Path dependency, regional variety and the dynamics of new firm creation in rooted and pioneering industries

Journal of Economic Geography (2022) **22**, 631–652
Advance Access Publication 9 October 2021
doi: 10.1093/jeg/lbab021

Path dependency, regional variety and the dynamics of new firm creation in rooted and pioneering industries

Carlo Corradini *,† *and Enrico Vanino* **

*University of Birmingham, Birmingham Business School, 116 Edgbaston Park Rd, B15 2TY, Birmingham, UK
**University of Sheffield, Department of Economics, 9 Mappin Street, S1 4DT, Sheffield, UK
†Correspondence to: *c.corradini@bham.ac.uk*

Abstract

This paper explores regional firm entry connecting insights on the role of localised path dependency with the analysis of variety in regional sectoral structures. Using data on 700 SIC5 industries across 174 NUTS3 regions in the UK between 2000 and 2014, we provide evidence of positive complementarities between industrial path dependence and regional-related variety for firm entry in rooted pre-existing industries. These are negative for entry of pioneering firms in industries new to the region, pointing to lock-in effects and the role of unrelated variety in fostering linkages and knowledge spillovers away from established trajectories.

Keywords: Relatedness, path dependency, related variety, entrepreneurship
JEL classifications: L26, O31, R11
Date submitted: 20 February 2020 **Editorial decision:** 2 May 2021 **Date Accepted:** 7 May 2021

1. Introduction

In this paper, we examine the process of new firm creation in the context of the ongoing debate on how regions transform their economies through the growth of new industries and other forms of new path development (Martin and Sunley, 2006; Boschma and Frenken, 2012; Grillitsch et al., 2018; Hassink et al., 2019). This question has attracted particular attention within evolutionary economic geography, where the emergence of new industries has been described as following a path dependent process defined by the relatedness between new activities entering the region and its pre-existing sectoral composition (Neffke et al., 2011; Boschma and Frenken, 2012; Balland et al., 2015). Within this stream of research, the importance of path dependence in the evolution of local industrial portfolio has been associated with processes of regional branching underlying the development of new industries (Neffke et al., 2011; Boschma et al., 2013; Xiao et al., 2018), as well as entry patterns of new technologies (Essletzbichler, 2015; Rigby, 2015; Tanner, 2015). Yet, the relationship between path dependence and regional entrepreneurship is still not clear. On one side, relatedness between new firms and regional activities may foster knowledge spillovers underpinning new ventures. Yet, this very process may hamper opportunities for entry away from established paths, reducing the development of new growth trajectories (Staber, 2005).

We consider these arguments in conjunction with a distinct but connected literature exploring the role of cognitive proximity within the sectoral composition of regions in shaping agglomeration externalities (Nooteboom, 2000; Boschma, 2005). Following this perspective, the traditional dichotomy of localisation economies and Jacobs's externalities has been revisited to disentangle the impact of different forms of variety across the local industrial structure. In particular, this strand of research has introduced the concept of related variety to depict the idea that the co-location of different sectors sharing commonalities and complementary competencies is conducive to knowledge spillovers underpinning regional growth and innovation. Conversely, unrelated variety reflects the presence of sectors characterised by dissimilar knowledge bases leading to less frequent, albeit potentially more radical, learning processes (Frenken et al., 2007; Boschma and Iammarino, 2009; Castaldi et al., 2015; Grillitsch et al., 2018). The concepts of related and unrelated variety have recently been applied to the analysis of entrepreneurial processes, offering some evidence of the complex yet important effects both may exert on regional firm entry (Bishop, 2012; Colombelli, 2016; Antonietti and Gambarotto, 2018).

Here, we argue that these two views provide complementary insights for the analysis of regional entrepreneurship. In fact, measures of industrial path dependence might not inform us of the level of variety within the regional sectoral structure or the dispersion of competencies around path dependent processes. At the same time, focusing on different layers of integration in the local knowledge base, regional-related and unrelated varieties do not capture the cognitive distance between the industry in which new firms start up and the rest of the regional economy. Take for instance the case of emerging green technology-based industries, whose entry has been associated with processes of regional path dependency (Tanner, 2015; Corradini, 2019). These may grow out of a cluster specialised in strongly related technologies such as chemical industry or material science, surrounded by diverse industries such as finance or tourism. Alternatively, they may draw from a wide set of engineering sectors partially related to green technologies, each one potentially defined by more or less related activities. Both dimensions are important to explain local entrepreneurship and should be taken jointly into account in order to provide a comprehensive analysis.

Accordingly, we define our analysis connecting the stream of research on industrial path dependence with the insights on related and unrelated varieties to explore their potential interplay for regional entrepreneurship. First, we offer novel evidence on the role of industrial path dependence in regional firm entry, measured by the degree of industry relatedness between the sector where new firms operate with respect to the existing regional industrial portfolio. Second, we posit that the specific characteristics of variety in the regional industrial base, as measured by related and unrelated varieties, may play a moderating role on the effect of industrial path dependency. Such framework is applied to the different cases of firm entry in 'rooted' pre-existing industries, as well as entry of 'pioneering' firms in industries new to the region,[1] reflecting different dynamics of industrial and structural change (Neffke et al., 2018).

Using a longitudinal dataset for almost 700 SIC5 industries across 174 NUTS3 regions in the UK for the period 2000–2014, we find the entry of new firms in rooted sectors is positively defined by industrial path dependence, while a negative effect is found for entry of pioneering firms in industries new to the region. At the same time, such effects are found to be moderated in different ways by related and unrelated variety. Cognitive

1 This concept is defined following Hausmann and Neffke (2019).

proximity within related variety facilitates learning opportunities and synergies among established activities, reinforcing the positive relationship between industrial path dependency and firm entry in rooted sectors. On the contrary, unrelated variety fosters connections between more distant knowledge, counterbalancing negative effects of industrial path dependence and supporting the emergence of pioneering firms operating in sectors new to the region. These relationships are shown to be heterogeneous across industries based on their knowledge intensity. In addition, we demonstrate that these results are robust to different methodologies and sensitivity tests.

The paper is organised as follows. In the next section, we present the theoretical framework for this study, combining insights on industrial path dependence with the literature on regional variety and entrepreneurship. The third section presents data and methods used in the econometric analysis. Results are reported and discussed in the fourth section. The fifth section concludes with implications of the study and some final remarks.

2. Literature review and theoretical framework

Regions have long been discussed as nodes of social connectivity facilitating knowledge exchange and information flows through the set of industrial interdependencies (Camagni, 1991; Storper, 1997; Bathelt and Glückler, 2003). These interactions define localised associative capabilities that underlie the emergence of knowledge spillovers and processes of new knowledge creation through collective learning (Asheim, 1996; Capello, 1999; Cooke and Morgan, 1999). Within these flows of localised knowledge spillovers there are also entrepreneurial opportunities (Feldman, 2001; Kwon and Arenius, 2010). The role of spatial proximity for new firm creation is formalised in the knowledge spillover theory of entrepreneurship (Audretsch and Lehmann, 2005; Acs et al., 2013), where entrepreneurial opportunities are defined as a function of localised knowledge created but left untapped by incumbent firms. This knowledge may be then recombined into new firms by other embedded actors, with human capital available in the region representing a key source of entrepreneurial absorptive capacity (Qian et al., 2012). The sources of entrepreneurial dynamism are not confined within regions, and may be also influenced by extra-regional inflows of knowledge, as suggested by the literature on global production networks (Coe et al., 2008; Crescenzi and Iammarino, 2017) and recent contributions in economic geography (Isaksen and Trippl, 2017; Hassink et al., 2019; Neffke et al., 2018).

These perspectives connect processes of new firm creation to established insights on the role of spatial proximity for the acquisition and diffusion of tacit knowledge, whose nature is highly localised (Gertler, 2003). At the same time, they are linked to a large body of research emphasising cognitive proximity as playing an important role for identifying value of untapped knowledge, and reducing uncertainty and transaction costs in knowledge transfer (Nooteboom, 2000; Boschma, 2005). Accordingly, relatedness between skills and knowledge bases leads to a wider set of opportunities for interactive learning and a more effective transmission of knowledge spillovers (Boschma, 2017). These insights have been connected to a view of innovation as a process of knowledge recombination (Weitzman, 1998; Fleming, 2001), considering regional technological change and entrepreneurship as the outcomes of new combinations between related and cognitively distant activities (Castaldi et al., 2015; Boschma, 2017; Colombelli and Quatraro, 2018).

