UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Padua Research Archive - Institutional Repository

The role of industry variety in the creation of innovative start-ups in Italy

(Article begins on next page)

CrossMark

The role of industry variety in the creation of innovative start-ups in Italy

Roberto Antonietti · Francesca Gambarotto

Abstract This paper aims to ascertain whether related and unrelated industry variety affects the creation of innovative as opposed to other start-ups in Italian local labor market areas. The analysis combines elements from the knowledge spillover theory of entrepreneurship, the recombinant growth approach, and evolutionary economic geography. Using data on Italian innovative start-ups created between 2012 and 2015, and on firms newly registered with the Italian Chambers of Commerce, and applying appropriate count data models, our estimates show that innovative start-ups are more frequently created in areas where unrelated variety is higher. This is because innovative start-ups find more opportunities to recombine different pieces of knowledge or maximize their portfolio of demand opportunities, in such a setting, whereas a higher related variety stimulates the creation of other types of new start-up, for which it is easier to combine similar, complementary knowledge sources. We also find that half of the effect of related and unrelated variety comes from the localization of (innovative) start-ups in large urban areas.

R. Antonietti (✉) · F. Gambarotto
"Marco Fanno" Department of Economics and Management, University of Padova, Via del Santo 33, 35123 Padova, Italy
e-mail: roberto.antonietti@unipd.it

F. Gambarotto
e-mail: francesca.gambarotto@unipd.it

1 Introduction

Innovative start-ups[1] are fundamental to innovation systems. They use technologies that increase productivity and explore consumer needs to produce new products. They stimulate inventiveness and competition, especially when conducive to radical innovations. They can be disruptive, or simply more efficient in producing better products or services, locally contributing to the economy's dynamism.

New businesses can affect employment and regional development in several ways. Start-ups may widen existing markets or create new ones (Audretsch 1995), generate a greater variety of products and problem solutions, accelerate structural change by competing with incumbent firms, and stimulate productivity by challenging established market positions. Start-ups can also have important regional effects on employment and economic growth (Fritsch and Weyh 2006; Fritsch and Mueller 2008). Understanding the local factors behind (innovative) start-up creation is consequently an important policy issue.

European economies were often based on systems devised more to support the search for scale economies and standardized production processes than to sustain

---

[1] A start-up can be defined as innovative when the entrepreneurial talent is used to recombine existing knowledge in innovative ways, especially new technologies, and when it invests in risky activities like R&D and patent production (Acs et al. 2009).

young creative firms.[2] The Smart Specialization Strategy adopted by the European Union strongly focuses instead on the drivers of the entrepreneurial discovery process, one of the most important of which is knowledge spillover.

The knowledge-spillover theory of entrepreneurship sees unexploited (local) knowledge as the main determinant of new firm creation (Audretsch et al. 2006; Acs et al. 2009). Particular importance is attributed to the opportunities left unexploited by local incumbent agents and to the amount of knowledge involved. According to another line of research, the recombinant knowledge approach, the type of knowledge that can be recombined matters too, as well as the amount (Bae and Koo 2008; Bishop 2012; Colombelli 2016; Colombelli and Quatraro 2017). In particular, areas with a greater variety of knowledge sources are more favorable for the generation of new ideas and their commercialization.

According to recent evolutionary economic geography literature (Frenken et al. 2007), we can distinguish between a related (within-industry) variety and an unrelated (between-industry) variety. The former favors new ventures in nearby knowledge areas, and it exploits network externalities to reduce investment risks and expand new business opportunities. The latter stimulates new firm formation through the exploration and recombination of a very diverse array of knowledge sources. A greater local knowledge diversity suggests more entrepreneurial opportunities, though they may be more risky and uncertain. The existing empirical literature on industry variety and entrepreneurship (Bae and Koo 2008; Bishop 2012; Colombelli 2016) does not adequately investigate the different influence of related and unrelated variety on different types of new firm. It only considers either generically defined start-ups or innovative start-ups, but these two types of business can be of a very different nature, or their creation would demand a different combination of knowledge sources.

This paper tries to fill this gap and extend the small business economics literature in two ways. First, we argue that related and unrelated variety affects different types of new business in different ways: unlike other types of new firm, innovative start-ups should be created more frequently where unrelated variety is higher. We investigate this relationship on a high level of geographical disaggregation, considering functional units based on actual travel-to-work flows rather than administrative regions. In so doing, we also distinguish large metropolitan areas from elsewhere, and we explore whether these areas have a specific role in attracting (innovative) start-ups.

Second, we provide additional support for recent efforts in the evolutionary economic geography literature to explain the drivers of regional diversification. Since Frenken et al. (2007), related variety and unrelated variety have revealed two distinct roles. Related variety stimulates a growth in employment because it improves the chances of new products or services being generated by combining technologically related activities. Unrelated variety, on the other hand, provides regions with a large portfolio of activities, thereby reducing the risk of further unemployment. Despite recent contributions, some points remain to be explained (Content and Frenken 2016). One concerns whether, and why, regions with a great unrelated variety can also yield product innovations, especially those with a radical content. We attribute this to the creation of innovative start-ups: the chance to recombine very different pieces of knowledge, plus the availability of a diversified portfolio of potential fields of application, provides the most fertile terrain for generating and commercializing radical innovations.

This paper is developed as follows: Section 2 reviews the literature on knowledge spillover and start-up creation; Section 3 presents the datasets used for our empirical analysis, the variables, and the estimation strategy; Section 4 discusses the results; and Section 5 concludes.

2 Relevant literature

Innovation is a social activity that demands the ability to recombine ideas, an entrepreneurial capacity to convert knowledge into new commercial products or services, and a favorable social milieu where profit-driven behavior and social value accumulation overlap (Cooke 2016; Kirzner 1997; Tödling et al. 2011).

---

[2] This is particularly true for Italy, which is still suffering from a lengthy period of economic stagnation, with a negative impact on the birth of innovative firms. In 2014, high-growth firms accounted for 6.8% of Italy's all active enterprises, and for 9.5% of the corresponding employment (the corresponding mean values for the EU are 9.2% and 13%, respectively), and they had 709,769 employees (as opposed to 2,961,954 in the UK, for instance, where high-growth firms account for 12.9% of all active firms and for a 19.3% of the corresponding employment). See: http://ec.europa.eu/eurostat/documents/2995521/7706167/4-26102016-AP-EN.pdf/20f0c515-ed43-45c3-ad6a-ca0b26b36de5.

