Utrecht University

School of Economics

Non-local Startups and Entrepreneurial Economies

Leonardo Mazzoni, Massimo Riccaboni & Erik Stam

U.S.E. Working Paper Series nr: 22-03

U.S.E. Research Institute
Kriekenpitplein 21-22, 3584 EC Utrecht, The Netherlands Tel: +31 30 253 9800, e-mail: use.ri@uu.nl www.uu.nl/use/research

Utrecht University **School of Economics**

U.S.E. Research Institute
Working Paper Series 22-03
ISSN: 2666-8238

Non-local Startups and Entrepreneurial Economies

Leonardo Mazzoni*
Massimo Riccaboni*
Erik Stam**
* IMT School for Advanced Studies, Lucca
**Utrecht School of Economics
Utrecht University

April 2022

Abstract
Research on entrepreneurship and economic development has emphasized the local bias of entrepreneurship. We question this local bias and show the relevance of startups founded by non-locals and interregional flows of non-local founders in Italy. In general, the quality of entrepreneurial economies is positively related to the creation of local startups, providing evidence for creation mechanisms. However, the majority of startups in Italy has at least one non-local founder. Using the Poisson Pseudo Maximum Likelihood gravity model and a novel machine learning application of the Poisson Pseudo Maximum Likelihood postestimation techniques based on LASSO, we show evidence that also interregional attraction and supply mechanisms are at work. Increasing the quality of the entrepreneurial economy leads to higher levels of entrepreneurship locally via attraction of entrepreneurial talent from other regions, but also increases the supply non-local founders to other higher quality entrepreneurial economies.

Keywords: entrepreneurial economies, entrepreneurial ecosystems, innovative startups, entrepreneurship, local bias, migration, gravity models, networks

JEL classification: D2, E02, L26, M13, O43, P00, R1, R23, R58

Acknowledgements: We would like to thank Bram Timmermans and Scott Stern, and the participants of seminars at the Utrecht University School of Economics, IMT School for Advanced Studies and ESADE for valuable comments on prior versions of this paper. And thanks to Tom Zylkin for providing us the code for the Lasso PPML technique.

Comments welcomed to: e.stam@uu.nl

Non-local Startups and Entrepreneurial Economies

1. Introduction

Many countries and regions aspire to stimulate entrepreneurship-led economic development. There is a multiplicity of policies to make countries, regions, and cities a better place for entrepreneurship (OECD 2021). Academic research has shown that entrepreneurship is largely a local event (Feldman, 2001; Michelacci & Silva, 2007; Fritsch & Storey, 2014), with entrepreneurs often being locationally inert (Figueiredo et al., 2002; Stam, 2007; Dahl & Sorenson, 2012). The home bias of founders (Figueiredo et al., 2002) can be explained by financial market mechanisms (Michelacci & Silva, 2007), cognitive biases in opportunity recognition (Sorenson & Audia, 2000), agglomeration economies (Glaeser et al., 2010), and social networks of founders (Stam, 2007; Dahl & Sorenson, 2009).

A key argument in urban and regional economics, and the geography of entrepreneurship in particular, is that there is a distinct set of local conditions that enables the creation of new firms – *creation mechanisms*. Creation mechanisms explain how the qualities of (regional) entrepreneurial economies affect the prevalence of (different types of) entrepreneurial activity. These studies focus on entrepreneurial activity in specific industries in particular regions (Glaeser & Kerr, 2009; Buenstorf & Klepper, 2010; Frenken et al., 2015), or on particular elements of entrepreneurial economies, for example, finance (Michelacci & Silva, 2007; Samila & Sorenson, 2011), culture (Fritsch & Wyrwich, 2014; Falck et al., 2017), demography (Bönte et al. 2009), social capital (Bauernschuster et al., 2010), agglomeration economies (Glaeser et al., 2010), or the overall quality of entrepreneurial economies and its entrepreneurial outputs (Guzman & Stern 2020; Stam & Van de Ven 2021; Leendertse et al. 2022). The substantial heterogeneity in the quality of entrepreneurial economies (Glaeser et al., 2010;

Chatterji et al., 2014) leads to highly uneven spatial distributions of entrepreneurship across cities, regions and countries, which is a driver of spatially uneven economic growth and development (Glaeser et al., 2015).

Many entrepreneurs stay in their region of origin, but a substantial group of entrepreneurs does migrate between countries (Azoulay et al. 2022) and regions (Reuschke & Van Ham, 2013). This questions whether the *creation mechanism* is enough to understand the uneven spatial distribution of entrepreneurship. To what extent do high-quality entrepreneurial economies not only nurture local founders (the creation mechanism) but also attract non-local founders? Despite the recent attention to the geographical mobility of startups (Bryan and Guzman, 2021), there is a paucity of insight into the flows of (potential) founders and the relationship with geographical contexts (see Bönte et al., 2007; Anelli et al., 2019; Bonaventura et al., 2020). We need more insight into the *attraction mechanisms*, explaining why some regions attract more (potential) founders than others. There may also be an interplay between the creation and attraction mechanisms, explaining where entrepreneurs are (partly) nurtured and leave to other more attractive places: *supply mechanisms*. The latter illustrates an escalator mechanism, when (nascent) entrepreneurs move from good to even better entrepreneurial economies (cf. the dominance of the US – in particular California – as attractor of Israeli startups – mostly from the Tel Aviv area: Conti & Guzman 2022).

The attraction of global cities as catalysts of innovation is a relatively well-documented phenomenon (see Bettencourt et al. 2007; Verginer and Riccaboni, 2021), but the other local entrepreneurship dynamics are still understudied. In the same way, interregional migration has been demonstrated to be a key driver of the labor market and professional network development, with effects both for regions of origin (brain drain) and for destination regions (brain gain), with the emergence of possible complementarities between the innovative capacity of migrants and locals (Bosetti et al., 2015; Faggian et al., 2017; Basile et al., 2019). Younger and more highly educated people are more likely to move over longer distances and might thus be attracted to high-quality entrepreneurial economies (cf. Anelli et al., 2019). This can be very relevant for young founders, attracted by creative and vivid environments, with better conditions to explore breakthrough ideas (Acemoglu et al., 2014; Bratti & Conti, 2018).

Many geography of entrepreneurship studies have analyzed local conditions with isolated elements and regional economies as closed systems. We conceptualize (entrepreneurial) economies as complex open systems (Arthur 2014; 2021), which enables us to take into account the multidimensional and interdependent nature of entrepreneurial economies and to investigate emergence within economies and effects on other economies. Emergence is tackled mainly with creation mechanisms, explaining why some places nurture more entrepreneurship than others. This is central to the new entrepreneurial ecosystem approach (Stam, 2015; Stam & Van der Ven, 2021; Leendertse et al., 2022) that provides a more suitable model to study entrepreneurship and its inherent uncertainty (Knight 1921) and out-of-equilibrium dynamism (Schumpeter 1934) than general equilibrium models (Lucas 1978).

Building upon and going beyond the current literature, we propose that the quality of entrepreneurial economies not only enables local startups (creation mechanisms), but that it is also an essential mechanism for attracting and supplying non-local founders. Not only the quality of the entrepreneurial economy of destination is essential for understanding the mobility of (potential) founders: also the quality of the entrepreneurial economy of origin matters, which may act as a supplier of non-local founders. Two insights from economic gravity models may be important here. First, a negative effect is expected of physical distance between two regions:

the more distant two regions are, the smaller the flows of (potential) founders between them. However, the literature that analyses spatial flows often reports a positive relationship between economic indicators of origin, like GDP, and the presence of a particular output (for example, trade; cf. Lewer & Van den Berg, 2008). The second and perhaps less intuitive insight from gravity models is that the larger the size of economic activity in particular regions, the less the distance between the two is an obstacle to exchanging goods, services, or human capital. For example, this explains the relatively large flows of people, goods, and services between the world cities London and New York, despite the considerable distance between them (Deruddder & Taylor, 2005). Applying this to an entrepreneurial setting raises the question of whether people leave regions with a very low-quality entrepreneurial economy out of necessity, as exemplified by the large-scale historical emigration out of the Mezzogiorno (Southern Italy; Hirschman 1970), or whether migrants are more likely to come from regions with relatively many economic activities and a relatively high-quality entrepreneurial economy.

This paper aims to provide insight into the role of entrepreneurial economies as creators, attractors, *and* suppliers of startup founders. The paper's central question is: to what extent and how does the quality of the entrepreneurial economies of origin and destination explain the prevalence of non-local startups? Do these non-local startups leave low-quality entrepreneurial economies, and are they created in high-quality entrepreneurial economies? This may reinforce the quality of the receiving region and diminish the quality of the area left behind (Anelli et al. 2019), and in the end, contribute to increased regional inequality, but perhaps overall (national) welfare increases.

With this paper, we provide novel insights into long-standing discussions on the role of entrepreneurship, geographic mobility, and economic development, going back to Hirschman

(1970), and connected to the recent wave of geography of entrepreneurship studies in economics (Bönte et al 2009; Glaeser et al 2010; Bryan & Guzman, 2021; Conti & Guzman 2022), management (Dahl & Sorenson 2009; 2012; Vedula & Kim 2019) and geography (Stam 2007; Reuschke & Van Ham 2013). In this paper, we tackle novel questions, with new methodologies in a multimethod study into the geographical mobility of founders between regions with heterogeneity in the quality of their entrepreneurial economies. We do this with data on innovative startups in Italian regions: an economy historically dominated by industrial districts with large interregional differences (Capozza et al., 2018), multiple economic core regions, very low entrepreneurship rates (GEM 2021), and low interregional mobility (Bonifazi et al. 2017). An ideal test site for studying the effect of heterogeneity in the quality of entrepreneurial economies on the presence of non-local startups.

In this paper, we build on entrepreneurial ecosystem studies measuring the quality of regional entrepreneurial economies (Stam & Van de Ven 2021; Leendertse et al. 2022) and improve these with more fine-grained data on 105 Italian NUTS 3 regions. An entrepreneurial ecosystem approach enables us to accommodate the complex systems nature of (entrepreneurial) economies (Arthur 2014; Wurth et al. 2022) and combined with a gravity model, it also allows us to analyze entrepreneurial economies as open systems to understand better and untangle the relevance of creation, attraction, and supplier mechanisms.

Analyzing Italian innovative startup data, reveals that the majority of innovative startups has at least one non-local founder. Why do (potential) founders leave their region of origin, and to which regions do they migrate? And what are the underlying mechanisms? Understanding the nature of these non-local startups and the effects of the quality of the entrepreneurial economy of their home region and destination region sheds new light on the mechanisms of (uneven) entrepreneurship-led economic development.

In section 2, we document the Entrepreneurial Ecosystem Index as an indicator of the quality of entrepreneurial economies and reveal its positive relationship with the prevalence of startups in Italian regions, correlating it against the number of innovative startups per capita and the PageRank of interregional flows of (nascent) founders, as a network measure of centrality. To obtain additional insights into the sequence of events in the geographic mobility of founders, we derive more fine-grained biographical data from the LinkedIn pages of a sample of non-local founders. In this way, we triangulate our analyses on the population of non-local startups understanding possible sequence patterns (considering education and working experiences of founders).

In section 3, we use a Poisson Pseudo Maximum Likelihood (PPML) gravity model to test the interregional flows of non-local founders and explain how the quality of entrepreneurial economies in the region of origin and destination affects these flows. To reinforce the findings, we rely on a novel machine learning application of the PPML postestimation techniques based on the Least Absolute Shrinkage and Selection Operator (LASSO) used to compute variables selection. In our main analysis, we focus on non-local innovative startups, and next to data on the region of origin of their founders, considering only startups entirely composed of members born outside the region.

Our analyses show that – controlling for a set of other variables – there is a positive relationship between the quality of the entrepreneurial economy and the presence of non-local startups. The quality of the entrepreneurial economy of both destination and origin regions is relevant to explain the mobility of (potential) founders between regions, respectively explained by attraction, creation, and supply mechanisms. This suggests that the quality of entrepreneurial economies not only stimulates the creation of local startups but also the attraction of non-local founders, and even the degree to which some regions act as suppliers of potential founders that move to other regions. Our findings have several implications for economic development policy, as discussed in section 4.

2. Data

2.1 Regional entrepreneurial economies

One cannot qualify the quality of entrepreneurial economies by just measuring their entrepreneurial outputs. This would lead to a tautology in which high-quality entrepreneurial economies are classified based on their outputs, and that regions with a high prevalence of entrepreneurship are classified as high-quality entrepreneurial economies. Such a tautology would be misleading for both scholars and policymakers in the identification process of the most relevant conditions for entrepreneurship-led economic development. Recently this gap has been addressed with the provision of entrepreneurial ecosystem metrics able to capture the granularity, interconnectedness, and systemic nature of entrepreneurial economies (Stam & Van der Ven, 2021; Leendertse et al., 2022). In this paper, we build on the entrepreneurial ecosystem approach (Stam 2015; Stam & Van der Ven 2021; Leendertse et al. 2021). We also operationalize ten key elements of entrepreneurial ecosystems around two layers: institutional arrangements (*Formal Institutions, Culture,* and *Networks*) and resource endowments (*Physical Infrastructure, Finance, Leadership, Talent, New Knowledge, Demand* and *Intermediate Service*). Considering the high variation between territories in terms of the quality of their entrepreneurial economy and the level of entrepreneurship outputs, we decided to adopt the European Union NUTS 3 level, analyzing 105 Italian provinces[1] (Iacobucci & Perugini, 2020; Iacobucci & Perugini, 2021). Metrics at this territorial level have been collected by different data sources: ISTAT, Italian Chambers of Commerce, EUROSTAT mainly from 2015 to 2019 (except for travel time to urban nodes reported for 2013). For five elements (*Formal Institutions*, *Culture*, *Networks*, *Leadership*, and *Intermediate Service*), we decide to take average values considering their structural character that make them less variable over time. In Table 1 we describe the data used to build each element.[2]

Table 1. The building blocks of the Entrepreneurial Ecosystem Index.

Element	**Measures**
Formal Institutions	The institutional quality index[3] developed by Nifo and Vecchione (2014) based on 5 groups of elementary indexes (evaluating corruption, governance, regulation, law enforcement, and social participation).
Culture	New firm formation rate (excluding the sole proprietorship firms), which reflects how common it is to create new business activity in a certain territory (Stam, 2013).
Networks	The number of Network Contracts between firms ("Rete Contratto"), established by Italian Law 33/2009 that represent an agreement tool that gives the possibility to firms to share one or more objectives and a common program, without creating a new legal entity (Leoncini et al., 2020). This policy tool has been mainly adopted by SMEs and therefore can be used as a proxy of connectedness degree within regions.
Physical Infrastructure	A composite indicator of three measures: a) travel time to urban centres, b) average speed in the NUTS 3 regional capital, and c) percentage of the population with a broadband subscription. The first two indicators have been intended as proxies of accessibility to measure the opportunity cost in terms of time, while the last one has been considered as a fundamental territorial prerequisite for businesses that highly rely on digital infrastructure to birth and prosper.
Finance	Exploiting information on the innovative source of financing (Venture capitalist, Project Finance and Crowdfunding) from the permanent

[1] Since the last territorial reclassification by ISTAT, Italian NUTS 3 level are currently represented by 111 regions. However, six regions belonging to Sardinia (Carbonia-Iglesias, Medio Campidano, Ogliastra, Olbia-Tempio and Sud Sardegna) and one region from Apulia (Barletta-Andria-Trani) have been excluded for their absence from many of the indicators used to build the Entrepreneurial Ecosystem index.

