# Spatial Agglomeration and Superstar Firms: Firm-level Patterns from Europe and the U.S.

Laura Alfaro
Harvard Business School
& NBER

Maggie X. Chen
George Washington University

Harald Fadinger
University of Mannheim
& CEPR

ECB Forum, 2019

# Backdrop

- Falling transportation costs, reduction of policy barriers, and technological progress: explosion of cross-border flows in goods, services, investments, and ideas.
  - Potential to reshape the landscape of economic geography and business network
    - European integration: predicated on the free movement of goods, services, labour and capital.
- A key driver of this phenomenon is the “superstar firms” (Rosen, 1981)
  - Very large, productive firms dominant in particular industries
    - Engage complex organization,
    - Transporting products, tasks, capital, and technology across countries.

# The Paradox of Globalization

- Geographic proximity could lead to agglomeration economies
  - Lower transport costs between inputs, labour-market and capital-goods-market externalities due to the proximity of firms with similar demands, and technology diffusion.

- Economic integration may lead to a reduction of geographic proximity/ agglomeration benefits
  - Movement of goods, people and ideas becomes easier.

- Dominance of superstar firms, industrial clusters, and cities despite reductions in transportation and communication costs and the competition implications of geographic concentration (Glaeser, 2010).

# What Do We Do

- We explore how firm heterogeneity affects the formation of an industrial landscape.
  - Significant productivity heterogeneity across firms within each industry and across countries (Helpman, Melitz, Yeaple, 20004).
    - Explore the geographic distribution of economic activities surrounding each plant (as the unit of observation).
- Examine how the degree of agglomeration varies with firm attributes (productivity, size and multinational status) and regional characteristics
  - Assess the potential benefits and costs provided by geographic proximity to productive firms
  - Relative to the effects of location fundamentals and the ability of regional policies to attract regional industry clusters centred around them.
- Compare agglomeration patterns in the U.S. with those in Europe
  - Eurozone: face deeper integration
    - Within Eurozone countries

# Questions

- Is there agglomeration around highly productive firms?
  - Is agglomeration driven by multinationals?
- In addition to firm productivity and internationalization:
  - What is the importance of internal markets and regional characteristics and policies?
- Does the Euro area share similar patterns to the U.S. and the rest of Europe?
  - How do countries within the Euro area compare?

# How Do We Address These Questions?
# Micro Index

- To take into account the role of firm heterogeneity,
  - Micro index of agglomeration for each individual plant,
    - Following an empirical methodology introduced by Duranton and Overman (2005) (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.
  - 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 (productivity) leads to different levels of ability to attract agglomeration.

# How Do We Address These Questions? Methodology

- To mitigate the concerns of reverse causality, we explore the dynamics in the data and examine the spatial relationship between incumbent and entrant plants.

  1) 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.

  2) Second, we identify the role of firm characteristics in determining the level of agglomeration by comparing plants located in the same disaggregated region.

