideally, in order to correctly identify the effect of the policy, I would assign students' homes to treated and control groups. As I do not have information student addresses, I assign students to treatment and control groups by defining geographical catchment areas around each school. Each area c is defined so as to guarantee that student i attending school s lives within the defined borders. In this way, I can assign the whole catchment area to a specific group and subsequently focus the analysis on treated and non-treated pupils living in neighbouring areas. De Simone (2013) developed a method to identify an area within which most resident students would be attending a specific school.

The strategy relies on two main features of the Italian lower secondary school system. First, as discussed in section 1.2.2, schools are characterised by a high degree of homogeneity, especially at the province level. Second, the enrolment follows rigid geographic criteria. Each school is assigned limited funds and a maximum number of students. When applying students exceed the available places, schools are allowed to select students according to a number of criteria, but are expected to take into account home-to-school distance as the main criterion. In summary, parents have little voice in the choice of the school among different institutions that accept students primarily on a distance basis. When choice is possible, the high homogeneity between institutions still guarantees allocation based mainly on geographical criteria. As a result, the design of the catchment areas, especially in rural areas, appears to be a suitable method with which to assign each student to a specific broadband area.

This strategy exploits the Italian census 2011, which provides information on population by age at a very low spatial scale (402,000 census areas). Specifically, for each school j, the association procedure consists of the following steps:

1. identify the school type (primary, lower secondary or upper secondary) and, consequently, the relevant student population in the census area (population aged 5-9 years, 10-14 years or 15-19 years respectively);
2. compute the distance between school j and the 400 nearest census areas;
3. for each school, neighbouring census areas are sorted by distance (in ascending order);
4. compute the areas' cumulative relevant population;
5. select the closest N areas so that the cumulative relevant population contains a multiple k of the number of students enrolled in school j.
6. For each school, I create two different catchment areas: a small catchment area ($k = 1$), where at least 80% of the enrolled students live and a large one ($k = 3$), where all enrolled students (and many non-enrolled) live.

Once the data have been extracted and the catchment areas defined, I can build a proxy for Broadband Access, BA_{ct}^{sh}, obtained as the weighted average of the broadband coverage measured in each catchment area c. The weights used are the share of students living in each Infratel area (aggregated from the smaller census areas) to the total catchment area.

$$BA_{ct}^{sh} = \sum_{i \in c} ICT_{ict} \left(\sum\nolimits_{s \in i} \frac{n_{sict}}{N_{ict}} \right)$$

where ICT_{ict} is the share of buildings with access to internet broadband in the Infratel area i, belonging to catchment area c , n_{sict} is the number of students living in the census area s and N_{ict} is the total number of students living in Infratel area i, in catchment area c.

In the analysis, I use two different values of k to define the treatment and control groups. Treated schools are defined as schools with a broadband coverage measured in the basic catchment area (k = 1) of above 70%. In Table A1. 5 (Appendix), I replicate the main results with 65% and 75% thresholds, confirming the results. Control schools are defined as schools with a broadband coverage measured in the large catchment area (k = 3) equal to 0%. Finally, I drop any large catchment areas where the share of white areas (clusters C and D) is lower than 95%. Figure 1. 2 reports the catchment areas built around a school in the Matera Province (Calabria region). The colours represent the number of students living in each census area, the red line defines the small catchment area (k = 1) and the blue line defines the large one. For both regions, the broadband coverage value is measured, following the above index, as the weighted average share of buildings covered, where the weights correspond to the relative share of students living in each census area.

1.5. Empirical Strategy

1.5.1 Basic model

In the analysis, I first estimate the following model:

$$y_{ist} = \beta BA_{ct} + \gamma X_{is} + \rho S_{st} + \theta W_c + \mu_v + \epsilon_{ist} \quad (6)$$

where y_{ist} represent student's i achievement in school s at time t, BA_{ct} is a dummy that takes value equal one if the weighted share of buildings located in catchment area c with access to high-speed internet broadband exceeds 70%, X_{is} is a vector of time-invariant individual characteristics, S_{st} is a vector of school-level characteristics, W_c is a vector of average socio-economic characteristics of the school's catchment area, μ_v are Province fixed effects and $\epsilon_{ist}$is the unobserved error term. In order to identify the causal effect of high-speed broadband on student achievement, I need to address various sources of bias. Many studies in the literature exploit data on internet usage, which often represents the only available information. The choice of this variable is problematic for a number of reasons. First, survey data, especially when covering technical information, are affected by attrition and measurement error. Moreover, internet usage is likely to be highly correlated with several variables related to the tested outcome, such as social background, profession, individual network and income. In this case, the treatment would not be orthogonal with respect to the unobservable individual characteristics (corr(BA_{ct}, ϵ_{ist}) $\neq$ 0). For all these reasons, I choose to focus on the broadband supply, measured as the number of houses with access to 30 Mbps internet broadband.

Broadband access measures are not exempted from endogeneity concerns. First, the empirical strategy must address possible selection bias. This may occur because broadband access can determine a self-selection of different groups across regions with different levels of coverage. Moreover, even excluding an active sorting based on internet speed, there are several reasons to believe that broadband rollout may be far from random. Exchange stations[6] are normally located close to central locations, where they can guarantee the best connectivity to high income households and offices. Faber et al. (2015) address this problem by adopting a neighbouring discontinuity design, able to guarantee a high degree of homogeneity between the treated and control group. This strategy is particularly effective in addressing endogeneity concerns, but it inevitably requires the analysis to focus on a small sample of the available data. Moreover, it is possible to perform a neighbouring discontinuity design only when the treatment and outcome variables share the same level of geographical detail. This is not the case for Italian student data, which can only be linked to school catchment areas. Data on distance from the closest local exchange station are exploited by Campante et al. (2017), who study the diffusion of access to high-speed internet using Italian municipal data from 1996 to 2013. The strategy is based on the assumption that the cost of providing ADSL-based broadband services varies depending on its relative position in the pre-existing voice telecommunications infrastructure. Since the pre-

[6] Exchange stations are physical infrastructure through which Internet service providers exchange Internet traffic between their networks.

existing infrastructure was not randomly distributed, the authors implicitly assume that the correlation between distance and unobserved municipal characteristics has not changed during the period considered, other than through the introduction of high-speed internet. In other words, firms and households may differ in terms of time-invariant unobservables, but are assumed to have, for instance, the same wage/productivity growth. This assumption appears to be relatively strong and would require strong supporting evidence to justify. In fact, the regional economics literature provides rich evidence of rising regional disparities in most developed countries, including Italy (A'Hearn and Venables, 2013).

Instead of relying the existing infrastructure, this paper exploits the specific design of the Italian 'National Ultra-Broadband Plan', which guarantees the timing of the rollout to be exogenous. This assumption is based on three features of the policy. First, the policy targets *white areas*, consisting in small-to-medium towns in the countryside (where the average population is 20,000) and there are no overlaps with the interests of private actors. Second, the introduction of the new fibre broadband does not depend on the pre-existing infrastructure, since no compatible infrastructure exists in *white areas* at the time the policy is implemented. Third, even though some geographical characteristics associated with the lack of coverage of any previous infrastructure may still influence the implementation costs, the policy aims to cover 100% of municipalities within three years. As a consequence, implementation costs can hardly be correlated with the rollout timing. On the other hand, an efficient implementation would imply a progressive geographical coverage. Moreover, the timing and the universal target undermine the risk of a spatial sorting of households based on treatment. These assumptions alone do not allow political bias in the implementation phase to be excluded. Local administrators may lobby to obtain full coverage before neighbouring municipalities, which would result in a selection bias. However, this issue does not appear to be particularly relevant in this context, since the program has been designed by the national Government and Italian political system and does not have a relevant representation at the local authority level. Moreover, Mayors in small towns lack the political power to deliver relevant changes to a national plan, especially when this kind of change would involve higher costs for the whole project. This issue is further addressed by using province dummies fixed effects. Italian Provinces (NUTS-3 regions according to the European nomenclature) are intermediate administrative divisions. They divide the Italian territories in 107 distinct regions, based on different geological and cultural characteristics. In order to test for within-province homogeneity, I report in Table A1. 6 a simple balancing test, regressing the treatment on all covariates and reporting the p-values and q-values obtained by adopting the Bonferroni correction[7]. Results exhibit no correlation between the policy and local geographical, economic and political variables, with the only exception of the share of population above 64 years[8]. Treatment areas appear to be characterised by a slightly higher share of elders. To account for this possible source of bias, in Table A1. 4, I perform a robustness exercise, repeating all the

[7] When multiple hypotheses are tested, the likelihood of rejecting a null increase. The Bonferroni correction tests each individual hypothesis at a significance level of α/m, where α is the overall desired significance level and m is the number of hypothesis.

[8] In the main specifications I add municipality level covariates to partially address any difference in socio-economic characteristics.

relevant analysis on a sample where every municipality having a share of elders higher than 20% (roughly 1/3 of the original sample) is excluded. Results are consistent with the main analysis.

1.5.2 Value added model

Extending the basic framework, student educational attainment can be described through a value-added model (VAM)

$$y_{ist} = \alpha y_{ist-1} + \beta BA_{ct} + \gamma X_{is} + \rho S_{st} + \theta W_s + \mu_v + \epsilon_{ist} \quad (7)$$

where y_{ist} and y_{ist-1} are, respectively, student i's achievement in school s at time t and $t-1$, BA_{ct} is the broadband dummy, X_{is} is the vector of time-invariant individual characteristics, S_{st} is a vector of time-variant school-level characteristics, W_s is a vector of average socio-economic characteristics of the school catchment area, μ_v are province fixed effects and ε_{ist} is the unobserved error term. Controlling for previous performance, I partially take into account potential time-invariant differences between treated and control students. Moreover, even assuming treatment to be uncorrelated with the performance at *t-1*, $(Y_{i0}, Y_{i1} \perp BA_{ct})$, the autoregressive specification allows me to investigate the heterogeneity of the treatment across different social groups. In particular, I can test the impact of the access to high-speed broadband on student performance, as depending on prior achievement, ethnicity, socio-economic background and nursery attendance.

In the economics of education literature, VAMs have been often used to measure the importance of productivity inputs (such as teacher quality or peer effects) on student performance. Recently, a number of concerns have been raised regarding the opportunity of using VAM models to estimate teacher quality (Kane & Staiger 2008; Hanushek & Rivkin, 2010; Rothstein, 2010; Chetty et al., 2014; Condie et al., 2014). Most of these studies have discussed the possible bias resulting from student sorting and the reliability of standardised tests scores as a proxy for student achievement. These issues do not appear to be relevant in the framework of this study. First, as explained in section 1.2.2, student sorting across schools appears to be a negligible phenomenon in Italy. On the other hand, although unobservable characteristics could, in principle, influence the cheating-corrected test scores, they are unlikely to be correlated with the rollout of the broadband infrastructure. To further address these concerns, in section 1.7 I show that moving home is not correlated with neither the treatment nor the outcome.
If student performance exhibits a mean reverting pattern at the tails of the previous performance distribution, the value-added model might fail to describe the learning process. Therefore, I test a quadratic specification:

$$y_{ist} = \alpha y_{ist-1} + \nu y_{ist-1}^2 + \beta BA_{ct} + \gamma X_{is} + \rho S_{st} + \theta W_{st} + \mu_v + \epsilon_{ist} \quad (8)$$

Figure 1. 4 suggests this this specification could better describe the autoregressive pattern studied.

1.6. Results

1.6.1 Baseline models

I start the investigation by studying whether the infrastructural policy had an effect on student achievements. In Table 1. 1, I regress students' performance in national standardised tests on the policy dummy and a set of covariates. In doing so, I distinguish between results the in literacy and numeracy section of the national standardised test.

Columns (1) and (2) illustrate results when no covariates are considered in the specification. High-speed internet broadband seems to have a small but significant effect on student performance in math, but no significant effect on literacy scores. Access to high-speed internet raises on average student scores in numeracy subjects by 0.88 on a 100-point scale. This corresponds to a 5% standard deviation change.

These results are confirmed in columns (3)-(4), when student-level variables are included. In this case, I find an increase in the coefficient size, that might suggest a small self-selection of low-performing groups in treated areas. This result could raise some concerns regarding the exogeneity of the timing of the roll-out of the broadband infrastructure. If for some reasons students in treated schools in the sample perform worse than those attending control school, results in columns (1)-(2) might be downward biased. In columns (5)-(6) class peer effects (mean class values for a set of student level characteristics) and school level variables are included. Here the size of the coefficient for the impact of broadband on numeracy tests is again closer to the one in Columns (1)-(2), whereas the policy is confirmed not to have any effects on literacy tests. Results are generally confirmed in column (7)-(8) where I add municipality-level variables. In Columns (9)-(10) I substitute the ESCS[9] variable with HISCED[10] and HISEI[11]. These two variables are the main components of the ESCS and adding them separately to the model could increase the overall information content. On the other hand, the ESCS is built also on additional information (see section 4.1) that are not available in the dataset. Again, results suggest a null effect of the policy on student literacy performance and a significantly lower (0.036 standard deviation), but still significant effect on numeracy scores.

In Table 1. 2, I estimate a simple value-added model of cognitive achievement, to investigate the effects of the rollout of the broadband infrastructure on students' learning trajectories.

In columns (1) and (2), the policy is found to have a relevant effect on student performance in the numeracy test, but no significant effect in the literacy assessment. Treated students are found to obtain, on average, 0.675 points more than control ones in the normalised test scores. When adding student-level covariates, the magnitude of the coefficient increases by 15%, suggesting again a limited overrepresentation of low performers or an underrepresentation of high performers in treated areas. Columns (3)-(6) reveal that, once considered the heterogeneity in class composition and school quality, the policy has a less relevant effect on math scores. The coefficient size decreases and becomes less significant. In columns (7) to (10) municipality-level covariates, HISCED and HISEI are included, but the results provide similar conclusions. The access to high-speed broadband has a small, but significant effect on student performance in

[9] Index of individual economic, social and cultural status
[10] Highest educational level of parents
[11] Highest occupational status of parents

numeracy (an increase of 0.5-0.8 test scores, corresponding to 3-5% of a standard deviation) but no effect on literacy tests. Students that attended Grade 5 in the academic year 2013/14 and spent the last year of the lower secondary school in a treated area were, on average, able to strengthen their mathematical skills more than those located in control areas.

These results are in contrast with the 'perfect zero-effect' found by Faber et al. (2017), analysing the effect of high-speed internet on student scores in the UK. In the same way, they contradict Vigdor et al. (2014), that find internet access in North Carolina to be associated with a 2.7% standard deviation decrease in numeracy scores and a 1% decrease for literacy. These results can also be analysed with respect to other policies aimed at improving ICT use in education.

This effect appears to be significantly lower than the one found by Comi et al. (2017), that analyse the ICT-related practices in a small sample of Northern Italian schools. However, the broader effect analysed in this study would justify a lower magnitude.

More generally, the policy is found to have limited effect on student performance when compared to the main policy investigated in the literature. For instance, according to recent literature reviews (Chingos, 2012; Jepsen, 2015), the effect of reducing class size by 10 students on educational performance ranges between 0 to 50% of a standard deviation. However, educational outcomes represent only a second order objective of this kind of infrastructural policy and should be considered together with medium-long term effects on other socio-economic outcomes.

Thus far, I assumed performance in grade 8 to be a linear function of performance in grade 5. In order to account for mean reverting patterns at the lower tale of the test score distribution (see Figure 1. 4), in Table 1. 3 I add a quadratic component to the estimation. The new model does not lead to significant changes with respect to the basic VAM, although the policy is found to have a marginally higher effect on numeracy scores (0.6-0.7 points).

1.6.2 Heterogeneity analysis

Table 1. 4 sheds further light on these findings by looking at the heterogeneous effect of the policy on test scores, controlling for student performance in grade 5 and their socio-economic background. From this point onwards, the analysis focusses exclusively on numeracy test scores, since – consistently with the literature - in previous tables broadband access was found to have a zero effect on literacy performance.

In column (1) I interact the main regressor with the performance in numeracy tests in grade 5. The interaction term is found to be not statistically significant, suggesting either a null effect of the policy on students' learning trajectories or a non-linear relationship between the two variables. In column (2) I further investigate this relationship, interacting the policy dummies with the quartile of the previous test score distribution. Results show a positive and significant effect on the first quartile, a negative effect on the third and a still negative, although non statistically significant on the second. This result is well explained in Figure 1. 5, showing a positive effect of the policy on low-performers in grade 5, a somewhat negative effect on median performers and no relevant effect on best performers. In column (3) I focus on the second relevant source of heterogeneity, namely the parental background. In this case, I find positive and significant contribution of ESCS index on the policy outcome. Interacting the main regressors I find a non-significant effect for students in the first ESCS quartile and a positive and significant effect for those in the remaining part of the distribution.

Many studies emphasise the importance of parents' education for students' mathematical achievements (among others, Lauder et al., 1999; Zimmer & Toma, 2000; Martins and Veiga, 2010; Vigdor et al., 2014). In countries where women are underrepresented in the labour market, mothers play the main role in supporting children's education. In Southern Italy, were women occupation rate is on average below 40%, MISCED is expected be a strong predictor of students' performance. Therefore, in column (5), I interact the policy-dummy with the 1-5 MISCED scale, that describes the level of education of students' mothers. Results provide strong evidence of an important role of this factor on the effectiveness of the policy. Broadband access is found to decrease by 1.5 percentage points the scores of students whose mother has only a primary school degree, whereas it is positive for the other groups, increasing educational scores up to 3 percentage points, corresponding to 0.18% of a standard deviation.
Finally, I investigate school-related factors that might facilitate students' adoption of information devices. In column (6) I test whether students whose teachers were involved in IT training programs in the three years before the policy benefited more from the access to the broadband. In this case, confirming a positive role of trained teachers (the large majority of teachers considered) on the effectiveness of the policy. This aspect deserves to be better analysed in future work, exploiting richer school-level data. Overall, results highlight an important nexus between the effectiveness of the policy and students' socioeconomic background. Low performers whose parents are sufficiently educated, might benefit the most from the introduction of new information technologies.

In Table 1. 5, I further investigate this dynamic, interacting the main regressor with performance in grade 5, within a quadratic VA framework. Column (1) confirms the results of the basic VAM. In columns (2) and (3) I distinguish between students above and below the mean ESCS level. As expected, low performers with above average parental background benefit the most from the policy. These results are consistent with the patterns illustrated in Figure 1. 5.

1.7. Channels

Results suggest a positive and significant effect of the policy on educational achievements.
Consistently with the literature on the ICT effect on student performance, the findings suggest that the access to high-speed internet at home could positively affect students' productivity. However, the empirical design does not allow to disentangle the home-internet access channel from the potential benefit of better computer access at school.
Home and school access to high-speed internet are likely to be complementary. Comi et al. (2017) highlight that ICT per se might not necessarily be beneficial for student achievement. On the contrary, ICT-related teaching methods are likely to increase student performance in math. This

prediction seems consistent with the findings of this study. Access to ICT can positively affect student performance when associated with a rich parental background and IT-trained teachers.

An alternative explanation is that these results are just second-order effects of the policy. The rollout of the high-speed broadband is likely to affect the local economies in two ways. In the short-run, the investment is expected to have a positive effect on employment and wages in some related industries. In the long-run, the new infrastructure might permanently increase firm productivity and benefit the whole economy. On the other hand, an increase in employment and wages might improve families' social status and have non-negligible effects on local public finances. I assume the time span of the study to be too short to expect any relevant economic effects. However, in order to rule out this hypothesis, in Tables Table A1. 7 – A1.8, I regress some socio-economic variables on the policy dummy. I find a significant negative effect on employment and a positive effect on unemployment rate in the short period (1 to 3 years). The results are not surprising, since the treated areas are likely to be located in poorer regions compared to control ones. Once included region FE in the specification, results become insignificant, confirming the absence of a direct effect of the policy on local labour market.

In Table A1. 9, I regress the number of taxpayers (proxy for the labour force), the average income, personal income and municipality tax revenues on the policy dummy. The former variable is particularly important since, while suggesting an improvement in average socio-economic conditions in a specific neighbourhood, it could also be associated with higher contributions to public schools (that receive on average 2% funds from local government). Results suggest no effect of the policy on any of these variables.

The policy outcome could simply result from sorting over space of high-performers. Advantaged families might choose to move in a neighbouring town to get access to high speed broadband. Alternatively, they might simply face higher commuting time in order to have their children attend a school with better access to ICT technologies. As suggested before, this appears rather unlikely, since the policy design is such that the difference between treatment and control areas is expected not to last more than 3 years. As a result, it is hard to imagine a family moving in order to anticipate something they would get in the relatively short period in their own town. Moreover, Italy is known to be one of the countries with lower mobility rate in Europe. Few people move and when this happens, it is to move to the richer North, rather than to a neighbouring town. In Table A1. 10, I regress population change over the policy period for town overall population and for the three cohorts interested by the policy. Again, no significant effect is found. In Table A1. 11, I investigate movers' performance. As with the socio-economic outcomes, I find a negative and significant effect in the OLS, that becomes non-significant once region FE are included. Overall, I find no evidence supporting the hypothesis of high performer sorting in treated areas.

1.8 Robustness Checks

The estimation of the causal effect of the policy requires the correct definition of treatment and control groups. Most treated students are expected to have access to high-speed broadband, whereas no student in the control group should receive the treatment. To clearly identify the two groups, I choose to drop from the sample intermediate situations, in which broadband coverage is above 0% but still significantly below the full coverage. The choice of the minimum threshold necessary to access the treatment group is based on the trade-off between estimation precision and sample size. In Table 1. 1 – 1.9, I adopt a threshold of 70% coverage, which allows me to analyse at least 27,000 students.

Table A1. 5 reports the results for the basic and VAM regressions using the 65% and 75% thresholds. Columns (1) and (6) can be compared with column (7) in Table 1. 1. Reducing the sample to areas where 75% of houses had access to the broadband, increases the policy coefficient by 15%. By contrast, lowering the threshold to 65% leads to a marginal decrease in coefficient magnitude. Overall, there is no significant deviation from the main analysis. Columns (2) and (7) report the results of the same estimation reported in Column (7), Table 1. 2.

Again, lowering the threshold, I obtain a slightly lower, although still significant, coefficient, whereas the opposite happens when I conduct a more conservative analysis. A similar result is shown in columns (3) and (8) reporting the interaction between the main regressor and the ESCS index. A difference is found in columns (4)-(5) and (9)-(10), where I report the estimations for the quadratic models. In this case both analysis report coefficients of higher magnitude with respect to those recorded in the main analysis. However, results confirm the overall pattern shown in the previous tables.

1.9. Conclusions

In this study, I tested the impact of access to high-speed internet broadband on educational achievements. To this aim, I exploited a large infrastructural program implemented by the Italian government. Available data allowed me to investigate the heterogeneous effect of the treatment with respect to students' performance in previous grades and their family background. Overall, broadband access is found to have a positive and significant effect on educational achievements. Consistently with the literature, ICTs play a relevant role in numeracy subjects but do not seem to be relevant for literacy. The effect largely depends on the parental socio-economic background and prior performance. Low performers in grade 5, when benefiting from a rich cultural background, might take advantage of the new learning devices available thanks to the new infrastructure in order to reduce the achievement gap with their peers. On the other hand, the productivity gain does not seem to characterise pupils with a poor background. This might be due to the fact that, without parental supervision, the online-gaming effect might overcome any possible positive effect on learning productivity. Alternatively, it might simply reflect a financial constraint, with poorer families not able to provide their children with the hardware necessary to benefit from the new infrastructure.

Although more work is required to better understand the underlying mechanisms, the preliminary empirical evidence may have relevant policy implications. ICT upgrading programs can be beneficial for students but need to be accompanied by training programs and other policies aimed at allowing disadvantaged students to access the benefits. More generally, infrastructural policies, as well as other place-based policies, are likely to affect students' overall performance and inequalities in the educational attainment.

To my knowledge, this work represents a first evaluation of the impact of the Italian Ultra-broadband policies on second order outcomes. More work is required to shed lights on the main drivers associated with the effectiveness of the policy. On one hand, the progression of the broadband rollout will allow to compare various 'cohorts' of treated municipalities conducting an event study or analogous estimation strategies. On the other hand, the recent release by the Invalsi of the students' identification number has made possible to researcher to track pupils' performance over time. In the future, it will be possible to exploit network analysis strategies to study the effect of the policy though treated students moving to a different school.

References

A'Hearn, B., & Venables, A. J. (2013). Regional disparities: internal geography and external trade. In *The Oxford handbook of the Italian economy since Unification*.

Aesaert, K., & Van Braak, J. (2015). Gender and socioeconomic related differences in performance-based ICT competences. *Computers & Education*, 84, 8-25.

Aesaert, K., van Braak, J., Van Nijlen, D., & Vanderlinde, R. (2015). Primary school pupils' ICT competences: Extensive model and scale development. *Computers & Education*, 81, 326-344.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics* 130(4), 1781-1824.

Ahlfeldt, Gabriel, Pantelis Koutroumpis, and Tommaso Valletti (2017). Speed 2.0: Evaluating access to universal digital highways. *Journal of the European Economic Association* 15(3), 586-625.

Barrera-Osorio, Felipe, & Linden, Leigh L. (2009). The use and misuse of computers in education: Evidence from a randomized experiment in Colombia. *The World Bank Policy Research Working Paper Series* 4836.

Beltran, D. O., Das, K. K., & Fairlie, R. W. (2006). Do Home Computers Improve Educational Outcomes? Evidence from Matched Current Population Surveys and the National Longitudinal Survey of Youth 1997. National Poverty Center Working Paper Series 06-01. *National Poverty Center, University of Michigan*.

Bauernschuster, Stefan, Oliver Falck, and Ludger Woessmann (2014). Surfing alone The Internet and social capital: Evidence from an unforeseeable technological mistake. *Journal of Public Economics* 117 : 73-89.

Bellou, A. (2015). The impact of Internet diffusion on marriage rates: evidence from the broadband market. *Journal of Population Economics*, 28(2), 265-297

Bertoni, M., Brunello, G., & Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools. *Journal of Public Economics*, 104, 65-77

Brunello, G., Checchi, D. (2005). School quality and family background in Italy. *Economics of Education Review*, 24(5), 563577.

Bhuller, Manudeep, et al. (2013). Broadband internet: An information superhighway to sex crime. *Review of Economic Studies* 80(4) 1237-1266.

Campante, F., Durante, R., & Sobbrio, F. (2017). Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*, *16*(4), 1094-1136

Checchi, D., & Flabbi, L. (2013). Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks. *Rivista di Politica Economica*, (3), 7-57

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, *9*(3), 295-320.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593-2632

Condie, S., Lefgren, L., & Sims, D. (2014). Teacher heterogeneity, value-added and education policy. *Economics of Education Review*, 40, 76-92

De Simone, Gianfranco, and Andrea Gavosto (2013). Patterns of value-added creation in the transition from primary to lower secondary education in Italy. *XXVIII national conference of labour economics*, Rome.

De Simone, G. (2013). Render unto primary the things which are primary's: Inherited and fresh learning divides in Italian lower secondary education. *Economics of Education Review*, *35*, 12-23.

Faber, B., Sanchis-Guarner, R., & Weinhardt, F. (2015). ICT and education: Evidence from student home addresses (No. w21306). *National Bureau of Economic Research*.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, *104*(7), 2238-65.

Fairlie, R. W., Beltran, D. O., & Das, K. K. (2010). Home computers and educational outcomes: Evidence from the NLSY97 and CPS. *Economic inquiry*, *48*(3), 771-792.

Fairlie, R. W., & Robinson, J. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, *5*(3), 211-40.

Fiorini, M. (2010). The effect of home computer use on children's cognitive and non-cognitive skills. *Economics of Education review* 29(1), 55-72.

Gibbons, S., Machin, S., & Silva, O. (2013). Valuing school quality using boundary discontinuities. *Journal of Urban Economics*, 75, 15-28

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100(2), 267-71

Ichino, A., & Tabellini, G. (2014). Freeing the Italian school system. *Labour Economics*, 30, 113-128

Infratel (2017), National Broadband Plan and National ultrawide band Plan, Rome.

INVALSI, 2012a. Sistema Nazionale di Valutazione A.S. 2012/2013 Rilevazione degli apprendimenti.

INVALSI, 2015a. Sistema Nazionale di Valutazione A.S. 2015/2016 Rilevazione degli apprendimenti.

ISTAT (2011). Italian general census. Italian National Institute of Statistics, Rome.

Kane, T. J., & Staiger, D. O. (2008). Are teacher-level value-added estimates biased? An experimental validation of non-experimental estimates. In *National Conference on Value-Added Modelling*, Madison, WI (Vol. 47, pp. 106-118)

OECD (2013) PISA 2012 results in focus: what 15-year-olds know and what they can do with what they know. Oecd - Paris

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education finance and policy*, 4(4), 537-571

Tondeur, J., Van Braak, J., & Valcke, M. (2007). Towards a typology of computer use in primary education. *Journal of Computer Assisted Learning*, 23(3), 197-206

Varian, Hal R. (2010). Computer mediated transactions. *American Economic Review* 100, no. 2: 1-10.

Verbeek, Marno, and Francis Vella (2005). Estimating dynamic models from repeated cross-sections. *Journal of econometrics* 127(1), 83-102.

Vigdor, Jacob L., Helen F. Ladd, and Erika Martinez (2014). Scaling the digital divide: Home computer technology and student achievement. *Economic Inquiry* 52(3), 1103-1119.

Tables

Table 1. 1 Basic Model

VARIABLES	(1) Num.	(2) Lit.	(3) Num.	(4) Lit.	(5) Num.	(6) Lit.	(7) Num.	(8) Lit.	(9) Num.	(10) Lit.
Broadband Access	0.758***	0.0402	0.880***	0.184	0.696***	0.143	0.702***	0.184	0.568**	0.0254
	(0.256)	(0.229)	(0.248)	(0.218)	(0.253)	(0.223)	(0.256)	(0.225)	(0.263)	(0.232)
Male			0.312	-5.04***	0.298	-5.15***	0.288	-5.16***	0.138	-5.321***
			(0.212)	(0.185)	(0.222)	(0.194)	(0.222)	(0.194)	(0.228)	(0.199)
Nursey			1.181***	-0.548	1.271***	-0.481	1.084**	-0.535	1.189**	-0.566
			(0.452)	(0.388)	(0.453)	(0.39)	(0.455)	(0.391)	(0.53)	(0.45)
Non-EU Father			0.398	-0.83	0.467	-0.775	0.492	-0.859	0.55	-1.269
			(0.869)	(0.748)	(0.871)	(0.75)	(0.874)	(0.75)	(0.899)	(0.779)
Non-EU Mother			0.899	-0.423	1.113	-0.353	1.097	-0.367	0.991	-0.47
			(0.745)	(0.645)	(0.745)	(0.647)	(0.745)	(0.647)	(0.763)	(0.662)
ESCS			4.402***	4.18***	4.375***	4.2***	4.383***	4.19***		
			(0.107)	(0.0927)	(0.107)	(0.0932)	(0.108)	(0.0941)		
HISCED									3.710***	3.586***
									(0.108)	(0.094)
HISEI									1.156***	1.090***
									(0.131)	(0.115)
I Gen. migrant			-0.1	-0.132	-0.329	0.185	-0.35	0.196	-0.369	-0.00095
			(1.273)	(1.066)	(1.346)	(1.124)	(1.341)	(1.125)	(1.421)	(1.215)
II Gen. migrant			-2.464**	-0.669	-2.491**	-0.445	-2.560**	-0.409	-2.937**	-0.798
			(1.2)	(1.035)	(1.241)	(1.069)	(1.243)	(1.068)	(1.301)	(1.12)
Early enrolled			-6.203***	-6.13***	-6.015***	-6.16***	-5.985***	-6.17***	-5.874***	-6.117***
			(1.149)	(1.07)	(1.147)	(1.076)	(1.142)	(1.079)	(1.187)	(1.125)
Constant	50.51***	62.20***	49.90***	65.96***	48.24***	64.97***	61.34***	78.79***	50.50***	67.79***
	(0.157)	(0.142)	(0.467)	(0.402)	(1.04)	(0.904)	(5.229)	(4.35)	(5.488)	(4.543)
Obs.	27,697	27,669	27,659	27,631	27,384	27,356	27,384	27,356	25,561	25,533
R-squared	0.028	0.011	0.086	0.103	0.089	0.104	0.091	0.105	0.102	0.118
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Peer Effects	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
School variables	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
Geographic variables	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES

Notes: This table presents the results of the linear regression model reported in equation (6). The dependent variables are students' numeracy and literacy scores in grade 8. Sample includes all students located in white areas, where broadband coverage recorded was below 5% or above 70% (see section 1.4.2). The ESCS index is a proxy for student individual economic, social and cultural status, based on known and unknown family characteristics. HISCED and HISEI measure, respectively, the highest educational level and occupational status of students' parents. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 1. 2 Value-added model

	(1) Num.	(2) Lit.	(3) Num.	(4) Lit.	(5) Num.	(6) Lit.	(7) Num.	(8) Lit.	(9) Num.	(10) Lit.
Broadband Access	0.675***	-0.0740	0.779***	0.0394	0.593**	0.0316	0.596**	0.105	0.574**	0.0240
	(0.239)	(0.206)	(0.235)	(0.200)	(0.239)	(0.205)	(0.242)	(0.207)	(0.249)	(0.213)
Test score - Grade 5	0.392***	0.421***	0.354***	0.375***	0.356***	0.376***	0.356***	0.377***	0.353***	0.372***
	(0.007)	(0.006)	(0.0067)	(0.0056)	(0.00669)	(0.00585)	(0.00669)	(0.00584)	(0.00693)	(0.00612)
Male			0.104	-4.055***	0.0658	-4.178***	0.0575	-4.182***	-0.0920	-4.286***
			(0.201)	(0.172)	(0.210)	(0.179)	(0.210)	(0.179)	(0.216)	(0.185)
Nursey			0.639	-0.832**	0.728*	-0.765**	0.544	-0.800**	0.599	-0.720*
			(0.434)	(0.359)	(0.435)	(0.362)	(0.437)	(0.363)	(0.511)	(0.421)
Non-EU Father			0.517	-0.559	0.574	-0.466	0.576	-0.566	0.823	-0.696
			(0.809)	(0.689)	(0.811)	(0.689)	(0.815)	(0.690)	(0.832)	(0.716)
Non-EU Mother			1.129	-0.0784	1.339*	-0.0477	1.297*	-0.0784	1.206*	-0.212
			(0.699)	(0.586)	(0.697)	(0.584)	(0.697)	(0.584)	(0.717)	(0.602)
ESCS			3.408***	2.776***	3.380***	2.799***	3.367***	2.772***		
			(0.104)	(0.0900)	(0.104)	(0.0905)	(0.105)	(0.0911)		
HISCED									2.861***	2.378***
									(0.104)	(0.0902)
HISEI									0.892***	0.777***
									(0.125)	(0.108)
I Gen. migrant			0.237	0.393	-0.408	0.115	-0.438	0.122	-0.392	-0.0781
			(1.164)	(1.002)	(1.238)	(1.041)	(1.237)	(1.042)	(1.306)	(1.108)
II Gen. migrant			-2.458**	0.0228	-2.750**	-0.0154	-2.788**	0.0398	-3.170***	-0.348
			(1.113)	(0.977)	(1.149)	(1.000)	(1.151)	(1.000)	(1.208)	(1.053)
Early enrolled			-5.07***	-4.37***	-4.926***	-4.389***	-4.868***	-4.364***	-5.014***	-4.091***
			(1.068)	(0.998)	(1.069)	(0.995)	(1.064)	(0.996)	(1.100)	(1.040)
Late Enrolled			0.201	-0.233	0.154	-0.250	0.181	-0.248	-0.0189	-0.399*
			(0.281)	(0.232)	(0.282)	(0.234)	(0.282)	(0.233)	(0.291)	(0.242)
Constant	26.32***	37.00***	28.43***	42.97***	26.64***	42.63***	41.78***	60.02***	36.02***	53.86***
	(0.436)	(0.380)	(0.604)	(0.522)	(1.065)	(0.910)	(4.934)	(4.046)	(5.169)	(4.167)
Obs.	27,697	27,182	27,659	27,149	27,384	26,875	27,384	26,875	25,561	25,121
R-squared	0.146	0.197	0.180	0.240	0.184	0.242	0.185	0.244	0.193	0.250
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Peer Effects	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
School variables	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
Geographic variables	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES

Notes: This table presents regression results of the value-added model reported in equation (7). The dependent variables are students' numeracy and literacy scores in grade 8. Controls include students' performance in grade 5. The sample covers all students located in white areas, where broadband coverage recorded was below 5% or above 70% (see section 1.4.2). The ESCS index provides a proxy of student individual economic, social and cultural status, based on known and unknown family characteristics. HISCED and HISEI measure, respectively, the highest educational level and occupational status of students' parents. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 1. 3: Quadratic value-added model

	(1) Num.	(2) Lit.	(3) Num.	(4) Lit.	(5) Num.	(6) Lit.	(7) Num.	(8) Lit.	(9) Num.	(10) Lit.
Broadband Access	0.682***	-0.0163	0.781***	0.0797	0.599**	0.0700	0.623***	0.137	0.616**	0.0662
	(0.237)	(0.204)	(0.232)	(0.199)	(0.237)	(0.203)	(0.240)	(0.206)	(0.247)	(0.212)
Test score - Grade 5	-0.602***	-0.255***	-0.564***	-0.206***	-0.561***	-0.207***	-0.561***	-0.207***	-0.553***	-0.206***
	(0.0431)	(0.0327)	(0.0421)	(0.0316)	(0.0421)	(0.0318)	(0.0421)	(0.0318)	(0.0439)	(0.0334)
(Test score - Grade 5)^2	0.0083***	0.0059***	0.0077***	0.0051***	0.0077***	0.0051***	0.0077***	0.0052***	0.0076***	0.0051***
	(0.00035)	(0.00027)	(0.00035)	(0.00027)	(0.00035)	(0.00027)	(0.00035)	(0.00027)	(0.00036)	(0.00028)
Male			0.0502	-3.947***	-0.000425	-4.073***	-0.00909	-4.077***	-0.149	-4.179***
			(0.199)	(0.170)	(0.208)	(0.178)	(0.208)	(0.178)	(0.214)	(0.183)
Nursey			0.551	-0.726**	0.631	-0.670*	0.459	-0.708**	0.523	-0.683
			(0.427)	(0.357)	(0.429)	(0.360)	(0.431)	(0.361)	(0.505)	(0.419)
Non-EU Father			0.567	-0.446	0.612	-0.360	0.600	-0.460	0.843	-0.589
			(0.808)	(0.685)	(0.809)	(0.685)	(0.813)	(0.686)	(0.829)	(0.709)
Non-EU Mother			1.164*	-0.0688	1.375**	-0.0291	1.335*	-0.0579	1.223*	-0.194
			(0.697)	(0.582)	(0.695)	(0.580)	(0.695)	(0.581)	(0.716)	(0.598)
ESCS			3.275***	2.610***	3.248***	2.632***	3.232***	2.610***		
			(0.103)	(0.0893)	(0.104)	(0.0898)	(0.104)	(0.0904)		
HISCED									2.761***	2.276***
									(0.103)	(0.0896)
HISEI									0.861***	0.727***
									(0.124)	(0.107)
I Gen. migrant			0.199	0.110	-0.341	-0.0994	-0.365	-0.0900	-0.427	-0.299
			(1.157)	(0.987)	(1.231)	(1.030)	(1.231)	(1.031)	(1.301)	(1.094)
II Gen. migrant			-2.411**	-0.171	-2.767**	-0.155	-2.799**	-0.102	-3.153***	-0.430
			(1.093)	(0.960)	(1.131)	(0.983)	(1.132)	(0.983)	(1.186)	(1.035)
Early enrolled			-5.022***	-4.443***	-4.857***	-4.463***	-4.806***	-4.442***	-4.857***	-4.227***
			(1.048)	(1.004)	(1.048)	(1.003)	(1.043)	(1.004)	(1.086)	(1.047)
Late Enrolled			0.246	-0.168	0.204	-0.185	0.225	-0.185	0.0279	-0.337
			(0.278)	(0.231)	(0.279)	(0.232)	(0.279)	(0.232)	(0.289)	(0.240)
Constant	53.55***	54.37***	53.63***	57.63***	51.76***	57.06***	69.80***	75.03***	63.81***	69.46***
	(1.255)	(0.938)	(1.290)	(0.964)	(1.550)	(1.211)	(5.100)	(4.091)	(5.353)	(4.231)
Observations	27,697	27,182	27,659	27,149	27,384	26,875	27,384	26,875	25,561	25,121
R-squared	0.163	0.212	0.195	0.251	0.199	0.254	0.201	0.255	0.208	0.261
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Peer Effects	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES
School variables	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
Geographic variables	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES

Notes: This table presents regression results of the quadratic value-added model reported in equation (8). The dependent variables are students' numeracy and literacy scores in grade 8. Controls include students' performance in grade 5. The sample covers all students located in white areas, where broadband coverage recorded was below 5% or above 70% (see Section *1.4.2*). The ESCS index provides a proxy of student individual economic, social and cultural status, based on known and unknown family characteristics. HISCED and HISEI measure, respectively, the highest educational level and occupational status of students' parents. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 1. 4: VAM - Interactions

VARIABLES	(1) Num.	(2) Num.	(3) Num.	(4) Num.	(5) Num.	(6) Num.
Broadband Access	1.442*	1.233***	0.693***	-0.554	-1.481*	0.150
	(0.845)	(0.432)	(0.251)	(0.425)	(0.886)	(0.251)
Test score - Grade 5	0.362***		0.350***	0.351***	0.349***	0.343***
	(0.00882)		(0.00693)	(0.00693)	(0.00693)	(0.00650)
Broadband Access x Test score - Grade 5	-0.0136					
	(0.0130)					
Test score - Grade 5, quartiles:						
II quartile		3.944***				
		(0.372)				
III quartile		8.728***				
		(0.383)				
IV quartile		15.85***				
		(0.399)				
Broadband Access x II quartile		-0.896				
		(0.562)				
Broadband Access x III quartile		-1.429**				
		(0.573)				
Broadband Access x IV quartile		-0.156				
		(0.590)				
Broadband Access x ESCS			0.477**			
			(0.206)			
ESCS, quartiles:						
II quartile				-0.625		
				(0.411)		
III quartile				1.327***		
				(0.444)		
IV quartile				2.644***		
				(0.512)		
Broadband Access x II quartile				2.077***		
				(0.579)		
Broadband Access x III quartile				1.175**		
				(0.580)		
Broadband Access x IV quartile				1.462**		
				(0.595)		
MISCED (1-5)						
MISCED=2					1.163*	
					(0.688)	
MISCED=3					0.772	
					(0.944)	
MISCED=4					2.603***	
					(0.796)	
MISCED=5					5.500***	
					(0.976)	
Broadband Access x MISCED=2					1.704*	
					(0.936)	
Broadband Access x MISCED=3					2.125*	
					(1.285)	
Broadband Access x MISCED=4					2.606***	
					(0.941)	
Broadband Access x MISCED=5					3.015***	
					(1.097)	
IT Training						0.130
						(0.513)
Broadband Access x IT Training						3.877***
						(0.722)
Constant	41.46***	57.59***	40.27***	38.01***	42.57***	45.26***
	(4.941)	(4.949)	(5.187)	(5.175)	(5.293)	(5.268)
Observations	27,384	27,384	25,531	25,561	25,531	25,531
Province FE	YES	YES	YES	YES	YES	YES
Individual variables	YES	YES	YES	YES	YES	YES
Peer Effects	YES	YES	YES	YES	YES	YES
School & geographic variables	YES	YES	YES	YES	YES	YES

Notes: This table presents regression results of the value-added model reported in equation (7). The dependent variable corresponds to students' numeracy scores in grade 8. In columns (1) and (2) the broadband access dummy is interacted with previous performance. In column (3), (4) and (5) the main explanatory variable is interacted, respectively, with family socio-economic scores, the occupational status of the mother and with teachers' IT training in the previous years. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

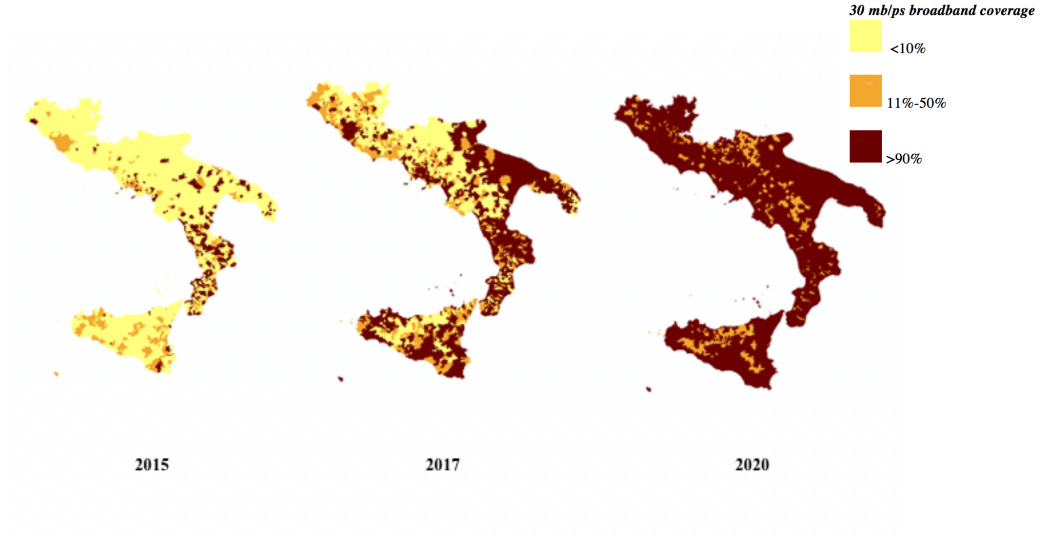
Table 1. 5: Interactions – quadratic model (Math)

	(1)	(2)	(3)
VARIABLES	all	ESCS<0	ESCS>0
Broadband Access	8.068***	7.419**	11.51***
	(2.465)	(3.190)	(3.966)
Test score - Grade 5	-0.430***	-0.273***	-0.581***
	(0.0579)	(0.0785)	(0.0877)
Broadband Access x Test score - Grade 5	-0.258***	-0.245**	-0.355***
	(0.0856)	(0.113)	(0.134)
(Test score - Grade 5)^2	0.00659***	0.00523***	0.00781***
	(0.000482)	(0.000668)	(0.000714)
Broadband Access x (Test score - Grade 5)^2	0.00213***	0.00212**	0.00277**
	(0.000712)	(0.000964)	(0.00109)
Constant	37.02***	35.13***	43.39***
	(1.793)	(2.366)	(2.941)
Observations	25,747	14,261	11,486
R-squared	0.200	0.147	0.186
Province FE	YES	YES	YES
Peer Effects	YES	YES	YES
School variables	YES	YES	YES
Geographic variables	YES	YES	YES

Notes: This table presents regression results of the quadratic value-added model reported in equation (8). The dependent variable corresponds to students' numeracy scores in grade 8. The broadband dummy is interacted with students' performance recorded in grade 5. Results in columns (2) and (3) concern, respectively, student below and above the average family socio-economic score. The sample covers all students located in white areas, where broadband coverage recorded was below 5% or above 70% (see Section *1.4.2*). All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

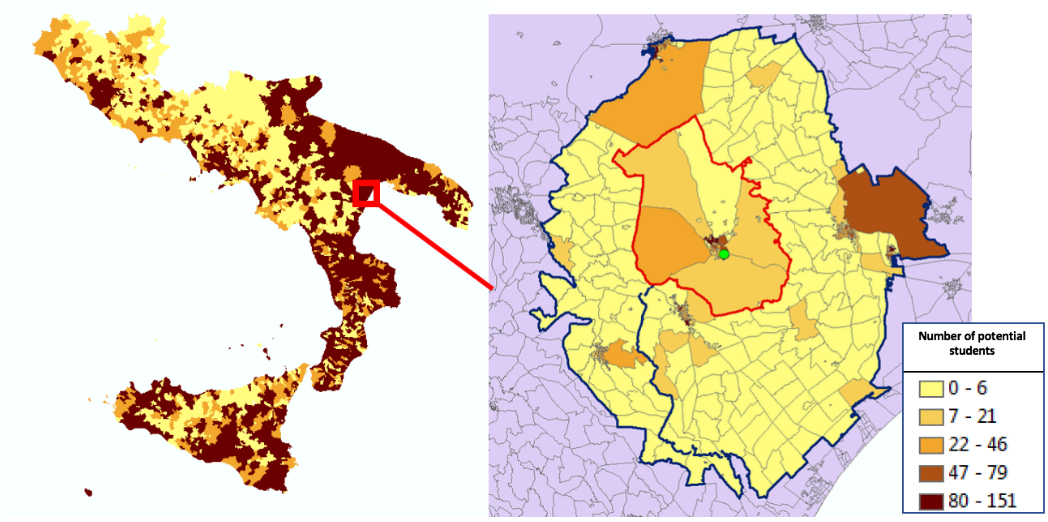
Figures

Figure 1. 1: Broadband coverage



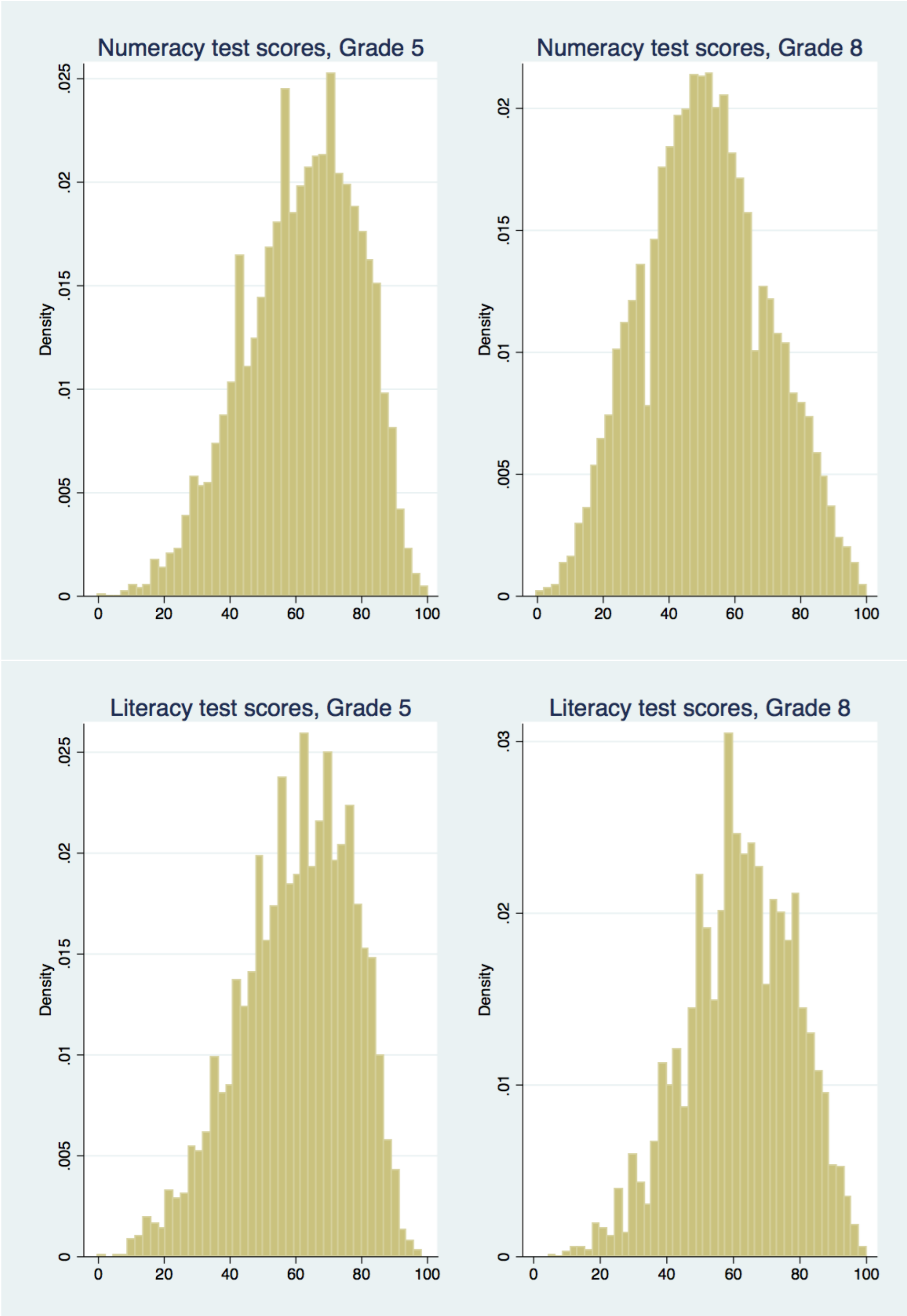
Notes: the maps show the progressive evolution of 30 mb/ps coverage in Southern Italy. For every year, I access sub-municipality level data recorded at the end of January.