Within this framework, the literature has explored two different yet connected mechanisms through which cognitive proximity affects regional economic dynamics. A large stream of research has delved into the specific composition of the sectoral structure of regions, to distinguish between the effects of commonalities in related industries and those exerted by the presence of diverse knowledge bases in unrelated industries, pointing to different effects of related and unrelated varieties for employment and growth (Frenken et al., 2007; Boschma and Iammarino, 2009), resilience (Boschma, 2015), as well as innovation activities (Castaldi et al., 2015; Miguelez and Moreno, 2018). At the same time, industrial relatedness has been identified as playing a significant role in the process of regional branching and diversification into new growth paths (Boschma and Frenken, 2012; Boschma, 2017). In this context, technological and cognitive proximity lead to the presence of regional path dependence in the evolution of industries, defined by the relatedness between new sectors and the existing regional industrial structure (Neffke et al., 2011).

These perspectives should be combined in order to discuss regional firm entry as the result of the interplay between industrial path dependency and variety in the existing sectoral structure of regions. In particular, we argue that related and unrelated varieties do not necessarily affect firm entry directly, but rather play a moderating effect on the relationship between industrial path dependence and firm entry in rooted or in pioneering industries.

2.1. Industrial path dependency and firm entry

The literature on evolutionary economic geography has indicated how regional structural change is significantly defined by a path dependent dimension shaped by the heterogeneity of knowledge across regions (Martin and Sunley, 2006; Boschma and Martin, 2010). As resources accumulate unevenly over time, they define a spatially bounded set of capabilities and learning opportunities, which translates in the evolution of industries as a function of the knowledge base available in the region. This defines the dynamics of structural change as following a branching process, where new industries grow out of related competences and skills within the pre-existing industry portfolio (Boschma and Frenken, 2012; Balland et al., 2015). Evidence of a significant relationship between the process of regional branching and localised path dependency is discussed in the seminal contribution by Neffke et al. (2011), who show the likelihood of industry entry to be defined by the knowledge relatedness with pre-existing industries. These findings have been corroborated by further studies showing the effect of cohesion between related industries, as well as product relatedness, on the link between incumbent industries and industry entry (Boschma et al., 2013; Essletzbichler, 2015). A similar process has been found when looking at the relationship between technological relatedness in the existing knowledge space and the entry of new technologies across US cities (Rigby, 2015).

Similar mechanisms apply to firm entry. In line with the path and place dependent nature of structural change discussed in this literature, new ventures can be seen as more likely to emerge from localised spillovers related to the existing portfolio of industries within a region. At the same time, the presence of path dependence, reflecting relatedness between sectors where new firms emerge and the existing regional industrial structure, may lead to core rigidity and cognitive lock-in effects (Nooteboom, 2000; Boschma, 2005), creating analogous trade-offs as those identified between exploitation and exploration in innovative search (March, 1991; Fleming, 2001). This equally reflects the tension between processes of adaptation and adaptability in regional resilience (Boschma, 2015). In this sense, path dependency may lead to narrower search breath, reducing the set of entrepreneurial opportunities from more distant knowledge domains. This limits the potential for entrepreneurs to bring novelty in the region, reinforcing path dependence rather than supporting the development of new growth paths (Staber, 2005; Hassink et al., 2019). Accordingly, industrial path dependence may effectively reduce the likelihood of firm entry in sectors which were not present before within the region, which we label as 'pioneering' firms following Hausmann and Neffke (2019). Recent studies have explored these insights looking at the case of specific industries and the entry of green start-ups. These suggest that new technology-based industries may arise even if unrelated to pre-existing activities (Tanner, 2015). Similarly, they provide evidence of an inverted-U relationship between relatedness to green technologies and new start-ups in that sector, suggesting that excessive path dependency limits the space for cross-fertilisation in regional knowledge (Corradini, 2019). Generalising this evidence, Neffke et al. (2018) present a theoretical framework where structural change is enhanced by economic agents that are less reliant on established regional capabilities.

To summarise, industrial path dependence may allow for a more effective transmission of knowledge and ideas along related trajectories, thereby fostering entry in rooted industries. At the same time, these very dynamics may reduce the transmission of spillovers across a wider set of knowledge domains, hampering pioneering entrepreneurial ventures into sectors new to the region. In line with these arguments, we put forward the following hypotheses:

H1a: Industrial path dependence is positively related to entry of new firms in pre-existing rooted industries.

H1b: Industrial path dependence is negatively related to entry of pioneering firms in industries new to the region.

2.2. The moderating effect of regional variety

The effects exerted by industrial path dependence are not independent of the structure of the regional economy. Regions characterised by related variety have been identified as being more conducive to learning opportunities due to the presence of diverse yet close capabilities and knowledge, providing greater adaptation along existing paths (Frenken et al., 2007). At the same time, similarly to the discussion on industrial path dependency in the previous section, related variety may reduce search breadth and the scope of knowledge spillovers, increasing the likelihood of lock-in effects within established sectors. Conversely, unrelated variety generates more challenges in absorbing effectively knowledge across more 'distant' sectors, but also provides more adaptability to move towards new paths and a higher breakthrough potential (Boschma, 2015; Grillitsch et al., 2018). In line with this, Castaldi et al. (2015) have shown that unrelated variety increases the likelihood of radical innovation, as the distance between knowledge domains in the region allows for more novel recombination. Similarly, Corradini and De Propris (2015) find that greater technological diversification improves regional combinatorial opportunities, fostering the emergence of new innovators.

A few studies have recently started exploring the degree of variety in the extant industry portfolio of regions to understand the process of new firm creation. Looking at entrepreneurship rates in the UK, Bishop (2012) finds that both related and unrelated varieties in knowledge-intensive sectors exert a positive effect. At the same time, he does not find a statistically significant effect of broader measures of diversity. Colombelli (2016) also finds a positive effect for both regional-related and unrelated variety on innovative start-ups. Similarly, Colombelli and Quatraro (2018) find a positive effect for both related and unrelated variety looking at medium and high technology new manufacturing firms, noting the former yields a stronger effect than the latter. They also find similar effects looking at the specific case of green start-ups (Colombelli and Quatraro, 2019). Also considering the case of Italy, Antonietti and Gambarotto (2018) identify a positive effect for both characteristics when looking at the number of new start-ups. However, only unrelated variety is found to be significantly associated with innovative new firms.

Combining these insights with the perspective of regional firm entry as inherently shaped by an industrial path dependent process described in Section 2.1, we argue that the specific characteristics of regional variety affect the diffusion and identification of entrepreneurial opportunities along the evolutionary trajectory of regional economies. Cognitive proximity within related variety may foster knowledge spillovers and learning opportunities along established sectors, effectively reinforcing the relationship between industrial path dependence and firm entry. However, this also hampers the likelihood of a radical change of direction in the process of regional branching at the firm level. For opposite reasons, unrelated variety may provide limited coherence for new companies arising in pre-existing sectors. Yet, it also allows for processes of cross-fertilisation across a broader set of more distant sectors, effectively counterbalancing the negative effect of relatedness with pre-existing industries on the emergence of pioneering firms operating in sectors new to the region. Accordingly, we hypothesise the following:

H2a: Related variety positively moderates the effect of industrial path dependence on firm entry in rooted pre-existing sectors, but negatively moderates its effect on pioneering firms in industries new to the region.

H2b: Unrelated variety negatively moderates the effect of industrial path dependence on firm entry in rooted pre-existing sectors, but positively moderates its effect on pioneering firms in industries new to the region.

3. Data and methods

3.1. Data and variables measurement

We employ a longitudinal dataset of almost 700 SIC5 industries across 174 NUTS3 regions in the UK for the period 2000–2014.[2] We retrieve information about regions' industrial portfolio by collapsing plant-level data from the ONS Business Structure Database (BSD) covering the entire population of plants in the UK (ONS, 2017).[3] The BSD data provide information on plants' age, ownership, employment, industrial classification at the SIC5 level and postcode at the street level used to group plants by NUTS3 classification.