# How Do We Address These Questions?
# Data

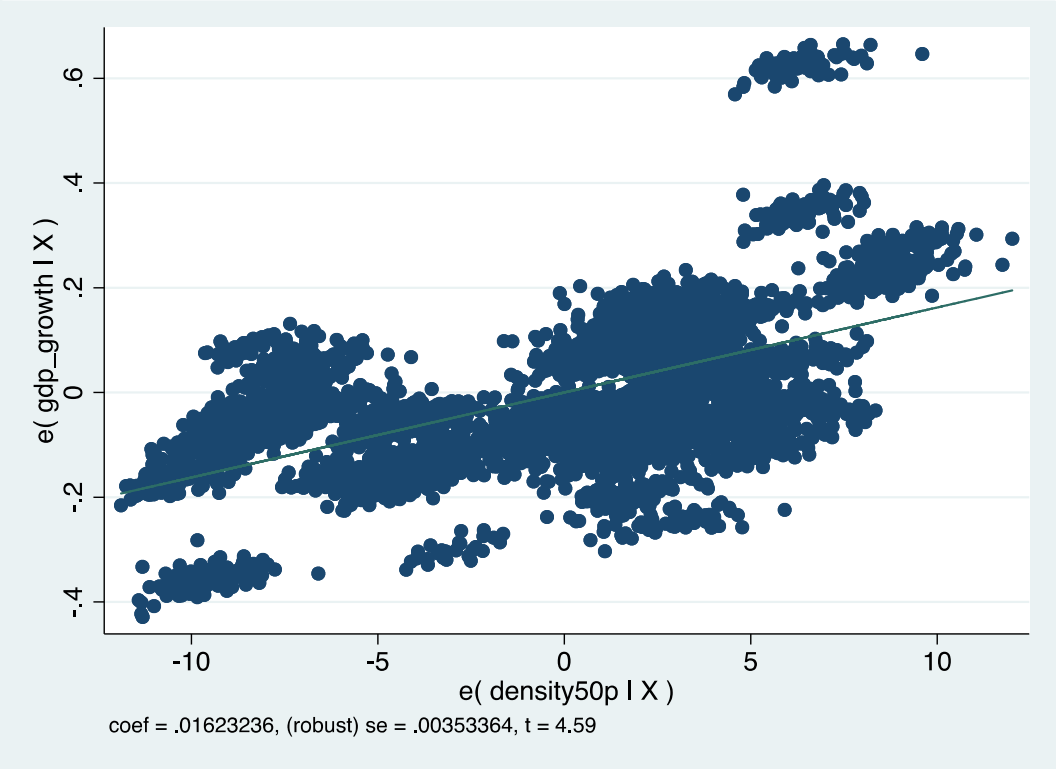
- We employ D&B data set: worldwide establishment dataset that provides detailed location, ownership and activity information (Alfaro and Chen, 2014, 2019; Alfaro, Conconi, Fadinger and Newman, 2016).
  - Dataset reports the physical address of each establishment to obtain latitude and longitude codes for each establishment (geocodes).
    - Information enables us to construct a global index of agglomeration using continuous metrics of space.
- We then construct the index of agglomeration based on the distance each pair of manufacturing establishments.

# Overview of Findings: Agglomeration and Star Firms

- Firms are far from equal within each industrial cluster
  - 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.
  - Eurozone: MNC plants attract more agglomeration than domestically owned plants.
- Region attributes also play an important role:
  - Regional attributes account for about 30-70 percent of the agglomeration.

# Why Does This Matter?
# Growth and Agglomeration

Eurozone Countries: Regional GDP Growth Rates (2005-2017)




$$regional\ growth_r \equiv \beta_1 Density_{k,r} + \gamma' X_r + \delta_k + \varepsilon_i$$

Eurozone: One-standard-deviation increase in agglomeration is associated with a 6-percentage-point increase in growth; (non euro, 3 percentage point).

# Related Literature

- Literature in trade examining agglomeration (Head et al., 1995; Ottaviano and Puga 1998; Ottaviano and Thisse, 2004; Redding, 2010, 2011.
- The urban economics literature on domestic agglomeration (Ellison, Gleaser 1997; Ellison, Glaeser, Kerr, 2009; DO, 2005 2008);
  - Agglomeration important in explaining increase in industry productivity and innovation (Goldfarb and Greenstein, 2016);
  - 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 to employment density of 4.5% to 5% in the US.
  - Gutierrez and Philippon (2017) decreasing domestic competition related to lower US investment.
  - Super-stars, MNCs: Hornbeck and Moretti (2010) agglomeration of a new "Million Dollar Plant; Alfaro and Chen (2016, 2019) distinct agglomeration patterns of MNC, with location playing an important role.
- Examine micro agglomeration patterns by constructing and exploring plant-level agglomeration indices.

# Outline

- Introduction and Motivation
- Quantifying Agglomeration: Methodology and Data
- Empirical evidence
  - Stylized Facts
- Concluding Remarks
  - Preliminary Implications and Future Research

# Digression: Measuring Localization

- "Measuring spatial concentration of activity is a far less trivial exercise that might seem at first sight" (Head and Mayer, 2004)
  - The measurement problem: space is continuous.
- DO's index:
  - Treats space as continuous,
  - Construction based on a counterfactual approach
    - Controls for the effect of location factors---such as market size, natural resources, and policies---that apply to all manufacturing plants.
- The index exhibits key properties:
  - Comparable across industries
  - Controls for the overall location patterns of manufacturing
  - Accounts for industrial concentration
  - Unbiased with respect to scale and aggregation (spatial continuity)
  - Indication of the significance of the results

# Methodology

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

1) 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.

2) **Counterfactual density** functions based on establishments in the same industry to control for factors that affect all plants in the industry.

3) **Agglomeration index** to measure the extent to which an establishment in a given industry attracts agglomeration at a threshold distance relative to the counterfactuals.