Figure 1. 2: Large (blue) and small (red) catchment areas



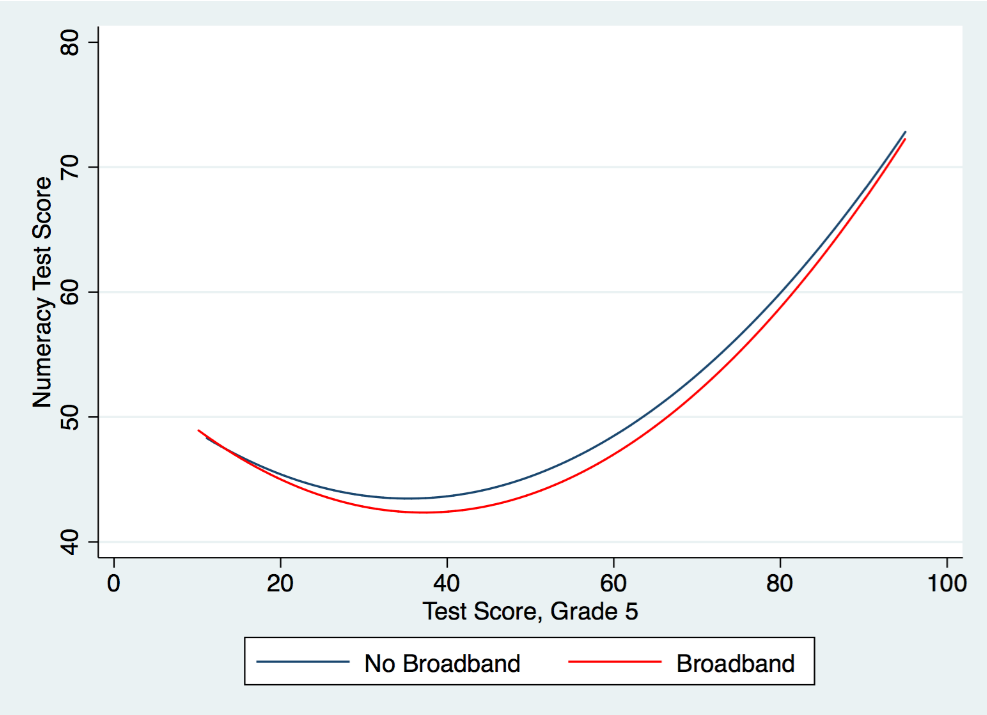
Notes: the map shows a large catchment area (blue line) and a small catchment area (red line) built progressively linking adjacent Census areas, starting from the one where the school is located (green dot). Overall, small catchment areas cover neighbouring census areas where is recorded an 8-grade student population equal or greater than the 8-grade student population enrolled in the reference school. On the other hand, large catchment areas cover territories that host a population at least three times larger than the number of students enrolled.

Figure 1. 3: Test score distributions



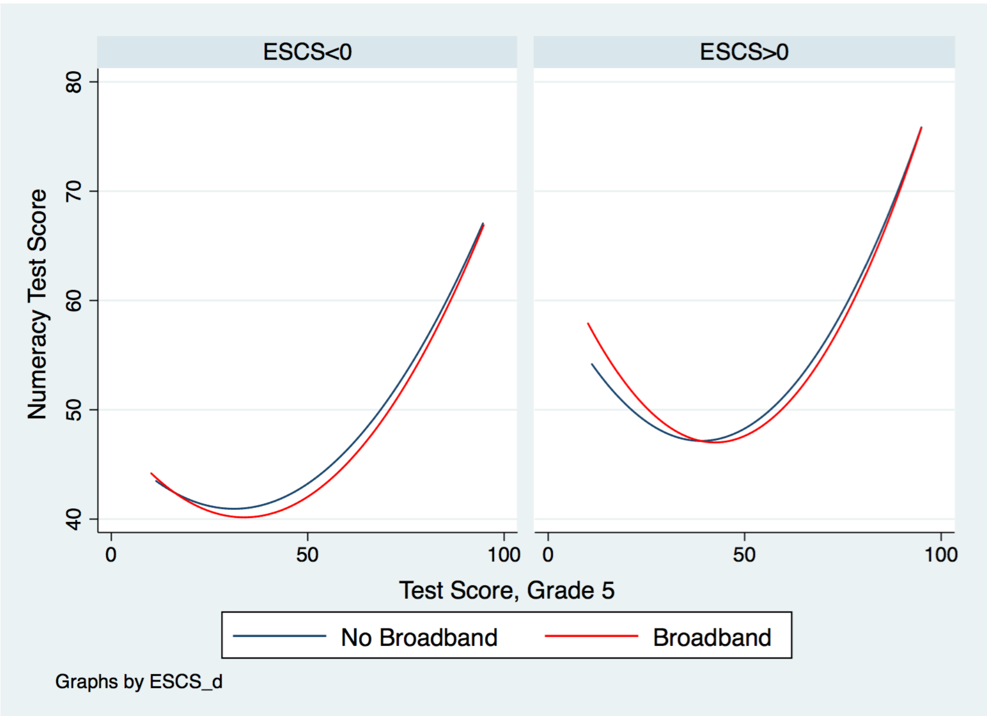
Notes: The histograms report the literacy and numeracy test score distribution for the same cohort of students in grade 5 (2013/2014) and grade 8 (2016/2017).

Figure 1. 4 Grade 8 numeracy scores by performance in grade 5 and treatment



Notes. The graph reports the quadratic prediction for 8-grade students' test scores based on previous performance in grade 5.

Figure 1. 5: Grade 8 test scores by previous performance, treatment and family background



Notes. The graph reports the quadratic prediction for students' test scores based on previous performance and family background (ESCS).

Appendix

Table A1. 1 Subjects

Subjects	Hours per week	Hours per year
Italian Language, History, Geography	9	297
In-depth studies in literary subjects	1	33
Mathematics and Science	6	198
Technology	2	66
English	3	99
Second foreign language	2	66
Art and design	2	66
Sport science	2	66
Music	2	66
Catholic religious education	1	33
Total	30	990

Notes. The table reports hours studied by subject in Italian lower-secondary school.

Table A1. 2 Final Dataset

Variables	count		mean		sd		min		max	
	0	1	0	1	yes	no	yes	no	yes	no
Test score - Grade 5	15,531	12,730	61.84	61.55	17.15	17	0	0	98	100
Male	15,531	12,730	0.49	0.49	0.5	0.5	0	0	1	1
Nursey	15,531	12,730	0.94	0.94	0.24	0.23	0	0	1	1
Non-EU Father	15,531	12,730	0.02	0.02	0.13	0.15	0	0	1	1
Non-EU Mother	15,531	12,730	0.03	0.02	0.14	0.17	0	0	1	1
ESCS[12]	15,505	12,718	-0.09	-0.21	1.02	0.99	-3	-3	3	3
HISCED[13]	14,513	12,119	3.39	3.34	1.16	1.13	1	1	5	5
HISEI[14]	14,463	12,176	3.77	3.71	0.93	0.92	1	1	5	5
I Gen. migrant	15,531	12,730	0.01	0.01	0.08	0.1	0	0	1	1
II Gen. migrant	15,531	12,730	0.01	0.01	0.09	0.12	0	0	1	1
Early enrolled	15,531	12,730	0.01	0.01	0.09	0.1	0	0	1	1
Late Enrolled	15,531	12,730	0.15	0.17	0.37	0.36	0	0	1	1
Male (class share)	15,531	12,730	0.49	0.49	0.14	0.15	0	0	1	1
I Gen. migrant (class share)	15,531	12,730	0.01	0.01	0.02	0.03	0	0	0	1
II Gen. migrant (class share)	15,531	12,730	0.01	0.01	0.03	0.04	0	0	0	0
Class Size	15,531	12,730	20.48	21.95	3.76	4.52	5	2	32	32
IT Training	15,531	12,730	0.88	0.85	0.31	0.27	0	0	1	1
Inclusion Policies	15,271	12,710	5.01	4.93	0.94	1.01	2	1	7	7
HR Management	15,271	12,710	4.59	4.56	1.01	0.98	2	1	7	7
evaluation	15,271	12,710	4.67	4.77	1.01	0.96	2	1	7	7
Student Monitoring	15,271	12,710	4.64	4.74	1	0.99	3	2	7	7
Strategic Guidance	15,271	12,710	4.75	4.8	0.93	0.98	3	1	7	7
ln City size	15,531	12,730	9.61	9.93	0.78	1.21	7.1	6.5	13	13
Ln Avg. Income	15,531	12,730	9.01	8.96	0.36	0.37	5.7	3.9	14	17
Income below 10000, share	15,531	12,730	1.87	1.85	0.91	0.84	0.6	0	12	12
income above 55000, share	15,531	12,730	43.7	43.05	5.93	5.88	20	18.7	62	64
Population below 15	15,531	12,730	14.95	16.54	2.48	2.23	10.6	8.1	24	21
Population above 64	15,531	12,730	18.54	15.98	3.9	4.05	7.3	8.1	27	34
Altitude	15,531	12,730	3.7	4.18	1.04	1.25	1	1	5	5

Notes. The table reports descriptive statistics for the main variables of interest.

[12] Index of individual economic, social and cultural status
[13] Highest educational level of parents
[14] Highest occupational status of parents

Table A1. 3 Groups (Math)

	(1)	(2)	(3)	(4)
VARIABLES	Foreigner		Nursery	
	YES	NO	YES	NO
Broadband Access	2.393	0.672***	0.697***	-4.121***
	(1.865)	(0.235)	(0.238)	(1.532)
Male	1.591	0.183	0.153	0.865
	(1.428)	(0.201)	(0.206)	(0.822)
Nursery	-0.0656	0.611		
	(2.075)	(0.439)		
Non-EU father	4.344*	0.900	0.584	-3.492
	(2.285)	(0.850)	(0.834)	(3.202)
Non-EU mother	-2.305	1.565**	1.243*	2.074
	(2.317)	(0.726)	(0.727)	(2.541)
ESCS	1.665**	3.488***	3.489***	3.361***
	(0.837)	(0.104)	(0.106)	(0.467)
Foreigner			-2.802**	-0.291
			(1.157)	(3.912)
Early Enrolled	0.245	-6.213***	-5.209***	-3.654
	(1.903)	(1.224)	(1.156)	(2.772)
Late Enrolled	2.156	0.153	0.148	0.889
	(2.223)	(0.280)	(0.290)	(1.048)
Test score - Grade 5	0.405***	0.340***	0.344***	0.285***
	(0.0460)	(0.00653)	(0.00669)	(0.0268)
Constant	21.68***	29.58***	29.98***	34.04***
	(3.424)	(0.601)	(0.455)	(1.903)
Observations	557	27,990	26,553	1,670
R-squared	0.202	0.182	0.185	0.179
Province FE	YES	YES	YES	YES

Notes. This table presents the main results of the value added model reported in equation (7), highlighting the different policy effects for foreign students and students that did not attend nursery school. The dependent variables is students' numeracy score in grade 8. Sample includes all students located in white areas, where broadband coverage recorded was below 5% or above 70%. The ESCS index is a proxy for student individual economic, social and cultural status, based on known and unknown family characteristics. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 4 No municipalities with share of population older than 65 >25%

VARIABLES	(1) BASIC MAT	(2) VAM MAT	(3) INTER MAT	(4) VAM MAT Q	(5) INTER MAT Q
Broadband Access	0.777**	0.867***	0.832***	0.886***	7.587***
	(0.317)	(0.285)	(0.309)	(0.283)	(2.942)
ESCS			1.186***		
			(0.233)		
Broadband Access * ESCS[15]			0.620**		
			(0.243)		
Test score, Grade 5		0.373***	0.330***	-0.581***	-0.369***
		(0.00792)	(0.00830)	(0.0511)	(0.0742)
(Test score, Grade 5)^2				0.00805***	0.00576***
				(0.000425)	(0.000620)
Broadband Access * Test score, Grade 5					-0.282***
					(0.102)
Broadband Access * (Test score, Grade 5)^2					0.00260***
					(0.000853)
Constant	55.55***	26.86***	49.21***	52.92***	35.52***
	(6.635)	(0.518)	(6.366)	(1.484)	(2.259)
Observations	20,302	20,211	18,594	19,732	18,216
R-squared	0.114	0.143	0.198	0.159	0.198
Province FE	YES	YES	YES	YES	YES
Peer Effects	YES	NO	YES	NO	YES
School variables	YES	NO	YES	NO	YES
Geographic variables	YES	NO	YES	NO	YES

Notes. The table reports the main regression results presented in Tables 1.1-1.5, estimated using a reduced sample, where all municipality where the share of population was above 25% are excluded. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

[15] Index of individual economic, social and cultural status

Table A1. 5 Alternative thresholds

	Broadband Coverage>65%					Broadband Coverage>75%				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Basic	VAM	INT	Q VAM	INT– Q VAM	Basic	VAM	INT	Q VAM	INT– Q VAM
Broadband Access	0.543*	0.634***	0.637**	0.732***	9.026***	0.661*	0.710***	0.651***	0.603***	9.537***
	-0.272	-0.25	-0.26	-0.247	-2.579	-0.255	-0.232	-0.244	-0.229	-2.403
ESCS			1.519***					1.444***		
			-0.202					-0.193		
Broadband Access * ESCS			0.444**					0.398**		
			-0.216					-0.202		
Test score - Grade 5		0.384***	0.341***	-0.613***	-0.442***		0.389***	0.347***	-0.643***	-0.444***
		-0.007	-0.00725	-0.0451	-0.0593		-0.0065	-0.0068	-0.0421	-0.0574
(Test score - Grade 5)^2				0.0084***	0.0067***				0.0087***	0.0067***
				-0.00037	-0.00049				-0.00035	-0.00048
Broadband Access * Test score - Grade 5					-0.280***					-0.305***
					-0.0896					-0.0835
Broadband Access * (Test score - Grade 5)^2					0.0022***					0.0024***
					-0.00075					-0.0007
Constant	54.37***	26.63***	44.82***	53.96***	36.95***	45.18***	26.57***	33.09***	54.82***	37.58***
	-5.645	-0.455	-5.33	-1.313	-1.842	-5.273	-0.428	-4.97	-1.226	-1.775
Observations	25,652	25,652	23,565	25,652	23,783	28,787	28,787	26,515	28,787	26,760
R-squared	0.101	0.139	0.192	0.157	0.193	0.1	0.144	0.193	0.163	0.199
Province FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Peer Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographic variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes. The table reports the main regression results presented in Tables 1.1-1.5, estimated using two different thresholds to identify the treatment: broadband overage>65% and >70%. Baseline analysis are conducted with a 70% threshold. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 6 Balancing test

Dependent variable:	Broadband access							
	OLS				FE			
	Coeff.	se	p_value	q_value	Coeff.	se	p_value	q_value
Test score, Grade 5- math	-0.318	0.196	0.105	1	0.211	0.227	0.352	1
Test score, Grade 5- literacy	0.160	0.203	0.427	1	0.497	0.234	0.033	1
Male	-0.002	0.016	0.890	1	-0.019	0.024	0.432	1
Nursey	0.007	0.019	0.722	1	0.032	0.026	0.226	1
Non-EU Father	-0.012	0.006	0.052	1	-0.001	0.009	0.895	1
Non-EU Mother	-0.017	0.007	0.020	0.629	-0.001	0.010	0.899	1
ESCS[16]	0.002	0.065	0.803	1	0.027	0.084	0.732	1
HISCED[17]	0.001	0.045	0.981	1	0.038	0.065	0.561	1
HISEI[18]	-0.013	0.037	0.733	1	0.013	0.053	0.805	1
II Gen. migrant	-0.007	0.005	0.135	1	-0.001	0.007	0.838	1
II Gen. migrant (class share)	-0.006	0.004	0.125	1	-0.001	0.005	0.826	1
Early enrolled	-0.001	0.005	0.799	1	-0.002	0.007	0.737	1
Late Enrolled	-0.006	0.013	0.650	1	-0.022	0.019	0.244	1
Male (class share)	-0.011	0.015	0.477	1	-0.020	0.022	0.367	1
I Gen. migrant	0.000	0.003	0.990	1	0.000	0.004	0.912	1
I Gen. migrant (class share)	0.000	0.003	0.966	1	0.000	0.004	0.971	1
Class Size	2.357	0.547	0.000	0.001	1.406	0.710	0.048	1
IT Training	0.02	0.030	0.475	1	0.058	0.044	0.186	1
I Gen. migrant	0.000	0.003	0.990	1	0.000	0.004	0.912	1
I Gen. migrant (class share)	0.000	0.003	0.966	1	0.000	0.004	0.971	1
Inclusion Policies	0.125	0.104	0.231	1	-0.107	0.144	0.456	1
HR Management	-0.024	0.012	0.047	1	-0.029	0.013	0.030	1
evaluation	0.281	0.113	0.013	0.409	-0.001	0.157	0.997	1
Student Monitoring	0.134	0.111	0.230	1	0.137	0.158	0.389	1
Strategic Guidance	0.196	0.114	0.085	1	-0.061	0.159	0.702	1
altitude	0.463	0.165	0.005	0.174	0.222	0.169	0.190	1
Class Size	2.357	0.547	0.000	0.001	1.406	0.710	0.048	1
Early enrolled	-0.001	0.005	0.799	1	-0.002	0.007	0.737	1
evaluation	0.281	0.113	0.013	0.409	-0.001	0.157	0.997	1
ln Avg. Income	-0.037	0.017	0.036	1	0.004	0.022	0.856	1
Income below 10000, share	-0.771	0.923	0.404	1	-0.939	0.987	0.342	1
income above 55000, share	0.141	0.131	0.281	1	0.087	0.149	0.560	1
Population above 64	-2.572	0.048	0	0	-0.581	0.045	0.000	0
Altitude	0.4783	0.0140	0	0	-0.036	0.013	0.004	0.134

Notes. The table reports a simple balancing test, studying the relation between the treatment dummy and the main observables. Standard p-value and Bonferroni q-value are reported for the OLS and the FE specification.

[16] Index of individual economic, social and cultural status
[17] Highest educational level of parents
[18] Highest occupational status of parents

Table A1. 7 Employment

VARIABLES	(1) Employment rate 1d	(2) Employment rate 2d	(3) Employment rate 3d	(4) Employment rate 1d	(5) Employment rate 2d	(6) Employment rate 3d
Broadband	-0.00536***	-0.00711**	-0.00446	0.000848	-0.00104	-0.00244
	(0.00201)	(0.00359)	(0.00367)	(0.00357)	(0.00262)	(0.00184)
Constant	0.551***	0.769***	0.689***	0.665***	0.741***	0.537***
	(0.0358)	(0.0638)	(0.0653)	(0.0597)	(0.0438)	(0.0308)
Observations	415	415	415	415	415	415
R-squared	0.017	0.009	0.004	0.190	0.545	0.293
Region FE	NO	NO	NO	YES	YES	YES

Notes: This table presents regression results for a simple OLS specification studying the relation between the policy and the local employment rate recorded 1-3 years after the treatment. The results support the assumption that, within provinces, the treatment does not have a significant effect on employment rates. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 8 Unemployment

VARIABLES	(1) Unemployment rate 1d	(2) Unemployment rate 2d	(3) Unemployment rate 3d	(4) Unemployment rate 1d	(5) Unemployment rate 2d	(6) Unemployment rate 3d
Broadband	0.00350	0.0124**	0.00921***	-0.00121	0.00117	0.00201
	(0.00356)	(0.00493)	(0.00297)	(0.00287)	(0.00374)	(0.00274)
Constant	0.164***	-1.347***	-0.278***	0.186***	-1.295***	-0.245***
	(0.0633)	(0.0876)	(0.0528)	(0.0480)	(0.0624)	(0.0458)
Observations	415	415	415	415	415	415
R-squared	0.002	0.015	0.023	0.443	0.513	0.285
Region FE	NO	NO	NO	YES	YES	YES

Notes: This table presents regression results for a simple OLS specification studying the relation between the policy and the local unemployment rate recorded 1-3 years after the treatment. The results support the assumption that, within provinces, the treatment does not have a significant effect on unemployment rates. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 9 Income

VARIABLES	(1) Taxpayers 1d	(2) Average Employee Income 1d	(3) Average Personal Employer Income 1d	(4) Municipality Tax Revenue 1d	(5) Taxpayers 3d	(6) Average Employee Income 3d	(7) Average Personal Employer Income 3d	(8) Municipality Tax Revenue 3d
Broadband	-0.000	-0.00005	0.00155	0.00740	-0.0008	0.00002	0.00137	0.00497
	(0.000)	(0.00015)	(0.00099)	(0.00590)	(0.00011)	(0.00016)	(0.00118)	(0.00686)
Constant	-0.0249***	0.0133***	0.189***	0.378***	-0.0250***	0.0130***	0.190***	0.389***
	(0.00184)	(0.00267)	(0.0180)	(0.103)	(0.00169)	(0.00250)	(0.0185)	(0.106)
Observations	413	413	365	410	413	413	365	410
R-squared	0.002	0.000	0.007	0.004	0.240	0.206	0.063	0.058
Province FE	NO	NO	NO	NO	YES	YES	YES	YES

Notes: This table presents regression results for a simple linear model studying the relation between the policy and four different proxies of local income. The results support the assumption that, within provinces, the policy does not have a significant effect on local income. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 10 Mobility

VARIABLES	(1) Pop change	(2) 2000-born change	(3) 2001-born change	(4) 2002-born change
Broadband	-0.0001	0.0003	0.00007	-0.0005
	(0.000113)	(0.000419)	(0.000313)	(0.00185)
Constant	-0.00831***	0.00870	-0.00198	0.0303
	(0.00179)	(0.00661)	(0.00494)	(0.0291)
Observations	415	415	415	415
R-squared	0.259	0.037	0.065	0.024
Province FE	YES	YES	YES	YES

Notes: This table presents regression results for a simple OLS specification studying the relation between treatment and mobility patterns characterising the entire population and the three student cohorts considered in this study. The results support the assumption that, within provinces, the policy does not have a significant effect on mobility patterns. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 11 Sorting

VARIABLES	(1) Num.	(2) Num.	(3) Num.	(4) Num.	(5) Num.	(6) Num.	(7) Num.	(8) Num.
Mover	-2.29***	-2.227***	-0.240	-0.518				
	(0.386)	(0.420)	(0.579)	(0.617)				
Mover (from control to treatment areas)					-2.413***	-2.265***	-0.380	-0.804
					(0.538)	(0.587)	(0.758)	(0.803)
Male		0.574***		0.452***		0.570***		0.452***
		(0.190)		(0.164)		(0.190)		(0.164)
Full Time		0.389**		0.829		0.388**		0.826
		(0.195)		(0.625)		(0.195)		(0.625)
EU born		-0.692		-0.680		-0.717		-0.686
		(0.989)		(0.860)		(0.989)		(0.860)
Nursery		1.541***		1.246**		1.547***		1.249**
		(0.304)		(0.594)		(0.305)		(0.594)
Italian Parents		-0.354		0.243		-0.325		0.243
		(0.331)		(0.298)		(0.331)		(0.298)
HISEI		1.029***		0.822***		1.026***		0.823***
		(0.111)		(0.0978)		(0.111)		(0.0978)
HISCED		2.335***		2.402***		2.332***		2.402***
		(0.0930)		(0.0837)		(0.0931)		(0.0837)
Constant	61.84***	48.86***	61.73***	48.83***	61.78***	48.79***	61.72***	48.82***
	(0.0927)	(0.570)	(0.0818)	(0.782)	(0.0914)	(0.570)	(0.0777)	(0.781)
Observations	36,843	31,484	36,843	31,484	36,843	31,484	36,843	31,484
R-squared	0.001	0.038	0.357	0.381	0.001	0.037	0.357	0.381
School FE	NO	NO	YES	YES	NO	NO	YES	YES

Notes: This table presents regression results for a simple linear model studying the performance of students changing municipality of residents between grade 5 and grade 8. These students constitute less than 0.1% of the population and are excluded from the sample in the baseline analysis. However, a higher performance could produce positive externalities on classmates in treated areas. The results confirm the main results are not driven simply by spatial sorting of high performers. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A1. 12 Primary School

VARIABLES	(1) Primary Literacy	(2) Primary Literacy	(3) Primary Literacy	(5) Primary Numeracy	(6) Primary Numeracy	(7) Primary Numeracy
Male	1.732***	1.759***	1.753***	-2.814***	-2.826***	-2.836***
	(0.0539)	(0.0580)	(0.0544)	(0.0536)	(0.0575)	(0.0550)
Nursey	0.810***	1.074***	2.023***	0.654***	0.842***	1.869***
	(0.107)	(0.122)	(0.205)	(0.108)	(0.122)	(0.213)
Non-EU Father	-1.787***	-1.722***	-1.715***	-2.000***	-1.912***	-1.858***
	(0.172)	(0.186)	(0.178)	(0.173)	(0.186)	(0.181)
Non-EU Mother	-1.209***	-0.844***	-0.925***	-1.647***	-1.311***	-1.297***
	(0.151)	(0.162)	(0.156)	(0.151)	(0.162)	(0.158)
BFMJ[19]		-0.186***	-0.151***		-0.334***	-0.276***
		(0.0422)	(0.0404)		(0.0420)	(0.0409)
BMMJ[20]		0.250***	0.253***		0.111***	0.130***
		(0.0300)	(0.0287)		(0.0297)	(0.0289)
MISCED[21]		1.437***	1.468***		1.648***	1.647***
		(0.0375)	(0.0359)		(0.0374)	(0.0364)
FISCED[22]		1.171***	1.238***		1.373***	1.379***
		(0.0334)	(0.0322)		(0.0332)	(0.0326)
ESCS[23]	2.641***	1.558***	1.568***	3.230***	1.955***	1.948***
	(0.0738)	(0.0596)	(0.0570)	(0.0732)	(0.0590)	(0.0574)
I Gen. migrant	-2.212***	-2.119***	-1.870***	-4.656***	-4.700***	-4.581***
	(0.224)	(0.250)	(0.241)	(0.226)	(0.253)	(0.248)
II Gen. migrant	-1.579***	-1.597***	-1.499***	-2.823***	-2.912***	-2.918***
	(0.182)	(0.203)	(0.196)	(0.183)	(0.204)	(0.200)
Early enrolled	-1.280***	-1.769***	-0.408*	-1.322***	-1.849***	-0.864***
	(0.264)	(0.278)	(0.244)	(0.255)	(0.266)	(0.237)
Late Enrolled	-4.293***	-4.227***	-3.889***	-5.426***	-5.407***	-5.078***
	(0.230)	(0.252)	(0.247)	(0.234)	(0.256)	(0.252)
Male (class share)		-0.505**	-0.341		-0.0856	-0.0452
		(0.249)	(0.280)		(0.245)	(0.282)
I Gen. migrant (class share)		-2.706***	-2.839***		-2.806***	-1.934***
		(0.583)	(0.659)		(0.572)	(0.654)
II Gen. migrant (class share)		-1.584***	-2.639***		-1.018***	-2.036***
		(0.389)	(0.477)		(0.381)	(0.479)
HISCED	1.646***			1.810***		
	(0.0482)			(0.0481)		
HISEI	-0.602***			-0.902***		
	(0.0580)			(0.0577)		
Constant	59.25***	54.59***	53.21***	60.61***	54.86***	53.59***
	(0.351)	(0.304)	(0.339)	(0.349)	(0.301)	(0.344)
Observations	338,497	307,200	307,200	331,693	301,258	301,258
R-squared	0.104	0.107	0.231	0.128	0.131	0.225
Province FE	YES	YES	YES	YES	YES	YES
School FE	NO	NO	YES	NO	NO	YES

Notes: This table presents regression results for a simple linear model studying the relation between several individual characteristics and test scores in grade 5. All regressions include 27 Province fixed effects. Robust standard errors are clustered at the school level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

[19] Occupational status of the father
[20] Occupational status of the mother
[21] Educational level of the mother
[22] Educational level of the father
[23] Index of individual economic, social and cultural status

Chapter 2: Out of the darkness: Re-allocation of confiscated real estate mafia assets

2.1 Introduction

Urban areas are often characterised by pockets of poverty, crime, and marginalisation (Rosenthal & Ross, 2015). In light of that, addressing urban deprivation by means of effective regeneration measures represents a key challenge for policymakers (Bailey & Robertson, 1997). In particular, crucial objectives for interventions aimed at fostering the overall quality of cities - especially in underprivileged neighbourhoods - include tackling criminal activities and improving public spaces and housing (Atkinson & Helms, 2007; Koster & van Ommeren, 2019). Yet, evidence on the effectiveness of urban renewal policies of this kind is very limited.

This paper focuses on a large-scale, nationwide policy intended to reach the double goal of opposing organised crime and also contributing to the revitalisation of local urban areas. The Italian law allows to seize any real estate asset previously owned by organised crime members or affiliates and, through re-allocations, re-assign these assets to local communities by converting them into public housing amenities (e.g. centres for disadvantaged groups, green spaces, police stations). The intention of re-allocations, as conceived by the Italian legislator, is to contribute to the development and revitalisation of local areas in which they are made. As such, this measure not only acts as a crucial device for the appropriation of relevant resources from criminal activities, but also allows its redistribution to local communities. In this way, it contributes to eradicate criminal organisations in the areas where they are most rooted and prevent their spreading in territories selected by criminal groups for investment and money laundering, while also providing new opportunities and facilities to the residents of neighbourhoods plagued by the mafia. The buildings re-assigned to the citizenry, in their new role, should stimulate the development of a 'culture of legality', favour local entrepreneurship, and help recovering disadvantaged people from their conditions (Falcone et al., 2016).

While some descriptive and anecdotal evidence exists on the use and application of the policy (Camera dei Deputati, 2009; 2018; Transcrime, 2013; Falcone et al., 2016), this evidence tells little on its actual effectiveness. When confiscated assets are discussed in the media, the monetary value of the assets is systematically presented (e.g. Repubblica 2019), but other local effects - let alone overall capitalisation effects - are seldom considered. In spite of the fact that policies to recover organised crime assets are widely diffused in several countries across the world[24], the re-allocation measure adopted by the Italian State has thus far been ignored by the academic literature. Whether and how real estate asset confiscations and re-allocations have had an impact on the wider society has not yet been investigated.

In this paper, we aim to fill this gap and investigate whether the re-allocation of confiscated mafia real estate assets produces any external effects on the local territory where such initiative takes place. Following the literature evaluating the impact of urban renewal policies, we capture the spillover effects of the intervention by looking at how the monetary value of buildings in the surrounding of confiscated and re-allocated assets responds to the implementation of the policy. The evidence produced by previous studies assessing the external effects of regeneration policy measures is mixed. While some works reveal that localised investments to revitalise urban areas have converted into higher house prices of neighbouring buildings (Santiago, Galster, & Tatian, 2001; Schwartz et al., 2006; Rossi-Hansberg, Sarte, & Owens, 2010; Ooi and Le, 2013; Koster & van Ommeren, 2019), others find that they have no effect on the property value of surrounding areas (Lee et al., 1999; Ahlfeldt et al., 2017). It is worth noticing that almost all these studies focus on specific neighbourhood of single cities where the programme has been implemented, producing hardly generalizable findings[25]. In contrast to that approach, we perform our analysis on the entire Italian territory, thus focusing on a very large and highly heterogeneous context. Hence, the main contribution of our work relates to the peculiarity of the intervention we examine, aiming to improve neighbourhoods by both increasing the stock of amenities and by tackling organised crime, as well as to the size and the spatial scale of the policy initiative.

Furthermore, our analysis is based on a unique database which allows to better identify the policy's impact. We exploit detailed information on the exact location and timing of over 15,000 confiscated and re-allocated properties in Italy and investigate their spillover effect in two ways adopting difference-in-differences empirical settings. First, we develop a panel model estimating how micro-aggregated local housing markets across the entire Italian territory respond to real estate asset confiscation and re-allocation. Second, exploiting information on over 60,000 geo-localised house sale points in the 55 major Italian cities, we provide a close examination of the impact of re-allocations on the housing value of neighbouring buildings, as well as a detailed investigation of the spatial decay of the estimated effect. These two empirical strategies complement each other. The first covers the whole of the Italian territory, focuses on a longer time period (2005-2016) but is relatively less geographically accurate, given that it is based on local area data. The second presents a more limited temporal span (2011-2017) but the precision and accuracy of the analysis are higher due to the use of georeferenced real estate data as units of observation, and to the possibility of accounting for a very large set of buildings characteristics as controls. This setting allows to minimise any issue of selection by local area/building

---

[24] According to the Asset Recovery Office of the European Commission, organised crime assets worth over 4 billion euros have been recovered in Europe alone in 2014 (the last year for which data is available) (ARO, 2014). Of this amount, over 1.6 bn was recovered in Italy.

[25] The only exception if the recent contribution by Koster and van Ommeren (2019), estimating the external benefits of a programme improving the quality of public housing in 83 deprived neighbourhoods throughout the Netherlands.

characteristics, as well as to control for any potentially confounding housing market dynamics. In addition, detailed information on confiscated / re-allocated assets make it possible to separately identify the two aspects of the policy (i.e. the confiscation and the re-allocation).

The results reveal strong and robust evidence of an external effect of re-allocations on neighbouring properties, noticeably increasing their value following the conversion of confiscated buildings into new amenities. This finding, consistent across estimation methodologies, reveals that for every building re-allocated in a territory, surrounding properties increase their monetary value between 0.4% and 0.8%. Examining the temporal dynamics of the effect by means of an event study, we show that the impact materialises the year following the re-allocation. Additionally, we demonstrate that this effect decays with distance and becomes insignificant 350m from the re-allocated building. When sub-dividing our sample into mafia and non-mafia regions, we show that the effect is driven by mafia-rigged areas, where the majority of confiscations and re-allocations take place. Furthermore, the impact appears stronger in deprived neighbourhoods. This suggests that the legislator's intent to improve the quality of neighbourhoods where the mafia presence is more pronounced may be effective.

A number of channels may be driving the uncovered effect. On the one hand, property values are directly influenced by the stock of amenities of the kind of those chosen for the re-allocations. A higher provision of green spaces, cultural facilities, social engagement centres, and similar buildings are expected to positively affect the monetary value of the neighbourhood they are in (Gibbons & Machin, 2008; Gibbons et al., 2014). On the other hand, the confiscation / re-allocation policy can influence disamenities such as the level of violence and crime, whose reduction also increases property prices (Gibbons, 2004, Linden & Rockoff, 2008, Ihlanfeldt & Mayock, 2010). Other mechanisms which may be triggered by the policy have to do with housing market dynamics, i.e. variations in the supply of real estate properties. An increase (decrease) in housing supply would reduce (increase) house prices (Glaeser et al., 2005; Caldera & Johansson, 2013).

Our research, in addition to contributing to the literature on urban renewal policy evaluation, adds up to the growing studies on organised crime (e.g. Acemoglu et al., 2013; Barone & Narciso, 2015; Pinotti, 2015; Alesina et al., 2018; Di Cataldo & Mastrorocco, 2018; Pinotti & Stanig, 2018) and, more specifically, to the recently developing literature studying the societal implications of public initiatives against criminal organisations. The most widely analysed policy is the Italian law allowing the dissolution of local governments upon clear evidence of links between mafia clans and local public officials. Acconcia et al. (2014) exploit the temporary contraction in public investment occurring in post-dissolution periods to obtain estimates of the fiscal multiplier for Italian provinces. Daniele and Geys (2015) and Galletta (2017) demonstrate that dissolutions affect the quality of elected politicians and the proportion of public investments in neighbouring municipalities. Another examined policy is the accomplice-witnesses regulation. Acconcia et al. (2009) show the policy to be more effective the less efficient the prosecution system and the higher the internal cohesion of mafia organisations, while Garoupa (2007) analyses the policy within a principal-agent theoretical environment. No study has yet looked at confiscations and re-allocations of mafia real estate assets as we do in our paper.

The remainder of the paper is organised as follows. Section 2.2 describes the legislative measures we aim to evaluate, providing some key descriptive statistics. Section 2.3 presents our data. Section 2.4 introduces our empirical strategy at local housing market (OMI zone) and sale-point (micro) levels. Section 2.5 presents our findings. Section 2.6 concludes.

2.2 Institutional background: confiscation and re-allocation of mafia assets

The rise in mafia activities throughout the 1980s and a series of violent attacks led the Italian government to introduce a set of tougher anti-mafia measures. On 13 September 1982, in the aftermath of the murders of politician Pio La Torre and anti-mafia prefect Carlo Alberto Dalla Chiesa in Palermo, the national Parliament approved the 'Rognoni-La Torre' law (646/82), which represented a turning point in the fight against organised crime. This bill introduced two key measures fighting mafia activities, namely the inclusion in the Penal Code of membership of a mafia-type criminal organisation as a crime independent of other criminal acts (so-called 416-bis article), and the possibility for the courts to confiscate any asset belonging to members of the criminal associations, as well as to relatives, partners and other subjects who in the previous five years played a cover-up role for criminal organisations. Any individual condemned with article 416-bis would immediately get their assets seized. The seizure may be converted into confiscation by the judges. To make law enforcement quick and effective, the law granted the judiciary full access to bank records in order to follow money trails.

The 'Rognoni-La Torre' law (646/1982) prescribes four steps to obtain the final confiscation:

1. The properties of suspects of belonging to mafia groups are scrutinised by the competent tribunal;
2. The seizure is decided upon by a panel of 3 judges. The asset goes under judiciary administration;
3. The judges provide a motivation for confiscation. The asset goes under first degree confiscation;
4. If appealed, the confiscation decision is reviewed by the Court of Appeal. The order can be 'revocation'[26] or confirmation (second degree confiscation).

The possibility of confiscating mafia-related goods and properties represents an extremely powerful tool in the hands of the Italian State in its fight against criminal organisations. Real estate asset confiscation is nowadays recognised as a fundamental instrument contributing to eradicate the pervasive presence of the mafia in the areas where it is most deeply rooted (Dalla Chiesa et al., 2016; Falcone et al., 2016). This is because real estate properties have a strong symbolic meaning for criminal groups. They are a physical representation of their power on the local territory and are often chosen by mafia families for their meetings. In addition, considering the large share of liquidity laundered by mafia groups into real estate properties - more than 50%

[26] Of all the confiscated buildings, only 14 have been 'revoked'. This suggests that judge bribing, even if taking place, is ineffective and plays little role as a confounder of our analysis

of illegal mafia profits are reinvested into the legal economy, with real estate as one of the preferred sectors of investment (Transcrime, 2015) - the confiscation policy is a way to harm their business model and earnings.

A fundamental step in the management procedure of seized assets is their re-allocation to a new use by 'returning them to the citizenry' (Frigerio & Pati, 2007). This is operated by the Italian State, after the confiscation has been completed. The procedure of re-allocation, already introduced in the 646/82 law, has been regulated more clearly in 1996, when law 109/96 has been promulgated. As can be seen in Figure 2. 1, the number of re-allocations has increased drastically in the aftermath of the approval of the 1996 law, and the large majority of re-allocations have occurred in the last few years.

Figure 2. 2 illustrates the geographical location of re-allocated properties across the Italian national territory. The confiscated and re-allocated mafia assets seem to be concentrated in metropolitan urban areas. Clusters can be observed in cities such as Milan, Rome, Naples, Reggio-Calabria and Palermo. A concentration of assets also seems to emerge in Southern Italian cities, with fewer clusters in Northern cities and even less in the central regions of Italy. The regions of Sicily, Puglia, Calabria and Campania also present higher concentrations of confiscated assets, which comes as no surprise given the publicised presence of mafia in these regions.

The approval of the 1996 law on re-allocation was the result of lobbying activity from the anti-mafia association *Libera*, asking for a faster management of confiscated assets and the possibility to use re-allocated goods for social purposes. As a result of that, the law lists a whole set of different uses for the re-allocated assets. The two broader categories are: 'social use' and 'institutional, justice and public order' (Figure 2. 3). The former category includes conversions of buildings into: anti-mafia/non-for-profit associations, senior centres, under18 centres, disable centres, health care centres, sport centres, green spaces. The latter includes: tribunal, police station, centre for migrants, archive, council houses. The logic of the policy is to use re-allocated assets to establish the principle of legality precisely where the control of the mafia is most entrenched, for example with the creation of police stations. Alternatively, buildings re-allocated for social use (e.g. by creating centres for employment-seekers) may contribute to provide concrete alternatives for individuals potentially attracted by organised crime. In all cases, the main principle behind this measure is the possibility for re-allocated assets to contribute to the regeneration of a local area and/or to become a fundamental resource in the fight against criminal organisations.

2.2.1 Local areas of confiscation/re-allocation

By exploiting 2011 Census data, it is possible to descriptively examine the characteristics of the areas where confiscated and re-allocated buildings are located. In order to do that, we construct a dataset at the level of Census areas for the entire Italian territory and focus on micro-areas with 100 or more inhabitants. For each of them, we are able to say whether there have been confiscations/re-allocations. We subsequently test for the correlation between a treatment dummy (taking value 1 if in a given Census area there has been at least one episode of confiscation and re-allocation, and 0 otherwise) and a number of Census characteristics. The results of this test are reported in Table A2. 4 in the Appendix. As visible in the table, territories where the policy has been applied are relatively smaller in size, as shown by the negative association between the treatment dummy and the log population variable, and they have a higher proportion of unemployed people, of families renting their house, and buildings in bad conditions. All in all, this evidence seems to suggest that, as hypothesised, the policy is most often being implemented in underprivileged territories[27]. Crucially, when we replicate the empirical test with the inclusion of local housing market fixed effects (OMI FE), it can be noted that none of the Census variables returns a significant coefficient, indicating that area characteristics are balanced in this case.

2.2.2 Re-allocation timing

The implementation of law 109/96 and the creation in 2010 of a National Authority for Mafia-Confiscated Assets (hereafter ANBSC) has contributed to speed up the application of the law, progressively increasing the number of confiscated real estate assets being re-allocated. Yet, the average time between confiscation and re-allocation has been of over 8 years even after 1996, with only 83 properties in total being re-allocated in the same year or the year following the confiscation, as visible in Table A2. 1. The average length of the re-allocation procedure is sharply varying across the national territory, as illustrated in Figure A2. 1 in the Appendix, with no clear identifiable geographical pattern. Table A2. 3, reporting the count and share of re-allocations by political colour of local governments over the 1998-2017 period, suggests that the length of the re-allocation procedures is unrelated with the political colour of the municipal government where the asset is located. The proportion of buildings taking either less than 10 years or 10 years or more to re-allocate is almost the same for each government type[28].

Next, we examine how the length of re-allocation procedure correlates with the characteristics of local areas and the type of building being assigned to a new use. Table A2. 4 reports the results of an exercise testing for the correlation between the duration of re-allocation procedures, computed as the difference between the year of re-allocation and the year of confiscation, and a number of variables measured either at the Census level or at the level of re-allocated building.

[27] In Italy, approximately 72% of the houses are owned by residents. As a result, being rented is often a condition of more disadvantaged families balanced in this case.

[28] Comparing column (4) with column (2) of Table A2. 3, it also appears that re-allocations occur less than proportionally under governments run by civic lists - i.e. politicians with no clear ideological affiliation - than in governments ruled by left-wing, right-wing, or centre governments. As a consequence, it appears important to account for the political colour of the local governments in our analysis, which we do as we control for any municipality time-varying characteristics by means of municipality-year fixed effects.

The correlation between these variables and the length of re-allocations is estimated first by accounting for re-allocation year fixed effects, then including local housing market (OMI) fixed effects. Table A2. 4 illustrates that re-allocations tend to take longer in territories with higher unemployment, i.e. in more deprived territories where it may be presumed that courts are relatively less efficient. However, as fixed effects are included in the model, none of the local characteristics emerges as significantly associated with the policy implementation timing. Furthermore, re-allocations take generally longer for buildings assigned to institutional use, while they take less time for buildings assigned to social use. Again, this correlation disappears with the inclusion of fixed effects in the model[29].

2.2.3 Heterogeneity of re-allocated assets and policy impact

Mafia organisations generally own both operational and economic assets. The former are critical resources to exercise sovereignty over their market, whereas the latter are investments and money laundering machines. Operational assets such as real estate properties serve both as inputs for the illicit activities, insurance system against detection for the family of the members of the organisation and institutional signals for the entire community. The different role played by these assets suggests to differentiate between different regions where they are located. Another important source of heterogeneity is linked to the different types of re-allocation.

The policy's impact can materialise in various ways. First, the confiscation/re-allocation may be weakening criminal organisations. Asset seizure and confiscation might have a direct effect on the mafia's economic power, and act as a deterrent reducing *ex ante* its size[30]. In addition, the policy might be particularly effective when complemented with plea-bargaining and other forms of amnesty for the agents, since it counterbalances the potential savings in labour cost for the organisation, with a higher punishment in case of detection for the employer. Decreasing the welfare of individuals linked by strong ties to the mafia could constitute a relevant deterrent even in absence of detection. Moreover, the simple confiscation could have *per se* an effect on the perception of impunity that often characterise criminal organisations. A weaker presence of criminal groups is expected to materialise into a higher value of buildings in the area where confiscations take place (Gibbons 2004, Linden & Rockoff 2008, Ihlanfeldt & Mayock 2010).

Second, the re-allocation measure could serve as an extraordinary engagement device for the local community (Falcone et al., 2016). Non-profit organisations could use assets located in critical areas to organise bottom-up initiatives and sustain institutional change. This process may contribute to the revitalisation of the targeted areas also through the attraction of competitive

---

[29] This exercise has been reproduced also by including fixed effects for local Court instead of OMI fixed effects, obtaining very similar results.

[30] This second dynamic is consistent with the model proposed by Garoupa (2007), where a higher punishment for the employer fosters a decrease in the number of agents and in information diffusion.

firms and skilled workers (Storper & Venables, 2004). All this would capitalise into higher house prices in the neighbourhood area.

Third, the involvement of local authorities in the decision process regarding the allocation of the real estate asset could counterbalance the influence exerted by the criminal organisation. A corrupted politician might selfishly decide to support the fight against the organisation, in order to obtain political support thanks to the provision of a new public asset.
All these sources of heterogeneity are investigated in section 2.4.3.

2.3 Data

The empirical analysis relies on a novel dataset constructed from a wide range of sources. First, data on confiscated and re-allocated real estate assets have been extracted from the National Agency for the Administration and Destination of Seized and Confiscated Assets from Organised Crime (ANBSC). This includes detailed information on all the 15,698 re-allocated buildings on the whole Italian territory with their full address, the date of confiscation and re-allocation, the type of building and of re-allocation, the local court responsible for completing the procedure, the administrative entity responsible to manage the building once re-allocated. Each asset has been correctly geolocalised. Of these buildings, a relatively small portion is sold on the housing market (746) or demolished (2). These assets are dropped from our sample.

In the first part of our analysis, we use housing transaction data at a micro-aggregated zone level (Osservatorio del Mercato Immobiliare, or OMI), a spatial division of the Italian territory defined by the Italian Revenue Agency. OMI zones are smaller than neighbourhoods and correspond to functional local housing markets, i.e. homogeneous real estate markets for similar property types[31]. The dataset spans from 2006 to 2016. For each OMI zone of Italy and for each real estate asset typology, the dataset includes maximum and minimum selling prices of properties. Following (Manzoli et al.), we compute simple averages between the minimum and the maximum price for each OMI zone – asset type. The values, computed for each quarter, are subsequently averaged at the year level. Within each OMI, the square deviation is usually lower than 1.5. OMI areas are drawn at the infra-municipality level, based on similar socio-economic and urban characteristics, building infrastructures and quality. All these features are crucial to determine prices[32] (Budiakivska & Casolaro, 2018).

We decide not to exploit all the information of the OMI dataset and to consider the value of prices only for the most representative categories, i.e. civil properties in normal state of conservation which are usually private residential buildings (excluding chalet, villas and boxes). We retain over 38,000 OMI zones per year from 2005 to 2016, 1718 of which have had at least one episode of re-allocation over the analysed period. Figure 2. 4 zooms into three major Italian cities, Milan, Naples, and Rome, to show their OMI zones and re-allocations.

The second part of our analysis exploits 53,728 geo-localised house sale points, spanning from 2011 to 2017 and collected from Immobiliare.it, the biggest Italian real estate website. These data are based on real estate properties sold in the 55 major Italian cities[33], with homogeneous

[31] According to the National Real Estate Agency, OMI areas are defined as: '*a continuous portion of the municipal area that reflects a homogeneous section of the local real estate market, where there are uniform prices for similar economic and socio-environmental conditions. This uniformity is translated into homogeneity in the positional, urban, historical-environmental, socio-economic characteristics of the settlements, as well as in the provision of services and urban infrastructure*'. In each OMI zone, the lowest unitary market value recorded for each building type should not be lower than 50% of the value recorded the most expensive asset in the same category.

[32] The prices reported in the OMI dataset are obtained from various sources, principally the analysis of actual prices specified in administrative archives or quoted by market operators. In cases of missing observations, the data is integrated with assessments of local experts aimed at correcting imperfections or attributing a reference price whenever the low number of transactions limits the representativeness of the reported values.

[33] These are: Alessandria, Ancona, Aosta, ,Ascoli Piceno, Bari, Bergamo, Bologna, Bolzano, Brescia, Cagliari, Campobasso, Caserta, Catania, Catanzaro, Cosenza, Firenze, Foggia, Genova, Isernia, La Spezia, L'Aquila, Latina, Livorno, Matera, Messina, Milano, Modena, Monza, Napoli, Novara, Nuoro, Padova, Palermo, Parma, Perugia, Pesaro, Pescara,

coverage of the website across different cities as shown in Figure 2. 5. The dataset does not provide actual selling prices but asking prices that we use as proxies for the actual transaction prices[34]. The files have been then compiled, cleaned and checked for duplicates through the website unique identifier for each ad[35]. Finally, some of the missing values were filled by using the textual description of the ads. A recent paper by Loberto et al. (2018) which focuses on the comparison between Immobiliare.it data and OMI data provided by the real estate market observatory of the Italian Tax Office, found the Immobiliare.it data provides an appropriate picture of the Italian housing market, consistent with official sources.

The micro-level dataset includes a wide range of structural attributes including floor space (m2), building height, type of property (studio, apartment, house, villa), the number bedrooms and bathrooms, floor, the date of construction, garage or parking facility and the type of heating an energy consumption.

In addition, a long list of controls is collected from the Italian census (2011), the Italian National Geoportal of the Environment, the Real Estate Observatory of the *Agenzia del Territorio* (AT), the Ministry of Education and Open Street Map. These include a series of controls for pre-existing amenities (i.e. already in place before re-allocations) such as typology of buildings on the street of the asset, distance to a range of natural and commercial amenities, distance to parking and transport controls, as well as the locations of schools (see Table A2. 5). Labour market, education, real estate quality and demographic data collected for the 2011 Italian Census were also obtained from the Italian Institute of Statistics (ISTAT). Descriptive statistics for treatment and control variables are reported in the Appendix (Table A2. 6, A2.7, and A2.8).

Pordenone, Potenza, Prato, Reggio di Calabria, Roma, Salerno, Sassari, Savona, Taranto, Teramo, Terni, Torino, Trento, Trieste, Udine, Venezia, Verona, Viterbo.

[34] Following Loberto et al. (2018), we assume that the removal of the ad corresponds to the sale of the property

[35] When a change of price was tracked, the final most conservative price was recorded.

2.4 Empirical Strategy

In order to correctly estimate the effect of the confiscation and re-allocation of Mafia assets, we develop two complementary empirical strategies. First, we focus on the longitudinal trends of local homogeneous housing markets (OMI), exploiting the 2005-2016 time period and considering the entire Italian territory. This difference-in-differences strategy allows us to first test for a significant policy effect on micro-aggregated local housing markets. Next, we perform our analysis at the level of sale point, further testing for the spillover effect of the policy on house prices, capturing the spatial decay of the estimated effect and investigating the heterogeneous treatment effect. This hedonic pricing model is estimated as a repeated cross-sectional difference-in-differences.

2.4.1 OMI areas

First, we analyse the effect of confiscation / re-allocation policies on property prices aggregated at the OMI area level. Average values are computed starting from the minimum and maximum market values per zone to obtain average euro / m^2 house prices.

In order to test for the effect of confiscation and re-allocation of real estate assets on house prices, we rely on a differences-in-differences panel model accounting for the timing of confiscation and re-allocation of one or more properties in each OMI zone.