The ability of entrepreneurs to generate and commercialize new ideas relies on the availability of local resources, such as physical, human, and financial capital, or transport and digital infrastructure. According to the knowledge spillover theory of entrepreneurship (Audretsch 1995; Acs et al. 2009), knowledge spillovers are the most important of all these resources, especially those originating from knowledge opportunities left unexploited by incumbent agents. Taking this approach, the public nature of knowledge implies that greater amounts of knowledge will coincide with more opportunities for knowledge to spill over from incumbent to new activities, and therefore with a higher likelihood of new firms being generated. This is particularly true when knowledge sources and new firms are in close proximity (Audretsch and Lehmann 2005).

Starting with Glaeser et al. (1992), the empirical literature identifies various ways in which knowledge spillovers can materialize. One is through the colocation of firms and universities (Anselin et al. 2000; Audretsch et al. 2004; Bonaccorsi et al. 2014; Ghio et al. 2016). Others emerge from interactions between local human capital and entrepreneurs (Acs and Armington 2004; Glaeser et al. 2010), from high concentrations of private R&D departments (Audretsch and Feldman 1996; Wieser 2005; Hall et al. 2010), and because of the density of economic activities, taken as a proxy for urbanization economies (Carlino et al. 2007).

Other studies have stressed the heterogeneous nature of knowledge and suggested that we should consider the nature of the local knowledge stock, rather than its size, when linking entrepreneurship with economic development. According to the recombinant growth model (Weitzman 1998), economic growth originates neither from the total amount of knowledge available, nor from the ability to generate new ideas, but from the ability to recombine, or cross-pollinate, an ever-increasing quantity of fruitful ideas.

Frenken et al. (2007) develop the concept of "related variety" to investigate what types of connection affect innovation and economic development. Drawing on Jacobs (1969), innovation is conceived as a recombinant process that "*necessarily builds on a pre-existing variety of knowledge and artefacts that are being combined in new ways, leading to new products and services*" (Content and Frenken 2016, p. 3). Since then, many studies have examined the different influence of related and unrelated variety on economic outcomes. In general, related variety emerges as a driver of growth in employment and export diversification, while unrelated variety helps to contain unemployment through the diversification of a region's industry portfolio.

This literature leaves some questions unanswered, however, particularly as concerns how radical innovation can originate from unrelated variety (Content and Frenken 2016; Boschma 2017). The question is relevant because relatedness is generally considered more helpful for the purpose of recombining knowledge into new commercial products, given the greater knowledge spillover stemming from complementarities and shared competences. Diversity can also be relevant, however, because the existence of unrelated knowledge sources leaves room for a creative recombination of ideas (Castaldi et al. 2015). This means that unrelated variety can potentially stimulate radical innovation and structural change in the process (Neffke et al. 2014).

The direct link between related/unrelated variety and local entrepreneurship was first analyzed in Bae and Koo (2008). Relying on the Schumpeterian distinction between invention and commercialization, they posited that relatedness and diversity have a different influence on incumbent and nascent entrepreneurs. While incumbent firms are better endowed with financial and organizational resources, so they can exploit existing knowledge spillovers more efficiently for their commercial purposes, new firms (or inventors) can benefit more from the diversity of the accessible knowledge thanks to a highly diversified local demand and a higher likelihood of their recombining existing knowledge sources to generate new products and services. The authors' estimates for the US electronic components and communication equipment sector confirmed that the number of new start-ups increases proportionally with both relatedness and local knowledge diversity.

Bishop (2012) analyzed the relationship between knowledge diversity and entrepreneurship in local authority districts in Great Britain, finding again that both related and unrelated knowledge diversity positively affected the birth rate of new firms.

Merging data from the innovative start-ups directory with patent information at NUTS3 regional level, Colombelli (2016) investigated whether the characteristics of the local knowledge base affected the number of innovative start-ups. The knowledge base was defined in terms of size, related versus unrelated variety, coherence, and cognitive distance. The results show that innovative newcomers benefit from locally available unexploited technological knowledge, and both the

related and the unrelated variety of the local technologies have a positive impact on the generation of innovative start-ups.

The above-mentioned works considered either the total number of start-ups or only the innovative start-ups as the dependent variable, but these two entities differ and their generation process can be affected by the composition of local knowledge in different ways. Concerning the Italian case, Finaldi Russo et al. (2016) showed that, by comparison with other start-ups, the innovative start-ups are smaller, their product commercialization rate is lower, they make more intensive use of intangible assets, they have a greater liquidity, and a stronger propensity for investment, a lower profitability and cash flow, but a higher sales growth. These differences are smaller, but still significant, even when innovative start-ups are compared with other start-ups in high-tech sectors. Using this information, the authors conclude that "*innovative start-ups are presumably pursuing truly new projects that require time to reach the commercialization phase*" (Finaldi Russo et al. 2016, p. 13).

We posit that related and unrelated variety might have two distinct effects on the generation of new activities. In line with previous literature, we would generally expect the number of new start-up firms to be larger where both related and unrelated variety are higher. We also expect related variety to be more relevant because it is easier for new activities to combine local complementary knowledge sources. Unrelated variety should matter more in the creation of innovative start-ups: on the supply side, their invention-based activity relies on recombining very different knowledge inputs; on the demand side, their proliferation should be facilitated where the variety of unexploited demand opportunities is greater, and where the portfolio of potential applications and customers is highly diversified, thus minimizing the business risk typical of highly innovative activities.

Our paper complements the analyses conducted by Bae and Koo (2008), Bishop (2012), and Colombelli (2016), and extends them in two directions. First, it distinguishes between the effects of related versus unrelated variety on innovative start-ups as opposed to other types. Second, it uses a finer territorial unit of analysis, namely the local labor market area (LLMA), defined according to actual travel-to-work flows rather than administrative rules.

The context of analysis is Italy, where a new law introduced in 2012 identified innovative start-ups as young, small firms with a strong commitment to research and innovation. Italy is also an interesting scenario because of its marked geographical variability in start-up creation and distribution of knowledge sources.