We consider two distinct but related dimensions of firm creation at the region-industry level. First, we calculate No.Rooted_{irt} as the number of new plants established every year t in each region r in a pre-existing industry i. Secondly, we measure No.Pioneers_{irt} as the

2 From this analysis we have excluded all industries in the SIC classifications 75 (Public Administration and Defence), 80 (Education), 85 (Health and Social Work) and 90–99 (Other Community, Social and Personal Service Activities), which follow different location motivations, other than the traditional market forces, such as political decisions or public services provision. Additional robustness tests performed including all industries are consistent and available from the authors upon request.

3 The annual BSD dataset is a live register of data based on the annual abstracts from the Inter-Departmental Business Register (IDBR), accessed through the UK Data Service and collected by HM Revenue and Customs via VAT and Pay as You Earn (PAYE) records.

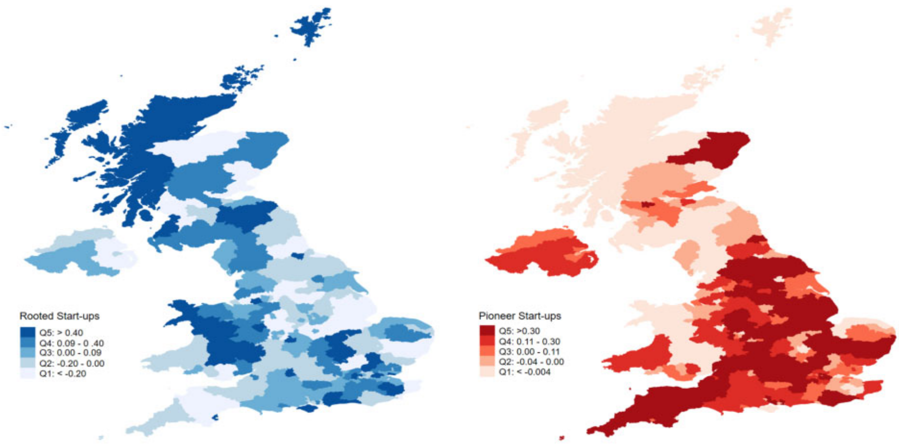

Figure 1. Distribution of rooted and pioneer start-ups across NUTS3 regions in the UK.
Notes: Statistics based on data from the ONS BSD for the period 2000–2014. Rooted start-ups are the logged number of start-ups in pre-existing industries per region. Pioneer start-ups are the logged number of new firms in industries which were not present before in the region. Statistics reported are the residuals of regressing the logged number of rooted and pioneering start-ups against the logged level of regional population, region and year fixed effects, classified according to the five quintiles of their distributions.

number of new businesses established each year in new pioneering industries, thus only considering new establishments in sectors which were not present before in the region (Hausmann and Neffke, 2019). Figure 1 reports the NUTS3 regional distribution of rooted and pioneering start-ups conditional on the population level in the region. It is possible to notice that rooted entrepreneurship is heterogeneously distributed across the UK, with high levels both in urban and in rural regions, while on the contrary, pioneering entrepreneurship is mostly concentrated in UK.

We estimate path dependency between industries within the same region following the methodology proposed by Breschi et al. (2003) based on co-occurrence analysis, as seminally started by Jaffe (1989) and broadly developed since (Teece et al., 1994; Hidalgo et al., 2007; Bryce and Winter, 2009). In our case, we investigate the frequency with which industries i and j co-locate across regions relative to all other industries. Co-occurrence analysis measures the relatedness between two industries by assessing whether they are often found together in the same local economic entity. The assumption made is that the frequency by which two industries are jointly located in the same regions can be interpreted as a sign of the strength of their relationship, in terms of production processes implemented, inputs of production used, technologies developed, skills required and final markets targeted (Neffke et al., 2011; Delgado et al., 2015; Hausmann et al., 2021).

We indicate the number of co-occurrences between SIC 5-digit industries i and j across NUTS 3-digit regions r as $C_{ir}C_{jr}$. By considering joint occurrences in all possible pairs of industrial classifications, we obtain a square symmetrical matrix of co-occurrences C, whose generic cell C_{ij} reports the weighted number of times these industries are jointly located in the same regions. This matrix of co-occurrences can then be used to derive a measure of relatedness between industries using the cosine index S_{ij}, which measures the angular separation between the vectors representing the co-occurrences of industries i and j. As the simple correlation coefficient, the cosine index provides a measure of the similarity between two industries in terms of their mutual relationships with all the other sectors, with the advantage of being symmetric. The final measure S_{ij} is greater the more the two industries i and j co-occur across the same regions[4]:

$$S_{ij} = \frac{\sum_r C_{ir} C_{jr}}{\sqrt{\sum_r C_{ir}^2} \sqrt{\sum_r C_{jr}^2}}. \tag{1}$$

The cosine index S_{ij} gives us a bilateral measure of relationship between two industries i and j. However, we are interested in evaluating how an industry is related to the regional portfolio of all industries operating in each region. Therefore, following the Neffke et al. (2011) measure of industrial closeness, we develop a measure of industry i path dependence with the rest of the region r's industrial portfolio, PD_{ir}, which considers the weighted average of the cosine index between sector i and all the other industries j in region r, weighted by the employment share of each industry j in region r. We limit this measure only to those sectors that are most closely related to industry i, specifically including only industries j in the top quartile of the bilateral relatedness distribution of industry i.[5]

To measure the regional dimension of industrial relatedness, we follow the approach proposed by Frenken et al. (2007), distinguishing a regional industrial portfolio in related and unrelated varieties. Accordingly, regional unrelated variety will be indicated by the entropy of the SIC 2-digit distribution, while related variety will be given by the weighted sum of the entropy at the 5-digit level within each 2-digit classification. Formally, unrelated variety UV_r is given by

$$\mathrm{UV}_r = \sum_{g=1}^{G} P_g \log_2 \left(\frac{1}{P_g} \right), \tag{2}$$

where P_g is the 2-digit share given by the sum of all 5-digit shares p_i. Related variety RV_r instead will be given by the weighted sum of entropy within each 2-digit sector:

$$\mathrm{RV}_r = \sum_{g=1}^{G} P_g H_g, \tag{3}$$

where H_g is the measure of entropy given by

$$H_g = \sum_i \frac{p_i}{P_g} \log_2 \left(\frac{1}{\frac{p_i}{P_g}} \right). \tag{4}$$

4 As a robustness test, in our analysis, we use alternative measures of industrial relatedness instead of the cosine index, such as the Teece et al. (1994) index of industrial relatedness and the Neffke et al. (2011) measure of revealed relatedness. The use of different industrial relatedness measures yields consistent results which are available upon request.

5 We followed this approach in order to reduce the risk of outliers and 'noise' arising from the bilateral relatedness between industries which are only loosely related. We test the sensitivity of our results to this threshold by including in different specifications all industries in the top 50th and 75th percentiles of the bilateral relatedness distribution. Results are consistent and available from the authors upon request.

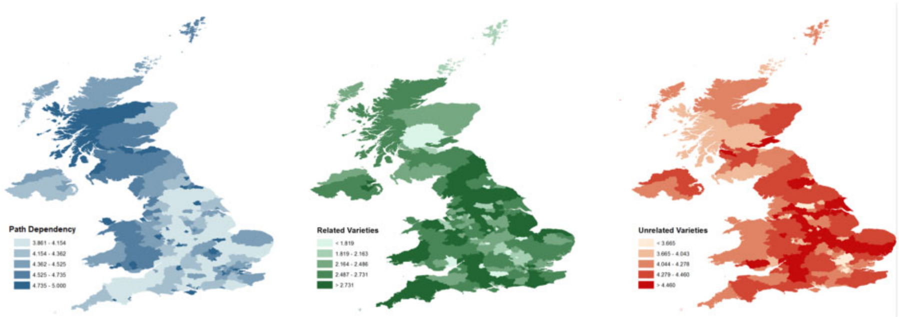

Figure 2. Distribution of path dependency, related and unrelated varieties across NUTS3 regions in the UK.