The estimated model is as follows:

$$lnp_{jt} = \alpha C_{jt} + \beta R_{jt} + \sum_{k=1}^{n} \gamma_k X_{jkt} + \delta_j + \lambda_t + e_{jt} \tag{1}$$

Where lnp_{jt}, the natural logarithm of average housing prices per square meter in OMI j and year t, is a function of a different set of variables. The two key variables in the model are the treatment variable C_{jt}, switching on for OMI j in the year(s) when confiscation(s) took place, until the moment of the re-allocation, and the treatment variable R_{jt} switching on from the moment in which a confiscated property has been re-assigned to a new use until the end of the sample period. As per our hypotheses, we expect a general increase in house prices in 'treated' OMI areas during the post-re-allocation period. This model captures the extensive margin effect of confiscations / re-allocations.

To control for different sources of heterogeneity in the sample, we exploit time-variant variables (X_{jkt}) retrieved from the 2011 Italian Census. We control for the number of properties in each area, the status of the buildings and other socio-economic conditions of the household living there (unemployment, level of education). In all specifications, we include time (λ_t) and OMI (δ_j) fixed effects. Year dummies allow to control for significant sudden generalised shocks in the Italian housing market, while OMI dummies account for any time-invariant factors at the level of local housing markets[36]. Furthermore, standard errors are clustered at the level of municipality, so to correct for the presence of spatial autocorrelation. The model is estimated for the 2005-2016 period.

In order to isolate the effect of confiscations and re-allocations, we focus exclusively on OMI zones having experienced only *one* episode of confiscation(s) or re-allocation(s) in time. That is, we exclude all OMI zones where confiscations and re-allocations have occurred over multiple years. The single episodes of treatment we consider may involve more than one single building confiscated/re-allocated if the confiscations/re-allocations of buildings in that OMI area have been established all in the same moment. To minimise the effects of confiscations on re-allocations, we test our findings by excluding all OMI zones where the re-allocation took less than 10 years to be completed.

2.4.2 Sale-point analysis

In our main specification, we estimate a hedonic pricing model using micro geo-localised data at the level of sold building. Although this is considered the ideal approach in the hedonic literature, few studies have used this strategy to explore the impact of public policies as punctually localised as the one under consideration in this paper. Moreover, our dataset is novel in terms of size and spatial detail for the Italian territory. In line with other policy evaluations (e.g. Ahlfeldt et al., 2017), our first assumption lies in expecting a very localised effect of confiscated assets on surrounding real estates.

Using geographic information system (GIS), we begin by drawing perimeters up to 500m radii around each of the re-allocated assets. These buffers roughly correspond to an average 5 minutes walking distance from the real estate asset, spatially translating the expected local effect (EVSTUDIO, 2016; Gibbons & Machin, 2008). The buffers of 500m represent the maximum extent to which we expect to measure a local effect. Given the punctuality of the policy, we in fact expect externalities to be more localised, with radii varying between 100m to 500m from confiscated/re-allocated assets[37].

Figure 2. 7 provides an illustration of our approach. All sale points with no assets in the buffer zone act as controls, while sale points located in the same OMI area, with at least one confiscated asset within their buffer radius act as treated units. We drop from the sample any observation with no confiscated asset within 2km distance, excluding in this way the large majority of OMI areas with no treated units. Exploiting information on each building's sale date, we can exploit the timing of the re-allocation and identify the impact of the policy on the prices of buildings

---

[36] Adopting OMI zones as our unit of analysis allows to minimise unobserved heterogeneity potentially confounding our estimates, given that these geographical units correspond to functional local housing markets.

[37] In choosing our buffer radii we follow the literature on the evaluation of the spillover effects of urban renewal policies (i.a. Linden & Rockoff 2008; Schwartz et al., 2006; Rossi-Hansberg et al., 2010; Ahlfeldt et al., 2017)

inside the buffer and being sold *after* the re-allocation took place. This method allows us a highly accurate focus on the neighbourhood of the confiscated and re-allocated asset, identifying with precision the treatment area.

To compute the external impact of the confiscated and re-allocated real estate assets we estimate the following hedonic pricing model:

$$\ln p_{ijmt} = \beta_1 C_{i,t-n}(d) + \beta_1 R_{i,t-n}(d) + \beta_3 C_{i,t+n}(d) + \beta_4 R_{i,t+n}(d) + \rho X_i + \delta_j + \theta_{mt} + e_{ijmt} \quad (2)$$

where $\ln p_{ijmt}$ is the natural logarithm of house price per m^2 of real estate property i in OMI zone j, municipality m, sold in year t. $C_{i,t-n}$ is a treatment indicator, defined as number of buildings confiscated within a radius d from building i in year t-n (n=1,2,3) before it was sold. Similarly, $R_{i,t-n}$ is a treatment indicator defined as the number of buildings re-allocated within distance d from building i in year t-n. The two treatment variables capture the intensive margin effect of confiscations and re-allocations on house prices of neighbouring buildings.

The variables $C_{i,t+n}$ and $R_{i,t+n}$ (n=0,1) are post-treatment covariates, that allow us to account for pre-treatment differences in housing prices. X_i is a vector of structural and amenity controls of property *i*, the latter which were constructed from multiple geographical datasets for all the Italian territory and e_{ijmt} is the error term for property *i*. We compute distances to a large range of amenities as specified in the data section (including distance to city CBD) to account for omitted variable bias. We also control for socio-economic conditions by census tract from the 2011 Italian Census. Although our temporal dimension is shorter than for our OMI analysis, we control for local time-invariant factors and for common shocks, adopting OMI zone (δ_j) and municipality-year (θ_{mt}) fixed effects. The model is estimated for the 2011-2019 period, for every distance d = {50, 100, 150, 200, 250, 300, 350, 400, 450, 500}. Standard errors are clustered at the OMI zone level so to correct for the presence of spatial autocorrelation.

This research design allows to separate the effect on property values of confiscation or re-assignment of real estate assets from correlated location effects (Koster et al., 2012; Noonan & Krupka, 2011).

2.4.3 Estimation issues

In order to correctly identify the effect of confiscation/re-allocations on housing prices, a number of estimation issues need to be addressed.

First, we need to consider potential problems of selection. According to Transcrime (2017), mafia organisations tend to invest more often in territories they control. If housing prices in these areas have peculiar trends for reasons not associated with the analysed policy, our results may be biased.

Second, the application of the policy may depend on the quality of public institutions. In areas where public authorities are more likely to be captured by criminal organisations through bribes and/or where the re-allocation procedure takes more time to be completed, we expect a lower density of seized (and re-allocated) assets. Figure A2. 1 in the Appendix shows no clear geographical/regional pattern in relation to the efficiency of local courts responsible for re-allocations, suggesting that court efficiency is semi-random. Re-allocation procedures exhibit a high degree of heterogeneity, with no clear differences in the average duration between Northern and Southern Italian regions. However, Table A2. 3 shows some evidence that the duration of the re-allocation procedure may vary depending on the political colour of the local government administrating the municipality where the asset is located.

In order to deal with these issues, we include a number of controls in our models. To start with, we always include OMI zone fixed effects in the estimates. As mentioned above, OMI are micro-geographical areas, smaller than neighbourhoods, characterised by homogeneous real estate markets. Areas are revealed at the infra-municipality level, sharing similar socio-economic and urban characteristics, building infrastructures and quality, namely the features which are crucial to determine house prices (Budiakivska & Casolaro, 2018). In Table A2. 2, we exploit data retrieved from the 2011 Italian Census to test the balancing properties of our setting on a number of local area characteristics, finding no significant difference within OMI areas (when OMI fixed effects are controlled for), confirming the homogeneity of these geographical units.

As a further test for that, we also verify if OMI areas can be considered as homogeneous units for less 'tangible' characteristics such as social capital[38]. To study the endowment of social capital within OMI areas we follow Putnam's (1993) seminal contribution and more recent literature (Peri, 2004; Guiso et al., 2004; Buonanno et al., 2009) and exploit variation in voter turnout within OMI areas as a proxy for civic engagement. We are able to measure this variable at the level of polling station in the four largest Italian cities: Rome, Milan, Naples, and Palermo, which are also those with most confiscated and re-allocated assets (see Figure 2. 2). To minimise any distortion of electoral competition from organised crime (more common for elections held at the municipal level) we focus on the 2009 European Elections[39]. Our assumption is that differences in voter

[38] Scholars sub-divide social capital into bridging and bonding, the former referring to linkages between different groups in society, while the latter referring to strong ties within the same groups. In Southern Italy, a lower level of bridging and an excess of bonding social capital has been connected with the activity of criminal organisations (Trigilia, 2001)

[39] European Elections are known to be hardly influenced by criminal organisations, due to the size of electoral constituencies (the Italian territory is divided in 5 macro-constituency). Moreover, in contrast to mayors and

turnout between treated polling areas and areas where no confiscation is recorded would undermine our claim of institutional homogeneity of neighbourhoods within OMI areas. The results shown in Table A2. 9 unveil a negative association between voter turnout and re-allocations, which however becomes insignificant when OMI fixed effects are accounted for, thus again confirming the homogeneity of OMI zones.

In addition to OMI fixed effects, our hedonic models control for Census area characteristics, further minimising any potential confounder within OMI areas. Moreover, the specifications account for generalised shocks in housing markets by means of year fixed effects, as well as for any municipality-specific characteristics varying over time with municipality-year fixed effects. The latter control allows to account for any change in the political composition of the local government potentially influencing the timing of the policy and its implementation. To conclude, the very large set of control variables at the level of building - including a number of variables identifying pre-existing amenities - further minimises the possibility that any observed policy effect is due to non-random characteristics of the local area where the policy is put in place.

Finally, we include in our model two pre-treatment variables, measuring the assets confiscated the same year of the transaction and the following year[40]. In this way, we test for pre-treatment differences in housing prices between treatment and control groups.

Another possible issue relates to the fact that our study focuses on a policy being implemented in two steps: first the confiscation, and then the re-allocation. In order to minimise any possible effect of confiscations on re-allocations, our analysis focuses on re-allocations taking ten years or more to be completed. The 'double' treatment may give rise to one additional concern, namely the fact that the confiscation affects other outcomes such as labour mobility. To minimise this issue, we test the impact of the policy within a very limited distance from the treatment site, as low as 150m, where the probability of any labour/firm relocation is unlikely to be more concentrated than in the outer ring.

Municipality/regional councillors, Members of the European Parliament do not have the power to affect the allocation of funds at the local level.

[40] According to Frigerio and Pati (2009) and Transcrime (2017), the large majority of assets become operative between 6 months and 18 months after the reallocation time. Fort this reason, we do not expect any treatment effects in the treatment year.

2.5 Results

2.5.1 OMI-level analysis

We begin by performing the analysis at the level of OMI areas, focusing on the whole Italian territory and relying on a panel dataset between 2005 and 2016. The OMI dataset includes information on house prices - our dependent variable - for a large variety of real estate properties. In order to obtain comparable observations and minimise heterogeneity, we perform our estimates by focusing on the monetary value of the most common type of property in Italy, i.e. civic houses , further restricting the analysis to those whose quality status is classified as 'normal' by the Italian land registry. While this strategy marginally reduces the number of OMI areas in the sample, it prevents differences in property prices to be driven by the diverse composition of buildings in a given area.

We restrict our analysis to OMI zones having experienced confiscation and/or re-allocation events only once over the full period of implementation of the policy (1982 to date). In other words, we exclude from the sample all local areas having experienced multiple episodes of confiscation/re-allocation. The results of the difference-in-differences analysis are reported in Table 2. 1.

We begin by testing the relationship between confiscation and property prices. The first specification in column (1) only includes the treatment variable accounting for whether an OMI zone has experienced a confiscation of one or more real estate assets at any point in time during 2005-2016. This variable switches on in the year of confiscation until the moment of the re-allocation. In column (2) we exclude all re-allocation years from the analysis. In both cases, the coefficient is not statistically significant, suggesting that house prices have not varied significantly in the aftermath of a confiscation episode.

Next, we test the effect of re-allocation on OMI zones house prices. In column (3) we include the treatment variable for re-allocation, switching on at the time of the re-allocation episode in the OMI zone. This specification considers all re-allocated buildings, regardless of the time it took to re-allocate them, while in column (4) we focus our attention only on re-allocation that took 10 or more years to be completed. Finally, in column (5) we include both treatment variables at the same time. It can be seen that in all cases the estimates return a positive and strongly significant coefficient, indicating that the selling price of houses within OMI areas in which the re-allocation took place increased in the aftermath of the re-allocation. In our favourite specification, a reallocation of a single asset is associated with a 4.2% increase in property prices in the OMI area.

It must be noted that, as all the sold re-allocated buildings are dropped from our sample, these estimates are testing the effect of real estate assets which are appropriated and managed by public institutions (mainly municipalities). Therefore, the observed increase in value in the OMI zones is due to a higher price of the buildings in the same local housing market of the re-allocated one(s).

In Figure 2. 8 we examine the timing of the estimated re-allocation effect. We perform an event study (Angrist & Pishke, 2008) by including a full set of leads and lags dummy variables for the entire period before the treatment year and during the treatment, using the year before the re-allocation as reference category[41]. As before, the sample is restricted to OMI zones having experienced only one re-allocation in time. The figure reports the coefficients for each year pre/during treatment with 90% confidence intervals, providing further evidence on a positive and significant effect of the re-allocation event. In all years before the re-allocation, there is no significant difference in house prices between treated OMI zones (i.e. those in which real estate assets will be re-allocated) and other OMI zones, as all coefficients specifically referring to years prior to the re-allocation are not statistically different from zero. The significant difference in prices emerges in the following years, already visible in the first post-treatment year.

2.5.2 Sale-point analysis

Having shown some evidence of a significant re-allocation effect of the value of buildings surrounding those re-allocated, we further examine this relationship with micro-level data. Table 2. 2, A2.10, and A2.11 report the results for the hedonic analysis conducted at the sale point level, using different radii to define the treatment area.

Results for the model estimated at a distance threshold of 250m are reported in Table 2. 2. The first specification in column (1) includes structural controls and OMI/year fixed effects only. It can be seen that the estimate returns a positive and significant coefficient one and three years after the treatment kicks in. Results are consistent in column (2)-(4), where we progressively add building, pre-existing amenity and socio-economic controls. No significant difference between treatment and control groups are recorded in the treatment year and before. It must be noted that, as no information is available on the exact period of the year when each property is re-allocated, re-allocations in t_0 might happen prior to the housing sale event. As a result, it is not surprising to find no significant result at t_0, consistently with the event study in Figure 2. 8. Overall, the regression results suggest positive and lasting effect of the re-allocation policy.

In column (5) we extend the specification to include municipality-year FE, in order to control for city-level exogenous shocks. Doing that, we implicitly rule out any municipal-level treatment effect. While this strategy is expected not to affect results concerning the largest cities in our sample, medium-size urban areas might still record an overall benefit from the policy. Nevertheless, results appear consistent with previous estimates. The coefficient in column (5) is positive and significant in the year following the treatment and in the third year after the treatment. Once identified the time trend in the event study, we estimate the overall effect of the policy. Column (6) only includes a cumulative treatment proxy, corresponding to the sum of the neighbouring assets re-allocated over the 3-year period. Finally, in column (7) we include in our specification a similar proxy for confiscated assets. Once again, the estimates report a positive

[41] While this implies including dummies up to 11 years before and during the treatment, the reliability of estimated coefficients reduces for years far away from the start of the treatment, as the number of observations for each year is inevitably lower.

and significant coefficient for re-allocations, while insignificant for confiscations. Overall, the findings are consistent with the existence of a positive externality arising from the reallocation of confiscated assets. For each asset confiscated in the previous 3 years, neighbouring property prices are expected to rise by 0.4%. Although the results generally confirm the dynamic found in the OMI analysis for confiscation and re-allocation events, the magnitude is significantly lower. This difference is probably due to the specific features of the two empirical strategies. The OMI analysis focuses on OMI areas where only one confiscation/reallocation event took place. By contrast, the sale-point analysis allows for multiple treatment. The difference might be partially explained with decreasing returns of confiscated/reallocated assets. Moreover, the second strategy guarantees a more precise identification of the policy effect. Controlling for property observable characteristics and unobservable time-invariant area characteristics, we better identify the effect of the confiscation policy on the properties located in the immediate neighbourhood around the seized assets. Positive effects on property prices taking place beyond the distance threshold chosen are likely to determine a downward bias in our estimates.

Despite the fact that a proper cost-benefit analysis is beyond the scope of this study, it is possible to discuss the magnitude of the policy effect. To our knowledge, this is the first study to investigate the impact of confiscation policies on property prices. As a result, it is not immediate benchmark to compare our results with. However, our result can be compared with studies analysing the effect of crime at a similar spatial scale. Thaler (1978) finds that a one standard deviation increase in the incidence of property crime reduces home values by about 3 percent. A more significant effect is reported by Gibbons (2004), that finds a standard deviation decrease in local density of criminal damage to be associated with a 10% price increase in the average Inner London property.

Our results can then be analysed in relation to studies investigating the effect of local amenities on property prices. Machin (2011) reviews 11 studies investigating the nexus between school quality and housing prices, finding a median change of 4% in housing prices following a standard deviation change in school quality. Similarly, the presence of sex offenders reduce property prices by 2-4% (Linden and Rockoff, 2008; Pope, 2008). On the other hand, changes in toxic emissions from industrial plants is associated with a 10% change in house price (Currie et al., 2015).

With respect to different amenities, our estimates appear to be significantly lower. However, the policy considered is likely to be significantly cheaper for local authorities. Moreover, the strategic position of confiscated assets, mostly located in deprived neighbourhood, is such that the policy is likely to particularly benefit deprived social groups.

In Table A2. 10 in the Appendix we report regression results for the hedonic micro-level model estimated within a radius of 150m, that we consider the minimum area of analysis, on the basis of our sample size and the related literature (e.g. Rossi-Hansberg et al., 2010). The basic specification in column (1) reports positive and significant coefficients for the third year following the treatment. The results are generally confirmed in magnitude and significance while adding to the specification the full set of housing sale level controls. While estimating the cumulative treatment in column (6) and (7), the treatment coefficient is higher than the one estimated with a

250m radius. Overall, at 150m distance, we again find evidence of a positive effect of the re-allocation policy.

This result, obtained with such a small distance from the treatment point, allows to further minimise any potential concern of endogeneity due to the presence of time-varying confounding factors at the OMI level. If, for instance, the confiscation has activated some dynamics we are not explicitly accounting for in the model (e.g. related to labour mobility), this may bias our estimates. However, the likelihood that these dynamics are stronger within the 150 metres from the treatment sites than in the rest of the OMI area is extremely low.

In Table A2. 11 we investigate treatment effects at 500m radius. Columns (1) to (5) report results for our main specification. Overall, we find some evidence of a positive effects in the years after the treatment. The coefficients exhibit a lower magnitude with respects to the one estimated in the 250m distance specification. Consistently, the coefficient is lower than with other threshold distances when considering the cumulative treatment and non-significant. Conversely, we do find some evidence of a negative effect of confiscations on house prices.

In order to investigate the distance decay of the policy, in Figure 2. 9 we combine together the estimated coefficients from 100 to 500m, with relative confidence intervals, controlling for confiscation and all other set of controls and fixed effects. The Figure allows to appreciate the spatial decay characterising the cumulative treatment. The coefficients are monotonically decreasing, with a larger standard error up to 150m due to the lower sample size. Overall, the policy is found to have a positive and significant effect up to 350m. At a radius of 350m the policy still has a positive effect, but the declining coefficient suggest the transactions localised further than the 300m threshold to be less affected. At 400m distance the coefficient is still positive, but no longer significant.

In Table A2. 12 , we test the robustness of these results by including in the model a control for the buffer zone. If there are time-invariant characteristics which are specifically located at a 100m to 500m distance from the re-allocated real estate asset and have an influence on house prices, this would act as an omitted variable and bias our estimates. The specification including a buffer zone dummy variable is fully controlling for that. The inclusion of this control leaves the main results virtually unaltered, as the re-allocation retains significance and the magnitude of the coefficients is lower as we move away from the treatment point. Interestingly, the buffer zone dummy is statistically insignificant up until 300m from the re-allocated asset, suggesting no generalised difference in house prices in the treatment areas vis-à-vis the untreated area within OMI zones.

2.5.3 Where is the effect stronger?

To conclude our analysis, we further characterise the estimated external impact of re-allocations on the value of surrounding real estate buildings by testing *where* is this impact stronger.

To begin with, the results commented above suggest that the impact of re-allocations on property prices is larger the higher the number of re-allocations - i.e. in presence of a higher density of re-allocated buildings. Hence, we may expect that the policy would produce the larger impacts in areas where organised crime groups are more rooted and where they invest the most. While we cannot directly measure the presence of organised crime, their strongholds are well known. Campania, Calabria, Sicily and Puglia are the regions in which Italian criminal groups have their roots[42] (Transcrime, 2013). More generally, criminal groups tend to prosper in more deprived areas, where public institutions are often perceived as weak and distant, the provision of public services is sometimes deficient, and employment opportunities are lower.
To test this hypothesis, we exploit the geographical extension of our dataset. We sub-divide our sample into regions of high mafia intensity (Campania, Calabria, Puglia and Sicily) and all remaining regions. The results, shown in Table 2. 3, indicate that the effect we obtain appears to be driven by the regions where organised crime has a stronger presence. As shown in Figure 2. 2, these regions are also those where the majority of re-allocations have been made.

Finally, we estimate the model by focusing on specific areas, selected on the basis of their characteristics. In particular, we attempt capture the degree of urban deprivation by means of two indicators related to labour market conditions and real estate characteristics of the area. We begin by sub-dividing the sample among OMI with average recorded unemployment rate above and below the 75th percentile of the national distribution (Census data). As visible in Table 2. 4, the effect of re-allocations on house prices only appears for areas with higher unemployment levels. Next, we sub-divide the sample according to the proportion of buildings classified as in 'bad' conditions, again relying on Census 2011 information. Once again, the estimated effect appears to be driven by the most deprived and disadvantaged neighbourhoods.

[42] While organised crime is spread across the entire Italian territory (and beyond), it still maintains its strongest presence in the areas where it was originally formed. According to Transcrime (2013), the Cosa Nostra (Sicily), 'Ndrangheta (Calabria), Camorra (Campania) and Sacra Corona Unita (Puglia) preserve their strongholds in their regions of origin. The cities in our sample belonging to the four regions of high mafia intensity are: Bari, Foggia, Taranto (Puglia); Napoli, Caserta, Salerno (Campania); Catanzaro, Cosenza, Reggio di Calabria (Calabria); Palermo, Messina, Catania (Sicily)

2.5.4 Channels

In this study, we presented evidence that confiscation and reallocation policies can be powerful tools to regenerate deprive neighbourhoods. However, we have not discussed what are the mechanisms driving the increase in housing prices. In part, the capitalisation of re-allocations into higher house prices of surrounding buildings may be due to a safer environment, 'cleaner' from the activity of criminal organisations. This kind of dynamic would be consistent with the fact that a stronger effect is visible in mafia-rigged regions, where the larger proportion of mafia investment into real estate are made (Transcrime, 2013). The effect we obtain may also be the result of the improved view of a previously more deprived and less attractive neighbourhood, thanks to the new amenities. This explanation is linked to the fact that the majority of re-allocations take place in local areas characterised by a high share of buildings in bad state, and that the effect of the policy is stronger in more disadvantaged areas.

In absence of detailed geocoded data, we are not able to investigate in depth the underlying channels. However, in this section we run a simple exercise that could provide at least some indications regarding the mechanisms that explain our results. We exploit 2013-2018 annual reports produced by the DIA, the Anti-Mafia Investigation Directorate, that provide very detailed information on the territories under the influence of mafia organisation in 5 Southern cities (Naples, Reggio Calabria, Palermo, Messina). In particular, the DIA maps the power exerted by each single mafia family on the territory. The data are updated every year and make it possible to follow the evolution of mafia presence in small neighbourhoods and even in single streets (see Figure 2. 11).

Thanks to these data, we have constructed a street-level panel dataset on organised crime presence in Naples (see Figure 2. 12). Exploiting DIA data, we estimate the following model:

$$Camorra_{st} = \alpha C_{st} + \beta R_{st} + \sigma_s + \lambda_t + \delta_{zt} + e_{st} \quad (3)$$

where $Camorra_{st}$ is the number of camorra families active in street s, year t, or a dummy for camorra activity, C_{st} is confiscation dummy and R_{st} is the treatment dummy. The confiscation variable switches on from the year of confiscation(s) taking place in street s to the year before the re-allocation, while the re-allocation variable switches on from the year of confiscation(s) taking place in street s to the year before the re-allocation. The specification includes time-invariant street-specific factors (σ_s), time shocks (λ_t) and OMI-year FE (δ_{zt}). Due to the high heterogeneity recorded in street-level data, we only focus on a 100/200 meters radius from each road.

This empirical strategy can be interpreted in two ways. The analysis of the effect of confiscation on Mafia presence could provide some evidence on the actual effectiveness of the State in eradicating criminal organisations. However, criminal organisation activities are likely to be affected by unobservable factors correlated with public enforcement.

On the other hand, assuming the timing of reallocation to be exogenous with respect to mafia activity on the territory, the re-allocation of mafia assets could be seen as a 'deterrent technology'[43]. The signal given to local communities by the re-allocation to a new use of an asset previously associated with the power of the organisation on the territory, as well as the new services provided by the asset itself, could both increase the costs associated with the control of the territory. Considering the strong assumptions required to claim causality in such a framework, we perform it as a descriptive exercise to test our predictions that part of the effect of the policy on property price is due to changes in organised crime behaviour.

In Table 2. 5, we regress the number of family operating in one street over the cumulative re-allocation variable, using a 100 meters radius. In the baseline specification (column (1)), confiscation is found to have a positive effect on the number of families per road, whereas the opposite effect is found for re-allocation events. When OMI-year FE are included (column (2)), the confiscation event becomes insignificant, while the coefficient for re-allocations, although losing magnitude, remains negative and significant. In Column (3) we focus only on treated roads, exploiting variation over time in the treatment variable. Results are consistent with the previous specification. Overall, confiscation is found to have an ambiguous effect on the influence of mafia families on Naples roads, while a significant and negative effect is found for the re-allocation of former mafia assets. I columns (4)-(6) we estimate the same specification using a 200 meters radius. Results are confirmed, but the coefficient magnitudes for re-allocation significantly declines.

Results suggest that re-allocation policies could decrease mafia activity in the neighbourhood where the reallocation takes place. However, the number of family could be a poor proxy for the actual power of Mafia on a territory. In Table 2. 6 we run a similar exercise, but this time we use a simple dummy that takes value of 1 if mafia presence is recorded in the street and 0 otherwise. The previous results are generally confirmed. Confiscation has a positive or no effect on the presence of mafia, while a significant negative effect is found following the re-allocation of confiscated assets. In this case however, when we focus on treated roads only, we find a positive and significant effect of confiscation while the negative effect of re-allocation events become less significant.

This exercise, conducted on a single city, does not allow us to draw conclusions regarding the effect of this policy on mafia activity. However, it supports our hypothesis that at least part of the regeneration effect of the re-allocation policy is obtained with the eradication of the pervasive presence of the mafia in the treated areas. On the other hand, the positive effect of the confiscation event on Mafia presence could be explained with the clan wars that often follow the conviction of important members of the organisation.

[43] Due to the lack of detailed geocoded data on deterrent policies and crime activities, most papers in the literature only focus on measures of expenditure in each policy. Two relevant exceptions are Draca et al. (2011) and Bell et al. (2014).

2.6 Conclusions

In an effort to tackle criminal organisations, the Italian State allows for the possibility to seize and confiscate real estate properties previously belonging to mafia groups. Such policy, widely considered as one of the most crucial tools to undermine the power of organised crime in local areas, entails the re-allocation of confiscated assets to a new use, supposedly contributing to the revitalisation of the territory in which this policy intervention takes place.

This paper assesses the extent to which re-allocations contribute to such regeneration process by testing their external effects on the monetary value of properties in the surrounding areas. Our estimates, performed at different geographical units of analysis and making use of unique micro-level datasets, unveil a robust positive relationship between re-allocation cases and the property price of neighbouring buildings. The increase is equal to 0.5%-0.8% per each re-allocated building, lasting a minimum of three years following the re-allocation.

This finding suggests that, as hypothesised (and as expected by the Italian legislator), re-allocations lead to significant spillover effects that add value to the whole territory where they are implemented. Such effect is visible in the range of up to 350m from each episode of re-allocation. The impact is stronger in more deprived neighbourhoods and in regions characterised by a stronger presence of criminal organisations.

With the available data, we are not currently able to investigate to what extent the observed effect is due to the eradication of the presence of criminal organisations or to a simple amenity effect. However, the exercise conducted at the street level suggests that at least a part of the effect could be associated with a reduction of the depressive effect of mafia activity on local economy.

In all cases, what emerges with clarity from our study is that the policy of re-allocating real estate assets recovered from criminal organisations has the important effect of increasing the value of local neighbourhoods where such buildings are located. The policy we have assessed is not explicitly characterised as 'place-based' in nature, in the sense that it is not specifically intended for poor neighbourhoods, but rather can be implemented in both more and less developed areas. Nevertheless, we have shown that its primary application has been in local areas characterised by high unemployment and more unattractive buildings. Furthermore, its effect is noticeably larger in cities where the presence of organised crime is stronger. Hence, this suggest that an effective and rapid implementation of the re-allocation policy may favour the revitalisation of urban areas at higher disadvantage where mafia groups hold the upper hand.

References

Acconcia, A. Giancarlo Corsetti, and Simonelli, S. (2014). Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment. *American Economic Review* 104.7, 2185-2209.

Acconcia, A. and Patrick Rey (2009). Accomplice Witnesses, Organized Crime and Corruption: Theory and Evidence from Italy. *Scandinavian Journal of Economics* 116, pp. 1116-1159.

Acemoglu, D. James A. Robinson, and Rafael J. Santos (2013). The monopoly of violence: Evidence from Colombia. *Journal of the European Economic Association* 11, pp. 5-44.

Ahlfeldt, G.M., S. Redding, D. Sturm, N. Wolf (2015). The Economics of Density: Evidence From the Berlin Wall *Econometrica* 83, pp. 2127-2189.

Ahlfeldt, G.M., Wolfgang Maennig, and Felix J. Richter (2017). Urban renewal after the Berlin wall: A place-based policy evaluation. *Journal of Economic Geography* 17, pp. 129-156.

Ahlfeldt, G.M. and Nancy Holman. (2018). Distinctively Different: A New Approach to Valuing Architectural Amenities. *Economic Journal* 128, pp. 1-33.

Alesina, A., Salvatore Piccolo, and Paolo Pinotti. (2017) Organized Crime, Violence, and Politics. *Review of Economic Studies*.

Angrist, J. and Jorn-Steffen Pischke (2008). Mostly Harmless Econometrics: An Empiricist's Companion. *Princeton University Press*

Atkinson, R. and Gesa Helms (2007). Securing an urban renaissance: Crime, community, and British urban policy. *Policy Press*

Bailey, N. and Robertson, D. (1997). Housing renewal, urban policy and Gentrification. *Urban Studies*.

Barone, G. and Narciso, G. (2015). The effect of mafia on public transfers. *Journal of Urban Economics*.

Bell, B., Jaitman, L., & Machin, S. (2014). Crime deterrence: Evidence from the London 2011 riots. *The Economic Journal*, 124(576), 480-506.

Budiakivska, V. and Casolaro, V. (2018). Please in my back yard: the private and public benefits of a new tram line in Florence. *Temi di Discussione, Bank of Italy* 1161.

Buonanno, P. Daniel Montolio, and Paolo Vanin (2009). Does Social Capital Reduce Crime? *Journal of Law Economics* 52.1 , pp. 145-170.

Caldera, A. and Asa Johansson. (2013). The price responsiveness of housing supply in OECD countries. *Journal of Housing Economics* 22, pp. 231-249

Dalla Chiesa, Nando (2016). Il riuso sociale dei beni confiscati. Le criticità del modello lombardo. *Rivista di Studi e Ricerche sulla criminalità organizzata* 2.2: 15-25.

Daniele, G. and Benny Geys. (2015) Organised Crime, Institutions and Political Quality: Empirical Evidence from Italian Municipalities. *Economic Journal* 125, pp. F233-F255.

Di Cataldo, M. and Nicola Mastrorocco. (2018) Organised Crime, Captured Politicians and the Allocation of Public Resources. *SSRN Electronic Journal* 1556-5068.

Draca, M., Machin, S., & Witt, R. (2011). Panic on the streets of London: Police, crime, and the july 2005 terror attacks. *American Economic Review*, 101(5), 2157-81.

Falcone C.R., Giannone T. e Iandolo F. (2016). BeneItalia. Economia, welfare, cultura, etica: la generazione di valori nell'uso sociale dei beni confiscati alle mafie. *Edizioni Gruppo Abele*

Frigerio, L., and Pati, D. (2007). L'uso sociale dei beni confiscati–Programma di formazione sull'utilizzazione e la gestione dei beni confiscati alla criminalità organizzata. *Libera–Associazioni, nomi e numeri contro le mafie, Roma*, 60

Galletta, S. (2017) Law enforcement, municipal budgets and spillover effects: Evidence from a quasi-experiment in Italy. *Journal of Urban Economics* 101, pp. 90-105.

Garoupa, N. (2007). Optimal law enforcement and criminal organization *Journal of Economic Behavior and Organization* 63.3, pp. 461-474.

Gibbons, S. (2004). The costs of urban property crime. *Economic Journal* 114.

Gibbons, S. and Stephen Machin. (2008). Valuing school quality, better transport, and lower crime: Evidence from house prices. *Oxford Review of Economic Policy* 24.1, pp. 99-119.

Gibbons, S., Susana Mourato, and Guilherme M. Resende. (2014) The Amenity Value of English Nature: A Hedonic Price Approach. *Environmental and Resource Economics* 57, pp. 175-196.

Glaeser, E., Joseph Gyourko, and Raven E. Saks. (2005). Why have housing prices gone up? *American Economic Review* 95, pp. 329-333.

Guiso, L., Sapienza, P., and Zingales, L. (2004). The role of social capital in financial development. *American economic review*, *94*(3), 526-556.

Ihlanfeldt, K. and Tom Mayock. (2010). Panel data estimates of the effects of different types of crime on housing prices. *Regional Science and Urban Economics* 40.2-3, pp. 161-172.

Koster, H. R., van Ommeren, J., and Rietveld, P. (2012). Bombs, boundaries and buildings: A regression-discontinuity approach to measure costs of housing supply restrictions. *Regional Science and Urban Economics*, *42*(4), 631-641.

Koster, H. R., and Van Ommeren, J. (2019). Place-based policies and the housing market. *Review of Economics and Statistics*, *101*(3), 400-414.

Lee, P. and Alan Murie. (1999). Spatial and Social Divisions within British Cities: Beyond Residualisation. *Housing Studies* 14, pp. 625-640.

Linden, L. and Jonah E. Rockoff. (2008). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *American Economic Review* 98:3, 1103-1127

Loberto, M. A. L., and Pangallo, M.. (2018). The potential of big housing data: an application to the Italian real-estate market. *Temi di discussione (Economic working papers) 1171. Bank of Italy.*

Manzoli, Elisabetta, and Sauro Mocetti (2019). The house price gradient: evidence from Italian cities. *Italian Economic Journal*: 1-25

Noonan, D. S., & Krupka, D. J. (2011). Making—or picking—winners: Evidence of internal and external price effects in historic preservation policies. *Real Estate Economics*, 39(2), 379-407.

Ooi, J. and Thao T.T. Le. (2013). The spillover effects of infill developments on local housing prices. *Regional Science and Urban Economics* 43.6, pp. 850-861.

Peri, Giovanni. (2004). Socio-Cultural Variables and Economic Success: Evidence from Italian Provinces 1951-1991. *The B.E. Journal of Macroeconomics* 4 pp. 1-36.

Pinotti, P. (2015). The Economic Costs of Organised Crime: Evidence from Southern Italy. *Economic Journal* 125.586 (2015), F203-F232.

Pinotti, P. and Piero Stanig. (2018) Sowing the Mafia: A Natural Experiment Bocconi, mimeo.

Rosenthal, S. S., Ross, S. L. (2015) Change and persistence in the economic status of neighborhoods and cities. In J. V. Henderson, G. Duranton and W. C. Strange (eds) *Handbook of Regional and Urban Economics*, Vol. 5, pp. 1047-1120. Elsevier.

Rossi-Hansberg, E., Sarte, P. D., and Owens III, R. (2010). Housing externalities. *Journal of political Economy*, *118*(3), 485-535.

Santiago, A., George C Galster, and Tatian, P.. (2001.) Assessing the property value impacts of the dispersed subsidy housing program in Denver. *Journal of Policy Analysis and Management* 20.1, pp. 65-88.

Schwartz, A.E., I. Ellen, I. Voivu, M. Schill (2006). The external effects of place-based subsidized housing. *Regional Science and Urban Economics* 36, pp. 679-707.

Storper, M. and A. J. Venables. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography* 4, pp. 351-370.

Transcrime. (2011, 2013, 2015, 2017). Gli investimenti delle mafie, rapporto di ricerca.

Trigilia, C. (2001). Social capital and local development. *European journal of social theory*, 4(4), 427-442.

Tables

Table 2. 1: OMI-level estimates

Log euro per m^2	(1)	(2)	(3)	(4)	(5)
Confiscation	0.0180	0.0179			0.0159
	(0.0145)	(0.0138)			(0.0186)
Re-allocation			0.0374***	0.0421**	0.0421**
			(0.0142)	(0.0187)	(0.0186)
Census controls	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Re-all. time	Any	No re-all. years	Any	10+	10+
Observations	255,995	253,080	253,562	251,779	251,779
R-squared	0.965	0.965	0.965	0.965	0.966

Notes. The table reports the estimation results for the difference-in-difference model testing the relationship between confiscation/re-allocation and property prices (see Section 2.4.1). The dependent variable is the average price per m^2 recorded for private properties in each OMI area. The confiscation dummy switches on in the year of confiscation until the time of reallocation. Consistently, the re-allocation dummy equals one from the year of re-allocation onwards. In columns (1) and (3), the analysis covers the whole sample. In column (2) re-allocation years are excluded from sample. In columns (4)-(5) only re-allocations taking 10 or more years from confiscation are included. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 2. 2: Sale point analysis – d=250

Log euro per m²	Buffer radius: 250m						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 year before re-allocation	0.00249	0.00211	0.00188	0.00301	0.00212		
	(0.00319)	(0.00315)	(0.00315)	(0.00322)	(0.00332)		
Re-allocation year	-0.00085	-0.00107	-0.00119	-2.57e-05	-0.00261		
	(0.00313)	(0.00312)	(0.00312)	(0.00282)	(0.00294)		
1 year after re-allocation	0.0083***	0.0079***	0.0076***	0.0081***	0.00556**		
	(0.00260)	(0.00255)	(0.00263)	(0.00254)	(0.00241)		
2 years after re-allocation	0.00299	0.00271	0.00253	0.00331	0.00130		
	(0.00260)	(0.00256)	(0.00259)	(0.00236)	(0.00265)		
3 years after re-allocation	0.0071***	0.0064***	0.0063***	0.0064***	0.0058***		
	(0.00225)	(0.00220)	(0.00219)	(0.00181)	(0.00164)		
Re-allocation						0.00383**	0.00379**
						(0.00150)	(0.00151)
Confiscation							-0.00438
							(0.00276)
Structural controls	✓	✓	✓	✓	✓	✓	✓
Building controls		✓	✓	✓	✓	✓	✓
Amenity controls			✓	✓	✓	✓	✓
Socio-econ. controls				✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓	✓	✓
Municipality-year FE					✓	✓	✓
Observations	52,526	52,513	52,513	52,513	51,906	51,906	51,906
R-squared	0.768	0.769	0.771	0.777	0.784	0.784	0.784

Notes. The table reports the estimation results for hedonic analysis presented in Section 2.4.2. The dependent variable is the price recorded for each sale point i in year t. The main explanatory variables are the number of confiscation/reallocation events taking place n years before/after the transaction. Columns (1)-(5) report the effect of property re-allocation 1-3 years after the event. Housing prices differences recorded the same year or the year before make it possible to account for pre-treatment differences in housing prices. Columns (6) and (7) report the cumulative effect of confiscation and reallocation events on housing prices. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 2. 3: Sale point analysis – regional heterogeneity

Dep. variable: Log euro per m^2	Campania, Calabria, Puglia, Sicily			Other Italian regions		
	100m	250m	500m	100m	250m	500m
	(1)	(2)	(3)	(4)	(5)	(6)
Re-allocation	0.00718**	0.00427**	0.00137	-0.00985	0.00431	-0.00210
	(0.00329)	(0.00190)	(0.00106)	(0.00948)	(0.00776)	(0.00776)
Confiscation	0.00302	0.00155	0.00155	0.00691	-0.00604	-0.00833
	(0.00290)	(0.00185)	(0.00185)	(0.0340)	(0.0115)	(0.0068)
Structural controls	✓	✓	✓	✓	✓	✓
Building controls	✓	✓	✓	✓	✓	✓
Amenity controls	✓	✓	✓	✓	✓	✓
Socio-econ. controls	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓	✓
Municipality-year FE	✓	✓	✓	✓	✓	✓
Observations	11,891	11,891	11,891	40,015	40,015	40,015
R-squared	0.719	0.719	0.719	0.787	0.787	0.787

Notes. The table reports the estimation results for hedonic analysis presented in Section 2.4.2. The dependent variable is the price recorded for each sale point *i* in year *t*. The main explanatory variables are the number of confiscation/reallocation events taking place *n* years before/after the transaction. Columns (1)-(3) report the effect of property re-allocation 1-3 years after confiscation/re-allocation events in the regions where mafia organisations are more rooted. Column (4)-(6) report the estimates for the remaining territory. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 2. 4: Sale point analysis – local deprivation

Log euro per m^2	Buffer: 250m			
	Unemployment		Bad real estate conditions	
	High (>75th perc.)	Low (<75th perc.)	High (>75th perc.)	Low (<75th perc.)
	(1)	(2)	(3)	(4)
Re-allocation	0.00385**	0.00185	0.00519***	0.00270
	(0.00181)	(0.00181)	(0.00158)	(0.00353)
Confiscation	-0.000690	0.00184	-0.00354	-0.0176
	(0.00285)	(0.00674)	(0.00288)	(0.0324)
Structural / building / amenity controls	✓	✓	✓	✓
Socio-econ. controls	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓
Municipality-year FE	✓	✓	✓	✓
Observations	7,028	44,878	14,836	37,070
R-squared	0.601	0.769	0.794	0.772

Notes. The table reports the estimation results for hedonic analysis presented in Section 2.4.2. The dependent variable is the price recorded for each sale point *i* in year *t*. The main explanatory variables are the number of confiscation/reallocation events taking place *n* years before/after the transaction. Columns (1)-(3) report the effect of property re-allocation 1-3 years after confiscation/re-allocation events took place. The model is estimated using different samples, based on two proxies for local deprivation. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 2. 5: Street-level analysis: number of Mafia families

Dependent variable: number of Mafia families	100 m buffer			200 m buffer		
	(1)	(2)	(3)	(4)	(5)	(6)
confiscation	0.311***	-0.0989	0.0703	0.570***	0.0696	0.166**
	(0.0664)	(0.0613)	(0.0647)	(0.0502)	(0.0603)	(0.0681)
re-allocation	-0.583***	-0.153***	-0.364***	-0.435***	-0.0964***	-0.210***
	(0.0444)	(0.0301)	(0.0380)	(0.0318)	(0.0207)	(0.0252)
Observations	84,174	84,162	9,036	84,174	84,162	20,748
R-squared	0.849	0.942	0.928	0.849	0.942	0.933

Notes. The table reports the estimation results for linear regression model presented in Section 2.5.4. The dependent variable is the number of camorra families recorded in each road. In columns (1), (2), (4) and (5) the sample covers all Naples roads, whereas in column (3) and (6) it is restricted only to roads with confiscations/re-allocations. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

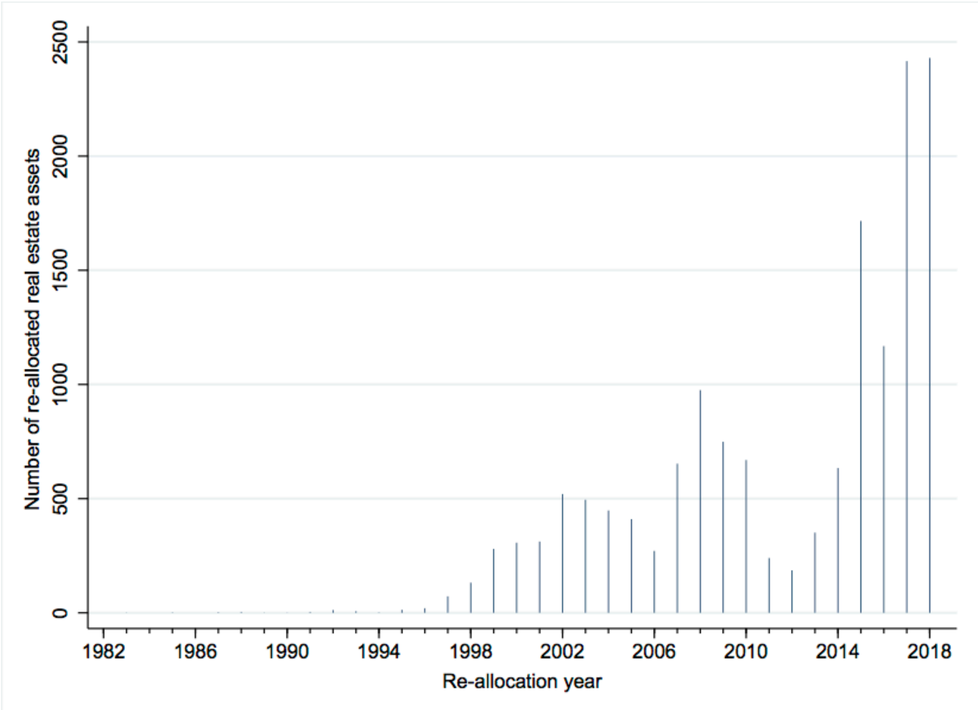
Table 2. 6: Street-level analysis: Mafia presence

Dependent variable: Mafia dummy	100 m buffer			200 m buffer		
	(1)	(2)	(3)	(4)	(5)	(6)
confiscation	0.0307***	0.00125	0.0187***	0.0386***	0.0208**	0.0362***
	(0.00370)	(0.00384)	(0.00692)	(0.00214)	(0.00822)	(0.0124)
re-allocation	-0.0182***	-0.0201***	-0.0135*	-0.0142***	-0.0152***	-0.00831*
	(0.00586)	(0.00538)	(0.00710)	(0.00437)	(0.00417)	(0.00483)
Observations	84,588	84,576	9,054	84,588	84,576	20,880
R-squared	0.874	0.944	0.936	0.874	0.944	0.935

Notes. The table reports the estimation results for linear regression model presented in Section 2.5.4. The dependent variable is a dummy equal to 1 if mafia activity is recorded in the street. In columns (1), (2), (4) and (5) the sample covers all Naples roads, whereas in column (3) and (6) it is restricted only to roads with confiscations/re-allocations. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

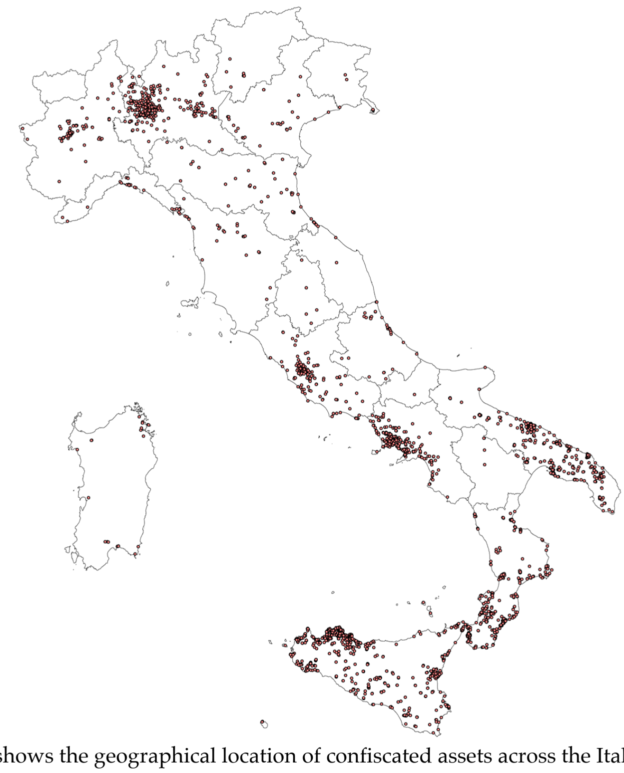
Figures

Figure 2. 1: Re-allocated real estate assets by year



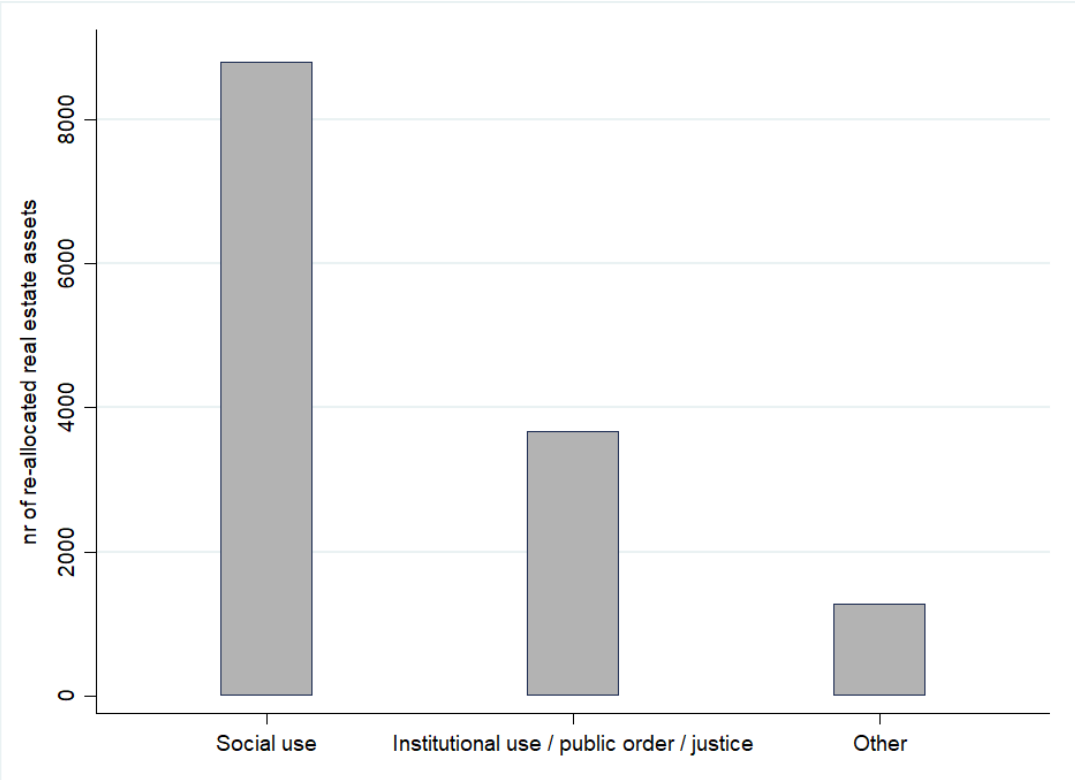
The figure shows the number of assets re-allocated by year, over the period 1982-2018.