3 Empirical analysis

3.1 Data

Data on the number of innovative start-ups were obtained from the registers of the Italian Chambers of Commerce and the Italian Ministry for Economic Development, and specifically from the online directory of "innovative start-ups." The definition of innovative start-up was established in the Italian Legislative Decree n. 221/2012 (the so-called "Growth 2.0" decree). To be considered an innovative start-up, a firm[3] has to meet a number of specific requirements. It must have an annual turnover of less than 5 million Euro, be resident in Italy, and have been active for less than 48 months (60 months since Legislative decree n. 3/2015). Most of the social capital must be owned by individuals, and must not pay dividends. It cannot be the outcome of a merger or acquisition, and it must focus on the generation and/or commercialization of new products or services of high technological value. The innovative start-up also has to satisfy at least one of the following additional criteria: a significant proportion (at least one third) of its employees must be highly qualified (with a PhD or Master's degree); and it must spend at least 15% of its budget on R&D,[4] or own at least one patent, license, or original computer program. The benefits for companies registered as innovative start-ups include cheaper and easier administrative start-up procedures, tax benefits for investors in their equity, zero-interest rate loans from public agencies, the chance to use flexible employment contracts, tax credits on highly skilled personnel, support for internationalization strategies, and easier failure procedures.

---

[3] The applicant firm can be a new or already existing company, but few innovative start-ups were born before 2012; the majority were established and registered with the Chambers of Commerce between 2012 and 2015.

[4] The rule is the following: R&D expenditures should be equal at least to 15% of the highest value between the total production cost and the total value of production.

We consider the number of innovative start-ups active in Italian LLMA between December 2012 and May 2015. LLMA are identified by the Italian Statistical Institute (ISTAT) using an algorithm based on actual travel-to-work flows. The two main advantages of using LLMA are that they enable a more precise measurement of spatially bounded knowledge spillovers, and they span different regions and provinces instead of reflecting strict administrative borders. Using 2011 population census data, the ISTAT identified 611 LLMA.

Our sample includes 3883 innovative start-ups. As mentioned in Section 2, Finaldi Russo et al. (2016) showed that they differ significantly from other start-ups in terms of size, commercialization, and performance: innovative start-ups are essentially still inventing, while other start-ups are more oriented towards commercializing their invention. Table 1 shows the industry distribution of innovative start-ups. Almost 80% of them are involved in knowledge-intensive business services, such as computer-related and professional activities. In the manufacturing sector, they belong largely to the medium-high-tech and high-tech industries (according to the OECD classification).

Examples of Italian innovative start-ups that combine elements from very different sectors include the following: firms that produce smart metering systems (nanotechnologies used in the housing, energy, software, and engineering sectors), and software applications for virtually managing queues in public offices; firms that use fruit waste to produce textiles; firms producing drones for the monitoring of vineyards and farmland.

Table 1 Distribution of innovative start-ups by industry and NUTS 1 region

Industry (NACE rev. 2)	N	%
A—Agriculture and fishing	13	0.33
C—Manufacturing	647	16.67
- Low-tech manufacturing	*129*	*19.94*
- Medium-low tech manufacturing	*68*	*10.51*
- Medium-high tech manufacturing	*215*	*33.23*
- High-tech manufacturing	*235*	*36.32*
D + E—Energy	60	1.55
F—Building	42	1.08
G-S—Services	3102	79.89
- J—Information and communication services	*1627*	*52.45*
- M—Professional activities	*1096*	*35.33*
Total	3883	100.0

Data on other start-ups come from the Movimprese archives managed by the Italian Chambers of Commerce. This dataset provides yearly information on the stock of existing and newly registered firms and shutdowns in Italy. For the 2012–15 period, information was collected on the number of firms newly-registered each year in Italy's municipalities,[5] the sum of which gave us the stock of newly registered firms for each municipality over the whole period. Then, the municipalities were pooled into LLMA using a conversion table provided by the ISTAT.

Figure 1 shows the geographical distribution of innovative start-ups (left map), as compared with that of other firms newly registered in Italy (right map). Both types of start-up are widespread all over the country, but slightly more concentrated in the north, especially in Lombardy and Emilia-Romagna. It is noteworthy that, compared with the other start-ups, the innovative start-ups are more concentrated in the largest metropolitan areas of the country, such as Milan, Rome, Naples, and Turin. In other words, innovative start-ups are an urban phenomenon.[6]

3.2 Model and variables

The model used for our estimations is as follows:

$$N_{i2012-15} = \beta_0 + \beta_1 RV_{i2011} + \beta_2 UV_{i2011} + X'_{i2011}\beta_3 + \mu_r + \varepsilon_{i2012-15} \qquad (1)$$

where, in a first specification, N is the number of innovative start-ups (*NISU*) in the LLMA i during the period 2012–2015, then, in a second specification, it is the number of other start-ups (*NSU*) located in LLMA i during the same period of time, after discounting the number of innovative start-ups.[7] The terms RV and UV represent the related and unrelated variety in 2011, while X is a vector of additional variables observed at LLMA level and measured in the census year 2011.

Related and unrelated variety indicators are taken from Frenken et al. (2007). The former measures the

---

[5] Unfortunately, the information available does not allow for a distinction between the sectors involved, so the analysis focuses on the number of firms newly registered in each LLMA.

[6] The pairwise correlation between the number of innovative start-ups and the number of newly-registered firms is 0.92.

