Spatial Agglomeration and Superstar Firms: Firm-level Patterns from Europe and US

By Laura Alfaro, Maggie X. Chen and Harald Fadinger[1]

Abstract

We characterize the agglomeration patterns of industries and plants in Europe, distinguishing eurozone countries, and the U.S. Using a micro-level index, we quantify the degree of geographic concentration in industrial activities and explore how firm heterogeneity, industry attributes, and location fundamentals jointly explain the observed patterns. Our analysis shows that there is a clear hub-and-spoke structure in the geographic distribution. Larger and more productive plants, especially the superstars of each industry, are more centred than their smaller, less productive counterparts. The greater agglomeration surrounding superstars is particularly pronounced in the Eurozone but not present in the rest of Europe. Location fundamentals also play an important role and can sometimes mitigate the importance of agglomeration economies around large firms. Regions with different levels of economic development, including education and technology, exhibit distinct agglomeration patterns. The findings suggest heterogeneity in the ability of regional policies to build superstar-centred industry clusters.

1 Introduction

In recent decades, falling transportation costs, dismantled policy barriers, and rapid technological progress have precipitated an explosion of cross-border flows in goods, services, investments, and ideas. This phenomenon, particularly so in Europe where European integration has been predicated on the free movement of goods, services, labour and capital, can rapidly reshape the landscape of economic geography and business network. A key driver of this phenomenon is the "superstar firms", a group, first coined by Rosen (1981)[2], consisting of the very large, productive firms that have come to dominate particular industries.[3] Engaging in increasingly complex organization decisions at home and abroad and transporting

[1] Harvard business School, lalfaro@hbs.edu; George Washington University, xchen@gwu.edu; University of Mannheim, harald.fadinger@uni-mannheim.de. We thank Zineb El Melouki, Sarah Jeong, Christopher Montgomery, and Diane Tchawa for valuable research support.

[2] Rosen (1981).

[3] In this paper we define super star firms by size and activities relative to their peers. For other characterization see Autor et al. (2017). "The Fall of the Labor Share and the Rise of Superstar Firms."

products, tasks, capital, and technology across countries, super star firms have risen to the centre of globalization and industrial activities.[4]

The dominance of few firms ---multinational firms (MNCs) in particular--- and the emergence of industrial clusters often surrounding them have been an important paradox of globalization. On the one hand, it is long recognized that geographic proximity could lead to agglomeration economies including lower transport costs between input suppliers and final good producers (vertical linkages), horizontal labour-market and capital-good-market externalities due to proximity of firms with similar demand for labour and capital goods, and technology diffusion occurring at close distances. These agglomeration economies can be particularly strong around superstar and multinational firms as these firms tend to be more productive as well as more intensive in capital and knowledge. On the other hand, as the movement of goods, people and ideas becomes easier through economic integration, the benefits of agglomeration economies are expected to decline. In contrast, however, as noted by Glaeser (2010), we observe continuing dominance of superstar firms, industrial clusters, and cities despite reductions in transportation and communication costs and the competition implications of geographic concentration.

In this paper, we characterize the agglomeration patterns of industries and firms. In a sharp departure from the existing literature, instead of assuming firms are created equal, we treat each plant as the unit of observation and explore the geographic distribution of economic activities surrounding each firm. Existing evidence shows that there is significant productivity heterogeneity across firms within each industry and across countries.[5] We explore how this heterogeneity, in conjunction with benefits of agglomeration, affects the formation of industrial landscape.

Our analysis compares agglomeration patterns in the U.S. with those in the European Union. Within the European Union, we separate the Eurozone from other countries, since the Eurozone economies face deeper integration of capital markets compared to other EU economies due to the common currency. Specifically, we ask: Is there agglomeration around highly productive firms? Is agglomeration driven by multinationals? Does the Euro area share similar patterns to the U.S. and the rest of Europe? In addition to firm productivity and internationalization, what is the importance of internal market and regional characteristics and policies, such population, income, and region-specific human capital and R&D investment?

Examining how the degree of agglomeration varies with firm attributes including, productivity, size and multinational status and regional characteristics allows us to assess the potential benefits and costs provided by geographic proximity to the superstars relative to the effects of location fundamentals and the ability of regional policies to attract regional industry clusters centred around super-star firms.

4 As Mayer and Ottaviano (2007) note “internationalized firms are superstars.” They are bigger, generate higher value added, pay higher wages.

5 Existing evidence (see, e.g., Helpman, Melitz and Yeaple 2004; Alfaro, Charlton, Kanczuk, 2009; Chen and Moore, 2010; Bloom et al, 2016, Alfaro and Chen, 2018) shows that there is significant productivity heterogeneity within each industry in particular between multinational and non-multinational firms.

To take into account the role of firm heterogeneity, we develop a new micro index of agglomeration and measure the level of agglomeration centring each individual plant, following an empirical methodology introduced by Duranton and Overman (2005) (henceforth, DO) and extended in Alfaro and Chen (2014, 2019). This index treats space as a continuous metric and identifies agglomeration at the most disaggregated level. It is constructed using precise latitude and longitude information of each establishment and the distance between each pair of establishments.

Based on the index, we study how the ability to attract agglomeration varies across plants and how firm heterogeneity, reflected in productivity and size, leads to different levels of ability to attract agglomeration. Specifically, we examine how a given plant's characteristics (such as size, productivity, age, foreign ownership, and the number of products) and its industry's characteristics (such as capital intensity, skilled-labour intensity, and R&D intensity) might jointly explain the extent of agglomeration centring around the plant. This step constitutes a sharp departure from the existing literature which has focused primarily on aggregate-level agglomeration and assumed all nodes in the cluster are created equal.

To mitigate the concerns of reverse causality, we explore the dynamics in the data and examine the spatial relationship between incumbent and entrant plants. We measure the distance between each pair of incumbent and entrant firms and construct the micro index to capture the degree to which entrants agglomerate towards each individual incumbent. Exploring the agglomeration between new and existing plants enables us to mitigate the potential reverse causality between firm characteristics and the level of agglomeration. Second, we identify the role of firm characteristics in determining the level of agglomeration by comparing plants located in the same disaggregated region.

To achieve the goal, we employ a unique worldwide establishment-level dataset, WorldBase, that provides detailed physical location, ownership, and activity information for manufacturing plants in more than 100 countries. The dataset's detailed location and operation information for over 43 million plants, including multinational and domestic, offshore and headquarters establishments, makes it possible to compare the agglomeration of different types of establishment. We use the plant-level physical location information in our data to obtain latitude and longitude codes for each establishment and compute the distance between each pair of establishments within the plant’s primary industry. We then construct the index of agglomeration based on the distance each pair of establishments.

Our analysis shows that firms are far from equal within each industrial cluster. There is a clear hub-and-spoke structure in the geographic concentration of industrial activities. More productive and larger establishments are more centred by other firms than their smaller, less productive counterparts. The greater agglomeration surrounding superstar firms is most pronounced in the Eurozone followed by the U.S. In the non-Eurozone European countries, superstar plants actually attract less agglomeration. In the U.S. and in the Eurozone, MNC establishments also attract significantly more agglomeration than domestically owned plants, while this is not the case in Europe outside of the Eurozone.

The different patterns in Eurozone and non-Eurozone European countries (Eastern Europe primarily) could reflect the different scope of agglomeration economies in these regions. In Eurozone and the U.S., economic activities by superstar firms and multinationals are likely to involve more skill and capital intensive upstream tasks such as component production, while in Eastern Europe affiliates of superstar and multinational firms are more likely to engage in unskilled-labour intensive downstream tasks such as assembly where there are limited positive agglomeration economies and more negative factor- and product-market competition effects.

Region attributes also play an important role. In fact, the majority of the variation in agglomeration patterns remain to be driven by regional location fundamentals such as market access and production cost. Specifically, the regional attributes account for about 30-70 percent of the agglomeration. Exploring the heterogeneous role of superstar firms, we find that higher regional human capital levels are associated with more agglomeration around larger and more productive plants, in particular in Europe. In contrast, larger regional R&D spending is associated with less agglomeration of economic activity around these plants in Europe.

Several policy implications emerge from these results. The preliminary results suggest that policies aimed to build industrial zones and foreign investment should take into account the different abilities of firms to stimulate new entrepreneurship activities. Firms with better performance and superior economic characteristics such as greater productivity can help attract more entrants and generate a domino effect in the formation of industrial clusters. An incentive structure whereby favourable incentives are offered first to potential hub firms could be more effective than a uniform incentive system. However, the design of such an incentive structure should be cautious and carefully devised to assess the potential of agglomeration economies across regions and industries.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 discusses the methodology and the data. Section 4 describes the patterns observed and presents the emerging stylized facts. The last section concludes.

2 Overview of the Literature

This paper is closely related to several strands of literature.

First, the paper builds on the existing economic geography literature that examines domestic agglomeration.[6] The agglomeration of economic activity, as long recognized by regional and urban economists and economic historians, is one of the salient features of economic development.

Transaction costs broadly defined--and to include cost of transmitting goods and information--affect not only firms' decisions to geographically separate production tasks but also the decisions to locate next to one another. They also affect firms’

[6] See Ottaviano and Puga (1998), Head and Mayer (2004), Ottaviano and Thisse (2004), Rosenthal and Strange (2004), Duranton and Puga (2004), Puga (2010), and Redding (2010, 2011) for excellent reviews of these literatures.

productivity. Geographic proximity to large firms implies more intense competition in final good, input and factor markets. The competition lowers the prices of final goods and raises input and factor costs which may lead to less successful (productive) firms exiting from the market. Output prices are also key determinants of firm's organization choices and vertical integration (Alfaro, Conconi, Fadinger, Newman, 2016). Firm boundaries affect incentives and thus firm's productivity and hence industry performance as well.

Agglomeration can also induce costs by, for example, increasing labor and land prices. On the other hand, proximity may imply benefits.[7] Agglomeration economies, which stress the benefits of geographic proximity between individuals or firms in realizing product- and factor-market externalities and technology diffusion, play a particularly important role. These benefits include lower transport costs between input suppliers and final good producers (vertical linkages), labour-market and capital-good-market externalities due to the proximity of firms with similar demand for labour and capital goods (common pool of resources), and technology diffusion thanks to low costs of technology transfer at close distance.

An overview of this vast literature is beyond the scope of our paper; we focus below on empirical studies most closely related to our analysis.[8] An extensive body of research examines the distribution of population and production across space and the economic characteristics and effects of spatial concentrations. Important literature in urban economics, led by Ellison and Glaeser (1997, 1999), Rosenthal and Strange (2001, 2003), Duranton and Overman (2005, henceforth DO, 2008), Ellison, Glaeser and Kerr (2009), Alfaro and Chen (2010, 2019) have examined patterns of agglomeration as a function of industry characteristics. These studies shed light on the role of Marshallian agglomeration forces in explaining cross-industry variation in spatial concentration in function of industry characteristic.

Concentration and Agglomeration: Europe and US

The following recent papers evaluate the effects of agglomeration on innovation and productivity in Europe and the U.S.

Evidence on the co-location of industries in the U.S. shows that firms locate near industries that are suppliers or customers (Ellison, Glaeser, and Kerr, 2007; Kolko,

[7] Marshall (1890) first introduced the idea that concentrations of economic factors, such as knowledge, labour, and inputs, can generate positive externalities. Three factors have been emphasized by these studies: market access to suppliers and customers, labour market pooling, and technology spillovers. One set of theories about agglomeration economies emphasizes the gains that come from reduced costs of moving goods across space (Krugman, 1991a). A second set of theories emphasizes labour market pooling and the benefits of moving people across firms (Marshall, 1890). A third set argues that cities speed the flow of ideas, which creates human capital at the individual level and facilitates innovation (Jacobs, 1968). Some of these theories emphasize the benefits that come from co-location of diverse firms; others emphasize the gains from single-industry agglomerations.

[8] Another important strand of empirical literature concerns one of the key theoretical predictions of new economic geography models, that is, factor prices should vary systematically across locations with respect to market access. See, for example, Redding and Venables (2004) and Hanson (2005), Ahlfeldt et al. (2012).

2000).[9] Rosenthal and Strange (2001) find both labor market pooling and input-output linkages to have a positive impact on agglomeration. The effect of knowledge spillovers is also significant, but mostly at the local level.

Several recent studies contribute an understanding of U.S. agglomeration trends since the early 2000s. Buzard et al. (2017) map the zip codes of 1700 private R&D labs and identify four major clusters in the Northeast Corridor (Boston, New York, Philadelphia, and Washington, D.C.,) and three major clusters in California (Bay Area, Los Angeles, and San Diego).

Gutierrez and Philippon (2017) review firm clusters to test two main hypotheses that explain concentration and low investment among U.S. industries – decreasing domestic competition and the efficient scale of operation. By comparing these U.S. industries to Europe, they conclude that the efficient scale of operation cannot be the main explanation for concentration. The paper also conducts tests to show that decreasing domestic competition in the US causes low investment, concluding that it caused a shortfall of non-residential capital of 5-10% by 2016.

Additional studies attempt to measure the impact of agglomeration on industry productivity. For example, Forman, Goldfarb and Greenstein (2016) study the Bay Area's increase in innovation from 4% of all successful US patent applications in 1976 to 16% in 2008, and attribute this growth to co-agglomeration in invention across technologies.

The different areas of focus for these studies highlight various explanations for the relationship between concentration and productivity. Rigby and Essletzbichler (2002) touch on several key explanations. Using U.S. Census data of U.S. metropolitan firms, they present evidence that multiple subfactors of spatial concentration affect productivity in different ways, including input-output linkages, occupational distribution, and embodied technological spillovers.

Andersson, Burgess and Lane (2007) apply a U.S. data set to quantify the benefits of agglomeration on the matching of workers and jobs, showing positive effects of thicker urban labor markets on assortative matching in terms of worker and firm quality. Using the US Census and National Longitudinal Survey of Youth, Bacolod, Blum and Strange (2009) similarly review impacts of concentration on workers. The authors conclude that agglomeration has a larger impact on wages and productivity for work that requires thinking and social interaction rather than manual labor, thus contributing to the knowledge spillovers theory.