Figure 2. 2: Re-allocations in Italy



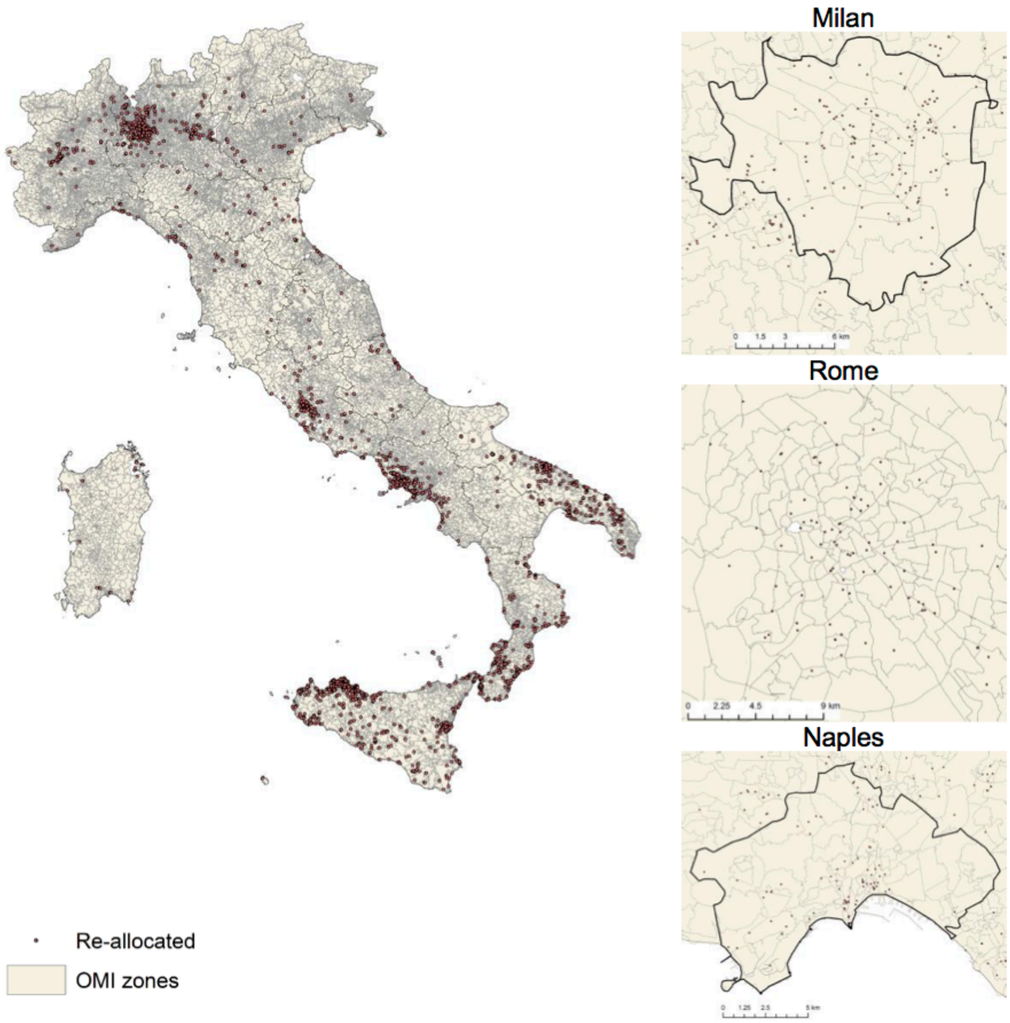
The figure shows the geographical location of confiscated assets across the Italian territory

Figure 2. 3: Re-allocation types



The figure shows the distribution of re-allocated assets by broader category.

Figure 2. 4: : OMI zones (Italy, Milan, Rome, Naples)

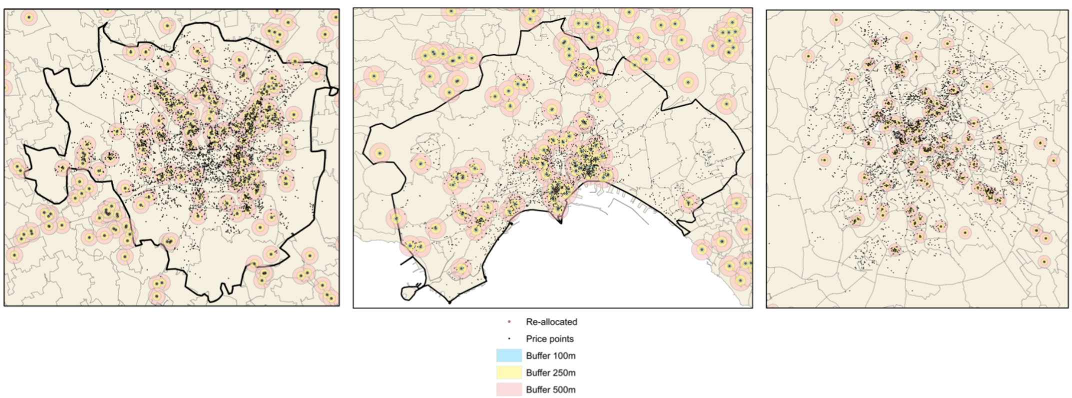


The figure shows re-allocated assets and OMI zones for the whole Italian territory and, in detail, for the three largest cities

Figure 2. 5: : Sale points in Italian cities

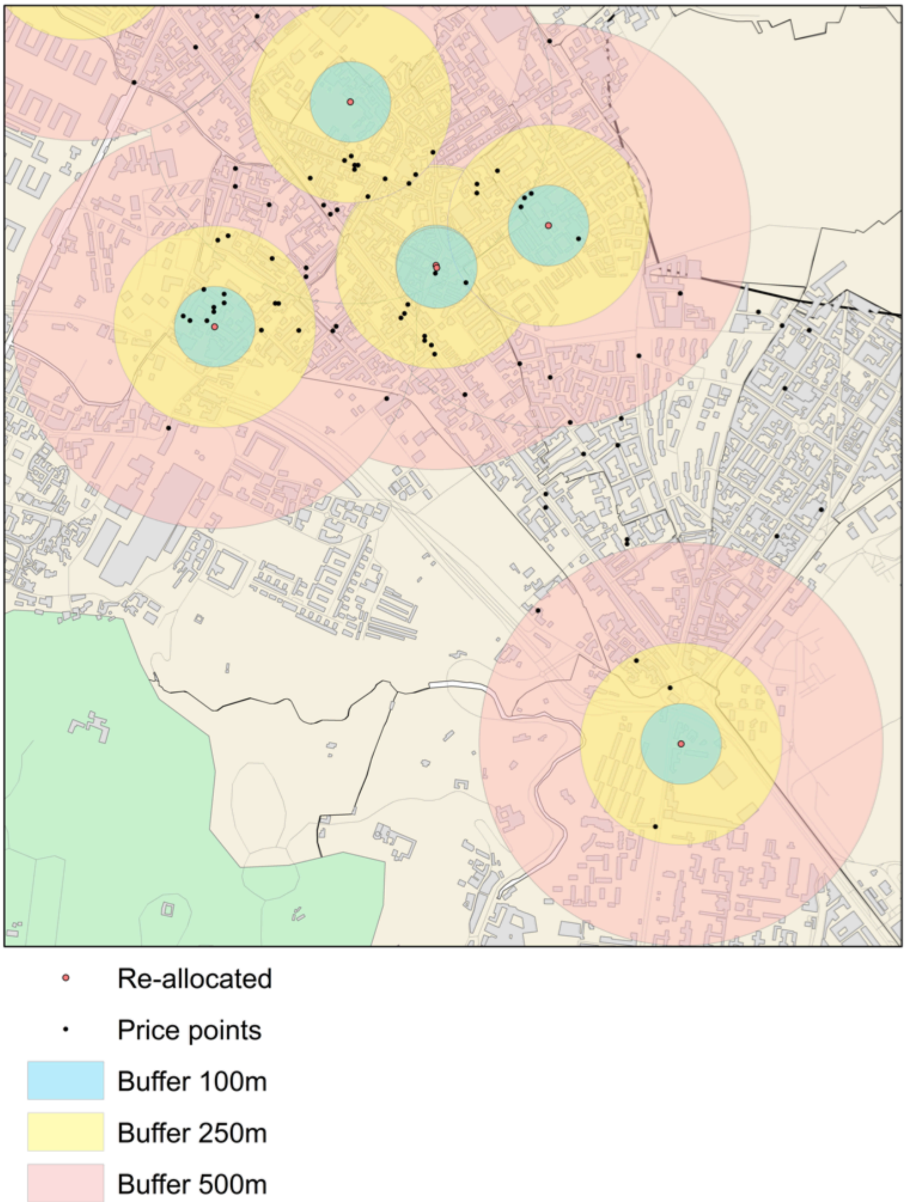
The figure shows the distribution of sale points across the 55 main Italian cities.

Figure 2. 6: : Sale points in major Italian cities



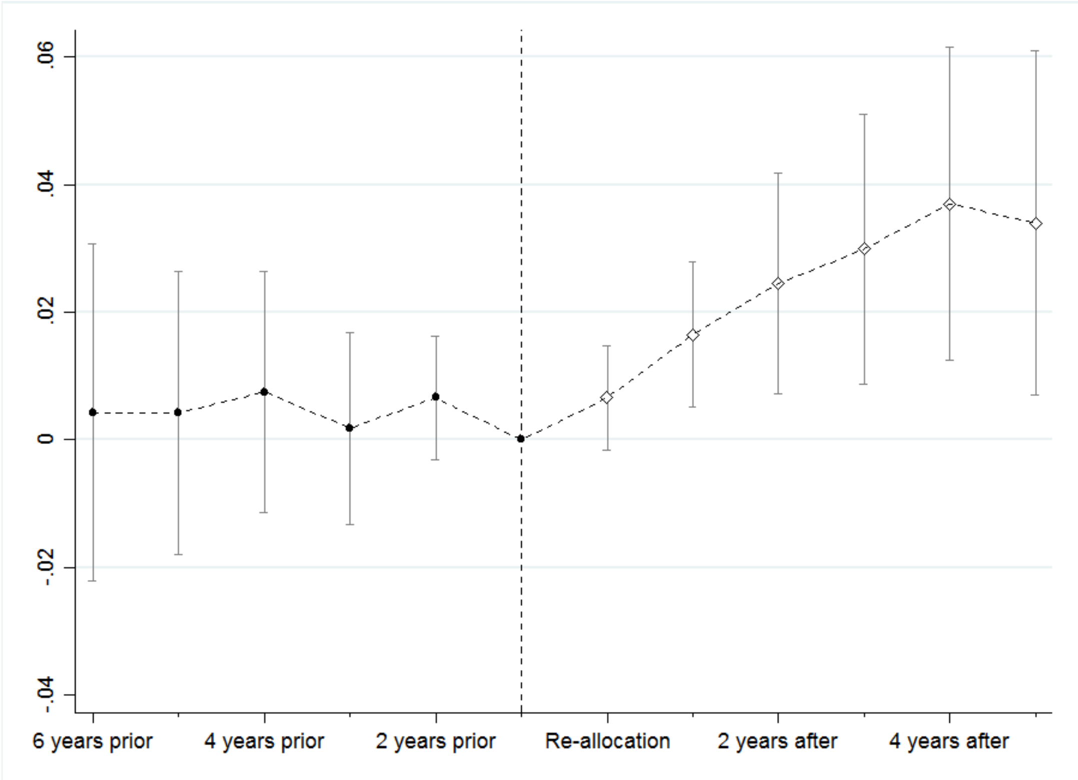
The figure shows 100m, 250m and 500m buffers around each sale point recorded in the three main Italian cities

Figure 2. 7: Buffer zones



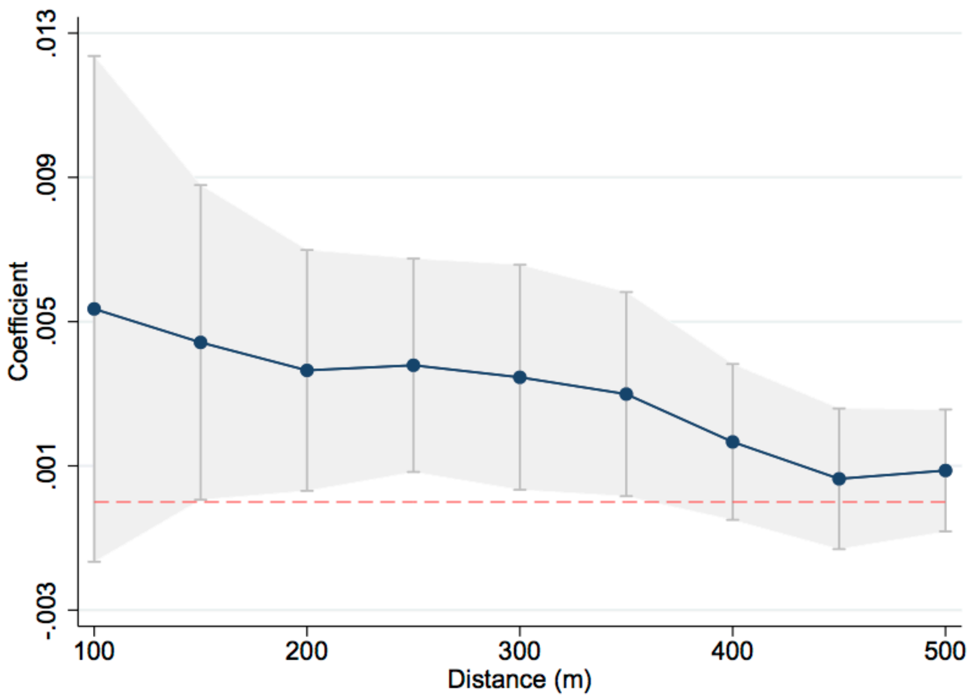
The figure provides a detailed representation of our methodology. Red dots represent re-allocated assets, black dots are sale points and blue, yellow and red areas represent, respectively the 100m, 250m and 500m buffer measured around each re-allocation. The point sales that in a given year fall inside a certain buffer *d* are said to be treated at distance *d*.

Figure 2. 8: Event study – re-allocation



The figure reports coefficients and confidence intervals for the event study conducted using OMI-level data.

Figure 2. 9: : Sale point analysis – distance decay



The figure reports coefficients and confidence intervals estimated in the hedonic specification. The figure allows to appreciate spatial decay characterising the cumulative treatment.

Figure 2. 10: Mafia families in Naples, 2013

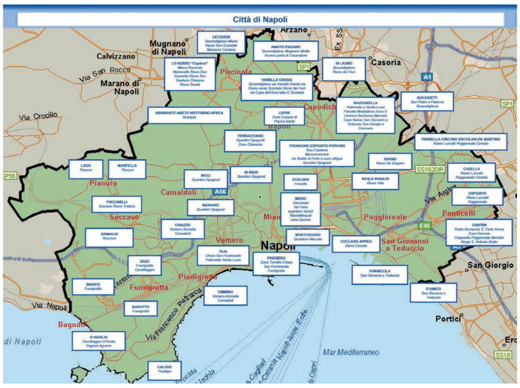
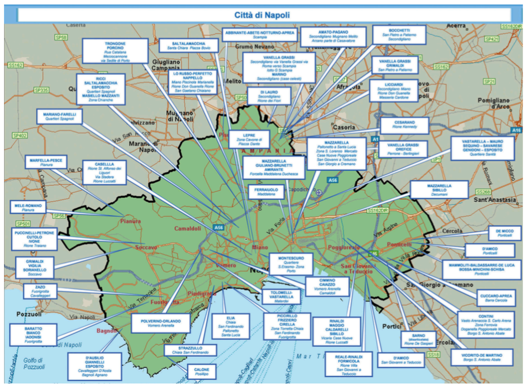
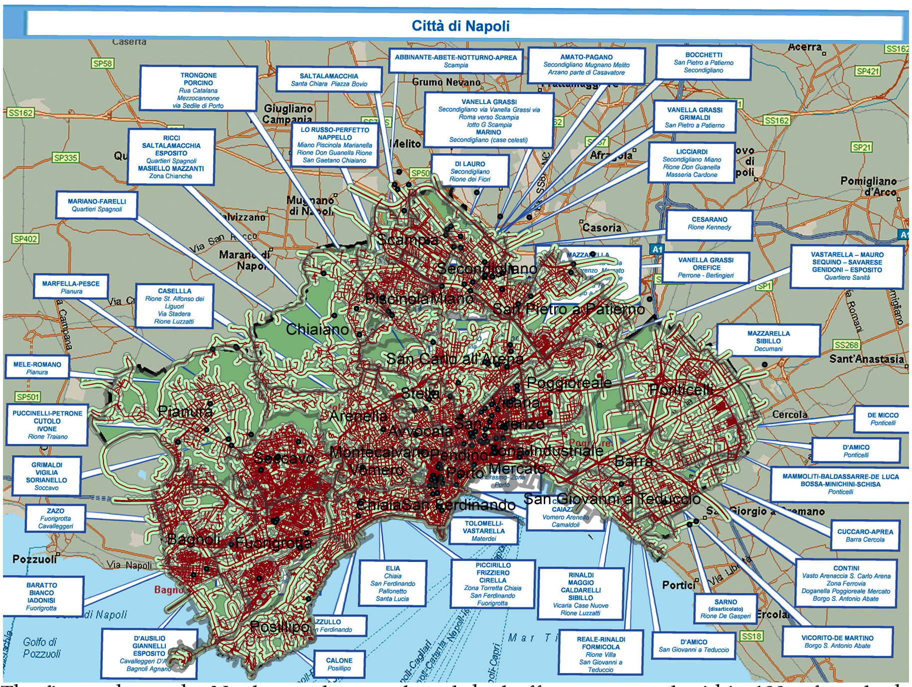


Figure 2. 11: Mafia families in Naples, 2018



The figure retrieved from Transcrime (2018) shows the spatial distribution of mafia families in Naples in 2013 and 2018

Figure 2. 12: Street-level dataset



The figure shows the Naples road network and the buffer constructed within 100m from both side of each road

Appendix

Table A2. 1: Timing of re-allocations

	Years between confiscation and re-allocation					
	0-1	2-3	4-5	6-7	8-9	10+
Number of re-allocated real estate properties	83	603	1,684	2,830	2,585	7,913
% of total re-allocated	0.5	3.8	10.7	18.0	16.5	50.4

Source: own elaboration with ANBSC data.

Table A2. 2: Re-allocation and local area characteristics

	Local area characteristics:				
Dep. variable: Re-allocation	Ln pop	Illiterate pop	Unemployed	Rented pop	Buildings bad conditions
	(1)	(2)	(3)	(4)	(5)
	-0.0153*	0.00851	0.01758**	0.00695***	0.00441**
	(0.00859)	(0.00674)	(0.00710)	(0.00229)	(0.00182)
Observations	123,718	123,718	123,718	123,718	123,648
R-squared	0.001	0.001	0.007	0.017	0.002
	-0.000189	0.000228	0.000092	6.53e-05	-0.00337
	(0.000682)	(0.000369)	(0.000250)	(6.36e-05)	(0.00660)
OMI FE	✓	✓	✓	✓	✓
Observations	121,174	121,174	121,174	121,174	121,107
R-squared	0.913	0.913	0.913	0.913	0.913

Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Dependent variable: Re-allocation dummy. Local area conditions: Log population, percentage of residents with tertiary education, percentage of illiterate population, percentage of unemployed, percentage of foreigners, Buildings being occupied and used as percentage of total in local area, buildings in excellent conditions as percentage of total in local area, buildings in bad conditions as percentage of total in local area.

Table A2. 3: Local governments and re-allocation duration

	Italy local Governments 1998-2017		Re-allocations 1998-2017		Re-allocations timing: 0-9 years		Re-allocations timing: 10+ years	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Party colour	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Right	5,886	14.3	2,436	26.9	1,256	27.9	1,777	39.2
Centre	5,158	12.6	595	6.6	305	6.8	290	6.4
Left	9,950	24.3	3,359	37.2	1,582	35.2	1,180	26.1
5Star	425	1.1	290	3.2	49	1.1	241	5.3
Civic list	23,664	57.7	2,280	25.3	1,332	29.7	948	20.9
Dissolved	274	0.7	300	3.3	202	4.5	98	2.1

Party colour: ideological leaning/party type of municipal governments during 1998-2017 in Italy. Civic lists: electoral lists/parties different from national parties, often created ad hoc for local elections. Right, Centre and Left include civic lists of that political colour. Civic list includes both ideologically identifiable lists and non-identifiable lists. Dissolved: municipal governments dissolved for any reason, such as collusion/corruption, financial disarray, vote of no confidence.

Table A2. 4: Re-allocation and local area characteristics

	Local area characteristics					Re-allocated building characteristics			
Dep. variable: Re-allocation timing	Ln pop	Illiterate pop	Unemployed	Rented pop	Buildings bad conditions	Social use	Institutional	Residential buildings	Terrains
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	6.23e-06	0.0274	0.429***	0.0460	0.0746	-1.967***	2.440**	-0.269	-0.386
	(0.306)	(0.102)	(0.105)	(0.0342)	(0.0732)	(0.652)	(1.023)	(0.493)	(0.540)
Re-allocation year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	5,999	5,999	5,999	5,999	5,999	8,969	8,969	8,969	8,969
R-squared	0.005	0.005	0.006	0.005	0.006	0.009	0.009	0.007	0.007
	-0.284	0.0915	0.403	-0.0105	-0.0140	-1.862	1.883	-0.202	-2.458
	(0.531)	(0.168)	(0.402)	(0.0647)	(0.120)	(1.669)	(2.211)	(0.318)	(2.630)
Re-allocation year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	5,593	5,593	5,593	5,593	5,579	8,438	8,438	8,438	8,438
R-squared	0.078	0.078	0.078	0.078	0.077	0.089	0.089	0.089	0.089

Notes. The table illustrates the relation between the length of re-allocation procedure and characteristics of the area where the confiscation took place and of the asset. . Independent variable: columns (1)-(5): local area conditions. Log population, percentage of residents with tertiary education, percentage of illiterate population, percentage of unemployed, percentage of families being rented, buildings in bad conditions as percentage of total in local area. Columns (6)-(9): re-allocated building characteristics (dummy variables). Re-allocated for social use, re-allocated for institutional use, residential buildings, terrains. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A2. 5: Property characteristics

Type of data	Variables
Identifiers	Unique ad identifier, date in which the ad was created in the database, date in which the ad was removed from the database, date in which one of the characteristics of the ad was modified for the last time
Numerical	Price, floor area, rooms, bathrooms, year built
Categorical	Property type, kitchen type, heating type, maintenance status, floor, air conditioning, energy class
Type of building	Elevator, garage/parking spot, building category
Geographical	Longitude, Latitude, address
Temporal	Ad posted, ad removed, ad modified
Contractual	Foreclosure auction
Textual	Description

The table illustrates the main variable types available in the hedonic dataset

Table A2. 6: Descriptive statistics: treatment variables

Variable	Obs	Mean	Std. Dev.
OMI zones:			
Price €/m2	262,740	1188.5	778.9
Re-allocation	388,884	0.0166	0.128
Confiscation	388,884	0.0134	0.115
Sale points:			
Price €/m2	52,651	2415.3	1525.3
Re-allocation	52,651	0.166	1.269
Confiscation	52,651	0.0391	0.721
Re-allocation year	52,651	0.0487	0.608
1 year after re-allocation	52,651	0.0521	0.615
2 years after re-allocation	52,651	0.0339	0.594
3 years after re-allocation	52,651	0.0169	0.286
4 years after re-allocation	52,651	0.0142	0.197

The table reports descriptive statistics for the variables of interest.

Table A2. 7: Descriptive statistics: sale point characteristics

Variable	Mean	Std. Dev.
Distance to green area	6,647.6	4,305.6
Distance to beach max 20km	172,000	335,000
Distance to city viewpoint 1km	19,962.3	10,809.2
Distance to a University	50,317.5	27,780.2
Distance to bus, tram or metro	3,081.6	755.6
Distance to Intercity transport, railway	6,017.8	1,750.8
Distance to airport	17,593.4	17,172.7
Distance to commercial centre	25,858.5	14,489.2
Distance to church	729.5	406.9
Distance to state schools	6,896.7	994.2
Noise - within 500m of a highway	0.23	0.06
Dummy industrial area	0.16	0.03
Distance to factory	5,859.9	2,665.2
Distance to construction site	19,820.4	9,124.5
Month of offer	3.51	5.00
Lift dummy	0.49	0.41
Building height	8.04	14.05
Typology of building	1.24	2.62
Area of building	1,141.1	538.4
Average typology of building in street	0.66	2.71
Property up for auction	0.14	0.02
Type of property	0.71	4.02
Number of rooms	1.30	2.80
Number of bathrooms	0.69	1.51
Type of kitchen	0.70	1.46
Floor number	2.61	2.01
Parking with property	0.47	0.33
Periods year built	2.01	2.49
Property condition	1.08	2.19
Property heating type	0.73	0.93
Air conditioning	0.44	0.27
Energy Efficiency	0.83	0.87

The table reports descriptive statistics for the sale-point-level variables used in the analysis.

Table A2. 8: Descriptive Statistics: Census area characteristics

Variable	Obs	Mean	Std. Dev.
Population	123,718	341.7	265.8
% Illiterate population	123,718	0.94	1.43
% Unemployed	123,718	3.24	1.82
% Rented families	123,718	8.30	7.20
% Buildings bad conditions	123,648	1.15	4.28

The table reports descriptive statistics for census-area-level variables used in the analysis.

Table A2. 9: Re-allocation and electoral turnout

	Dep. variable: Re-allocation		*Dep. variable:* Re-allocation timing	
	(1)	(2)	(3)	(4)
Turnout	-0.168***	0.0115	-17.46***	-0.984
	(0.00935)	(0.0162)	(1.857)	(5.306)
OMI FE		✓		✓
Observations	26,898	26,898	633	633
R-squared	0.003	0.044	0.123	0.445

Notes. The table illustrates the relation between re-allocation events, duration of the re-allocation procedure and local turnout. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.
.

Table A2. 10: Sale point analysis – d=150

Log euro per m^2	Buffer radius: 150m						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 year before re-allocation	-0.000447	-0.00114	-0.00158	0.000269	-0.000654		
	(0.00531)	(0.00548)	(0.00543)	(0.00581)	(0.00569)		
Re-allocation year	-0.00376	-0.00439	-0.00475	-0.00263	-0.00556		
	(0.00600)	(0.00580)	(0.00580)	(0.00522)	(0.00542)		
1 year after re-allocation	0.00754	0.00692	0.00628	0.00756*	0.00574		
	(0.00514)	(0.00495)	(0.00499)	(0.00464)	(0.00398)		
2 years after re-allocation	0.00449	0.00400	0.00390	0.00433	0.00224		
	(0.00492)	(0.00484)	(0.00488)	(0.00431)	(0.00443)		
3 years after re-allocation	0.00549*	0.00503*	0.00487*	0.00494**	0.00549**		
	(0.00303)	(0.00294)	(0.00288)	(0.00222)	(0.00219)		
Re-allocation						0.00439**	0.00442**
						(0.00220)	(0.00222)
Confiscation							-0.00544
							(0.00376)
Structural controls	✓	✓	✓	✓	✓	✓	✓
Building controls		✓	✓	✓	✓	✓	✓
Amenity controls			✓	✓	✓	✓	✓
Socio-econ. controls				✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓	✓	✓
Municipality-year FE					✓	✓	✓
Observations	52,526	52,513	52,513	52,513	51,906	51,906	51,906
R-squared	0.768	0.769	0.771	0.777	0.784	0.784	0.784

Notes. The table reports the estimation results for hedonic analysis presented in Section 2.4.2. Columns (1)-(5) report the effect of property re-allocation taking place within 150m from the sale point. Housing prices differences recorded the same year or the year before make it possible to account for pre-treatment differences in housing prices. Columns (6) and (7) report the cumulative effect of confiscation and reallocation events on housing prices. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A2. 11: Sale point analysis – d=500

Log euro per m²	Buffer radius: 500m						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 year before re-allocation	0.00220	0.00215	0.00194	0.00248	0.00115		
	(0.00175)	(0.00172)	(0.00173)	(0.00176)	(0.00203)		
Re-allocation year	0.00202	0.00212	0.00198	0.00235*	0.000879		
	(0.00139)	(0.00134)	(0.00133)	(0.00126)	(0.00130)		
1 year after re-allocation	0.00311**	0.00277**	0.00254*	0.00314**	0.00139		
	(0.00146)	(0.00140)	(0.00145)	(0.00140)	(0.00117)		
2 years after re-allocation	0.00180	0.00186	0.00174	0.00229**	8.38e-05		
	(0.00119)	(0.00116)	(0.00118)	(0.00104)	(0.00110)		
3 years after re-allocation	0.00254*	0.00241*	0.00226*	0.00312***	0.00307***		
	(0.00137)	(0.00133)	(0.00132)	(0.00114)	(0.00113)		
Re-allocation						0.000954	0.000872
						(0.00086)	(0.00086)
Confiscation							-0.0035**
							(0.00151)
Structural controls	✓	✓	✓	✓	✓	✓	✓
Building controls		✓	✓	✓	✓	✓	✓
Amenity controls			✓	✓	✓	✓	✓
Socio-econ. controls				✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓	✓	✓
Municipality-year FE					✓	✓	✓
Observations	52,526	52,513	52,513	52,513	51,906	51,906	51,906
R-squared	0.768	0.769	0.771	0.777	0.784	0.784	0.784

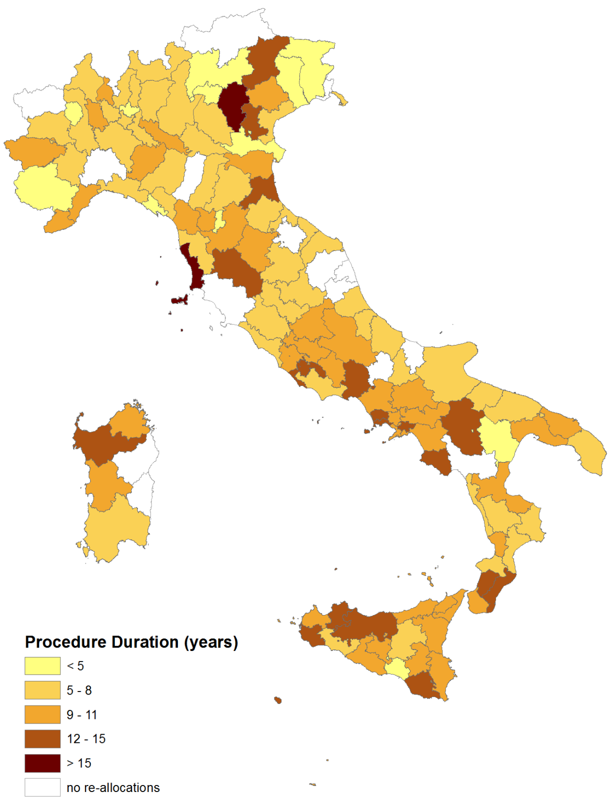
Notes. The table reports the estimation results for hedonic analysis presented in Section 2.4.2. Columns (1)-(5) report the effect of property re-allocation taking place within 500m from the sale poin. Housing prices differences recorded the same year or the year before make it possible to account for pre-treatment differences in housing prices. Columns (6) and (7) report the cumulative effect of confiscation and reallocation events on housing prices. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table A2. 12: Sale point analysis controlling for buffer zone

Dep. variable: Log euro per m²	100m	200m	300m	400m	500m
	(1)	(2)	(3)	(4)	(5)
Buffer zone	-0.00762	-0.0188	-0.0192	-0.0301*	-0.0328*
	(0.0112)	(0.0169)	(0.0149)	(0.0172)	(0.0175)
Re-allocation	0.00706**	0.00402**	0.00335**	0.00222**	0.00175*
	(0.00342)	(0.00171)	(0.00153)	(0.00110)	(0.00088)
Structural controls	✓	✓	✓	✓	✓
Building controls	✓	✓	✓	✓	✓
Amenity controls	✓	✓	✓	✓	✓
Socio-econ. controls	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
OMI FE	✓	✓	✓	✓	✓
Municipality-year FE	✓	✓	✓	✓	✓
Observations	51,906	51,906	51,906	51,906	51,906
R-squared	0.784	0.784	0.784	0.784	0.784

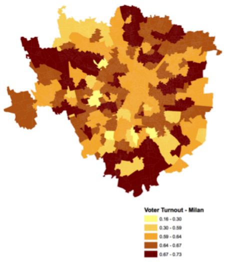
Notes In this table we test the robustness of these results by including in the model a control for the buffer zone, controlling for time-invariant characteristics located at a distance d from the sale point. Robust standard errors are clustered at the municipality level and reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Figure A2. 1: Re-allocation duration by Court

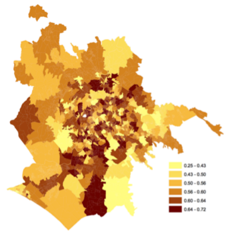


The figure shows the average time required for local cohorts to re-allocate confiscated mafia assets

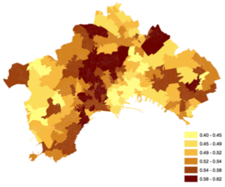
Figure A2. 2: Polling station areas



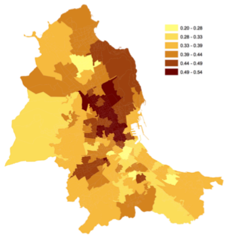
(a) Milan



(b) Rome



(a) Naples



(b) Palermo

The figures report the average voter turnout in the European elections by polling station area in four Italian cities

Chapter 3:
Within-industry co-agglomeration patterns

3.1 Introduction

This chapter analyses the spatial distribution of manufacturing and service firms in France and the United Kingdom. By exploiting a large micro-dataset covering the whole population of French and British establishments over the period 2008-2015, I analyse the specific co-agglomeration patterns that characterise various types of plants within each industry.

It is a fact that industries tend to be geographically concentrated. The benefits that firms gain by locating near each other have been studied by an extensive and heterogeneous literature (among others, Sveikauskas,1975; Moomaw, 1981; Carlton, 1983; Henderson, 1986; Nakamura, 1985; Glaeser et al, 1992; Henderson et al, 1994; and Ciccone & Hall, 1996). Marshall in 1890 was the first economist to theorise and analyse this specific dynamic; he argued that the cities foster firm productivity through labour market pooling, input sharing and technological spillovers.
Modern empirical literature on the topic has focused on the concept of agglomeration economies, i.e. the idea that firms benefit mutually from their presence on the same territory. These kinds of studies are usually referred to as 'urbanisation economies' (Jacobs externalities), when the benefits arise from the size of the overall market, and 'localisation economies' (Marshall-Arrow-Romer externalities), when they depend on geographic concentration at the industry level (Dicken & Lloyd, 1990; Glaeser et al., 1992).

In recent decades, a new branch of the literature has questioned the implicit assumption that agglomeration economies are driven by the same forces in different economic environments and industries. Several studies have separately analysed Marshall's agglomeration forces, offering empirical evidence for each of them (Fallick et al., 2006; Almazan et al., 2007; Holmes, 1999; Arzaghi & Henderson, 2008; Lin, 2012). Duranton and Puga (2004) further analysed the sources of agglomeration economies, proposing a different taxonomy on the basis of the concepts of sharing, matching and learning. Modelling each mechanism, they suggested a relevant role for firms' and workers' heterogeneity in determining the effective outcome of agglomeration forces.

Strange et al. (2006) analysed the three mechanisms in relation to both urbanisation and localisation economies, suggesting competitive instability and technological innovativeness to be associated with city size, and skill-orientation to be more relevant for industry clustering. Glaeser and Kerr (2009) and Rosenthal and Strange (2010) focused instead on firm dimension, theorising an organisational dimension for agglomeration effects. In order to disentangle the drivers responsible for agglomeration dynamics in different industries, Ellison and Glaeser (2010) proposed a new approach, focusing on the concept of coagglomeration, the tendency of industries to locate together. Following this approach, Faggio, Silva and Strange (2014) develop a unified framework to analyse heterogeneity in industrial agglomeration and its implications for the micro-foundations of agglomeration economies. On the basis of a comprehensive analysis, they question the external validity of the mixed results found in the literature. Specifically, they demonstrate that agglomeration economies and, in particular, the relative and absolute strength of the underlying drivers vary consistently with respect to the specific characteristics of industries and firms.

Other studies argue that agglomeration effects could be heterogeneous even within industries. The way external economies affect firm behaviour and performance would depend on various firm-specific characteristics. Békés, Kleinert and Toubal (2009), using Hungarian data, report a heterogeneous response from agglomeration spillovers, demonstrating that while the most productive firms gain substantially, laggards lose out on higher presence of foreign firms. Combes et al. (2012) develop and econometric strategy to distinguish firm selection from the other factors behind the productivity premium of denser areas, studying the distribution pattern of total factor productivity at the firm-level. The study concludes that the observed productivity advantages that characterise French metropolitan firms are not simply due to the selection of the most productive firms, but are also driven by external economies of scale. It also demonstrates that agglomeration effects increase with the firm productivity.

This study contributes to this literature by shedding new light on the specific agglomeration and co-agglomeration patterns that characterise certain plant types within industries. Plants are analysed with respect to size, ownership, demographic characteristics and other firm and plant-level dimensions. Results are discussed on the basis of the main predictions of the recent literature on agglomeration economies.

The literature on regional and urban economics has produced various measures to analyse the distribution of industries over space. Traditional discrete indices of spatial agglomeration, such as the one proposed by Ellison and Glaeser (2003), depend heavily on the spatial unit of analysis used. This well-known issue is commonly described in the literature as the modifiable area unit problem (Openshaw & Taylor, 1979, Arabia 1989, Briant et al., 2010). Although discrete indices can still provide interesting insights with respect to the actual distribution of a set of industries belonging to a common sector at a given moment in time, they fail to provide comparable estimates for cross-country or cross-sector analysis. A new class of continuous indices have been proposed as a valuable alternative to the traditional discrete measures. In this regard, the first contribution dates back to Ripley (1976). Ripley's K measures the spatial concentration or dispersion of points in space, compared to a complete spatial randomness benchmark. The K function is estimated by calculating the average number of neighbouring plants within circles of increasing radius around the establishment. Since it requires the calculation of bilateral distances between all spatial points analysed, computational constraints limited its application in the early urban economics literature. Marcon and Puech (2003) proposed a Besag L-transformation of Ripley's K, with a benchmark value of 0. This index and the different extensions provided by the literature (see Marcon and Puech, 2017 for a comprehensive review) do not generally allow the weight of points on the basis of meaningful characteristics, nor provide meaningful benchmarks other than the unrealistic random distribution. With respect to the agglomeration literature, Duranton and Overman (2005) define a set of properties a spatial agglomeration index should satisfy:

- be comparable across industries;
- control for the overall agglomeration of manufacturing;
- control for industrial concentration in the sense of Ellison and Glaeser (1997);
- be unbiased with respect to scale and aggregation; and
- provide an indication of the significance of the results

Duranton and Overman develop an index that satisfies all these properties. The DO function can be interpreted as the probability density to find a neighbouring spatial point at a given distance from the point of interest. The function is subsequently analysed against a confidence interval estimated through Monte Carlo simulations over 1,000 random draws from the overall sector.

In this study I use the DO function to analyse the spatial distribution of plants within industries. Its specific characteristics make it possible to compare plant-types across industries, countries and years.

Moreover, in contrast to main discrete indices, it provides meaningful values even when analysing small group of plants and establishments of different sizes. Following Behrens (2017), I integrate the function up to different distance threshold, to get the probability for a random firm pair bilateral distance to be lower or equal than *d*.

3.2. Data

Establishment- and firm-level data are retrieved from two sets of confidential datasets provided by the French and British Institutes of Statistics.

France

Data are retrieved from two administrative sources made available by the INSEE (the French national statistical agency) to researchers with access to CASD secure lab. The first source is the Declarations Annuelles de Donnes sociales (DADS). Every year all French workers are expected to submit to the tax authority information about all incomes earned over the previous 12 months. The INSEE receives these data from the tax authority, merges them with other information on workers and households provided by various sources, and uses them to produce statistics about employment and wages.
The data used in this process are recorded in an exhaustive dataset, where each observation corresponds to every job contract linking a worker to a firm in a given year. For every worker, the dataset provides job-spell level information on gross and net wage, employment periods, age, sex and skills, number of hours worked, type of contract and occupational category (Professions et categories socio-professionnelles, PCS 2003) at the 4-digit level.

The dataset FICUS/FARE, produced by INSEE/DGFiP, provides balance sheet information (output, capital, inputs, exports, number of employees, sector, etc.) for each firm registered in France from 1993 to 2016 (approximately 47 million observations). Data are retrieved from the compulsory reporting of firms and income statements to fiscal authorities in France, with no limiting threshold in terms of firm size or sales. By merging the dataset with the information provided by the Sirene (Système Informatique pour le Répertoire des Entreprises de leurs Etablissements) dataset, it is possible to assign address and precise spatial coordinates to 27 million establishments (the entire population of establishments that ever operated in France over the past 20 years). In this analysis, I drop from the sample firms subject to the micro-bin/micro-bic regimes, namely firms:

- whose principal activities are services relating to the category of business profits (BIC) or non-commercial profits (BNC)
- whose turnover does not exceed €32,600 (for 2011), excluding taxes
- which operate VAT free and do not perform an excluded activity (for example, the rental of equipment and durable consumer goods);
- which do not opt for the *régime du réel simplifié* tax system

The Liaison Financiére dataset (LIFI) provides data on all relevant financial linkages involving at least one firm located in France. For each observation, I can determine the country of residence of all foreign subsidiary/parent firms and the capital share. The median voting share owned upon acquisition is 99%, with the result that the acquisition event represents a near complete takeover of assets and control for the overwhelming majority of the sample.

Geocoded plant level data are retrieved from the Sirene database. Each plant is assigned to specific coordinates, corresponding to the address reported for fiscal purposes.
This information is missing for a small number of plants (<5%), located primarily in rural areas. In these cases, I assign the coordinate of the applicable *IRIS*. *IRIS* are the smallest French administrative unit (the French territory consists of over 49,000 IRIS areas) and plants with missing coordinates are expected to be located, on average, 400 m from their estimated location.

United Kingdom

Confidential establishments' data are retrieved from the ONS Business Structure Database 1997-2015 (Secure Access), that covers any business liable for value-added taxation and/or with at least one employee registered for tax collection. Overall, the dataset covers 99% of economic activity in the UK. The dataset provides information on postcode, ultimate owner, 5-digit industrial classification, date of creation, and number of employees for each single plant. I use this detail to assign to each plant active in England, Wales and Scotland to the coordinates of its postcode centroid. The raw data include approximately 2.8 million local units every year. I carry out a series of checks, excluding a number of anomalous cases and checking the consistency of plants postcodes, employment and sectors of activity over the years.

Firm level data are retrieved from the Annual Respondent Database (ARD) for the period 2001-2007 and from the Annual Business Survey for the remaining years. The two datasets provide balance sheet data for a large sample of firms with more than 20 employees. In particular, the survey is known to provide particularly good coverage of manufacturing firms.

3.3. Empirical Strategy

In this section I illustrate the empirical strategy used. Following Behrens at al. (2017), I extend the continuous agglomeration index developed by Duranton and Overman (2003) and I compute agglomeration and co-agglomeration measures for a set of industries and plant types within industries. Subsequently, the distribution of the plant-type/industry-specific agglomeration indices is analysed in a common framework, to test some relevant predictions proposed in the literature.

3.3.1 Continuous agglomeration indices

Duranton and Overman's localisation index (henceforth referred to as K_j) is a continuous estimator of localisation based on the distance between every pair of plants. Denoting d_{xy} the Euclidean distance between plants x and y, the estimator of the density of bilateral distances at any point d is:

$$K_j(d) = \frac{1}{n(n-1)h} \sum_{x=1}^{n-1} \sum_{y=x+1}^{n} f\left(\frac{d - d_{xy}}{h}\right)$$

where h is the bandwidth and f is the kernel function. $K_j(d)$ will be larger when the distance between many establishment pairs is approximately d.
A slightly more advanced – although much more computationally intensive – version of the index illustrates the distribution of bilateral distances between workers in a given industry.

$$K_j^w(d) = \frac{1}{h \sum_{x=1}^{n-1} \sum_{y=x+1}^{n} (e_x + e_y)} \sum_{x=1}^{n-1} \sum_{y=x+1}^{n} f\left(\frac{d - d_{xy}}{h}\right)$$

Where d_{xy} is the distance between plant x and plant y, h is the bandwidth (500 meters), e_x and e_y are the employment levels in the two establishments.

For each industry, $K_j^w(d)$ is subsequently compared with the counterfactual K-density estimated from 1,000 simulations of bilateral distances between n randomly sampled (distinct) establishments in the aggregate industry. Local confidence intervals are then estimated for any distance between 0 and $\bar{d}$.
For each industry j, they select the 5-th and 95-th percentile to obtain a lower 5% and an upper 5% confidence interval that denote $\underline{K^w}(d)$ and $\overline{K^w}(d)$ respectively. For each industry, when $K_j^w(d) > \overline{K^w}(d)$ for at least one $d \in [0, \bar{d}]$, the industry is said to exhibit localisation at distance $\bar{d}$ (at a 5% interval). By contrast, when $K_j^w(d) < \underline{K^w}(d)$ for at least one $d \in [0, \bar{d}]$, the industry is said to exhibit dispersion at distance $\bar{d}$ (at a 5% interval).

The authors propose two simple indices to summarise the distribution pattern of each industry:

Global localisation index:

$\gamma_j^w = \max\left(K_j^w(d) - \overline{K^w}(d), 0\right)$

Global dispersion index:

$\psi_j^w = \max\left(\underline{K^w}(d) - K_j^w(d), 0\right)$

Graphically, global localisation is detected when the weighted $K_j^w(d)$ function lies above its upper confidence band, whereas global dispersion is detected when it lies below the lower confidence band and never lies above the upper confidence band.

In Figure 3. 1, I report the DO function estimated for the Nace Rev.2 industry 2361 ('Manufacture of concrete products for construction purposes') in 2008. In that year the industry counted 1,158 plants and, consequently, 1,340,964 bilateral distances. The blue line reports the weighted probability of finding a bilateral distance in each 500m bin, between 0 and 180km. The dotted lines limit the confidence interval, obtained by means of a Monte-Carlo process, where the same procedure is randomly repeated 1,000 times, on each occasion drawing a sample of 1,158 plants from the overall population of manufacturing firms (850,000 plants). For any distance threshold, the upper boundary represents the 95th percentile of the agglomeration distribution of all random draws. Symmetrically, the lower bound corresponds to the 5th percentile of the distribution. The chart shows that the probability of a pair of firms to be at a distance d from each other range between 0.01% and 0.3%.

Comparing the function with the sector benchmark, we can say that the concrete industry is agglomerated up to a distance threshold of 50km, while it does not deviate from the probability distribution of sector for longer distances.

The weighted K-density proposed by Duranton and Overman describes the distribution of bilateral distances between workers in a given industry. In order to obtain an index providing information on the degree of agglomeration/dispersion, I integrate the distribution function, obtaining its cumulative (CDF) up to a distance d. The values obtained represent the weighted share of plant-pairs located less than distance d from each other. Alternatively, we can view this as the probability that two randomly drawn plants in an industry will be at most d km away from each other.

$$\overline{K_j^w}(\bar{d}) = \sum\nolimits_{d \leq \bar{d}} K_j^w(d)$$

This procedure was used by Behrens (2017) to study the agglomeration of Canadian manufacturing industries. The blue area in Figure 3. 2 corresponds to the $\overline{K_j^w}$ computed at a 180km threshold.

Figure 3. 1: DO function

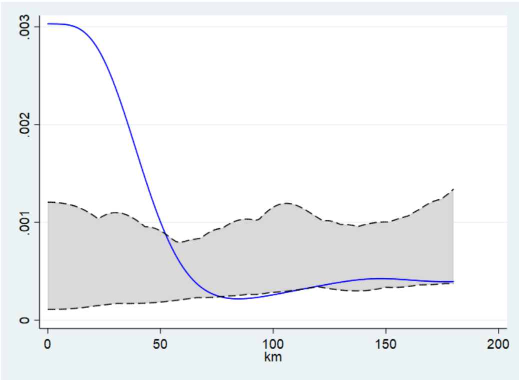
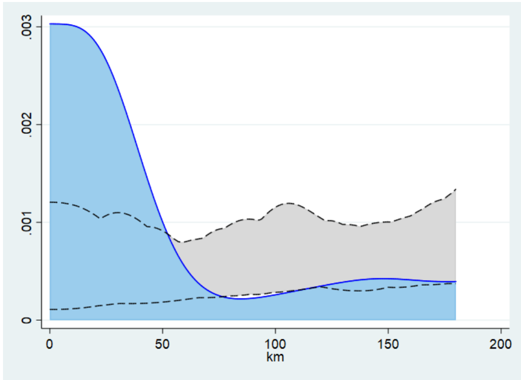


Figure 3. 2: CDF



The charts report the $\widehat{K}_w^j(d)$ agglomeration functions for the whole industry (blue line) and the confidence interval (gray area)

3.3.2 Localisation and co-localisation patterns

If the CDF can be directly used to study the location pattern of a given industry *j* in a given time *t*, a different approach is required in order to describe the location pattern of a specific group of plants *i* within an industry *j*.

Consider Figure 3. 3. Plants (blue and red dots) are clustered in three different areas. If we focus on group types and we consider the share of plants located within a distance *d* from each other, establishments *i* are found to be more agglomerated than the other plants in the area. As a matter of fact, 100% bilateral distances fall below the distance threshold. By contrast, if we measure the bilateral distances between all establishment in the industries, irrespective of the plant type, we find that *i* plants are somewhat dispersed than plants *p* (Figure 3. 4).

Figure 3. 3: Agglomeration

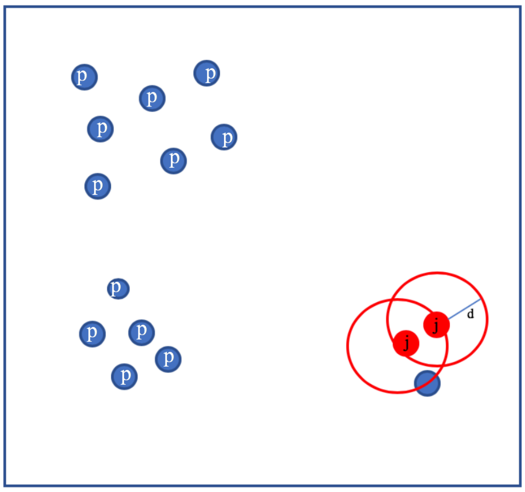
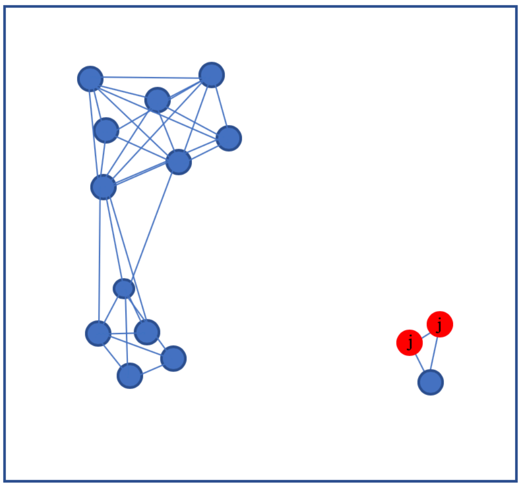


Figure 3. 4: Co-agglomeration



It is not hard to think about similar scenarios in the real world. For instance, we might find an industry characterised by a relevant share of multi-plant firms exhibiting high levels of agglomeration of headquarters (primarily located close to the CBD) and a certain degree of dispersion when the same headquarters are related with productive plants, located in different environments, far from the CBD. It follows that industry-level measures that do not take establishment heterogeneity into account could fail to detect agglomeration or co-agglomeration patterns characterising the plants that are more sensible to agglomeration and dispersion forces. Therefore, while studying spatial agglomeration within industries, it is possible to describe three main features:

I. The agglomeration of all plants in the industry: $\overline{K_J^w}$

II. The agglomeration of plants belonging to a certain type (group agglomeration): $\overline{K_{\iota J}^w}$

III. The co-agglomeration between i-type plants and all establishments in the industry (joint agglomeration): $\overline{K_{\iota J,J}^w}$

Within each industry, we can identify various types of firms and plants on the basis of demographic characteristics, size, productivity level, foreign status, etc… All these factors might determine a different spatial configuration with respect to the other plants located in the industry. Fig. 3.5 illustrates the distribution of headquarters in the industry 2825 'Manufacture of non-domestic cooling and ventilation techniques'. In this case, the green line corresponds to the distribution of all 700 plants (490,000 bilateral distances) in the industry. The red line corresponds to the $\overline{K_{\iota J}^w}(d)$ function estimated for headquarters only, while the blue line shows co-agglomeration $\overline{K_{\iota J,J}^w}(d)$ estimated for all bilateral distances between a headquarter and another plant in the industry. In the case of the cooling industry, headquarters exhibit a relevant agglomeration at short distances. However, when we consider the co-agglomeration with non-headquarter, the function almost mimics the more dispersed pattern that characterise the whole sector. I provide a further example in Figure 3. 6, that illustrates the distribution of the industry 2651 'Manufacture of instruments and appliances for measuring, testing and navigation'. In this case, the distribution of the whole industry is compared with the specific spatial pattern of high-productive firms. Interestingly, firms in the top 10% of the labour productivity distribution are significantly more dispersed than less productive ones up to a 70km distance.