Notes: Statistics based on data from the ONS BSD for the period 2000–2014. Path dependency is the measure of industrial relatedness between all sectors present in the region estimated using co-occurrence analysis. Related and unrelated varieties measures estimated following Frenken et al. (2007) approach. Regions classified according to the five quintiles of the variables' distribution.

The intuition behind this is that related variety measures the variety within each two digit class. By contrast, unrelated variety measures the extent to which a region is diversified in very different types of activity across 2-digit industry classes.[6]

Figure 2 presents a description of the spatial distribution of industrial path dependence, regional related and unrelated varieties across NUTS3 regions in the UK over the period 2000–2014. It is possible to notice that path dependence is particularly high around the main urban areas in the UK, while the related and unrelated varieties do not seem to follow a similar spatial distribution, being particularly strong in UK rather than the other countries.

3.2. Econometric methodology

In the analysis presented, we test our hypotheses by estimating the following model using a panel ordinary least square (OLS) methodology with region-industry and year-fixed effects[7]:

$$Y_{irt} = \begin{array}{l}\beta_0 + \beta_1 \mathrm{PD}_{irt} + \beta_2 \mathrm{RV}_{rt} + \beta_3 \mathrm{UV}_{rt} + \beta_4 \mathrm{PD}_{irt} \times \mathrm{RV}_{rt} + \beta_5 \mathrm{PD}_{irt} \times \mathrm{UV}_{rt} \\ + \beta_6 X_{irt} + \phi_{ir} + \phi_t + \epsilon_{irt}\end{array}. \tag{5}$$

In the above model, Y_{irt} represents our main dependent variables measuring entrepreneurial activity at the sector i (SIC5), region r (NUTS3) and year t level. In our baseline specifications, we consider two different measures: first, $\ln(\mathrm{Rooted}_{irt})$ measuring the natural log of the number of new plants established every year t in each region r and rooted industry i; secondly, measuring $\ln(\mathrm{Pioneers}_{irt})$ as the natural log of the number of new

6 Related and unrelated varieties are based on predefined hierarchical industry classifications based on main production activities of plants. It is worth noting that some firms may produce goods in different sectors. However, the use of granular plant level data mitigates this problem in our case, as economies of scope related to producing very different products matter much less at the plant level than at the firm level, making the plant main sector a reliable proxy of core competences (Neffke et al., 2011).

7 As a robustness test we have included in our baseline specifications industry-year fixed effects to control for potential industrial trends. In an additional robustness check, we have lagged the independent variables up to 5 year in respect to the dependent variables set at time t. Results are robust and available upon request.

Table 1. Summary statistics for the main variables included in the analysis

Variable name	Mean	SD		Mean	SD
No.Rooted	0.143	0.388	Ind.Saturation	0.806	0.116
Rooted Rate	0.027	0.021	Av.Est.Size	7.421	1.830
No.Pioneers	0.002	0.048	Ind.Density	0.381	0.508
Pioneers Rate	0.001	0.005	GDP	0.024	0.081
PD	4.447	2.254	Education	3.509	0.220
RV	2.445	0.396	Unemployment	6.014	1.999
UV	4.173	0.397	Aggl.Index	0.003	0.012

Note: Statistics based on data from the ONS BSD for the period 2000–2014.

start-ups established each year in new pioneering industries i, thus only considering new businesses established in sectors which were not present before within the region. Table A1 in the Online Appendix presents tests of robustness estimating the same models but considering as dependent variables the RootedRate_{irt} and the PioneerRate_{irt}, weighting the number of new start-ups in rooted and pioneering industries by the total population in each region.[8]

In the above specification, our main variable of interest is PD_{irt}, which measures the natural log of the industrial path dependence between sector i and all other sectors present in region r at time t, estimated using co-occurrence analysis as previously specified and used to test hypotheses H1a and H1b. To test hypotheses H2a and H2b of a moderating effect of variety within the regional industrial portfolio, we look at the interaction between industrial path dependence and region r related RV_{rt} and unrelated varieties UV_{rt}, following the approach proposed by Frenken et al. (2007) as previously detailed. This will give us a clear identification of the role played by specialisation and diversification.

We include in our specifications X_{irt}, a set of region–industry variables to control for other factors which could influence the level of entrepreneurship in the local economy. In this regard, we include the average size of establishments in each industry, the industrial density measured as the number of firms in each industry over the overall population in the region and the Ellison and Glaeser (1997) index of agglomeration at the industry-region level.[9] In addition, we include as controls GDP growth, the level of unemployment and the level of tertiary education attainment provided by EUROSTAT at the NUTS2

8 In further checks, we have estimated the models for No.Rooted_{irt} and No.Pioneers_{irt} using a panel Poisson model in order to take into account the continuous integer count nature of these dependent variables. For rooted industries, we calculate as well the net number of new firms established net of the number of closures, to control that the new ventures do not crowd out existing businesses. Results are consistent and available from the authors upon request.

9 The agglomeration index γ_{rst}^{EG} is measured using data from the BSD:

$$\gamma_{rst}^{\text{EG}} = \frac{G_{rst}/(1-\sum_{rst} x_{rst}^2) - H_{rst}}{1-H_{rst}},$$

where G_{rst} is the *Gini Index* of a region–industry rs at time t measured as $G_{rst} = \sum_{rst} (s_{rst} - x_{rt})^2$ and given by the share s_{rst} of industry s employment in region r and the share of total employment in region r x_{rt}, while H_{rst} is the Herfindahl Index of industrial concentration obtained as $H_{rst} = \sum_{rst} z_{rst}^2$ as the squared of the size of all establishments in industry s and region r at time t.

Table 2. Relationship between path dependence, related/unrelated varieties and entrepreneurship in rooted and pioneering industries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln(Rooted)				**ln(Pioneers)**			
PD	0.00193***		0.00193***	0.0267***	−0.00151***		−0.00151***	−0.00262***
	(0.000265)		(0.000265)	(0.00189)	(0.00009)		(0.00009)	(0.00059)
RV		0.000931	0.000895	−0.0469***		−0.000461	−0.000570	0.00169**
		(0.00303)	(0.00305)	(0.00269)		(0.000600)	(0.000605)	(0.000843)
UV		−0.000985	−0.000900	0.0555***		−0.00001	0.00004	−0.00252***
		(0.00288)	(0.00291)	(0.00278)		(0.00055)	(0.000559)	(0.000895)
PD#RV				0.0105***				−0.000507***
				(0.000570)				(0.000172)
PD#UV				−0.0122***				0.000568***
				(0.000550)				(0.000173)
Av.Est.Size	−0.000235	−0.000140	−0.000226	−0.000550	−0.00102***	−0.000740***	−0.00103***	−0.00101***
	(0.000844)	(0.000832)	(0.000847)	(0.000846)	(0.000151)	(0.000150)	(0.000151)	(0.000151)
Ind.Density	0.541***	0.539***	0.541***	0.539***	−0.0169***	−0.0169***	−0.0170***	−0.0169***
	(0.0103)	(0.0102)	(0.0103)	(0.0103)	(0.00173)	(0.00171)	(0.00173)	(0.00173)
Education	0.00683	0.00710	0.00688	0.00743	−0.00207*	−0.00201*	−0.00209*	−0.00211*
	(0.00667)	(0.00660)	(0.00666)	(0.00666)	(0.00123)	(0.00121)	(0.00123)	(0.00123)
Uenmpl.	0.000867**	0.000895**	0.000863**	0.000845*	0.000026	0.000027	0.000021	0.000022
	(0.000432)	(0.000428)	(0.000433)	(0.000433)	(0.00006)	(0.00006)	(0.000066)	(0.000067)
Aggl.Index	0.722***	0.701***	0.721***	0.690***	−0.194***	−0.201***	−0.194***	−0.192***
	(0.124)	(0.122)	(0.125)	(0.124)	(0.0232)	(0.0227)	(0.0231)	(0.0231)
GDP	0.00630	0.00581	0.00624	0.00638	−0.00586***	−0.00537***	−0.00586***	−0.00595***
	(0.00974)	(0.00965)	(0.00973)	(0.00973)	(0.00168)	(0.00165)	(0.00168)	(0.00168)
Ind.Saturation					0.0180***	0.0180***	0.0185***	0.0190***
					(0.00479)	(0.00478)	(0.00483)	(0.00485)
Observations	956,371	956,371	956,371	956,371	1,002,995	1,002,995	1,002,995	1,002,995
No.Reg-Ind.	88,137	88,137	88,137	88,137	92,397	92,397	92,397	92,397

Notes: Estimations based on data from the ONS BSD for the period 2000–2014 using an OLS estimator with region–industry (NUTS3 and SIC5 digit level) and year-fixed effects. Robust standard errors clustered at the region–industry level reported in parenthesis. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Variables included as previously defined.

regional level. Moreover, in the model for pioneering entrepreneurship, we include as well an index of industrial saturation indicating the number of industries already active over the total number of industries possible at the NUTS3 regional level. Finally, we include year ϕ_t and region-industry ϕ_{ir} fixed effects to control for year specific and region-industry specific time-invariant factors. Table 1 presents summary statistics about the main variables included in our analysis, including the relatedness and diversification measures, the entrepreneurship measures and the main control variables.