Hornbeck and Moretti (2010) also quantify agglomeration spillovers by tracking the impact of a new "Million Dollar Plant" on the total factor productivity (TFP) of incumbent plants in the same county. Five years after the new plant opening, TFP of incumbent plants in the "Million Dollar Plant" county is 12% higher than TFP of

[9] In a survey of the literature Glaeser and Gottlieb (2009) note there is abundant evidence that manufacturing firms in the U.S. choose location to reduce transport costs, but this does not seem to be an important part of urban comparative advantage today. The urban role in reducing transport costs seems to be more important for service firms. The largest body of evidence supports the view that cities succeed by spurring the transfer of information (skilled industries are more likely to locate in urban areas and skills predict urban success).

incumbent plants in other counties. Consistent with some theories of agglomeration economies, this effect is larger for incumbent plants that share similar labor and technology pools with the new plant. They also find a relative increase in skill-adjusted labor costs, indicating that the ultimate effect on profits is smaller than the direct increase in productivity.

Ciccone (2002) estimates that agglomeration effects on labor productivity in France, Germany, Italy, Spain, and the UK are slightly smaller than in the US, with an elasticity of labor productivity with respect to employment density of 4.5 percent compared to 5 percent in the US.

Overman and Puga (2009) examine the role of labor market pooling and input sharing in determining the spatial concentration of UK manufacturing establishments. They find sectors whose establishments experience more idiosyncratic employment volatility and use localized intermediate inputs to be more spatially concentrated.

Firms are generally more productive in larger cities. This trend is attributed to two main explanations – agglomeration economies (cities promote interactions) and firm selection (fierce competition weeds out unproductive firms). Several European studies support the agglomeration economies theory. For example, Helmers and Overman (2017) provide evidence that proximity to a large scientific research facility in the U.K. disproportionately benefits institutions that are closer to the infrastructure through improved distribution of knowledge. Additionally, Combes et al. (2012) use French establishment-level data to provide evidence in favor of the agglomeration theory and to challenge the hypothesis that firm selection explains productivity differences.

Fritsch and Changoluisa (2017) find support for the firm selection theory by assessing correlations between new start-ups and productivity. Using evidence across 71 West German planning regions, the study finds that new businesses – not just innovative, technologically advanced firms – induce higher productivity in incumbents. They do not find significant benefits generated by knowledge spillovers or the provision of better inputs, attributing productivity instead to fiercer market competition.

Similarly, Gordon and McCann (2005) agree that while agglomeration explains innovation dynamics in London, firms do not perceive advantages of informal information spillovers from agglomeration. Their analysis comes from surveys of London firms, so it would be interesting to compare perceived effects with actual effects of knowledge spillovers on productivity.

Alfaro, Conconi, Fadinger and Newman (2018) find the higher domestic prices the more vertically integrated are the firms producing that product in that country. The effect is larger precisely where organizational decisions ought to be more responsive to domestic prices, i.e., for firms that only serve the domestic market. These results suggest that policies that affect product prices can have direct effects on firm organization.[10]

By contrast, the role of firm attributes in shaping the formation of clusters remains mostly unknown.[11] However, the international trade literature has paid particular attention to the role of multinational firms and examined their agglomeration patterns, incentives and implications.

Agglomeration, Trade and Multinationals

MNCs are likely to exhibit different motives of agglomeration than domestic firms due to their greater revenue and productivity, vertically integrated production, and higher knowledge- and capital-intensities. In contrast to domestic production, which emphasizes domestic geography and natural advantage, multinational production stresses foreign market access and international comparative advantage. Moreover, as highlighted in a growing literature (e.g., Helpman, Melitz, and Yeaple, 2004; Antras and Helpman, 2004, 2008; Alfaro and Chen, 2018), the economic attributes and organization of multinationals are, by selection, distinctively different from average domestic firms. Thus, the advantage of proximity can differ dramatically between multinational and domestic firms.

Compared to domestic firms, multinationals are often the leading corporations in each industry with large volumes of sales and intermediate inputs. Externalities in the movement of workers from one job to another can also affect MNCs which are characterized by similar skill requirements and large expenditures on worker training. MNCs can have a particularly strong incentive to lure workers from one another because the workers tend to receive certain types of training that are well suited for working in most multinational firms (business practices, business culture, etc.). Moreover, MNCs' proximity to one another shields workers from the vicissitudes of firm-specific shocks. External scale economies can also arise in capital-good markets.[12] MNCs may also face significant market entry costs when relocating to a foreign country because of, for example, limited supplies of capital goods. An additional motive relates to the diffusion of technologies. given their technology intensity, technology diffusion from proximity to technologically linked firms and industries can be particularly attractive to MNCs. Technology can diffuse from one firm to another through movement of workers between companies, interaction

[10] The authors also study the effect of trade policy on the degree of organizational convergence across countries as the theory suggests that countries with similar domestic price levels should have firms with similar ownership structures. Differences in vertical integration across countries is significantly larger in sectors in which differences in domestic prices to be larger. Differences in vertical integration indices are smaller for country pairs engaged in regional trade agreements; this effect being stronger for customs unions, which impose common external tariffs vis-à-vis non-members and should thus be characterized by stronger price convergence.

[11] Most research, however, has tended to focus on the effects of industry characteristics and regional natural advantage, treating each industrial cluster as a homogeneous entity Rosenthal and Strange (2003) offer one of the few disaggregated analyses in this area.

[12] Geographically concentrated industries offer better support to providers of capital goods (e.g. producers of specialized components and providers of machinery maintenance) and reduce the risk of investment (due to, for example, the existence of resale markets). Local expansion of capital intensive activities can consequently lead to expansion of the supply of capital goods, thereby exerting a downward pressure on costs.

between those who perform similar jobs, or direct interaction between firms through technology sourcing.[13]

The literature has found consistent evidence that MNC agglomeration patterns differ from those of their domestic counterparts.

In the field of international trade, the advantage of proximity and low transport cost between customers and suppliers has received particular attention. A number of studies have examined the role of production linkages in multinationals' location decisions (see, e.g., Head, Ries and Swenson, 1995; Head and Mayer, 2004; Crozet, Mayer and Mucchielli, 2004; Blonigen, Ellis and Fausten, 2005; Bobonis and Shatz, 2007; Amiti and Javorcik, 2008; Debaere, Lee and Park, 2010). These studies show that MNCs with vertical linkages tend to agglomerate regionally in countries such as the U.S., China, and the EU.

A number of studies, including Head, Reis and Swenson (1995) and Blonigen, Ellis and Fausten (2005), exploit the Japanese institution of vertical keiretsu and examine the location interdependence of vertically linked Japanese plants. The evidence there suggests that members of the same keiretsu tend to choose the same states in the United States.

For example, Head, Ries, and Swenson (1995) estimate the location choices of Japanese firms who set manufacturing factories in the US during the period 1980-1992. They find that Japanese investments do not mimic domestic plants; rather, their agglomeration is driven by positive externalities of colocation rather than fundamental forces (such as infrastructure, natural resources, and labor). The authors note that the dependence of Japanese manufacturers on the "just-in-time" inventory system exerts a particularly strong incentive for vertically linked Japanese firms to agglomerate.

Head and Mayer (2004) study the location choices of Japanese firms in Europe, and find that regions with a greater market potential (larger number of existing foreign affiliates) are more likely to be selected by multinationals. The authors find fundamental forces (market potential) to matter. In particular, the authors find a 10 percent increase in a region's market potential to increase the likelihood of multinational entry by 3 to 11 percent. However, these forces do not fully explain location choices as they can also be driven by forces of agglomeration.

Crozet, Mayer and Mucchielli (2004 and Bobonis and Shatz(2007), study the determinants of location choices by foreign investors in France and in the U.S., respectively, finding evidence of clustering. The authors find that targeted policies influence foreign investments while regional or state level policies do not seem to affect the location of FDI. Crozet, Mayer and Mucchielli (2004) find agglomeration forces to be an important determinant of foreign firm investments in France, while Bobonis and Shatz (2007), using data on the U.S. state-level stock of foreign-owned

[13] This has been noted by Barba Navaretti and Venables (2004), who predict that MNCs may benefit from setting up affiliates in proximity to other MNCs with advanced technology (e.g., "so-called centres of excellence"). Affiliates can benefit from technology spillovers, which can then be transferred to other parts of the company.

property, plant, and equipment (PPE), find agglomeration to be an important externality.

Alfaro and Chen (2014, 2019) assess the different patterns underlying the global agglomeration of multinational and non-multinational firms using a spatially continuous index of agglomeration and a unique worldwide plant-level dataset from World Base. The analysis shows that the offshore agglomeration patterns of MNCs are distinctively different from those of their headquarters and their domestic counterparts.

3 Data

3.1 Firm Data Cross-Country Coverage

Our empirical analysis uses a unique worldwide establishment dataset, WorldBase, that covers more public and private establishments in more than 100 countries and territories.

WorldBase is compiled by Dun & Bradstreet (D&B), a leading source of commercial credit and marketing information since 1845. D&B--presently operating in over a dozen countries either directly or through affiliates, agents, and associated business partners---compiles data from a wide range of sources including public registries, partner firms, telephone directory records, and websites.[14]

WorldBase reports, for each establishment in the dataset, detailed information on location, ownership, and economic activities. Four categories of information are used in this paper: (i) industry information including the four-digit SIC code of the primary industry in which each establishment operates; (ii) ownership information including headquarters, domestic parent, global parent, status (for example, joint venture and partnership), and position in the hierarchy (for example, branch, division, and headquarters); (iii) detailed location information for both establishment and headquarters; and (iv) operational information including sales, employment, and year started.

D&B's WorldBase is, in our view, an ideal data source for the research question proposed in this study. It's broad coverage and detailed plant location information enables us to examine agglomeration on a global and continuous scale. Viewing agglomeration on a continuous scale is important in light of the increasing geographic agglomeration occurring across regional and country borders as we explain in detail in the next section.[15] In addition, the database reports detailed

[14] For more information, see: http://www.dnb.com/us/about/db_database/dnbinfoquality.html. The dataset used in this paper was acquired from D&B with disclosure restrictions. See Alfaro, Conconi, Fadinger and Newman (2016) for a detailed description of the data.

[15] Examples of cross-border clusters include the metalworking and electrical-engineering cluster involving Germany and German-speaking Switzerland; an electric-machinery cluster involving Switzerland and Italy; a biotech cluster spreading across Germany, Switzerland, and France; an automobile industry cluster that crosses the border of Germany and Slovakia; the Ontario-Canada-Michigan-US (Windsor-Detroit) auto cluster; and the South US-Northeastern -Mexico cluster.

information for multinational and non-multinational, offshore and headquarters establishments. This makes it possible to compare agglomeration patterns across different types of establishment and to investigate how the economic geography of production evolves with forms of firm organization.

In this paper, we restrict analysis to comparisons between countries in Europe and the United States. Appendix Tables A1a lists the countries included. We use the 2004/5 vintage of the data set.[16] In the main analysis, we limit to manufacturing sectors for tractability.

In terms of the final sample, an establishment is deemed an MNC foreign subsidiary if it satisfies two criteria: (i) it reports to a global parent firm, and (ii) the headquarters or the global parent firm is located in a different country. The parent is defined as an entity that has legal and financial responsibility for another establishment. We drop establishments with zero or missing employment values and industries with fewer than 10 observations.[17]

3.2 Geocode Information

Using postal code information of each plant in the data set we obtain latitude and longitude codes for each establishment using different methods.

We obtained data from the Geocoding Databases for Europe, a Database including latitudes and longitudes of cities and postcodes of most European countries for free download and from GeoNames, a website of geographical database covers all countries.[18] We also use the software Google's Geocoding API services, well known as an industry standard for transportation data, to verified the data. The software provides more accurate geocode information than most alternative sources.

We apply the Haversine formula to the geocode data to compute the great-circle distance between each pair of establishments. We limit the analysis to firms within a given 3-digit manufacturing sectors for computational reasons.

3.3 Additional Data

We examine activity at the region, rather than the country, level and include a series of regional characteristics, such as market size, natural and comparative advantages, as additional regressors to capture the effect of regional location fundamentals.

[16] We also preliminarily explored related patterns using the 09/12/18 vintages.

[17] Requiring positive employment helps to exclude establishments registered exclusively for tax purposes.

[18] The websites are: https://www.clearlyandsimply.com/clearly_and_simply/2010/10/geocoding -databases-for-europe.html and geonames.org, respectively.

For Europe, the data was compiled from the Eurostat Regional Database at the NUTS 2 level of disaggregation. For the US we obtained information at primarily the state or province level.[19]

The regional characteristics systematically available across countries and included in our final sample are GDP per capita, population density, schooling (percentage of labour force with more than secondary education), all measured in 2004 or the closest year available (to mitigate causality concerns). We also include regional R&D expenditure. We also use the OECD STAN data and NBER-CES Manufacturing Industry Database to construct industry's capital and skilled-labour intensities, which are defined as, respectively, the ratio of investment and of non-production workers' payroll to value added. Each industry's R&D intensity is measured using the median firm's ratio of R&D expenditure relative to value added based on the COMPUTSTAT database.[20]

4 Methodology

In this section, we describe the empirical methodology we use to quantify the global agglomeration of firms.

We compute plant-level agglomeration densities to measure the degree to which a plant is proximate to other plants following an empirical methodology introduced by Duranton and Overman (2005) and extended in Alfaro and Chen (2014, 2019). The index contains information on the extent of localization by industry and the spatial scales at which it takes space. In contrast to traditional indices, which tend to define agglomeration as the amount of activity taking place in a particular geographic unit, this spatially continuous index separates agglomeration from the general geographic concentration and is unbiased with respect to the scale of geographic units and the level of spatial aggregation.