[7] As a robustness test, Eq. (1) was also estimated using the gross number of new start-ups as the dependent variable. The results did not change.
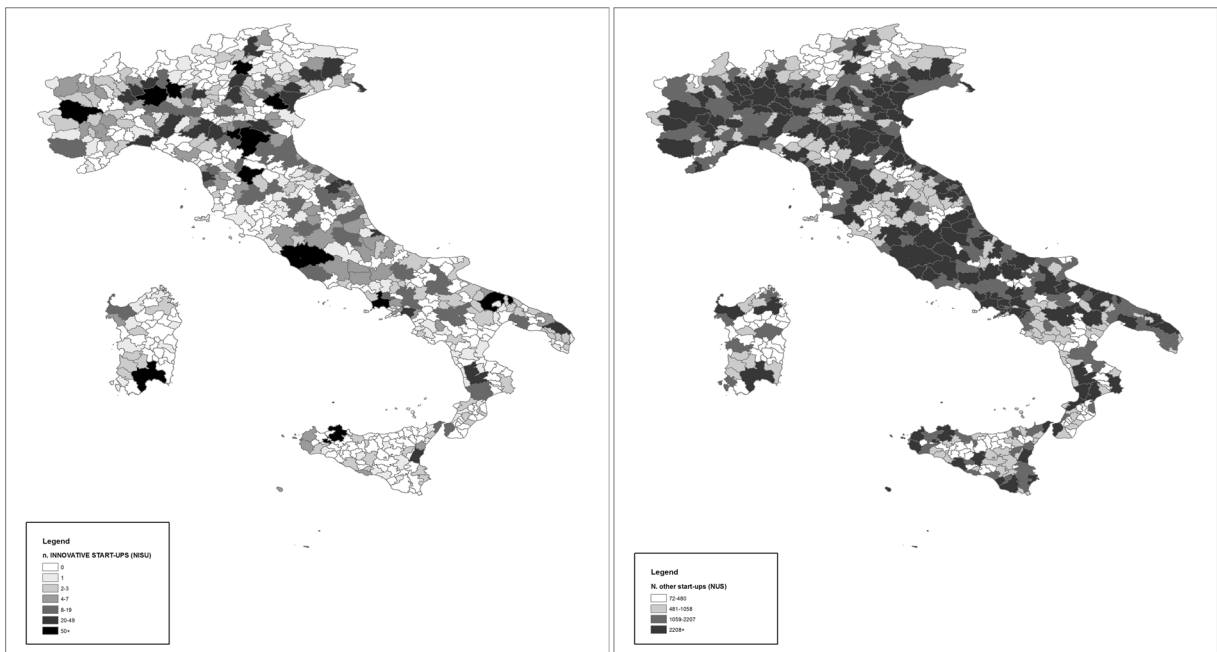
Fig. 1 Map of the geographical distribution of innovative start-ups (left) and of other start-ups in Italy (right)

weighted sum of the entropy within each two-digit industry in a LLMA and captures knowledge spillovers between firms producing and selling related products and services. The latter captures the degree of entropy between the two-digit sectors and is a measure of industry diversification at LLMA level:

$$RV_i = \sum_{j=1}^{J} P_j H_j \qquad (2)$$

where P_j represents the two-digit employment shares $P_j = \sum_{k \in S_j} p_k$, k is the five-digit industry falling under the two-digit industry S_j ($j = 1 \dots J$), and p_k represents the five-digit employment shares, and:

$$H_j = \sum_{k \in S_j} \frac{p_k}{P_j} \log_2 \left(\frac{P_j}{p_k} \right); \qquad (3)$$

$$UV_i = \sum_{j=1}^{J} P_j \log_2 \left(\frac{1}{P_j} \right).$$

The following variables are included in vector X. First, we control for the size of the LLMA using the number of incumbent plants in year 2011 (# *PLANTS*). We prefer to add this variable on the right-hand side of Eq. (1), instead of using it as the denominator of the

dependent variable N, in order to clarify the magnitude and statistical significance of the size effect on N. We expect both types of start-up to be located in larger LLMA, which are characterized by a higher local demand and a greater presence of local suppliers and potential knowledge sources.

Second, we control for the level of human capital in the area. We use two variables: the first is a dummy that takes a value of 1 if there is a university within the LLMA (*UNIV*); the second is the share of employees holding a university degree (*HK*). LLMA with a better-qualified human capital should generally be an ideal ecosystem for the generation and proliferation of innovative start-ups. This happens for two reasons: because a university acts as an incubator of innovative start-ups and spin-offs (Ghio et al. 2016); and because of the availability of a highly-qualified workforce that can create new, innovative activities (possibly after registering a patent), or serve as a pool of specialized labor that innovative entrepreneurs can recruit. We consequently expect both variables *UNIV* and *HK* to correlate positively with *NISU* and *NSU*. We also include a dummy for the presence of incubators in the LLMA (*INCUBA-TOR*), and we expect it to positively affect *NISU* (Colombelli 2016).

The capability of a local area to generate new (innovative) firms may also depend on its degree of

trade openness: areas where imports exceed exports may suffer employment and business losses because of foreign competition, whereas areas where exports exceed imports may benefit from new business opportunities (Autor et al. 2013; Donoso et al. 2015). Using readily available information provided by the ISTAT for the census year 2011, we define two dummy variables: one takes a value of 1 if the LLMA is a net importer, i.e. if it imports more goods and services than it exports (IMPORT), while the other takes a value of 1 if the LLMA is a net exporter of goods and services (EXPORT). A dummy that takes a value of 1 when imports equal exports represents the term of reference.

We also include the local unemployment rate (UNEMP) in 2011, computed as the proportion of unemployed individuals out of the total labor force in the LLMA. Its impact on N is ambiguous (Audretsch and Vivarelli 1996; Bishop 2012). On the one hand, higher unemployment implies a human resource potential that could be the target of regional entrepreneurship policies, and higher unemployment can also lower the opportunity costs of becoming an entrepreneur, so we might expect a positive relationship between UNEMP and N. On the other hand, LLMA with higher unemployment rates could be a sign of economically depressed areas, which would be unfavorable for the birth of (innovative) start-ups to grow due to a lack of resources, in which case a negative correlation between UNEMP and N could be expected.

Another two attributes to consider are the quantity and the spatial dimension of the relationships occurring within each LLMA. The former is captured by an index (FLOWS) that measures relational intensity, provided by the ISTAT: this value is the percentage of (commuting) flows that connect different municipalities within a LLMA (after discounting the commuters who live and work within the LLMA) out of the total possible flows. The index varies between 0 (i.e., the case of a LLMA where nobody commutes across municipalities) and 1 (when everyone commutes outside their municipality of residence), so the higher the index, the larger the proportion of people circulating within a LLMA. This variable can consequently capture the quality of the local transport system and/or better job matching opportunities, so we would expect a positive correlation between FLOWS and N, i.e., there should be more innovative start-ups in the more dynamic areas, where people move around more easily, find or change job and exchange ideas.