4. Results

4.1. Main results

To disentangle the impact of industrial path dependence and regional related/unrelated varieties from other confounding factors, we report in Table 2 the results from the OLS fixed-effects regression. Columns 1–4 show coefficients for rooted firm entry, while Columns 5–8 report the estimates for pioneering firms. Starting in Column 1, we only introduce the variable for industrial path dependence to capture its relationship with firm entry. We find a positive and statistically significant impact, where a 1 standard deviation increase in the level of path dependence corresponds to a 16.7% growth in the number of start-ups in rooted industries.[10] This result extends previous evidence at the industry level (Neffke et al., 2011), supporting hypothesis H1a of new firm creation being facilitated by the cognitive proximity between the sector where these firms emerge and the regional industrial portfolio. Conversely, in line with hypothesis H1b, the coefficient of industrial path dependence is statistically significant but negative in the case of pioneering firms (Column 5), where a 1 standard deviation increase in path dependence is related with almost a third less new firms entering in pioneering industries.[11] These two findings combined show that industrial path dependency connects new firm entry to a set of existing related sectors, limiting instead the opportunity for diversification in new industries. It is worth noting that our finding of a negative relationship between path dependency and pioneering start-ups should not be interpreted as in contrast with previous empirical literature emphasising the role of relatedness in the entry of new technologies and the emergence of new revealed comparative advantages (Neffke et al. 2011; Boschma et al. 2013, 2015). Focusing on the volume of new firms, we interpret our findings in line with recent perspectives suggesting that a stronger connection to existing industries may reduce spillovers from more distant knowledge domains and incentives for entrepreneurs to push in new directions. At the same time, pioneering start-ups that break away from path dependence have higher failure rates and the anchor points they provide for structural change may be more likely to lead to the emergence of new industries and comparative advantages when closer to the existing regional structure (Neffke et al., 2018).

When we consider previously explored variables of regional-related and unrelated varieties in Columns 2 and 6, we obtain similar results to recent studies pointing to a positive effect of related variety on rooted firm entry (Colombelli, 2016; Antonietti and Gambarotto, 2018). However, the coefficients for regional-related and unrelated variety are

10 The standardised coefficient for path dependence is 0.0043 which roughly corresponds to 1.005 more firms per sector–region, where the median number of new rooted firms in a SIC5 industry and NUTS3 region per year is 6.

11 The standardised coefficient for path dependence in this case is −0.0034 which roughly corresponds to 1.003 less pioneering firms per sector–region, where the median number of new pioneering firms in a SIC5 industry and NUTS3 region per year is 3.

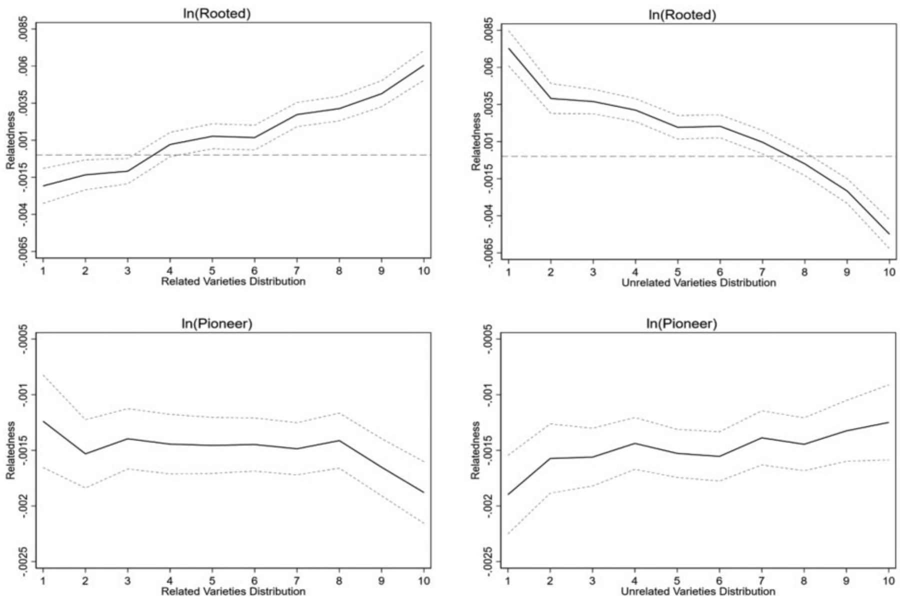

Figure 3. Marginal effect of path dependence on the number of start-ups in rooted and pioneering industries across the distributions of related and unrelated varieties.
Notes: Estimations based on data from the ONS BSD for the period 2000–2014 using an OLS estimator with region–industry (NUTS3 and SIC5 digit level) and year-fixed effects. Confidence intervals at the 95% level reported as dashed lines. Variables included as previously defined.

not statistically significant. In our framework, this reflects the idea that regional variety may not affect entry directly, but only through moderating the effect of industrial path dependence. These results are confirmed once we include both industrial path dependence and regional varieties measures together in Columns 3 and 7. To understand this relationship, we explore the interaction between industrial path dependence and both related and unrelated regional varieties in predicting entrepreneurship using the models reported in Columns 4 and 8. When looking at the interaction terms, we observe a positive sign for RV on rooted entry, pointing to a positive moderating effect, while a negative moderating effect for the interaction term of UV. The opposite holds for pioneering entry, in line with H2a and H2b. For a more intuitive interpretation of this model, in Figure 3, we report the marginal effects of industrial path dependence on rooted and pioneering start-ups across the decile distribution of regional-related variety on the left and for unrelated variety on the right.

For regions characterised by higher levels of related variety, we observe an increasingly stronger effect of industrial path dependence on entrepreneurship in rooted industries. For instance, regions in the top decile of the related variety distribution experience a positive effect of path dependence on entrepreneurship in rooted industries three times larger than firms at the median of the related variety distribution. At the same time, the opposite takes place for pioneering firms, since for regions in the top decile of the related varieties distribution the negative effect of path dependence on pioneering entrepreneurship is a third larger than in regions at the median. These findings support hypothesis H2a, pointing to regional-related variety enhancing learning opportunities along established trajectories, and at the same time impeding recombinations that move away from industrial path dependency. We find distinctly opposite results when observing marginal effects for path dependence across the distribution of unrelated variety, supporting H2b. In this case, there are decreasing effects of path dependence on firm entry in rooted industries for increasingly higher levels of unrelated variety, especially in regions in the top quartile of the unrelated variety distribution, where industrial path dependence becomes negatively related with firm entry in rooted industries. This evidence indicates that regional dissimilarities between industries negatively moderate learning opportunities offered by industrial path dependence for firm entry. Yet, higher levels of unrelated variety in the region allow for processes of cross-fertilisation across a wider set of skills and competencies in more distant and emerging sectors, almost halving the negative impact of path dependency on the creation of pioneering firms for regions highly diversified in respect to regions highly specialised. These marked relationships further highlight that entrepreneurial activities cannot be explained solely observing variety within the existing regional sectoral structure. As pointed out by Balland et al. (2015), this should be considered in conjunction with a dynamic perspective reflecting elements of relatedness and cognitive proximity connecting new activities to pre-existing industries in the region.