As noted in Head and Mayer (2004b), measurement of agglomeration is a central challenge in the economic geography literature. There has been a continuous effort to design an index that accurately reflects the agglomeration of economic activities.

[19] For the US, population and education attainment data were collected from the U.S. Census; GDP and income/compensation statistics were collected from the Bureau of Economic Analysis; roadway statistics were from the Federal Highway Administration; employment data was collected from the Bureau of Labour Statistics, all at the state level. Port data was from World Port Source, and tax rates were compiled from Ernst and Young, Deloitte, KPMG, and the World Bank's Doing Business report.

[20] In additional robustness we also use upstreamness (average distance from final use) measures from Antràs and Chor (2013). Constructing the proxies of agglomeration economies using the U.S. industry account data is motivated by three considerations. First, compared to firm-level input-output, factor demand, or technological information, industry-level production, factor and technology linkages reflect standardized production technologies and are relatively stable over time, limiting the potential for the measures to endogenously respond to agglomeration. Second, using the U.S. as the reference further mitigates the possibility of endogenous production, factor, and technology linkage measures, even though the assumption that the U.S. production structure carries over to other countries\could potentially bias our empirical analysis against finding a significant relationship. Third, the U.S. industry accounts are more disaggregated than most other countries', enabling us to dissect linkages between disaggregated product categories.

Most existing indices have tended to equalize concentration and colocation with activities located in the same administrative or geographic region (measured by the number of firms or the size of production in the region). Three issues arise with these measures. First, these indices can be strongly driven by industrial concentration. Industries with a small number of establishments may appear spatially concentrated when they are not. Second, many indices cannot separate general geographic concentration due to location attractiveness from agglomeration. Manufacturing plants can be attracted to the same location because of location characteristics but this is interpreted as agglomeration. The index developed by Ellison and Glaeser (1997) provides a solution to the above two issues.

Duranton and Overman (2005)[21] address the unresolved issue of the dependence of the existing indices on the level and method of geographic disaggregation, and develop a "continuous-space concentration index". By equating concentration with activities in the same region, previous indices omit concentrated activities separated by administrative or geographic borders while overestimating the degree of concentration within the same administrative or geographic units.

DO's index thus exhibits several properties essential to agglomeration measures. It is comparable across industries and captures cross-industry variation in the level of agglomeration. The index controls for industrial concentration within each industry. Its construction is based on a counterfactual approach and controls for the effect of location factors (such as market size, natural resources, and policies) that apply to all manufacturing plants. By taking into account spatial continuity, the index is unbiased with respect to the scale and aggregation of geographic units. In addition, the index offers an indication of the statistical significance of agglomeration. The estimated parameters of the variables represent the net effect, cost and benefits, of similar factor demand structures on agglomeration decisions.

However, the construction of this index poses two constraints. First, the index requires detailed physical location information for each establishment. As described above, the WorldBase dataset, supplemented by a geocoding software, satisfies this requirement. Second, the approach is extremely computationally intensive, especially for large datasets.

4.1 Agglomeration Indices

The empirical procedure to construct the agglomeration index has three steps.

In the first step, we estimate an actual geographic density function for each establishment in a given industry based on the distance to every other plant in the same industry that was established after the establishment date of the incumbent plant. In the second step, we obtain counterfactual density functions based on establishments in the same industry to control for factors that affect all plants in the industry. In the last step, we construct the agglomeration index to measure the extent

[21] Duranton and Overman (2005), DO, construct an index to measure the significance of agglomeration in the U. K. DO's index has been adapted by other studies such as EGK's measurement of the agglomeration of U.S. pairwise industries.

to which an establishment in a given industry attracts agglomeration at a threshold distance relative to the counterfactuals.

Step 1: Kernel Estimator

We first estimate an actual geographic density function for each establishment in a given industry.

First, we obtain, for each establishment i with primary industry k, the kernel estimator of bilateral distances at any point d $(\mathrm{i.e.}, f_i(\mathrm{d}))$. Formally, we obtain

$$f_i(d) = \frac{1}{n_i h} \sum_{j:density_k(T)>0} K\left(\frac{d-d_{ij}}{h}\right), \qquad (1)$$

where n_i is the cardinality of i's industry cluster, h is bandwidth, and K the kernel function. All kernel estimates are calculated using a Gaussian kernel with the bandwidth set to minimize the mean integrated squared error.

Note that even when the locations of nearly all establishments are known with a high degree of precision (as is the case with the data we use, as described below), distance---and estimated trade cost---are only approximations of the true trade cost between establishments. One source of systematic error, for example, is that the travel time for any given distance might differ between low- and high-density areas. Given the potential noise in the measurement of trade costs, we follow DO in adopting kernel smoothing when estimating the distribution function. We limit the analysis to firms within the same 3-digit sector to ease the computation burden.

Step 2: Counterfactuals

In the second stage, we construct a counterfactual kernel estimator for each establishment, i.e., $\bar{f}_i(d)$. We use here the mean kernel estimates of each industry as the counterfactual. This enables us to control for all factors common to establishments in the same industry and to focus on each establishment's deviation from its average counterpart.

We compare the kernel estimators at various distance thresholds. We focus on 50km but we also considered lower thresholds (10, 20) and higher distance thresholds, such as 200, 400, and 800 km.

Step 3: Agglomeration Density

Finally, we construct the density index for each establishment, i.e.,

$$density_i(T) \equiv \sum_{d=0}^{T} \left(\mathrm{f_i}(\mathrm{d}) - \bar{f}_i(d)\right). \qquad (2)$$

This index captures the relative probability that other establishments agglomerate with i, as opposed to i's counterfactuals, within distance T.

Establishments with the greatest density are the hubs of each cluster whereas those with relatively low densities emerge in the periphery.

4.2 Empirical Procedure

With the plant-level agglomeration densities at hand, we measure the degree to which a plant is proximate to other plants and examine how plant characteristics (such as productivity, ownership structure, size, age and the number of products), and industry characteristics (such as capital intensity, skilled-labour intensity, and R&D intensity) might jointly explain the extent of agglomeration centred around each plant

We run the following specification:

$$Density_i(T) \equiv \alpha + \beta\theta_i + \gamma Z_r + D_k + \varepsilon_i, \qquad (3)$$

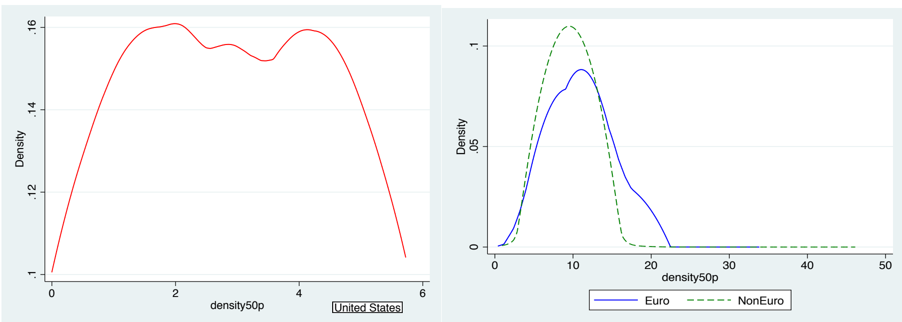
where $Density_i(T)$ is the estimated density of establishment i's network that captures the probability of other establishments agglomerating around i, as opposed to i's counterfactuals in the same host country and industry, within a threshold distance T.

We obtain estimates of $Density_i(T)$ based on the previously methodology for different thresholds (50, 100, 200, 400 km). For our baseline results, we report the estimation results based on plant-level agglomeration indices at 50 km.

On the right hand side of equation (3) we use labour productivity as our main measure of firm performance.

We include a vector of industry dummies, represented by D_k, to control for industry specific factors. We include series of geographical controls Z_r, to control for regional variables. This enables us to focus on the effect of heterogeneity in determining the extent of agglomeration.

Figure 1: Densities Across Regions



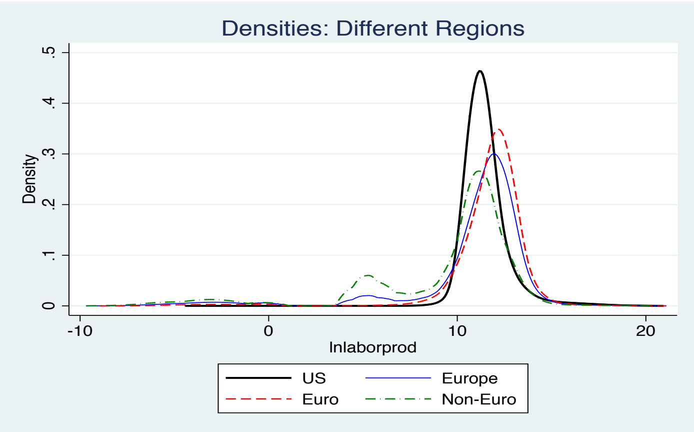
Panel A: United States Panel

Panel B: Euro and Non-Euro countries

Appendix table A2a provides main descriptive statistics of the agglomeration indices by region. The plant-level agglomeration index captures the probability relative to the industry average to get an entrant in the same 3-digit sector within less than T km from the location of the incumbent. Because the relative entry probability has a low baseline value, we scale the percentages by 100 for better readability.[22] Two stylized facts emerge from the Table. The mean agglomeration density at 50 km is 2.9[23] in the U.S., 11.6 in the Eurozone and 14.6 in the rest of Europe. Thus, Europe features more agglomeration compared to the U.S. Comparing 20km and 50km indices, mean agglomerations close to double in all regions.

Appendix Tables A2b and 2c provide summary statistics for the main variables of analysis. Our main explanatory variable of interest will be plant performance, measured as labour productivity (in U.S. Dollars). Figure 2 shows the distribution of plant-level log labour productivity across different regions. Labour productivity is approximately log normal both in Europe and in the U.S. In our sample, plant-level labour productivity in manufacturing is slightly larger in Europe than in the U.S. Within Europe, the Eurozone has significantly more productive plants than the rest of Europe, where plants have on average lower productivity than in the U.S.

Figure 2: Productivity: Densities Across Regions



5 Results

5.1 Agglomeration and Firm Performance

We first investigate the relationship between the density of economic activity within a 50-kilometer distance around the location of the plant (within a given 3-digit sector) and plant-level characteristics. We run the following regression at the plant level:

$$Density_{i,k}(T) \equiv \beta_1 \mathrm{performance}_i + \beta_2 \mathrm{multinational}_i + \beta_3 \log(\mathrm{age}_i) + \beta_4 \mathrm{multiproduct}_i + \gamma' \mathrm{X}_r + \delta_k + \varepsilon_i, \quad (4)$$

where $\mathrm{performance}_i$ is log(labour productivity), $\mathrm{multinational}_i$ is a dummy for multinational affiliate, $\mathrm{multiproduct}_f$ is a dummy for the plant being active in multiple

[22] The scale of the agglomeration index is driven by the scope of the dataset and the empirical methodology. Because we take into account the distance of all establishment pairs across continents, kernel estimates at each distance level will be low.

[23] This corresponds to 0.03, 0.12 and 0.14 percentage points, respectively.

4-digit industries, δ_k is a 3-digit industry fixed effect. The industry fixed affects control for differences in industry factors which may affect the relationship between regional economic density and plant performance. Finally, X_r is a vector of region controls, that includes, regional population density and per capita GDP, the fraction of the population who have successfully completed post-secondary education in the regional population and regional R&D spending (in logs). These regional variables control for fundamental factors, as well as policies that may affect regional productivity and thereby impact both on economic activity and firm performance.

To allow for geographic heterogeneity of the effects, we always estimate the regressions separately for four macro regions: the U.S., Europe, and we also separate Europe into the Eurozone and other European economies.

We first present results for log(labour productivity) as a measure of firm performance (Table 1). We find that the degree of agglomeration varies sharply across plants in the same industry. At 50 km, labour productivity matters. Both in the U.S. and in Europe (even though the relationship is weaker), there is a positive association between the density of economic activity and plant-level (log) labour productivity. Plants with larger labour productivity tend to attract significantly more agglomeration. However, within Europe there is a stark difference between Eurozone countries, where the association between labour productivity and agglomeration is very strong and non-Eurozone countries, where this relationship is much weaker, unless a full set of regional controls is included. In terms of economic magnitudes, a one-standard-deviation change in log labour productivity, is associated with an increase in economic density by around 0.05 standard deviations in the U.S., 0.1 standard deviations in the Eurozone and 0.005 standard deviations in the rest of Europe.[24]

There is also more agglomeration around older firms in the U.S. and in the Eurozone. The age control is positive in the U.S. and in the Eurozone but negative in Europe outside of the Eurozone. Finally, the number of products produced by each plant has a significant effect on agglomeration. There is more agglomeration around multiproduct firms.

[TABLE 1 HERE]

Multinational Firms Similarly, both in the U.S. and in Europe, there is more agglomeration of economic activity around affiliates of multinational companies and the effect is of similar magnitude in both regions. Again, this similarity hides substantial heterogeneity within Europe: while the association between agglomeration and the plant's multinational status is positive and large in the Eurozone, the same relationship is negative outside of the Eurozone, i.e. in non-Eurozone Europe, multinationals attract less agglomeration. In terms of economic magnitudes, in the U.S. multinational status of the plant is associated with a 0.22 standard-deviation increase in agglomeration. In the Eurozone, the increase in

[24] These numbers are based on the coefficients from the first specification for each region and are computed using the standard deviations of the explanatory and dependent variables reported in Appendix Table A.1.a. In particular, 0.05=0,09*1.05/1.72 for the US, 0.1= 0.24*2.13/5.2 for the Eurozone and 0.005=-0.005*4.16/4.32 for the rest of Europe.

agglomeration corresponds to 0.37 standard deviations of agglomeration, while in Europe outside of the Eurozone multinational status is associated with a decrease in agglomeration of 0.35 standard deviations.[25]

Finding 1: There is more agglomeration of economic activity around more productive plants, in particular in the U.S. and in the Eurozone.