The latter attribute is captured by means of an index of LLMA self-containment (SELF), provided by the ISTAT, which amounts to the minimum value between a self-containment index on the labor demand side (SELF_D), and one on the labor supply side (SELF_S). The first is the ratio of people living and working in the LLMA (after discounting those who work at home, the homeless and those who work in other countries) to the total number of people who work in the same LLMA (again after discounting those who work at home, the homeless, and those working in other countries). The second is the ratio of people living and working in the LLMA (after discounting those who work at home, the homeless, and those working in other countries) to the total number of people who live in the same LLMA (after discounting those who work at home, the homeless, and those who work in other countries). SELF amounts to the local area's minimum amount of self-containment: the higher the index, the more it can be considered a "market," where production, consumption and social activities are spatially concentrated. Such a variable should correlate positively with N: a very self-contained LLMA should have a higher concentration of potential market opportunities than a scarcely self-contained LLMA.

We also add the population density of the LLMA (DEN), which we use to capture urbanization economies. Denser urban areas should stimulate the creation of innovative activities because they act as incubators during the earliest stages of their development (Duranton and Puga 2001), or due to the spatial concentration of innovation inputs such as R&D laboratories, pools of scientists, financial capital, and public services (Carlino et al. 2007). According to Jacobs (1969), denser areas offer better chances of cross-fertilization among a variety of different knowledge sources. For all these reasons, we would expect a positive correlation between DEN and N.

To distinguish cultural from industry variety, we use the share of foreign citizens (FOR) living in the LLMA: the higher this share, the greater the cultural diversity of the LLMA, and the higher the consequent chances of new businesses being created. The main reasons lie in a higher possibility for cross-fertilization of ideas in culturally diversified environments and a higher divergence in the appraisal of new projects that provide an incentive for individuals to start a new venture (Jacobs 1969; Audretsch et al. 2010). Another plausible explanation could be that of ethnic segregation: if some ethnic

groups are discriminated in the local labor market, it can be that opening a new activity is driven by necessity rather than by an entrepreneurial strategy. Alternatively, this could happen if ethnic groups are characterized by an average lower education than domestic entrepreneurs are. If this is the case, we should find *FOR* to be relevant only for *NSU*, but not for *NISU*.

Finally, we include two variables measuring the specialization of the LLMA in manufacturing activities (*SPEC MAN*), and knowledge-intensive business services (*SPEC KIBS*) (Bishop 2012). We capture specialization using the location quotient for each industry, computed as the ratio between the share of employment in manufacturing (KIBS) in each LLMA, and the share of employment in manufacturing (KIBS) in Italy in 2011. The higher the location quotient, the greater the specialization of the LLMA in manufacturing and KIBS, respectively. Both types of activity can generate knowledge spillovers. On the supply side, a high share of manufacturing and KIBS employment can be a proxy for the presence of a dense network of local suppliers and knowledge sources. On the demand side, a high specialization in manufacturing or KIBS could help new firms to benefit promptly from economies of scale in production thanks to the presence of a large mass of potential customers. We therefore expect both variables to correlate positively with *N*, although the literature seems to emphasize the role of KIBS, rather than of manufacturing, in generating knowledge spillovers (Doloreux and Shearmur 2012).

To enable comparisons between the average marginal effects, we standardized each continuous variable at zero mean and unit standard deviation. Tables 2 and 3 show the summary statistics of all our variables, and their pairwise correlations, respectively.

Finally, we include 19 regional dummies (μ_r) at NUTS 2 level to control for region-specific fixed effects related to regional institutional quality, among other things, or to being a target of national or European industrial policies.

3.3 Empirical strategy

When estimating Eq. (1) for *NISU*, two problems arise. First, since *NISU* is the discrete, and non-negative, number of innovative start-ups located in each LLMA, we cannot estimate Eq. (1) using Ordinary Least Squares (OLS). Second, we have 261 LLMA (42.72% of the sample) with zero innovative start-ups in the

Table 2 Summary statistics

	Mean	Std. dev.	Min	Max
NISU	6.355	30.02	0	581
NSU	2475.3	7621.6	72	114,716
# PLANTS	7868.1	23,532.2	214	385,491
UNIV	0.095	0.294	0	1
HK	0.084	0.023	0.031	0.176
INCUBATOR	0.043	0.202	0	1
IMPORT	0.501	0.500	0	1
EXPORT	0.249	0.433	0	1
UNEMP	0.119	0.061	0.015	0.275
FLOWS	0.257	0.145	0.002	0.661
SELF	0.757	0.074	0.573	0.941
FOR	6591.2	22,080.8	12	370,018
DEN	205.67	292.25	10.39	3104.9
SPEC MAN	0.970	0.538	0.078	2.750
SPEC KIBS	0.576	0.263	0.074	1.961
RV	2.223	0.333	0.985	2.980
UV	4.622	0.391	2.939	5.390

reference period. To cope with the first issue, we estimate Eq. (1) using a count data model—a negative binomial model to be specific—which enables a solution to be found for the problem of data over-dispersion that arises when the variance of the observed distribution of the count variable is larger than the mean. The second issue demands the use of either a zero-inflated (ZINB), or a hurdle (HNB) version of the negative binomial. Both models enable a distinction between the process that generates the excess of zeros and the process that generates the positive outcomes, but the two differ in the way in which the nature of the zeros is interpreted.

The ZINB assumes that the zeros may originate from "sampling" or be "structural," the former meaning that they occur by chance, while the latter are due to a particular structure of the data, and are therefore not random. In our case, it is as if a LLMA were to remain without any innovative start-ups for some random reason, or if it were unable to host any innovative start-ups for some specific reason. The ZINB requires a logit estimate to predict the excess of zeros and a negative binomial estimate to predict the positive outcomes. The HNB model assumes instead that all zeros are structural, while the positive outcome originates from sampling and follows a truncated negative binomial distribution.