With respect to our control variables, our results reflect evidence from the literature on rooted start-up entry. Industrial density and agglomeration effects are positive and significant, underlying the importance of interaction as found in previous studies (Armington and Acs, 2002; Fotopoulos, 2013). The same holds for unemployment (Fotopoulos, 2013; Audretsch et al., 2015a). Conversely, the sign of coefficients for pioneering firms is negative for the control variables. This suggests the same elements that foster interaction and knowledge spillovers underlying rooted entrepreneurial opportunities, such as industrial density and agglomeration effects, also strengthen path-dependency with respect to existing knowledge, industrial competencies and skills.

Following previous evidence of significant heterogeneity characterising new start-ups across different industries (Audretsch et al., 2015b), and of diverse effects of regional variety on innovative firms (Antonietti and Gambarotto, 2018), we also disaggregate our analysis focusing on firm entry in high-tech and knowledge intensive industries (labelled *High-Tech*) versus low-tech and non-knowledge intensive industries (labelled *Low-Tech*).[12] Results are reported in Table 3. Columns 1 and 2 show that High-Tech industries exhibit similar effects of industrial path dependence and regional varieties for both start-ups in rooted and pioneering industries as reported in Table 2. However, we note the impact of path dependence is slightly smaller. This effect might be driven by the fact that entrepreneurial processes in knowledge intensive industries may be inherently defined by stronger search capabilities partially counterbalancing dynamics of industrial path dependence. Conversely, we notice that the role of related/unrelated varieties and their interaction with path dependence for pioneering entry is more marked for knowledge intensive industries, highlighting the importance of access to diverse knowledge in order to foster new knowledge intensive economic activities in a region (Xiao et al., 2018). This is also reflected in the results for rooted and pioneering start-ups in Low-Tech industries (columns 3 and 4).

12 Following the EUROSTAT classification, High-tech firms have SIC codes (2003) equal to: (24) chemicals and pharmaceuticals; (29) machinery and engines; (30) computers and office machinery; (31) electrical machinery; (32) IT and communication equipment; (33) medical, precision and optical instruments; (34) motor vehicles; (35) transport equipment; (61) water transports; (62) air transports; (64) post and telecommunications; (65) financial intermediation; (66) insurance; (67) auxiliary activities to financial intermediation; (70) real estate; (71) renting of machinery and equipment; (72) computer-related activities; (73) research and development and (74) other business activities.

Table 3. Relationship between path dependence, related/unrelated varieties and start-ups in high and low-tech rooted and pioneering industries

	(1) **High-tech** ln(Rooted)	(2) ln(Pioneer)	(3) **Low-tech** ln(Rooted)	(4) ln(Pioneer)
PD	0.0197***	−0.00400***	0.0307***	−0.00111*
	(0.00341)	(0.00109)	(0.00226)	(0.000677)
RV	−0.0414***	0.00534***	−0.0496***	−0.00110
	(0.00486)	(0.00165)	(0.00328)	(0.000910)
UV	0.0454***	−0.00648***	0.0606***	0.000721
	(0.00520)	(0.00168)	(0.00333)	(0.00101)
PD#RV	0.00848***	−0.00143***	0.0113***	−0.000184
	(0.000996)	(0.000346)	(0.000700)	(0.000176)
PD#UV	−0.00956***	0.00140***	−0.0136***	0.00009
	(0.000987)	(0.000338)	(0.000667)	(0.000184)
Observations	330,624	348,485	625,747	654,510
No. Reg-Ind.	30,875	32,154	57,262	60,243

Notes: Estimations based on data from the ONS BSD for the period 2000–2014 using an OLS estimator with region–industry (NUTS3 and SIC5 digit level) and year-fixed effects. Robust standard errors clustered at the region–industry level reported in parenthesis. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Control variables are included as previously defined.

While the signs of the coefficients are unchanged, we observe higher coefficients for rooted start-ups indicating that these industries are more clearly defined by the existing regional structure. For pioneering entry in Low-Tech industries, the sign of coefficients is also in line with our main results. However, only industrial path dependence is found to be statistically significant. This reflects the lack of evidence on the relationship between regional variety and entry for non-innovative firms (Antonietti and Gambarotto, 2018), suggesting that the role of knowledge recombination through variety for pioneering entry is less effective in less complex activities.

4.2. Robustness tests

In addition to the methodological tests previously discussed, we perform several supplementary sensitivity analyses in order to validate the robustness of our main findings. First, we control whether spillovers and spatial externalities might be a concern in our case. As a matter of facts, entrepreneurship could be affected not only by path dependency with industries in the area and regional varieties, but also by linkages with the industrial structure of neighbouring regions. Thus, we first control for the spatial autocorrelation of our main independent variables. Figure A1 in the Online Appendix reports the results of the Moran's I test of spatial autocorrelation for these variables (Moran, 1950). We find evidence of statistically significant spatial autocorrelation only in the case of unrelated varieties (Moran's I index of 0.032, z-score of 4.031 and a p-value of 0.000), while path dependence and related varieties do not seem to be spatially autocorrelated.[13]

13 In particular, we find evidence of significant negative spatial autocorrelation for unrelated varieties in the South-East and East of England, where both regions with the most diverse industrial structure (London,

Table 4. Relationship between path dependence, related/unrelated varieties and entrepreneurship in rooted and pioneering industries: robustness tests

	(1)	(2)	(3)	(4)
	Spillovers		**Hidalgo**	
	ln(Rooted)	ln(Pioneers)	ln(Rooted)	ln(Pioneers)
PD	0.0272***	−0.00272***	0.360*	−0.363***
	(0.00190)	(0.000594)	(0.219)	(0.0705)
RV	−0.0460***	0.00159*	−0.0807*	0.149***
	(0.00269)	(0.000846)	(0.0477)	(0.0169)
UV	0.0555***	−0.00259***	0.0945**	−0.167***
	(0.00280)	(0.000899)	(0.0470)	(0.0165)
PD#RV	0.0105***	−0.000502***	0.0967	−0.214***
	(0.000571)	(0.000173)	(0.0682)	(0.0249)
PD#UV	−0.0124***	0.000584***	−0.119*	0.238***
	(0.000551)	(0.000174)	(0.0671)	(0.0242)
PD spillover	0.0149***	−0.000428		
	(0.00258)	(0.00108)		
RV spillover	0.0323***	−0.00689***		
	(0.00429)	(0.00115)		
UV spillover	−0.0281***	0.00533***		
	(0.00448)	(0.00111)		
Observations	945,780	991,144	828,994	879,928
No. Reg-Ind.	87,058	91,106	82,405	87,468

Notes: Estimations based on data from the ONS BSD for the period 2000–2014 using an OLS estimator with region–industry (NUTS3 and SIC5 digit level) and year-fixed effects. Robust standard errors clustered at the region–industry level reported in parenthesis. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Variables included as previously defined.

Given the presence of spatial autocorrelation for unrelated varieties, we build measures of spatial externalities for industrial path dependence and regional-related/unrelated varieties, in order to control for potential spillover effects originating from neighbouring regions. For path dependence, we achieve this by following Neffke et al. (2011) methodology, as previously explained, and developing a measure PDSpillover_{ir} averaging the cosine index of bilateral relatedness between each industry i in region r and all other industries j in neighbouring regions s. We weight each bilateral industrial relatedness measure by the geographical proximity between each pair of regions r and s,[14] and we limit this measure only to the sectors that are most closely related to industry i and to regions proximate to region r included in the top quartile of the respective distributions. In addition, we create externalities measures for related (RVSpillover_r) and unrelated varieties (UVSpillover_r) by averaging the related/unrelated varieties indexes of neighbouring regions s weighted by the geographical proximity between regions r and s (again considering only regions in the top quartile of the geographical proximity distribution for each region r).

Oxford, Cambridge and Sussex) and regions with homogeneous traditional industries (Lincolnshire, Norfolk, Kent and Hampshire) are located. On the contrary, evidence of positive spatial autocorrelation is found in the South-west and in Wales, regions which are similarly characterised by traditional low-tech sectors.