[2] For the data sources we used see Table A.1 in the Appendix.

[3] For more details on the composition of the Institutional Quality Index please visit the website of the authors Nifo and Vecchione: https://sites.google.com/site/institutionalqualityindex/home?authuser=0

	census by ISTAT (2018), we build a proxy for the local financial development (cf. Michelacci & Silva, 2007).
Leadership	Leadership is still scarcely measured at the territorial level, despite its increasing importance to understanding path creation dynamics (Grillitsch & Sotarauta, 2019). Thanks to the availability of the CORDIS database, which contains information on the Research and Innovation Program Horizon2020 projects and participants, we follow Leendertse et al. (2022) in considering Italian participants that cover the role of project coordinator. In this way we measure the capacity of the territories to coordinate and attract sources of innovation.
Talent	This is an important measure to understand the human capital that can nurture the ecosystem. We build a composite indicator, considering the level of education of people (percentage of graduates and Ph.D.) and the engagement of firms in training activities to acquire new skills and competencies.
New Knowledge	The contribution of the ecosystem in terms of research and innovation, measured with intramural expenditure on activities related to R&D.
Demand	The potential internal market of the ecosystems measured with GDP per capita
Intermediate Service	identifies the availability of business services that can nurture the activity of startups, sustaining them in consulting activities across different levels (e.g., legal, financial, strategical).[4] As a proxy, we use the percentage of firms in knowledge-intensive market services in line with Leendertse et al. (2022).

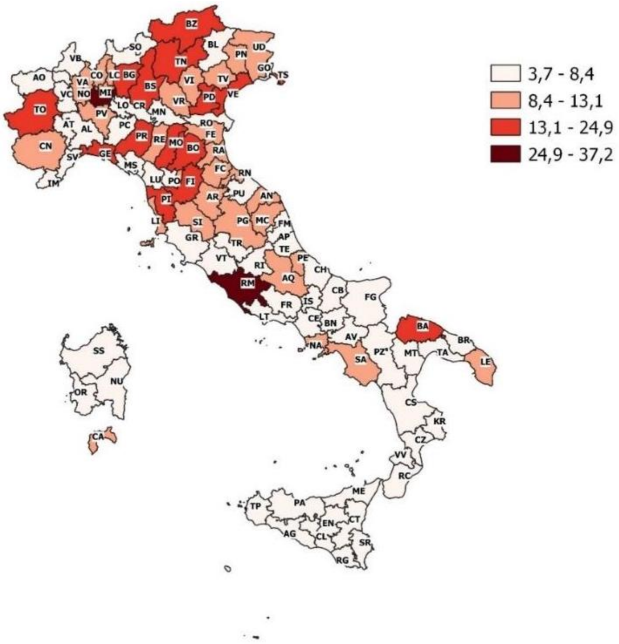
To provide a better interpretation of our results (more robust to outliers), we set a cap for the maximum values (four times the standard deviation). Then we standardized the composite indicators (Physical Infrastructure and Talent), normalized the ten elements by dividing them by their mean (to facilitate the inter-regional comparison) and finally we assembled the Entrepreneurial Ecosystem Index in an additive way.[5] The results show a high interconnectedness between elements and their aggregate measure (correlation is 0.42 on average, see the Appendix; cf. Leendertse et al. 2022). Looking at the performance of regions, the highest values of the Entrepreneurial Ecosystem Index are spatially concentrated in the

[4] In the original version of the Entrepreneurial Ecosystem Index (Leendertse et al., 2022) the number of incubators was also considered. However, we decided not to include this information in our analysis since most of the Italian NUTS 3 regions do not have any incubator (about 74%).

[5] For more details on the methodological steps related to normalization, standardization and capping see Leendertse et al. (2022).

north of Italy (with few exceptions like Rome and Bari), while the lowest values are in the most southern regions and the islands. Without surprise, Milan and Rome obtain the highest Entrepreneurial Ecosystem Index values (respectively 37.2 and 33.3) followed by Bologna (see Figure 1 and Table A.2 in the Appendix).

Figure 1. The Entrepreneurial Ecosystem Index of Italian regions[6]



2.2 Local and non-local innovative startups

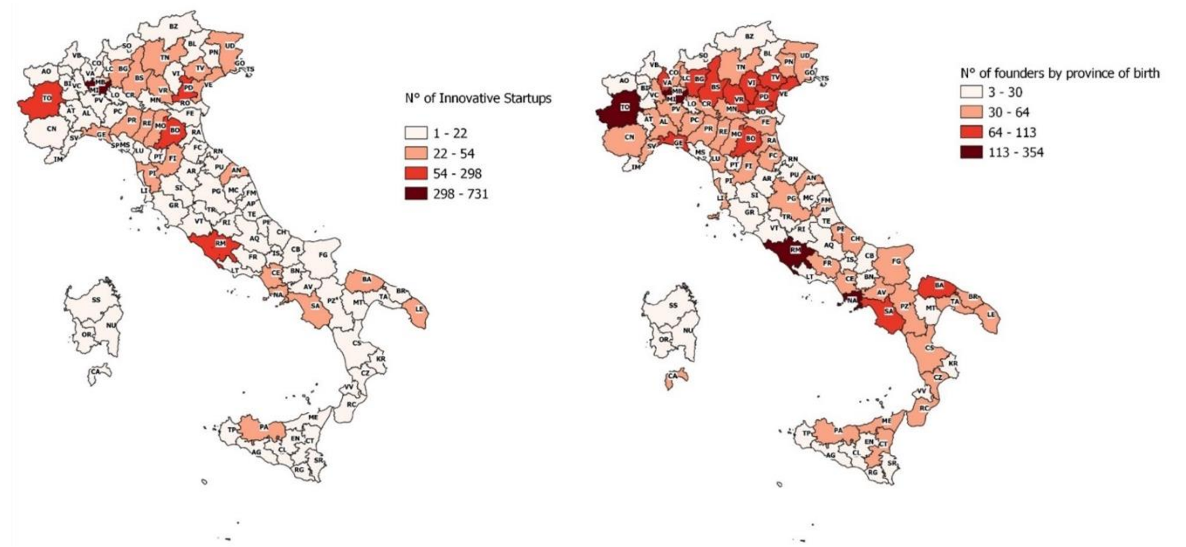
In our analysis, we focus on Innovative Startups delineated by a recent entrepreneurship policy in Italy, the so-called Start Up Act (Decree Law 179/2021).[7] This policy aims to enhance the prevalence of innovative startups, bridging the gap with other more innovative OECD countries (Biancalani et al., 2022; Grilli et al., 2022).

[6] The intervals for all the geographical maps have been computed using Jenks Natural Breaks, which have the advantage of reducing the variance within groups and, therefore, representing together the more similar elements.

[7] According to article 25 of Decree-Law no. 179/2012, the notion of Innovative Startup meet the following criteria (MISE, 2019): new company or incorporated for less than five years; headquarter in Italy; annual turnover lower than €5 billion; no profits' distribution; mission innovative oriented; no result of split-up or a company merger; able to satisfy at least one of the following innovation indicators: 1) expenses in R&D and innovation are at least 15% of the yearly costs; 2) 1/3 of employees with a Ph.D. or 2/3 with a master's degree are the holder, depositary, or licensee of a registered patent or software.

We selected from the Italian Chambers of Commerce database all the innovative startups founded from 2016 to 2019. From this initial sample (7,529 innovative startups), we search on ORBIS the innovative startups with available information on founders.[8] We found 4,838 firms with 9,737 founders. Since we are interested in the role of entrepreneurial economies as suppliers and attractors of startup founders we focus on non-local founders of innovative startups. Excluding founders born in the same NUTS 3 region where the startup was founded, we end up with 5,004 non-local founders (51.4% of all founders) of 2,779 Innovative Startups (57.5% of all firms). If we consider innovative startups composed of only non-local co-founder(s) the number lowers to 2,743 founders distributed in 1,660 firms (34.3% of all innovative startups).[9]

Figure 2. Innovative startups and non-local founders by Italian regions



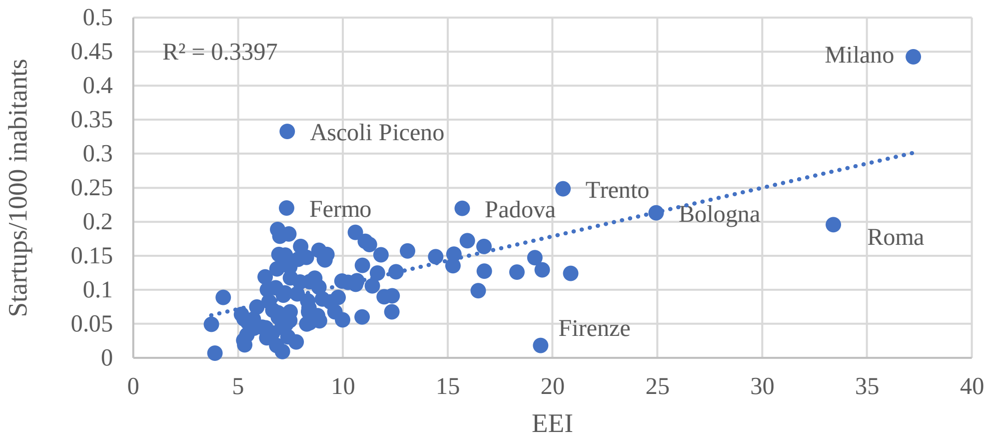
We test the robustness of the Entrepreneurial Ecosystem Index with the presence of Innovative Startups, based on the assumption that there should be positive relation between the quality of the entrepreneurial economy and this measure of productive entrepreneurship (Baumol 1990).

[8] In this paper, we consider founders the startup members who actively participate in the process of startup creation (e.g., Chief Executive Officer, Chief Technical Officer, Chief Operating Officer, Chief Scientific Officer, Chairman of the board, Majority shareholder), excluding external legal and commercial advisors.

[9] Within non local startups 1287 founders are below 45 years old (46% of the total) and 451 are represented by women (16 % of the total). About half of women (54 %) is below 45 years old.

The results show a strong and positive relationship between the Entrepreneurial Ecosystem Index and the prevalence of Innovative Startups in Italian regions: an increase of quality of entrepreneurial economies with 5 points on the Entrepreneurial Ecosystem Index tends to lead to an increase of 0.05 innovative startups per 1,000 inhabitants (see Figure 3). Negative outliers (with a much lower prevalence of innovative startups than expected) are Firenze and Roma, while Ascoli Piceno and Milano are positive outliers.

Figure 3. The correlation between the Entrepreneurial Ecosystem Index and Innovative Startups per capita of Italian regions



As the second step of descriptive analysis, to understand the relationship between the innovative startups and non-local founders represented in Figure 3, we analyze the interregional mobility network. More specifically, we consider the network in which nodes are Italian NUTS 3 regions, and directed links are drawn between the region where the non-local founder is born and the region where the startup is founded, setting the network directionality is in this order.[10] Links weights are the number of non-local startups moving from the region of origin to the one where the startup has been founded.

[10] We only include founders that are born in Italy, and thus exclude international immigrant entrepreneurs, which are of central importance in other studies on entrepreneurship and migration (Azoulay, 2022; Conti & Guzman, 2022).

As a measure of network centrality, we adopted the PageRank algorithm, useful to uncover influential nodes beyond their directed connection. The choice to rely on PageRank is due to its suitability to describe directed mobility networks that involve more than one possible movement.

Originally the PageRank has been used to identify the importance of webpages, according to the links received from other webpages (Page et al., 1999). In a nutshell, it assumes persons as "random surfers" that re-iterate the action to search different web pages. By analogy in this study, it measures the probability of a "random surfer" of a given agent, born in a region *j*, and her/his decision to move to a different region to create a startup. Technically, it is an iterative algorithm that assigns to each node a probability to be connected with the other nodes of a network, summing all the incoming links of a node *j* and dividing by the number of outgoing links of nodes *i* to n and represents a variant of eigenvector centrality for directed networks (Page et al., 1999). The formula of Page Rank is reported hereinafter:

$$PR(prov_j) = (1 - d) + d(PR(prov\ _i)/C(prov\ _i) \cdots PR(prov\ _n)/C(prov\ _n)) \qquad (1)$$

Where PR is the Page Rank, d is the dumping factor[11] and C represents the outgoing connections (interregional movement of non-local founders). The sum of all the PageRanks of the network (excluding the self-ties relationship) must be equal to one. The results show Milan, Rome, Turin and Bologna in the first places and Crotone, Imperia and Nuoro occupying the last positions of the ranking (see table A.2 in the Appendix).

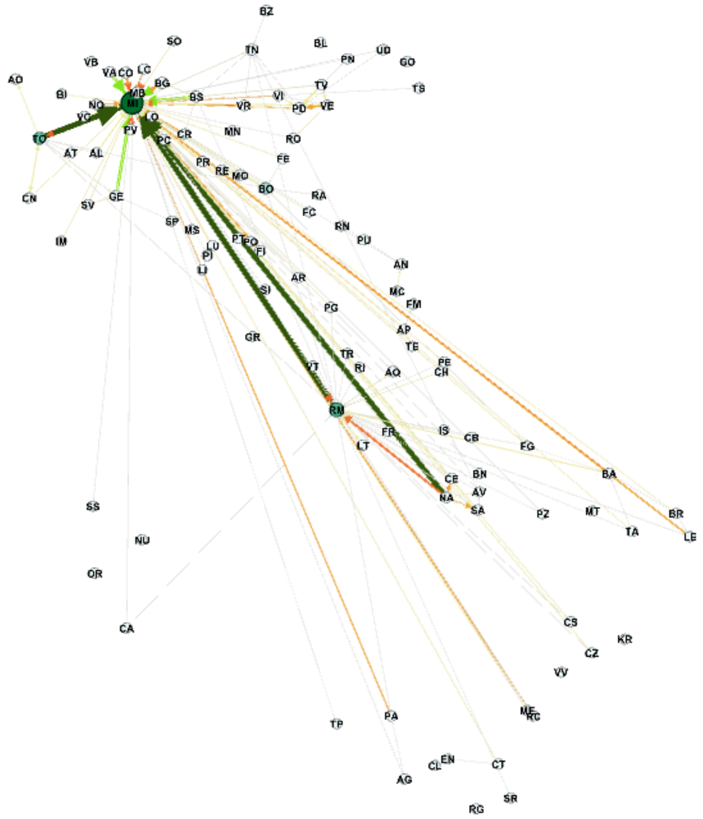
[11] The dumping factor represents the probability, at each step of the computation, that the random walk continues. The optimal value is usually set at 0.85 (Page et al., 1999). We calculated PageRank centrality using Gephi, a specialized software for network analysis.

Figure 4 shows the inter-regional mobility network of non-local founders. We embedded Page Rank estimation and the networks of non-local startup founders in a geographical layout using the Gephi software. The size of the node is proportional to its PageRank centrality and the thickness of links varies according to their intensity. Dark Green links represent the dyadic connections with more intense mobility (90-105), light green with medium-high (45-60), Orange with a medium intensity (21-40) and yellow with low intensity (10-20 links).