Table 3 Correlation matrix

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
[1] # PLANTS	1														
[2] UNIV	0.46	1													
[3] HK	0.36	0.58	1												
[4] INCUBATOR	0.51	0.43	0.34	1											
[5] IMPORT	−0.18	−0.09	−0.12	−0.18	1										
[6] EXPORT	0.22	0.10	0.04	0.22	−0.58	1									
[7] UNEMP	−0.05	−0.04	−0.10	−0.11	−0.55	−0.44	1								
[8] FLOWS	0.21	0.19	0.15	0.26	−0.50	0.43	−0.49	1							
[9] SELF	0.19	0.28	0.26	0.16	0.28	−0.22	0.11	−0.07	1						
[10] FOR	0.13	0.09	0.15	0.16	−0.60	0.54	−0.70	0.40	−0.24	1					
[11] DEN	0.49	0.30	0.28	0.29	−0.20	0.17	0.14	0.16	−0.07	0.05	1				
[12] SPEC MAN	−0.01	−0.07	−0.13	0.04	−0.66	0.59	−0.43	0.43	−0.40	0.53	0.01	1			
[13] SPEC KIBS	0.49	0.48	0.57	0.41	−0.31	0.27	−0.26	0.31	0.12	0.29	0.28	0.07	1		
[14] RV	0.25	0.27	0.27	0.16	−0.06	0.09	0.28	0.15	0.01	0.00	0.37	−0.08	0.36	1	
[15] UV	0.32	0.37	0.37	0.27	−0.45	0.31	−0.18	0.43	−0.01	0.33	0.29	0.27	0.55	0.58	1

The HNB implies a logit estimate for the probability of a non-zero observation, and a separate truncated negative binomial estimate to explain the positive outcomes.

The choice between the two models is based on traditional information criterion tests, like the AIC or BIC. The two models often produce very similar results, however, so the choice is based on reasons of convenience, without any strong theoretical justification. Table 4 shows that the performance of the two models is very similar, with the ZINB performing slightly better. It is hard to say for sure whether a LLMA randomly or deliberately chose to have no innovative start-ups between 2012 and 2015. We consequently opt for the ZINB model for our estimates.[8]

Since we observe no zeros, we use a standard negative binomial estimator to compute Eq. (1) for *NSU*.

Another issue is endogeneity. We rely on the fact that the Italian law on innovative start-ups was adopted at the end of 2012, while almost all of our innovative start-ups were established or registered with the Chambers of Commerce after 2012, and our regressors are all measured in 2011. We can consequently interpret the introduction of the legislation as a sort of policy shock, so any reverse causality between *NISU* (or *NSU*) and our measures of industry variety should be mitigated. In any case, our results should be considered more in terms of robust correlations rather than causal effects.

Finally, we control for the presence of multicollinearity between the dependent variables by estimating Eq. (1) with a linear probability model and quantifying the variance inflation factor (VIF).

4 Results

Table 5 shows the results of our estimations. The second column refers to the negative binomial estimate of Eq. (1) on *NSU*, while the third and fourth concern the ZINB estimates for *NISU*.

We can see from the second column that the estimated coefficients of *RV* and *UV* are both positive and highly significant. In line with previous research (Bishop 2012), new start-up firms are more common where both related and unrelated variety are higher. In line with our expectations, we find the average marginal effect of *RV* much higher than that of *UV*. A unit increase

[8] Hurdle negative binomial estimates are not reported due to word count restrictions, but are available upon request.

Table 4 Choice of model: ZINB versus HNB

Dependent variable: *NISU*	AIC	BIC
ZINB	2110.91	2424.38
HNB	2117.41	2430.88

Table 5 The impact of related and unrelated variety on the number of start-ups (NSU) and of innovative start-ups (NISU)

ESTIMATION METHOD:	NSU	NISU	
	Negative binomial	Zero inflation	Negative binomial
RV	0.342***	0.808	0.113
	(0.034)	(0.864)	(0.077)
Average marginal effect	1312.67		
UV	0.173***	1.150	0.349***
	(0.041)	(2.043)	(0.128)
Average marginal effect	665.57		2.541
# PLANTS	0.203**	−41.42*	0.078**
	(0.095)	(22.33)	(0.036)
UNIV	0.342***	2.956	0.613***
	(0.094)	(2.003)	(0.131)
HK	0.123***	−0.457	0.330***
	(0.035)	(1.186)	(0.077)
INCUBATORS	−0.022	4.588	0.441***
	(0.103)	(4.369)	(0.146)
IMPORT	−0.310***	−3.237	−0.297**
	(0.073)	(4.894)	(0.141)
EXPORT	0.160**	0.905	0.300**
	(0.065)	(0.876)	(0.115)
UNEMP	0.484***	−0.673	0.389
	(0.071)	(1.969)	(0.230)
FLOWS	0.142***	0.553	0.242***
	(0.028)	(1.824)	(0.058)
SELF	0.167***	1.225	0.252***
	(0.024)	(0.813)	(0.047)
FOR	0.128***	−0.951	0.126
	(0.037)	(2.924)	(0.146)
DEN	0.157***	0.506	0.109
	(0.049)	(1.216)	(0.078)
SPEC MAN	0.078*	−1.631	0.111
	(0.043)	(2.547)	(0.092)
SPEC KIBS	0.054*	−0.617	0.154***
	(0.028)	(0.944)	(0.048)
Regional dummies	Yes	Yes	Yes
N obs.	611	611	611
Nonzero obs.	611	350	350
α (over-dispersion)	0.196***	0.174***	
Vuong test (ZINB Vs NB)		5.04***	
Pseudo R^2	0.133		
VIF test			
RV (no regional dummies)		2.75 (2.23)	
UV (no regional dummies)		3.07 (2.76)	
Mean (no regional dummies)		6.10 (2.20)	

All continuous variables are standardized at zero mean and unit standard deviation. All the estimates include a constant term. Standard errors are clustered at LLMA level

***Significant at 1% level

**Significant at 5% level

*Significant at 10% level

in RV is related to an average increase of 1313 new firms, whereas the marginal effect of UV is around 666 newly registered firms.

Among the controls, we find NSU higher in larger LLMA, or in those containing a university and a larger endowment of human capital. There is also a higher NSU where exports exceed imports, confirming that import competition can be an obstacle to new firm creation. There are more new start-ups where unemployment is higher, where the LLMA is self-contained and characterized by intense travel-to-work flows. In line with previous research, a greater degree of cultural diversity correlates with a larger stock of new firms. Finally, there is a weak positive relation between NSU and specialization in both the manufacturing and the KIBS sectors.

A different picture emerges from the estimates for NISU. First, we can see from the third column that only one variable, the small size of the LLMA, explains the excess of zeros. Looking at the negative binomial estimates (in line with our hypothesis), we find that only the coefficient of UV is positive and statistically significant, whereas that of RV does not differ statistically from zero. So the number of innovative start-ups is affected by the amount of between-industry variety, but not by within-industry variety.