# Step 1: Kernel Estimator

- Calculate geographic density function for each establishment in a given industry.
- For each establishment i with primary industry k, we obtain the kernel estimator of bilateral distances at any point d (i. e. , $f_i(\mathrm{d})$).

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

- $n_i$ is the cardinality of $i$'s industry cluster, $h$ is bandwidth, and $K$ the kernel function; Gaussian kernel with the bandwidth set to minimize the mean integrated squared error.
- Given the potential noise in the measurement of trade costs:
  - Kernel smoothing when estimating the distribution function.
- We limit the analysis to firms within the same 3-digit manufacturing sector to ease the computation burden.

# Step 2: Counterfactuals

- Counterfactual kernel estimator for each establishment, i.e., $\bar{f}_i(d)$.
- We use the mean kernel estimates of each industry as the counterfactual.
  - Control for all factors common to establishments in the same industry and to focus on each establishment's deviation from its average counterpart

# Step 3: Agglomeration Density

- Density index for each establishment, i.e.,

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

- 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.
- We compare the kernel estimators at various distance thresholds.
  - We focus on 50km.
  - Lower thresholds (10km and 20km) and higher distances (200km, 400km).

# Data: The WorldBase Database

- The  index requires detailed physical location information for each establishment.
  - WorldBase dataset + geocoding.
  - 2004/05 and 2017/18
- We use four main categories of information for each establishment:
  - Primary industry information including four-digit SIC code
    - Analysis 3 digit
  - Ownership information including headquarter and global parent
    - Foreign-owned establishment: (i) reports to a global parent firm, (ii) the global parent is located in a different country
  - Operational information
    - Employment, sales

# Geodata

- Geocoding:
  - Geocoding Databases for Europe:  latitudes and longitudes of cities and postcodes of most European countries;
  - GeoNames, a website of geographical database covers all countries.
  - Google's Geocoding API services.
- 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 (limitation of the method).

# Regional Data

- We examine activity at the region level: Eurostat Regional Database at the NUTS 2 level of disaggregation.
- GDP per capita, population density, schooling (percentage of population with more than secondary education), all measured in 2004 or the closest year available (to mitigate causality concerns), regional R&D expenditure (public and private); population, tax policy; additional analysis (2016)
- OECD STAN data and NBER-CES Manufacturing Industry Database; Upstreamness (Antràs and Chor (2013).
- US: Bureau of Labour Statistics.

# Empirical Procedure

- Measure the degree to which a plant is proximate to other plants and examine how plant characteristics (productivity, ownership structure, size, age and the number of products) might explain the extent of agglomeration centred around each plant:

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

- $Density_i(T)$: 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.
  - Baseline results: plant-level agglomeration indices at 50 km.

- Labour productivity as our main measure of firm performance; Vector of industry dummies, represented by $D_k$, to control for industry specific factors; $Z_r$ different regional characteristics.

# Agglomeration and Firm Performance

## Are firm characteristics important in explaining agglomeration?

$$Density_{i,k}(T) \equiv \beta_1 \text{performance}_i + \beta_2 \log(\text{age}_i) + \beta_3 \text{multiproduct}_i + \delta_k + \varepsilon_i \quad (1)$$

- $\text{Performance}_i$ is log(labour productivity)
- $\text{Multiproduct}_f$ is a dummy for the plant being active in multiple 4-digit industries, $\delta_k$ is a 3-digit industry fixed effect; age (year started)
- Industry FE: industry factors which may affect the relationship between regional economic density and plant performance
- U.S., Eurozone and non Euro, European Union countries

# Agglomeration and Productivity

Table 1: Agglomeration and Firm Performance (Labor Productivity**)**

| | Density 50km (1) | Density 50km (2) | Density 50km (3) |
|---|---|---|---|
| ln(PrdL) | 0.0940*** | 0.2386*** | 0.005 |
| | (0.007) | (0.015) | (0.009) |
| ln(Age) | 0.2106*** | 0.1330*** | -1.895*** |
| | (0.009) | (0.035) | (0.040) |
| Multi-product | 0.0406*** | 0.7219*** | 0.6850*** |
| | (0.007) | (0.025) | (0.029) |
| Observations | 61,576 | 32,437 | 17,883 |
| R-squared | 0.053 | 0.16 | 0.305 |
| FE | Industry | Industry | Industry |
| Region Controls | No | No | No |
| Errors | Robust | Robust | Robust |
| Sample | US | 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.