Milan stands out as the most central region, attracting the most non-local founders from many different regions (even from other large regions like Rome and Turin). Also, it is worth noticing that a consistent part of Milan's outgoing links is directed towards high-quality entrepreneurial economies in its vicinity such as Monza-Brianza and Brescia.

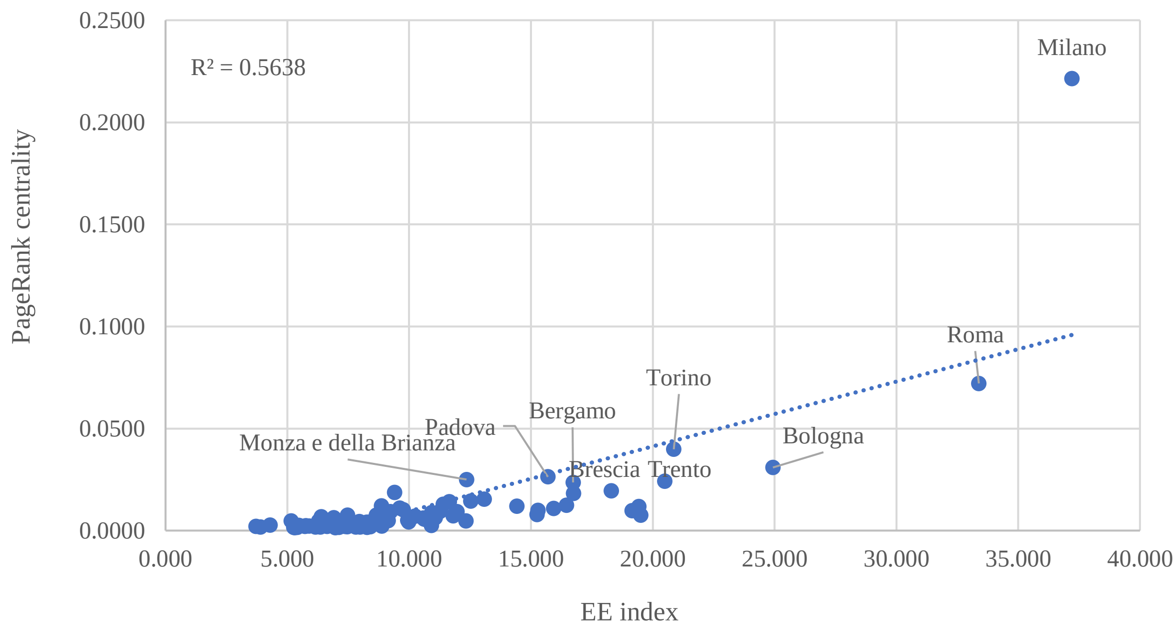
The Entrepreneurial Ecosystem Index turns out to be strongly correlated with the PageRank centrality of NUTS 3 regions in the network of movements of nonlocal founders between territories (see Figure 5). This is preliminary evidence that high-quality entrepreneurial economies might attract more non-local founders.

Figure 4. The inter-regional mobility network of non-local founders[12]



[12] The size of nodes is proportional to PageRank centrality. Link thickness is proportional to the number of non-local founders' movements. Dyadic links with less than 10 observations have been excluded due to visualization reasons.

Figure 5. Correlation Page Rank Centrality and Entrepreneurial Ecosystem Index



To further investigate this potential relationship, in Table 2 we divide NUTS 3 regions into four quartiles by Entrepreneurial Ecosystem Index values.

Table 2. Founders' flows between regions divided in four quantiles by Entrepreneurial Ecosystem Index values.

Origin (birth)/destination	**High EEI**	**Medium-high EEI**	**Medium-low EEI**	**Low EEI**	**Total**
High EEI	1865	362	141	137	2505
Medium-high EEI	860	134	64	35	1093
Medium-low EEI	486	99	68	43	696
Low EEI	519	84	44	63	710
Total	3730	679	317	278	5004

Most of non-local founders (1,865) move from high quality entrepreneurial economies to other high-quality entrepreneurial economies. This is largely driven by the large number of moves from Rome to Milan. Second, there is hardly any movement between the low, medium-low, and medium-high quality entrepreneurial economies. Third, high-quality entrepreneurial economies enjoy a positive net balance since incoming movements (3,730) are much more than outgoing movements to lower quality ecosystems (2,505). Lastly, there is still a considerable group of founders that have moved from high-quality entrepreneurial economies to medium-high (e.g., Milan to Brescia) and to medium-low quality entrepreneurial economies (e.g., Milan to Lecco).

2.3 The domestic mobility of startup founders: insights from LinkedIn

Our dataset contains information about the founder's birthplace and the location of the creation of the startup but does not provide information on potential movements between those two events. In recent ethnographic research, Sontag (2018, p. 149) identifies six migration strategies: deliberate step-by-step progress; careful planning; seizing opportunities as they come; pushing the boundary; adjusting to the situation; family compromise. Complex migration patterns, including transmigration and circular migration, are described, with many work and personal arrangements. Migration patterns present some "shadow" cases, such for instance:

- daily commuting or local mobility: the founder moves in a neighboring region to start up a business, keeping connections with her place of origin and supporting a lower cost for relocation in case she remains in the same region;
- management reasons. For managerial (or strategical) purposes, legal and operational headquarters can be in different places, but founders can continue to work in the region of origin;
- people formally born in a territory but who lived, studied, and worked in the location where they founded the startup;
- serial entrepreneurs that participate in the founding process of more than one startup.

To get insight into potential intermediate moves to other regions and the reasons for moving, we analyze a subsample of founders, using LinkedIn data on their biography. Social media have been recently used to track the mobility of entrepreneurs (Butler et al., 2020). However, especially in countries like Italy, where the usage of new social media is still limited, LinkedIn offers only biased and partial coverage (Credit et al., 2018). Only one-third of the founders we searched have a profile on LinkedIn.

We manually scraped LinkedIn profiles of non-local founders until reaching theoretical saturation (Patton, 1990) to derive possible mobility patterns. Manual search is less error-prone and feasible given the general restriction to scrape and crawl LinkedIn data using APIs. Figure 7 summarizes the possible identified patterns with a hypothetical decision tree. As evident, creating a startup can originate with different sequences and steps (cf. Stam 2007). Here we identify 14 paths that lead to startup creation only considering the declination of two elements (education and work experience). From the result of the sequence analyses displayed in Figure 6, we can see how 39% of the scraped profiles followed the path in the left part of the tree (education in the same place of birth) and 61% in the right one (education in a different place). These paths show that the decision to start a (non-local) business is most often made after the home-region is left for educational or career reasons. Only 5% of the founders of non-local startups seems to have left the region of origin for setting up a business in another region. This confirms the stylized fact in the geography of entrepreneurship that most entrepreneurs start a business in the region in which they live (have been educated) and/or work (Stam, 2007). There are multiple paths that precede the creation of a non-local startup, and entrepreneurial careers can vary according to age and career path. Young founders are more likely to start a non-local business after they have moved due to educational reasons, while middle-aged founders are likely to have gone through several work career stages, possibly in different regions, before they start a business.

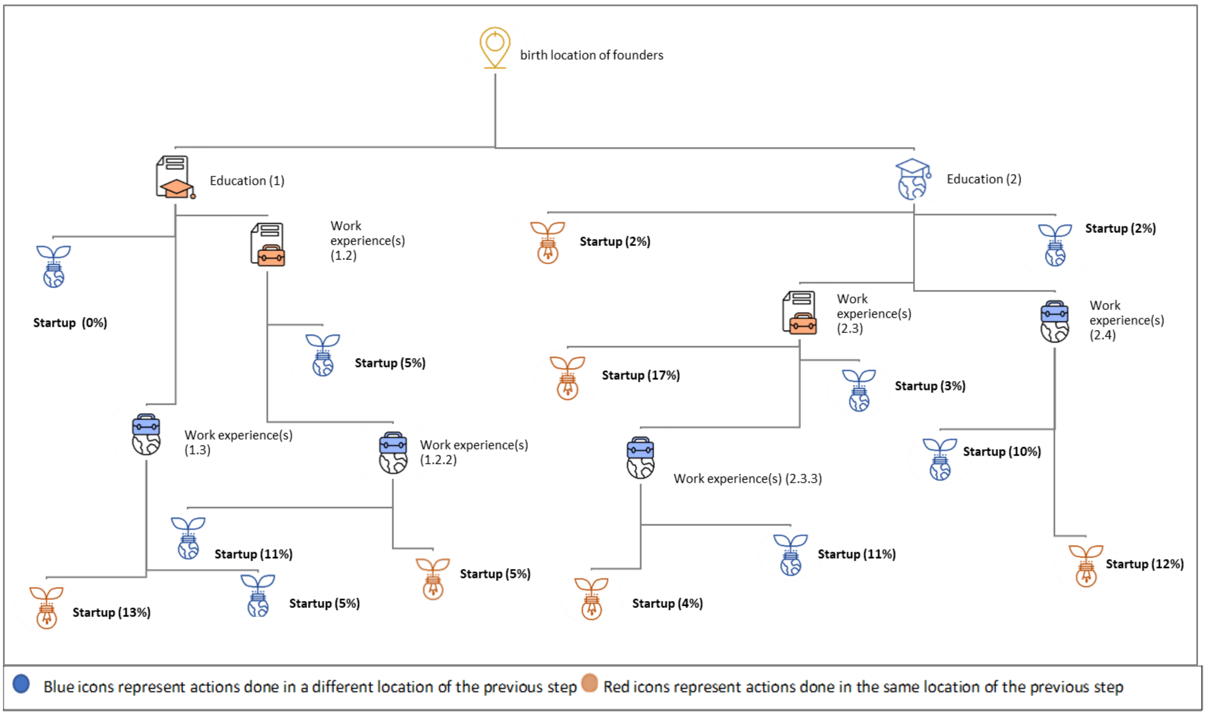
As an illustration, two examples of well-known American entrepreneurs can be given: Marc Andreessen (co-founder of Netscape) and Michael Mauldin (co-founder of Lycos). Marc

Andreessen was born in Iowa and raised in Wisconsin, and then moved to Illinois for his education (Education 2). After his education he move to California for work (Work experience 2.4) and stayed in California to start Mosaic (later rebranded as Netscape) (Startup in same location as Work experience, with nonlocal founder: 12% of the Italian innovative startups in our sample). Michael Mauldin was born and raised in Texas, where he also pursued his bachelor and master, but he moved to Pennsylvania for a PhD (Education 2), where he also later on started Lycos (Startup in same location as Education, with nonlocal founder: 2% of the Italian innovative startups in our sample). The fact that Lycos moved to Massachusetts within 2 years after creation does not count, because we only consider location of foundation. Such interregional moves of young firms are extremely rare (see Stam 2007).

We also provide two examples from our LinkedIn sample, to illustrate the possible sequences. The first one was born and raised in Sassari (Sardinia) and moved to Pisa (Tuscany) to pursue a university degree in physics. Then He moved to Milan for work (R&D department) (work experience 2.4 of figure 6), after to Berlin and San Francisco (covering similar role) and after 15 years he founded his company in the region of Monza-Brianza (specialized in Graphene Photonics).

The second one was born in Taranto and pursued a university degree in Modena in pharmacy. Then she moved to Venice due to new work experiences. After 3 years of experience as researcher she founded her company in the region of Treviso (specialized in food health and nutritional analysis).

Figure 6. Sequence analysis by stages - considering education and work experience



Source: Authors' elaboration using resources from Flaticon.com. Base Icons taken from: Ultimatearm, Freepik and Eucalyp.

2.4 Control Variables

Apart from the effects of the Entrepreneurial Ecosystem Index, we have also controlled for some factors that have been identified earlier as possible geographical mediating effects of startup creation and socio-economic measures of distances between two regions.

First, population density (population by km^2) is considered as a fundamental socio-economic mediating factor because of its capacity to capture supply-push and demand-pull factors. Population density serves also as a proxy for agglomeration economies (Piras, 2017; Capozza et al., 2018).

Second, we control for geographical distances between NUTS 3 regions, calculating distances between NUTS 3 regions centroids as the matrix of the driving hours, using the free API provided by https://openrouteservice.org/ and geocoding with R (for a similar approach see Cavallo et al., 2020b).

Third, we include a dummy variable of contiguity between NUTS 3 regions to check if to share a border can reduce movements between regions, as described in the previous section.

Fourth, we include GDP per capita, a standard measure to check the power of attraction between two countries, region and cities in gravity models (Tinbergen, 1962). Whenever we include GDP per capita as a control in a regression we omit the GDP component from the Entrepreneurial Ecosystem Index.

Fifth, we include number of STEM students every 1000 inhabitants to check for the presence of educational clusters in scientific and technical subjects, usually a very important “catchment effect” for founders of innovative startups (see our LinkedIn analysis and Anelli et al., 2019).

Sixth, we include a NUTS 2 dummy variable to check for possible institutional and spatial patterns (Bonaccorsi et al., 2014).

Table 3 reports the descriptive statistics for the variables we consider in our model. The correlation table is available in the Appendix.

Table 3. Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Non-local founders	10,920	.251	1.482	0	62
Pop. density (ln)	10,920	1	1.405	.141	9.556
Driving time (ln)	10,920	1.681	.766	-.916	3.109
NUTS 3 Contiguity	10,920	.043	.203	0	1
GDP p.c. (ln)	10,920	-.038	.278	-.567	.72
STEM/pop.	10,920	.04	.041	0	.234
EEI (ln)	10,920	2.084	.445	1.143	3.56
EEI (ln, w/o demand)	10,920	2.201	.422	1.313	3.616

2.5 Econometric Strategy

To measure inter-regional geographical mobility of (nascent) entrepreneurs, we relied on an econometric strategy widely used in migration studies to estimate bilateral flows. At this stage, we included data on the regions at origin (the birthplace of the founder) and destination (the startup's location). Given that we trace the movements of non-local founders between Italian NUTS 3 regions, we decided to rely on a specific stream of the literature that analyses domestic migration flows (Biagi et al., 2011; Poot et al., 2016; Piras, 2017). As in the case of trade and finance, the methodological starting point of this literature is a gravity model. This model originated in physics,[13] to explain the attraction between two objects (proportional to their mass and inversely proportional to their distance) has been applied to explain the different impact of the origin versus destination "forces" in determining the migration flows. The empirical specification of Gravity Equation applied to entrepreneurial decision to choose a region different from the one of birth can be estimated in the following terms

$$ln\left(SF_{ij}\right) = \beta_0 + \beta_1\, ln\, ln\, (Fd_i) + \beta_2\, ln\, ln\left(Fo_j\right) - \beta_3\, ln\left(R_{ij}\right) + \eta_{ij} \qquad (2)$$

Where SF_{ij} are the founders flows from the region of birth i to the place of destination j, function of: Fd_i, the "destination factors" (in our case the EEI of destination with the addition of the population density and GDP per capita at destination as control), Fo_j "origin factors" (in our case the EEI of origin, controlling for the population density and GDP per capita) and R_{ij} represent the vector of distances (in our case driving hours between regions and regional contiguity). C represents the constant term, $\beta_1, \beta_2, \beta_3$ the elasticities terms linking destination and origin regions and η_{ij} represents the error term.