Unlike the case of NSU, the stock of innovative start-ups is larger when there is an incubator in the LLMA, and when the LLMA specializes in knowledge-intensive activities. NISU are insensitive to the local unemployment rate, urbanization economies, specialization in manufacturing activities, and cultural diversity. As explained in Section 3.2, this could be driven by ethnic discrimination or by the lower education of foreign-born entrepreneurs.

Concerning the multicollinearity issue, the mean VIF is 6.10, but the VIF for RV and UV is around 3. Much of this value is due to the inclusion of the regional dummies. When they are excluded from the estimates, all the VIF values decrease, and the mean VIF drops below 2.5. We can therefore rule out any multicollinearity between the regressors.

As a robustness test, we compute Eq. (1) again after excluding the largest metropolitan LLMA, namely Milan, Rome, Naples, Turin, Florence, Bologna, Venice, Genoa, Bari, Palermo, and Catania. The number of innovative start-ups in these LLMA ranges between 40 and 581, and the number of other start-ups between 13,099 and 114,402. The RV and UV levels are both

higher in these areas too, with mean values of 2.6 and 5.2, respectively, as opposed to 2.2 and 4.6 in the other LLMA. Table 6 shows that the results remain the same in qualitative terms, but the average marginal effect of *UV* on *NISU* drops to 1093, while the effects of *RV* and *UV* on *NSU* drop to 641 and 377, respectively. In other words, both related and unrelated variety still matter for new firm creation, but larger urban areas exhibit a significant multiplier effect: almost half of the impact of industry variety on local entrepreneurship is explained by the metropolitan nature of the LLMA. It is worth noting that, when large urban areas are left out of the sample, the presence of a university is no longer statistically significant, while unemployment becomes significant at 5% level. We surmise that a university's size, or quality, influence the likelihood of it generating innovative entrepreneurial activities, and that the positive relationship between unemployment and entrepreneurship applies particularly to small areas.

5 Conclusions

This paper investigates the phenomenon of innovative start-ups by looking for features of Italian LLMA that facilitate their creation. Using data registered by the Italian Chambers of Commerce and count data models, three main findings emerge from our analysis. First, new start-ups are more likely where local levels of related and unrelated variety are higher, and the former has a much stronger effect than the latter. New businesses generally focus on the commercialization of incremental innovations, and are the outcome of similar knowledge sources being recombined in different ways. Second, innovative start-ups focus more on the early development of breakthrough innovations and emerge where unrelated variety is higher. The chance to combine very diverse knowledge sources and to serve a diversified portfolio of customers makes these risky activities more likely to be profitable. Third, much of the effect of industry variety comes from the localization of (innovative) start-ups in large metropolitan areas, where industry variety is usually higher than elsewhere.

From a theoretical perspective, this article confirms that the nature of localized knowledge is an important driver of new firm creation. It also provides further evidence of the importance of different types of knowledge for different types of start-up. In doing so, we implicitly assume that regional characteristics, like

Table 6 The impact of related and unrelated variety on the number of start-ups (*NSU*) and of innovative start-ups (*NISU*): excluding large metropolitan areas

	NSU	NISU	
ESTIMATION METHOD:	Negative binomial	Zero inflation	Negative binomial
RV	0.270***	0.568	0.049
	(0.031)	(0.887)	(0.070)
Average marginal effect	*640.99*		
UV	0.159***	1.245	0.295***
	(0.037)	(0.760)	(0.092)
Average marginal effect	*377.27*		*1.093*
# PLANTS	1.417***	− 49.84**	0.715***
	(0.241)	(19.61)	(0.171)
UNIV	− 0.041	3.439	0.165
	(0.076)	(2.203)	(0.127)
HK	0.106***	− 0.021	0.416***
	(0.033)	(0.590)	(0.072)
INCUBATORS	− 0.252**	6.446*	0.407***
	(0.116)	(3.604)	(0.148)
IMPORT	− 0.208***	− 2.933	− 0.143
	(0.066)	(1.631)	(0.123)
EXPORT	0.072	1.262	0.222**
	(0.052)	(1.300)	(0.108)
UNEMP	0.418***	− 1.016	0.340**
	(0.061)	(1.215)	(0.153)
FLOWS	0.087***	0.068	0.188***
	(0.025)	(0.593)	(0.053)
SELF	0.087***	1.416**	0.177***
	(0.023)	(0.654)	(0.045)
FOR	0.120***	− 0.887	0.152
	(0.033)	(0.833)	(0.104)
DEN	0.157***	0.631	0.069
	(0.049)	(0.590)	(0.063)
SPEC MAN	0.075**	− 1.623*	0.142*
	(0.037)	(0.890)	(0.073)
SPEC KIBS	0.032	− 0.942	0.132***
	(0.025)	(0.881)	(0.045)
Regional dummies	Yes	Yes	Yes
N obs.	600	600	600
Nonzero obs.	611	339	339
α (over-dispersion)	0.196***	0.125***	
Pseudo R^2	0.122		

All continuous variables are standardized at zero mean and unit standard deviation. All the estimates include a constant term. Standard errors are clustered at LLMA level

***Significant at 1% level

**Significant at 5% level

*Significant at 10% level

related and unrelated variety, are capable to affect microeconomic decisions like that of starting a new, innovative, activity. In the absence of microeconomic information on intra-firm relationships, this could represent a limit of the analysis but also a promising avenue for future research.

As for policy considerations, these findings suggest that innovation policies that target knowledge creation and knowledge-intensive entrepreneurship should first try to generate a diversified portfolio of industries and technologies, rather than reinforcing existing specializations. By stimulating technological relatedness, the smart specialization policies adopted by the European Union can be useful in helping to generate start-ups. But policies should try to support knowledge diversification to facilitate the diffusion of innovations and benefit from their potential employment effects. The present findings also confirm that large metropolitan areas are important for the diffusion of innovation through the creation of start-ups: urban or regional policies aiming for an efficient scale of cities can also work indirectly as innovation-driving policies.

References

Acs, Z. J., & Armington, C. (2004). The impact of geographic differences in human capital on service firm formation rates. *Journal of Urban Economics, 56*(2), 244–278. https://doi.org/10.1016/j.jue.2004.03.008.

Acs, Z. J., Braunerhjelm, P., Audretsch, D. B., & Carlsson, B. (2009). The knowledge spillover theory of entrepreneurship. *Small Business Economics, 32*, 15–30. https://doi.org/10.1007/s11187-008-9157-3.