# Agglomeration and Super Stars

**Do "superstar" firms attract additional agglomeration compared to more productive plants?**

$$Density_{i,k}(T) \equiv \beta_1 \text{performance}_i + \beta_2 \log(\text{age}_i) + \beta_3 \text{multiproduct}_i + \beta_4 \text{superstar}_i + \delta_k + \varepsilon_i \qquad (2)$$

- "Superstar": dummy equals one if a given plant belongs to the top 5% of labour productivity defined as plants that are within the top 5% of the labour productivity distribution within a given 2-digit sector within each region (robustness with 1%).

# Agglomeration and Super Stars

Table 2: Agglomeration and Firm Performance (Labor Productivity)
Super-Stars

| | Density 50km | Density 50km | Density 50km |
|---|---|---|---|
| | (1) | (2) | (3) |
| ln(PrdL) | 0.0746*** | 0.2072*** | 0.0364*** |
| | (0.010) | (0.015) | (0.009) |
| Super_ln(Prdl)_Sample | 0.1296*** | 1.2205*** | -2.3358*** |
| | (0.044) | (0.095) | (0.136) |
| ln(Age) | 0.2107*** | 0.1346*** | -1.8702*** |
| | (0.009) | (0.035) | (0.040) |
| Multi-product | 0.0405*** | 0.7248*** | 0.6821*** |
| | (0.007) | (0.025) | (0.029) |
| | | | |
| Observations | 61,576 | 32,437 | 17,883 |
| R-squared | 0.054 | 0.162 | 0.317 |
| FE | Industry | Industry | Industry |
| Region Controls | No | No | No |
| Errors | Robust | Robust | Robust |
| Sample | US | 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.

# Agglomeration, Super Stars and MNC

## Are results driven by Multinationals (larger, more productive, internationalized)?

- We add a $\text{multinational}_i$ is a dummy for multinational affiliate:

$$Density_{i,k}(T) \equiv \beta_1 \text{performance}_i + \beta_2 \text{multinational}_i + \beta_3 \log(\text{age}_i) + \beta_4 \text{multiproduct}_i + \beta_5 \text{superstar}_i + \delta_k + \varepsilon_i \quad (3)$$

- Ownership information including headquarter and global parent
  - Foreign-owned establishment: (i) reports to a global parent firm, (ii) the global parent is located in a different country

# Agglomeration, Super Size, MNC

Table 3: Agglomeration and Firm Performance (Labor Productivity)
Super Stars and MNCs

| | Density 50km (1) | Density 50km (2) | Density 50km (3) | Density 50km (4) | Density 50km (5) | Density 50km (6) |
|---|---|---|---|---|---|---|
| ln(PrdL) | 0.0746*** | 0.0527*** | 0.2072*** | 0.2633*** | 0.0364*** | 0.003 |
| | (0.010) | (0.010) | (0.015) | (0.014) | (0.009) | (0.008) |
| Suprlnprdl_Sample | 0.1296*** | 0.1029** | 1.2205*** | 0.5926*** | -2.3358*** | -1.6924*** |
| | (0.044) | (0.044) | (0.095) | (0.099) | (0.136) | (0.139) |
| ln(Age) | 0.2107*** | 0.2229*** | 0.1346*** | 0.1330*** | -1.8702*** | -1.9015*** |
| | (0.009) | (0.009) | (0.035) | (0.035) | (0.040) | (0.040) |
| Multi-product | 0.0405*** | 0.0401*** | 0.7248*** | 0.7210*** | 0.6821*** | 0.7922*** |
| | (0.007) | (0.007) | (0.025) | (0.024) | (0.029) | (0.030) |
| MNC | | 0.3789*** | | 1.8394*** | | -1.3613*** |
| | | (0.027) | | (0.086) | | (0.078) |
| Observations | 61,576 | 61,576 | 32,437 | 32,437 | 17,883 | 17,883 |
| R-squared | 0.054 | 0.056 | 0.162 | 0.173 | 0.317 | 0.33 |
| FE | Industry | Industry | Industry | Industry | Industry | Industry |
| Region Controls | No | No | No | No | No | No |
| Errors | Robust | Robust | Robust | Robust | Robust | Robust |
| Sample | US | US | Euro | 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.

# Findings: Agglomeration and Firms

- **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 superstar plants in the Eurozone, and in the U.S. There is less agglomeration of economic activity around superstar plants outside of the Eurozone.