Figure 3. 5: DO functions, Headquarters

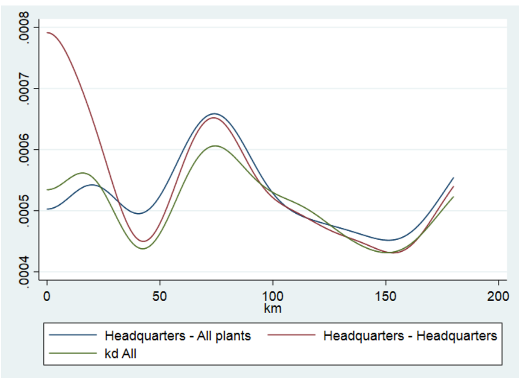
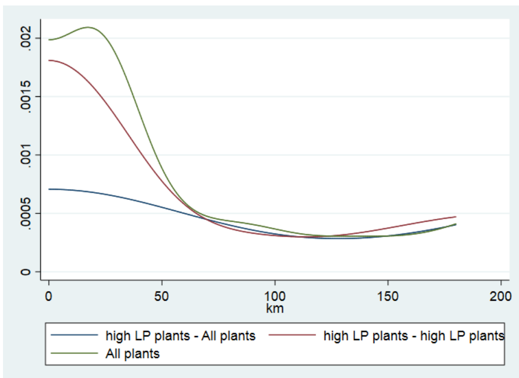


Figure 3. 6: DO functions, High LP



3.4. Agglomeration trends by macro-sectors

In this section, I analyse the evolution in agglomeration patterns of manufacturing and service firms in France and the United Kingdom. To this end, I estimate the $\overline{K_j^w}(d)$ of 124 manufacturing industries and 178 service industries in France and the United Kingdom in every year, over the period 2008-2015. Industry estimates are subsequently aggregated using employment weights.

First, it is worth noting certain geographical features that can explain the overall agglomeration (agglomeration up to 180km) in the two countries. The UK is characterised by a very agglomerated structure, with London alone accounting for more than 28% of British GDP and a very large share of service plants in many industries. Some traditional manufacturing industries are concentrated in a limited area that includes East and West Midlands, Wales and Yorkshire. France, on the other hand, has a more decentralised structure. Although Paris and the Ille de France represent a massive share of the whole economy, other important economic centres such as Lyon and Marseilles are located on the opposite side of the country, over the rural diagonal that cross the country from south-west to north-east. If the two countries cannot be compared in terms of levels, we can still compare the evolution of the respective economic geographies.

Figures 3.7-3.8 and 3.9-3.10 describe, respectively, the trends in local agglomeration patterns (up to 30km) of manufacturing and service sectors in France and the United Kingdom. In both countries, local agglomeration fluctuated over the period, with a downward trend inverted in 2012, in the aftermath of the debt crisis. France demonstrates a strikingly different pattern when plants are associated with their firm industry code, a dynamic probably associated with the general trend of domestic fragmentation. French service plants exhibit a U-shaped pattern, whereas a constant decline is registered in the UK.

Global agglomeration trends (Fig. 3.11-3.14) are more consistent, with a relevant decline in service spatial agglomeration up to 180km and a common pattern in manufacturing up to 2015, a year in which the UK experience a steady jump. Besides the fluctuating pattern that characterise the dynamic, spatial agglomeration levelled off or decreased over the period. This evidence is in line with other studies that identify a trend of de-concentration encompassing most manufacturing activities and business services in Germany (Dauth et al., 2018) and Canada (Beherens et al., 2013).

In the eight-year period under consideration, manufacturing was exposed to two main forces: technological change and trade liberalisations. On the one hand, new technologies caused a consistent reduction in labour demand for routine occupations, that once employed the large majority of British and French workers. On the other hand, reductions in trade and communication costs fostered foreign competition and offshoring. The progressive disappearance of most labour-intensive industries was complemented by the relocation of a significant part of production processes of most industries to cheaper countries. The general decline in the manufacturing sector and the consequent tertiarization process has recently been linked not only to job disruption but also to significant changes in the production process and the organisation of firms across industries. Fort (2013) analyses the recent rise of multi-plant firms in the US manufacturing sector. In addition to merger and acquisitions (M&A) decisions, a firm might decide to segment its production over space in order to access cheaper labour or to locate different plant types in areas where they can achieve the preferred distance with respect to customers, suppliers and competitors. She demonstrates that domestic fragmentation is far more prevalent than offshoring, especially when firms can access advanced communication technology. For all these reasons, it is not surprising to see an evolution in the aggregate spatial agglomeration patterns. The same dynamics described above have driven a constant expansion of Service sector in recent years. Unlike manufacturing, however, the extreme heterogeneity that characterises the sector in terms of tradability and B2C/B2B relations makes it more difficult to define common centripetal and centrifugal forces. On the one hand, the rise in non-tradable service industries could be associated with a more dispersed distribution of plants, aimed at reaching customers far from the city centres. In contrast, new technologies have the potential to increase the tradability of certain types of services, fostering spatial agglomeration. In addition, for certain industries the dispersion could simply be driven by the dynamic pattern characterising some manufacturing industries they are connected with by significant input-output linkages.

3.5 Industrial agglomeration patterns

In this section I briefly analyse the industrial location patterns of some manufacturing and service industries. In particular, I will focus on industries exhibiting highest and lowest agglomeration levels, according to the indices used.

In Table 3. 1, I list the manufacturing industries recording the highest and lowest average levels of spatial agglomeration at 30 and 180km distance in France and in the United Kingdom. The table shows that the two countries are not symmetrical in the agglomeration ranking. However, it is possible to identify some common patterns. Among the industries recording the highest levels of local agglomeration (share of bilateral distances below 30km), I find textile and wearing apparel industries, chemical industries, industries characterised by very low transport costs (Reproduction of recorded media) or technologically advanced products (Manufacture of communication equipment) and industries that require specific environments (Manufacture of wine from grapes).

These features are also common among industries exhibiting high degree of global agglomeration (share of bilateral distances below 180km), where we also find luxury product industries (Building of pleasure and sporting boats) and craft manufactures. Among the industries at the opposite site of the local agglomeration distribution we find manufactures of products with high storage or transport costs (manufacture of ice cream, wood, structural metal and concrete products) and industries requiring proximity to customers or dispersed geographic environments (Building of ships and floating structures). Similar patterns characterise most dispersed industries at 180km distance.

Table 3. 2 illustrates most agglomerated and dispersed service industries. In this case the patterns are even more evident. Services characterised by high tradability, such as television, sound-recording and book publishing, show the highest local and global agglomeration scores, together with the wholesale of luxury products.

The lowest local agglomeration levels are recorded for the wholesale of agricultural products and the services that are usually predominant in remote regions (Camping grounds and recreational vehicle parks, holiday and other short-stay accommodation). At 180km distance we also find the retail of products characterised by high frequency of consumption. In Appendix (Table A3. 1 and Table A3. 2), I report spatial agglomeration indices for all industries in the two economies, aggregated at the 2-digit industry level using employment weights.

Overall, results are consistent with previous studies (Ellison and Glaeser, 1997; Duranton and Overman, 2005, 2008). However, before moving to the main analysis, it is important to evaluate how the continuous agglomeration index performs with respect to the standard indices used in the urban economics literature. In Table 3. 3, I report the same agglomeration rankings obtained using the Ellison and Glaeser agglomeration index (EGI). The EGI is a discrete measure that normalise spatial agglomeration with the market concentration. The way the index is constructed and its economic meaning are significantly different from the agglomeration measures used in this study. However, industry level results appear very similar. The relation between the two measures can be further appreciated in Table 3. 4, where I analyse the linear relation between the estimates obtained with the two indices. Results confirm that the two measures reach very similar results at the industry level.

3.6 Within-industry agglomeration patters

In this Section I analyse the co-agglomeration patterns that characterise certain plant-types and I compare them with the distribution of overall industries. The analysis separately focusses on the two macro-sectors and is conducted for both French and British plants, to allow cross-country comparisons. For each plat-type, sector and country, I report two kernel density plots, one illustrating the distribution of industry specific agglomeration indices ($\overline{K_J^w}$) and the other showing the distribution of the co-agglomeration indices ($\overline{K_{\iota J,J}^w}$) estimated on the basis of the bilateral distances between type-i plants and all other establishments in the industry.

3.6.1 New entries and exiters

Spatial agglomeration of new-entries and plants that are going to quit the market the following year may result from spatial heterogeneity in turnover rates and/or from mobility patterns in business lifecycle.

Denser markets are generally associated with higher competitive pressure and productivity thresholds. Combes et al. (2012) extend the theoretical framework proposed by Melitz (2003) to justify the higher productivity threshold that characterises denser areas. This explanation, however, might be relevant only for industries characterised by high storability or transport costs. Industries facing negligible costs to ship their products to remote areas of the country should not be affected.

By contrast, the three traditional drivers of agglomeration – labour pooling, input sharing and knowledge spillovers (Duranton & Puga, 2002) – could affect differently entrants and incumbents. First, new firms are often the result of spinoffs from incumbents. Even when this is not the case, a new firm in the start-up phase often relies on the existing business environment. As a result, it is likely to locate in areas with good access to relevant labour markets and suppliers.

Heterogeneous agglomeration patterns may also result from firm lifecycle mobility. Firms that manufacture a new product are more likely to locate in so called 'nursery cities' (Duranton & Puga, 2000), which are large diversified urban areas, where they can 'test' different kinds of production process at a reasonable cost (thanks to localisation economies). Once a firm produces a prototype with its ideal production process and moves to mass-production, internal economies of scale make it less beneficial to remain in diversified areas, and the firms can relocate to the specialised cities, where all firms use the same production process. This second model would suggest that entrants exhibit lower industrial agglomeration patterns than incumbent firms.

The kernel density distributions in Figures 3.15-3.16 compare French entrants' spatial agglomeration to the agglomeration recorded for the sector as a whole. Both manufacturing and service firms exhibit a limited degree of local dispersion with respect to other plants. Moving to the global distribution (Fig. 3.17-3.18), we find a more heterogeneous pattern, with a higher share of plants at the opposite tails of the distribution.

UK plants in both manufacturing and service entrants sector exhibit a limited degree of spatial agglomeration (Fig. 3.19-3.20). This pattern disappears for manufacturing establishments at a

larger distance threshold, whereas service entrants keep exhibiting higher agglomeration than benchmark plants (Fig. 3.21-3.22).
More pronounced differences are found among French exiters, which exhibit a significant local dispersion (Fig. 3.23-3.24), whereas global location patterns are again heterogenous (Fig. 3.25-13.26).
In contrast, UK plants are characterised by limited local agglomeration in the manufacturing sector and global agglomeration in the service sector (Fig. 3.27-30).
Overall, industries and countries do not exhibit clear patterns in the spatial distribution of demographic events.

3.6.2 Large and small plants

In this section I investigate within-industry differences in spatial distribution between plants of different size. In particular, small plants are defined by a size equal to or lower than the 25th percentile of the industry plant-employment distribution recorded over the period. Similarly, establishments with employment equal to or greater than the 10th percentile are categorised as large plants.
Figures 3.31-3.32 demonstrate a clear local dispersion of both small manufacturing and service plants in France. Consistent results are obtained at a 180km distance (Fig. 3.33-3.34). British manufacturing plants exhibit a similar pattern (Fig. 3.35-3.36), whereas this is not the case for the service sector, where no clear pattern is recorded (Fig. 3.37-3.38).

Specular results are found for plants at the opposite tail of the size distribution. Large plants in France demonstrate high local and global agglomeration (Fig. 3.39-3.42). The patterns is partially confirmed by British plants, with the exception of global dispersion in Service sector (Fig. 3.43-3.46).

This spatial configuration could be explained by the trade-off between external economies of scale and proximity. Large manufacturing firms co-agglomerate in dense economic areas, where they can benefit from labour pooling and access to a multitude of suppliers, whereas smaller plants maximise proximity to local demand and lower labour costs.
The different spatial distributions in the service sector could be due to the specific agglomeration pattern that characterises British service plants, largely concentrated in the capital city, as against to the polycentric French system.
Kim (1995) analyses industrial agglomeration in the US, finding a positive correlation between concentration and average plant size per worker. Holmes and Stevens (2002) investigate within-industry differences, finding establishments located in denser areas to be larger than those located in remote regions. Lafourcade and Mion (2007) investigate whether the geographic distribution of manufacturing activities in Italy differs according to the scale of plants. They find a positive relation between size and co-agglomeration, explained by the capability of large plants to serve customers far beyond the boundaries of neighbouring markets. By contrast, small plants co-locate in wider areas, where they can share certain basic input but save in transport cost by locating close to local demand.

3.6.3 Single-plant and multi-plant firms

The industrial economics and trade literatures provide several theoretical frameworks to understand the way offshoring affects the economic environment. However, despite the increasing attention devoted to international fragmentation by media and by researchers, relatively few studies have analysed the domestic fragmentation of plants.
Markusen and Venables (2013) propose a simple theoretical framework to understand the fundamentals of firm fragmentation. According to the model, firms can be 'integrated', operating in one location, or 'fragmented', operating in various locations. Fragmented firms incur additional costs, but they benefit from higher labour productivity, thanks to functional agglomeration economies. Different combinations of fragmentation costs, local endowments and wage elasticities can determine different equilibria, ranging from a fully fragmented to a fully integrated industry. Fort (2013) empirically investigates how coordination and communication costs foster firm fragmentation and, conditionally on fragmentation, how they influence firms' choice to fragment production in the domestic market or offshore. Empirical evidence confirms a clear nexus between electronic communication and domestic fragmentation. Davis and Henderson (2007) analyse the determinants of headquarters agglomerations. Their results demonstrate that separation between production plants and headquarters is particularly beneficial for the firm when the headquarters can access a higher supply of differentiated local service inputs and locate close to other headquarters.

Figures 3.47-3.50 demonstrate that French single-plant French firms are characterised by location patterns identical to or slightly more dispersed than the overall plant distribution. No noticeable pattern is found for UK plants, although in this case both manufacturing and service single-plant firms show a limited degree of agglomeration (Fig. 3.51-3.54). In contrast, significant agglomeration patterns are found for establishment belonging to multi-plant firms in French and British manufacturing sectors and in the French service industries). An opposite pattern is found for British service multi-plant firms, that are noticeably more dispersed in the industrial spatial network (Fig. 3.55-3.62).

In Fig. 3.63-3.66 I focus on French plants, comparing the spatial agglomeration of headquarters only with the overall industry. Like other establishments belonging to multi-plant firms, headquarters demonstrate spatial concentration at both 30 and 180km distance. However, the difference with the benchmark is less significant, especially in the manufacturing sector.

Together these results confirm a tendency of multi-plant firms to collocate single establishments in more central areas of the industrial spatial network. Plant specialisation on their core business is therefore associated with a better access to inputs or local demand. However, this pattern is less evident for manufacturing headquarters, that are often located in dense urban areas, far from production plants.

3.6.4 Top decile of labour productivity

Thus far, I have analysed the way agglomeration patterns vary on the basis of plant-specific characteristics. When it comes to analysing the distribution of plants on the basis of firm-level characteristics, it is worth considering a number of issues that might limit the interpretation of the results. First, assigning the same productivity to different plants means not considering the combinations of input intensity, technologies and local organisation that might characterise plants belonging to the same company. A functional agglomeration of particularly efficient plants belonging to firms with a generally low productivity would not be recorded in this estimation strategy. Second, productivity is often associated with firm size. Larger firms are likely to be more productive, to offer higher wages and to engage in export activities (Altomonte et al., 2010). It follows that the location decision of a few large multi-product plants might affect the whole distribution of productivity over space.

The limitations of the analysis having been pointed out, the next section discusses Figures 3.67-3.74. At a 30km distance, most productive French and British firms are characterised by higher levels of agglomeration with respect to the whole industry. This is true for both the manufacturing and service sectors, with the latter exhibiting the more relevant differences (Fig. 3.67-3.68 and 3.71-3.72).

Most productive firms in France are also agglomerated at a 180 km distance (Fig. 3.69-3.70), while the pattern is less clear in the UK (3.73-3.74), where high performers are more represented at both tails of the distribution.

The empirical evidence could result from confounding factors, as suggested above, or be somewhat related to the micro-foundations of agglomeration economies. Among the three Marshallian forces, knowledge spillovers are generally associated with co-location patterns that characterise most productive and innovative firms. At the industry level, Faggio et al. (2019) find knowledge spillovers to be the main drivers of agglomeration patterns not only in the computer industry but also in less technological advanced sectors, such as the manufacturing of ceramic goods, the manufacturing of cement, lime and plaster and the preparation/spinning of textile fibres.

At the firm level, a significant literature in international trade and industrial economics has provided considerable empirical evidence and theoretical arguments supporting the presence in most industries of a limited group of frontier firms, characterised by larger size, a high degree of internationalisation and high productivity (Mayer & Ottaviano, 2008). My findings seem to support the importance of intra-industry disparities in productivity with respect to the overall spatial distribution of each industry.

3.6.4 Foreign firms

The final section illustrates the spatial distribution of foreign firms. Some studies suggest that location choice of foreign firms would be driven by the opportunity to maximise market and supplier access. For example, Amiti and Javorcik (2008) show that these features largely explain the spatial distribution of foreign firms in China. Even firms facing relatively low transport costs in the foreign countries could choose to locate close to other foreign firms more interested in the local economic environment. Head et al. (1995) examine the location choices of 751 Japanese manufacturing plants in the US, finding evidence of a significant co-location pattern. Bloigen et al. (2005) examine 1485 investment decisions of Japanese firms between 1985 and 2001, demonstrating that firms belonging to invest in the same foreign region.

On the other hand, the evidence provided is primarily based on large countries. It is possible that the limited size and the low transport costs of European countries would sizeably reduce agglomeration externalities for foreign firms. Figures 3.75-3.76 show that both foreign manufacturing and Service plants tend to co-locate at short distances. The pattern is confirmed at a 180 km distance. In contrast, the pattern is less clear-cut in the UK, where foreign manufacturing firms – and to some extent even service ones - are characterised by higher degrees of agglomeration, whereas no difference is found in terms of global agglomeration (Fig. 3.79-3.82).

3.7 Conclusions

In this chapter, I have proposed a composite strategy to analyse the dynamic agglomeration patterns of industries and specific firm types.

Using the continuous agglomeration measures proposed by Duranton and Overman (2006) and extended by Behrens (2017), I have investigated two large micro datasets covering the whole population of French and British establishments from 2008 to 2015. Despite some significant differences, the results are generally consistent, supporting the external validity of the main results. I find significant within-industry heterogeneity in spatial agglomeration. Large plants, and more productive and multi-plant firms are generally more agglomerated, although the difference with the overall distribution varies across sectors and countries.

On the other hand, I do not find specific location patterns for new entries and exiters. This suggests that the general dispersion trend characterising the two sectors might be driven primarily by between-firm reallocation of resources rather than demographic events.

From a methodological perspective, this study demonstrates that continuous agglomeration indices represent a valuable tool in cross-country studies. First, working on continuous space avoids any measurement issue related to scale and aggregation and makes it possible to compare countries characterised by different geographies. Secondly, the possibility to analyse at the same time spatial distributions at different spatial scale allows the researcher to choose a distance threshold that is at once meaningful from an economic point of view and suitable for the characteristics of the territory analysed. Finally, these indices produce valuable results even when applied to the analysis of small population of plants, making it possible to analyse within-firm heterogeneity in spatial location patterns.

There are several directions this work could be taken. First, this methodology should be extended and applied to relevant policy questions (in this regard, a short application is presented in Chapter 4). In particular, it would be interesting to develop a dynamic framework to investigate co-location patterns of different economic actors. For instance, continuous agglomeration measures could be used to study occupational dynamics across industries. Second, this approach can be used to test the external validity of the main result of the agglomeration literature, extending the analysis to more countries (i.e. Spain, Germany and Italy). Finally, work is required to further extend this measure beyond the continuous space dimension to allow them to map more significant economic dimensions, such as time and carbon emissions.

References

Almazan, A., De Motta, A., & Titman, S. (2007). Firm location and the creation and utilization of human capital. *The Review of Economic Studies*, *74*(4), 1305-1327.

Altomonte, Carlo, Alessandro Barattieri, and Armando Rungi (2014). Import penetration, intermediate inputs and productivity: evidence from Italian firms. *Rivista italiana degli economisti*19.1 45-66.

Amiti, M. and Smarzynska Javorcik, B. (2008). Trade costs and location of foreign firms in China, *Journal of Development Economics* 85(1-2): 129-149.

Arbia, G., (1989). Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems. *Kluwer*, Dordrecht.

Arbia, G., (2001). Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, 80, 411-424.

Arzaghi, M., & Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, *75*(4), 1011-1038.

Behrens, K., & Bougna, T. (2015). An anatomy of the geographical concentration of Canadian manufacturing industries. *Regional Science and Urban Economics*, 51, 47-69.

Behrens, K. (2016). Agglomeration and clusters: tools and insights from coagglomeration patterns. *Canadian Journal of Economics/Revue canadienne d'économique*, *49*(4), 1293-1339

Behrens, K. (2016). Agglomeration and clusters: Tools and insights from coagglomeration patterns. Canadian Journal of Economics/Revue canadienne d'économique, 49(4), 1293-1339.

Behrens, K., & Guillain, R. (2017). The determinants of coagglomeration: Evidence from functional employment patterns. *CEPR Discussion Papers* (No. 11884).

Békés, G., Kleinert, J., & Toubal, F. (2009). Spillovers from multinationals to heterogeneous domestic firms: Evidence from Hungary. *World Economy*, *32*(10), 1408-1433.

Békés, G., & Harasztosi, P. (2013). Agglomeration premium and trading activity of firms. *Regional Science and Urban Economics*, 43(1), 51-64.

Briant, A., Combes, P. P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?, *Journal of Urban Economics*, *67*(3), 287-302.

Blonigen, B. A., Ellis, C.J., and Fausten, D. (2005). Industrial Groupings and Foreign Direct Investment, *Journal of International Economics* 65(1): 75-91.

Carlton, D. W. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics* 65: 440-449.

Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity (No. w4313). *National Bureau of Economic Research*

Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, *80*(6), 2543-2594

Dauth, W., Fuchs, M., & Otto, A. (2018). Long-run processes of geographical concentration and dispersion: Evidence from Germany. *Papers in Regional Science*, *97*(3), 569-593.

Duranton, Gilles and Overman, H. (2005) Testing for localization using micro-geographic data. *The Review of Economic Studies* 72.4 (2005): 1077-1106.

Duranton, Gilles and Overman, H. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science* 48.(1), 213-243.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier

Ellison, G. and Glaeser, E.L. (1997). Geographic concentration in US manufacturing industries: a dartboard approach. Journal of political economy, 105(5), pp.889-927

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), 1195-1213

Faggio, G, Silva, O, WC Strange (2014). Heterogeneous agglomeration, *Review of Economics and Statistics*

Faggio, Giulia, Olmo Silva, and William C. Strange (2019). Tales of the City: What Do Agglomeration Cases Tell Us About Agglomeration in General? *Working Papers 19/10*, City University London.

Fallick, B., Fleischman, C., & Rebitzer, J. (2007). Job Hopping in Silicon Valley: The Microfoundations of a High Tech Industrial Cluster. *Review of Economics and Statistics*.

Fort, T. C. (2012). Firms' Organization of Global Production: Theory and Evidence (Doctoral dissertation).

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy*, *100*(6), 1126-1152.

Head, Keith, John Ries, and Deborah Swenson (1995). Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the US, *Journal of International Economics* 38(3-4): 223-247.

Henderson, J. V. (1986). Efficiency of resource usage and city size. *Journal of Urban economics*, *19*(1), 47-70

Henderson, J. V. (1994). *Externalities and industrial development* (No. w4730). National Bureau of Economic Research.

Henderson, J. V., & Ono, Y. (2008). Where do manufacturing firms locate their headquarters?. *Journal of Urban Economics*, *63*(2), 431-450

Lafourcade, M., & Mion, G. (2007). Concentration, agglomeration and the size of plants. *Regional Science and Urban Economics*, *37*(1), 46-68

Lin, J. (2012), Technological adaptation, cities, and new work, *Review of Economics and Statistics* 93, 554 - 574

Marcon, E., & Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, *3*(4), 409-428

Marcon, E., & Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, *62*, 56-67

Markusen, J. R., & Venables, A. J. (2013). *Functional specialization, sectoral specialization, and inter-city trade* (Doctoral dissertation, *The European Trade Study Group Working paper*)

Marshall, A. (1890). *Principles of economics. Vol. 1*.

Mayer, T., & Ottaviano, G. I. (2008). The happy few: The internationalisation of european firms. *Intereconomics*, *43*(3), 135-148

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*(6), 1695-1725

Nakamura, R. (1985). Agglomeration economies in urban manufacturing industries: A case of Japanese cities. *Journal of Urban Economics*, 17(1):108-124.

Openshaw, S. and P. J. Taylor, 1979. A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In N. Wrigley, ed. *Statistical Applications in the Spatial Sciences*, 127–144. London: Pion.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, *13*(2), 255-266

Rosenthal, S. S., & Strange, W. C. (2010). Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. In *Agglomeration economics* (pp. 277-302). University of Chicago Press

Strange, W., Hejazi, W., & Tang, J. (2006). The uncertain city: competitive instability, skills, innovation and the strategy of agglomeration. *Journal of Urban Economics*, *59*(3), 331-351

Sveikauskas, L. (1975). The productivity of cities. *The Quarterly Journal of Economics*, *89*(3), 393-413.

Tables

Table 3. 1: Most and least localised manufacturing industries, by country and distance threshold

	DO, 30km				DO, 180km			
	France		United Kingdom		France		United Kingdom	
Most agglomerated	1413 - Manufacture of other outerwear	0.2139	2341 - Manufacture of ceramic household and ornamental articles	0.4894	3012 - Building of pleasure and sporting boats	0.5704	2341 - Manufacture of ceramic household and ornamental articles	0.8175
	2630 - Manufacture of communication equipment	0.1985	1310 - Preparation and spinning of textile fibres	0.3257	2053 - Manufacture of essential oils	0.4634	1414 - Manufacture of underwear	0.769
	2053 - Manufacture of essential oils	0.1805	2013 - Manufacture of other inorganic basic chemicals	0.2181	2042 - Manufacture of perfumes and toilet preparations	0.4532	2594 - Manufacture of fasteners and screw machine products	0.7644
	3212 - Manufacture of jewellery and related articles	0.14928	1820 - Reproduction of recorded media	0.21335	2319 - Manufacture and processing of glass, including technical glassware	0.4336	1393 - Manufacture of carpets and rugs	0.7637
	1102 - Manufacture of wine from grape	0.1343	1520 - Manufacture of footwear	0.2128	2313 - Manufacture of hollow glass	0.3817	2451 - Casting of iron	0.7065
Least agglomerated	1052 - Manufacture of ice cream	0.0071	1052 - Manufacture of ice cream	0.019	1086 - Manufacture of homogenised food preparations and dietetic food	0.1374	3315 - Repair and maintenance of ships and boats	0.2392
	1610 - Sawmilling and planning of wood	0.0088	1610 - Sawmilling and planning of wood	0.019	2370 - Cutting, shaping and finishing of stone	0.1403	1020 - Processing and preserving of fish, crustaceans and molluscs	0.2518
	2511 - Manufacture of structural metal products	0.0104	3011 - Building of ships and floating structures	0.0205	2512 - Manufacture of doors and windows of metal	0.1423	1712 - Manufacture of paper and paperboard	0.328
	2593 - Manufacture of wire products, chain and springs	0.011	1623 - Manufacture of other builders' carpentry and joinery	0.0231	1082 - Manufacture of cocoa, chocolate and sugar confectionery	0.1436	2670 - Manufacture of optical instruments and photographic equipment	0.3288
	2361 - Manufacture of concrete products for construction purposes	0.0112	2611 - Manufacture of electronic components	0.0237	2363 - Manufacture of ready-mixed concrete	0.145	1610 - Sawmilling and planning of wood	0.333

Table 3. 2: Most and least localised service industries, by country and distance threshold

	DO, 30km				DO, 180km			
	France		United Kingdom		France		United Kingdom	
Most agglomerated	5911 - Motion picture, video and television programme production activities	0.656	5913 - Motion picture, video and television programme distribution activities	0.3065	5912 - Motion picture, video and television programme post-production activities	0.7646	5912 - Motion picture, video and television programme post-production activities	0.6888
	5912 - Motion picture, video and television programme post-production activities	0.6178	5920 - Sound recording and music publishing activities	0.2908	5811 - Book publishing	0.7037	5913 - Motion picture, video and television programme distribution activities	0.6755
	5814 - Publishing of journals and periodicals	0.6076	5911 - Motion picture, video and television programme production activities	0.2903	5814 - Publishing of journals and periodicals	0.6987	5920 - Sound recording and music publishing activities	0.6686
	5811 - Book publishing	0.5776	6020 - Television programming and broadcasting activities	0.2799	5920 - Sound recording and music publishing activities	0.694	7320 - Market research and public opinion polling	0.6345
	5920 - Sound recording and music publishing activities	0.594	4648 - Wholesale of watches and jewellery	0.205	6391 - News agency activities	0.6758	6202 - Computer consultancy activities	0.5987
Least agglomerated	4661 - Wholesale of agricultural machinery, equipment and supplies	0.0075	5530 - Camping grounds, recreational vehicle parks and trailer parks	0.0152	4674 - Wholesale of hardware, plumbing and heating equipment and supplies	0.1445	5520 - Holiday and other short-stay accommodation	0.2615
	4621 - Wholesale of grain, unmanufactured tobacco, seeds and animal feeds	0.0102	5520 - Holiday and other short-stay accommodation	0.0169	4764 - Retail sale of sporting equipment in specialised stores	0.1464	5530 - Camping grounds, recreational vehicle parks and trailer parks	0.2728
	7500 - Veterinary activities	0.0154	4621 - Wholesale of grain, unmanufactured tobacco, seeds and animal feeds	0.0174	4774 - Retail sale of medical and orthopaedic goods in specialised stores	0.151	4723 - Retail sale of fish, crustaceans and molluscs in specialised stores	0.2842
	4677 - Wholesale of waste and scrap	0.0163	4661 - Wholesale of agricultural machinery, equipment and supplies	0.0177	4532 - Retail trade of motor vehicle parts and accessories	0.1513	5222 - Service activities incidental to water transportation	0.2908
	4519 - Sale of other motor vehicles	0.0169	4623 - Wholesale of live animals	0.0183	4726 - Retail sale of tobacco products in specialised stores	0.152	5010 - Sea and coastal passenger water transport	0.2955

Table 3. 3: Five most and least localised manufacturing industries, by country and distance threshold (EGI index)

	EGI Zone d'emploi							
	Manufacturing sector				Service Sector			
	France		United Kingdom		France		United Kingdom	
Most agglomerated	2053 - Manufacture of essential oils	0.1793	2341 - Manufacture of ceramic household and ornamental articles	0.456	5912 - Motion picture, video and television programme post-production activities	0.3858	5913 - Motion picture, video and television programme distribution activities	0.3895
	1101 - Distilling, rectifying and blending of spirits	0.1353	1310 - Preparation and spinning of textile fibres	0.2132	5911 - Motion picture, video and television programme production activities	0.3398	5912 - Motion picture, video and television programme post-production activities	0.2842
	1439 - Manufacture of other knitted and crocheted apparel	0.1232	1439 - Manufacture of other knitted and crocheted apparel	0.169	5110 - Passenger air transport	0.336	6391 - News agency activities	0.2825
	2451 - Casting of iron	0.0992	1101 - Distilling, rectifying and blending of spirits	0.1078	5814 - Publishing of journals and periodicals	0.2663	5920 - Sound recording and music publishing activities	0.246
	1399 - Manufacture of other textiles n.e.c.	0.0906	1520 - Manufacture of footwear	0.1028	6020 - Television programming and broadcasting activities	0.2465	6020 - Television programming and broadcasting activities	0.2416
Least agglomerated	2051 - Manufacture of explosives	-0.143	2110 - Manufacture of basic pharmaceutical products	-0.004	7721 - Renting and leasing of recreational and sports goods	-0.012	4638 - Wholesale of other food, including fish, crustaceans and molluscs	-0.005
	2314 - Manufacture of glass fibres	-0.115	1107 - Manufacture of soft drinks; production of mineral waters and other bottled waters	-0.001	4636 - Wholesale of sugar and chocolate and sugar confectionery	-0.009	7430 - Translation and interpretation activities	-0.004
	2620 - Manufacture of computers and peripheral equipment	-0.004	1812 - Other printing	0.003	4613 - Agents involved in the sale of timber and building materials	-0.004	5914 - Motion picture projection activities	-0.002
	2740 - Manufacture of electric lighting equipment	-0.003	2059 - Manufacture of other chemical products n.e.c.	0.0035	6010 - Radio broadcasting	-0.004	4665 - Wholesale of office furniture	-0.002
	2041 - Manufacture of soap and detergents, cleaning and polishing preparations	-0.001	1813 - Pre-press and pre-media services	0.0047	5813 - Publishing of newspapers	-0.003	4743 - Retail sale of audio and video equipment in specialised stores	-0.001

Table 3. 4: DO and EGI agglomeration indices

Dependent variable: $\overline{\overline{K_J^w}}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	30km		180km		30km		180km		30km		180km	
VARIABLES	Commune	Zone d'emploi	Commune	Zone d'emploi	Commune	Zone d'emploi	Commune	Zone d'emploi	Commune	Zone d'emploi	Commune	Zone d'emploi
EGI	2.957***	1.432***	3.400***	1.497***	2.957***	1.432***	3.400***	1.497***	0.258***	0.465***	0.552***	0.674***
	(0.302)	(0.0477)	(0.280)	(0.0365)	(0.302)	(0.0477)	(0.280)	(0.0365)	(0.0775)	(0.0794)	(0.126)	(0.0858)
Constant	0.0823***	0.0606***	0.239***	0.218***	0.0823***	0.0606***	0.239***	0.218***	0.0971***	0.0862***	0.255***	0.240***
	(0.00231)	(0.00200)	(0.00273)	(0.00255)	(0.00231)	(0.00200)	(0.00273)	(0.00255)	(0.000456)	(0.00206)	(0.000720)	(0.00222)
Observations	2,400	2,400	2,400	2,400	2,400	2,400	2,400	2,400	2,400	2,400	2,400	2,400
R-squared	0.151	0.388	0.153	0.324	0.151	0.388	0.153	0.324	0.990	0.991	0.986	0.988
Industry FE	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes. The table reports the regression for a simple linear regression model testing the relation between the Duranton and Overman agglomeration index, measured at different distant thresh
Standard error are clustered at the 2-digit industry level. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Graphs

Figure 3. 7: Manufacturing local agglomeration trends, France

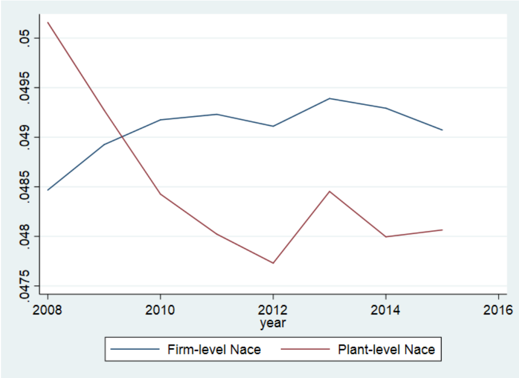
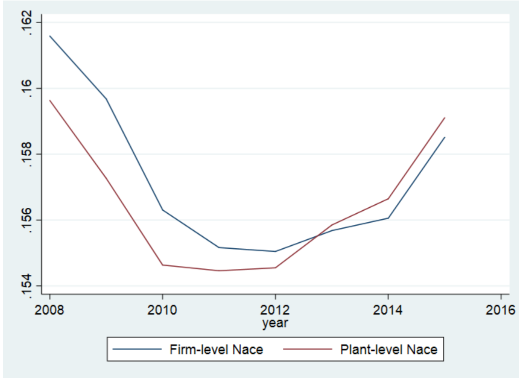


Figure 3. 8: Manufacturing local agglomeration trends, UK



Notes. The graphs show the evolution of the Weighted mean $\overline{\overline{K_J^w}}$, computed at a 30 km distance for all manufacturing industries

Figure 3. 9: Service local agglomeration trends, France

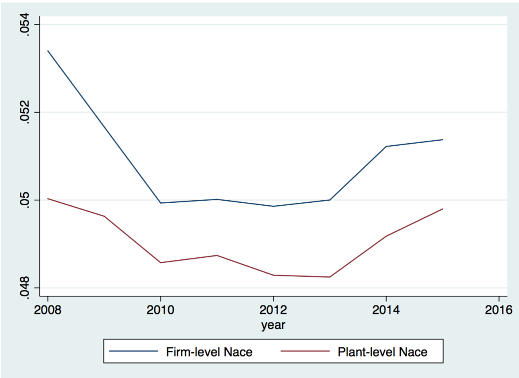
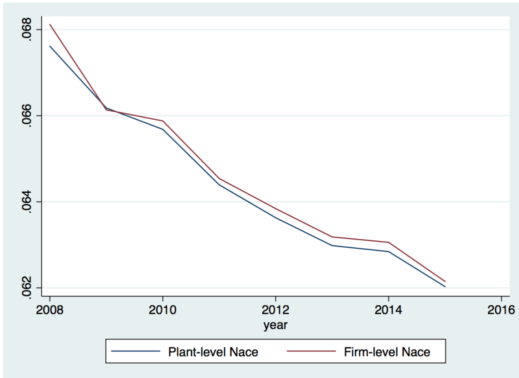


Figure 3. 10: Service local agglomeration trends, UK



Notes. The graphs show the evolution of the Weighted mean $\overline{\overline{K_J^w}}$, computed at a 30 km distance for all service industries

Figure 3. 11: Manufacturing global agglomeration trends, France

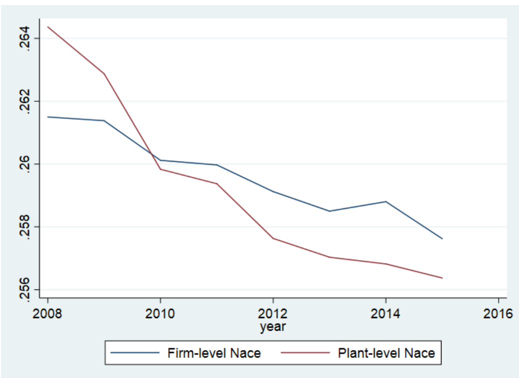
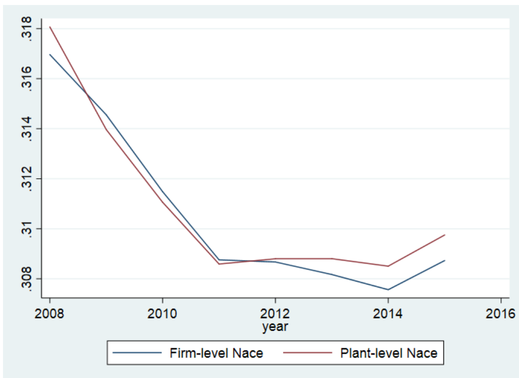


Figure 3. 12: Manufacturing global agglomeration trends, UK



Notes. The graphs show the evolution of the Weighted mean $\overline{\overline{K_J^w}}$, computed at a 180 km distance for all manufacturing industries

Figure 3. 13: Service global agglomeration trends, France

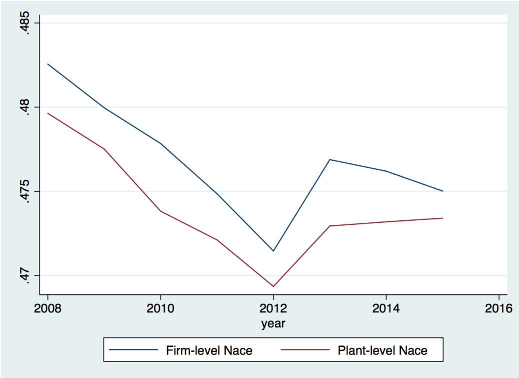
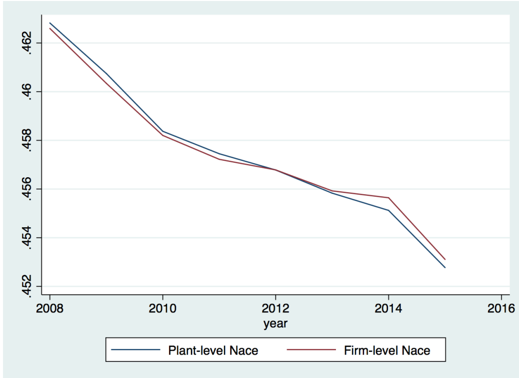


Figure 3. 14: Service global agglomeration trends, UK



Notes. The graphs show the evolution of the Weighted mean $\overline{\overline{K_j^w}}$, computed at a 180 km distance for all service industries

New entries – France

Figure 3. 15: New entries, Manufacturing, d=30km, France

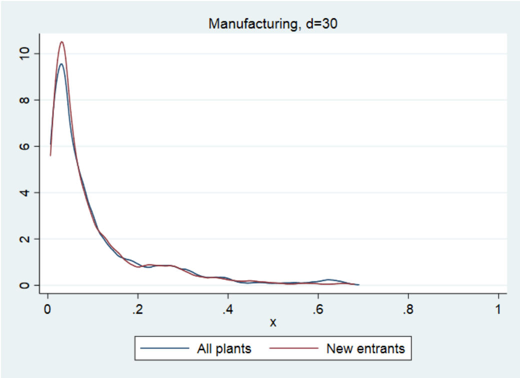
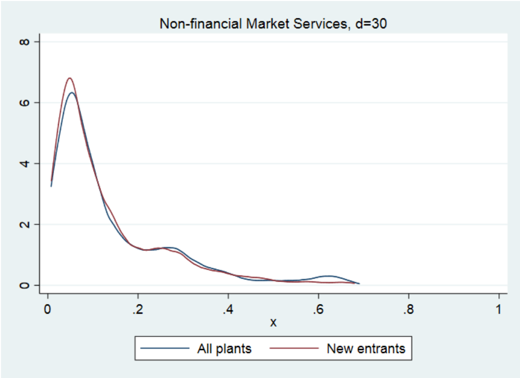


Figure 3. 16: New entries, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between new entries and all other plants in the industry.

Figure 3. 17: New entries, Manufacturing, d=180km, France

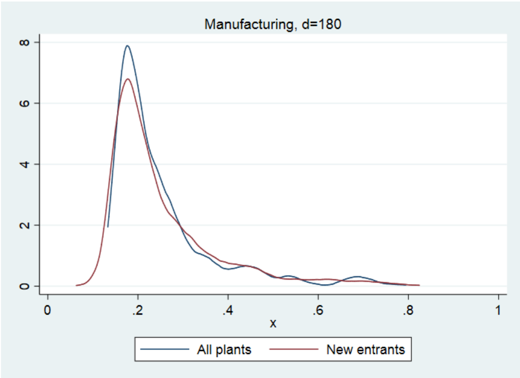
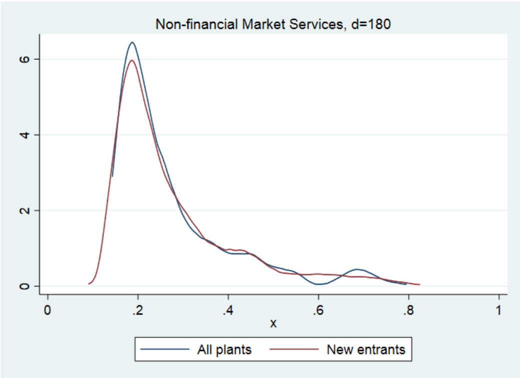


Figure 3. 18: New entries, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between new entries and all other plants in the industry.

New entries– United Kingdom

Figure 3. 19: New entries, Manufacturing, d=30km, UK

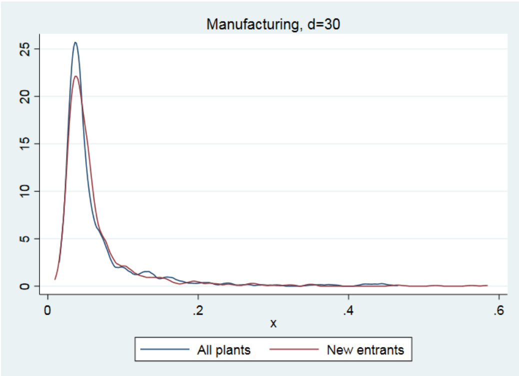
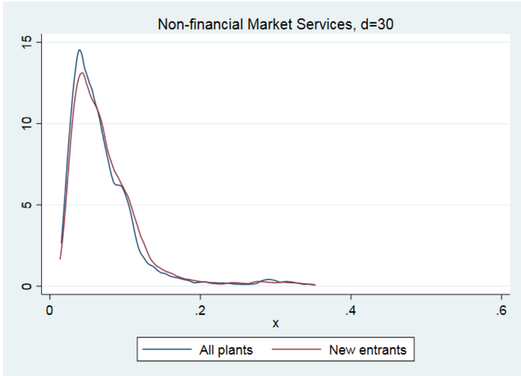


Figure 3. 20: New entries, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between new entries and all other plants in the industry.

Figure 3. 21: New entries, Manufacturing, d=180km, UK

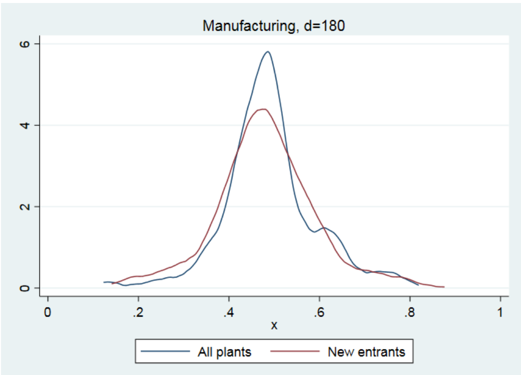
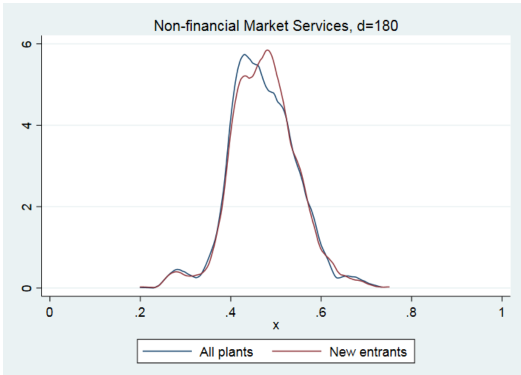


Figure 3. 22: New entries, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between new entries and all other plants in the industry.

Exiters – France

Figure 3. 23: Exiters, Manufacturing, d=30km, France

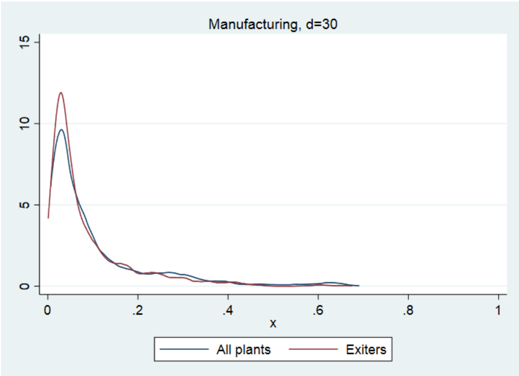
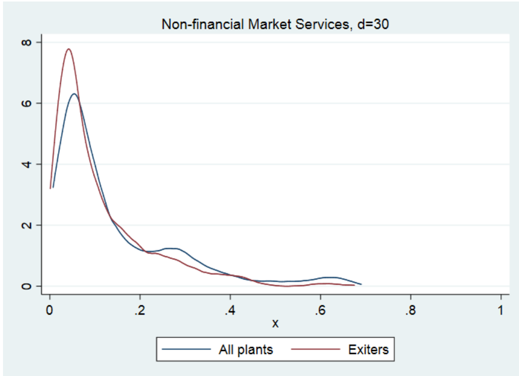


Figure 3. 24: Exiters, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants that exit the market and all other plants in the industry.

Figure 3. 25: Exiters, Manufacturing, d=180km, France

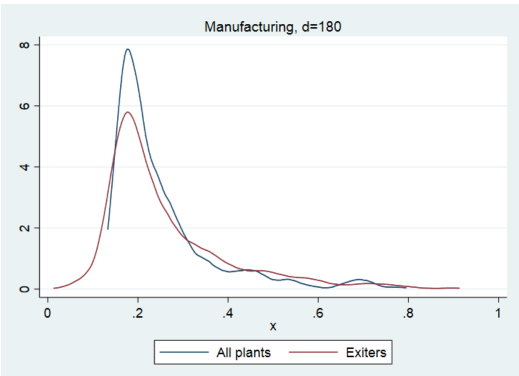
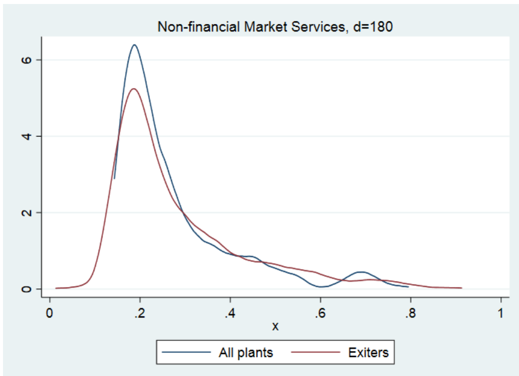


Figure 3. 26: Exiters, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants that exit the market and all other plants in the industry.

Exiters – United Kingdom

Figure 3. 27: Exiters, Manufacturing, d=30km, UK

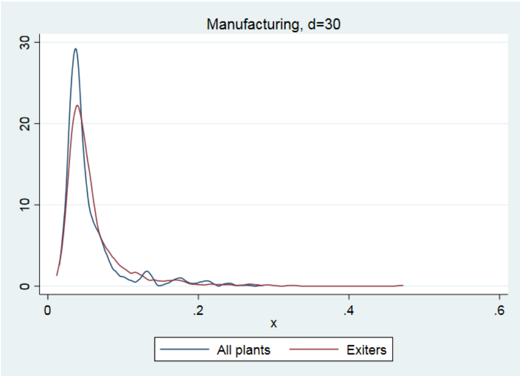
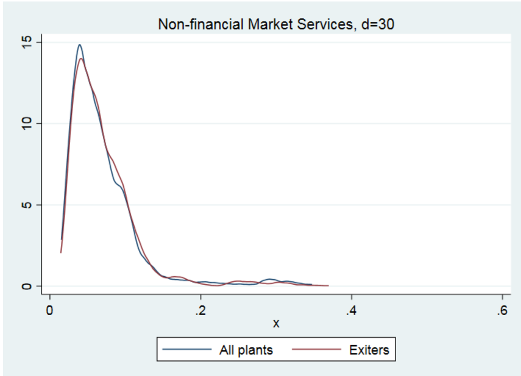


Figure 3. 28: Exiters, Manufacturing, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants that exit the market and all other plants in the industry.

Figure 3. 29: Exiters, Manufacturing, d=180km, UK

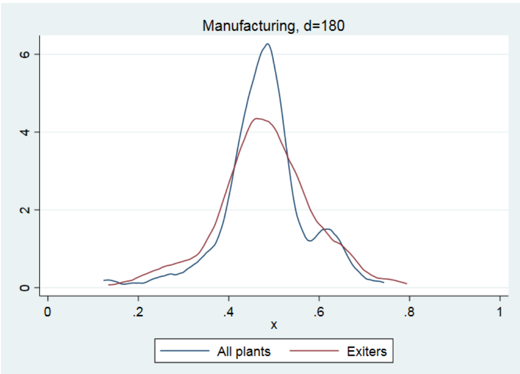
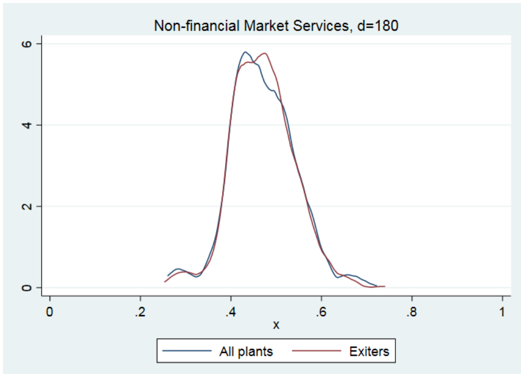


Figure 3. 30: Exiters, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants that exit the market and all other plants in the industry.