The log transformation of the gravity equation shows two major weaknesses (Silva & Tenreyro, 2006):

1) it does not allow to account for zeros (i.e., absence of movements between two regions *i* and *j*);
2) in the presence of heteroskedasticity the errors will be correlated with the covariates, producing thus inconsistent estimations.

[13] The application of physics in social sciences can be attributed to John Q. Stewart with the establishment of the 'social physics' school (Stewart, 1950).

A standard approach to cope with missing movements had been to drop zeros from the estimation, estimating the log-linear form with OLS. However, consider the case of two regions (*i*,*j*), both with low "gravitational" force (low Entrepreneurial Ecosystem Index) and a resulting value of zero occurs (that is, no entrepreneur decides to locate her/his startup from region *i* to *j*). This information can be seen as coherent with the structural characters of the regions and therefore essential to keep to avoid sample selection bias.[14]

Hence, to maintain the integrity of the sample and avoid heteroskedasticity issues, we follow Silva and Tenreyro (2006) approach, which consists in estimating the gravity equations in a multiplicative form adopting a PPML model[15]. PPML is a count model particularly indicated for zero-inflated models and consistent in the presence of fixed effects, a desirable property to test the impact entrepreneurial ecosystems index (Fally, 2015). In the light of this, the specification (and for the properties of log and exp) equation 2 becomes

$$SF_{ij} = exp[\beta_0 + \beta_1 \, ln \, ln \, (Fd_i) + \beta_2 \, ln \, ln \, (Fo_j) - \beta_3 \, ln \, (R_{ij})]\eta_{ij} \quad (3)$$

Following Silva and Tenreyro (2011), we rescale our dependent variables to avoid converge issues of PPML. We consider different model versions with different definitions of non-local startup founders moving from i to j (SF_{ij}). To avoid spurious effects through the contribution of local founders to startups, in our baseline model, we consider only non-local startups. Therefore, we count only the movements of non-local founders who gave birth to non-local startups. We test our model for a broader set of non-local founders as a robustness check. In one alternative setting, we consider all non-local founders. This is equivalent to considering the contribution to most Italian startups, which have at least one non-local founder. As an intermediate case, we also consider movements to give birth to a startup in which the majority

[14] In our case, considering the possible set of relations (105x104=10,920) the presence of zeros is dominant and excluding them would reduce the sample to 1,113 observations.

[15] The term "Pseudo" is attributed to the fact that PPML works regardless the data distribution (Silva & Tenreyro, 2006).

of the founders are non-local. To further investigate the importance of the Entrepreneurial Ecosystem Index in our gravity model, we adopt the LASSO PPML postestimation techniques (Breinlich et al., 2021). Finally, we run separate regressions by age (non-local founders above and below 45 years old) and gender (male and female non-local founders).

3. Results

Our gravity model tests the role of the Entrepreneurial Ecosystem Index to explain the mobility pattern of non-local founders of startups in Italy. In our baseline setting, we focus on non-local startups, i.e., startup companies whose founders are all born in different NUTS 3 regions than the one in which the startup is founded. Table 4 summarizes our results. Models 2 and 3 strongly support the claim that the quality of the entrepreneurial economy is essential to attract non-local startups (attraction effect). We also find a positive and significant role of the quality of the ecosystem of origin of the founders (supply effect). This effect is less pronounced than for the quality of the entrepreneurial economy where the startup is founded, confirming the escalator mechanism described in the introduction.

Table 4. Gravity model of interregional flows of non-local innovative startups (all founders are non-local) between Italian regions

	Non-local startups (all founders are non-local)		
	(1)	(2)	(3)
	b/se	b/se	b/se
EEI (ln), foundation		**2.4448***	**1.8029***
		(0.0922)	(0.1351)
EEI (ln), origin		**1.6418***	**1.4249***
		(0.1214)	(0.1787)
Pop. density (ln), foundation	0.4319***	0.1056***	0.0927***
	(0.0296)	(0.0195)	(0.0199)
Pop. density (ln), origin	0.2527***	0.0912*	0.0834*
	(0.0348)	(0.0359)	(0.0376)
Driving time (ln)	-0.8959***	-0.9583***	-0.9616***
	(0.0980)	(0.0777)	(0.0770)
Contiguity, NUTS 3	0.7240***	0.8052***	0.8081***
	(0.1807)	(0.1327)	(0.1344)
GDP p.c. (ln), foundation			1.2849***
			(0.3575)
GDP p.c. (ln), origin			0.7344
			(0.5050)
STEM/pop., foundation			2.2450*
			(0.9988)
STEM/pop., origin			-2.5260**
			(0.9010)
Constant	-0.4485	-8.2167***	-5.6698***
	(0.3878)	(0.4489)	(0.5969)
NUTS 2 dummy	YES	YES	YES
N	10,920	10,920	10,920
R^2	0.202	0.559	0.561
Pseudo Log-likelihood	-5372.9567	-3970.2291	-3951.1587
VIF max	1.23	1.39	3.18

The robustness of the Entrepreneurial Ecosystem Index is confirmed in model 3 with the control variables GDP[16] and the presence of STEM graduates that could in some way shadow the effect of Index.

As expected, density, as a proxy of urban agglomerations, plays a positive and significant role in origin and destination regions, with a lower significance level for region of origin in the full version of the model (3).

The negative values of driving times highlight the geographical impact of distance, explained also by the positive impact of NUTS 3 regional contiguity across all the 3 models.

[16] Entrepreneurial Ecosystem Index in model 3, 5, 7,9, 11 ,13, 15 is tested without demand component that is equal to GDP p.c. regional value.

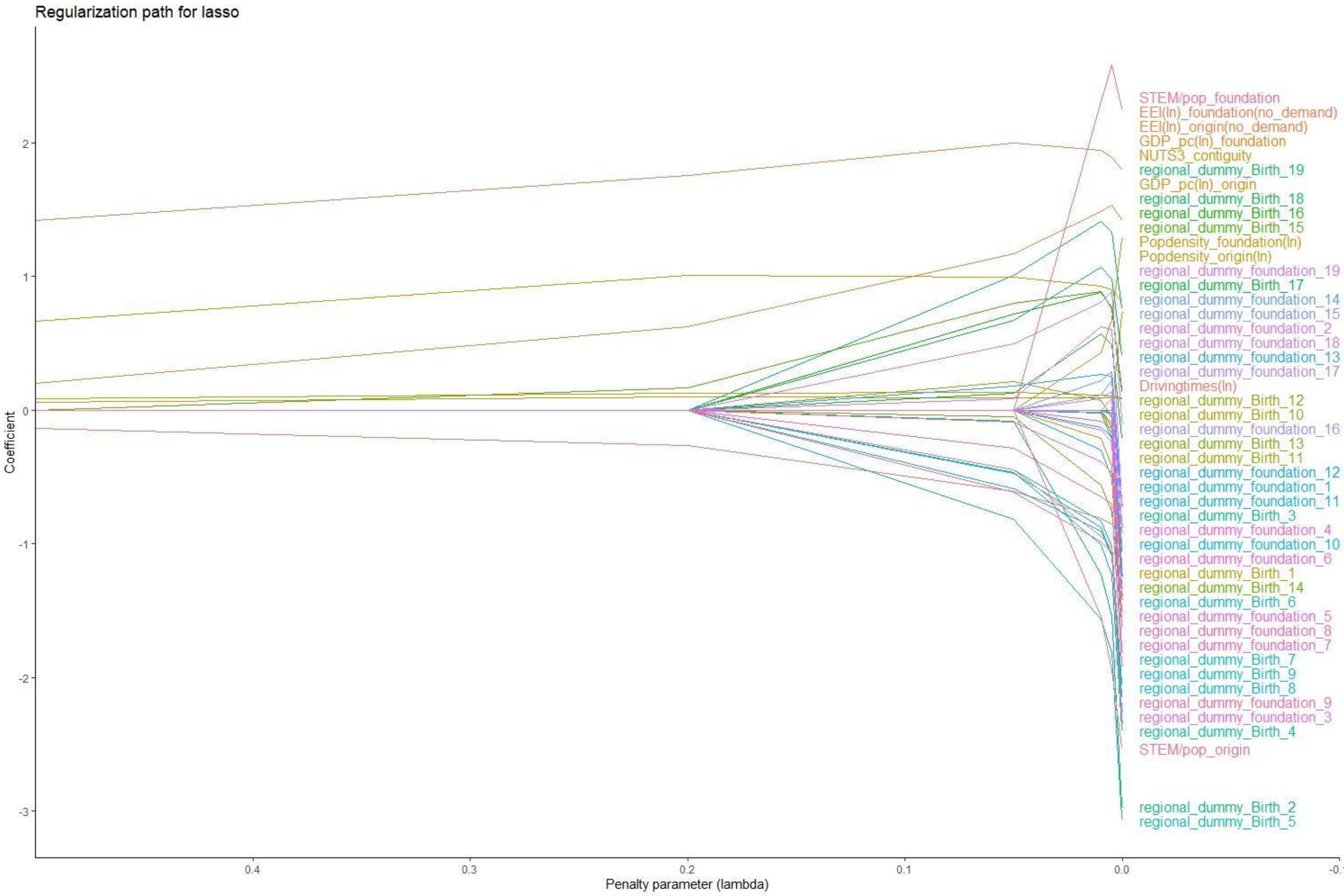
GDP per capita is significant in the destination region, working as a proxy of regional wealth, but is not significant as a push mechanism, a fact that could help to interpret flows of founders as less dependent on the general economic conditions.

The effect of STEM is interesting for the negative push effect registered, in other words scarcity of graduated in STEM subjects can be interpreted as a factor that motivate people to move from origin regions towards regions equipped with technic and scientific universities (this is valid especially for young people, see model 9 in Table 6).

We check for the presence of multicollinearity issues, performing a VIF test and recording values from 1.23 to 3.18, well below the threshold recommended by the literature (10 and 5 in the most parsimonious settings).

To analyze the role of the Entrepreneurial Ecosystem Index in our gravity model, we adopt the LASSO PPML postestimation techniques following the approach developed by Breinlich et al. (2021). LASSO is used in model selection to identify λ values at which variables are excluded from the model. For each λ value the set of non-zero coefficients are identified: the higher is λ the smaller is the set of variables with non-zero coefficient. With a long list of variables (as for our case for the presence of NUTS 2 dummies), we typically see some variables removed at a certain threshold λ.

Figure 7. PPML LASSO of model 3



The results (see figure 7 and Table A.4 in the appendix) confirm that Entrepreneurial Ecosystem Index of destination and origin are both within the threshold of non-zero coefficients (considering λ cross-validation rule[17]), respectively occupying 1 and 3 position in the coefficient rank. Therefore, we can reinforce the thesis that the quality of the ecosystems is an essential component of the gravity model of non-local founder movements.

---

[17] For model 3: λ Cross-validation rule that minimizes RMSE = 0.3 with seed set at the value of 5000 (see table A.5 in the appendix). We test the robustness of the findings also with a stricter variant than the Cross Validation method, the plug-in method (Belloni et al., 2012), which automatically computes the optimal penalty threshold (for details see the penppml R package developed by Breinlich and colleagues, available here). Also in this case the Entrepreneurial Ecosystems index of destination and origin are both included within the lists of non-zero coefficients.

To corroborate our findings, in Table 5, we report the gravity model results with alternative definitions of non-local founders. Models 4 and 5 consider all non-local founders' movements, including also founders who give birth to startups with local founders. Models 6 and 7 instead limit the analysis to the case in which the majority of founders is not local. All the baseline model results concerning the quality of the entrepreneurial economies are confirmed. Among the control variables, it is worth noticing that when we consider all the non-local founders in model 5, population density and STEM at origin are no more significant. In this model, only controls at destination remain significant, whereas the role of the quality of the entrepreneurial economy in the region of origin stays positive and largely significant.

Table 5. Robustness check: alternative definitions of non-local innovative startup founders' mobility between Italian regions

	at least 1 founder non-local		majority of founders non-local	
	(4)	(5)	(6)	(7)
	b/se	b/se	b/se	b/se
EEI (ln), foundation	**2.6188***	**1.9497***	**2.5655***	**1.8901***
	(0.0876)	(0.1125)	(0.0887)	(0.1166)
EEI (ln), origin	**1.6988***	**1.3919***	**1.6795***	**1.3985***
	(0.1207)	(0.1693)	(0.1197)	(0.1663)
Pop. density (ln), foundation	0.0829***	0.0679***	0.0900***	0.0759***
	(0.0180)	(0.0182)	(0.0184)	(0.0187)
Pop. density (ln), origin	0.0694	0.0610	0.0708	0.0630
	(0.0379)	(0.0400)	(0.0377)	(0.0397)
Driving time (ln)	-0.9722***	-0.9786***	-0.9809***	-0.9871***
	(0.0716)	(0.0706)	(0.0733)	(0.0723)
Contiguity, NUTS 3	0.8278***	0.8272***	0.8063***	0.8048***
	(0.1238)	(0.1254)	(0.1255)	(0.1266)
GDP p.c. (ln), foundation		1.3342***		1.3215***
		(0.3278)		(0.3300)
GDP p.c. (ln), origin		0.9044		0.8368
		(0.4919)		(0.4900)
STEM/pop., foundation		2.5093**		2.7487***
		(0.7836)		(0.8266)
STEM/pop., origin		-1.5776		-1.7058*
		(0.8076)		(0.8033)
Constant	-8.2685***	-5.4973***	-8.1884***	-5.4916***
	(0.4303)	(0.5341)	(0.4347)	(0.5450)
NUTS 2 dummy	YES	YES	YES	YES
N	10,920	10,920	10,920	10,920
r2	0.593	0.598	0.597	0.603
Pseudo Log-likelihood	-5660.4268	-5627.6903	-5397.2146	-5365.515
VIF max	1.39	3.18	1.39	3.18

The role played by the quality of entrepreneurial economies in the region of origin and destination might be different for young and senior non-founders. To see how the effects change in different cohorts of non-local founders, we split them into two balanced groups by age (below and above 45 years old).[18] Results in Table 6 show that, as expected, the quality of the entrepreneurial economy in the region of origin is more relevant for senior non-local founders. This is generally true also for other characteristics of the region of origin of non-local founders, whereas the effect of distance is less pronounced for seniors.

When we distinguish non-local founders by gender (Table 7), we find that the attractive force of the quality of the entrepreneurial economy where startup companies are founded is stronger for women. This is interesting since female entrepreneurs tend to stay closer to their birthplaces, as highlighted by the stronger regional contiguity and distance effects in models 14 and 15. Also, we find that GDP per capita and population density at destination are not significant for females. All in all, we can conclude that the quality of entrepreneurial economies is particularly relevant in motivating female entrepreneurs to move out of their region of origin and establish innovative startups.

To furtherly improve the reliability of our results we test the models in other two versions. First, we exclude intra NUTS 2 regional mobility (between NUTS 3 regions within the same NUTS 2) to reduce the probability of impact of daily commuting shadow cases. Second, we exclude the Milan region, as the main attractor of non-local startups. In both cases the results of the model still hold, showing positive and significant effect of both ecosystem of origin and destination.