- **Finding 3:** There is more agglomeration of economic activity around affiliates of multinationals, in particular in the U.S. and in the Eurozone.

# Agglomeration and Regional Policies

## What is the role of regional characteristics?

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

- $X_r$ is a vector of region controls: 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, public and private (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.

# Agglomeration and Regional Variables

Table 4: Agglomeration, Super-Size and Regional Variables

| | Density 50km (1) | Density 50km (2) |
|---|---|---|
| ln(PrdL) | 0.1072*** | 0.1319*** |
| | (0.011) | (0.007) |
| SuprlnPrdl_Sample | 0.3475*** | -0.4644*** |
| | (0.102) | (0.090) |
| ln(Age) | -0.2127*** | -0.5362*** |
| | (0.031) | (0.029) |
| Multi-product | 0.6707*** | 0.0734*** |
| | (0.023) | (0.023) |
| MNC | 1.8760*** | -0.6852*** |
| | (0.087) | (0.047) |
| ln(gdp) | 0.2422 | -2.5562*** |
| | (2.692) | (0.667) |
| ln(pop. density) | 0.1540 | 0.8700* |
| | (0.772) | (0.464) |
| ln (post sec.) | 0.9858 | -4.7620*** |
| | (1.195) | (1.031) |
| ln(R&D) | 1.59493** | 0.071 |
| | (25.086) | (6.962) |
| Observations | 32,437 | 17,883 |
| R-squared | 0.336 | 0.721 |
| FE | Industry | Industry |
| Region Controls | Yes | Yes |
| Errors | Cluster | Cluster |
| Sample | 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. Non clustered errors for firm level variables.

# Attracting Agglomeration: Role of Regional Policies

**Can we think of regional policies that will foster agglomeration around high-performance plants?**

- We augment our previous specification by interacting firm performance measures with regional variables.

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

- Interactions of firm-level log labour productivity with: post-secondary schooling, and regional R&D spending (in logs).

# Attracting Agglomeration

Table 5: Agglomeration and Regional Policies

| | Density 50km (3) | Density 50km (4) | Density 50km (5) | Density 50km (6) |
|---|---|---|---|---|
| ln(PrdL) | 0.27066*** | 0.11374*** | 0.07189*** | 0.06865*** |
| | (0.014) | (0.011) | (0.007) | (0.008) |
| SuprlnPrdl_Sample | -1.46954*** | 9.85923*** | -0.58565** | -4.60661*** |
| | (0.302) | (0.631) | (0.285) | (0.561) |
| ln(Age) | 0.12016*** | -0.19584*** | -0.85779*** | -1.20865*** |
| | (0.034) | (0.031) | (0.033) | (0.038) |
| Multi-product | 0.75211*** | 0.71122*** | 0.47165*** | 0.32663*** |
| | (0.025) | (0.022) | (0.024) | (0.029) |
| MNC | 1.71708*** | 1.69078*** | -0.63605*** | -1.26213*** |
| | (0.088) | (0.085) | (0.058) | (0.068) |
| ln(post sec) | 0.80952 | | -6.16803*** | |
| | (1.767) | | (0.997) | |
| SprlnPrdl_Smp×ln(post sec) | -1.17943*** | | 0.124 | |
| | (0.181) | | (0.190) | |
| ln(R&D) | | 1.65066*** | | -1.40821*** |
| | | (0.505) | | (0.181) |
| SprlnPrdl_Smp×R&D | | -1.33195*** | | 0.57839*** |
| | | (0.081) | | (0.077) |
| Observations | 32,437 | 32,437 | 17,883 | 17,883 |
| R-squared | 0.177 | 0.336 | 0.627 | 0.486 |
| FE | Industry | Industry | Industry | Industry |
| Region Controls | NO | NO | NO | NO |
| Errors | Cluster | Cluster | Robust | Robust |
| Sample | Euro | Euro | 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. Non clustered errors for firm level variables.

# Findings: Agglomeration and Regional Policy

- **Finding 4:** Higher levels of regional R&D spending are associated with more agglomeration in manufacturing in Europe, in particular inside of the Eurozone.

- **Finding 5:** Better location fundamentals lowers the gains from, and hence the incentives of, agglomeration, especially around superstar.