Small plants – France

Figure 3. 31: Small plants, Manufacturing, d=30km, France

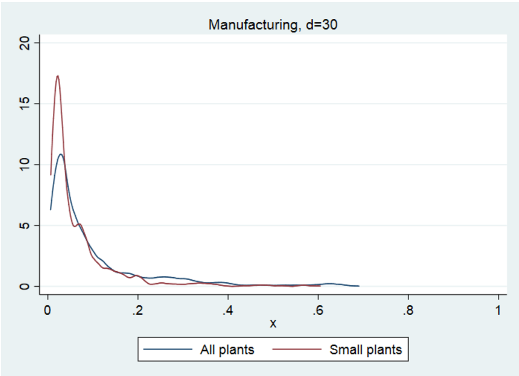
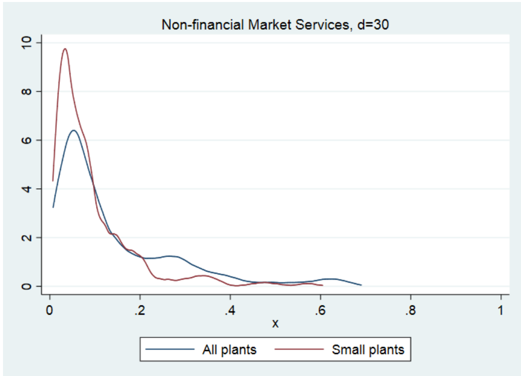


Figure 3. 32: Small plants, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between small plants (below the 25th percentile of the employment distribution) and all other plants in the industry.

Figure 3. 33: Small plants, Manufacturing, d=180km, France

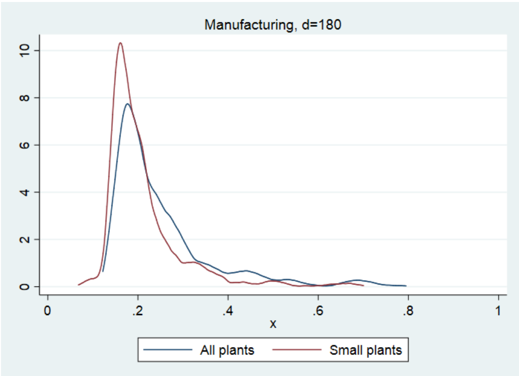
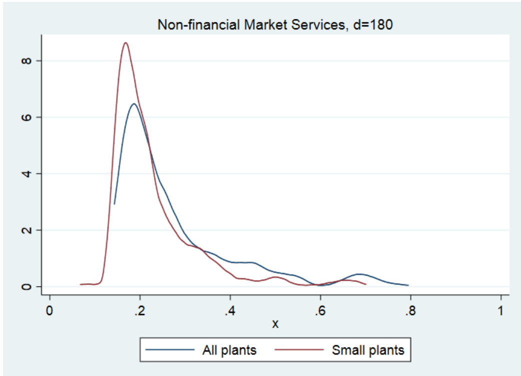


Figure 3. 34: Small plants, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between small plants (below the 25th percentile of the employment distribution) and all other plants in the industry.

Small plants – United Kingdom

Figure 3. 35: Small plants, Manufacturing, d=30km, UK

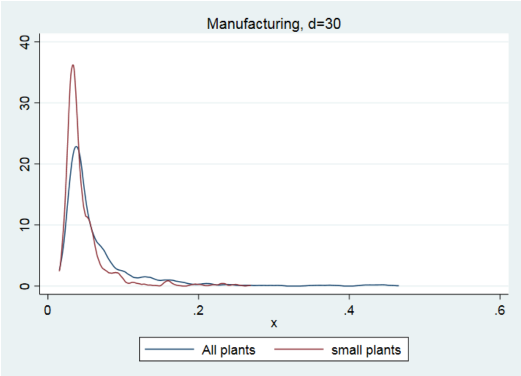
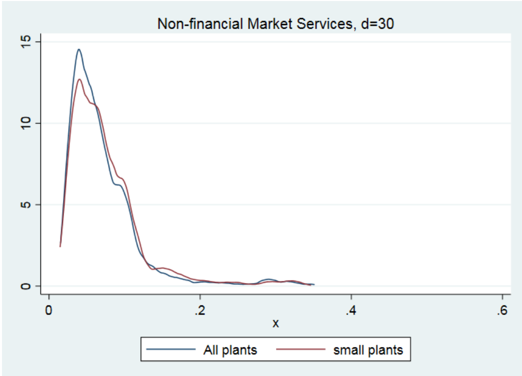


Figure 3. 36: Small plants, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between small plants (below the 25th percentile of the employment distribution) and all other plants in the industry.

Figure 3. 37: Small plants, Manufacturing, d=180km, UK

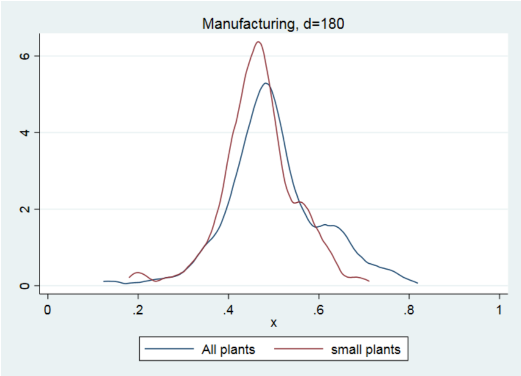
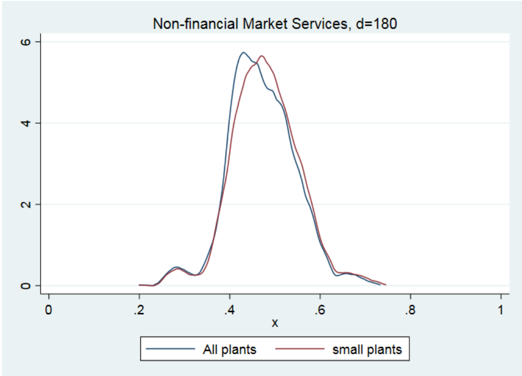


Figure 3. 38: Small plants, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between small plants (below the 25th percentile of the employment distribution) and all other plants in the industry.

Large plants – France

Figure 3. 39: Large plants, Manufacturing, d=30km, France

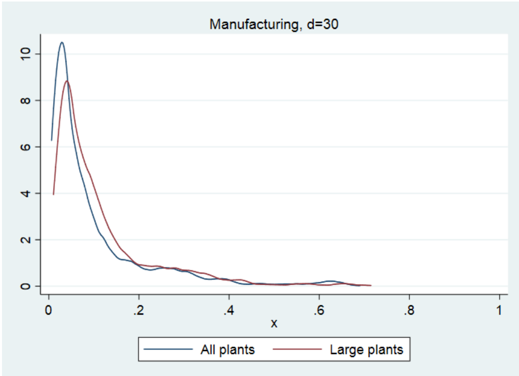
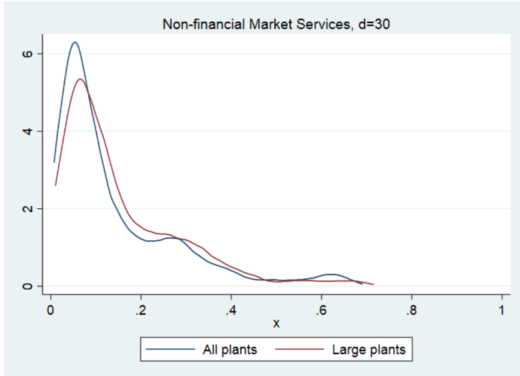


Figure 3. 40: Large plants, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between large plants (above the 90^{th} percentile of the employment distribution) and all other plants in the industry.

Figure 3. 41: Large plants, Manufacturing, d=180km, France

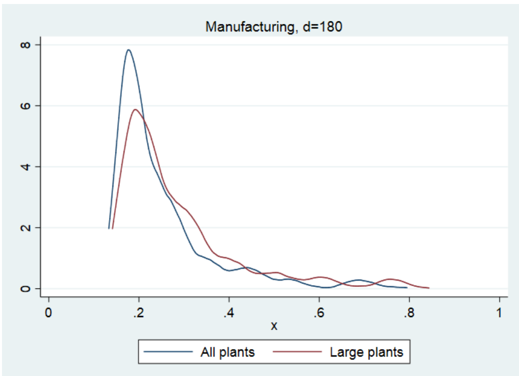
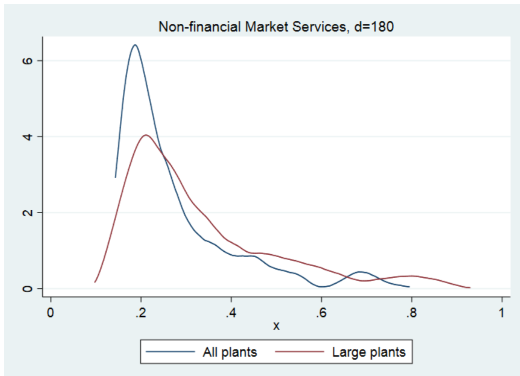


Figure 3. 42: Large plants, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between large plants (above the 90^{th} percentile of the employment distribution) and all other plants in the industry.

Large plants – United Kingdom

Figure 3. 43: Large plants, Manufacturing, d=180km, UK

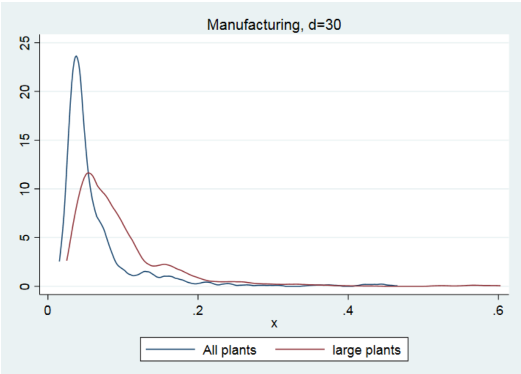
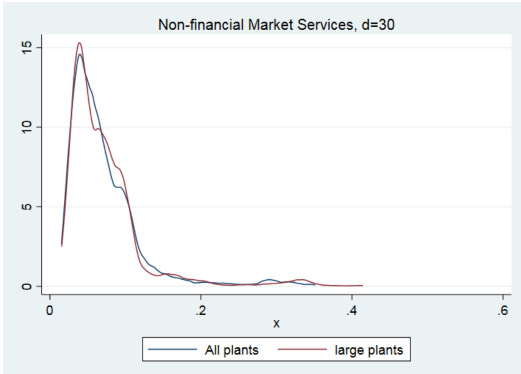


Figure 3. 44: Large plants, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between large plants (above the 90th percentile of the employment distribution) and all other plants in the industry.

Figure 3. 45: Large plants, Manufacturing, d=180km, UK

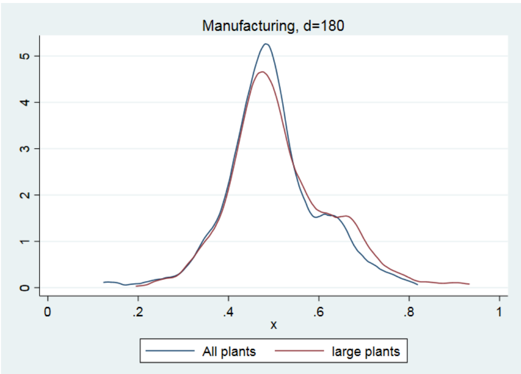
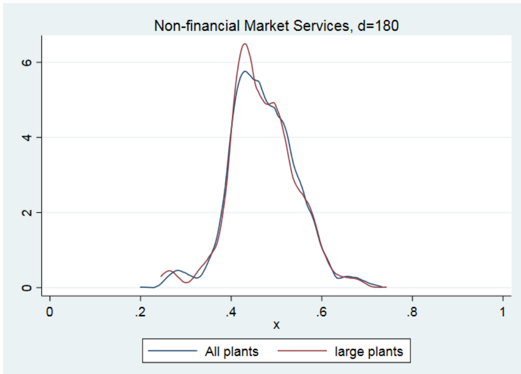


Figure 3. 46: Large plants, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between large plants (above the 90th percentile of the employment distribution) and all other plants in the industry.

Single-plant firms – France

Figure 3. 47: Single-plant firms, Manufacturing, d=30km, France

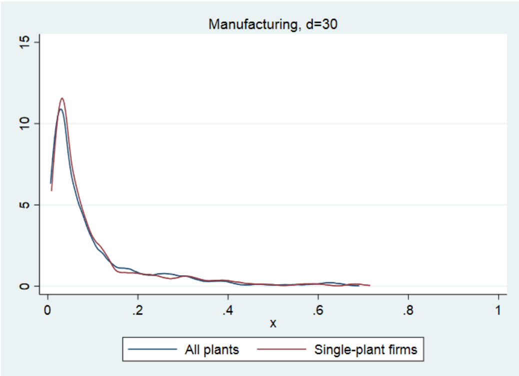
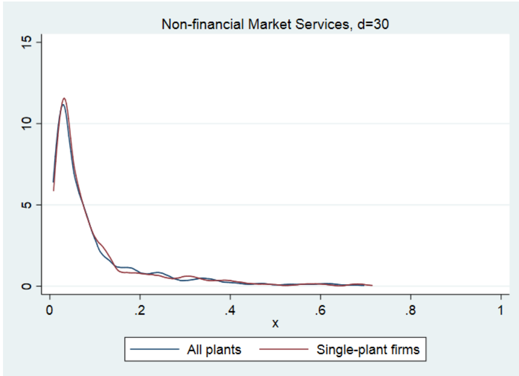


Figure 3. 48: Single-plant firms, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between single-plant firms and all other plants in the industry.

Figure 3. 49: Single-plant firms, Manufacturing, d=180km, France

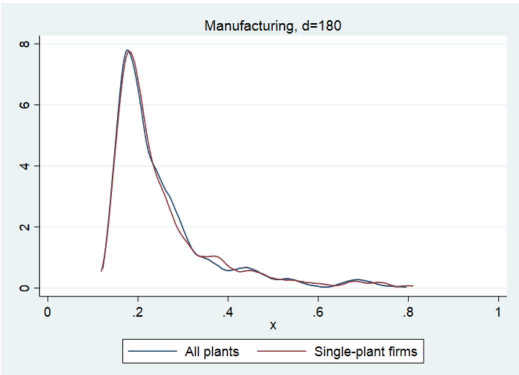
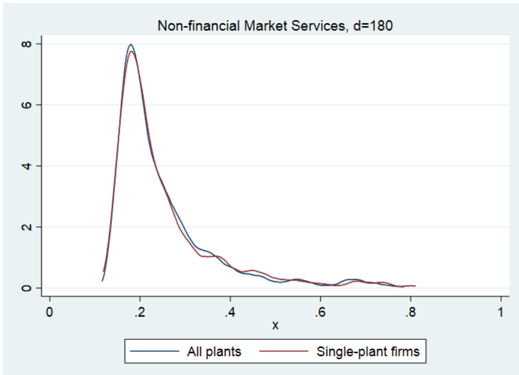


Figure 3. 50: Single-plant firms, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between single-plant firms and all other plants in the industry.

Single-plant firms – United Kingdom

Figure 3. 51: Single-plant firms, Manufacturing, d=30km, UK

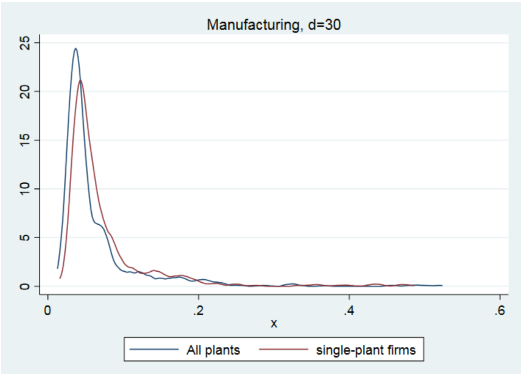
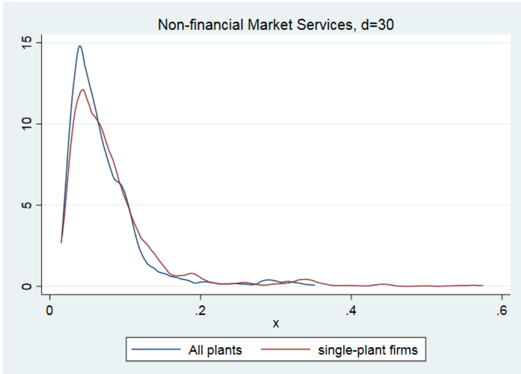


Figure 3. 52: Single-plant firms, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between single-plant firms and all other plants in the industry.

Figure 3. 53: Single-plant firms, Manufacturing, d=180km, UK

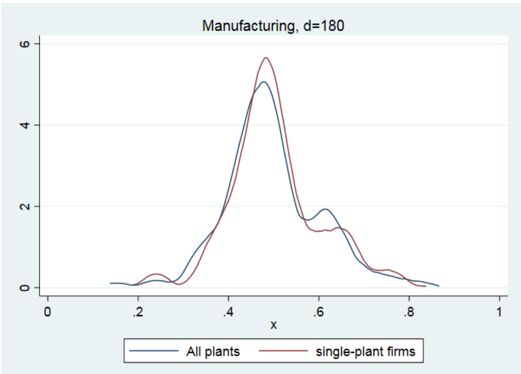
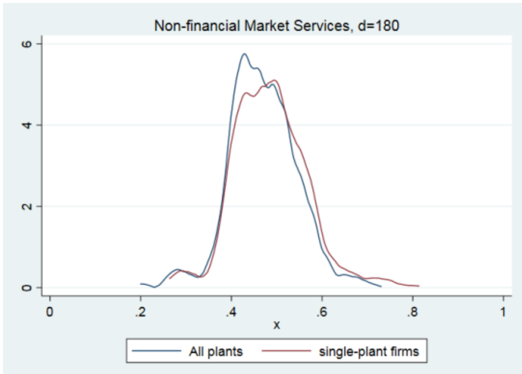


Figure 3. 54, Single-plant firms, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between single-plant firms and all other plants in the industry.

Multi-plant firms – France

Figure 3. 55: Multi-plant firms, Manufacturing, d=30km, France

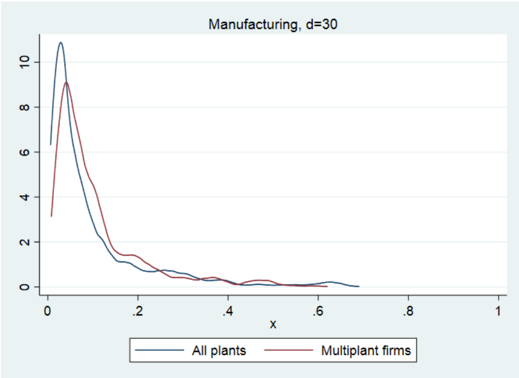
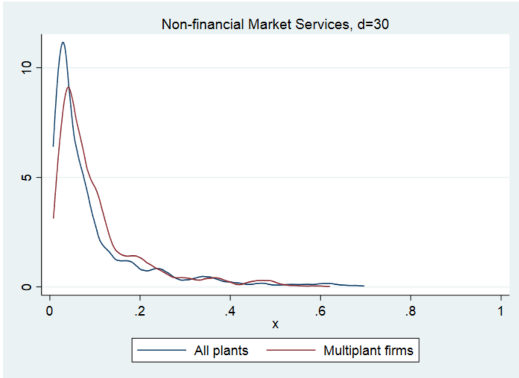


Figure 3. 56: Multi-plant firms, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants belonging to multi-plant firms and all other plants in the industry.

Figure 3. 57: Multi-plant firms, Manufacturing, d=180km, France

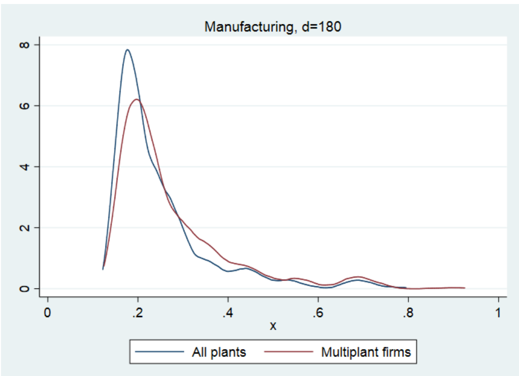
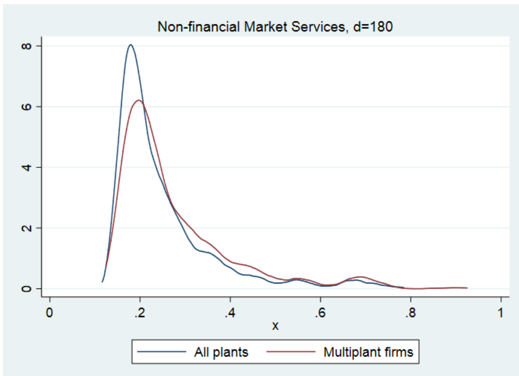


Figure 3. 58: Multi-plant firms, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants belonging to multi-plant firms and all other plants in the industry.

Figure 3. 59: Multi-plant firms, Manufacturing, d=30km, UK

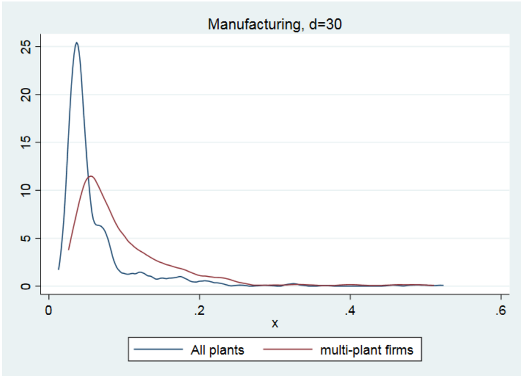
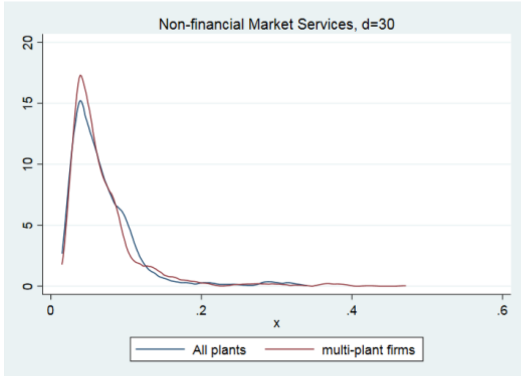


Figure 3. 60: Multi-plant firms, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants belonging to multi-plant firms and all other plants in the industry.

Figure 3. 61: Multi-plant firms, Manufacturing, d=180km, UK

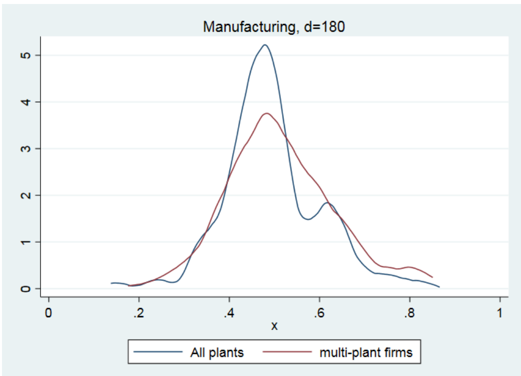
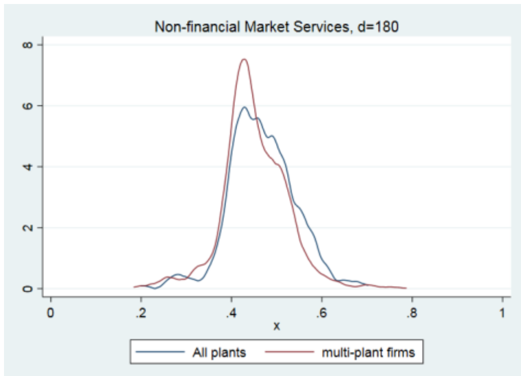


Figure 3. 62: Multi-plant firms, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between plants belonging to multi-plant firms and all other plants in the industry.

Headquarters – France

Figure 3. 63: Headquarters, Manufacturing, d=30km, France

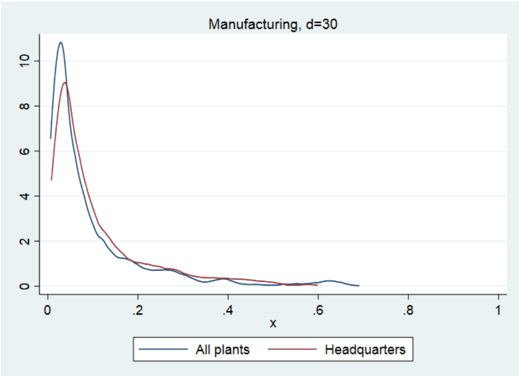
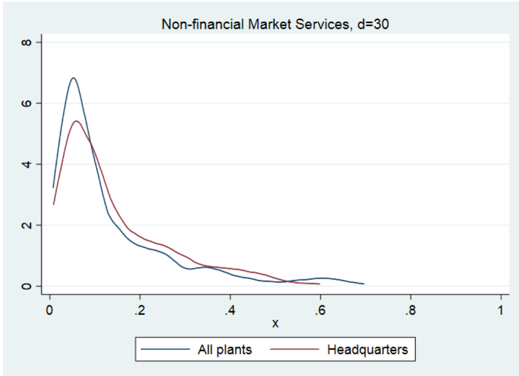


Figure 3. 64: Headquarters, Manufacturing, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between headquarters and all other plants in the industry.

Figure 3. 65: Headquarters, Manufacturing, d=180km, France

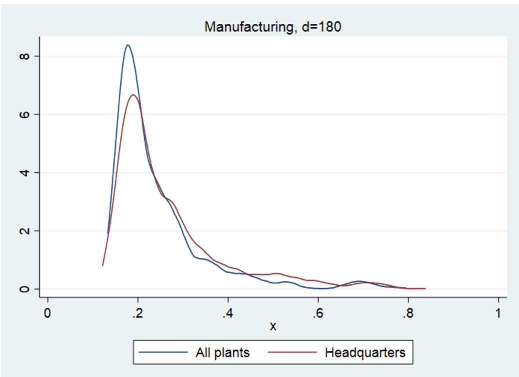
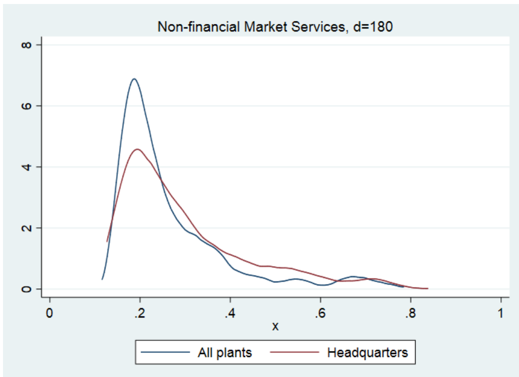


Figure 3. 66: Headquarters, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between headquarters and all other plants in the industry.

High-productive firms – France

Figure 3. 67: High productive firms, Manufacturing, d=30km, France

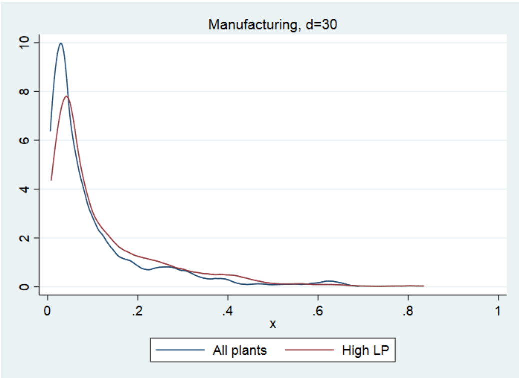
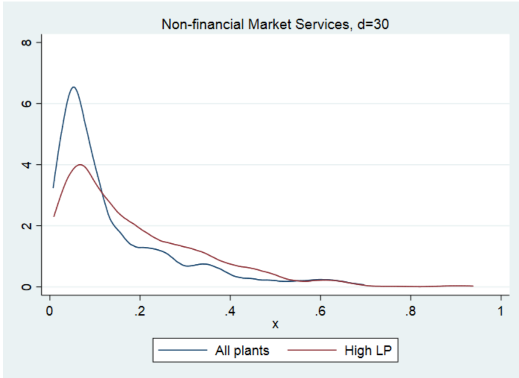


Figure 3. 68: High productive firms, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between high productive firms (above the 90th percentile of productivity distribution) and all other plants in the industry.

Figure 3. 69: High productive firms, Manufacturing, d=180km, France

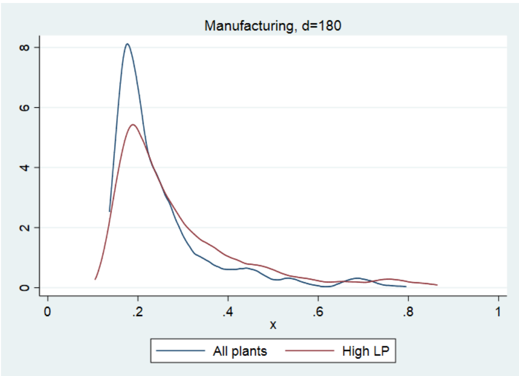
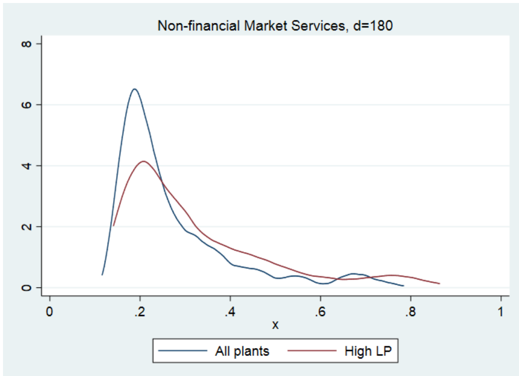


Figure 3. 70: High productive firms, Service, d=180km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between high productive firms (above the 90th percentile of productivity distribution) and all other plants in the industry.

High-productive firms – United Kingdom

Figure 3. 71: High productive firms, Manufacturing, d=30km, UK

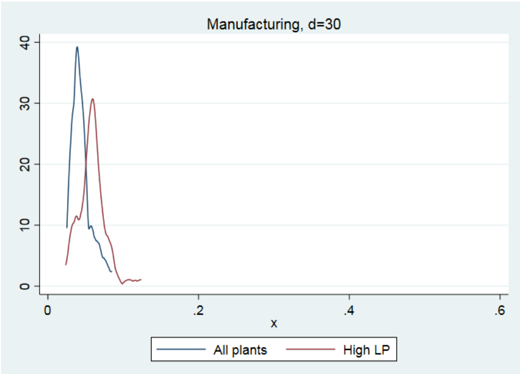
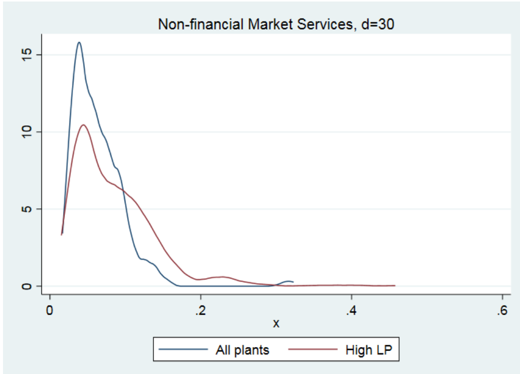


Figure 3. 72: High productive firms, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between high productive firms (above the 90th percentile of productivity distribution) and all other plants in the industry.

Figure 3. 73: High productive firms, Manufacturing, d=180km, UK

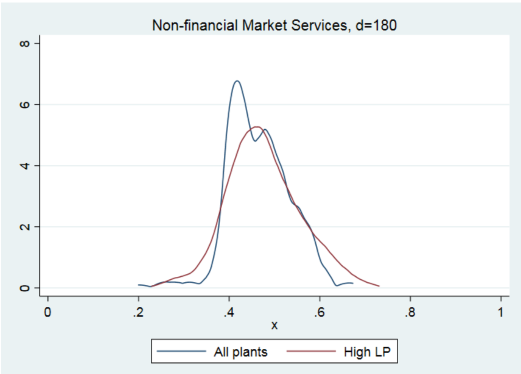
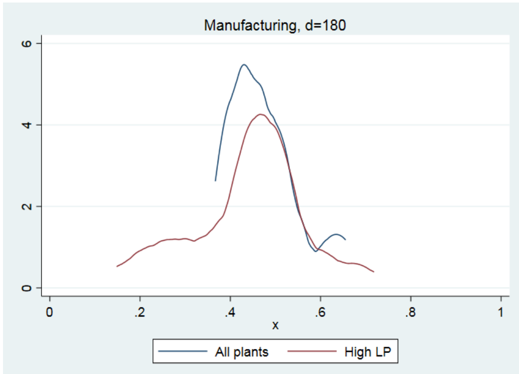


Figure 3. 74: High productive firms, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances between high productive firms (above the 90th percentile of productivity distribution) and all other plants in the industry.

Foreign firms – France

Figure 3. 75: Foreign firms, Manufacturing, d=30km, France

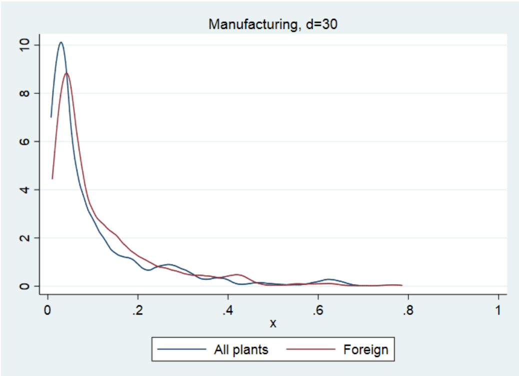
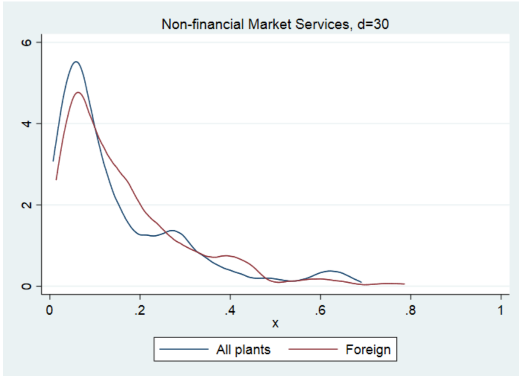


Figure 3. 76: Foreign firms, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances plants belonging to foreign firms and all other plants in the industry.

Figure 3. 77: Foreign firms, Manufacturing, d=180km, France

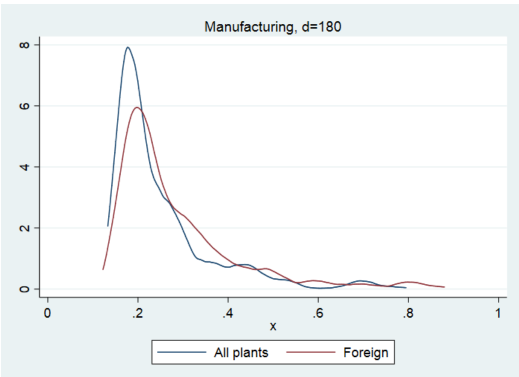
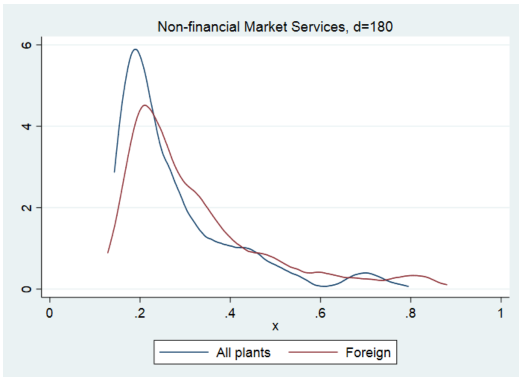


Figure 3. 78: Foreign firms, Service, d=30km, France



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances plants belonging to foreign firms and all other plants in the industry.

Foreign firms – United Kingdom

Figure 3. 79: Foreign firms, Manufacturing, d=30km, UK

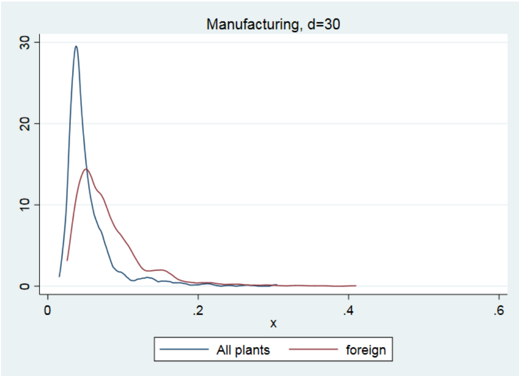
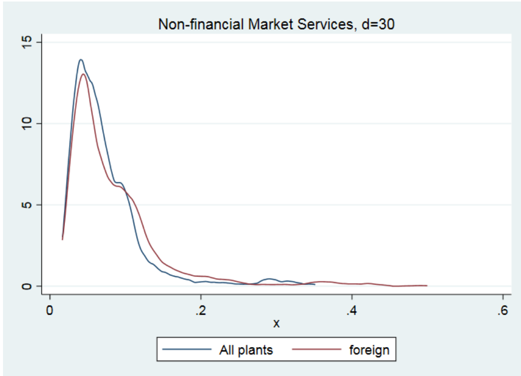


Figure 3. 80: Foreign firms, Service, d=30km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances plants belonging to foreign firms and all other plants in the industry.

Figure 3. 81: Foreign firms, Manufacturing, d=180km, UK

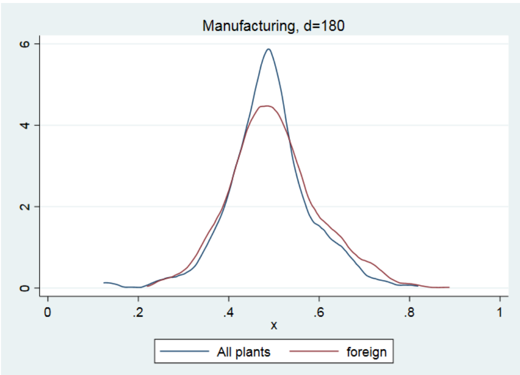
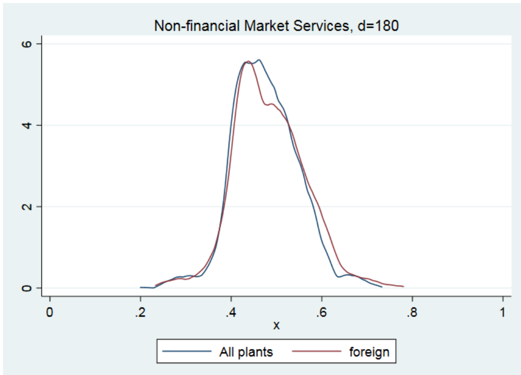


Figure 3. 82: Foreign firms, Service, d=180km, UK



Notes. The graphs show the distribution of industry-specific agglomeration indices and the distribution of co-agglomeration indices, based on bilateral distances plants belonging to foreign firms and all other plants in the industry.

Appendix

Table A3. 1: Agglomeration by 2-digit service industry

30km				180km			
France		UK		France		UK	
Industry group	Kd	Industry group	Kd	Industry group	Kd	Industry group	Kd
75 - Veterinary activities	0.0154	55 - Accommodation	0.0268	75 - Veterinary activities	0.1454	55 - Accommodation	0.3058
45 - Wholesale and retail trade and repair of motor vehicles and motorcycles	0.0256	75 - Veterinary activities	0.0269	45 - Wholesale and retail trade and repair of motor vehicles and motorcycles	0.1667	50 - Water transport	0.3621
47 - Retail trade, except of motor vehicles and motorcycles	0.0408	45 - Wholesale and retail trade and repair of motor vehicles and motorcycles	0.0327	47 - Retail trade, except of motor vehicles and motorcycles	0.1724	75 - Veterinary activities	0.3961
50 - Water transport	0.0414	53 - Postal and courier activities	0.0338	55 - Accommodation	0.1915	53 - Postal and courier activities	0.4117
49 - Land transport and transport via pipelines	0.0679	77 - Rental and leasing activities	0.0391	50 - Water transport	0.2083	47 - Retail trade, except of motor vehicles and motorcycles	0.4168
55 - Accommodation	0.0767	47 - Retail trade, except of motor vehicles and motorcycles	0.0417	49 - Land transport and transport via pipelines	0.2132	77 - Rental and leasing activities	0.4286
52 - Warehousing and support activities for transportation	0.0828	72 - Scientific research and development	0.05	71 - Architectural and engineering activities; technical testing and analysis	0.224	71 - Architectural and engineering activities; technical testing and analysis	0.4291
77 - Rental and leasing activities	0.0875	52 - Warehousing and support activities for transportation	0.0505	77 - Rental and leasing activities	0.2311	49 - Land transport and transport via pipelines	0.4361
46 - Wholesale trade, except of motor vehicles and motorcycles	0.1021	81 - Services to buildings and landscape activities	0.0588	68 - Real estate activities	0.2393	81 - Services to buildings and landscape activities	0.4409
68 - Real estate activities	0.1051	82 - Office administrative, office support and other business support activities	0.0593	72 - Scientific research and development	0.2427	52 - Warehousing and support activities for transportation	0.4604
72 - Scientific research and development	0.1074	46 - Wholesale trade, except of motor vehicles and motorcycles	0.0595	46 - Wholesale trade, except of motor vehicles and motorcycles	0.2436	45 - Wholesale and retail trade and repair of motor vehicles and motorcycles	0.4605
56 - Food and beverage service activities	0.1104	71 - Architectural and engineering activities; technical testing and analysis	0.0604	52 - Warehousing and support activities for transportation	0.2538	79 - Travel agency, tour operator and other reservation service and related activities	0.4621
71 - Architectural and engineering activities; technical testing and analysis	0.1128	49 - Land transport and transport via pipelines	0.0651	56 - Food and beverage service activities	0.2545	82 - Office administrative, office support and other business support activities	0.4724

81 - Services to buildings and landscape activities	0.1145	50 - Water transport	0.066	81 - Services to buildings and landscape activities	0.2547	46 - Wholesale trade, except of motor vehicles and motorcycles	0.4776
82 - Office administrative, office support and other business support activities	0.1286	61 - Telecommunications	0.066	78 - Employment activities	0.2579	72 - Scientific research and development	0.481
78 - Employment activities	0.1295	80 - Security and investigation activities	0.0724	82 - Office administrative, office support and other business support activities	0.2721	56 - Food and beverage service activities	0.4902
53 - Postal and courier activities	0.1467	70 - Activities of head offices; management consultancy activities	0.0799	61 - Telecommunications	0.2824	61 - Telecommunications	0.496
80 - Security and investigation activities	0.1527	79 - Travel agency, tour operator and other reservation service and related activities	0.08	53 - Postal and courier activities	0.2899	80 - Security and investigation activities	0.5006
61 - Telecommunications	0.1689	58 - Publishing activities	0.0865	80 - Security and investigation activities	0.3007	68 - Real estate activities	0.5032
74 - Other professional, scientific and technical activities	0.2025	56 - Food and beverage service activities	0.0902	79 - Travel agency, tour operator and other reservation service and related activities	0.3081	51 - Air transport	0.5058
79 - Travel agency, tour operator and other reservation service and related activities	0.2068	51 - Air transport	0.0922	74 - Other professional, scientific and technical activities	0.3158	58 - Publishing activities	0.5087
63 - Information service activities	0.2532	63 - Information service activities	0.0944	73 - Advertising and market research	0.3815	70 - Activities of head offices; management consultancy activities	0.5198
73 - Advertising and market research	0.2707	68 - Real estate activities	0.109	63 - Information service activities	0.3924	63 - Information service activities	0.5306
60 - Programming and broadcasting activities	0.3373	78 - Employment activities	0.1108	58 - Publishing activities	0.4348	78 - Employment activities	0.5472
58 - Publishing activities	0.3389	74 - Other professional, scientific and technical activities	0.1422	70 - Activities of head offices; management consultancy activities	0.4493	60 - Programming and broadcasting activities	0.5534
70 - Activities of head offices; management consultancy activities	0.3554	73 - Advertising and market research	0.1449	60 - Programming and broadcasting activities	0.5131	74 - Other professional, scientific and technical activities	0.5535
51 - Air transport	0.411	60 - Programming and broadcasting activities	0.1751	59 - Motion picture, video and television programme production, sound recording and music publishing activities	0.6454	73 - Advertising and market research	0.585
59 - Motion picture, video and television programme production, sound recording and music publishing activities	0.5556	59 - Motion picture, video and television programme production, sound recording and music publishing activities	0.243	51 - Air transport	0.6604	59 - Motion picture, video and television programme production, sound recording and music publishing activities	0.5879

Table A3. 2: Agglomeration by 2-digit manufacturing industry

30km				180km			
France		UK		France		UK	
Industry group	Kd	Industry group	Kd	Industry group	Kd	Industry group	Kd
16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.0138	16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.025	16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.1605	21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations	0.3939
22 - Manufacture of rubber and plastic products	0.0174	26 - Manufacture of computer, electronic and optical products	0.032	17 - Manufacture of paper and paper products	0.1708	16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.4108
25 - Manufacture of fabricated metal products, except machinery and equipment	0.0182	27 - Manufacture of electrical equipment	0.0344	31 - Manufacture of furniture	0.1774	33 - Repair and installation of machinery and equipment	0.4189
17 - Manufacture of paper and paper products	0.0188	17 - Manufacture of paper and paper products	0.0355	25 - Manufacture of fabricated metal products, except machinery and equipment	0.1788	17 - Manufacture of paper and paper products	0.4385
10 - Manufacture of food products	0.0203	22 - Manufacture of rubber and plastic products	0.0359	33 - Repair and installation of machinery and equipment	0.1837	10 - Manufacture of food products	0.4556
28 - Manufacture of machinery and equipment n.e.c.	0.027	21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations	0.0371	18 - Printing and reproduction of recorded media	0.1953	32 - Other manufacturing	0.4623
23 - Manufacture of other non-metallic mineral products	0.0275	33 - Repair and installation of machinery and equipment	0.0415	28 - Manufacture of machinery and equipment n.e.c.	0.1973	26 - Manufacture of computer, electronic and optical products	0.4675
33 - Repair and installation of machinery and equipment	0.033	10 - Manufacture of food products	0.0422	10 - Manufacture of food products	0.2031	30 - Manufacture of other transport equipment	0.4687
27 - Manufacture of electrical equipment	0.0354	28 - Manufacture of machinery and equipment n.e.c.	0.0447	22 - Manufacture of rubber and plastic products	0.2061	27 - Manufacture of electrical equipment	0.4709
24 - Manufacture of basic metals	0.0359	32 - Other manufacturing	0.0452	27 - Manufacture of electrical equipment	0.2069	20 - Manufacture of chemicals and chemical products	0.474
18 - Printing and reproduction of recorded media	0.038	31 - Manufacture of furniture	0.0466	32 - Other manufacturing	0.2075	28 - Manufacture of machinery and equipment n.e.c.	0.4786

15 - Manufacture of leather and related products	0.0435	30 - Manufacture of other transport equipment	0.0495	11 - Manufacture of beverages	0.2175	22 - Manufacture of rubber and plastic products	0.4832
32 - Other manufacturing	0.0488	25 - Manufacture of fabricated metal products, except machinery and equipment	0.0525	15 - Manufacture of leather and related products	0.2242	18 - Printing and reproduction of recorded media	0.4839
13 - Manufacture of textiles	0.0567	18 - Printing and reproduction of recorded media	0.0526	23 - Manufacture of other non-metallic mineral products	0.2243	31 - Manufacture of furniture	0.5029
29 - Manufacture of motor vehicles, trailers and semi-trailers	0.0589	29 - Manufacture of motor vehicles, trailers and semi-trailers	0.0632	24 - Manufacture of basic metals	0.2386	25 - Manufacture of fabricated metal products, except machinery and equipment	0.5051
20 - Manufacture of chemicals and chemical products	0.065	20 - Manufacture of chemicals and chemical products	0.0636	13 - Manufacture of textiles	0.2444	24 - Manufacture of basic metals	0.5095
21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations	0.0734	24 - Manufacture of basic metals	0.0651	30 - Manufacture of other transport equipment	0.2447	11 - Manufacture of beverages	0.5162
11 - Manufacture of beverages	0.0746	11 - Manufacture of beverages	0.0805	29 - Manufacture of motor vehicles, trailers and semi-trailers	0.2469	23 - Manufacture of other non-metallic mineral products	0.5227
30 - Manufacture of other transport equipment	0.0829	13 - Manufacture of textiles	0.101	26 - Manufacture of computer, electronic and optical products	0.2663	13 - Manufacture of textiles	0.5477
26 - Manufacture of computer, electronic and optical products	0.1061	14 - Manufacture of wearing apparel	0.1161	14 - Manufacture of wearing apparel	0.269	14 - Manufacture of wearing apparel	0.5498
14 - Manufacture of wearing apparel	0.1378	15 - Manufacture of leather and related products	0.1687	20 - Manufacture of chemicals and chemical products	0.2889	29 - Manufacture of motor vehicles, trailers and semi-trailers	0.593
31 - Manufacture of furniture	0.1894	23 - Manufacture of other non-metallic mineral products	0.84	21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations	0.2999	15 - Manufacture of leather and related products	0.6345

Chapter 4: Heterogeneous Agglomeration and Trade Shocks

4.1. Introduction

This chapter analyses the impact of import competition from low-income countries on the location decisions of plants between and within industries, interpreting agglomeration and co-agglomeration estimates within a common framework.

Over the past two decades, most advanced economies experienced a rapid increase in import penetration from emerging economies. The BRICS countries[44] - thanks to market friendly reforms and lower tariffs - managed to capture a significant share of world exports in several manufacturing industries. This phenomenon, complementing and driving the technological change, is known to have further fostered the process of tertiarization of western economies. Trade theory identifies labour abundant countries as a likely source of disruption in high-wage labour markets. Many empirical studies suggest that China's access to the World Trade Organisation (WTO) would have accelerated the decline of the manufacturing sector in high-income countries (Bernard, Jensen, & Schott, 2006; Autor et al., 2013; Pessoa, 2016; Foliano & Riley, 2017). Import penetration externalities can even go beyond the affected industries. Indeed, other sectors are affected because they are linked to the target sectors via labour market reallocation (Bloom et al. 2015; Acemoglu et a., 2016), Keynesian-type aggregate demand spillovers (Author et al., 2011; Acemoglu et al., 2016) and upstream/downstream linkages (Altomonte et al., 20014; Heise et al., 2015; Acemoglu et al., 2016). On the other hand, Chinese import competition has been demonstrated to increase innovation within surviving firms. Bloom et al. (2015) find that import competition leads to an increase in the speed of technical change and an overall reallocation of employment towards the most technologically advanced firms. More generally, empirical evidence indicates that the resulting Schumpeterian process could enhance productivity, both through within- and between-firm reallocation effects (Bernard, Jensen, and Schott, 2006; Altomonte et al., 2013).

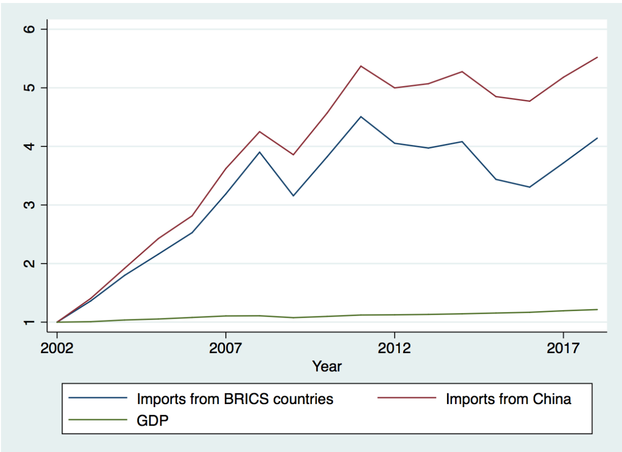
[44] Brazil, Russia, India, China and South Africa

The distributional effects of globalisation have recently come into renewed public focus. In Europe and North America, the political geography - once defined to a large extent by regional identities and social status - was completely reshaped by a clear divide between cities and countryside.

Once again, trade has been linked to the increase in disparities across space. Between 2001 and 2011 French imports from China grew five-fold. After a five-year stagnation, mostly coinciding with the debt crisis, imports started to increase again in 2016. Over the 15-year period, the manufacturing sector experienced an overall 25% decline in employment, and even larger reductions in the industries more exposed to foreign competition.

This study aims to shed new light on the distributional effect of trade shocks on the economic geography of manufacturing firms.