Anselin, L., Varga, A., & Acs, Z. J. (2000). Geographic and sectoral characteristics of academic knowledge externalities. *Papers in Regional Science, 79*(4), 435–443. https://doi.org/10.1111/j.1435-5597.2000.tb01766.x.

Audretsch, D. B. (1995). *Innovation and industry evolution*. Cambridge: MIT Press.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review, 96*(3), 630–640.

Audretsch, D. B., & Lehmann, E. (2005). Does the knowledge spillover theory of entrepreneurship hold for regions? *Research Policy, 34*, 1191–1202. https://doi.org/10.1016/j.respol.2005.03.012.

Audretsch, D. B., & Vivarelli, M. (1996). Determinants of new-firm startups in Italy. *Empirica, 23*, 91–105.

Audretsch, D. B., Lehmann, E., & Warming, S. (2004). University spillovers: does the kind of science matter? *Industry and Innovation, 11*(3), 193–206. https://doi.org/10.1080/1366271042000265375.

Audretsch, D. B., Keilbach, M. C., & Lehmann, E. E. (2006). *Entrepreneurship and economic growth*. Oxford: Oxford University Press.

Audretsch, D. B., Dohse, D., & Niebuhr, A. (2010). Cultural diversity and entrepreneurship: a regional analysis for Germany. *Annals of Regional Science, 45*(1), 55–85. https://doi.org/10.1007/s00168-009-0291-x.

Autor, D., Dorn, D., & Hanson, G. H. (2013). The China syndrome: local labor market effects of import competition in the United States. *American Economic Review, 103*, 2121–2168. https://doi.org/10.1257/aer.103.6.2121.

Bae, J., & Koo, J. (2008). The nature of local knowledge and new firm formation. *Industrial and Corporate Change, 18*(3), 473–496. https://doi.org/10.1093/icc/dtn017.

Bishop, P. (2012). Knowledge, diversity and entrepreneurship: a spatial analysis of new firm formation in Great Britain. *Entrepreneurship and Regional Development, 24*(7–8), 641–660. https://doi.org/10.1080/08985626.2011.617786.

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. https://doi.org/10.1007/s11187-013-9536-2.

Boschma, R. (2017). Relatedness as driver of regional diversification: a research agenda. *Regional Studies, 51*(3), 351–364. https://doi.org/10.1080/00343404.2016.1254767.

Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics, 61*(3), 389–419.

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies, 49*(5), 767–781. https://doi.org/10.1080/00343404.2014.940305.

Colombelli, A. (2016). The impact of local knowledge bases on the creation of innovative start-ups in Italy. *Small Business Economics, 47*, 383–396. https://doi.org/10.1007/s11187-016-9722-0.

Colombelli, A., & Quatraro, F. (2017). Green start-ups and local knowledge spillovers from clean and dirty technologies, *Small Business Economics*, Online First, 1–20, doi: https://doi.org/10.1007/s11187-017-9934-y.

Content, J., & Frenken, K. (2016). Related variety and economic development: a literature review. *European Planning Studies, 24*(12), 2097–2112. https://doi.org/10.1080/09654313.2016.1246517.

Cooke, P. (2016). The virtues of variety in regional innovation systems and entrepreneurial ecosystems. *Journal of Open Innovation: Technology, Market and Complexity, 2*, 3–19. https://doi.org/10.1186/s40852-016-0036-x.

Doloreux, D., & Shearmur, R. (2012). Collaboration, information and the geography of innovation in knowledge intensive business services. *Journal of Economic Geography, 12*(1), 79–105. https://doi.org/10.1093/jeg/lbr003.

Donoso, V., Martin, V., & Minondo, A. (2015). Do differences in the exposure to Chinese imports lead to differences in labour market outcomes? An analysis for Spanish provinces. *Regional Studies, 49*(10), 1746–1764. https://doi.org/10.1080/00343404.2013.879982.

Duranton, G., & Puga, D. (2001). Nursery cities: urban diversity, process innovation, and the life cycle of products. *American Economic Review, 91*(5), 1454–1477.

Finaldi Russo, P., Magri, S., & Rampazzi, C. (2016). Innovative start-ups in Italy: their special features and the effects of the 2012 law. *Politica Economica/Journal of Economic Policy, 32*(2), 297–330. https://doi.org/10.1429/83967.

Frenken, K., van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth, Regional Studies, 418(5), 685–697, doi: https://doi.org/10.1080/00343400601120296.

Fritsch, M., & Mueller, P. (2008). The effects of new business formation on regional development over time. *Regional Studies, 38*, 961–975. https://doi.org/10.1007/s11187-007-9067-9.

Fritsch, M., & Weyh, A. (2006). How large are the direct employment effects of new businesses? An empirical investigation for West Germany. *Small Business Economics, 27*(2), 245–260. https://doi.org/10.1007/s11187-006-0005-z.

Ghio, N., Guerini, M., & Rossi-Lamastra, C. (2016). University knowledge and the creation of innovative start-ups: an analysis of the Italian case. *Small Business Economics, 47*, 293–311. https://doi.org/10.1007/s11187-016-9720-2.

Glaeser, E. L., Kallal, H. D., Scheinknam, J. S., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*, 1126–1152.

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

Hall, B. H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to R&D. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (pp. 1033–1082). Amsterdam: North Holland.

Jacobs, J. (1969). *The economy of cities*. New York: Random House.

Kirzner, I. (1997). Entrepreneurial discovery and the competitive market process: an Austrian approach. *Journal of Economic Literature, 35*(1), 60–85.

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2014). Agents of structural change. The role of firms and entrepreneurs on regional diversification. In *Papers in evolutionary economic geography* (Vol. 14/10). Utrecht University Section of Economic Geography.

Tödling, F., Prud'homme van Reine, P., & Dörhöfer, S. (2011). Open innovation and regional culture—findings from different industrial and regional settings. *European Planning Studies, 19*(11), 1885–1907. https://doi.org/10.1080/09654313.2011.618688.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics, 113*(2), 331–360.

Wieser, R. (2005). Research and development productivity and spillovers: empirical evidence at the firm level. *Journal of Economic Surveys, 19*(4), 587–621. https://doi.org/10.1111/j.0950-0804.2005.00260.x.