Finding 2: There is more agglomeration of economic activity around affiliates of multinationals, in particular in the U.S. and in the Eurozone, but not in Europe outside of the Eurozone.

In Appendix Table A3 we show that the above results are robust to measuring the density of economic activity using a 100-kilometer or 200-kilometer distance around the plant instead of using a 50-kilometer distance.

Superstar Firms Next, we test if "superstar" firms, defined as plants that are within the top 5% or top 1% of the labour productivity distribution within a given 2-digit sector within each region[26], attract additional agglomeration compared to more productive plants. We thus add to our previous specification a "superstar" dummy that equals one if a given plant belongs to the top 5% of labour productivity in a 2-digit sector. (We also use a top 1% cut-off in an alternative specification, which delivers similar results. Results available upon request.). The regression specification is now given by:

$$Density_{i,k}(T) \equiv \beta_1 \mathrm{performance}_i + \beta_2 \mathrm{multinational}_i + \beta_3 \log(\mathrm{age}_i) + \beta_4 \mathrm{multiproduct}_i + \beta_5 \mathrm{superstar}_i + \gamma' \mathrm{X}_r + \delta_k + \varepsilon_i \qquad (5)$$

The results of this specification are presented in Table 2. In the U.S., the superstar dummy is positive and significant, indicating that superstar plants attract additional agglomeration compared to plants that are simply more productive (around 0.08 standard deviations more agglomeration). Similarly, in Europe superstar status is also associated with additional agglomeration but this hides heterogeneity. In the Eurozone, the superstar dummy is positive, highly significant and very large (it corresponds to an around 0.33 standard-deviations increase in agglomeration). Around half of the superstar effect is driven by multinationals, as can be seen from the next column, where the superstar-dummy decreases from 1.2 to 0.6 once we add the multinational dummy. This result suggests that smaller plants decide to agglomerate towards highly productive plants, in particular affiliates of MNCs, and that that benefits of being close to these more internationalized firms outweigh the costs. However, this is not the case for plants outside of the Eurozone, where the association between of superstars and agglomeration is strongly negative

[TABLE 2 HERE]

[25] These numbers are based on the coefficients from the first specification for each region and are computed using the standard deviations of the dependent variables reported in Appendix Table A.1.a. In particular, 0.22=0.38/1.72 for the US, 0.37= 1.93/5.2 for the Eurozone and -0.35=--1.57/4.32 for the rest of Europe.

[26] We calculate supersize firms for US, Europe, eurozone and non eurozone countries.

Finding 3: There is more agglomeration of economic activity around superstar plants in the Eurozone, and in the U.S. There is less agglomeration y of economic activity around superstar plants in Europe outside of the Eurozone.

Overall, the association between plant performance and agglomeration is stronger in the U.S. and in the Eurozone compared to Europe outside of the Eurozone, indicating that plants in the first two regions benefit more from spillover effects from their high-performance competitors.

5.2 Agglomeration, Regional Policies, and Firm Performance

We now address the question if there are any regional policies, such as investment in human capital or R&D spending that are associated with more agglomeration of economic activity or if these effects are driven more by standard agglomeration forces, such as population or income.

Looking again at Table 1, we add regional control variables in column 3. First, note that these regional variables explain a major share of the variation in agglomeration. The R-squared of the first two specifications, which include only plant-level controls and industry dummies, is around 0.05 to 0.3, i.e. these variables explain around 5 to 30 percent of the variance of plant-level agglomeration. When adding the regional controls, the R-squared increases to between 0.3 and 0.7, so that these regional variables explain an additional 15 to 40 percent of the variation in agglomeration across plants.

Focusing on the role of individual regional variables, the association between income per capita and agglomeration is weak in the U.S. and in the Eurozone and not statistically significant. By contrast, the relationship is negative in Europe outside of the Eurozone, where richer regions attract significantly less agglomeration.

Moreover, in all macro regions there is a positive relationship between population density and agglomeration, even though it is less significant in the Eurozone than in the U.S. and in Europe outside of the Eurozone. In terms of economic magnitudes, a one-standard-deviation increase in regional population density is associated with a 0.7 standard-deviation increase in agglomeration in the U.S., while these effects are weaker in Europe. (0.03 s.d. in the Eurozone, and 0.25 s.d. in the rest of Europe.)

In terms of policy variables, both regional human capital and R&D investment have significant effects on agglomeration. The association between the fraction of the population with post-secondary education and agglomeration is negative in all regions and statistically significant in the U.S. and in Europe outside of the Eurozone: a one-standard-deviation increase in this variable reduces plant-level agglomeration by 0.17 standard deviations in the U.S., by 0.28 standard deviation in the Eurozone and by 0.3 standard deviations in the rest of Europe. Finally, regional R&D spending has no significant impact on agglomeration in the U.S., while the same variable is highly positively correlated with agglomeration in Europe. Separating again the Eurozone from the rest of Europe, we see that results for Europe are driven by the Eurozone economies. In terms of economic magnitudes: a one-standard-deviation increase in regional R&D spending is associate with a 0.2 standard-deviation decrease in agglomeration in the U.S., a remarkable 2 standard-deviation increase in the Eurozone and a 0.08 standard deviation increase in the rest of Europe.

Overall, regional R&D investment plays a prominent role in explaining differences in agglomeration across European regions, while regional variation in human capital is a more important driver of agglomeration in the U.S.

Finding 4: Higher levels of regional human capital are associated with less agglomeration of economic activity in manufacturing in all macro regions (US; Eurozone, rest of Europe); Higher levels of regional R&D spending are associated with more agglomeration in manufacturing in Europe, in particular inside of the Eurozone.

Next, we investigate the role of regional variables in attracting agglomeration around high-performance plants. To this end, we augment our previous specification by interacting firm performance measures with regional variables. In particular, we use interactions of firm-level log labour productivity with: population density, GDP per capita, post-secondary schooling, and regional R&D spending (in logs). We first add the regional variables and their interaction with firm performance one by one and then include them simultaneously in the regressions. The modified regression specification is:

$$Density_{i,k}(T) \equiv \beta_1 \mathrm{performance}_i + \beta_2 \mathrm{multinational}_i + \beta_3 \log(\mathrm{age}_i) + \beta_4 \mathrm{multiproduct}_i + \beta_5 \mathrm{performance}_i \mathrm{X}_r + \gamma' \mathrm{X}_r + \delta_k + \varepsilon_i \qquad (6)$$

The results of this specification are presented in Tables 3a and 3b.

[TABLE 3a and 3b HERE]

We emphasize here on the role of policy variables and first focus on regional human capital. The direct impact of post-secondary schooling is negative in all macro regions except for the Eurozone (at least in the specifications including all region controls), whereas the interaction of regional human capital with plant-level labour productivity varies across regions. It is negative in the U.S. and in the Eurozone and positive in the rest of Europe. Thus, while in the U.S. and in the Eurozone a higher level of regional human capital is associated with relatively less agglomeration around highly productive plants, in Europe outside of the Eurozone it induces relatively more agglomeration around high-performance plants.

Turning to regional R&D spending, the direct impact of regional R&D on agglomeration is positive in the U.S. and in the Eurozone and negative in the rest of Europe. The sign of the interaction term with labour productivity varies across regions as well: it is negative in the U.S. and in the Eurozone and positive in Europe outside of the Eurozone. Hence, higher regional R&D spending is associated with comparatively weaker agglomeration effects around high-productivity plants in the

U.S. and in the Eurozone and more agglomeration around high-productivity plants in Europe outside of the Eurozone.[27]

In Appendix Tables A4a and A4b we repeat the same specifications using interactions with superstar dummies. Again, results are similar.

Finding 5: Higher levels of regional human capital and R&D spending are associated with more agglomeration. High-performance plants in the U.S. and in the Eurozone, however, are able to rely less on the regional policies, while they are associated with more agglomeration around high-performance plants in Europe outside of the Eurozone.

5.3 Agglomeration, Firm Performance and Sector Characteristics

We now investigate how sector characteristics affect agglomeration. In particular, we focus on the sectoral intensity in non-production workers (a proxy for skill intensity), R&D intensity, and capital intensity (measured as investment over value added). Thus, instead of including sector fixed effects, we first directly control for these sector characteristics. The regression specification is now given by:

$$Density_{i,k}(T) \equiv \beta_1 \text{performance}_i + \beta_2 \text{multinational}_i + \beta_3 \log(\text{age}_i) + \beta_4 \text{multiproduct}_i + \gamma' \text{X}_r + \mu' \text{s}_k + \varepsilon_i \qquad (7)$$

where s_k is the characteristic of sector i (skill intensity, R&D intensity, capital intensity).

Results for each macro region are reported in the first two specifications of Table 4.

[TABLE 4 HERE]

The first specification only includes firm and sector controls, while the second one adds regional controls. As is obvious from the R-squared of these specifications, firm and sector characteristics explain only 2 to 19 percent of the variation in agglomeration across firms, while region characteristics mostly account for the major share of the variation in this variable. Thus, while sector characteristics are not that important for explaining variation in agglomeration patterns, they do have some impact on agglomeration. We first discuss the role of R&D intensity. The coefficient on this variable is negative in the U.S. and in Europe outside of the Eurozone, indicating that more R&D intensive sectors attract less agglomeration in these regions. Inside the Eurozone, this variable has no significant effect on agglomeration. The impact of higher skill intensity is also heterogeneous across regions: while more skill intensive sectors attract less agglomeration in the US, the opposite is the case in Europe, both inside and outside of the Eurozone. Finally,

[27] Among other variables, tax policy had a positive relation with agglomeration in eurozone countries (not significant) while negative and significant in noneuro countries. In all regions, firms tend to relatively agglomerate more in urban areas than in intermediate and rural ones.

more capital-intensive sectors feature more agglomeration, as indicated by the positive significant coefficients on sectoral capital intensity.[28]

Finding 6: More capital-intensive sectors attract more agglomeration in all regions and more skill-intensive sector attract more agglomeration in Europe, particularly in the Eurozone

Next, we turn to the question of how sector characteristics interact with plant performance in shaping agglomeration patterns. We thus modify our regression specification and include an interaction term between plant performance and sector characteristics:

$$Density_{i,k}(T) \equiv \beta_1 \text{performance}_i + \beta_2 \text{multinational}_i + \beta_3 \log(\text{age}_i) + \beta_4 \text{multiproduct}_i + \beta_5 \text{performance}_i s_k + \gamma' X_r + \delta_k + \varepsilon_i \quad (8)$$

where s_k is the characteristic of sector i (skill intensity, R&D intensity, capital intensity). When including sector fixed effects, the direct impact of sector characteristics is absorbed by them.

Results for these regressions are presented in the third, fourth and fifth specifications of Table 4. These results however need to be analysed with caution as they tend not to be significant to a full specification of fixed effects and better proxies may be warranted.[29] Focusing R&D intensity, for example, we find that the interaction term between plant-level labour productivity and R&D intensity is not statistically significant in most specifications albeit positive for the eurozone. Further disentangling the role of private and public research may be an interesting venue for future research.

6 Agglomeration and Growth

After having investigated the role of plant performance and regional policies for agglomeration, we now briefly investigate the relationship between regional agglomeration and growth. To this end, we average our firm-level agglomeration index at 50 km at the industry-region level (either taking simple averages or sales-weighted averages). We also compute regional GDP growth rates between 2005 and 2017 for each European NUTS2 region and each U.S. province and run the following regression at the region level:

$$growth_r \equiv \beta_1 \text{Density}_{k,r} + \gamma' X_r + \delta_k + \varepsilon_i \quad (9)$$

We add a set of regional controls measured in the initial year 2004: the level of GDP per capita, population density, the fraction of the population with more than

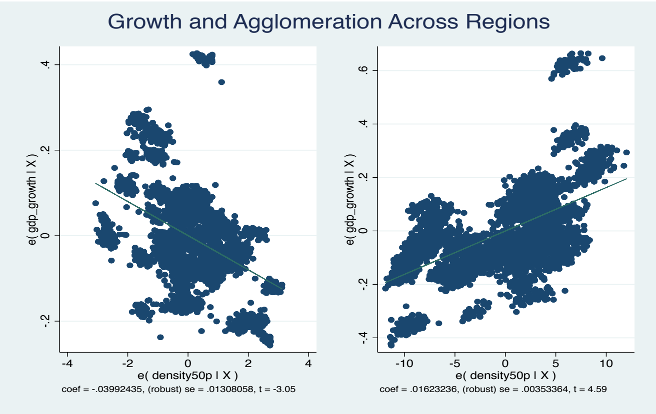
[28] When analysing the relation with the distance to final goods, firms with higher upstreamness tend to agglomerate relatively more, while more productive firms tend to agglomerate relatively less the higher the upstreamness.

[29] Comparing industries at 2 digits, firms tend to agglomeration around manufacturing sectors more linked to natural resources sectors (e.g. wood/furniture in Europe, tobacco / textiles in US.) with somewhat stronger effects in the US. There are strong agglomeration around capital intensive industries (chemicals, rubber, metals and machinery, etc.) and transportation in particular in the Eurozone regions.

secondary education and R&D expenditure. As we showed before, these regional controls are correlated with agglomeration and are potentially also drivers of GDP growth. Since the dependent variable varies only at the region level, while the explanatory variable of interest varies at the industry-region level, we cluster standard errors at the region level. Since we only have cross-section data on agglomeration available, omitted variables are of course a concern and one should be careful not to interpret these relationships as causal.

