



HAL
open science

Regional recombinant novelty, related and unrelated technologies: a patent-level approach

Anne Plunket, Felipe Starosta de Waldemar

► **To cite this version:**

Anne Plunket, Felipe Starosta de Waldemar. Regional recombinant novelty, related and unrelated technologies: a patent-level approach. 2022. hal-03567051

HAL Id: hal-03567051

<https://universite-paris-saclay.hal.science/hal-03567051>

Preprint submitted on 11 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Regional recombinant novelty, related and unrelated technologies: a patent-level approach

Anne Plunket

anne.plunket@universite-paris-saclay.fr

Felipe Starosta de Waldemar

felipe.starosta-de-waldemar@universite-paris-saclay.fr

Université Paris Saclay, RITM

December 2021

Abstract

This paper investigates the impact of regional technological relatedness on the emergence of recombinant novelty (i.e. new combinations of subclasses occurring for the first time) in French regions using patent data over 1990-2010. We find that relatedness favors incremental innovations which reuse already applied combinations, whereas increasing levels of relatedness reduces the likelihood of novelty. However, the impact is less negative when combined technologies are new, unrelated or not locally specialized because it facilitates learning and technological recombination. We also find that universities and large incumbents are less dependent on relatedness than small and novel players to create novelty.

Keywords: innovation, recombinant novelty, technological relatedness, Regional development

JEL = O31, R11, O33

1. Introduction

It is a well-established fact that regional economic development builds on the local available capabilities in terms of inputs and knowledge (Hidalgo et al., 2007, 2018). When regions benefit from a diverse set of related activities or industries, they enhance their strengths in specific areas by orienting their investments and thus fostering comparative advantages. This development path builds on the principle of relatedness which has been shown as one of the core mechanisms on which regions rely their future growth (Balland et al., 2018). It is based on the structural role of diversity and similarity: new activities enter more easily in a region when they require similar knowledge and inputs as those already present (Crespo et al., 2017).

Boschma et al., (2015) and Kogler et al., (2013) have demonstrated that this principle of relatedness also applies to technologies. They provide empirical evidence that the entry (and exit) of technologies in cities over time is shaped by the existing (or declining) technological knowledge base of that city. A new technology is more likely to enter a city when it is cognitively related to technological domains in which the region is already specialized.

This result is of particular importance when considering regional innovations and their capacity to be more creative. In the Schumpeterian tradition, innovation is the result of a process of recombination which “refers to the way old ideas can be reconfigured in new ways to make new ideas” (Weitzman, 1998, p.333). The result of this process can lead both to incremental and radical innovations given the underlying search process. Because exploration is costly and uncertain, most innovations are incremental and follow a well-known trajectory in which search occurs in close cognitive proximity. It builds on reusing and improving known technological combinations (Carnabuci and Operti, 2013). In a regional context, it means that innovations tend to rely on local specialization and expertise which are diffused and improved through localized knowledge spillovers and learning. In contrast, more novel combinations may lead to radical innovations and tend to follow a more uncertain and disruptive path. The invention can be considered as radical when it combines previously unconnected technologies or components (Fleming, 2007; Verhoeven et al., 2016). The underlying search process is more exploratory as it deviates from usual paths and combines technological domains that are cognitively more distant (Arts and Fleming, 2018; Schoenmakers and Duysters, 2010). Applied to a regional context, novel innovations can be defined as a combination of technologies that occur for the first time in a region (Montresor et al., 2020).

If the entry of new technologies has been shown to be crucial for innovation as it feeds local technological diversity and enhances the opportunities to combine them, it is less clear whether this principle also supports more radical types of innovation (Boschma, 2017). And in fact a number of scholars argue that unlike most (incremental) innovations, relatedness may not be sufficient to explain the occurrence of radicalness which combines rather unrelated technologies (Castaldi et al., 2015; Hesse and Fornahl, 2020).

The aim of this paper is to contribute to this literature by considering the relationship between novelty and relatedness and focus on the regional dynamic underlying the

emergence of these combinations. More specifically, we study whether novelty combines technologies that build on the regional technological portfolio using the relatedness density indicator. However, our paper departs from this literature by its methodology and empirical strategy. First, unlike studies based on aggregate data, we explore the microeconomic process underlying patent novelty by considering individual inventions produced by inventors located in a given region. Focusing on patents and their underlying technological combinations, we control for patent specific characteristics which allows to focus on regional determinants of novelty. The approach is very similar to microeconomic approaches investigating spillovers arising from agglomeration economies such as urbanization or localization on individual innovations; in our case, the variable of interest is relatedness density. Second, previous studies show that the likelihood, for a new technology, of entering a region increases when the technology is related to the local knowledge base. However, a technology rarely enters the region on its own, it is combined with other technologies within a patent invented in a region. Our empirical strategy enables to control for the characteristics of these other technologies by considering whether they build upon existing trajectories or instead depart from it when introducing novelty. More precisely, patents may combine technologies 1) that enter the region for the first time, 2) that are already in the region but do not have comparative advantage or 3) are unrelated with each other and make its entry less likely and more difficult. Our aim is to consider how relatedness interacts with these technological characteristics and facilitate or instead limit the occurrence of novelty.