As a robustness check, to separate general attractiveness from “Entrepreneurial Economy attractiveness”, we consider interregional migration flows of the general population as another control variables for the interregional migration flows of non-local founders (see Table A.6 in

[18] Results are robust when alternative thresholds are used (i.e., 40 to 50 years old).

the appendix). As expected, interregional migration flows have a strong and positive effect on the patterns of mobility of non-local founders in all the models we considered. However, the positive effect of the quality of the entrepreneurial economy at origin and destination still hold. Again, we find that their role is stronger for women and, in relative terms, the quality of the entrepreneurial economy at origin (destination) is more important for senior (young) founders.[19]

Table 6. Split regressions for young (up to 45 years old) and senior (above 45) non-local founders

	YOUNG		SENIOR	
	(8)	(9)	(10)	(11)
	b/se	b/se	b/se	b/se
EEI (ln), foundation	**2.5705*****	**1.8580*****	**2.3367*****	**1.7440*****
	(0.1119)	(0.1714)	(0.1134)	(0.1689)
EEI (ln), origin	**1.5188*****	**1.3166*****	**1.7371*****	**1.5233*****
	(0.1249)	(0.1987)	(0.1383)	(0.2118)
Pop. density (ln), foundation	0.1024***	0.0946***	0.1080***	0.0914***
	(0.0230)	(0.0250)	(0.0229)	(0.0232)
Pop. density (ln), origin	0.0785*	0.0731*	0.1019**	0.0918*
	(0.0346)	(0.0359)	(0.0392)	(0.0413)
Driving time (ln)	-1.0537***	-1.0611***	-0.9046***	-0.9039***
	(0.0984)	(0.0984)	(0.0888)	(0.0880)
Contiguity, NUTS 3	0.8275***	0.8209***	0.7624***	0.7766***
	(0.1474)	(0.1474)	(0.1618)	(0.1649)
GDP p.c. (ln), foundation		1.2049**		1.3550**
		(0.4576)		(0.4367)
GDP p.c. (ln), origin		0.5919		0.8244
		(0.5309)		(0.6137)
STEM/pop., foundation		4.5826**		0.6075
		(1.4412)		(1.1975)
STEM/pop., origin		-1.6606		-3.3382**
		(1.1568)		(1.0687)
Constant	-8.9670***	-6.6061***	-8.7827***	-6.1271***
	(0.5803)	(0.7663)	(0.5016)	(0.7050)
NUTS 2 dummy	YES	YES	YES	YES
N	10920	10920	10920	10920
r2	0.531	0.534	0.489	0.491
Pseudo Log-likelihood	-2328.8908	-2316.5884	-2679.541	-2668.1684
VIF max	1.39	3.18	1.39	3.18

[19] In line with the systems logic of the entrepreneurial ecosystem approach we have used the composite Entrepreneurial Ecosystem Index to capture the quality of entrepreneurial economies. This might come at the expense of more isolated analyses of the importance of individual elements. We use machine learning approaches to better analyze the effects of individual entrepreneurial ecosystem elements. In the appendix (table A.7 and A.8) we report the PPML LASSO with the entrepreneurial ecosystem decomposed into 10 elements (20 in total considering origin and destination). This exercise suggests that at $\lambda = 0.1$ (the value that minimizes RMSE) some entrepreneurial ecosystem elements matter more than others (INTERMEDIATE SERVICES_foundation, NETWORKS_birth, NETWORKS_foundation, LEADERSHIP_foundation).

Table 7. Split regressions: male and female non-local founders

	MALE		FEMALE	
	(12)	(13)	(14)	(15)
	b/se	b/se	b/se	b/se
EEI (ln), foundation	**2.4042*** **	**1.7517*** **	**2.6654*** **	**2.0809*** **
	(0.0961)	(0.1520)	(0.1836)	(0.2512)
EEI (ln), origin	**1.6271*** **	**1.4041*** **	**1.7296*** **	**1.5314*** **
	(0.1297)	(0.1882)	(0.1547)	(0.2661)
Pop. density (ln), foundation	0.1096***	0.0954***	0.0858*	0.0804
	(0.0202)	(0.0206)	(0.0392)	(0.0431)
Pop. density (ln), origin	0.0874*	0.0792	0.1073**	0.1011**
	(0.0390)	(0.0410)	(0.0341)	(0.0350)
Driving time (ln)	-0.9192***	-0.9224***	-1.1524***	-1.1574***
	(0.0836)	(0.0825)	(0.1261)	(0.1272)
Contiguity, NUTS 3	0.7825***	0.7861***	0.9165***	0.9154***
	(0.1439)	(0.1454)	(0.1956)	(0.1974)
GDP p.c. (ln), foundation		1.3415***		0.9715
		(0.3912)		(0.6589)
GDP p.c. (ln), origin		0.7463		0.7314
		(0.5458)		(0.7023)
STEM/pop., foundation		2.0714		3.3077
		(1.1308)		(1.8692)
STEM/pop., origin		-2.4095*		-3.2933
		(0.9484)		(2.1421)
Constant	-8.1655***	-5.5697***	-12.4302***	-10.1071***
	(0.4714)	(0.6367)	(1.1898)	(1.3790)
NUTS 2 dummy	YES	YES	YES	YES
N	10920	10920	10610	10610
r2	0.511	0.514	0.483	0.480
Pseudo Log-likelihood	-3596.1208	-3580.1995	-1091.2005	-1087.1943
VIF max	1.39	3.18	1.39	3.18

4. Conclusions and Policy Implications

To what extent and how does the quality of the entrepreneurial economies of origin and destination explain the prevalence of non-local startups? We first tested the assumption that the quality of entrepreneurial economies is strongly related to subsequent startup activity, with new data of Italian regions. After improving an entrepreneurial ecosystem composite index to measure the quality of local entrepreneurial economies (at NUTS 3 level) in Italy, we test it against startups per capita and network centrality (Page Rank). This assumption was confirmed and was used as a starting point to test attraction and supply mechanisms concerning the quality of destination and origin entrepreneurial economy in relation to the prevalence of non-local startups. The majority of the startups turned out to have at least one non-local founder, indicating that even in Italy there is no strong local bias of founders.

The results of the econometric and machine learning analyses point out positive effects of the quality of both destination and origin entrepreneurial economies of non-local founders against a relevant set of control variables (density, geographical distances, GDP, and STEM graduates). Both attraction and supply mechanisms seem to be at work, impacting the economy of destination and origin. Founders reveal to come from relatively high-quality entrepreneurial economies, setting up new firms in even higher quality entrepreneurial economies. Regions with a low-quality entrepreneurial ecosystem not only have low probability to "create" local innovative startups, they are also less likely to "supply" nascent entrepreneurs to higher quality entrepreneurial economies.

This paper provides important policy insights for entrepreneurship-led economic development. The recent economics of entrepreneurship literature has shown that policy for an entrepreneurial economy is more likely to be effective than entrepreneurship policy, as policy can better improve entrepreneurial ecosystems to enable productive entrepreneurship than target particular (potential) entrepreneurs (Audretsch & Thurik, 2001; Thurik et al., 2013; Stam, 2015). Until now, most studies and policies assumed entrepreneurship to be essentially a local phenomenon. Our study does not question the value of local contexts for entrepreneurship, but provides novel, more nuanced insights on this. We show the double edged-sword of improving entrepreneurial ecosystems: on the one hand, it stimulates the creation of startups; on the other hand, it also seems to increase the outflow of (potential) founders. Increasing the quality of entrepreneurial economies is likely to increase the creation of innovative startups in all regions, but with potentially increasing the inflow of non-local founders into the highest quality entrepreneurial economies, which might lead to the strongest increase of entrepreneurial activity in a few regions with the highest quality entrepreneurial economies. Positive sorting into migration, with migrant entrepreneurs realizing more successful firms, might even further strengthen this cumulative causation (cf. Conti & Guzman, 2022). This may partly explain the non-linear relation between the quality of entrepreneurial economies and entrepreneurial outputs (Leendertse et al., 2022). This might be good for the national economy at large, and in particular for the highest quality entrepreneurial regional economies (in Italy especially Milan), but at the expense of regions with a high but not the highest quality of entrepreneurial economy (in Italy especially Torino, Rome, and Naples). The outcomes of our analyses thus suggest that increasing the quality of the entrepreneurial economy not only leads to higher levels of entrepreneurship locally, but may also supply non-local startups to higher quality entrepreneurial economies. This may improve national welfare more than welfare at the region of origin.

Concerning the data collection on the relationship between entrepreneurship and entrepreneurial ecosystems, we need more longitudinal data of individual (firm) careers and of regional economies, in order to better trace the dynamic and net effects (creation, supply, and attraction) of increases in the quality of entrepreneurial economies over time. This might also provide insight into the net effects of founders leaving the region of origin (Anelli et al., 2019) and the improved access of remaining entrepreneurs to knowledge accumulated in these destination regions (Agrawal et al., 2011), or even the return of diaspora entrepreneurs (Saxenian, 2007). At the micro level, there is a need for more insight into the decision-making mechanisms of (potential) founders over time, and the performance of their firms. This could be done with data that allows us to trace the personal history of founders from birth to the act of startup creation, with the help of multi-stage location choice models. Notwithstanding, the startup location choice process is more complex to model than standard single choices, as composed by personal and work motivations of each startup member. We provided some first insights into these personal histories of founders with LinkedIn data, which allows observing sequential patterns of the startup founding process. Longitudinal data of individuals, in particular about their education, work and location decisions, and of their firms, in combination with data on their contexts, is needed to fully understand the micro level mechanisms and the resulting macro effects.

5. References

Acemoglu, D., Akcigit, U., & Celik, M. A. (2014). *Young, restless and creative: Openness to disruption and creative innovations* (No. w19894). National Bureau of Economic Research.

Agrawal, A., Kapur, D., McHale, J., & Oettl, A. (2011). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1), 43-55.

Anelli, M., Basso, G., Ippedico, G., & Peri, G. (2019). Youth drain, entrepreneurship and innovation (No. w26055). *National Bureau of Economic Research*.

Arthur, W. B. (2014). *Complexity and the Economy*. Oxford University Press.

Arthur, W. B. (2021). Foundations of complexity economics. *Nature Reviews Physics* 3(2), 136-145.

Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2022) Immigration and Entrepreneurship in the United States. *American Economic Review: Insights*, 4(1): 71-88.

Baltzopoulos, A., & Broström, A. (2013). Attractors of entrepreneurial activity: Universities, regions and alumni entrepreneurs. *Regional Studies*, 47(6), 934-949.

Basile, R., Girardi, A., Mantuano, M., & Russo, G. (2019). Interregional migration of human capital and unemployment dynamics: evidence from Italian provinces. *German Economic Review*, *20*(4), e385-e414.

Bauernschuster, S., Falck, O., & Heblich, S. (2010). Social capital access and entrepreneurship. *Journal of Economic Behavior & Organization*, 76(3), 821-833.

Baumol, W. J. (1990) Entrepreneurship: Productive, unproductive, and destructive, *Journal of Political Economy*, 98(5), pp. 893–921.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80(6), 2369-2429.

Bettencourt, L. M., Lobo, J., & Strumsky, D. (2007). Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy* 36(1), 107-120.

Biagi, B., Faggian, A., & McCann, P. (2011). Long and short distance migration in Italy: the role of economic, social and environmental characteristics. *Spatial Economic Analysis*, *6*(1), 111-131.

Biancalani, F., Czarnitzki, D. & Riccaboni, M. (2022) The Italian Start Up Act: a microeconometric program evaluation. *Small Business Economics* 58, 1699–1720. https://doi.org/10.1007/s11187-021-00468-7

Bonaccorsi, A., Colombo, M. G., Guerini, M., & Rossi-Lamastra, C. (2014). The impact of local and external university knowledge on the creation of knowledge-intensive firms: Evidence from the Italian case. *Small Business Economics*, *43*(2), 261-287.

Bonaventura, M., Ciotti, V., Panzarasa, P., Liverani, S., Lacasa, L., & Latora, V. (2020). Predicting success in the worldwide start-up network. *Scientific reports*, *10*(1), 1-6.

Bonifazi, C., Heins, F., & Tucci, E. (2017). Italy: internal migration in a low-mobility country. In *Internal Migration in the Developed World* (pp. 242-262). Routledge.

Bönte, W., Falck, O., & Heblich, S. (2009). The impact of regional age structure on entrepreneurship. *Economic Geography*, 85(3), 269-287.

Bosetti, V., Cattaneo, C., & Verdolini, E. (2015). Migration of skilled workers and innovation: A European perspective. *Journal of International Economics*, *96*(2), 311-322.

Bratti, M. and Conti, C., (2018). Immigration, Population Diversity and Innovation of Italian regions, *Regional Studies*, vol. 52(7), pages 934-947.

Breinlich, H., Corradi, V.,Rocha, N., Ruta, M., Santos Silva, J.M.C., Zylkin, T., (2021) “Machine Learning in International Trade Research – Evaluating the Impact of Trade Agreements”, World Bank Group –Development Research Group & Macroeconomics, Trade and Investment Global Practice Policy Research Working Paper 9629.

Buenstorf, G., & Klepper, S. (2010). Why does entry cluster geographically? Evidence from the US tire industry. *Journal of Urban Economics*, 68(2), 103-114.

Butler, J. S., Garg, R., & Stephens, B. (2020). Social networks, funding, and regional advantages in technology entrepreneurship: an empirical analysis. *Information Systems Research*, 31(1), 198-216.

Bryan, K. and Guzman, J., (2021). Entrepreneurial Migration (August 30, 2021). Available at SSRN: https://ssrn.com/abstract=3914010

Chatterji, A., Glaeser, E., & Kerr, W. (2014). Clusters of entrepreneurship and innovation. *Innovation policy and the economy*, 14(1), 129-166.

Chin, J. T. (2020). Location choice of new business establishments: Understanding the local context and neighborhood conditions in the United States. *Sustainability*, 12(2), 501.

Capozza, C., Salomone, S., & Somma, E. (2018). Local industrial structure, agglomeration economies and the creation of innovative start-ups: evidence from the Italian case. *Entrepreneurship & Regional Development*, *30*(7-8), 749-775.

Cavallo, A., Ghezzi, A., Colombelli, A., & Casali, G. L. (2020). Agglomeration dynamics of innovative start-ups in Italy beyond the industrial district era. *International Entrepreneurship and Management Journal*, 16(1), 239-262.

Colombelli, A. (2016). The impact of local knowledge bases on the creation of innovative start-ups in Italy. *Small Business Economics*, 47(2), 383-396.

Conti, A., & Guzman, J. A. (2022). What is the US comparative advantage in entrepreneurship? Evidence from Israeli migration to the United States. *The Review of Economics and Statistics*, forthcoming.

Credit, K., Mack, E. A., & Mayer, H. (2018). State of the field: Data and metrics for geographic analyses of entrepreneurial ecosystems, *Geography compass*, 12(9), e12380.

Dahl, M. S., & Sorenson, O. (2009). The embedded entrepreneur. *European Management Review*, 6(3), 172-181.

Dahl, M. S., & Sorenson, O. (2012). Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management Science*, 58(6), 1059-1071.

Derudder, B., & Taylor, P. (2005). The cliquishness of world cities. *Global Networks*, 5(1), 71-91.