# Agglomeration After the Crisis

**Has the agglomeration around highly productive firms remained stable in Eurozone and no Eurozone countries?**

$$Density_i(T) \equiv \alpha + \beta\theta_i + \gamma Z_r + D_k + \varepsilon_i, \qquad (6)$$

- $Density_i(T)$: estimated density of establishment i's network using 2018 vintage D&B.
  - Baseline results: plant-level agglomeration indices at 50 km, manufacturing
  - Different groups of countries in Europe.
    - Eurozone
    - Euro Crisis: Ireland, Greece, Italy, Portugal, Spain

# Patterns after Crisis

Table 6: Agglomeration: Patterns after Crisis

| | $Density_{18}$ 50km (1) | $Density_{18}$ 50km (2) | $Density_{18}$ 50km (3) | $Density_{18}$ 50km (4) |
|---|---|---|---|---|
| ln(PrdL) | 0.08073*** | 0.06420*** | 0.13409*** | -0.01383** |
| | (0.004) | (0.005) | (0.006) | (0.003) |
| MNC | 0.14407*** | 0.07301*** | 0.11397*** | 0.1319 |
| | (0.009) | (0.011) | (0.013) | (0.144) |
| Observations | 204,836 | 96,316 | 108,941 | 154,999 |
| R-squared | 0.059 | 0.12 | 0.121 | 0.117 |
| Firm Controls | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Region Controls | Yes | Yes | Yes | Yes |
| Errors | Robust | Robust | Robust | Robust |
| Sample | Euro | Non crisis Euro | Ireland, Greece, Italy, Portugal, Spain | nonEuro |

| | $Density_{18}$ 50km (1) | $Density_{18}$ 50km (2) | $Density_{18}$ 50km (3) | $Density_{18}$ 50km (4) |
|---|---|---|---|---|
| ln(PrdL) | 0.08974*** | 0.06264*** | 0.16725*** | -0.02674** |
| | (0.004) | (0.005) | (0.007) | (0.004) |
| Suprlnprdl_Sample | -0.09376*** | 0.01867 | -0.28999*** | 0.13392* |
| | (0.018) | (0.026) | (0.023) | (0.068) |
| MNC | 0.14344*** | 0.07314*** | 0.11237*** | 0.1293 |
| | (0.009) | (0.011) | (0.013) | (0.113) |
| Observations | 204,836 | 96,316 | 108,941 | 154,999 |
| R-squared | 0.059 | 0.12 | 0.122 | 0.117 |
| Firm Controls | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Region Controls | Yes | Yes | Yes | Yes |
| Errors | Robust | Robust | Robust | Robust |
| Sample | Euro | Non crisis Euro | Euro Crisis | 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.

# Summary

- Firm characteristics matter: Some firms are more centred than others
  - Firms, including MNCs, are far from equal within each industrial cluster.
  - Larger and more productive establishments are centred with more agglomeration than their smaller, less productive counterparts.
    - Reflecting greater potential spillovers from leading firms due perhaps to the more complex activities engaged by these firms in the regions.
- Region attributes play an important role.
  - Better location fundamentals (such as human capital and R&D spending) could weaken the incentive to agglomerate around super large firms.
- Movement of goods, people and ideas have become easier through integration:
  - Continuing dominance of certain firms, despite reductions in transportation and communication costs.

# Preliminary Policy Implications

- Firm heterogeneity: Policies aimed to build industrial zones and foreign investment should take into account the different abilities of firms to stimulate new entrepreneurship activities (“one size fits all?”).
- Firms with better performance and superior characteristics can help attract more entrants, generate a domino effect in the formation of industrial clusters.
  - Euro countries, superstar and MNC firms in area more likely to specialize in upstream, knowledge and capital intensive tasks.
- An incentive structure whereby favourable incentives are offered first to potential hub firms could be more effective than a uniform incentive system.
  - Consideration should be given to the interdependence of firm’s location decisions especially in policy making aimed at influencing FDI.
  - The design of such an incentive structure should be carefully devised to assess the potential of agglomeration economies across regions and industries.

# Related Research

- Understanding the mechanisms through which superstar firms attract agglomeration differently in different regions (complement/substitutes).
  - Role and interaction of private (within firm) and public policies
    - E.g. firm level R&D versus public?
  - Relation of capital market integration and capital goods externalities
    - Evidence of agglomeration around capital intensive sectors
      - What is the role of deeper local capital market around star firms/agglomerated centers?
      - Role of lending to underperforming firms (cluster dynamism)?
- How agglomeration patterns change with the rise of protectionism and uncertainty?