14 We measure geographical proximity as the normalised value of the inverse of the square root of the Euclidean distance between the centroids of each NUTS3 regions combination.

We then include these measures into our econometric model to control for the presence of potential spillovers. The results of this test reported in Columns 1 and 2 of Table 4 show that our main findings are not affected by spatial externalities, and are consistent with the main findings shown in Table 2. In particular, Column 1 shows that both the direct and indirect measures of industrial path dependence are statistically significant to explain rooted entrepreneurship. Also for entrepreneurship in pioneering industries the main findings are robust to controlling for spatial externalities in Column 2. Despite indirect measures of industrial path dependence do not significantly affect entrepreneurship in these industries, the results confirm the negative relationship between path dependence and pioneering entrepreneurship, which is negatively mediated by the regional degree of related variety, while positively mediated by regional unrelated variety.

In addition, a vast empirical literature has recently corroborated the 'principle of relatedness', showing how many different, but related, measures of relatedness are all valid to describe the probability of regions to enter an economic activity as a function of the number of related activities present in that location, with pros and cons linked to the different spatial scale, type of economic activities and variety of institutional regimes (Hidalgo et al., 2018). However, to make sure that our results are not driven by the uniqueness of the industrial path dependence measure employed in this study, we replicate our baseline analysis using as an alternative the related density measure developed by Hidalgo et al. (2007).[15] Columns 3 and 4 in Table 4 report the results for this sensitivity test, showing a remarkably similar outcome to our main findings estimated using the Neffke et al. (2011) measure of industrial closeness. This is a further evidence not only of the robustness of our findings, but also of the 'principle of relatedness', since different measures of industrial proximity yield very similar results.

5. Conclusions

This paper contributes to the literature on the regional determinants of entrepreneurship by exploring the dynamics between relatedness and firm entry through an evolutionary economic geography perspective. Building on a unified framework connecting insights on the role of spatial path dependency in the emergence of new industries (Neffke et al., 2011; Boschma, 2017) with the literature exploring the impact of economic variety within the industrial structure of regions (Frenken et al., 2007; Castaldi et al., 2015), we argue that knowledge spillovers and learning opportunities underpinning the formation of new companies are co-defined by the relatedness between new firms and incumbent industries, as

15 The concept of 'proximity' ϕ_{ij} between industry i and j developed by Hidalgo et al. (2007) is measured as the minimum of the pairwise conditional probabilities of a region having a comparative advantage in industry i given that it has a comparative advantage also in industry j:

$$\phi_{ij} = \min\{P(\mathrm{RCAe}_i|\mathrm{RCAe}_j), P(\mathrm{RCAe}_j|\mathrm{RCAe}_i)\},$$

where RCA stands for revealed comparative advantage, which measures whether region r has a larger industry i, as a share of its total employment, than the average region (if RCA > 1):

$$\mathrm{RCA}_{ri} = \frac{e_{ri}}{\sum_i e_{ri}} / \frac{\sum_r e_{ri}}{\sum_{ri} e_{ri}}$$

well as the degree of related and unrelated variety that characterises the sectoral structure of regions.

We analyse these elements using a longitudinal dataset of almost 700 SIC5 industries across 174 NUTS3 regions in the UK for the period 2000–2014. Our findings indicate that industrial path dependency, reflecting the relatedness between the sector where the new firms operate and the existing regional industrial portfolio, positively defines the process of new firm creation in established industries. Extending the previous firm-level evidence on industry dynamics (Neffke et al., 2011; Essletzbichler, 2015), and on firms in specific industries (Corradini, 2019), this points to the role of cognitive proximity in enhancing learning processes along evolutionary trajectories for new rooted firms within the region. Thus, spillovers along defined paths effectively support firm creation, providing fundamental vitality to the regional economy. At the same time, this very process negatively affects the entry of firms in sectors new to the region. In this sense, industrial path dependency limits opportunities for regional spillovers across a wider set of knowledge domains, which underpin the emergence of pioneering firms. These findings imply a process of gradual change (Boschma, 2017), where the role of entrepreneurial activities to bring radical novelty in the region remains subdued. This may exacerbate lock-in effects, with entrepreneurship reinforcing path dependency rather than functioning as a driver of new path development (Staber, 2005; Hassink et al., 2019).

These relationships are significantly moderated by the degree of related and unrelated variety in the regional context, adding to previous evidence on their complex role on regional entrepreneurship (Colombelli, 2016; Antonietti and Gambarotto, 2018). In particular, trade-offs are likely to occur between the advantages of related variety for a more effective support of firm entry in incumbent industries as opposed to the advantages of unrelated variety in supporting the entry of pioneering firms moving the region into new sectors. The former may provide a wider yet related set of learning opportunities for new firms whose activities are closer to the existing sectoral structure, reinforcing the positive impact of path dependence for entrepreneurship in rooted industries. Conversely, unrelated variety may ease the effects of path dependence by promoting a platform for spillovers across more diversified knowledge sources anchored in different institutional settings. This could hamper entry in rooted industries. At the same time, it reduces lock-in effects and support entry of pioneering firms in industries new to the region. In this sense, our findings reflect previous research on the role of unrelated variety for breakthrough innovation and new path development (Castaldi et al., 2015; Grillitsch et al., 2018). These results contribute to the literature suggesting the importance of considering the relatedness defined by path dependence in regional structural change together with the degree of variety across regional structures (Balland et al., 2015).

While the paper provides some novel insights on the relationship between path dependence and variety within regions for firm entry, several aspects require further analysis. Following established theories of regional entrepreneurship, our analysis makes the assumption of localised effects for knowledge spillovers and learning opportunities. Yet, further evidence is required to better understand exogenous sources of knowledge flows, extra-regional collaborations as well as migration (Barzotto et al., 2019; Bettin et al., 2019; Hassink et al., 2019). Similarly, our analysis does not allow to explore the differential role of local and non-local entrepreneurs, nor the long-term structural change pioneering entry may provide (Neffke et al., 2018). Also, our results are based on the analysis of spatial and cognitive proximity and should be complemented by further work on other types of proximity as underlined by Boschma (2005), and of institutional elements

(Boschma, 2017; Isaksen and Trippl, 2017). Finally, our work builds on a sectoral classification of employment to define regional capabilities and relatedness, and should be complemented by analyses looking at the relatedness between specific skills and tasks as better data become available. We underline these are important and interesting directions for future research.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

Acknowledgments

The authors would like to thank Dieter Kogler and participants at the RSA 2019 and at the GEOINNO 2020 conferences for comments and feedback. The statistical data used here are from the Office of National Statistics (ONS) and are Crown copyright and reproduced with the permission of the controller of HMSO and Queens Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. The analysis upon which this paper is based uses research datasets which may not exactly reproduce National Statistics aggregates.

References

Acs, Z. J., Audretsch, D. B., Lehmann, E. E. (2013) The knowledge spillover theory of entrepreneurship. *Small Business Economics*, 41: 757–774.
Antonietti, R., Gambarotto, F. (2018) The role of industry variety in the creation of innovative start--ups in Italy. *Small Business Economics, 54:* 1–13.
Armington, C., Acs, Z. J. (2002) The determinants of regional variation in new firm formation. *Regional Studies*, 36: 33–45.
Asheim, B. (1996) Industrial districts as 'learning regions': a condition for prosperity. *European Planning Studies*, 4: 379–400.
Audretsch, D. B., Dohse, D., Niebuhr, A. (2015a) Regional unemployment structure and new firm formation. *Papers in Regional Science*, 94: S115–S138.
Audretsch, D. B., Kuratko, D. F., Link, A. N. (2015b) Making sense of the elusive paradigm of entrepreneurship. *Small Business Economics*, 45: 703–712.
Audretsch, D. B., Lehmann, E. E. (2005) Does the knowledge spillover theory of entrepreneurship hold for regions? *Research Policy*, 34: 1191–1202.
Balland, P.-A., Boschma, R., Frenken, K. (2015) Proximity and innovation: from statics to dynamics. *Regional Studies*, 49: 907–920.
Barzotto, M., Corradini, C., Fai, F. M., Labory, S., Tomlinson, P. R. (2019) Enhancing innovative capabilities in lagging regions: an extra-regional collaborative approach to RIS3. *Cambridge Journal of Regions, Economy and Society*, 12: 213–232.
Bathelt, H., Glückler, J. (2003) Toward a relational economic geography. *Journal of Economic Geography*, 3: 117–144.
Bettin, G., Bianchi, P., Nicolli, F., Ramaciotti, L., Rizzo, U. (2019) Migration, ethnic concentration and firm entry: evidence from Italian regions. *Regional Studies*, 53: 55–66.
Bishop, P. (2012) Knowledge, diversity and entrepreneurship: a spatial analysis of new firm formation in Great Britain. *Entrepreneurship & Regional Development*, 24: 641–660.
Boschma, R. (2005) Proximity and innovation: a critical assessment. *Regional Studies*, 39: 61–74.
Boschma, R. (2015) Towards an evolutionary perspective on regional resilience. *Regional Studies*, 49: 733–751.
Boschma, R. (2017) Relatedness as driver of regional diversification: a research agenda. *Regional Studies*, 51: 351–364.