We operationalize the relationship between novelty and relatedness by using the OECD REGPAT database (March 2018 Edition) and the EPO and PCT patent applications at the NUTS 2 level. We focus on a subsample of patents produced by French-resident inventors over 1990-2010 and consider for each region and each technological subclass pair used in a patent whether it enters the region for the first time, and if not, how many times it has been used previously. We do also use the OECD REGPAT database to characterize the technological space of regions in terms of specialization patterns based on 18 countries and compute the relatedness density indicator which reveals the extent to which a technological class is related to the existing space.

While it is admitted that the entry of new technologies within a region is driven by its degree of relatedness to its technological portfolio, our results indicate that this relationship does not apply to novel technological combinations. Our findings suggest that relatedness density tends to favor the production of inventions that reinforce local specialization and the reuse of already applied combinations which translates into a negative impact on recombinant novelty. More specifically, relatedness density has an inverted quadratic relationship with novelty: for small levels of relatedness density, it increases the likelihood of novel combinations but for higher levels the impact is decreasing. By further investigating the determinants of novelty, we find that novel combinations occur when it combines technologies that are new to the region, unrelated or do not rely on local comparative advantage. However, the likelihood of novelty is less negative if these technologies are related to the local capabilities as it facilitates learning and knowledge spillovers. Said differently,

novel inventions combine technologies for which the region has an expertise with technologies that are more cognitively distant and unrelated to the local portfolio. Finally, we do also consider the impact of different types of actors and explore whether they rely differently on relatedness. Our findings suggest that small and novel players unlike large incumbents and universities tend to build on the local technological capabilities to create novelty.

The rest of the paper is structured as follows. In the next section, we explain the underlying mechanisms and the hypothesis we test based on previous literature. Section 3 unveils the data, our empirical strategy and the construction of variables. Section 4 presents descriptive statistics, section 5 our empirical results. The last section concludes.

2. Literature review

Innovation processes are usually geographically localized and constrained by the local knowledge base available to organizations and their inventors. The regional knowledge portfolio offers opportunities for novel combinations by broadening the available knowledge base and geographical proximity facilitates learning and lowers the cost of searching for new technologies and moving into related ones (Boschma et al. 2015). The question raised in this paper is whether and under which conditions relatedness favors or inhibits the creation of genuinely novel innovations.

2.1 Recombinant search process and the sources of novelty

Innovations are not all alike and they differ by their degree of novelty which is best understood using the recombinant approach (Weitzman, 1998; Fleming, 2001; Verhoeven et al., 2016). The Schumpeterian tradition conceptualizes the innovation process as the outcome of a recombination of existing or new technologies, knowledge elements or components (Nelson and Winter, 1982, p. 130; Fleming, 2001; Weitzman, 1998). When successful, this recombinant process does not necessarily lead to genuinely novel inventions and in fact most inventions build on incremental improvements of formerly used combinations (Arts and Fleming, 2018). The capacity of organizations to innovate and find new technological combinations depends on their underlying search mechanisms, that is, whether they explore new and distant knowledge or instead exploit and deepen their prior knowledge (March, 1991).

The literature converges to claim that genuine novelty relies primarily on distant search and exploration as it offers higher opportunities for cross-fertilization and new combinations. First, the degree of novelty of an invention depends on the combination of previously unconnected technologies or components (Fleming, 2007; Verhoeven et al., 2016). The invention is considered as radical when it is composed by a relatively high number of new combinations. This definition highlights the introduction of genuinely new combinations which may also potentially generate new areas of development as compared to already

existing combinations which may rather indicate a refinement or an incremental improvement. Second, novelty can also be conceived as the combination of technologies distributed over different technological domains (Keijl et al., 2016; Nemet and Johnson, 2012). For a given invention, the larger the distance between the technological domains and the larger the degree of novelty. Both definitions refer to technological brokering of ideas and technologies that may belong to similar or separate technological fields or industries.

Even though distant search and exploration offer higher opportunities for cross-fertilization, they are also more uncertain and increase the variability of outcomes, failures as well as successes (Fleming, 2001). For these reasons, most inventions reuse familiar technologies as they benefit from experience and learning and they rely on local search in closely related technologies within organizational and regional boundaries (Rosenkopf and Nerkar, 2001) as interactive learning is more effective and can be realized at lower cost.

In order to better formalize this distinction between local and distant search and whether inventors reinforce regional specialization or instead explore new trajectories, we decompose and characterize the technologies combined in each invention. First, technologies can be related or unrelated with each other: related means they are commonly combined and appear frequently within patent documents in a region otherwise they are unrelated, more difficult and riskier to interact, and inventors may lack the needed competences (Caviggioli, 2016; Li et al., 2021). Second, individual technologies may enter the region for the first time; we know that they are more likely to enter the region if they are related to the local knowledge base as in Boschma et al. (2015). Third, the combined technologies may belong to the regional specialization or not. If they do, they tend to reinforce and deepen the current regional technological trajectory. Otherwise, the combination may potentially open new technological developments.