[TABLE 5 HERE]

Figure 3: Agglomeration and Growth



Results are presented in Table 6 and Figure 3. Some interesting patterns emerge. In the U.S., regional GDP growth correlates negatively with agglomeration, while the same relationship is positive in Europe. In the U.S., a one-standard-deviation increase in agglomeration leads to a 6-percentage-point reduction in total regional growth over the 12-year period. The association between growth and agglomeration is particularly strong within the Eurozone, where a one-standard-deviation increase in agglomeration is associated with a 6-percentage-point increase in growth. Outside of the Eurozone, the effect of a one-standard deviation increase in agglomeration is only around half as large (3 percentage points). Thus, higher levels of agglomeration in manufacturing are associated with faster regional GDP growth in Europe but not in the U.S. This may, of course, reflect the fact that in the U.S. regions specializing in manufacturing were overall in decline during the 2000s, while some of the most strongly performing regions in core Europe were heavily specialized in manufacturing.

7 Conclusion

Our analysis shows that firms, including multinationals, are far from equal within each industrial cluster. Some firms are significantly more centred than others. Different groups of factors are expected to explain the heterogeneous level of agglomeration across firms. First, firm characteristics matter. Larger and more productive establishments are centred with more agglomeration than their smaller, less productive counterparts, in particular in the U.S. and in the Eurozone.

Agglomeration around superstar plants is most pronounced in the Eurozone. In the non-Eurozone European countries, superstar plants actually attract less agglomeration. In the U.S. and in the Eurozone, MNC establishments also attract significantly more agglomeration than domestically owned plants, while this is not the case in Europe outside of the Eurozone.

Overall, in US and Eurozone there is more agglomeration around high performance plants (productivity, multinationals), reflecting greater potential spillovers from leading firms due perhaps to the more skill- and capital-intensive activities engaged by these firms in the regions. In contrast, the economic activities of superstar firms are more likely to be unskilled-labour intensive outside of the Eurozone, limiting the scope of agglomeration economies.

Region attributes also play an important role. Regions with different levels of economic development including education and technology exhibiting distinct agglomeration patterns. Specifically, we find that location characteristics such as human capital levels and R&D spending could sometimes weaken the incentive to agglomerate around large firms. Higher regional human capital levels are associated with less agglomeration around larger and more productive plants, in particular in the U.S. and in the Eurozone. Similarly, larger regional R&D spending is associated with less agglomeration of economic activity around these plants in Europe. Regarding the concentration of economic activity around affiliates of multinationals, we uncover that both in the U.S. and in Europe outside of the Eurozone larger regional R&D investment is associated with more agglomeration around these plants. This is not true in the Eurozone, where this association is negative. Similarly, outside of the Eurozone and in the U.S. larger regional human capital levels are associated with more agglomeration around multinationals.

In terms of policies, R&D leads to more agglomeration in manufacturing in Europe, while more human capital leads to less agglomeration, probably by inducing specialization outside of the manufacturing sector. However, R&D spending does not induce agglomeration around high productivity plants. Data measurement is of course a concern and better proxies may help solve this puzzle. However, exploring and disentangling the role of public and private spending may shed further light into this question.

We present suggestive evidence of a positive relation between agglomeration and growth in Europe, but not in the U.S. Of course, we caution in terms of causality implications. We conjecture that the negative effect in U.S. may be driven by a decline in regions that are specializing in manufacturing and display high levels of agglomeration in these industries. In the Eurozone, the relationship between growth and agglomeration in manufacturing was positive: core regions (e.g. Bavaria) that are heavily specialized and agglomerated in manufacturing fared relatively well..

Overall, although, as the movement of goods, people and ideas have become easier through economic integration, we observe continuing dominance of certain firms, industrial clusters and cities despite reductions in transportation and communication costs.

8 References

Ahlfeldt, G. and Wendland, N (2012), "How Polycentric is a Monocentric City? Centers, Spillovers and Hysteresis", Journal of Economic Geography, Vol. 13, No 1, pp. 53-83.

Alfaro, L. and Chen, M.X. (2018), "Selection and Market Reallocation: Productivity Gains from Multinational Production", American Economic Journal: Economic Policy, Vol. 10, No 2, pp. 1–38.

Alfaro, L., Charlton, A. and Kanczuk, F (2009), "Plant-Size Distribution and Cross-Country Income Differences", NBER International Seminar on Macroeconomics, No 14060.

Alfaro, L. and Chen, M.X. (2019), "Location Fundamentals, Agglomeration Economies, and the Geography of Multinational Firms", The Oxford Handbook of Structural Transformation, Oxford University Press.

Alfaro, L. and Chen, M.X. (2018), "Selection and Market Reallocation: Productivity Gains from Multinational Production", American Economic Journal: Economic Policy, Vol. 10, No 2, pp. 1–38.

Alfaro, L. and Chen, M.X. (2014), "The Global Agglomeration of Multinational Firms", Journal of International Economics, Vol. 94, No 2, pp. 263–276.

Alfaro, L. and Chen, M.X. (2010), "Multinational Firms, Agglomeration, and Global Networks", Vox.org.

Alfaro, L., Conconi, P., Fadinger, H. and Newman, A. (2016), "Do Prices Determine Vertical Integration?", Review of Economic Studies, Vol. 83, No 3, pp. 855–888

Amiti, M. and Javorcik, B. (2008), "Trade costs and location of foreign firms in China", Journal of Development Economics, Vol. 85, No 1-2, pp. 129-149.

Andersson, F., Burgess, S. and Lane, J. (2007), "Cities, matching and the productivity gains of agglomeration", Journal of Urban Economics, Vol. 61, No 1, pp. 112-128.

Antràs, P. and Helpman, E. (2004), "Global sourcing", Journal of Political Economy, Vol. 112, No 3, pp. 552-580.

Antràs, P. and Helpman, E. (2008), “Contractual Frictions and Global Sourcing”, The Organization of Firms in a Global Economy, pp. 9-54.

Autor, D., Dorn, D., Katz, L, Patterson, C. and Van Reenen, J. (2017), "The Fall of the Labor Share and the Rise of Superstar Firms”, NBER Working Paper, No 23396.

Bacolod, M., Blum, B.S. and Strange, W.C. (2009), “Skills in the city”, Journal of Urban Economics, Vol. 65, No 2, pp. 136-153.

Blonigen, B., Ellis, C. and Fausten, D. (2005), "Industrial groupings and foreign direct investment", Journal of International Economics, Vol. 65, No 1, pp. 75-91.

Bloom, N., Draca, M. and Reenen, J.V. (2016), "Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity", The Review of Economic Studies, Vol. 83, No 1, pp. 87-117.

Bobonis, G.J. and Shatz, H.J. (2007), "Agglomeration, Adjustment, and State Policies in the Location of Foreign Direct Investment in the United States", The Review of Economics and Statistics, Vol. 89, No 1, pp. 30-43.

Bureau of Economic Analysis (2002), “Benchmark Input-Output Accounts of the U.S. Economy, 2002”.

Buzard, K., Carlino, G.A., Hunt, R.M., Carr, J.K. and Smith, T.E. (2017), “The Agglomeration of American R&D Labs”, Journal of Urban Economics, Vol. 101, pp. 14-26.

Chen, M.X., and Moore, M. (2010), "Location decision of heterogeneous multinational firms", Journal of International Economics, Vol. 80, No 2, pp. 188-199.

Ciccone, A. (2002), “Agglomeration effects in Europe”, European Economic Review, Vol. 46, No 2, pp. 213-227.

Combes, P.P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (2012), “The productivity advantages of large cities: Distinguishing agglomeration from firm selection”, Econometrica, Vol. 80, No 6, pp. 2543-2594.

Crozet, M., Mayer, T. and Mucchielli, J.L. (2004), "How do firms agglomerate? A study of FDI in France", Regional Science and Urban Economics, Vol. 34, No 1, pp. 27-54.

Debaere, P., Lee, J. and Paik, M. (2010), "Agglomeration, backward and forward linkages: Evidence from South Korean investment in China", Canadian Journal of Economics, Vol. 43, No 2, pp. 520-546.

Duranton, G. and Overman, H. (2005), "Testing for localization using micro-geographic data", The Review of Economic Studies, Vol. 72, No 4, pp. 1077-1106.

Duranton, G. and Overman, H. (2008), "Exploring the detailed location patterns of UK manufacturing industries using microgeographic data", Journal of Regional Science, Vol. 48, No 1, pp. 213-243.

Duranton, G. and Puga, D. (2004), "Micro-foundations of urban agglomeration economies", Handbook of Regional and Urban Economics, Vol. 4, pp. 2063-2117.

Ellison, G. and Glaeser, E.L. (1997), "Geographic concentration in US manufacturing industries: a dartboard approach", Journal of Political Economy, Vol. 105, No 5, pp. 889-927.

Ellison, G. and Glaeser, E.L. (1999), "The geographic concentration of industry: does natural advantage explain agglomeration?", American Economic Review, Vol. 89, No 2, pp. 311-316.

Ellison, G., Glaeser, E.L. and Kerr, W. (2010), " What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", American Economic Review, Vol. 100, No 3, pp. 1195-1213.

Forman, C., Goldfarb, A. and Greenstein, S. (2016), “Agglomeration of invention in the Bay Area: Not just ICT”, American Economic Review, Vol. 106, No 5, pp. 146-51.

Fritsch, M. and Changoluisa, J. (2017), “New business formation and the productivity of manufacturing incumbents: Effects and mechanisms”, Journal of Business Venturing, Vol. 32, No 3, pp. 237-259.

Glaeser, E. (2010), Agglomeration economics, University of Chicago Press.

Glaeser, E.L. and Gottlieb, J.D. (2009), "The wealth of cities: Agglomeration economies and spatial equilibrium in the United States", Journal of Economic Literature, Vol. 47, No 4, pp. 983-1028.

Gordon, I.R. and McCann, P. (2005), “Innovation, agglomeration, and regional development”, Journal of Economic Geography, Vol. 5, No 5, pp. 523-543.

Greenstone, M., Hornbeck, R. and Moretti, E. (2010), “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings”, Journal of Political Economy, Vol. 118, No 3, pp. 536-598.

Gutiérrez, G. and Philippon, T. (2017), “Declining Competition and Investment in the US”, NBER, No 23583.

Hanson, Gordon H. (2004), "Market potential, increasing returns and geographic concentration", Journal of International Economics, Vol. 67, No 1, pp. 1-24.

Head, K. and Mayer, T. (2004), "Market potential and the location of Japanese investment in the European Union", Review of Economics and Statistics, Vol. 86, No 4, pp. 959-972.

Head, K. and Mayer, T. (2004), "The Empirics of Agglomeration and Trade”, Handbook of Regional and Urban Economics, Vol. 4, pp. 2609-2669.

Head, K., Ries, J. and Swenson, D. (1995), "Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States", Journal of International Economics, Vol. 38, No 3-4, pp. 223-247.

Helmers, C. and Overman, H.G. (2017), “My precious! The location and diffusion of scientific research: evidence from the Synchrotron Diamond Light Source”, The Economic Journal, Vol. 127, No 604, pp. 2006-2040.

Helpman, E., Melitz, M.J. and Yeaple, S.R. (2004), "Export versus FDI with Heterogeneous Firms", American Economic Review, Vol. 94, No 1, pp. 300-316.

Kolko, Jed (2000), "The death of cities? The death of distance? Evidence from the geography of commercial Internet usage", The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy, The MIT Press, pp. 73-98.

Krugman, P. (1991), "Increasing Returns and Economic Geography", Journal of Political Economy, Vol. 99, No 3, pp. 483-499.

Markusen, J. (2002), Multinational Firms and the Theory of International Trade, The MIT Press.

Mayer, T. and Ottaviano, G. (2007), “The Happy Few: The Internationalisation of European Firms. New Facts Based on Firm-Level Evidence”, Bruegel blueprint series, Vol. 3.

Marshall, A. (1890), Principles of Economics, Macmillan and Co.

Navaretti, G.B. and Venables, A.J. (2004), "Host country effects: conceptual framework and the evidence", Multinational Firms in the World Economy, Princeton University Press, pp. 151-182.

Ottaviano, G.M., and Puga, D. (1998), "Agglomeration in the Global Economy: A Survey of the ‘New Economic Geography’", World Economy, Vol. 21, No 6, pp. 707-731.

Ottaviano, G.M. and Thisse, J.F. (2004), "Agglomeration and Economic Geography”, Handbook of Regional and Urban Economics, Vol. 4, pp. 2563-2608.

Puga, D. (2010), "The Magnitude and Causes of Agglomeration Economies", Journal of Regional Science, Vol. 50, No 1, pp. 203-219.

Redding, S. (2011), “Economic Geography: a Review of the Theoretical and Empirical Literature”, The Palgrave Handbook of International Trade, Macmillan Publishers Ltd.

Redding, S.J. (2010), "The Empirics of New Economic Geography", Journal of Regional Science, Vol. 50, No 1, pp. 297-311.

Redding, S.J. (2011), "Theories of heterogeneous firms and trade", Annual Review of Economics, Vol. 3, No 1, pp. 77-105.

Redding, S.J. and Venables, A.J. (2004), "Economic geography and international inequality", Journal of International Economics, Vol. 62, No 1, pp. 53-82.

Rigby, D.L. and Essletzbichler, J. (2002), "Agglomeration economies and productivity differences in US cities", Journal of Economic Geography, Vol. 2, No 4, pp. 407-432.

Rosen, S. (1981), "The Economics of Superstars", The American Economic Review, Vol. 71, No 5, pp. 845-858.

Rosenthal, S. and Strange, W.C. (2004), "Evidence on the nature and sources of agglomeration economies", Handbook of Regional and Urban Economics, Vol. 4, pp. 2119-2171.

Rosenthal, S. and Strange, W.C. (2003), "Geography, industrial organization, and agglomeration", Review of Economics and Statistics, Vol. 85, No 2, pp. 377-393.

Rosenthal, S. and Strange, W.C. (2001), "The Determinants of Agglomeration", Journal of Urban Economics, Vol. 50, No 2, pp. 191-229.