Figure 4. 1: Imports from China and BRICS countries



The chart compares the evolution of imports from China and BRICS countries to the evolution of France GDP

4.2. Literature

4.2.1 Import competition

A vast body of theoretical and empirical literature support the existence of a positive relationship between trade shocks and productivity. Melitz (2003), introducing product differentiation and inter-regional trade into the framework of industry dynamics proposed by Hopenhayn (1992), predicts a pro-competitive effect of trade liberalisation. Melitz and Ottaviano (2008) extend this framework, incorporating variable price-cost mark-ups and demonstrate that larger markets attract more firms, which makes competition tougher and ultimately leads less productive firms to exit. While these models take only between-firm productivity changes into account, new studies highlight a possible within-firm productivity gain. Bernard, Redding and Scott (2011) further develop the main intuition of the models, reinterpreting the dynamic for multi-product firms. In the new framework, liberalisation fosters productivity growth within and across firms, by inducing firms to reallocate resources from least to most productive products and forcing least productive firms to exit.

As far as empirical contributions are concerned, considerable evidence has been provided to demonstrate that significant effects of trade on firm-level productivity (see Tybout and Westbrook, 1995; Krishna and Mitra, 1998; Pavcnick, 2002; Fernandes, 2003; and Topalova, 2004, among others). Other studies explore the 'vertical' channels through which a trade shock can enhance productivity, using measures of inter-industry integration such as imported input, input tariffs or import competition in the upstream sector. Amiti and Konings (2005) study the effect of both input and output tariffs on productivity for a sample of Indonesian firms. The study demonstrates that a 10% reduction in output tariffs increases productivity by 1%, whereas a 10% reduction in input tariffs increases it by 3% for average firms, and by 11% for input-importing firms. Altomonte et al. (2014) test the impact of import penetration in the same industries and in upstream industries on the productivity of Italian manufacturing firms between 1996 and 2003. They find import penetration to have a positive impact on TFP, especially for penetration in upstream industries and imports from European countries. Furthermore, firm openness to trade is positively correlated with a positive pro-competitive effect on productivity. Olper et al (2014) test the impact of import penetration on 25 European countries and 9 food industries between 1995 and 2008, controlling for market size, country and sector heterogeneities, and for the endogeneity of import competition. Using different dynamic panel estimators, import penetration is found to be systematically positively related to productivity. Again, the positive relationship seems almost exclusively driven by penetration in final product markets coming from high-income countries.

In addition to firm and product selection dynamics, the productivity gains associated with import competition can be explained by trade-induced technical change, in the form of innovations and the adoption of existing technologies. Bloom et al. (2016) provide a useful review of the theoretical literature. On the one side, the pro-competitive effects driven by foreign firms could reduce agency costs (Schmidt, 1997) and increase the optimal scale of production, reducing the opportunity cost of innovation. On the other hand, the same forces might reduce firm mark-ups and thus the quasi-rents from innovation (Aghion et al., 2005). Technical change may also be driven by knowledge spillovers from foreign firms (e.g., Coe and Helpman, 1995 or Acharaya and Keller, 2008). However, this channel works only when the foreign competitors are closer to the technological frontier in the affected industries. Bloom et al. (2015) have recently proposed a 'trapped factor' model of innovation that might well explain the relation between import penetration and innovative behaviour. The intuition is that when factors of production are temporarily trapped within firms due to moving frictions, an unexpected decline in demand can reduce the opportunity cost of their application to the innovation process. Bloom et al. (2016) test the model exploiting the China shock on 12 European countries. In line with traditional models, they find that import competition fosters factor reallocation toward the most productive and innovative firms. However, they also provide evidence of a within-firm increase in the absolute volume of patents, confirming the predictions of the trapped factor model.

Import competition could also affect domestic firms via changes in the skill mix. Lu and Ng (2013) suggest three channels through which import penetration can be associated with skill composition. The first one directly follows empirical results on trade-induced technical change. Both within- and between-firm increase in innovation could lead to the adoption of new technologies and, therefore, to higher demand for skilled workers. Tougher competition could also encourage firms to change their input mix, creating the conditions for offshoring dynamics. As before, workers employed in routine tasks would pay the highest price, whereas non-routine interactive occupations would get a zero or even a positive effect. A final channel is a change in the output mix. Khandelwal (2010) suggests that in some sectors the best response to foreign competition from low-income countries is to switch the production process towards high-quality products. Holmes and Steven (2014) develop a theoretical framework to explain the resilience of small US plants with respect to foreign shocks. They suggest that flexibility of production and customer interaction make these businesses less likely to lose market share to the benefit of large foreign competitors. Both these mechanisms would encourage affected domestic firms to change their skill mix and production processes.

A third way import competition might affect firms is through changes in their internal organisation. Bloom et al. (2010) suggest that competition would foster firm decentralisation. The result can be explained theoretically with lower agency costs making decentralisation more convenient. Lower costs of decentralisation can also derive from the reduced margin for cannibalisation between competing plants (Alonso et al., 2008).

Overall, a rich theoretical and empirical literature highlights different ways in which foreign competition change structure and behaviour of domestic firms. Changes in productivity distribution, innovative behaviour, internal structure, size and skill mix are all likely to have some effects on dispersion and agglomeration forces that determine the spatial distribution of the economic activity. Surprisingly, the firm-level literature has so far provided little evidence on the spatial implications of these dynamics. Instead, more attention has been devoted to the spatial location of exporters and multinational firms.

4.2.2 Agglomeration and international trade

Since Bernard and Jensen's (1999) seminal empirical paper on US exporters, several studies have provided evidence on the fact that trading activity is often concentrated in small spatial areas. Furthermore, firms engaging in international trade differ from non-trading firms in many respects. These firms generally employ more high-skilled workers, pay higher wages and are more productive than firms operating in the domestic market only. Mayer and Ottaviano (2008) offer significant evidence regarding the peculiar advantages that characterise the 'happy few' involved in international trade. While new economic geography and trade models (Krugman 1991, Krugman & Venables, 1996) have widely studied the role of trade costs and economies of scale in defining the uneven spatial distribution of economic activity, scant attention has been given to the relationship between trading behaviour and agglomeration. Most significantly of all, while the core-periphery model and several extensions provided in the literature have focused on regional agglomeration and dispersion dynamics, few models have attempted to explain the relation between internal agglomeration and external trade. An important contribution was made by Redding (2012), who extended a recent class of quantitative international trade models to incorporate factor mobility across regions within countries. If labour is perfectly mobile, regions with higher locational fundamentals and land quality experience larger increases in their population until real wages are again equalised in the new equilibrium following the reduction in trade barriers. As a result, external trade liberalisation leads to an endogenous internal reallocation of workers.

Further theoretical and empirical evidence has been provided by Fajgelbaum and Redding (2014), who studied Argentina's integration into world markets in the late-nineteenth century. Their study demonstrates the role of internal geography in shaping the effects of international trade liberalisations. Koenig (2005) argues that exporters benefits from being in the vicinity of 'first geography factors' such as bodies of water or national borders. French overseas traders are found to locate near those borders and ports that provide access to their respective partner countries. The empirical literature has given less attention to the so called 'second-geography factors': proximity of other firms and density of economic environment. A relevant exception is presented by Bekes and Harasztosi (2009), who argue that externalities that determine agglomeration premia for firms are affected by the firms' involvement in trade. They maintain that the performance of internationalised firms in more densely populated environments differs from that of firms operating within the domestic market only. Using a large panel of Hungarian manufacturing firms, they demonstrate that international traders benefit twice as much as non-traders from agglomeration in productivity. Sun et al. (2012) demonstrate that agglomeration has a final positive effect on firms' export possibilities and sales. Higher natural-productivity firms gain a higher productivity premium from agglomeration. Empirical results based on data from Chinese industrial enterprises confirm the theoretical results.

This study takes a different perspective, investigating to what extent import competition changes the geography of domestic economic activity. Exploiting geocoded plant-level data, covering the whole population of manufacturing firms, and international trade data retrieved from Comtrade dataset, I study the effect of import penetration from China on the spatial location patterns of industries and specific types of plants.

4.3. Data

Data are retrieved from two administrative sources made available by the INSEE (the French national statistical agency) to researchers with access to CASD secure lab. The first source is the Declarations Annuelles de Donnes sociales (DADS). Every year all French workers are expected to submit to the tax authority information about all incomes earned over the previous 12 months. The INSEE receives these data from the tax authority, merges them with other information on workers and households provided by various sources, and uses them to produce statistics about employment and wages.

The data used in this process are recorded in an exhaustive dataset, where each observation corresponds to every job contract linking a worker to a firm in a given year. For every worker, the dataset provides job-spell level information on gross and net wage, employment periods, age, sex and skills, number of hours worked, type of contract and occupational category (Professions et categories socio-professionnelles, PCS 2003) at the 4-digit level.

The dataset FICUS/FARE, produced by INSEE/DGFiP, provides balance sheet information (output, capital, inputs, exports, number of employees, sector, etc.) for each firm registered in France from 1993 to 2016 (approximately 47 million observations). Data are retrieved from the compulsory reporting of firms and income statements to fiscal authorities in France, with no limiting threshold in terms of firm size or sales. By merging the dataset with the information provided by the Sirene (Système Informatique pour le Répertoire des Entreprises de leurs Etablissements) dataset, it is possible to assign address and precise spatial coordinates to 27 million establishments (the entire population of establishments that ever operated in France over the past 20 years). In this analysis, I drop from the sample firms subject to the micro-bin/micro-bic regimes, namely firms:

- whose principal activities are services relating to the category of business profits (BIC) or non-commercial profits (BNC)
- whose turnover does not exceed €32,600 (for 2011), excluding taxes
- which operate VAT free and do not perform an excluded activity (for example, the rental of equipment and durable consumer goods);
- which do not opt for the *régime du réel simplifié* tax system

The Liaison Financiére dataset (LIFI) provides data on all relevant financial linkages involving at least one firm located in France. For each observation, I can determine the country of residence of all foreign subsidiary/parent firms and the capital share. The median voting share owned upon acquisition is 99%, with the result that the acquisition event represents a near complete takeover of assets and control for the overwhelming majority of the sample.

Geocoded plant level data are retrieved from the Sirene database. Each plant is assigned to specific coordinates, corresponding to the address reported for fiscal purposes.
This information is missing for a small number of plants (<5%), located primarily in rural areas. In these cases, I assign the coordinate of the applicable *IRIS*. *IRIS* are the smallest French administrative unit (the French territory consists of over 49,000 IRIS areas) and plants with missing coordinates are expected to be located, on average, 400 m from their estimated location.

4.4 Empirical Strategy

My goal is to test the effect of trade shocks on the spatial agglomeration of industries and different plant-types. To this end, I first compute an index for spatial agglomeration. Following Behrens at al. (2017), I extend the continuous agglomeration measures developed by Duranton and Overman (2003) [DO] and compute agglomeration and co-agglomeration measures for a set of industries and plant-types within industries. The functions are then linked to a simple index, obtained integrating the DO function. In contrast to the previous chapter, I also replicate the analysis using an alternative measure, considering only the intervals in which the DO function lies above or below its confidence interval.

An import penetration measure is computed exploiting industry-product data retrieved from *Comtrade* and instrumenting the endogenous variable with world import supply for each product-country combination.

4.4.1. Main Specification

In the main estimation equation, I model local agglomeration as a function of time-invariant industry-specific characteristics, θ_j, common shocks, δ_t, and the import penetration from China, IP_{jt}^{CN}.

$$Aggl = \alpha + \beta_1 IP_{jt}^{CN}(d) + \theta_j + \delta_t + e_{jt} \qquad (1)$$

The same regression is estimated multiple times for different levels of *d*, ranging between 30 km, considered a good proxy for local clustering, and 180 km, which can be associated with global agglomeration.

4.4.2 Continuous agglomeration indices

The weighted K-density proposed by Duranton and Overman describes the distribution of bilateral distances between workers in a given industry (see Section 3.3.1 in Chapter 3). In order to obtain an index providing information on the degree of agglomeration/dispersion, I follow two different strategies.

First, consistently with the previous chapter, I integrate the distribution function, obtaining its cumulative (CDF) up to a distance *d*. The values obtained represent the weighted share of plant-pairs located less than distance *d* from each other. Alternatively, we can view this as the probability that two randomly drawn plants in an industry will be at most *d* km away.

$$\overline{K_j^w}(\bar{d}) = \sum_{d \le \bar{d}} K_j^w(d)$$

This procedure was used by Behrens (2017) to study the agglomeration of Canadian manufacturing industries. Although it represents a useful index to synthetize the information content provided by the DO function, it fails to account for the overall pattern of concentration in the manufacturing sector as a whole. Therefore, I develop a different index, consisting in the difference between the $K_j^w(d)$ lying above (below) the upper (lower) confidence band. In this way I am able to compute two new indices:

$$\Gamma_j^w = \sum_{d \le \bar{d}} \gamma_j^w(\bar{d})$$

Where $\gamma_j^w = \max\left(K_j^w(d) - \overline{K_w}(d), 0\right)$

$$\Psi_j^w = \sum_{d \le \bar{d}} \psi_j^w(\bar{d})$$

Where $\psi_j^w = \max\left(\underline{K_w}(d) - K_j^w(d), 0\right)$

The adjusted CDF, based on the index proposed by Behrens and Bougna (2014), is computed as:

$$\overline{\overline{K_j^w}}(\bar{d}) = \begin{cases} \Gamma_j^w(\bar{d}) & \text{if } g \text{ is significantly aggregated} \\ 0 & \textit{if g is neither aggregated, nor dispersed} \\ -\psi_j^w(\bar{d}) & \text{if } g \text{ is significantly dispersed} \end{cases}$$

To understand the behaviour of the index, we can consider the industry $K_j^w(d)$ (blue line) in Figure 4. 3. Its DO function continuously lies above the upper confidence interval up to roughly 50 km. As a result, the $\overline{\overline{K_j^w}}(d)$ will be greater than 0 for every distance level and increasing up to d=50. The $\overline{\overline{K_j^w}}(d)$ at 60 km distance will be equal to the blue area, between the agglomeration function and the upper confidence interval.

Figure 4. 2: Behrens (2017) agglomeration index

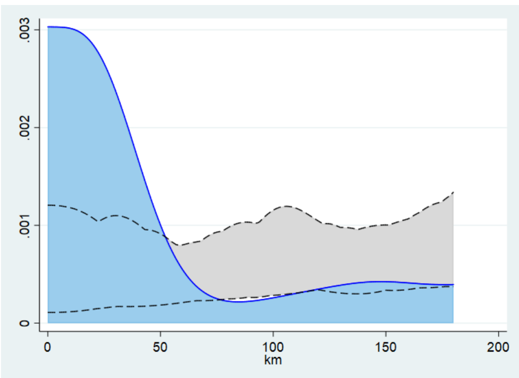
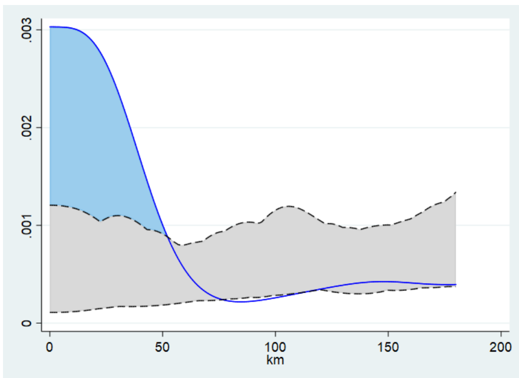


Figure 4. 3: Adjusted agglomeration index



4.4.3 Localisation and co-localisation patterns

In Section 3.3.2, I showed that between- and within-industry agglomeration patterns can be analysed using three different indices:

I. The agglomeration of all plants in the industry: $\overline{K_J^w}$

II. The agglomeration of plants belonging to a certain type (group agglomeration): $\overline{K_{\iota J}^w}$

III. The co-agglomeration between *i*-type plants and all plants in the industry (joint agglomeration): $\overline{K_{\iota J,J}^w}$

Within each industry, we can identify various types of firms and plants on the basis of demographic characteristics, size, productivity level, foreign status, etc… All this factor might determine a different spatial configuration with respect to the other plants located in the industry. The three indices make it possible to assign each industry-group type to a specific spatial pattern:

a. $\overline{K_{\iota J,J}^w} > \overline{K_J^w}$; $\overline{K_{\iota J}^w} > \overline{K_J^w}$

In this case, group *i* is on average more agglomerated than its reference industry. This is the case when *i*-type plants are overrepresented in industry *j* local clusters and, at the same time, they tend to locate close to each-other. This second feature can be obtained with a lower representation outside local clusters or through the concentration of *i*-type plants in some, but not all clusters. This pattern could describe some manufacturing industries characterised by clusters of SMEs, that located close to each other in order to share some common functions.

b. $\overline{K_{\iota J,J}^w} > \overline{K_J^w}$; $\overline{K_{\iota J}^w} < \overline{K_J^w}$

A group of plants might exhibit high degrees of co-agglomeration with respect to the whole industry but low group-level spatial agglomeration. This spatial distribution could describe plants characterised by high complementarity with other establishments in the sector. Establishments with e central position in cluster networks and characterised by high internal economies of scale show a similar pattern.

c. $\overline{K_{\iota J,J}^w} < \overline{K_J^w}$; $\overline{K_{\iota J}^w} > \overline{K_J^w}$

A specific plant-type can exhibit low degree of co-agglomeration and high levels of group-level spatial concentration. This spatial configuration could characterise, for instance, manufacturing headquarters located close to each other in a large city, far from productive establishments, accounting for most industry *j* plants.

d. $\overline{K_{\iota J,J}^w} < \overline{K_J^w}$; $\overline{K_{\iota J}^w} < \overline{K_J^w}$

Finally, *i* establishments could be more dispersed than other plants in industry *j* both in terms of group agglomeration and group-industry co-agglomeration. This pattern could describe exiters and new entries in a sector where dispersed plants are characterised by higher turnout.

The examples illustrated above are not meant to be exhaustive and indicative of all possible drivers of each scenario. By contrast, they serve as examples to understand the centrifugal and centripetal forces that together shape the geography of different industries.

4.4.4 Import Penetration

Import penetration analysis is performed using data retrieved from the UN Comtrade database. The import penetration proxy, defined as in Bernard, Jensen and Schott (2011), is given by:

$$IP_{jt}^{CN} = \sum_{p} \frac{IMP_{pjct}^{CN}}{IMP_{pjt} + PROD_{pjct} - EXP_{pjct}}$$

where IMP_{pjct}^{CN} and IMP_{pjt} represent respectively the value of product p (industry *j)* imports from China and all countries, $PROD_{pt}$ is product p domestic production and EXP_{pt} represents the exports. The index, aggregated at the 4-digit industry level, is a ratio between the imports from China and the estimated local demand.

4.4.5 Instrumental Strategy

A potential concern in my empirical strategy is that unobserved factors, such as technology shocks or institutional reforms might be correlated with both changes in product-level Chinese imports and labour demand. To address this issue, I use the interaction between initial share of products imported from China to each French industry *j* with the world export supply of product p from China (excluding exports to France). The instrument is subsequently normalised using world export supply of the same product.

$$I_{jt} = \sum_{p} \frac{WES_{pjt}^{CN}}{WES_{pjt}^{W}} \frac{Imp_{pj,t-n}^{CN}}{Imp_{j,t-n}^{CN}}$$

Where WES_{pjt}^{CN} is the world export supply of product p in year *t* form China, WES_{pjt}^{ALL} is the overall value of the same product traded worldwide, $Imp_{j,t-n}^{CN}$ is the overall value imported from industry *j* in pre-sample period and $Imp_{pj,t-n}^{CN}$ is the value of product p only imported in the same period. World export supply data are based on Comtrade data at the HS6 product level. Similar strategies have been used in the literature (Altomonte et al., 2006; Bloom et al., 2012, Hummels et al., 2014, among others) to isolate the exogenous supply-driven components of import shocks. First stage results are reported in Table A4. 1.

4.5 Results

Import penetration and manufacturing employment

Table 4. 1 presents regression results describing some general effects of import penetration that have been studied in depth by the trade literature. In columns (1), foreign competition from China is found to have a large negative effect on industrial employment. The result is consistent with the recent literature (Autor et al., 2013; Acemoglu et al., 2016), suggesting that trade with developing economies is correlated with a more rapid decline in manufacturing employment, especially in labour-intensive industries. Results are consistent when considering the IV specification (column 2). According to the trade literature, trade liberalisation fosters pro-competitive effects in the domestic market, leading to an increase in domestic market concentration (Melitz, 2003; Melitz & Ottaviano, 2007). The results in column (3) confirm the predictions. The effect is found to be even greater when the explanatory variable is instrumented (column 4).

Import penetration and between-industry spatial agglomeration patterns

Import penetration from emerging economies accelerates the employment disruption in manufacturing sectors, leading to an input reallocation towards the largest firms, which are more resilient to foreign competition. I now investigate how this dynamic affects the spatial distribution of economic activities. To do so, in Table 4. 2 I regress the continuous agglomeration index computed up to distance d on the import penetration from China. In other words, I test whether the import penetration is correlated with the probability of finding two plants within a certain distance *d*.

Overall, foreign competition is found to foster spatial agglomeration in the manufacturing sector. However, I find heterogeneous results with respect to the scale considered. Column (1) presents the impact of import penetration on the spatial agglomeration up to a 30 km distance, measured as the simple CDF of the $K_j^w(d)$ function. I can interpret this lower threshold of the bilateral distance distributions as an index for the share of economic activities localised in industrial clusters. One standard deviation increase in the share of local demand covered by Chinese suppliers is associated with a 0.05% increase in the share of establishments located within 30km from each other.

The results suggest that clusters increase their relative share of employment in France. On the opposite side of the spatial distribution, column (6) presents the impact on the 'global agglomeration', up to 180 km. This index measures the tendency of firms to localise their plants in certain macro-areas (such as NUTS2 regions). One standard deviation increase in Chinese import penetration determines a 0.056% increase in the share of plant pair bilateral distances lower or equal to 180km.

This result suggest that import penetration leads to a rise in global agglomeration, largely driven by variations at shorter distances. Considering the intermediate thresholds of the bilateral distance distribution, I notice that coefficients rise up to a distance of 90 km and then progressively decline. This suggests that centripetal and centrifugal forces are affected up to 90 km distance. In the same Table, I test the effect of foreign competition on a different proxy of spatial agglomeration, namely the adjusted CDF proposed in *Section 4.1*. Results are consistent, although the magnitude of the coefficient is constant above 60km, suggesting the main change happens below that threshold.

Table 4. 3 presents the results of the IV specification. Instrumenting the explanatory variable with product-country specific world input supply, the pattern described above becomes even more evident. Import penetration is found to have a positive and significant effect on spatial agglomeration at any distance between 30 km and 180km. One standard deviation increase in the share of domestic demand covered by Chinese suppliers leads to a 0.2% increase in the share of plant pairs located within 30km from each other over the total bilateral distribution in the industry. The effect seems particularly relevant at intermediate distances, with the coefficient peaking at 90km and then slowly declining. In particular, the magnitude doubles between 30 and 60km, suggesting a relevant effect up to that distance threshold, and then remains constant.
A similar pattern is found when I consider spatial agglomeration/dispersion outside the sector confidence interval, with coefficients reduced by approximately 25%.

In Table 4. 4 and Table 4. 5 I replicate the same analysis focussing on all the BRICS countries (Brazil, Russia, India, China, South Africa). Again, the results confirm a positive effect of foreign competition from emerging countries on the spatial agglomeration of manufacturing plants. The results are consistent when I adopt the adjusted agglomeration index and when I instrument the endogenous variable.

Overall, foreign competition from China is found to have a non-negligible effect on the spatial distribution of economic activity. One standard deviation change in import penetration is associated to a 0.05%-0.2% increase in the share of firms clustered within 30km from each other and to a 0.05%-0.4% increase in the number of firms agglomerated within 180 km from each other. In chapter 3, I showed that in the period 2008-2015 the overall French manufacturing sector experienced a decline by 0.2% in local agglomeration and a 0.8% decline in global agglomeration. The results suggest import competition could have significantly slowed down the process of spatial dispersion characterising the manufacturing sector.

Import penetration and within-industry spatial agglomeration patterns

I now proceed to analyse the relation between import penetration from emerging countries and the spatial agglomeration of specific industrial firm types. To this end, I refer to the four possible patterns described in Section 4.4. Exogenous trade shocks can impact agglomeration forces, changing the degree to which plants co-locate with other establishment of the same type and/or with all plants belonging to the same industry.

In Section B1.1 I focus on plant demographics. Figure 4. 4 and Figure 4. 5 report regression coefficients and confidence intervals for, respectively, the effect of import penetration from China on the degree of co-agglomeration between entrants and the incumbent plant and the group-specific agglomeration of this category. Firms entering the market are not affected by the higher competition in their location decision with respect to the existing economic geography. However, a small, although insignificant, negative effect is found for the spatial co-location of new entrants.

The estimates for firms leaving the market exhibit a positive and increasing pattern (Figure 4. 6 and Figure 4. 7). Above 120km, the p-value is below 0.1. In this case, the result for group co-agglomeration is perfectly symmetrical.

In Section B1.2 I report the same estimates for the alternative specification in which the adjusted CDF is used as a dependent variable ($\overline{\overline{K_J^w}}(d)$). Results are confirmed for new entries, but the positive and significant effect on agglomeration of exiters disappears. Overall, foreign competition does not seem to have any effects on the spatial distribution of demographic events.

In Section B2.1 I focus on plant size. Import penetration from China is found to have a positive and significant effect on both joint co-location with the rest of the sector and the group agglomeration of large plants (Figure 4. 12 and Figure 4. 13), especially up to 60k distance. By contrast, small plants show an inverted U-shaped pattern. The only significant result is the negative effect on group-specific agglomeration at very short distance (Figure 4. 15).

In section B2.2 I analyse the $\overline{\overline{K_J^w}}(d)$ estimates. The pattern discussed above with respect to large plants is confirmed and so is the null-effect on the co-agglomeration of small plants with the rest of industry (Figure 4. 16 and Figure 4. 17) The coefficient for small plant agglomeration becomes largely negative, although still non-significant (Figure 4. 18 and Figure 4. 19).
In summary, the positive effect of import penetration on spatial agglomeration is driven primarily by the largest plants, with employment equal to or greater than the 10th percentile, whereas little or no effect is found for the small firms, with employment equal to or lower than the 25th percentile. These results seem consistent with the predictions of Holmes and Steven (2011). Large plants are more affected by import competition and consequently engage in a process of spatial agglomeration to benefit from external economies of scale.

In Section B3.1 I analyse the relation between foreign competition and the spatial agglomeration of firms with different internal spatial organisation.
In order to understand the way in which industries respond to foreign competition, it is important to study the behaviour of multi-plant firms. If it is true that market concentration should naturally foster spatial agglomeration, the rapid rise of multi-plant firms could be associated with a higher dispersion of plants belonging to the same firm. It is possible that larger companies expanding through M&As maintain the spatial distributions of plants previously operating as autonomous firms. In addition, as suggested by Bloom et al. (2010), larger competition could reduce agency costs and allow large firms to decentralise different functions and locate each establishment in the region that best fits its skill/input mix.

In Figure 4. 20 and Figure 4. 21 I report coefficients and confidence intervals of the IV specification, analysing the relation between import penetration and the agglomeration of single-plant firms. The charts suggest that foreign competition has a zero effect on the agglomeration of this type of firm, with some evidence of a positive effect only at very large distances. By contrast, multi-plant firms are found to exhibit higher agglomeration in response to foreign shocks. This is true for the joint agglomeration with the rest of the industry and the agglomeration of multi-plant firms only, up to 120 km (Figure 4. 22 and Figure 4. 23).

An even clearer pattern is found for headquarters, which react to import penetration with higher agglomeration at any distance threshold (Figure 4. 24 and Figure 4. 25).

In Section B3.2, I report regression coefficients and confidence intervals for the $\overline{\overline{K_J^w}}(d)$ specification, which generally confirm previous results. Overall, competition from China seems to affect only multi-plant firms. This might be consistent with an important role of the internal organisation of firms in explaining the overall dynamic patterns that characterise the affected industries.

In Section B4.1, I consider how import penetration affects the spatial agglomeration of the most productive firms. In this case, the pattern is relatively clear. An increase in foreign competition does not affect the group agglomeration and co-agglomeration for the most productive firms (Figure 4. 32 and Figure 4. 33). By contrast, we find an extremely positive and significant coefficient for low productive firms (below the 25^{th} percentile of the productivity distribution). However, these results hold only for co-agglomeration patterns, whereas no such a result is found for group agglomeration. Once again, results are largely confirmed by $\overline{\overline{K_J^w}}(d)$ estimates.

The results could be explained in three ways. First, the positive effect on firm productivity highlighted by the international trade literature (Altomonte et al. 2008; Bloom et al., 2012) could reduce the overall dispersion of the productivity distribution, helping more dispersed and less productive firms to catch-up with the clustered frontier.

Second, an increase in industrial concentration obtained by means of a reallocation of resources towards the most productive firms could lead the productivity frontier to assume a spatial pattern more consistent with the overall distribution. In other words, factor reallocation of dispersed low-productive plants towards high-productive plants could foster spatial agglomeration at the lower tale of the productivity distribution. Finally, the tougher competition could decrease the benefits arising from knowledge spillovers for the firms at the productivity frontier, leading some firms to relocate their establishment far from their competitors.

In Sections B5.1 and B5.2 I finally consider multinational firms. I find some evidence of a positive relation between supply shocks and multinational co-agglomeration with domestic firms, but results are not significant (Figure 4. 40 and Figure 4. 42). By contrast, import penetration has no effect on group agglomeration of multinational firms (Figure 4. 41 and Figure 4. 43). This result could depend on the higher resilience of multinational firms that already compete in international markets and are less affected than domestic firms by new competitors accessing in the market. The rise in co-agglomeration would be then primarily driven by domestic firms.

In the Appendix I replicate all the analyses, using import penetration from all BRICS countries. Results generally confirm the main findings relative to import penetration from China only.

4.6 Conclusions

In this study, I proposed a composite strategy to analyse the way industries and some specific establishment types respond to import competition from low-income countries.
Extending the continuous agglomeration measures proposed by Duranton and Overman (2006) and developed by Behrens (2017), I investigate a large micro dataset covering the whole population of French establishments from 2008 to 2015. Foreign competition from China is found to increase spatial agglomeration in affected industries. Instrumenting the main explanatory variable with the world market share of exporting countries leads to similar results. The outcome is not clearly explained by the location of new entries and firms quitting the market. It follows, therefore, that the main trend is probably driven by reallocation of resources across plants rather than by plant demographic events. Pro-competitive shocks are also found to affect plant location patterns within industries. Large, multi-plant and low productive firms respond to pro-competitive shock increasing spatial agglomeration, whereas no effect is found for small plants and single-plant, multinational and low productive firms. Similar results are found when I consider import penetration from all BRICS countries. These findings highlight the fact that, behind the negative effect on aggregate employment and the positive effects on productivity and innovative behaviour documented by the literature, foreign competition is likely to change the internal geography of firms.

The results have a number of policy implications. First, foreign competition is found to partially offset the spatial dispersion trend that characterise the manufacturing sector, with relevant implications on sector productivity and spatial inequalities. Second, import penetration does not have a homogenous effect on industries, but has instead a particularly significant effect on certain plant types. As a result, limited modifications in the spatial locations of entire industries could mask deeper changes in the nature of economic activities across regions.

Further research is needed to shed new light on these mechanisms. A possible extension would be to adapt to this framework the approach proposed by Faggio et al. (2018) to study to what extent import penetration affects the co-agglomeration patterns between plant-types belonging to different industries. A second line of research could focus instead on functional agglomeration patterns, studying relation between foreign competition and the spatial distribution of individual workers on the basis of location, industry and occupational category.

References

Acemoglu, Daron, and David Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics 4* : 1043-1171.

Almazan, A., De Motta, A., & Titman, S. (2007). Firm location and the creation and utilization of human capital. *The Review of Economic Studies*, *74*(4), 1305-1327.

Acharya, R. C., & Keller, W. (2009). Technology transfer through imports. *Canadian Journal of Economics/Revue canadienne d'économique*, *42*(4), 1411-1448

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An inverted-U relationship. *The Quarterly Journal of Economics*, *120*(2), 701-728

Alonso, R., Dessein, W., & Matouschek, N. (2008). When does coordination require centralisation?. *American Economic Review*, *98*(1), 145-79

Altomonte, C., Aquilante, T., Békés, G., & Ottaviano, G. I. (2013). Internationalization and innovation of firms: evidence and policy. *Economic policy*, *28*(76), 663-700

Altomonte, Carlo, Alessandro Barattieri, and Armando Rungi (2014). Import penetration, intermediate inputs and productivity: evidence from Italian firms. *Rivista italiana degli economisti*19.1 45-66.

Amiti, M., & Konings, J. (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. *American Economic Review*, *97*(5), 1611-1638

Autor, D., (2013). The 'task approach' to labor markets: an overview. No. w18711. *National Bureau of Economic Research.*

Autor, D., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *The American Economic Review,* 103(6), 2121-2168. ISO 690

Behrens, K., & Bougna, T. (2015). An anatomy of the geographical concentration of Canadian manufacturing industries. *Regional Science and Urban Economics*, 51, 47-69.

Behrens, K. (2016). Agglomeration and clusters: tools and insights from coagglomeration patterns. *Canadian Journal of Economics/Revue canadienne d'économique*, *49*(4), 1293-1339

Behrens, K. (2016). Agglomeration and clusters: Tools and insights from coagglomeration patterns. Canadian Journal of Economics/Revue canadienne d'économique, 49(4), 1293-1339.

Behrens, K., & Guillain, R. (2017). The determinants of coagglomeration: Evidence from functional employment patterns. *CEPR Discussion Papers* (No. 11884).

Békés, G., & Harasztosi, P. (2013). Agglomeration premium and trading activity of firms. *Regional Science and Urban Economics*, *43*(1), 51-64

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2009). The margins of US trade. *American Economic Review*, *99*(2), 487-93

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott (2006). Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of US manufacturing plants. Journal of international Economics 68.1, 219-237.

Bernard, A.B., Redding, S., Schott (2006) CEP Discussion Paper No 769 December Multi-Product Firms and Trade Liberalization

Bernard, A. B., & Jensen, J. B. (1999). Exceptional exporter performance: cause, effect, or both?. *Journal of international economics*, *47*(1), 1-25

Bloom, N., Romer, P. M., Terry, S. J., & Van Reenen, J. (2013). A trapped-factors model of innovation. *American Economic Review*, *103*(3), 208-13

Bloom, Nicholas, Mirko Draca, and John Van Reenen (2016). Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity. *The Review of Economic Studies* 83.1 87-117.

Coe, D. T., & Helpman, E. (1995). International r&d spillovers. *European economic review*, *39*(5), 859-887

Duranton, Gilles and Overman, H. (2005) Testing for localization using micro-geographic data. *The Review of Economic Studies* 72.4 (2005): 1077-1106.

Duranton, Gilles and Overman, H. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science* 48.(1), 213-243.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier

Fajgelbaum Pablo, Redding Stephen J. (2014) External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914. *CEP Discussion Paper No 1273 June 2014*

Fernandes, A. (2003). Trade policy, trade volumes, and plant-level productivity in Colombian manufacturing industries. *The World Bank.*

Foliano, F., & Riley, R. (2017). International Trade and UK de-industrialisation. *National Institute Economic Review*, *242*(1), R3-R13

Heise, S., Pierce, J. R., Schaur, G., & Schott, P. K. (2015). *Trade policy and the structure of supply chains*. mimeo

Holmes, T. J., & Stevens, J. J. (2014). An alternative theory of the plant size distribution, with geography and intra-and international trade. *Journal of Political Economy*, *122*(2), 369-421.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, 1127-1150

Hummels, D., Jørgensen, R., Munch, J., & Xiang, C. (2014). The wage effects of offshoring: Evidence from Danish matched worker-firm data. *American Economic Review*, *104*(6), 1597-1629.

Koenig, P. (2009). Agglomeration and the export decisions of French firms. *Journal of Urban Economics*, 66(3), 186-195.

Krishna P, Mitra D. Trade liberalization, market discipline and productivity growth: new evidence from India. *Journal of development Economics*. 1998 Aug 1;56(2):447-62.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483-499

Krugman, P., & Venables, A. J. (1996). Integration, specialization, and adjustment. *European economic review*, *40*(3-5), 959-967

Khandelwal, A. (2010). The long and short (of) quality ladders. *The Review of Economic Studies*, *77*(4), 1450-1476

Lu, Y., & Ng, T. (2013). Import competition and skill content in US manufacturing industries. *Review of Economics and Statistics*, *95*(4), 1404-1417

Maggioni, D. (2013). Productivity dispersion and its determinants: the role of import penetration. *Journal of Industry, Competition and Trade*, *13*(4), 537-561

Marcon, E., & Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, *3*(4), 409-428

Marcon, E., & Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, *62*, 56-67

Mayer, T., & Ottaviano, G. I. (2008). The happy few: The internationalisation of European firms. *Intereconomics*, *43*(3), 135-148

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*(6), 1695-1725

Olper, A., Pacca, L., & Curzi, D. (2014). Trade, import competition and productivity growth in the food industry. *Food Policy*, *49*, 71-83

Pavcnik N. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *The Review of Economic Studies*. 2002 Jan 1;69(1):245-76.

Pessoa, J. P. (2016). International Competition and Labor Market Adjustment. *CEP Discussion Paper No 1411 March 2016*

Redding Stephen J., Goods Trade (2012), Factor Mobility and Welfare, *NBER Working Paper* No. 18008

Schmidt, K. M. (1997). Managerial incentives and product market competition. *The Review of Economic Studies*, *64*(2), 191-213

Tippins, N. T., Hilton, M. L. (2010). A database for a changing economy: review of the occupational information network (O* NET). *National Academies Press.*

Topalova P, Khandelwal A. Trade liberalization and firm productivity: The case of India (2011). *Review of economics and statistics*. Aug 1;93(3):995-1009.

Tybout, James R., and M. Daniel Westbrook (1995). Trade liberalization and the dimensions of efficiency change in Mexican manufacturing industries. *Journal of International Economics* 39.1-2

Tables

Table 4. 1: Employment and market concentration

VARIABLES	(1)	(2)	(3)	(4)
	ln Employment	ln Employment	ln Herfindal Index	ln Herfindal Index
China IP	-0.115***	-0.118**	0.133***	0.243***
	-0.0137	-0.0148	-0.0388	-0.0461
Observations	1305	1305	1305	1305
R-squared	0.799	0.799	0.804	0.816
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
IV	NO	YES	NO	YES

Notes. The table reports the regression results for a simple linear model testing the relation between industrial employment, concentration and import penetration from China. The analysis is conducted over 124 4-digit manufacturing industries over the period 2008-2015. Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 4. 2: Import penetration from China and spatial agglomeration, FE

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	CDF (d=30)	CDF (d=60)	CDF (d=90)	CDF (d=120)	CDF (d=150)	CDF (d=180)
			$\overline{K_J^w}(\mathrm{d})$			
China IP	0.0498***	0.0686***	0.0694***	0.0614***	0.0574***	0.0559***
	(0.0150)	(0.0198)	(0.0198)	(0.0180)	(0.0167)	(0.0164)
Observations	791	791	791	791	791	791
R-squared	0.963	0.963	0.967	0.972	0.977	0.977
			$\overline{\overline{K_J^w}}(d)$			
China IP	0.0448***	0.0551***	0.0556***	0.0554***	0.0554***	0.0538***
	(0.0149)	(0.0174)	(0.0175)	(0.0173)	(0.0172)	(0.0168)
Observations	791	791	791	791	791	791
R-squared	0.942	0.929	0.885	0.871	0.856	0.809
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
IV	NO	NO	NO	NO	NO	NO

Notes. The table reports the estimation results for the model presented in section 4.4.1, testing the effect of import penetration form China on spatial agglomeration. The dependent variables are the simple CDF of the DO function, $\overline{K_J^w}(\mathrm{d})$ and the adjusted CDF,$\overline{\overline{K_J^w}}(d)$, presented in section 4.4.2. The import penetration variable is standardised to have zero mean and unit standard deviation. The model is estimated at different distance thresholds, between 30km and 180km. The analysis is conducted over 124 4-digit manufacturing industries over the period 2008-2015. Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 4. 3: Import penetration from China and spatial agglomeration, FE-IV

VARIABLES	(1) CDF (d=30)	(2) CDF (d=60)	(3) CDF (d=90)	(4) CDF (d=120)	(5) CDF (d=150)	(6) CDF (d=180)
			$\overline{K_J^w}(\mathrm{d})$			
China IP	0.207***	0.402***	0.483***	0.467***	0.429***	0.434***
	(0.0715)	(0.113)	(0.140)	(0.153)	(0.157)	(0.168)
Observations	791	791	791	791	791	791
R-squared	0.798	0.768	0.757	0.752	0.770	0.728
			$\overline{\overline{K_J^w}}(d)$			
China IP	0.152***	0.294***	0.298***	0.293***	0.287***	0.290***
	(0.0586)	(0.0793)	(0.0796)	(0.0785)	(0.0767)	(0.0765)
Observations	791	791	791	791	791	791
R-squared	0.942	0.928	0.885	0.87	0.856	0.809
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
IV	YES	YES	YES	YES	YES	YES

Notes. The table reports the estimation results for the model presented in section 4.4.1, testing the effect of import penetration form China on spatial agglomeration. The dependent variables are the simple CDF of the DO function, $\overline{K_J^w}(\mathrm{d})$ and the adjusted CDF,$\overline{\overline{K_J^w}}(d)$, presented in section 4.4.2. Import penetration from China, standardised to have have zero mean and unit standard deviation, is instrumented using world export supply. The model is estimated at different distance thresholds, between 30km and 180km. The analysis is conducted over 124 4-digit manufacturing industries over the period 2008-2015. Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 4. 4: Import penetration from BRICS countries and spatial agglomeration, FE

	(1)	(2)	(3)	(4)	(5)	(6)
				CDF	CDF	CDF
VARIABLES	CDF (d=30)	CDF (d=60)	CDF (d=90)	(d=120)	(d=150)	(d=180)
$\overline{K_J^w}(\mathrm{d})$						
BRICS IP	0.0326***	0.0423***	0.0417***	0.0370***	0.0360***	0.0359***
	(0.0103)	(0.0131)	(0.0126)	(0.0114)	(0.0108)	(0.0109)
Observations	791	791	791	791	791	791
R-squared	0.963	0.962	0.966	0.972	0.977	0.977
$\overline{\overline{K_J^w}}(d)$						
BRICS IP	0.0300***	0.0345***	0.0346***	0.0345***	0.0345***	0.0332***
	(0.00999)	(0.0110)	(0.0111)	(0.0110)	(0.0109)	(0.0106)
Observations	791	791	791	791	791	791
R-squared	0.942	0.929	0.885	0.871	0.856	0.809
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
IV	NO	NO	NO	NO	NO	NO

Notes. The table reports the estimation results for the model presented in section 4.4.1, testing the effect of import penetration form BRICS countries (Brazil, Russia, India, China, South Africa) on spatial agglomeration. The dependent variables are the simple CDF of the DO function, $\overline{K_J^w}(\mathrm{d})$ and the adjusted CDF,$\overline{\overline{K_J^w}}(d)$, presented in section 4.4.2. The import penetration variable is standardised to have zero mean and unit standard deviation. The model is estimated at different distance thresholds, between 30km and 180km. The analysis is conducted over 124 4-digit manufacturing industries over the period 2008-2015. Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table 4. 5: Import penetration from BRICS countries and spatial agglomeration, FE-IV

VARIABLES	(1) CDF (d=30)	(2) CDF (d=60)	(3) CDF (d=90)	(4) CDF (d=120)	(5) CDF (d=150)	(6) CDF (d=180)
$\overline{K_J^w}(\mathrm{d})$						
BRICS IP	0.177***	0.340***	0.405***	0.387***	0.352***	0.356**
	(0.0624)	(0.0995)	(0.120)	(0.129)	(0.132)	(0.140)
Observations	791	791	791	791	791	791
R-squared	0.798	0.768	0.757	0.752	0.770	0.728
$\overline{\overline{K_J^w}}(d)$						
BRICS IP	0.133**	0.254***	0.255***	0.251***	0.245***	0.247***
	(0.0526)	(0.0746)	(0.0751)	(0.0739)	(0.0727)	(0.0725)
Observations	791	791	791	791	791	791
R-squared	0.942	0.928	0.885	0.87	0.856	0.809
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
IV	YES	YES	YES	YES	YES	YES

Notes. The table reports the estimation results for the model presented in section 4.4.1, testing the effect of import penetration form BRICS countries (Brazil, Russia, India, China, South Africa) on spatial agglomeration. The dependent variables are the simple CDF of the DO function, $\overline{K_J^w}(\mathrm{d})$ and the adjusted CDF,$\overline{\overline{K_J^w}}(d)$, presented in section 4.4.2. Import penetration from BRICS countries, standardised to have zero mean and unit standard deviation, is instrumented using world export supply. The model is estimated at different distance thresholds, between 30km and 180km. The analysis is conducted over 124 4-digit manufacturing industries over the period 2008-2015. Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

B. Graphical Analysis – Import penetration from China

B.1.1 Entrants and exiters

Figure 4. 4: Entrants (group-population)

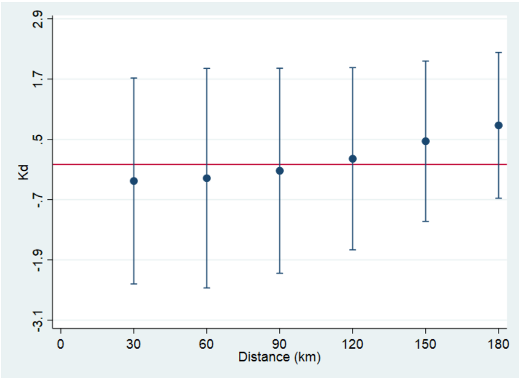
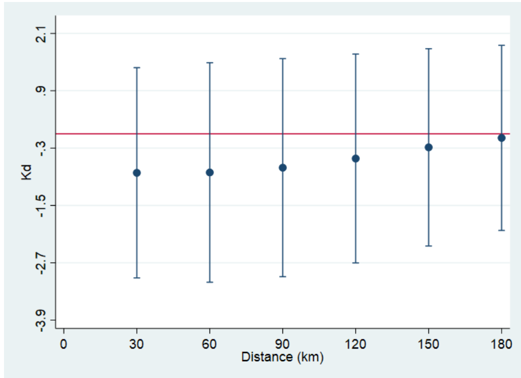


Figure 4. 5: Entrants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for new entrants in the market. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 6: Exiters (group-population)

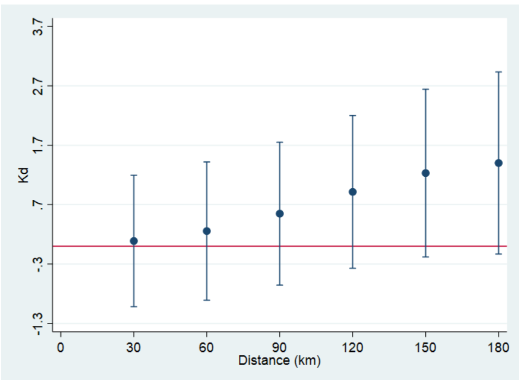
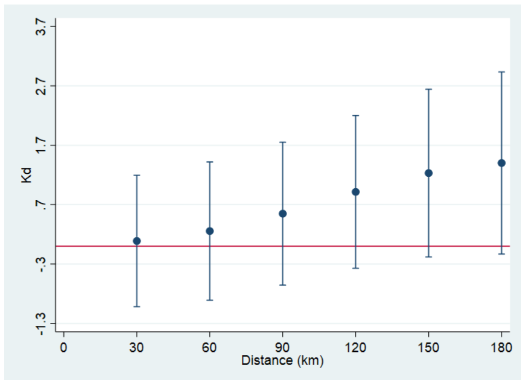


Figure 4. 7: Exiters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for plants that left the market. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.1.2 Entrants and exiters

Figure 4. 8: Entrants (group-population)

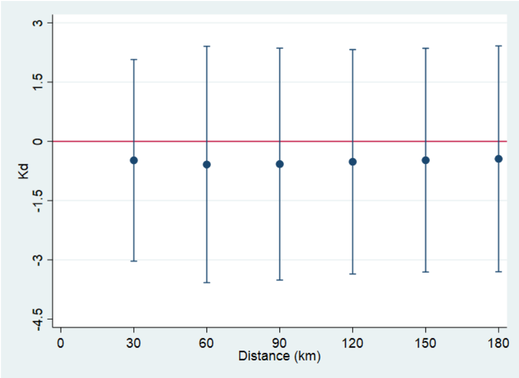
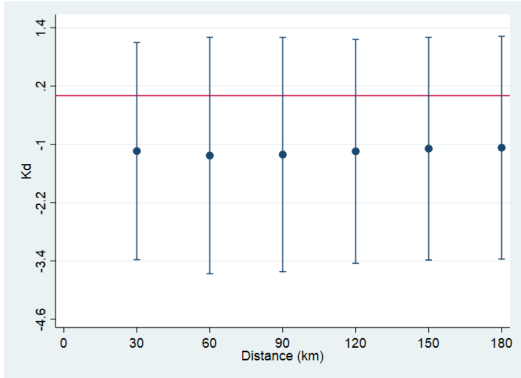


Figure 4. 9: Entrants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ *and* $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for plants that enter the market. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 10: Exiters (group-population)

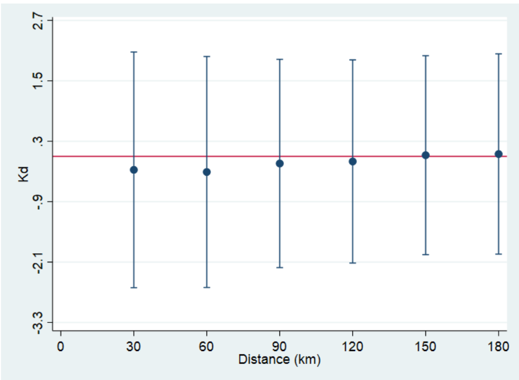
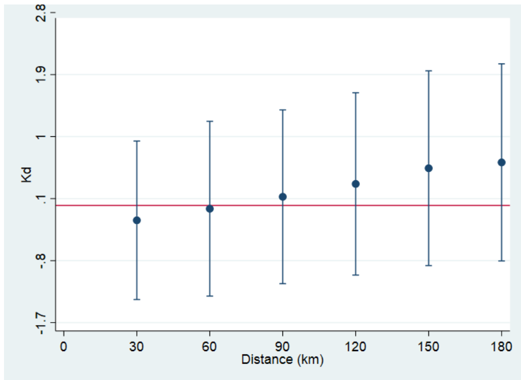


Figure 4. 11: Exiters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ *and* $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for plants that left the market. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.2.1 Small and large plants

Figure 4. 12: Large plants (group-population)

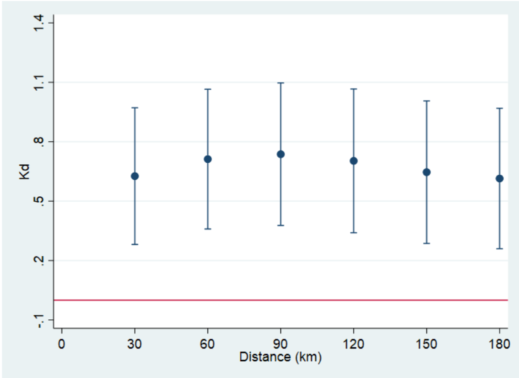
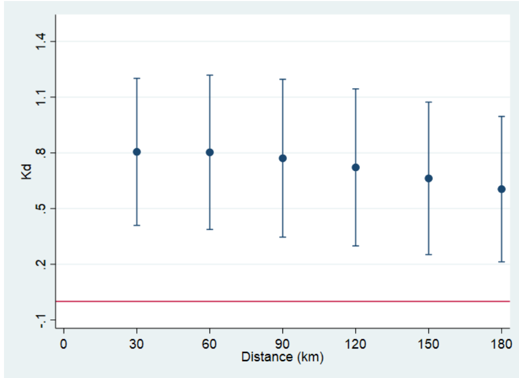


Figure 4. 13: Large Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{IJ,J}}$ and $\overline{K^{w}_{IJ}}$, estimated for large plants (above the 90th percentile of the size distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 14: Small plants (group-population)

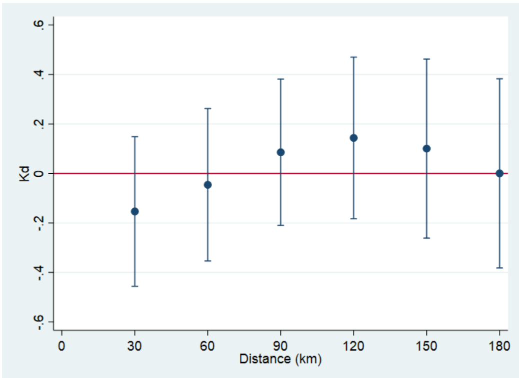
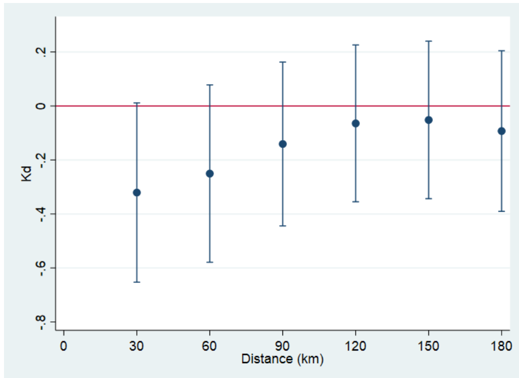