Faggian, A., Rajbhandari, I., & Dotzel, K. R. (2017). The interregional migration of human capital and its regional consequences: a review. *Regional Studies*, *51*(1), 128-143.

Falck, O., Gold, R., & Heblich, S. (2017). Lifting the iron curtain: school-age education and entrepreneurial intentions. *Journal of Economic Geography*, 17(5), 1111-1148.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, *97*(1), 76-85.

Figueiredo, O., Guimaraes, P., & Woodward, D. (2002). Home-field advantage: location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, 52(2), 341-361.

Frenken, K., Cefis, E., & Stam, E. (2015). Industrial dynamics and clusters: a survey. *Regional Studies*, 49(1), 10-27.

Fritsch, M., & Storey, D. J. (2014). Entrepreneurship in a regional context: Historical roots, recent developments and future challenges. *Regional Studies*, 48(6), 939-954.

Fritsch, M., & Wyrwich, M. (2014). The long persistence of regional levels of entrepreneurship: Germany, 1925–2005. *Regional Studies*, 48(6), 955-973.

GEM (2021) *Global Entrepreneurship Monitor Report 2020*.

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?. *Journal of Economics & Management Strategy*, 18(3), 623-663.

Glaeser, E. L., Kerr, W. R., & Ponzetto, G. A. (2010). Clusters of entrepreneurship. *Journal of Urban Economics*, 67(1), 150-168.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2), 498-520.

Grilli, L., Mrkajic, B. & Giraudo, E. (2022). Industrial policy, innovative entrepreneurship, and the human capital of founders. *Small Business Economics* https://doi.org/10.1007/s11187-022-00611-y

Grillitsch, M. & Sotarauta, M. (2019). Trinity of change agency, regional development paths and opportunity spaces. *Progress in Human Geography*, 0309132519853870.

Guzman, J., & Stern, S. (2020). The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US States, 1988–2014. *American Economic Journal: Economic Policy*, 12(4), 212-43.

Hilton, B., Sood, A. P., & Evans, T. S. (2020). Predictive limitations of spatial interaction models: a non-Gaussian analysis. *Scientific Reports*, *10*(1), 1-10.

Hirschman, A. O. (1970). Exit, voice, and loyalty: Responses to decline in firms, organizations, and states (Vol. 25). *Harvard University Press*.

Iacobucci, D., & Perugini, F. (2020). Entrepreneurial Ecosystems in Italy. *L'industria*, *41*(2), 239-267.

Iacobucci, D., & Perugini, F. (2021). Entrepreneurial ecosystems and economic resilience at local level. *Entrepreneurship & Regional Development*, 1-28.

Italian Ministry of Economic Development - MISE (2019), *The Italian Startup Act Italy's policy framework to support innovative startups – executive summary*, Directorate General for Industrial Policy, Competitiveness and SMEs.

Knight, F.H. (1921). *Risk*, *Uncertainty and Profit*. New York: Houghton Mifflin.

Knoben, J., Ponds, R., & Van Oort, F. (2011). Employment from new firm formation in the Netherlands: Agglomeration economies and the Knowledge Spillover Theory of Entrepreneurship. *Entrepreneurship and Regional Development*, 23(3-4), 135-157.

Larsson, J. P., Wennberg, K., Wiklund, J., & Wright, M. (2017). Location choices of graduate entrepreneurs. *Research Policy*, 46(8), 1490-1504.

Leendertse, J., Schrijvers, M., & Stam, E. (2022). Measure twice, cut once: Entrepreneurial ecosystem metrics. *Research Policy*, https://doi.org/10.1016/j.respol.2021.104336.

Leoncini, R., Vecchiato, G., & Zamparini, L. (2020). Triggering cooperation among firms: an empirical assessment of the Italian Network Contract Law. *Economia Politica*, *37*(2), 357-380.

Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. *Economics Letters*, 99(1), 164-167.

Lucas, R. E. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics*, 9(3): 508–523

Maggioni, M. A., Uberti, T. E., & Usai, S. (2011). Treating patents as relational data: knowledge transfers and spillovers across Italian provinces. *Industry and Innovation*, *18*(1), 39-67.

Mason, C. M., & Harrison, R. T. (2006). After the exit: Acquisitions, entrepreneurial recycling and regional economic development. *Regional Studies*, 40(1), 55-73.

Michelacci, C., & Silva, O. (2007). Why so many local entrepreneurs? *Review of Economics and Statistics*, 89(4), 615-633.

Nifo, A., & Vecchione, G. (2014). Do institutions play a role in skilled migration? The case of Italy. *Regional Studies*, *48*(10), 1628-1649.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web.* Stanford InfoLab.

Patton, M. 1990. *Qualitative Evaluation and Research Methods*. Newbury Park, CA, SAGE Publications

Piras, R. (2017). A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches. *Papers in Regional Science*, *96*(3), 571-602.

Poot, J., Alimi, O., Cameron, M. P., & Maré, D. C. (2016). The gravity model of migration: the successful comeback of an ageing superstar in regional science. *Investigaciones Regionales= Journal of Regional Research*, (36), 63-68.

Reuschke, D., & Van Ham, M. (2013). Testing the 'residential rootedness' hypothesis of self-employment for Germany and the UK. *Environment and Planning A*, 45(5), 1219-1239.

Samila, S., & Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *Review of Economics and Statistics*, 93(1), 338-349.

Saxenian, A. (2007). *The new argonauts: Regional advantage in a global economy*. Harvard University Press.

Schumpeter, J. A. (1934). *The Theory of Economic Development*. Cambridge, MA: Harvard University Press

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and statistics*, *88*(4), 641-658.

Silva, J. S., & Tenreyro, S. (2011). Poisson: Some convergence issues. *The Stata Journal*, *11*(2), 207-212.

Sontag, K. (2018). *Mobile entrepreneurs: An ethnographic study of the migration of the highly skilled*, Budrich UniPress Ltd, Opladen, Germany

Sorenson, O., & Audia, P. G. (2000). The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940–1989. *American Journal of Sociology*, 106(2), 424-462.

Stam, E. (2007). Why butterflies don 't leave: Locational behavior of entrepreneurial firms. *Economic Geography*, 83(1), 27-50.

Stam, E. (2013). Knowledge and entrepreneurial employees: A country-level analysis. *Small Business Economics*, *41*(4), 887-898.

Stam, E. (2015). Entrepreneurial ecosystems and regional policy: a sympathetic critique. *European Planning Studies*, *23*(9), 1759-1769.

Stam, E., & Van de Ven, A. (2021). Entrepreneurial ecosystem elements. *Small Business Economics*, 56(2), 809-832.

Stewart, J. Q. (1950). The development of social physics. *American Journal of Physics*, *18*(5), 239-253.

Tinbergen, J., (1962)., *Shaping the World Economy*, New York: The Twentieth Century Fund.

Vedula, S., & Kim, P. H. (2019). Gimme shelter or fade away: the impact of regional entrepreneurial ecosystem quality on venture survival. *Industrial and Corporate Change*, 28(4), 827-854.

Verginer, L., & Riccaboni, M. (2021). Talent goes to global cities: The world network of scientists' mobility. *Research Policy*, *50*(1), 104127.

Wurth, B., Stam, E., & Spigel, B. (2022). Toward an entrepreneurial ecosystem research program. *Entrepreneurship Theory and Practice*, https://doi.org/10.1177/1042258721998948.

6. Appendix

Table A.1 Data sources for the Entrepreneurial Ecosystem Index, Non-local founders and control variables

Variable	Description	Source	Year(s)
EEI. Element 1: FORMAL INSTITUTIONS	Institutional Quality Index based on the World Governance Indicator (WGI)	https://sites.google.com/site/institutionalqualityindex/dataset	Avg 2015-2018
EEI. Element 2: ENTREPRENEURSHIP CULTURE	Number of new firms per capita (excluding the "sole proprietorship" firms). Average number of new firms in the period	ITALIAN CHAMBERS OF COMMERCE & ISTAT	Avg 2015-2019
EEI. Element 3: NETWORKS	Number of contracts between firms ("rete contratto")	ITALIAN CHAMBERS OF COMMERCE - http://contrattidirete.registroimprese.it/reti/	2010-2020
EEI. Element 4: PHYSICAL INFRASTRUCTURE	travel time to urban nodes	ISTAT	2013
	percentage of population with a broadband subscription	ISTAT	2017
	Average speed per km of public road transport in the NUTS 3 regional capitals	ISTAT	2017
EEI. Element 5: FINANCE	number of firms with at least 3 employees that rely on Venture Capital Funds as a main source of financing	ISTAT	2018
	number of firms with at least 3 employees that rely on Crowdfunding as a main source of financing	ISTAT	2018
	number of firms with at least 3 employees that rely on project financing as a main source of financing	ISTAT	2018
EEI. Element 6: LEADERSHIP	The number of coordinators on H2020 innovation projects per capita (per thousand inhabitants)	CORDIS Database	from 2014 to 2019
EEI. Element 7: TALENT	percentage of population that completed tertiary education	ISTAT	2018
	percentage of population that obtained a PhD	ISTAT	2018
	percentage of firms with at least 10 employees engaged in training activities (excluding the compulsory ones)	ISTAT	2018
	percentage of firms with at least 10 employees that have invested in digital technologies	ISTAT	2018

Variable	Description	Source	Year(s)
EEI. Element 8: NEW KNOWLEDGE	percentage of firms that conduct intramural R&D activities	ISTAT	2018
EEI. Element 9: DEMAND	GDP per capita (thousand value)	EUROSTAT	2017
EEI. Element 10: INTERMEDIATE SERVICES	percentage of firms in knowledge intensive market services over the total business population	ITALIAN CHAMBERS OF COMMERCE	Avg 2015-2019
Non-local founders	Number of founders born outside the NUTS 3 regions of startup foundation	ORBIS	2016-2019
Pop. density (ln)	Population density	ISTAT	2018
Driving time	distances between NUTS 3 regions centroids	OPENROUTESERVICE	2019
NUTS 3 Contiguity	dummy variable for shared borders	ISTAT	-
GDP p.c. (ln)	GDP per capita	EUROSTAT	2017
STEM/pop.	number of STEM students every 1000 inhabitants	MIUR	Avg. 2016-2019
NUTS 2		ISTAT	-
Migration (btw NUTS 3 regions)	Internal migration between Italian NUTS 3regions	ISTAT	Avg. 2016-2019

Figure A.1 The correlation matrix between the elements of the Entrepreneurial Ecosystem Index

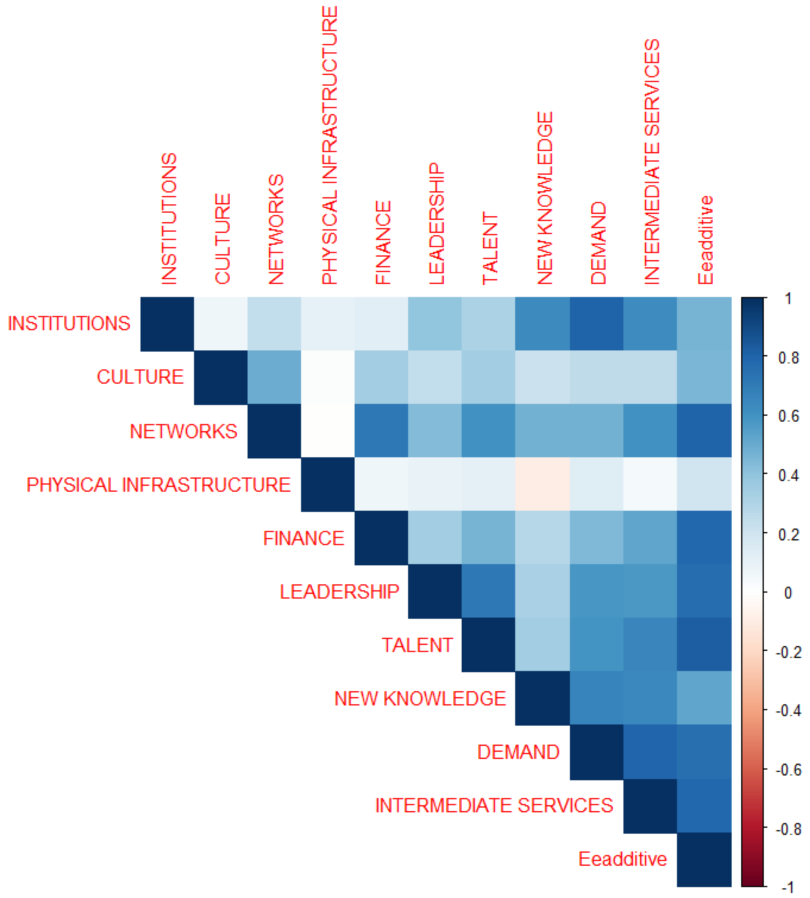


Table A.2. Summary of the main features of Italian NUTS 3 regions ranked by EE index: innovative startups, network centrality, and EE index

Ran k	**NUTS 3 region**	**code_regione**	**NUTS 2 Region**	**EE index**	**EE index quartile**	**Pagerank**	**Average age founders born in the region**	**Indegree**	**Outdegree**	**Nuber of Innovative Startups 2016-19**	**Innovative Startups 2016_19 (sample_all_ext)**	**Innovative Startups 2016_19 (sample_at_least_oneext)**
1	Milano	MI	LOMBARDIA	37,20	high	0,2214	50,25	1527	354	1362	435	731
2	Roma	RM	LAZIO	33,39	high	0,0721	48,58	444	303	764	187	298
3	Bologna	BO	EMILIA-ROMAGNA	24,93	high	0,0309	47,59	180	67	204	61	92
4	Torino	TO	PIEMONTE	20,86	high	0,0398	48,87	185	233	269	47	97
5	Trento	TN	TRENTINO-ALTO ADIGE/SÜDTIROL	20,49	high	0,0242	48,11	113	37	128	35	54
6	Trieste	TS	FRIULI-VENEZIA GIULIA	19,50	high	0,0076	50,17	33	28	28	8	14
7	Firenze	FI	TOSCANA	19,43	high	0,0118	44,89	61	64	93	15	40
8	Bolzano/Bozen	BZ	TRENTINO-ALTO ADIGE/SÜDTIROL	19,16	high	0,0098	47,68	34	26	73	16	20
9	Brescia	BS	LOMBARDIA	18,29	high	0,0195	48,39	88	100	137	17	42
10	Genova	GE	LIGURIA	16,75	high	0,0183	51,31	90	113	98	14	43
11	Bergamo	BG	LOMBARDIA	16,73	high	0,0236	47,19	83	103	172	33	51
12	Venezia	VE	VENETO	16,46	high	0,0124	47,52	59	91	75	17	28
13	Pisa	PI	TOSCANA	15,93	high	0,0109	48,14	55	22	70	24	33
14	Padova	PD	VENETO	15,70	high	0,0263	47,60	144	87	204	44	78
15	Parma	PR	EMILIA-ROMAGNA	15,29	high	0,0098	49,58	48	34	65	14	24
16	Bari	BA	PUGLIA	15,25	high	0,0079	45,21	53	104	145	23	40
17	Modena	MO	EMILIA-ROMAGNA	14,43	high	0,0120	51,26	47	44	100	7	27
18	Verona	VR	VENETO	13,08	high	0,0155	45,21	74	87	137	22	38
19	Reggio nell'Emilia	RE	EMILIA-ROMAGNA	12,53	high	0,0144	51,02	69	42	61	21	32
20	Monza e della Brianza	MB	LOMBARDIA	12,35	high	0,0250	43,37	82	60	71	32	41