Table 1: Agglomeration and Firm Performance (Labor Productivity)

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)
ln(PrdL)	0.09408***	0.06805***	0.01859***	0.01859*	-0.04290***	-0.03772***	0.00969*	0.010
	(0.007)	(0.007)	(0.005)	(0.009)	(0.005)	(0.006)	(0.006)	(0.037)
ln(Age)	0.21068***	0.22304***	0.13650***	0.13650***	-0.53899***	-0.53674***	-0.29663***	-0.29663**
	(0.009)	(0.009)	(0.007)	(0.032)	(0.026)	(0.026)	(0.025)	(0.144)
Multi-product	0.04066***	0.04027***	0.05196***	0.05196***	0.81172***	0.80459***	0.55945***	0.55945***
	(0.007)	(0.007)	(0.005)	(0.012)	(0.017)	(0.017)	(0.017)	(0.143)
MNC		0.38181***	0.22952***	0.22952***		0.22456***	0.47727***	0.47727*
		(0.027)	(0.021)	(0.079)		(0.055)	(0.050)	(0.285)
ln(gdp)			-0.18615***	-0.1862			-4.33669***	-4.33669***
			(0.005)	(0.177)			(0.057)	(1.235)
ln(pop. density)			1.41340***	1.41340***			0.51240***	0.512
			(0.008)	(0.216)			(0.021)	(0.451)
ln(post sec.)			-2.57529***	-2.57529*			-1.83313***	(1.833)
			(0.036)	(1.422)			(0.048)	(1.121)
ln(R&D)			-0.11513***	-0.1151			1.57731***	1.57731***
			(0.005)	(0.269)			(0.023)	(0.557)
Observations	61,576	61,576	61,576	61,576	54,242	54,242	54,242	54,242
R-squared	0.053	0.056	0.417	0.417	0.156	0.156	0.298	0.298
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	Yes	Yes	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Cluster
Sample	US	US	US	US	Europe	Europe	Europe	Europe

	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
ln(PrdL)	0.23861***	0.28060***	0.11738***	0.11738**	0.005	-0.02309***	0.12514***	0.12514***
	(0.015)	(0.014)	(0.011)	(0.049)	(0.009)	(0.008)	(0.007)	(0.041)
ln(Age)	0.13308***	0.13227***	-0.21414***	(0.214)	-1.89508***	-1.92338***	-0.53795***	-0.53795***
	(0.035)	(0.035)	(0.031)	(0.141)	(0.040)	(0.040)	(0.029)	(0.171)
Multi-product	0.72191***	0.71949***	0.67006***	0.67006***	0.68508***	0.81183***	0.07631***	0.076
	(0.025)	(0.024)	(0.023)	(0.224)	(0.029)	(0.030)	(0.023)	(0.078)
MNC		1.93294***	1.92717***	1.92717***		-1.57871***	-0.74273***	-0.74273***
		(0.084)	(0.085)	(0.446)		(0.076)	(0.046)	(0.222)
ln(gdp)			0.25020*	0.250			-2.56060***	-2.56060***
			(0.150)	(2.692)			(0.058)	(0.670)
ln(pop. density)			0.15905***	0.159			0.87311***	0.87311*
			(0.039)	(0.772)			(0.024)	(0.465)
ln(post sec.)			-0.96999***	-0.9700			-4.77043***	-4.77043***
			(0.063)	(1.191)			(0.082)	(1.033)
ln(R&D)			1.59231***	1.59231**			0.06695***	0.067
			(0.031)	(0.639)			(0.024)	(0.477)
Observations	32,437	32,437	32,437	32,437	17,883	17,883	17,883	17,883
R-squared	0.16	0.173	0.336	0.336	0.305	0.325	0.72	0.72
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	Yes	Yes	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Cluster
Sample	Euro	Euro	Euro	Euro	Non Euro	Non Euro	Non Euro	Non Euro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 2: Agglomeration and Firm Performance (Super Productive Firms)

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)
ln(PrdL)	0.07461***	0.05279***	0.0052	0.0052	-0.05006***	-0.04506***	-0.0006	-0.0006
	(0.010)	(0.010)	(0.007)	(0.015)	(0.006)	(0.006)	(0.006)	(0.037)
Suprlnprdl_Sample	0.12969***	0.10293**	0.09027***	0.090	0.45759***	0.39269***	0.56643***	0.56643**
	(0.044)	(0.044)	(0.034)	(0.073)	(0.077)	(0.079)	(0.077)	(0.243)
ln(Age)	0.21074***	0.22299***	0.13645***	0.13645***	-0.53999***	-0.53811***	-0.29617***	-0.29617**
	(0.009)	(0.009)	(0.007)	(0.032)	(0.026)	(0.026)	(0.025)	(0.144)
Multi-product	0.04052***	0.04016***	0.05187***	0.05187***	0.81384***	0.80805***	0.56398***	0.56398***
	(0.007)	(0.007)	(0.005)	(0.012)	(0.017)	(0.017)	(0.017)	(0.143)
MNC		0.37898***	0.22704***	0.22704***		0.17284***	0.40501***	0.405
		(0.027)	(0.021)	(0.078)		(0.056)	(0.051)	(0.269)
ln(gdp)			-0.18596***	-0.1860			-4.34216***	-4.34216***
			(0.005)	(0.177)			(0.058)	(1.235)
ln(pop. density)			1.41325***	1.41325***			0.50897***	0.509
			(0.008)	(0.216)			(0.021)	(0.452)
ln (post sec.)			-2.57358***	-2.57358*			-1.84583***	-1.8458
			(0.036)	(1.422)			(0.048)	(1.121)
ln(R&D)			-0.11498***	-0.1150			1.57691***	1.57691***
			(0.005)	(0.269)			(0.022)	(0.558)
Observations	61,576	61,576	61,576	61,576	54,242	54,242	54,242	54,242
R-squared	0.054	0.056	0.417	0.417	0.156	0.156	0.299	0.299
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	Yes	Yes	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Cluster
Sample	US	US	US	US	Europe	Europe	Europe	Europe
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(PrdL)	0.20723***	0.26334***	0.10724***	0.10724**	0.03640***	0.003	0.13191***	0.13191***
	(0.015)	(0.014)	(0.011)	(0.051)	(0.009)	(0.008)	(0.007)	(0.043)
Suprlnprdl_Sample	1.22054***	0.59265***	0.34758***	0.348	-2.33584***	-1.69249***	-0.46442***	-0.46442**
	(0.095)	(0.099)	(0.102)	(0.321)	(0.136)	(0.139)	(0.090)	(0.193)
ln(Age)	0.13467***	0.13308***	-0.21278***	-0.2128	-1.87027***	-1.90150***	-0.53622***	-0.53622***
	(0.035)	(0.035)	(0.031)	(0.141)	(0.040)	(0.040)	(0.029)	(0.170)
Multi-product	0.72482***	0.72102***	0.67072***	0.67072***	0.68213***	0.79224***	0.07347***	0.073
	(0.025)	(0.024)	(0.023)	(0.225)	(0.029)	(0.030)	(0.023)	(0.077)
MNC		1.83943***	1.87607***	1.87607***		-1.36135***	-0.68520***	-0.68520***
		(0.086)	(0.087)	(0.420)		(0.078)	(0.047)	(0.214)
ln(gdp)			0.242	0.242			-2.55623***	-2.55623***
			(0.150)	(2.692)			(0.058)	(0.667)
ln(pop. density)			0.15396***	0.154			0.87007***	0.87007*
			(0.039)	(0.772)			(0.024)	(0.464)
ln (post sec.)			-0.98584***	-0.9858			-4.76200***	-4.76200***
			(0.063)	(1.195)			(0.081)	(1.031)
ln(R&D)			1.59493***	1.59493**			0.07109***	0.071
			(0.031)	(0.640)			(0.024)	(0.477)
Observations	32,437	32,437	32,437	32,437	17,883	17,883	17,883	17,883
R-squared	0.162	0.173	0.336	0.336	0.317	0.33	0.721	0.721
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	Yes	Yes	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Cluster
Sample	Euro	Euro	Euro	Euro	Non Euro	Non Euro	Non Euro	Non Euro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 3a: Agglomeration and Regional Policies

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	-0.22604***	0.04856**	0.17842***	0.11600***	-0.17901***	(0.179)	(0.134)	-0.55783***	0.24367***	-0.33729***	0.93699***	0.937
	(0.045)	(0.021)	(0.053)	(0.018)	(0.054)	(0.168)	(0.083)	(0.032)	(0.020)	(0.023)	(0.137)	(1.199)
ln(Age)	0.22036***	0.13902***	0.21892***	0.21818***	0.13654***	0.13654***	-0.24398***	-0.58876***	-0.42265***	-0.63873***	-0.29307***	-0.29307**
	(0.009)	(0.007)	(0.009)	(0.009)	(0.007)	(0.032)	(0.027)	(0.025)	(0.026)	(0.026)	(0.025)	(0.144)
Multi-product	0.04034***	0.06207***	0.04428***	0.04635***	0.05222***	0.05222***	0.75656***	0.78355***	0.78263***	0.81770***	0.57072***	0.57072***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.005)	(0.012)	(0.018)	(0.017)	(0.018)	(0.018)	(0.017)	(0.141)
MNC	0.38411***	0.24043***	0.37019***	0.38113***	0.22938***	0.22938***	0.18345***	0.32095***	0.59930***	0.26817***	0.51650***	0.51650*
	(0.027)	(0.022)	(0.027)	(0.026)	(0.021)	(0.079)	(0.052)	(0.054)	(0.054)	(0.056)	(0.050)	(0.294)
ln(gdp)	-0.22715***				-0.12009**	-0.1201	-1.67588***				-3.38028***	-3.38028*
	(0.050)				(0.057)	(0.156)	(0.087)				(0.168)	(1.794)
ln(pop. density)		1.22439***			1.21339***	1.21339***		-0.52560***			0.31767***	0.318
		(0.055)			(0.089)	(0.307)		(0.064)			(0.065)	(0.775)
ln(post sec.)			0.091		-1.17978***	-1.1798			-3.36852***		-3.35799***	(3.358)
			(0.361)		(0.393)	(1.270)			(0.156)		(0.192)	(2.186)
ln(R&D)				0.41654***	-0.0258	-0.0258				-0.21002***	1.49053***	1.49053**
				(0.063)	(0.055)	(0.232)				(0.035)	(0.058)	(0.656)
ln(PrdL)×ln(gdp)	0.02779***				-0.0058	-0.0058	0.01531*				-0.08734***	-0.0873
	(0.004)				(0.005)	(0.011)	(0.008)				(0.015)	(0.118)
ln(PrdL)×ln(pop.den)		-0.00953**			0.01760**	0.0176		0.09611***			0.01883***	0.0188
		(0.005)			(0.008)	(0.028)		(0.005)			(0.005)	(0.042)
ln(PrdL)×ln(post sec.)			0.06824**		-0.12236***	-0.1224			0.17089***		0.13640***	0.1364
			(0.031)		(0.034)	(0.101)			(0.013)		(0.016)	(0.155)
ln(PrdL)×R&D				-0.01728***	-0.0078	-0.0078				0.04400***	0.00886*	0.0089
				(0.005)	(0.005)	(0.013)				(0.003)	(0.005)	(0.035)
Observations	61,576	61,576	61,576	61,576	61,576	61,576	54,242	54,242	54,242	54,242	54,242	54,242
R-squared	0.057	0.368	0.064	0.078	0.417	0.417	0.184	0.172	0.175	0.164	0.3	0.3
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Robust	Robust	Cluster
Sample	US	US	US	US	US	US	Europe	Europe	Europe	Europe	Europe	Europe

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 3b: Agglomeration and Regional Policies, cont.

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	-3.97285***	0.58685***	-0.81569***	0.72606***	-7.15737***	-7.15737**	2.43264***	-0.28359***	-0.36415***	0.42759***	2.50078***	2.50078***
	(0.380)	(0.060)	(0.056)	(0.085)	(0.685)	(2.789)	(0.063)	(0.042)	(0.020)	(0.024)	(0.105)	(0.704)
ln(Age)	-0.22000***	0.032	0.09380***	-0.20200***	-0.22113***	(0.221)	-0.59002***	-1.85288***	-0.81568***	-1.19501***	-0.47756***	-0.47756***
	(0.033)	(0.034)	(0.035)	(0.031)	(0.031)	(0.146)	(0.030)	(0.041)	(0.032)	(0.038)	(0.026)	(0.164)
Multi-product	0.68826***	0.62182***	0.69274***	0.71685***	0.62133***	0.62133***	-0.26969***	0.77328***	0.44203***	0.30345***	(0.014)	(0.014)
	(0.023)	(0.024)	(0.025)	(0.022)	(0.023)	(0.209)	(0.025)	(0.030)	(0.024)	(0.028)	(0.021)	(0.064)
MNC	1.94785***	1.81344***	1.68648***	1.70560***	1.91559***	1.91559***	-1.40310***	-1.54855***	-0.84425***	-1.43264***	-0.82095***	-0.82095***
	(0.089)	(0.082)	(0.087)	(0.082)	(0.086)	(0.453)	(0.052)	(0.076)	(0.057)	(0.067)	(0.046)	(0.198)
ln(gdp)	(0.376)				-8.42119***	-8.42119*	-2.02373***				-0.46120***	(0.461)
	(0.427)				(0.873)	(4.793)	(0.070)				(0.111)	(0.811)
ln(pop. density)		1.87286***			0.45019**	0.450		-0.77910***			0.86340***	0.863
		(0.120)			(0.223)	(1.385)		(0.072)			(0.066)	(0.766)
ln(post sec.)			8.94962***		3.74739***	3.747			-3.84030***		-4.71355***	-4.71355***
			(0.504)		(0.454)	(2.757)			(0.130)		(0.189)	(0.948)
ln(R&D)				2.59340***	3.70569***	3.70569***				-0.74015***	-0.18509***	-0.1851
				(0.137)	(0.249)	(1.193)				(0.041)	(0.070)	(0.629)
ln(PrdL)×ln(gdp)	0.40217***				0.78241***	0.78241**	-0.24260***				-0.25349***	-0.25349***
	(0.036)				(0.075)	(0.299)	(0.006)				(0.010)	(0.067)
ln(PrdL)×ln(pop. den)		-0.05567***			-0.0229	-0.0229		0.04698***			-0.0035	-0.0035
		(0.010)			(0.019)	(0.086)		(0.007)			(0.006)	(0.038)
ln(PrdL)×ln(post sec.)			-0.70540***		-0.40559***	-0.40559**			-0.26223***		0.03694**	0.0369
			(0.042)		(0.037)	(0.178)			(0.012)		(0.019)	(0.084)
ln(PrdL)×R&D				-0.08806***	-0.19138***	-0.19138**				-0.07034***	0.02461***	0.0246
				(0.012)	(0.021)	(0.073)				(0.004)	(0.007)	(0.033)
Observations	32,437	32,437	32,437	32,437	32,437	32,437	17,883	17,883	17,883	17,883	17,883	17,883
R-squared	0.256	0.212	0.196	0.333	0.355	0.355	0.678	0.335	0.638	0.496	0.754	0.754
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Errors	Robust	Robust	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Robust	Robust	Cluster
Sample	Euro	Euro	Euro	Euro	Euro	Euro	NonEuro	NonEuro	NonEuro	NonEuro	NonEuro	NonEuro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 4a: Agglomeration and Sector Characteristics