Boschma, R., Balland, P.-A., Kogler, D. F. (2015) Relatedness and technological change in cities: the rise and fall of technological knowledge in us metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24: 223–250.
Boschma, R., Frenken, K. (2012) *4 technological relatedness and regional branching*. In: Beyond Territory:Dynamic Geographies of Innovation and Knowledge Creation, pp. 78–95. London: Routledge.
Boschma, R., Iammarino, S. (2009) Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, 85: 289–311.
Boschma, R., Martin, R. (2010) *The Handbook of Evolutionary Economic Geography*. Edward Elgar Publishing.
Boschma, R., Minondo, A., Navarro, M. (2013) The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. *Economic Geography*, 89: 29–51.
Breschi, S., Lissoni, F., Malerba, F. (2003) Knowledge-relatedness in firm technological diversification. *Research Policy*, 32: 69–87.
Bryce, D. J., Winter, S. (2009) A general interindustry relatedness index. *Management Science*, 55: 1570–1585.
Camagni, R. (1991) *Innovation Networks: Spatial Perspectives*. Belhaven-Pinter.
Capello, R. (1999) Spatial transfer of knowledge in high technology milieux: learning versus collective learning processes. *Regional Studies*, 33: 353–365.
Castaldi, C., Frenken, K., Los, B. (2015) Related variety, unrelated variety and technological breakthroughs: an analysis of us state-level patenting. *Regional Studies*, 49: 767–781.
Coe, N. M., Dicken, P., Hess, M. (2008) Global production networks: realizing the potential. *Journal of Economic Geography*, 8: 271–295.
Colombelli, A. (2016) The impact of local knowledge bases on the creation of innovative start-ups in Italy. *Small Business Economics*, 47: 383–396.
Colombelli, A., Quatraro, F. (2018) New firm formation and regional knowledge production modes: Italian evidence. *Research Policy*, 47: 139–157.
Colombelli, A., Quatraro, F. (2019) Green start-ups and local knowledge spillovers from clean and dirty technologies. *Small Business Economics*, 52: 773–792.
Cooke, P., Morgan, K. (1999) *The Associational Economy: Firms, Regions, and Innovation*. Oxford University Press.
Corradini, C. (2019) Location determinants of green technological entry: evidence from European regions. *Small Business Economics*, 52: 845–858.
Corradini, C., De Propris, L. (2015) Technological diversification and new innovators in European regions: evidence from patent data. *Environment and Planning A*, 47: 2170–86.
Crescenzi, R., Iammarino, S. (2017) Global investments and regional development trajectories: the missing links. *Regional Studies*, 51:97–115.
Delgado, M., Porter, M. E., Stern, S. (2015) Defining clusters of related industries. *Journal of Economic Geography*, 16:1–38.
Ellison, G., Glaeser, E. L. (1997) Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105: 889–927.
Essletzbichler, J. (2015) Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies*, 49: 752–766.
Feldman, M. P. (2001) The entrepreneurial event revisited: firm formation in a regional context. *Industrial and Corporate Change*, 10: 861–891.
Fleming, L. (2001) Recombinant uncertainty in technological search. *Management Science*, 47: 117–132.
Fotopoulos, G. (2013) On the spatial stickiness of UK new firm formation rates. *Journal of Economic Geography*, 14: 651–679.
Frenken, K., Van Oort, F., Verburg, T. (2007) Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41: 685–697.
Gertler, M. S. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3: 75–99.
Grillitsch, M., Asheim, B., Trippl, M. (2018) Unrelated knowledge combinations: the unexplored potential for regional industrial path development. *Cambridge Journal of Regions, Economy and Society*, 11: 257–274.

Hassink, R., Isaksen, A., Trippl, M. (2019) Towards a comprehensive understanding of new regional industrial path development. *Regional Studies*, 53: 1636–1645.
Hausmann, R., Neffke, F. M. (2019) The workforce of pioneer plants: the role of worker mobility in the diffusion of industries. *Research Policy*, 48: 628–648.
Hausmann, R., Stock, D. P., Y ild ir im, M. A. (2021) Implied Comparative Advantage. *Research Policy*, p. 104143.
Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., Zhu, S. (2018) The principle of relatedness. In A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, and Y. Bar-Yam (eds) *Unifying Themes in Complex Systems IX*, pp. 451–457, Cham: Springer International Publishing.
Hidalgo, C. A., Klinger, B., Barabási, A.-L., Hausmann, R. (2007) The product Space Conditions the Development of Nations. *Science*, 317: 482–487.
Isaksen, A., Trippl, M. (2017) Exogenously led and policy-supported new path development in peripheral regions: analytical and synthetic routes. *Economic Geography*, 93: 436–457.
Jaffe, A. B. (1989) Real effects of academic research. *American Economic Review*, 79: 957–970.
Kwon, S.-W., Arenius, P. (2010) Nations of entrepreneurs: a social capital perspective. *Journal of Business Venturing*, 25: 315–330.
March, J. G. (1991) Exploration and exploitation in organizational learning. *Organization Science*, 2: 71–87.
Martin, R., Sunley, P. (2006) Path dependence and regional economic evolution. *Journal of Economic Geography*, 6: 395–437.
Miguelez, E., Moreno, R. (2018) Relatedness, external linkages and regional innovation in Europe. *Regional Studies*, 52: 688–701.
Moran, P. A. P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, 37: 17–23.
Neffke, F., Hartog, M., Boschma, R., Henning, M. (2018) Agents of structural change: the role of firms and entrepreneurs in regional diversification. *Economic Geography*, 94: 23–48.
Neffke, F., Henning, M., Boschma, R. (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87: 237–265.
Nooteboom, B. (2000) *Learning and Innovation in Organizations and Economies*. Oxford: OUP.
ONS. (2017) *Business Structure Database, 1997–2017: Secure Access* [data collection]. 9th edn. UK Data Service. sn: 6697. Available at: 10.5255/ukda-sn-6697-9.
Qian, H., Acs, Z. J., Stough, R. R. (2012) Regional systems of entrepreneurship: the nexus of human capital, knowledge and new firm formation. *Journal of Economic Geography*, 13: 559–587.
Rigby, D. L. (2015) Technological relatedness and knowledge space: entry and exit of US cities from patent classes. *Regional Studies*, 49: 1922–1937.
Staber, U. (2005) Entrepreneurship as a source of path dependency. In Fuchs, G. and Philip, S. (eds) *Rethinking Regional Innovation and Change*, pp. 107–126. New York, NY: Springer.
Storper, M. (1997) *The Regional World: Territorial Development in a Global Economy*. Guilford Press.
Tanner, A. N. (2015) The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases. *Journal of Economic Geography*, 16: 611–635.
Teece, D., Rumelt, R., Dosi, G., Winter, S. (1994) Understanding corporate coherence: theory and evidence. *Journal of Economic Behavior & Organization*, 23: 1–30.
Weitzman, M. L. (1998) Recombinant growth. *The Quarterly Journal of Economics*, 113: 331–360.
Xiao, J., Boschma, R., Andersson, M. (2018) Industrial diversification in Europe: the differentiated role of relatedness. *Economic Geography*, 94: 514–549.