Ran k	NUTS 3 region	code_regione	NUTS 2 Region	EE index	EE index quartile	Pagerank	Average age founders born in the region	Indegree	Outdegree	Nuber of Innovative Startups 2016-19	Innovative Startups 2016_19 (sample_all_ext)	Innovative Startups 2016_19 (sample_at_least_oneext)
21	Siena	SI	TOSCANA	12,33	high	0,0047	51,47	15	15	19	8	11
22	Vicenza	VI	VENETO	11,96	high	0,0094	46,51	35	71	67	5	20
23	Pordenone	PN	FRIULI-VENEZIA GIULIA	11,81	high	0,0072	46,79	31	32	44	10	18
24	Treviso	TV	VENETO	11,65	high	0,0141	45,17	58	80	105	16	31
25	Napoli	NA	CAMPANIA	11,40	high	0,0129	47,26	73	275	301	25	47
26	Ancona	AN	MARCHE	11,26	high	0,0095	46,30	49	33	78	12	28
27	Perugia	PG	UMBRIA	11,07	medium-high	0,0064	47,24	29	45	107	11	21
28	Udine	UD	FRIULI-VENEZIA GIULIA	10,93	medium-high	0,0088	49,19	41	50	69	13	27
29	Livorno	LI	TOSCANA	10,91	medium-high	0,0025	51,03	5	32	18	4	5
30	Cagliari	CA	SARDEGNA	10,67	medium-high	0,0055	46,95	17	39	46	9	11
31	Ravenna	RA	EMILIA-ROMAGNA	10,61	medium-high	0,0057	49,19	27	35	39	10	16
32	Rimini	RN	EMILIA-ROMAGNA	10,58	medium-high	0,0062	46,26	30	22	62	14	18
33	Cuneo	CN	PIEMONTE	10,24	medium-high	0,0070	47,67	37	40	66	11	20
34	Forlì-Cesena	FC	EMILIA-ROMAGNA	9,98	medium-high	0,0042	43,68	14	34	38	7	10
35	Ferrara	FE	EMILIA-ROMAGNA	9,95	medium-high	0,0051	48,60	14	39	20	2	7
36	Pavia	PV	LOMBARDIA	9,76	medium-high	0,0103	48,49	46	60	47	10	19
37	Varese	VA	LOMBARDIA	9,61	medium-high	0,0110	47,07	33	100	55	10	19
38	Como	CO	LOMBARDIA	9,40	medium-high	0,0187	49,58	61	55	46	15	20
39	Salerno	SA	CAMPANIA	9,24	medium-high	0,0094	45,28	67	81	168	34	51
40	Macerata	MC	MARCHE	9,14	medium-high	0,0049	49,24	22	27	43	9	14
41	Pescara	PE	ABRUZZO	9,11	medium-high	0,0052	50,97	21	36	39	7	12
42	Lecco	LC	LOMBARDIA	9,02	medium-high	0,0067	48,54	14	38	30	4	7

Ran k	NUTS 3 region	code_regione	NUTS 2 Region	EE index	EE index quartile	Pagerank	Average age founders born in the region	Indegree	Outdegree	Nuber of Innovative Startups 2016-19	Innovative Startups 2016_19 (sample_all_ext)	Innovative Startups 2016_19 (sample_at_least_oneext)
43	Prato	PO	TOSCANA	8,88	medium-high	0,0023	44,67	7	9	12	5	6
44	L'Aquila	AQ	ABRUZZO	8,86	medium-high	0,0122	46,30	55	21	47	12	20
45	Novara	NO	PIEMONTE	8,84	medium-high	0,0087	44,53	39	40	33	11	19
46	Arezzo	AR	TOSCANA	8,79	medium-high	0,0032	55,04	9	24	18	3	5
47	Lecce	LE	PUGLIA	8,65	medium-high	0,0076	45,01	46	61	86	18	29
48	Mantova	MN	LOMBARDIA	8,42	medium-high	0,0023	47,64	4	43	22	3	4
49	Piacenza	PC	EMILIA-ROMAGNA	8,41	medium-high	0,0019	46,16	4	33	31	2	4
50	Cremona	CR	LOMBARDIA	8,37	medium-high	0,0036	48,05	13	53	26	4	8
51	Pesaro e Urbino	PU	MARCHE	8,37	medium-high	0,0041	48,03	13	28	41	6	9
52	Chieti	CH	ABRUZZO	8,36	medium-high	0,0030	46,70	11	48	23	5	7
53	Lucca	LU	TOSCANA	8,32	medium_low	0,0029	47,62	10	33	29	3	7
54	Sondrio	SO	LOMBARDIA	8,27	medium_low	0,0016	45,77	1	18	8	0	1
55	Pistoia	PT	TOSCANA	8,26	medium_low	0,0042	51,08	17	13	20	5	8
56	Foggia	FG	PUGLIA	7,99	medium_low	0,0017	47,23	4	64	37	1	3
57	Catania	CT	SICILIA	7,96	medium_low	0,0045	43,47	26	57	109	12	20
58	Avellino	AV	CAMPANIA	7,86	medium_low	0,0030	41,56	22	38	58	16	22
59	Gorizia	GO	FRIULI-VENEZIA GIULIA	7,81	medium_low	0,0018	46,33	3	12	12	3	3
60	Vercelli	VC	PIEMONTE	7,77	medium_low	0,0025	46,44	2	24	4	0	1
61	Matera	MT	BASILICATA	7,49	medium_low	0,0025	45,25	9	26	17	5	7
62	Latina	LT	LAZIO	7,48	medium_low	0,0075	44,69	27	28	38	12	15
63	Catanzaro	CZ	CALABRIA	7,47	medium_low	0,0033	47,51	13	36	44	6	11
64	La Spezia	SP	LIGURIA	7,46	medium_low	0,0019	50,48	3	22	10	2	3
65	Campobasso	CB	MOLISE	7,42	medium_low	0,0040	48,17	17	18	38	13	16

Rank	NUTS 3 region	code_regione	NUTS 2 Region	EE index	EE index quartile	Pagerank	Average age founders born in the region	Indegree	Outdegree	Nuber of Innovative Startups 2016-19	Innovative Startups 2016_19 (sample_all_ext)	Innovative Startups 2016_19 (sample_at_least_oneext)
66	Trapani	TP	SICILIA	7,37	medium_low	0,0020	47,86	7	18	12	5	6
67	Ascoli Piceno	AP	MARCHE	7,35	medium_low	0,0053	48,31	25	47	69	6	15
68	Terni	TR	UMBRIA	7,33	medium_low	0,0033	46,50	8	18	32	4	6
69	Fermo	FM	MARCHE	7,31	medium_low	0,0037	38,46	16	13	16	5	8
70	Belluno	BL	VENETO	7,28	medium_low	0,0022	43,55	2	19	10	1	2
71	Teramo	TE	ABRUZZO	7,26	medium_low	0,0041	48,96	23	22	43	13	19
72	Valle d'Aosta/Vallée d'Aoste	AO	VALLE D'AOSTA/VALLÉE D'AOSTE	7,23	medium_low	0,0036	50,75	15	5	12	3	5
73	Cosenza	CS	CALABRIA	7,15	medium_low	0,0029	47,70	19	44	60	12	15
74	Asti	AT	PIEMONTE	7,12	medium_low	0,0016	49,96	1	21	2	1	1
75	Frosinone	FR	LAZIO	7,11	medium_low	0,0026	44,40	10	32	19	3	5
76	Verbano-Cusio-Ossola	VB	PIEMONTE	7,09	medium_low	0,0025	48,40	7	14	7	2	4
77	Isernia	IS	MOLISE	6,99	medium_low	0,0034	42,11	11	8	15	5	7
78	Crotone	KR	CALABRIA	6,98	medium_low	0,0015	44,92	1	10	11	1	1
79	Potenza	PZ	BASILICATA	6,95	medium_low	0,0036	50,10	18	36	50	11	15
80	Alessandria	AL	PIEMONTE	6,91	Low	0,0063	51,30	18	41	21	7	8
81	Lodi	LO	LOMBARDIA	6,90	Low	0,0060	45,75	19	25	15	4	7
82	Rovigo	RO	VENETO	6,89	Low	0,0049	47,26	23	30	41	13	17
83	Benevento	BN	CAMPANIA	6,86	Low	0,0028	48,84	10	29	36	5	9
84	Grosseto	GR	TOSCANA	6,83	Low	0,0027	49,72	5	18	4	1	2
85	Biella	BI	PIEMONTE	6,80	Low	0,0031	44,32	8	25	17	3	5
86	Sassari	SS	SARDEGNA	6,66	Low	0,0028	47,68	10	20	31	3	6
87	Massa-Carrara	MS	TOSCANA	6,63	Low	0,0020	53,13	5	15	7	0	2

Ran k	NUTS 3 region	code_regione	NUTS 2 Region	EE index	EE index quartile	Pagerank	Average age founders born in the region	Indegree	Outdegree	Nuber of Innovative Startups 2016-19	Innovative Startups 2016_19 (sample_all_ext)	Innovative Startups 2016_19 (sample_at_least_oneext)
88	Messina	ME	SICILIA	6,46	Low	0,0032	48,47	12	46	43	6	10
89	Palermo	PA	SICILIA	6,39	Low	0,0068	45,57	32	60	121	15	27
90	Savona	SV	LIGURIA	6,36	Low	0,0017	49,42	2	43	7	2	2
91	Taranto	TA	PUGLIA	6,32	Low	0,0032	46,58	14	50	25	4	7
92	Caserta	CE	CAMPANIA	6,29	Low	0,0047	45,29	40	35	110	24	32
93	Viterbo	VT	LAZIO	6,16	Low	0,0018	51,67	2	16	11	2	2
94	Reggio di Calabria	RC	CALABRIA	5,90	Low	0,0022	46,01	8	63	43	2	6
95	Siracusa	SR	SICILIA	5,76	Low	0,0024	51,78	11	21	18	4	6
96	Brindisi	BR	PUGLIA	5,72	Low	0,0020	43,98	6	39	19	3	5
97	Rieti	RI	LAZIO	5,47	Low	0,0025	46,25	6	12	7	4	4
98	Nuoro	NU	SARDEGNA	5,41	Low	0,0015	48,65	1	15	7	0	1
99	Imperia	IM	LIGURIA	5,31	Low	0,0015	48,00	1	17	2	0	1
100	Ragusa	RG	SICILIA	5,29	Low	0,0018	46,13	5	13	17	4	5
101	Vibo Valentia	VV	CALABRIA	5,27	Low	0,0016	43,62	1	12	3	1	1
102	Oristano	OR	SARDEGNA	5,15	Low	0,0048	50,67	14	3	9	6	7
103	Caltanissetta	CL	SICILIA	4,29	Low	0,0026	51,50	14	17	20	8	11
104	Agrigento	AG	SICILIA	3,89	Low	0,0017	46,97	3	30	3	0	2
105	Enna	EN	SICILIA	3,72	Low	0,0020	48,20	8	15	8	2	2
/	total	/	/	/	/	1	47,67	5004	5004	7529	1660	2779

Table A.3. Correlation matrix variables gravity model

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) Non-local founders (all external)	1.000													
(2) Pop. density (ln), foundation	0.287	1.000												
(3) Pop. density (ln), origin	0.137	-0.010	1.000											
(4) Driving time (ln)	-0.157	-0.085	-0.084	1.000										
(5) Contiguity, NUTS 3	0.187	0.002	0.002	-0.415	1.000									
(6) GDP p.c. (ln), foundation	0.201	0.232	-0.002	-0.337	0.030	1.000								
(7) GDP p.c. (ln), origin	0.049	-0.002	0.232	-0.334	0.030	-0.010	1.000							
(8) STEM/pop., foundation	0.064	0.044	-0.000	-0.032	0.018	0.153	-0.001	1.000						
(9) STEM/pop., origin	0.021	-0.000	0.044	-0.033	0.018	-0.001	0.153	-0.010	1.000					
(10) EEI (ln), foundation	0.285	0.419	-0.004	-0.230	0.024	0.732	-0.007	0.401	-0.004	1.000				
(11) EEI (ln), origin	0.118	-0.004	0.419	-0.228	0.024	-0.007	0.732	-0.004	0.401	-0.010	1.000			
(12) Regional_transfer	0.596	0.218	0.220	-0.282	0.474	0.146	0.069	0.043	0.039	0.222	0.166	1.000		
(13) EEI (ln), foundation (w/o demand)	0.286	0.415	-0.004	-0.241	0.025	0.764	-0.007	0.389	-0.004	0.999	-0.010	0.221	1.000	
(14) EEI (ln), origin (w/o demand)	0.116	-0.004	0.415	-0.239	0.025	-0.007	0.764	-0.004	0.389	-0.010	0.999	0.163	-0.010	1.000

Table A.4 Performance of PPML LASSO (model 3) across different lambda threshold (β penalized coefficients are reported)

id	Variable	Lambda(0.5)	Lambda(0.4)	Lambda(0.3)	Lambda(0.2)	Lambda (0.1)	Lambda (0.075)	Lambda (0.05)	Lambda (0.025)	Lambda (0.01)	Lambda (0.005)	Lambda (0.0001)
1	NUTS3_contiguity	0,668	0,761	**0,875**	1,011	1,048	1,027	0,994	0,952	0,926	0,900	0,810
2	Drivingtimes(ln)	-0,136	-0,188	**-0,229**	-0,262	-0,409	-0,501	-0,604	-0,721	-0,808	-0,850	-0,959
3	GDP_pc(ln)_origin	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,715
4	GDP_pc(ln)_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,098	0,431	0,677	1,274
5	Popdensity_foundation(ln)	0,087	0,100	**0,114**	0,128	0,134	0,134	0,131	0,129	0,113	0,105	0,093
6	Popdensity_origin (ln)	0,062	0,077	**0,094**	0,103	0,091	0,093	0,094	0,095	0,092	0,091	0,084
7	EEI(ln)_origin(no_demand)	0,202	0,332	**0,467**	0,628	0,888	1,029	1,171	1,328	1,490	1,536	1,428
8	EEI(ln)_foundation(no_demand)	1,422	1,533	**1,644**	1,757	1,881	1,929	1,997	2,005	1,942	1,889	1,804
9	STEM/pop_origin	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,248	-1,541	-1,948	-2,516
10	STEM/pop_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,422	2,310	2,584	2,260
11	regional_dummy_Birth_1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,208	-0,421	-1,497
12	regional_dummy_foundation_1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-1,207
13	regional_dummy_Birth_2	0,000	0,000	0,000	0,000	0,000	0,000	-0,082	-0,695	-1,224	-1,538	-2,954
14	regional_dummy_foundation_2	0,000	0,000	0,000	0,000	0,000	0,000	0,089	0,501	0,626	0,607	-0,690
15	regional_dummy_Birth_3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,019	-0,218	-1,315
16	regional_dummy_foundation_3	0,000	0,000	0,000	0,000	0,000	-0,214	-0,444	-0,726	-0,942	-1,032	-2,311
17	regional_dummy_Birth_4	0,000	0,000	0,000	0,000	-0,016	-0,248	-0,465	-0,706	-1,005	-1,232	-2,378
18	regional_dummy_foundation_4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,034	-1,354
19	regional_dummy_Birth_5	0,000	0,000	0,000	0,000	-0,094	-0,463	-0,816	-1,188	-1,564	-1,814	-3,045
20	regional_dummy_foundation_5	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,148	-0,236	-1,589
21	regional_dummy_Birth_6	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,075	-0,299	-0,503	-1,597
22	regional_dummy_foundation_6	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,082	-0,133	-1,389
23	regional_dummy_Birth_7	0,000	0,000	0,000	0,000	0,000	-0,237	-0,448	-0,652	-0,831	-1,019	-2,040
24	regional_dummy_foundation_7	0,000	0,000	0,000	0,000	-0,016	-0,148	-0,282	-0,503	-0,647	-0,702	-1,889