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	0.06169***	0.02046***	0.16328***	0.12104***	0.12104**	0.04808	-0.09401***	-0.03181***	0.04525	0.12673***	0.12673	0.23696**
	(0.0070)	(0.0050)	(0.0250)	(0.0200)	(0.0480)	(0.0340)	(0.0060)	(0.0060)	(0.0340)	(0.0360)	(0.1120)	(0.0950)
ln(Age)	0.24847***	0.15381***	0.24917***	0.15407***	0.15407***	0.13499***	-0.61323***	-0.33897***	-0.61682***	-0.34182***	-0.34182**	-0.28708**
	(0.0090)	(0.0070)	(0.0090)	(0.0070)	(0.0380)	(0.0320)	(0.0270)	(0.0260)	(0.0270)	(0.0260)	(0.1650)	(0.1410)
Multi-product	0.03157***	0.04276***	0.03268***	0.04360***	0.04360***	0.04949***	0.78835***	0.52336***	0.78681***	0.52084***	0.52084***	0.56546***
	(0.0070)	(0.0060)	(0.0070)	(0.0060)	(0.0120)	(0.0120)	(0.0180)	(0.0170)	(0.0180)	(0.0170)	(0.1640)	(0.1430)
MNC	0.47598***	0.29357***	0.47496***	0.29539***	0.29539***	0.22216***	0.46814***	0.60019***	0.47387***	0.60421***	0.60421*	0.49067*
	(0.0260)	(0.0210)	(0.0260)	(0.0210)	(0.1060)	(0.0800)	(0.0510)	(0.0470)	(0.0510)	(0.0470)	(0.3440)	(0.2860)
R&Dint	-0.20587***	-0.09912***	-0.1670	0.0694	0.0694		-0.0237	-0.0657	-0.62161**	-0.4914	-0.4914	
	(0.0270)	(0.0230)	(0.2600)	(0.2350)	(0.3820)		(0.1160)	(0.1180)	(0.2870)	(0.3860)	(0.3950)	
Lint	-0.37049**	-0.36239***	2.0111	3.42879***	3.4288		2.39316***	2.64474***	7.69100***	10.70248***	10.70248***	
	(0.1500)	(0.1200)	(1.3860)	(1.0760)	(2.2090)		(0.4960)	(0.4500)	(1.3960)	(1.4240)	(4.0890)	
Kint	0.2791	0.67959***	14.57098***	11.40251***	11.40251**		11.59646***	10.69613***	14.87019***	10.34389***	10.34389**	
	(0.2920)	(0.2290)	(2.5510)	(2.0130)	(5.5610)		(1.0020)	(0.9200)	(4.8900)	(5.0760)	(10.0660)	
ln(gdp)		-0.21023***		-0.20998***	-0.2100	-0.1952		-5.05563***		-5.06883***	-5.06883***	-4.44447***
		(0.0050)		(0.0050)	(0.1850)	(0.1800)		(0.0620)		(0.0620)	(1.5260)	(1.2410)
ln(pop. density)		1.44220***		1.44169***	1.44169***	1.42943***		0.43643***		0.43415***	0.4342	0.4944
		(0.0080)		(0.0080)	(0.2260)	(0.2210)		(0.0220)		(0.0220)	(0.4910)	(0.4430)
ln(post sec.)		-2.76822***		-2.76638***	-2.76638*	-2.65414*		-1.51318***		-1.50598***	(1.5060)	-1.86513*
		(0.0380)		(0.0380)	(1.5010)	(1.4510)		(0.0490)		(0.0490)	(1.3240)	(1.1180)
ln(R&D)		-0.12760***		-0.12697***	-0.1270	-0.1226		1.81931***		1.81969***	1.81969***	1.62027***
		(0.0060)		(0.0060)	(0.2780)	(0.2710)		(0.0240)		(0.0240)	(0.6750)	(0.5550)
lnprdl×R&Dint			-0.0031	-0.0138	-0.0138	-0.0268			0.05190**	0.0366	0.0366	0.0252
			(0.0210)	(0.0190)	(0.0290)	(0.0260)			(0.0230)	(0.0320)	(0.0320)	(0.0270)
lnprdl×Lint			-0.20380*	-0.32508***	-0.32508*	-0.087			-0.48873***	-0.74505***	-0.74505**	-1.02405***
			(0.1180)	(0.0920)	(0.1780)	(0.1230)			(0.1150)	(0.1190)	(0.3450)	(0.3310)
lnprdl×Kint			-1.20786***	-0.90852***	-0.90852**	-0.27663			-1.20095***	-0.88270**	-0.8827	-1.31014**
			(0.2140)	(0.1690)	(0.4230)	(0.3290)			(0.4120)	(0.4330)	(0.8440)	(0.6520)
Observations	56,370	56,370	56,370	56,370	56,370	56,370	51,522	51,522	51,522	51,522	51,522	51,522
R-squared	0.023	0.398	0.023	0.398	0.398	0.415	0.046	0.223	0.046	0.224	0.224	0.306
FE	No	No	No	No	No	Industry	No	No	No	No	No	Industry
Errors	Robust	Robust	Robust	Robust	Cluster	Cluster	Robust	Robust	Robust	Robust	Cluster	Cluster
Sample	US	US	US	US	US	US	europe	europe	europe	europe	europe	europe

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 4b: Agglomeration and Sector Characteristics, cont.

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	0.34823***	0.13447***	0.90914***	0.51419***	0.51419***	0.54385***	-0.11066***	0.09966***	-0.16383***	0.19398***	0.19398*	0.26241**
	(0.0150)	(0.0110)	(0.0810)	(0.0590)	(0.1320)	(0.1330)	(0.0090)	(0.0070)	(0.0430)	(0.0380)	(0.1050)	(0.1010)
ln(Age)	0.16761***	-0.22430***	0.16499***	-0.22350***	(0.2235)	(0.2039)	-2.23574***	-0.56776***	-2.23230***	-0.57026***	-0.57026***	-0.53942***
	(0.0370)	(0.0320)	(0.0370)	(0.0320)	(0.1530)	(0.1420)	(0.0410)	(0.0300)	(0.0410)	(0.0290)	(0.1970)	(0.1710)
Multi-product	0.81203***	0.74049***	0.80588***	0.73702***	0.73702***	0.67610***	0.61136***	(0.0346)	0.61181***	-0.04117*	(0.0412)	0.0770
	(0.0250)	(0.0230)	(0.0250)	(0.0230)	(0.2620)	(0.2250)	(0.0300)	(0.0230)	(0.0300)	(0.0230)	(0.0890)	(0.0760)
MNC	2.38363***	2.11212***	2.37573***	2.10646***	2.10646***	1.91097***	-2.08544***	-0.85866***	-2.08323***	-0.85096***	-0.85096***	-0.75069***
	(0.0780)	(0.0820)	(0.0790)	(0.0820)	(0.4920)	(0.4450)	(0.0700)	(0.0430)	(0.0700)	(0.0420)	(0.2850)	(0.2340)
R&Dint	0.0466	(0.1373)	2.24314***	0.7884	0.7884		-0.33867**	0.0057	-0.79840***	0.1445	0.1445	
	(0.1490)	(0.1470)	(0.6930)	(0.5940)	(0.9970)		(0.1590)	(0.1130)	(0.2440)	(0.4420)	(0.3630)	
Lint	1.23301*	-2.29550***	15.40688***	9.08454***	9.08454**		4.30001***	5.43032***	5.49271***	13.64536***	13.64536***	
	(0.6550)	(0.5650)	(3.7480)	(2.6340)	(4.1810)		(0.7060)	(0.4320)	(1.9640)	(1.6260)	(4.3380)	
Kint	5.00202***	3.91867***	8.13496***	5.56671***	5.56671***		8.62763***	2.44183***	-2.495	-2.752	-2.752	
	(1.2040)	(1.0210)	(10.6280)	(7.9130)	(13.3770)		(1.4510)	(0.8930)	(4.0630)	(3.9610)	(4.8960)	
ln(gdp)		0.0835		0.0695	0.0695	0.1283		-2.59220***		-2.61697***	-2.61697***	-2.62365***
		(0.1640)		(0.1640)	(2.9150)	(2.7110)		(0.0600)		(0.0580)	(0.7030)	(0.6570)
ln(pop. density)		0.0380		0.0408	0.0408	0.1728		0.85203***		0.85170***	0.85170*	0.83870*
		(0.0400)		(0.0400)	(0.8040)	(0.7640)		(0.0260)		(0.0260)	(0.5020)	(0.4670)
ln(post sec.)		-0.76511***		-0.75386***	-0.7539	-1.0007		-5.31820***		-5.29935***	-5.29935***	-4.77322***
		(0.0610)		(0.0600)	(1.2800)	(1.1880)		(0.0830)		(0.0810)	(1.0410)	(1.0300)
ln(R&D)		1.83815***		1.82972***	1.82972**	1.62385**		0.13090***		0.12962***	0.1296	0.1131
		(0.0330)		(0.0330)	(0.7170)	(0.6410)		(0.0260)		(0.0260)	(0.5100)	(0.4650)
lnprdl×R&Dint			0.04823**	-0.0143	-0.0143	-0.0229			-0.17381***	-0.0727	-0.0727	-0.0458
			(0.0240)	(0.0390)	(0.0330)	(0.0320)			(0.0540)	(0.0460)	(0.0730)	(0.0560)
lnprdl×Lint			-0.12015	-0.86238***	-0.86238**	-1.01685***			-1.20853***	-0.97108***	-0.97108***	-1.52951***
			(0.1870)	(0.1510)	(0.3590)	(0.3400)			(0.3030)	(0.2130)	(0.3250)	(0.3890)
lnprdl×Kint			1.14590***	0.50066	0.50066	0.16692			-6.64238***	-4.13689***	-4.13689***	-3.57596***
			(0.3930)	(0.3710)	(0.4610)	(0.4600)			(0.8650)	(0.6430)	(1.1550)	(0.9280)
Observations	30,883	30,883	30,883	30,883	30,883	30,883	16,918	16,918	16,918	16,918	16,918	16,918
R-squared	0.072	0.294	0.077	0.296	0.296	0.343	0.189	0.7	0.19	0.702	0.702	0.725
FE	No	No	No	No	No	Industry	No	No	No	No	No	Industry
Errors	Robust	Robust	Robust	Robust	Cluster	Cluster	Robust	Robust	Robust	Robust	Cluster	Cluster
Sample	euro	euro	euro	euro	euro	euro	noneuro	noneuro	noneuro	noneuro	noneuro	noneuro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 5: Agglomeration and Growth

	Growth (1)	Growth (2)	Growth (3)	Growth (4)	Growth (5)	Growth (6)	Growth (7)	Growth (8)	Growth (9)
Density 50	-0.03691***	-0.03992***	0.01222***	0.01623***	0.00806*	0.01478**			
	(0.009)	(0.013)	(0.003)	(0.004)	(0.005)	(0.007)			
Weighted Density 50							-0.03969***	0.01619***	0.01456**
							(0.013)	(0.004)	(0.007)
Observations	4,947	4,947	6,421	5,164	5,369	4,857	4,947	5,164	4,857
R-squared	0.244	0.347	0.182	0.496	0.07	0.21	0.347	0.495	0.209
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Contols	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Errors	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
Sample	US	US	Euro	Euro	NonEuro	Non Euro	US	Euro	Non Euro

Notes: Real regional growth 2017-2005. Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Appendix Table A1: List of Countries

	Europe	EU	EU-Eurozone	Eurozone
AUSTRIA	1	1	0	1
BELGIUM	1	1	0	1
BULGARIA	1	1	1	0
CROATIA	1	1	1	0
CYPRUS	1	1	0	1
CZECH REP.	1	1	1	0
DENMARK	1	1	1	0
ESTONIA	1	1	0	1
FINLAND	1	1	0	1
FRANCE	1	1	0	1
GERMANY	1	1	0	1
GREECE	1	1	0	1
HUNGARY	1	1	1	0
IRELAND	1	1	0	1
ITALY	1	1	0	1
LITHUANIA	1	1	0	1
LUX.	1	1	0	1
NETHERL.	1	1	0	1
NORWAY	1	0	0	0
POLAND	1	1	1	0
PORTUGAL	1	1	0	1
ROMANIA	1	1	1	0
SLOVENIA	1	1	0	1
SLOVAKIA	1	1	0	1
SPAIN	1	1	0	1
SWEDEN	1	1	1	0
SWITZERL.	1	0	0	0
UK	1	1	1	0
LATVIA	1	1	0	1
US	0	0	0	0

Appendix Table A.2a: Descriptive Statistics: Agglomeration Densities

	Obs.	Mean	Sd	Min	Max
Threshold (T)			United States		
20km	61620	1.312	0.788	0.000	2.609
50km	61620	2.904	1.726	0.000	5.723
100km	61620	6.034	3.520	0.000	11.696
200k	61620	12.372	7.016	0.000	23.316
400k	61620	27.756	14.997	0.001	49.406
			Euro		
20km	32454	5.237	2.361	0.035	8.881
50km	32454	11.562	5.183	0.079	19.506
100km	32454	23.926	10.606	0.171	39.962
200k	32454	48.871	21.191	0.380	80.038
400k	32454	109.431	45.037	1.023	171.654
			Non Euro		
20km	17893	6.619	2.073	0.610	8.915
50km	17893	14.586	4.515	1.374	19.579
100km	17893	30.076	9.106	2.952	40.106
200k	17893	61.006	17.670	6.466	80.327
400k	17893	134.492	34.979	16.782	172.095

Notes: Density is the estimated distance kernel function of at 20, 50, 100, 200, 400km. See text for descriptions of the variables.