Figure 4. 15: Small Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{IJ,J}}$ and $\overline{K^{w}_{IJ}}$, estimated for small plants (below the 25th percentile of the size distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.2.2 Small and large plants

Figure 4. 16: Large plants (group-population)

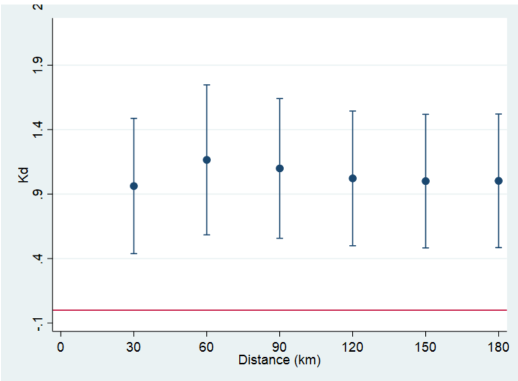
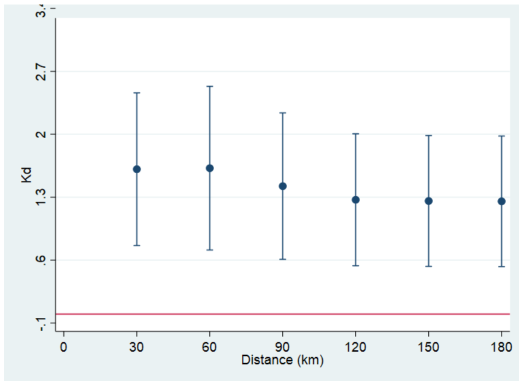


Figure 4. 17: Large Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{IJ,J}^{w}}(d)$ and $\overline{K_{IJ}^{w}}(d)$, estimated for large plants (above the 90th percentile of the size distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 18: Small plants (group-population)

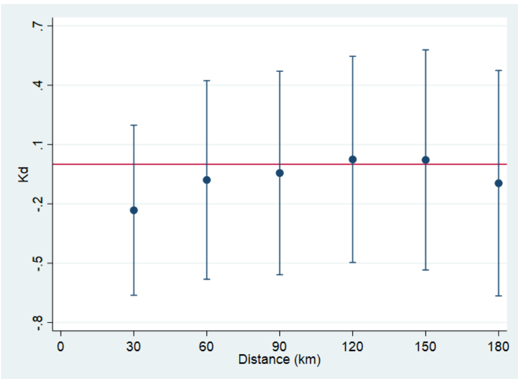
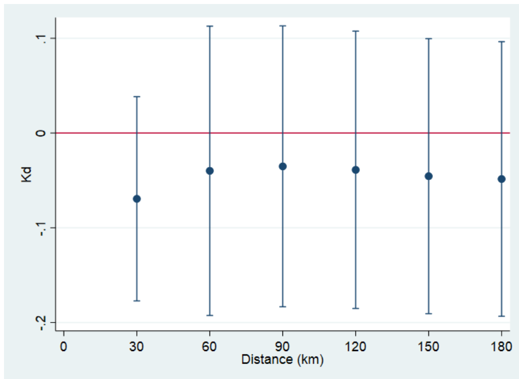


Figure 4. 19: Small Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{IJ,J}^{w}}}(d)$ and $\overline{\overline{K_{IJ}^{w}}}(d)$, estimated for small plants (below the 25th percentile of the size distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.3.1 Single-plant, Multiplant Firms and Heqdquarters

Figure 4. 20: Single-plant (group-population)

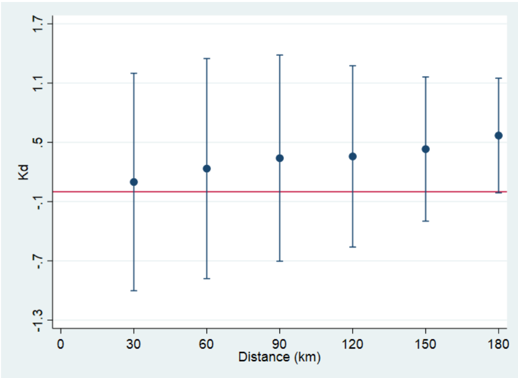
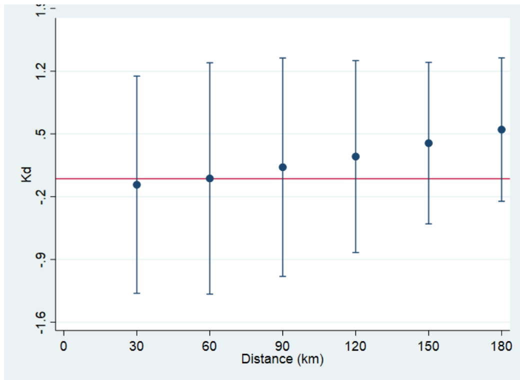


Figure 4. 21: Single-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{ij,j}}$ and $\overline{K^{w}_{ij}}$, estimated for single-plant firms. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 22: Multiplant (group-population)

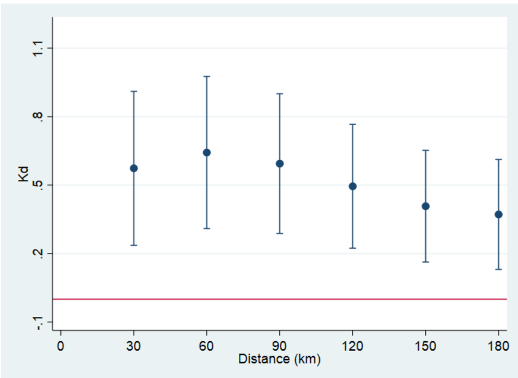
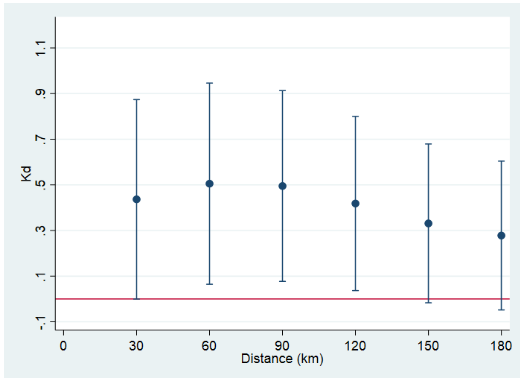


Figure 4. 23: Multiplant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{ij,j}}$ and $\overline{K^{w}_{ij}}$, estimated for multiplant firms. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 24: Headquarters (group-population)

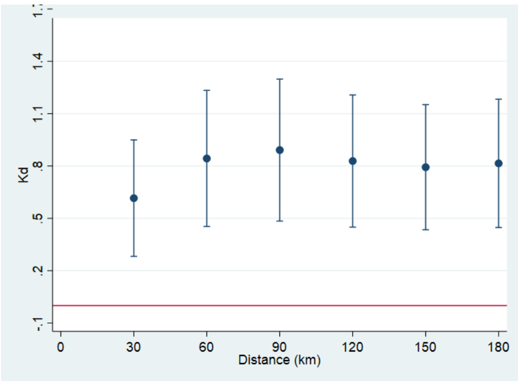
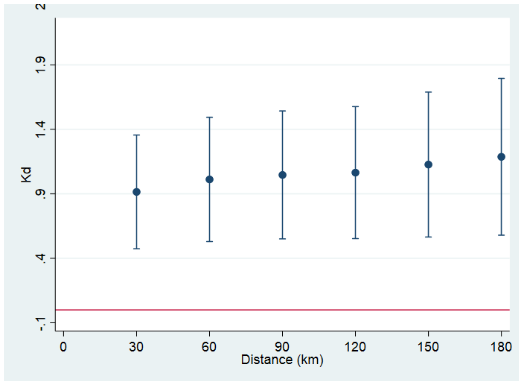


Figure 4. 25: Headquarters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^w_{ij,J}}$ and $\overline{K^w_{ij}}$, estimated for headquarter plants. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B3.2 Single-plant, Multiplant Firms and Heqdquarters

Figure 4. 26: Single-plant (group-population)

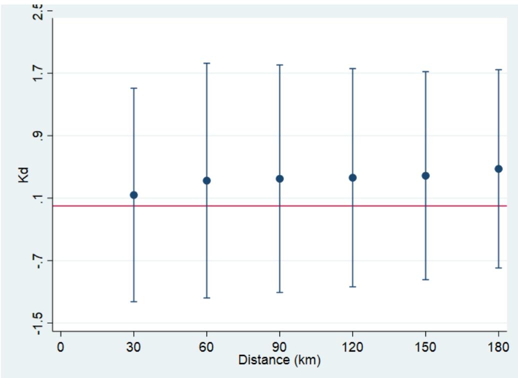
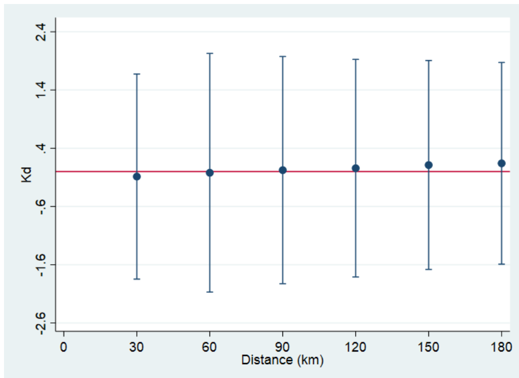


Figure 4. 27: Single-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^w_{ij,J}}}(d)$ *and* $\overline{\overline{K^w_{ij}}}(d)$, estimated for single-plant firms. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 28: Multiplant (group-population)

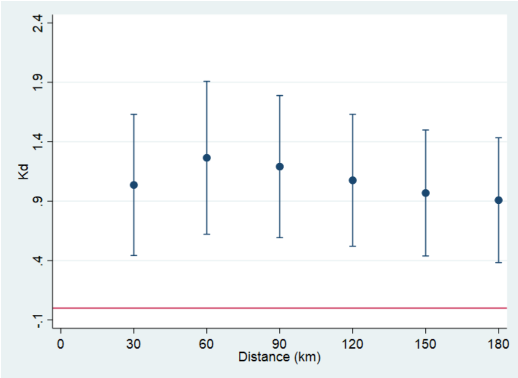
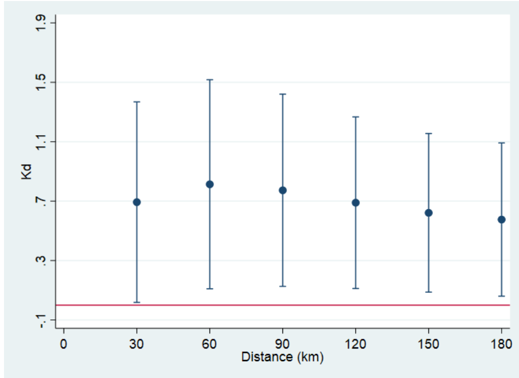


Figure 4. 29: Multiplant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{iJ,J}^{w}}}(d)$ and $\overline{\overline{K_{iJ}^{w}}}(d)$, estimated for multiplant firms. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 30: Headquarters (group-population)

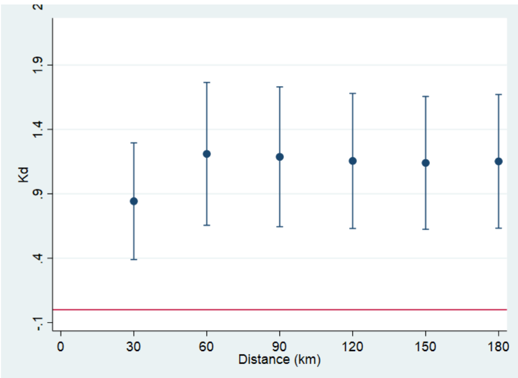
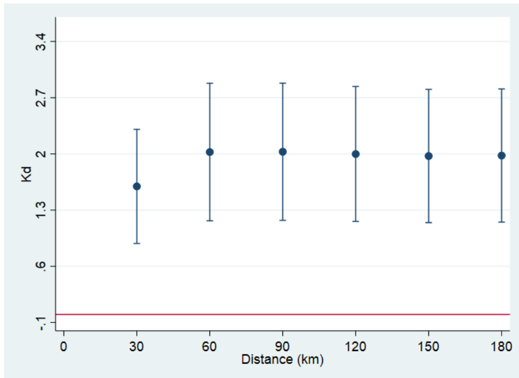


Figure 4. 31: Headquarters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{iJ,J}^{w}}}(d)$ and $\overline{\overline{K_{iJ}^{w}}}(d)$, estimated for headquarter plants. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.4.1 High/low LP

Figure 4. 32: High LP (group-population)

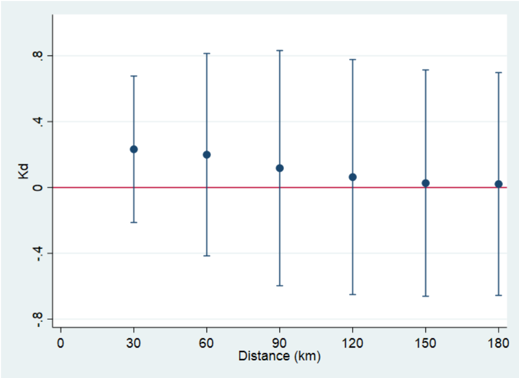
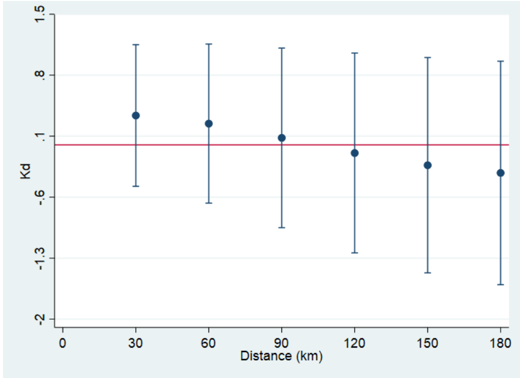


Figure 4. 33: High LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for high-productive firms (above the 90th percentile of the labour productivity distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 34: Low LP (group-population)

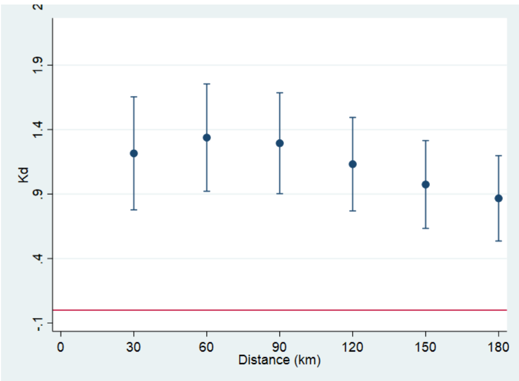
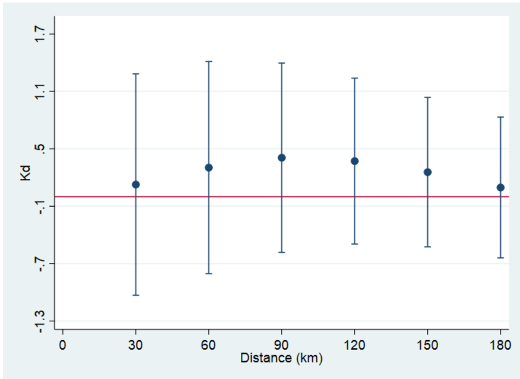


Figure 4. 35: Low LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for high-productive firms (below the 25th percentile of the labour productivity distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.4.2 High/low LP

Figure 4. 36: High LP (group-population)

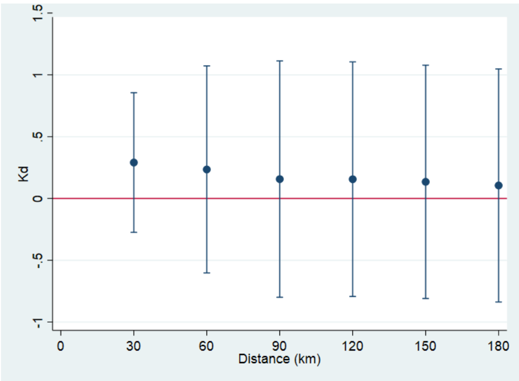
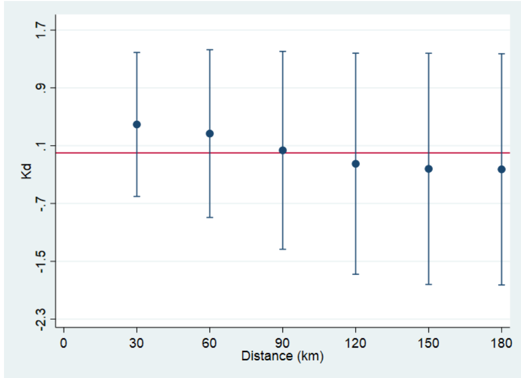


Figure 4. 37: High LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ *and* $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for high-productive firms (above the 90th percentile of the labour productivity distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure 4. 38: Low LP (group-population)

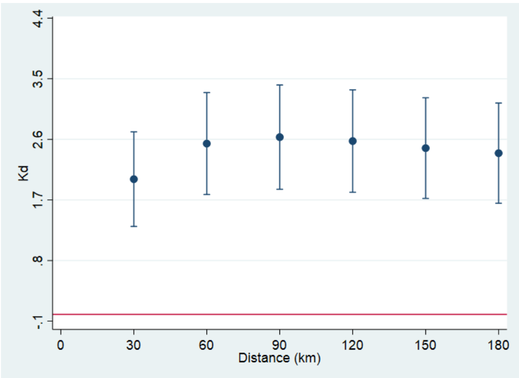
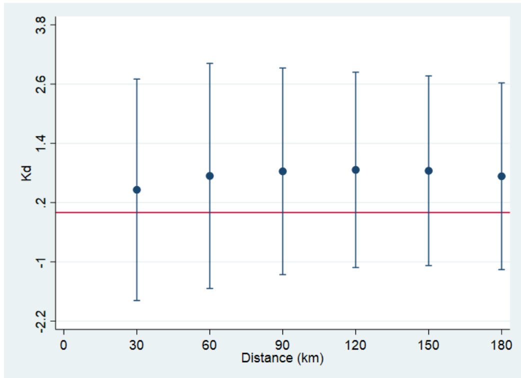


Figure 4. 39: Low LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ *and* $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for low-productive firms (below the 25th percentile of the labour productivity distribution). The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.5.1 MNEs

Figure 4. 40: MNEs (group-population)

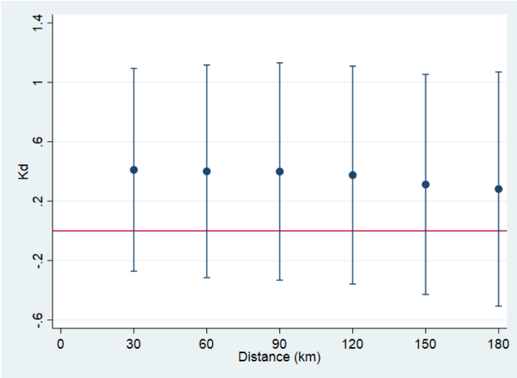
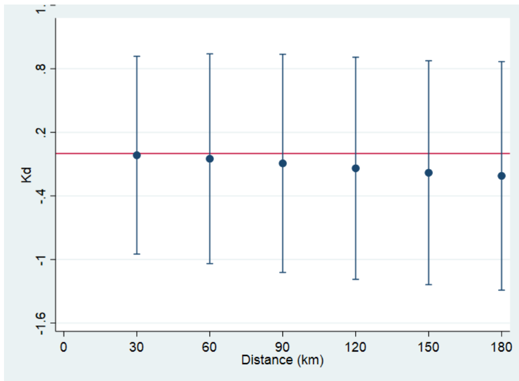


Figure 4. 41: MNEs (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for multinational enterprises. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

B.5.2 MNEs

Figure 4. 42: MNEs (group-population)

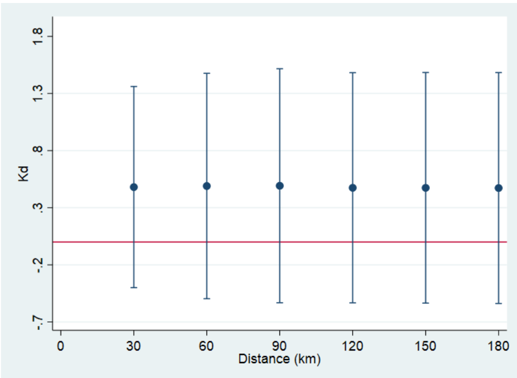
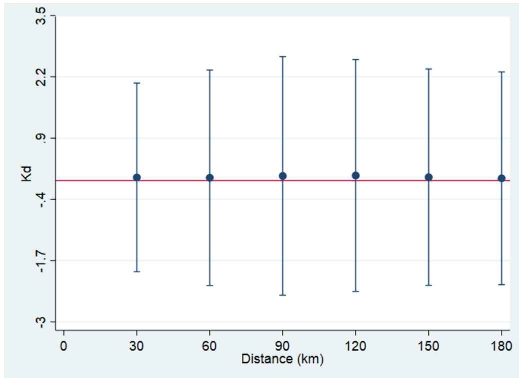


Figure 4. 43: MNEs (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ *and* $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for multinational enterprises. The explanatory variable is import penetration from China, instrumented by world supply (sections 4.4.4 and 4.4.5)

C. Appendix

C.1 Tables

Table A4. 1: First stage

VARIABLES	(1)	(2)
	China IP	BRICS IP
WES	0.219***	0.194***
	0.0259	0.0272
Observations	1305	1305
R-squared	0.68	0.57
Industry FE	YES	YES
Year FE	YES	YES
F	23.44	18.72

This table reports the first stage of the instrumental strategy presented in Section 4.4.5. Import penetration from China and BRICS countries is analysed in relation to the overall supply from these countries to the world (excluding France). Robust standard errors are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

C.1 Import penetration from BRICS

C.1.1 Entrants and exiters

Figure A4. 1: Entrants (group-population)

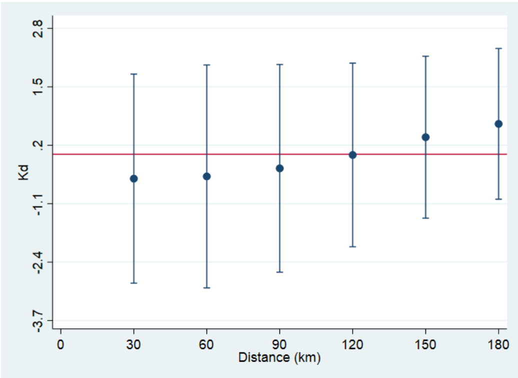
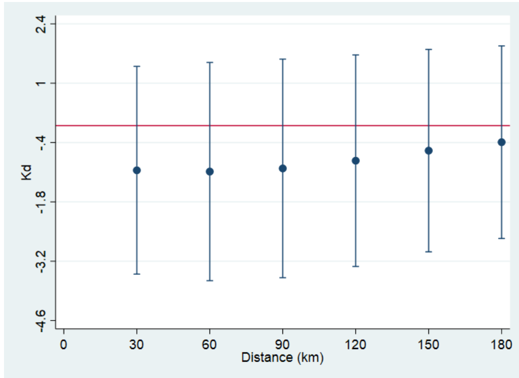


Figure A4. 2 Entrants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{iJ,J}}$ and $\overline{K^{w}_{iJ}}$, estimated for new entrants in the market. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 3: Exiters (group-population)

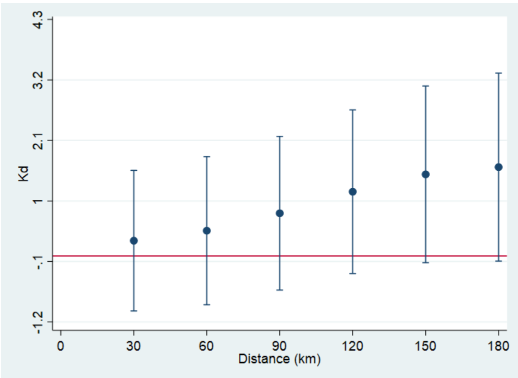
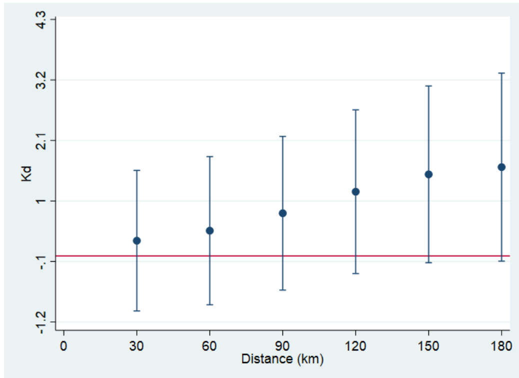


Figure A4. 4: Exiters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{J,J}}$ and $\overline{K^{w}_{iJ}}$, estimated for plants that left the market. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.1.2 Entrants and exiters

Figure A4. 5: Entrants (group-population)

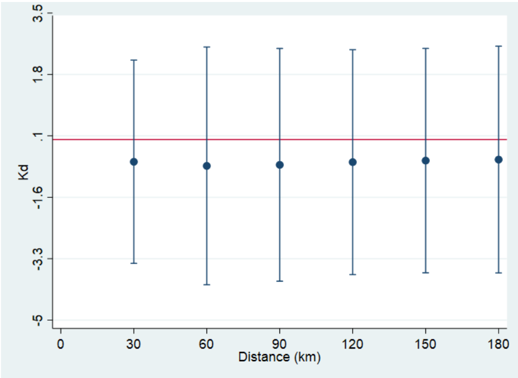
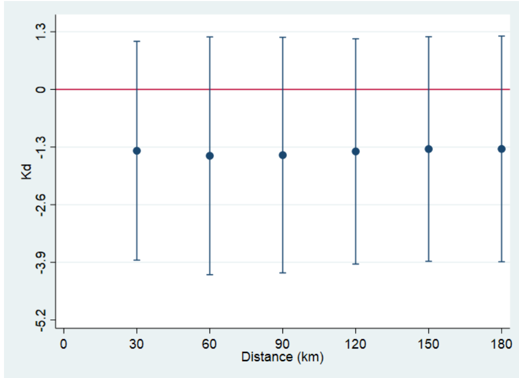


Figure A4. 6: Entrants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{IJ,J}^{w}}}(d)$ *and* $\overline{\overline{K_{IJ}^{w}}}(d)$, estimated for plants that enter the market. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 7: Exiters (group-population)

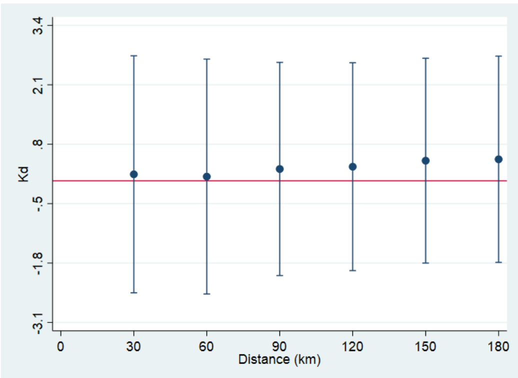
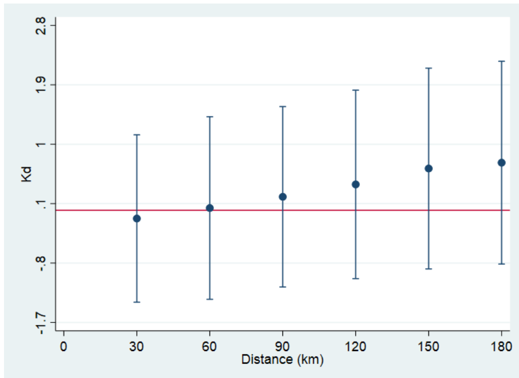


Figure A4. 8: Exiters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{IJ,J}^{w}}}(d)$ *and* $\overline{\overline{K_{IJ}^{w}}}(d)$, estimated for plants that left the market. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.2.1 Small and large plants

Figure A4. 9: Large plants (group-population)

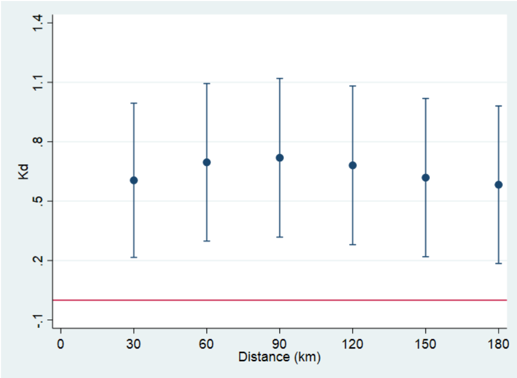
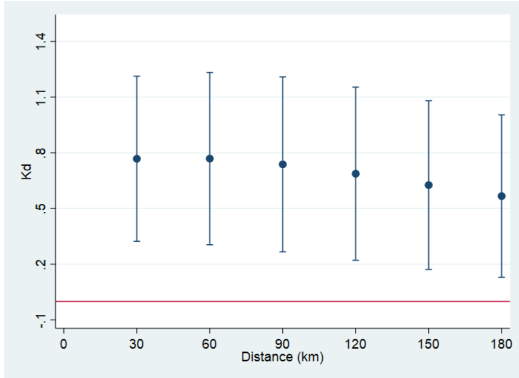


Figure A4. 10: Large Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{ij,J}}$ and $\overline{K^{w}_{ij}}$, estimated for large plants (above the 90th percentile of the size distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 11: Small plants (group-population)

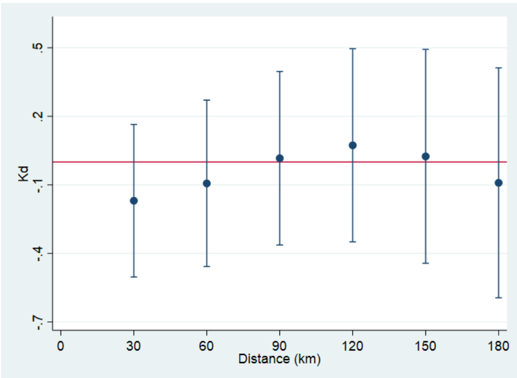
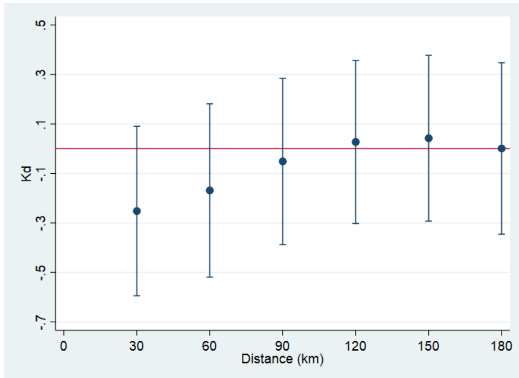


Figure A4. 12: Small Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^{w}_{ij,J}}$ and $\overline{K^{w}_{ij}}$, estimated for small plants (below the 25th percentile of the size distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.2.2 Small and large plants

Figure A4. 13: Large plants (group-population)

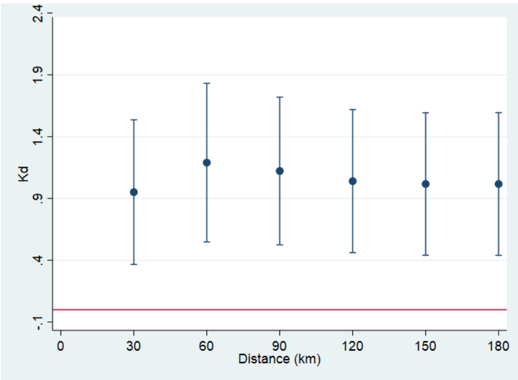
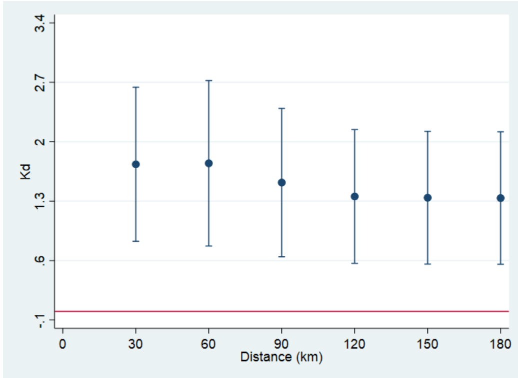


Figure A4. 14: Large Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^{w}_{ij,J}}}(d)$ and $\overline{\overline{K^{w}_{ij}}}(d)$, estimated for large plants (above the 90th percentile of the size distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 15: Small plants (group-population)

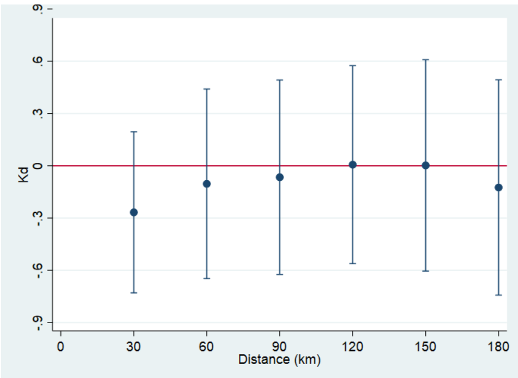
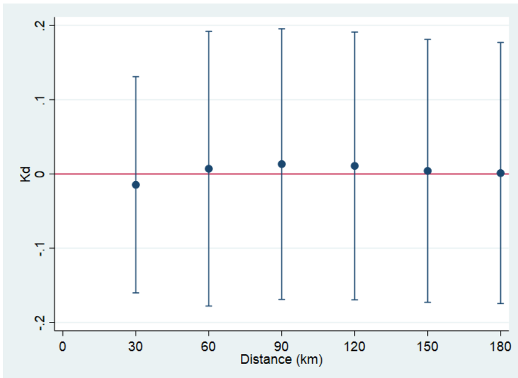


Figure A4. 16: Small Plants (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^{w}_{ij,J}}}(d)$ and $\overline{\overline{K^{w}_{ij}}}(d)$, estimated for small plants (below the 25th percentile of the size distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.3.1 Single-plant, Multiplant Firms and Heqdquarters

Figure A4. 17: Single-plant (group-population)

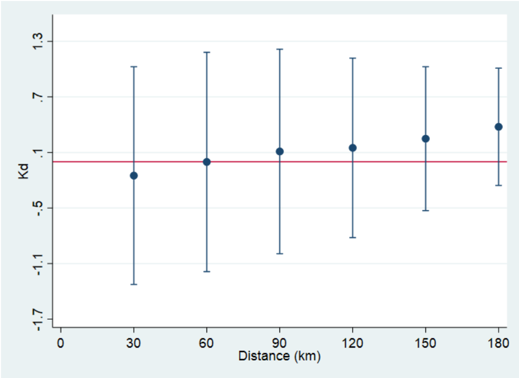
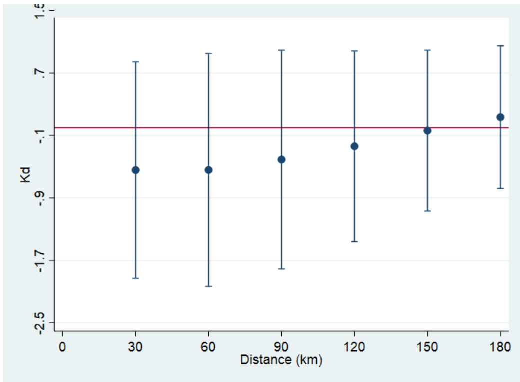


Figure A4. 18: Single-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for single-plant firms. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 19: Multi-plant (group-population)

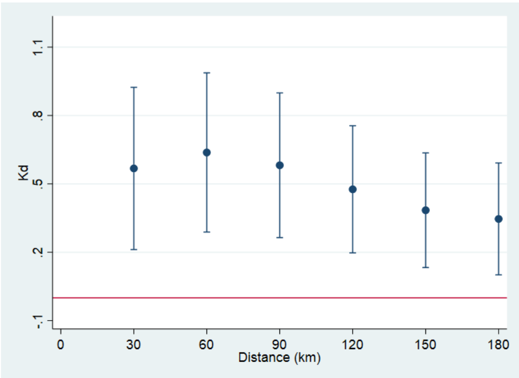
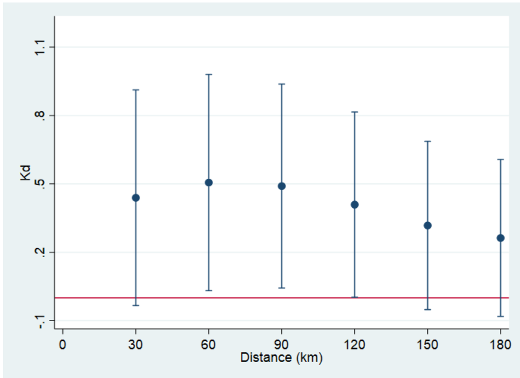


Figure A4. 20: Multi-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for multiplant firms. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 21: Headquarters (group-population)

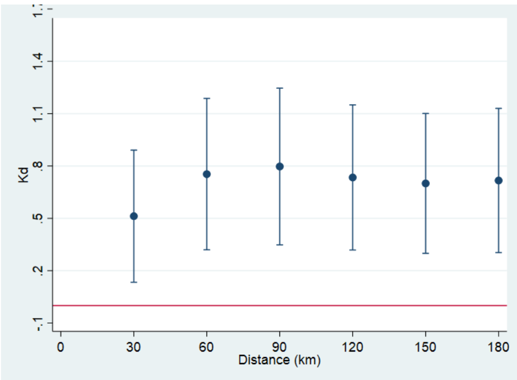
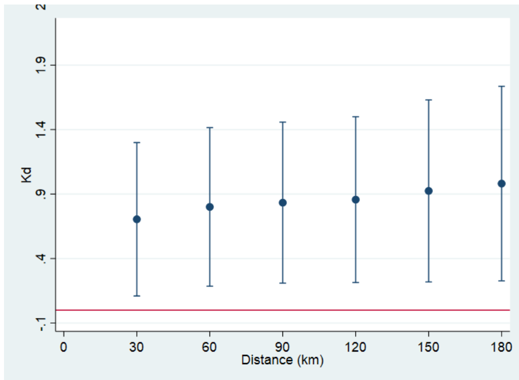


Figure A4. 22: Headquarters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ and $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for headquarter plants. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.3.2 Single-plant, Multiplant Firms and Heqdquarters

Figure A4. 23: Single-plant (group-population)

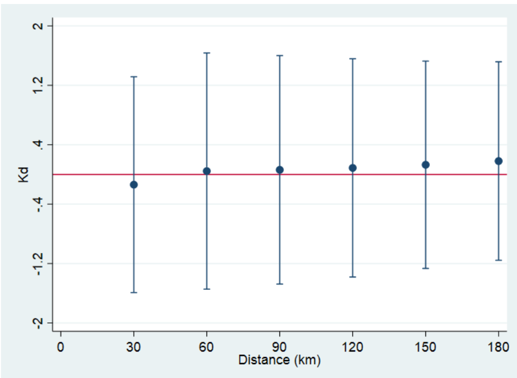
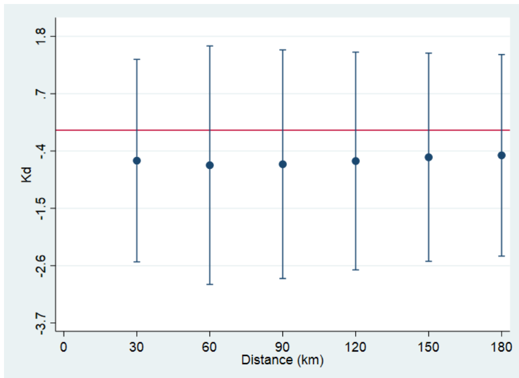


Figure A4. 24: Single-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ and $\overline{\overline{\ }}{}_{ij}^{w}(d)$, estimated for single-plant firms. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 25: Multi-plant (group-population)

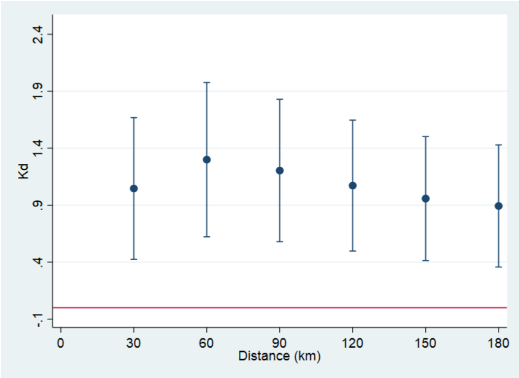
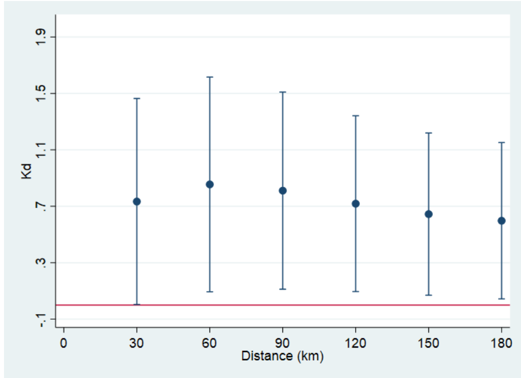


Figure A4. 26: Multi-plant (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^{w}_{ij,J}}}(d)$ *and* $\overline{\overline{K^{w}_{ij}}}(d)$, estimated for multiplant firms. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 27: Headquarters (group-population)

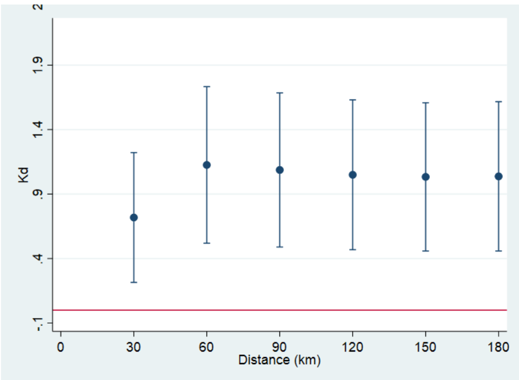
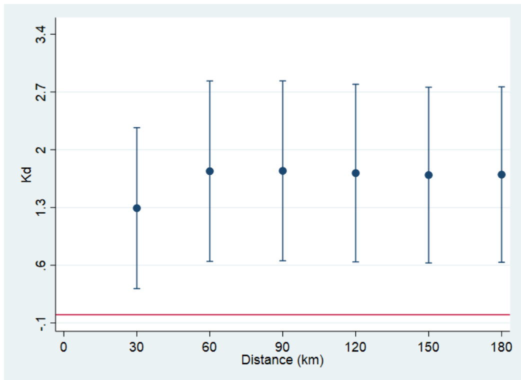


Figure A4. 28: Headquarters (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^{w}_{ij,J}}}(\quad)$ *and* $\overline{\overline{K^{w}_{ij}}}(d)$, estimated for headquarter plants. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.4.1 High/low LP

Figure A4. 29: High LP (group-population)

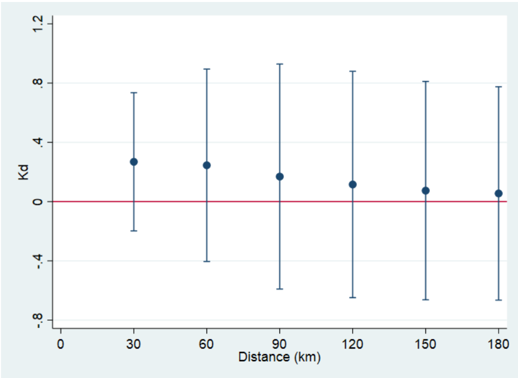
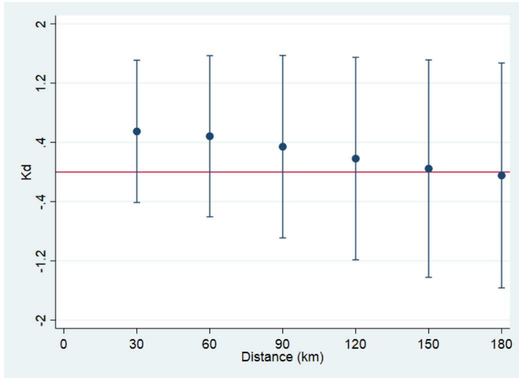


Figure A4. 30: High LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for high-productive firms (above the 90th percentile of the labour productivity distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 31: Low LP (group-population)

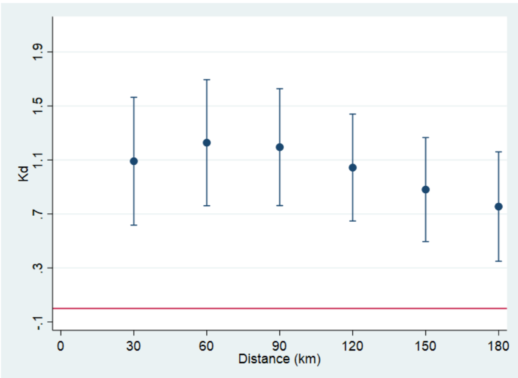
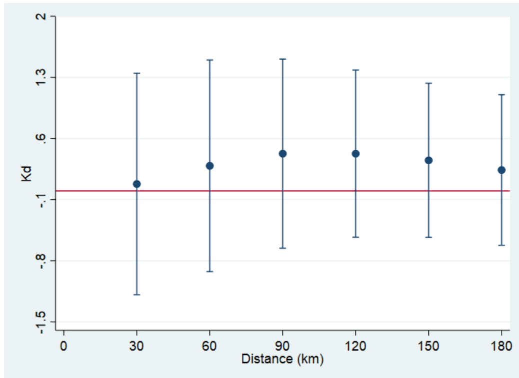


Figure A4. 32: Low LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K_{ij,J}^{w}}$ and $\overline{K_{ij}^{w}}$, estimated for high-productive firms (below the 25th percentile of the labour productivity distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.4.2 High/low LP

Figure A4. 33: High LP (group-population)

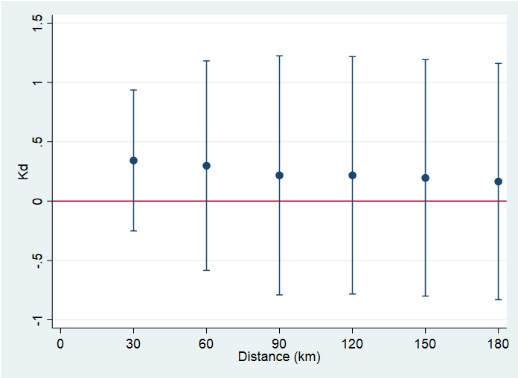
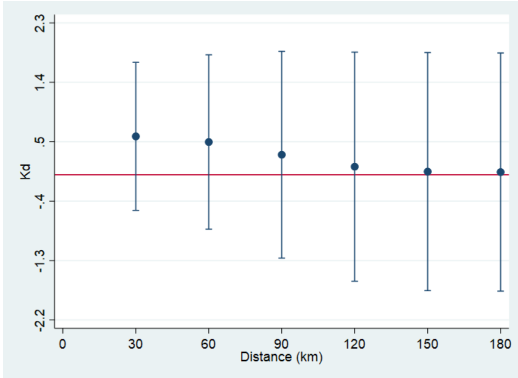


Figure A4. 34: High LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ and $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for high-productive firms (above the 90th percentile of the labour productivity distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

Figure A4. 35: Low LP (group-population)

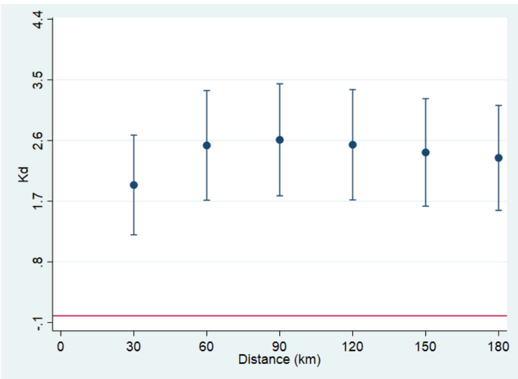
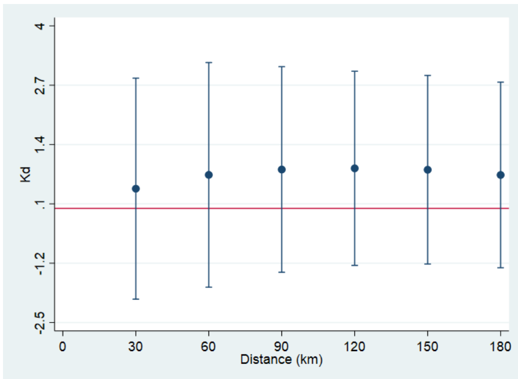


Figure A4. 36: Low LP (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K_{ij,J}^{w}}}(d)$ and $\overline{\overline{K_{ij}^{w}}}(d)$, estimated for low-productive firms (below the 25th percentile of the labour productivity distribution). The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.5.1 MNEs

Figure A4. 37: MNEs (group-population)

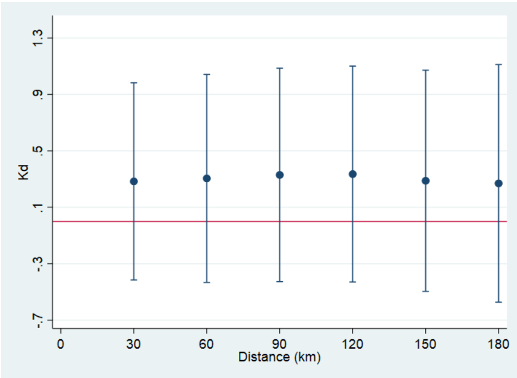
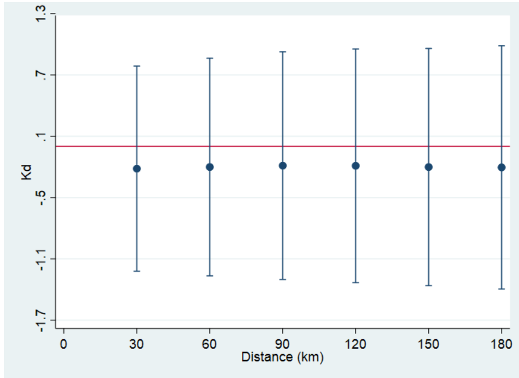


Figure A4. 38: MNEs (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{K^w_{ij,J}}$ and $\overline{K^w_{iJ}}$, estimated for multinational enterprises. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)

C.5.2 MNEs

Figure A4. 39: MNEs (group-population)

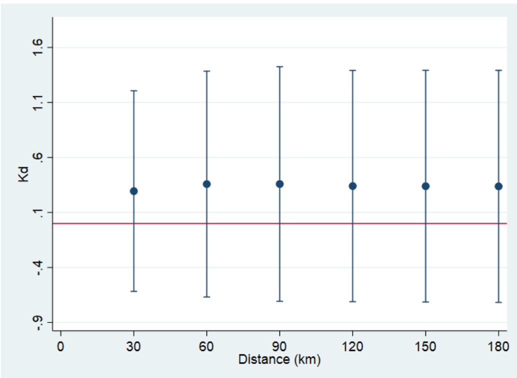
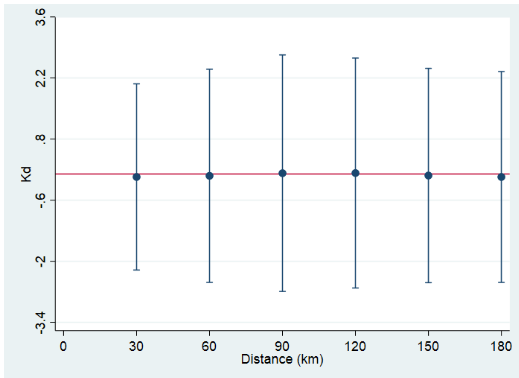


Figure A4. 40: MNEs (group-group)



The charts report coefficient estimates and confidence intervals for the specification 1 (Section 4.4.1), estimated for different distance thresholds (x axis). The dependent variables are, respectively, $\overline{\overline{K^w_{ij,J}}}(d)\ nd\ \overline{\overline{K^w_{iJ}}}(d)$, estimated for multinational enterprises. The explanatory variable is import penetration from BRICS countries, instrumented by world supply (sections 4.4.4 and 4.4.5)