25	regional_dummy_Birth_8	0,000	0,000	0,000	0,000	-0,081	-0,290	-0,471	-0,659	-0,878	-1,083	-2,238
26	regional_dummy_foundation_8	0,000	0,000	0,000	0,000	0,000	0,000	-0,088	-0,226	-0,382	-0,465	-1,781
27	regional_dummy_Birth_9	0,000	0,000	0,000	0,000	-0,264	-0,440	-0,586	-0,732	-0,902	-1,089	-2,131
28	regional_dummy_foundation_9	0,000	0,000	0,000	0,000	-0,364	-0,486	-0,617	-0,813	-0,987	-1,065	-2,305
29	regional_dummy_Birth_10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,102	-1,025
30	regional_dummy_foundation_10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,130	-0,195	-1,365
31	regional_dummy_Birth_11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,177	-1,149
32	regional_dummy_foundation_11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,019	-1,221
33	regional_dummy_Birth_12	0,000	0,000	0,000	0,000	0,223	0,208	0,212	0,201	0,076	-0,055	-1,003
34	regional_dummy_foundation_12	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,004	-1,207
35	regional_dummy_Birth_13	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,131	-1,054
36	regional_dummy_foundation_13	0,000	0,000	0,000	0,000	0,000	0,029	0,179	0,263	0,266	0,255	-0,853
37	regional_dummy_Birth_14	0,000	0,000	0,000	0,000	0,000	0,000	-0,045	-0,348	-0,558	-0,759	-1,566
38	regional_dummy_foundation_14	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,059	0,220	0,289	-0,682
39	regional_dummy_Birth_15	0,000	0,000	0,000	0,164	0,658	0,715	0,803	0,898	0,891	0,757	0,144
40	regional_dummy_foundation_15	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,113	0,223	-0,694
41	regional_dummy_Birth_16	0,000	0,000	0,000	0,000	0,470	0,580	0,722	0,873	0,883	0,764	0,152
42	regional_dummy_foundation_16	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,017	0,000	-1,008
43	regional_dummy_Birth_17	0,000	0,000	0,000	0,000	0,000	0,000	0,126	0,435	0,572	0,492	-0,205
44	regional_dummy_foundation_17	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,085	0,142	-0,889
45	regional_dummy_Birth_18	0,000	0,000	0,000	0,000	0,186	0,409	0,672	0,953	1,068	0,981	0,415
46	regional_dummy_foundation_18	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,086	-0,832
47	regional_dummy_Birth_19	0,000	0,000	0,000	0,000	0,514	0,744	1,010	1,292	1,415	1,335	0,770
48	regional_dummy_foundation_19	0,000	0,000	0,000	0,000	0,102	0,305	0,494	0,667	0,808	0,885	-0,070

Table A.5 Lambda Cross Validation of model 3

Lambda	RMSE
0.4000	90.28339
0.3000	**88.72941**
0.2000	91.55902
0.1000	101.87478
0.0750	102.67207
0.0500	104.45101
0.0250	109.44966
0.0100	112.75802
0.0050	114.72712
0.0001	116.96749
0.0000	116.98751

Table A.6 Gravity models with interregional migration flows as a control variable

	Non-local founders' movements						
	all external	at least 1 external	majority external	all external UNDER45	all external OVER45	all external female	all external male
	b/se	b/se	b/se	b/se	b/se	b/se	b/se
EEI (ln), foundation	**0.8597*****	**0.9875*****	**0.9355*****	**0.8645*****	**0.8439*****	**1.3809*****	**0.7570*****
	(0.1359)	**(0.1067)**	**(0.1123)**	**(0.1833)**	**(0.1766)**	**(0.2832)**	**(0.1520)**
EEI (ln), origin	**0.5957*****	**0.5325*****	**0.5463*****	**0.4967****	**0.6851*****	**0.8736*****	**0.5419*****
	(0.1376)	**(0.1070)**	**(0.1109)**	**(0.1748)**	**(0.1733)**	**(0.2565)**	**(0.1472)**
GDP p.c. (ln), foundation	1.2061***	1.3102***	1.2768***	1.1953**	1.2229**	0.9279	1.2597***
	(0.3431)	(0.3009)	(0.3062)	(0.4485)	(0.4315)	(0.6728)	(0.3741)
GDP p.c. (ln), origin	0.5259	0.7266*	0.6545	0.3510	0.6465	0.6113	0.5189
	(0.3941)	(0.3497)	(0.3583)	(0.4536)	(0.5043)	(0.6636)	(0.4272)
Pop. density (ln), foundation	0.0267	-0.0036	0.0062	0.0252	0.0281	0.0249	0.0277
	(0.0217)	(0.0198)	(0.0203)	(0.0266)	(0.0263)	(0.0479)	(0.0217)
Pop. density (ln), origin	0.0102	-0.0122	-0.0103	-0.0030	0.0215	0.0422	0.0034
	(0.0295)	(0.0297)	(0.0296)	(0.0274)	(0.0345)	(0.0322)	(0.0326)
STEM/pop., foundation	3.6712***	3.8826***	4.1031***	5.7903***	2.1370	4.2566*	3.6064***
	(0.9418)	(0.7088)	(0.7384)	(1.3358)	(1.2119)	(1.8405)	(1.0788)
STEM/pop., origin	-2.1205*	-1.0222	-1.1835	-1.3579	-2.8533**	-2.9116	-2.0010*
	(0.8596)	(0.6268)	(0.6805)	(1.1438)	(1.1039)	(2.2200)	(0.9094)
Driving time (ln)	-0.4494***	-0.4416***	-0.4501***	-0.5188***	-0.4088***	-0.7316***	-0.3927***
	(0.0625)	(0.0534)	(0.0554)	(0.0798)	(0.0755)	(0.1173)	(0.0674)
Contiguity, NUTS 3	-0.3545**	-0.3626***	-0.3759***	-0.3574*	-0.3653*	0.0476	-0.4346***
	(0.1143)	(0.0933)	(0.0967)	(0.1452)	(0.1488)	(0.2158)	(0.1248)
Interregional migration	0.7725***	0.7984***	0.7931***	0.7931***	0.7538***	0.5956***	0.8071***
	(0.0433)	(0.0361)	(0.0374)	(0.0580)	(0.0540)	(0.0824)	(0.0468)
NUTS 2 dummies				YES			
Constant	-6.0791***	-5.9684***	-5.9713***	-7.0169***	-6.5436***	-10.4082***	-5.9947***
	(0.5109)	(0.3989)	(0.4229)	(0.6852)	(0.6285)	(1.3358)	(0.5456)
N	10920	10920	10920	10920	10920	10610	10920
R2	0.726	0.794	0.791	0.665	0.647	0.542	0.687
Pseudo Log-likelihood	-3645.663	-5034.385	-4829.2458	-2170.100	-2511.946	-1058.179	-3301.110
VIF max	3.76	3.76	3.76	3.76	3.76	3.76	3.76

Table A.7 Performance of PPML LASSO (model 3 with decomposed version of EE index) across different lambda threshold (β penalized coefficients are reported)

Variable	lambda(0.4)	lambda(0.3)	lambda(0.2)	**lambda (0.1)**	lambda(0.075)	lambda(0.05)	lambda(0.025)	lambda(0.01)	lambda(0.005)	lambda(0.0001)

NUTS3_contiguity	0,688	0,785	0,889	**0,895**	0,884	0,903	0,923	0,931	0,928	0,868
Drivingtimes(ln)	-0,252	-0,317	-0,392	**-0,584**	-0,644	-0,720	-0,808	-0,869	-0,893	-0,966
Popdensity_foundation (ln)	0,000	0,001	0,020	**0,037**	0,040	0,044	0,037	0,020	0,012	0,009
Popdensity_origin(ln)	0,008	0,020	0,029	**0,017**	0,012	0,022	0,028	0,040	0,046	0,049
STEM/pop_origin	0,000	0,000	0,000	0,000	0,000	0,000	0,977	1,349	1,491	1,359
STEM/pop_foundation	0,000	0,000	0,000	0,000	0,389	1,484	2,724	3,428	3,625	3,801
INSTITUTIONS_birth	0,000	0,000	-0,062	**-0,574**	-0,677	-0,769	-0,801	-0,664	-0,606	-0,314
INSTITUTIONS_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,402
CULTURE_birth	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,076	0,058
CULTURE_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,037	0,070	0,082	0,048
NETWORKS_birth	0,305	0,355	0,413	**0,485**	0,505	0,493	0,497	0,494	0,486	0,489
NETWORKS_foundation	0,348	0,354	0,345	**0,330**	0,341	0,357	0,405	0,464	0,485	0,569
PHYSICAL INFRASTRUCTURE_birth	0,000	0,000	0,000	0,000	0,000	0,042	0,143	0,211	0,239	0,243
PHYSICAL INFRASTRUCTURE_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,033	-0,158
FINANCE_birth	0,000	0,000	0,000	**0,012**	0,015	0,037	0,050	0,079	0,091	0,092
FINANCE_foundation	0,017	0,027	0,043	**0,055**	0,052	0,049	0,049	0,037	0,030	0,025
LEADERSHIP_birth	0,000	0,000	0,000	0,000	0,000	0,002	0,008	0,044	0,065	0,083
LEADERSHIP_foundation	0,000	0,028	0,068	**0,111**	0,118	0,117	0,124	0,133	0,137	0,114
TALENT_birth	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,040	-0,077	-0,135
TALENT_foundation	0,070	0,062	0,051	**0,040**	0,034	0,025	0,017	0,004	0,002	0,025
NEW KNOWLEDGE_birth	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,024
NEW KNOWLEDGE_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,032	0,101
DEMAND_birth	0,000	0,000	0,000	0,000	0,000	0,000	-0,004	-0,212	-0,272	-0,189
DEMAND_foundation	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,636
INTERMEDIATE SERVICES_birth	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,151
INTERMEDIATE SERVICES_foundation	0,781	0,793	0,729	**0,720**	0,738	0,784	0,720	0,706	0,720	1,067
regional_dummy_Birth_1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,173	-0,279	-0,885
regional_dummy_foundation_1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	-1,167
regional_dummy_Birth_2	0,000	0,000	0,000	0,000	0,000	-0,092	-0,721	-1,118	-1,287	-1,972
regional_dummy_foundation_2	0,000	0,000	0,000	0,000	0,000	0,000	0,232	0,488	0,572	-0,148
regional_dummy_Birth_3	0,000	0,000	0,000	0,000	0,000	0,075	0,154	0,114	0,053	-0,442
regional_dummy_foundation_3	0,000	0,000	0,000	0,000	-0,120	-0,339	-0,574	-0,763	-0,838	-1,815
regional_dummy_Birth_4	0,000	0,000	0,000	0,000	0,000	-0,224	-0,504	-0,772	-0,898	-1,567
regional_dummy_foundation_4	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-1,252

regional_dummy_Birth_5	0,000	0,000	0,000	0,000	0,000	0,000	-0,072	-0,526	-0,750	-1,513
regional_dummy_foundation_5	0,000	0,000	0,000	0,059	0,215	0,408	0,569	0,653	0,682	-0,173
regional_dummy_Birth_6	0,000	0,000	0,000	0,000	0,000	0,049	0,067	0,000	-0,054	-0,753
regional_dummy_foundation_6	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,126	-0,212	-1,592
regional_dummy_Birth_7	0,000	0,000	0,000	0,000	0,000	0,000	-0,165	-0,323	-0,370	-0,989
regional_dummy_foundation_7	0,000	0,000	0,000	0,000	0,000	0,000	-0,116	-0,245	-0,316	-1,689
regional_dummy_Birth_8	0,000	0,000	0,000	0,000	0,000	-0,059	-0,292	-0,447	-0,520	-1,109
regional_dummy_foundation_8	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,095	-0,159	-1,300
regional_dummy_Birth_9	0,000	0,000	0,000	-0,113	-0,185	-0,313	-0,522	-0,727	-0,875	-1,416
regional_dummy_foundation_9	0,000	0,000	0,000	-0,100	-0,220	-0,350	-0,548	-0,744	-0,823	-1,885
regional_dummy_Birth_10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,089	-0,609
regional_dummy_foundation_10	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,171	-0,243	-1,462
regional_dummy_Birth_11	0,000	0,000	0,000	0,000	0,000	0,000	0,061	0,046	0,008	-0,517
regional_dummy_foundation_11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,008	-0,069	-1,363
regional_dummy_Birth_12	0,000	0,000	0,000	0,000	0,000	-0,154	-0,391	-0,653	-0,779	-1,184
regional_dummy_foundation_12	0,000	0,000	0,000	0,000	0,000	0,000	-0,206	-0,369	-0,422	-1,713
regional_dummy_Birth_13	0,000	0,000	0,000	0,000	0,000	0,000	-0,189	-0,335	-0,418	-0,839
regional_dummy_foundation_13	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-1,178
regional_dummy_Birth_14	0,000	0,000	0,000	0,000	0,000	-0,302	-0,725	-1,019	-1,166	-1,504
regional_dummy_foundation_14	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,971
regional_dummy_Birth_15	0,000	0,000	0,025	0,105	0,155	0,160	0,182	0,115	0,053	-0,168
regional_dummy_foundation_15	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-1,205
regional_dummy_Birth_16	0,000	0,000	0,000	0,000	0,001	0,070	0,121	0,072	0,027	-0,220
regional_dummy_foundation_16	0,000	0,000	0,000	0,000	0,000	0,000	-0,038	-0,260	-0,338	-1,601
regional_dummy_Birth_17	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,224
regional_dummy_foundation_17	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,039	-1,043
regional_dummy_Birth_18	0,000	0,000	0,000	0,000	0,000	0,000	0,072	0,166	0,197	0,149
regional_dummy_foundation_18	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,123	-0,181	-1,137
regional_dummy_Birth_19	0,000	0,000	0,000	0,002	0,135	0,255	0,405	0,492	0,512	0,415
regional_dummy_foundation_19	0,000	0,000	0,000	0,000	0,000	0,041	0,168	0,199	0,204	-0,761

Table A.8 Lambda Cross Validation of model 3 with decomposed version of EE index

Lambda	**RMSE**
0.4000	84.49247
0.3000	80.97842
0.2000	78.95541
0.1000	78.47412
0.0750	80.34562
0.0500	83.16648
0.0250	83.44096
0.0100	87.93589
0.0050	90.02294
0.0001	101.51646
0.0000	101.98145