Appendix Table A.2b: Descriptive Statistics: Size and Productivity

	Obs.	Mean	Sd	Min	Max
			United States		
ln(Sales)	61576	14.446	1.677	0.000	25.342
ln(PrdL)	61576	11.470	1.097	-4.394	20.905
			Euro		
ln(Sales)	32437	14.545	2.619	0.000	24.465
ln(PrdL)	32437	11.660	2.314	-8.960	21.070
			Non Euro		
ln(Sales)	17883	12.909	4.241	0.000	23.385
ln(PrdL)	17883	9.398	4.149	-9.669	20.771

Notes: Firm level Ln(Sales) and ln(Labor Productivity) from D&B.

Appendix Table A.2c: Descriptive Statistics: Regional Controls

	Obs.	Mean	Sd	Min	Max
			United States		
ln(Post Sec.)	61620	-1.725	0.171	-2.261	-1.115
ln(gdp)	61620	10.534	0.603	7.079	11.863
ln(pop. density)	61620	4.117	0.888	-0.968	8.048
ln(R&D)	61620	3.263	1.183	-1.897	4.765
			Euo		
ln(Post Sec.)	32454	-1.844	0.449	-2.688	-0.826
ln(gdp)	32454	10.276	0.389	8.156	11.139
ln(pop. density)	32454	5.424	0.856	2.851	8.737
ln(R&D)	32454	6.689	1.410	-0.186	9.582
			Non Euro		
ln(Post Sec.)	17893	-1.709	0.477	-2.688	-1.002
ln(gdp)	17893	9.675	0.738	7.768	10.970
ln(pop. density)	17893	5.432	1.299	1.194	8.460
ln(R&D)	17893	5.468	1.481	0.857	8.207

Notes: Regional data: post secondary education attainment, gdp per capital, population density and regional R&D investment. All data in logs.

Appendix Table 3: Agglomeration and Firm Performance:Different Density Thresholds (Labor Productivity)

	Density 50 (1)	Density 100 (2)	Density 200 (3)	Density 50 (4)	Density 100 (5)	Density 200 (6)	Density 50 (7)	Density 100 (8)	Density 200 (9)	Density 50 (10)	Density 100 (11)	Density 200 (12)
ln(PrdL)	0.068047***	0.001376***	0.002657***	-0.037722***	-0.000721***	-0.001244***	0.280603***	0.005802***	0.011816***	-0.023088***	-0.000454***	-0.000833***
	(0.007)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.014)	(0.000)	(0.001)	(0.008)	(0.000)	(0.000)
ln(Age)	0.223039***	0.004550***	0.009065***	-0.536736***	-0.010846***	-0.021125***	0.132269***	0.002685***	0.005273***	-1.923378***	-0.038771***	-0.075144***
	(0.009)	(0.000)	(0.000)	(0.026)	(0.001)	(0.001)	(0.035)	(0.001)	(0.001)	(0.040)	(0.001)	(0.002)
Multi-product	0.040268***	0.000837***	0.001742***	0.804593***	0.016394***	0.032447***	0.719491***	0.014714***	0.029338***	0.811830***	0.016560***	0.032825***
	(0.007)	(0.000)	(0.000)	(0.017)	(0.000)	(0.001)	(0.024)	(0.001)	(0.001)	(0.030)	(0.001)	(0.001)
MNC	0.381808***	0.007748***	0.015246***	0.224562***	0.004801***	0.010416***	1.932941***	0.039282***	0.077454***	-1.578712***	-0.031797***	-0.061538***
	(0.027)	(0.001)	(0.001)	(0.055)	(0.001)	(0.002)	(0.084)	(0.002)	(0.003)	(0.076)	(0.002)	(0.003)
Observations	61576	61576	61576	54242	54242	54242	32437	32437	32437	17883	17883	17883
R-squared	0.056	0.056	0.056	0.156	0.155	0.155	0.173	0.174	0.175	0.325	0.324	0.322
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	No	No	No	No	No	No	No	No	No	No	No	No
Errors	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust	Robust
Sample	US	US	US	Europe	Europe	Europe	Euro	Euro	Euro	Non-Euro	Non-Euro	Non-Euro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Appendix Table 4a: Agglomeration and Regional Policies, Super Size Firms, Productivity

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	0.05146***	-0.01271*	0.04397***	0.04066***	0.005	0.005	0.008	-0.03816***	-0.03943***	-0.06293***	0.000	0.000
	(0.010)	(0.008)	(0.010)	(0.009)	(0.007)	(0.015)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.037)
Suprlnprdl_Sample	-0.98215***	0.45332***	0.67957***	0.31900***	(0.176)	(0.176)	-7.47120***	-2.28878***	0.63324**	0.586	-30.49751***	-30.49751***
	(0.186)	(0.119)	(0.256)	(0.097)	(0.278)	(0.710)	(1.945)	(0.356)	(0.255)	(0.483)	(2.037)	(10.325)
ln(Age)	0.22054***	0.13886***	0.21885***	0.21784***	0.13649***	0.13649***	-0.24219***	-0.59519***	-0.43838***	-0.63721***	-0.29559***	-0.29559**
	(0.009)	(0.007)	(0.009)	(0.009)	(0.007)	(0.032)	(0.027)	(0.025)	(0.026)	(0.026)	(0.025)	(0.145)
Multi-product	0.04019***	0.06192***	0.04421***	0.04619***	0.05188***	0.05188***	0.76171***	0.79348***	0.76265***	0.80145***	0.54932***	0.54932***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.005)	(0.012)	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)	(0.139)
MNC	0.38160***	0.23653***	0.36663***	0.37683***	0.22723***	0.22723***	0.09263*	0.22900***	0.47745***	0.17946***	0.40796***	0.408
	(0.027)	(0.022)	(0.027)	(0.026)	(0.021)	(0.078)	(0.053)	(0.055)	(0.055)	(0.057)	(0.051)	(0.269)
Sprlnprdl_Smple_lgdp	0.10305***				0.003	0.003	0.77839***				3.57497***	3.57497***
	(0.018)				(0.025)	(0.053)	(0.185)				(0.211)	(1.132)
ln(gdp)	0.08597***				-0.18572***	(0.186)	-1.56375***				-4.49602***	-4.49602***
	(0.005)				(0.006)	(0.177)	(0.033)				(0.060)	(1.269)
Sprlnprdl_Smple_lpop		-0.07283***			(0.022)	(0.022)		0.46111***			0.59213***	0.592
		(0.026)			(0.041)	(0.111)		(0.062)			(0.059)	(0.377)
ln(pop density)		1.11895***			1.41464***	1.41464***		0.47388***			0.47307***	0.473
		(0.006)			(0.008)	(0.216)		(0.022)			(0.022)	(0.475)
Sprlnprdl_Smple_lpostsec			0.32912**		(0.131)	(0.131)			0.013		-0.31411**	(0.314)
			(0.150)		(0.177)	(0.369)			(0.159)		(0.151)	(0.751)
ln(post sec)			0.85745***		-2.56537***	-2.56537*			-1.55909***		-1.84353***	(1.844)
			(0.038)		(0.037)	(1.420)			(0.051)		(0.050)	(1.144)
Sprlnprdl_Smple_R&D				-0.05764**	0.03314	0.03314				-0.04032	-1.38220***	-1.38220***
				(0.027)	(0.024)	(0.055)				(0.062)	(0.071)	(0.459)
ln(R&D)				0.22160***	-0.11688***	-0.11688				0.24097***	1.64744***	1.64744***
				(0.006)	(0.006)	(0.268)				(0.014)	(0.023)	(0.575)
Observations	61,576	61,576	61,576	61,576	61,576	61,576	54,242	54,242	54,242	54,242	54,242	54,242
R-squared	0.057	0.368	0.064	0.078	0.417	0.417	0.184	0.167	0.173	0.161	0.303	0.303
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	NO	NO	NO	NO	Yes	Yes	NO	NO	NO	NO	Yes	Yes
Errors	Robust	Robust	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Robust	Robust	Cluster
Sample	US	US	US	US	US	US	Europe	Europe	Europe	Europe	Europe	Europe

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.

Appendix Table 4b: Agglomeration and Regional Policies, Super Size Firms, Productivity, cont.

	Density 50 (1)	Density 50 (2)	Density 50 (3)	Density 50 (4)	Density 50 (5)	Density 50 (6)	Density 50 (7)	Density 50 (8)	Density 50 (9)	Density 50 (10)	Density 50 (11)	Density 50 (12)
ln(PrdL)	0.18033***	0.26391***	0.27066***	0.11374***	0.10379***	0.10379**	0.11850***	0.001	0.07189***	0.06865***	0.13231***	0.13231***
	(0.012)	(0.013)	(0.014)	(0.011)	(0.011)	(0.052)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.043)
Suprlnprdl_Sample	14.33268***	3.06575***	-1.46954***	9.85923***	-21.72800***	(21.728)	(0.643)	-5.68172***	-0.58565**	-4.60661***	1.816	1.816
	(4.268)	(0.662)	(0.302)	(0.631)	(5.858)	(22.425)	(1.222)	(0.530)	(0.285)	(0.561)	(1.995)	(6.111)
ln(Age)	-0.20979***	0.033	0.12016***	-0.19584***	-0.20252***	(0.203)	-0.67290***	-1.82838***	-0.85779***	-1.20865***	-0.53277***	-0.53277***
	(0.033)	(0.034)	(0.034)	(0.031)	(0.031)	(0.141)	(0.032)	(0.041)	(0.033)	(0.038)	(0.029)	(0.171)
Multi-product	0.67863***	0.62139***	0.75211***	0.71122***	0.64939***	0.64939***	-0.19041***	0.75026***	0.47165***	0.32663***	0.07319***	0.073
	(0.023)	(0.024)	(0.025)	(0.022)	(0.023)	(0.216)	(0.026)	(0.030)	(0.024)	(0.029)	(0.023)	(0.077)
MNC	1.87926***	1.79429***	1.71708***	1.69078***	1.88552***	1.88552***	-1.25903***	-1.38350***	-0.63605***	-1.26213***	-0.67692***	-0.67692***
	(0.091)	(0.084)	(0.088)	(0.085)	(0.087)	(0.413)	(0.055)	(0.077)	(0.058)	(0.068)	(0.047)	(0.209)
Sprlnprdl_Smple×lgdp	-1.35177***				3.06392***	3.064	(0.005)				-0.73394***	(0.734)
	(0.406)				(0.643)	(2.345)	(0.122)				(0.157)	(0.533)
ln(gdp)	4.01115***				0.029	0.029	-4.10924***				-2.50349***	-2.50349***
	(0.127)				(0.155)	(2.743)	(0.042)				(0.060)	(0.677)
Sprlnprdl_Smple×lpop		-0.49712***			0.62160***	0.622		0.73231***			-0.16665**	(0.167)
		(0.113)			(0.181)	(0.473)		(0.086)			(0.075)	(0.290)
ln(pop density)		1.26421***			0.11019***	0.110		-0.37043***			0.87556***	0.87556*
		(0.036)			(0.040)	(0.788)		(0.033)			(0.025)	(0.485)
Sprlnprdl_Smple×lpostsec			-1.17943***		-0.40127**	(0.401)			0.124		-1.02347**	(1.023)
			(0.181)		(0.197)	(0.747)			(0.190)		(0.408)	(0.923)
ln(post sec)			0.80952***		-0.99203***	(0.992)			-6.16803***		-4.76332***	-4.76332***
			(0.072)		(0.065)	(1.216)			(0.055)		(0.083)	(1.012)
Sprlnprdl_Smple×R&D				-1.33195***	-1.94174***	-1.94174***				0.57839***	0.72916***	0.72916**
				(0.081)	(0.092)	(0.554)				(0.077)	(0.122)	(0.292)
ln(R&D)				1.65066***	1.70903***	1.70903**				-1.40821***	0.0376	0.0376
				(0.022)	(0.033)	(0.656)				(0.020)	(0.025)	(0.477)
Observations	32,437	32,437	32,437	32,437	32,437	32,437	17,883	17,883	17,883	17,883	17,883	17,883
R-squared	0.251	0.212	0.177	0.336	0.343	0.343	0.633	0.341	0.627	0.486	0.722	0.722
FE	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry	Industry
Region Controls	NO	NO	NO	NO	Yes	Yes	NO	NO	NO	NO	Yes	Yes
Errors	Robust	Robust	Robust	Robust	Robust	Cluster	Robust	Robust	Robust	Robust	Robust	Cluster
Sample	Euro	Euro	Euro	Euro	Euro	Euro	nonEuro	nonEuro	nonEuro	nonEuro	nonEuro	nonEuro

Notes: Density 50 is the estimated distance kernel function of at 50 km. See text for descriptions of the variables. *** p< 0.01, ** p< 0.05, * p< 0.1.