2.2 Deepening technological trajectories and regional path-dependence

Because actors have limited cognitive capabilities and try to reduce uncertainty and risk, they will primarily exploit their own knowledge and expand it through local search by exploring technological combinations that are related to their existing technology portfolio and regional network. As knowledge is tacit and sticky, sourcing knowledge within a region enables to reduce the cost of learning and knowledge absorption. This is the main result of the relatedness literature which demonstrates that the probability that a region enters – or exits – new technologies is a function of the number of related technologies that already exists in that region (Hidalgo et al., 2007, 2018; Boschma et al., 2014; Boschma et al., 2015; Balland et al., 2018).

As a consequence, the technological composition of a region affects the rate and direction of technical change which ends-up being a cumulative, path- and place-dependent process (Martin and Sunley, 2006; Rigby and Essletzbichler, 2008; Henning et al., 2013; Heimeriks and Boschma, 2014). Said differently, organizations are conditioned by the existing knowledge base and they search new technologies by diversifying in closely related technologies because

it lowers risk, uncertainty and facilitates learning processes (Arts and Fleming, 2018; Arts and Veugelers, 2015; Fleming, 2001; Rosenkopf and Nerkar, 2001). However, this path and place-dependent process will predominantly favor incremental refinements of already existing combinations instead of genuine novelty (Castaldi et al., 2015; Li et al., 2021) as these inventions tend to reinforce and deepen the current specialization and the existing technological trajectory. Thus, we have the following expectations:

Hypothesis 1a: Relatedness and local specialization increase the likelihood of reusing existing technological combinations.

Hypothesis 1b: Relatedness and local specialization reduce the likelihood of producing new combinations.

2.3 Path-breaking trajectories and novelty

The previous section just argued that relatedness reduces the likelihood of producing new combinations when the associated technologies reinforce and deepen the current specialization (hypothesis 1b). However, when these combinations introduce technological novelty that departs from local trajectories, there is a greater need for stronger embeddedness in the local knowledge base because relatedness facilitates the underlying learning and recombinant search process. Said differently, the impact of relatedness on the likelihood of producing new combinations will depend on their characteristics in terms of novelty and path-breaking trajectory relative to the local knowledge base. Three situations may be distinguished:

1. the combination introduces a technology subclass that is new to the region;
2. the combination is rather unusual in the sense that the combined subclasses are usually unrelated with each other;
3. the combination associates technologies that have no comparative advantage within the region ($RCA < 1$).

We now discuss each of these three scenarios and explain their moderating role on the impact of relatedness density on novelty, that is, why novel combinations are more likely to be produced if they are related to the local knowledge base.

First, an invention may combine technologies that enter the region for the first time, as depicted in Boschma et al. (2014 and 2015), with the difference that we consider a pair of technologies. If the novel technology is combined with subclasses that are related to the local knowledge base, it will be easier to deal with the learning challenge of exploring new technologies (Boschma et al., 2014; Boschma et al., 2015; Balland et al., 2018; Arts and Fleming, 2018). Building on local knowledge through collaborations or networks provides competences and expertise of how technologies may interact with each other, it reduces the cost of learning new competences and increases the likelihood of success.

Second, inventions may combine technologies in which the region has limited or no expertise, that is, for which the region has no revealed comparative advantage ($RCA < 1$). As in the previous case, the lack of prior knowledge may be compensated if the combined technologies have some proximity with the local knowledge base as it facilitates the learning and recombinant search process.

Third, the new invention may combine technologies that are unrelated with each other in the sense that the combination of two subclasses are seldomly found together on a patent document within a region¹. When an invention combines technologies that are unrelated with each other, it requires by definition completely different capabilities and their combination can be difficult especially if actors lack absorptive capacity to integrate and combine too distant bodies of knowledge effectively. However, unrelated technologies are more likely to be recombined if they are individually strongly present in the same region or related to the local knowledge base as technological distance is easier to bridge within geographical boundaries (Janssen and Frenken, 2019; Li *et al.*, 2021). When these combinations occur successfully they may even help create new technological paths and generate competitive advantage as these novel competences are more difficult to imitate than innovations stemming from related knowledge as argued by Janssen and Frenken (2019). Finally, Castaldi *et al.* (2015) also find that innovations in general draw on technologies that are locally related whereas more breakthrough inventions tend to rely on rather unrelated diversity.

Compared to the baseline effect (Hypothesis 1b) in which relatedness has rather a negative impact on the introduction of novelty, its impact is moderated when individual technologies do not belong to the local knowledge base. In consequence, we hypothesize that relatedness facilitates the emergence of technologies and technological combinations which depart from the local trajectory because it facilitates learning and the recombinant search process.

Hypothesis 2a. Relatedness increases the likelihood of novelty when the combined technologies are new to the region

Hypothesis 2b. Relatedness increases the likelihood of novelty when the combined technologies have no comparative advantage ($RCA < 1$)

Hypothesis 2c. Relatedness increases the likelihood of novelty when the combined technologies are unrelated.

¹ Here we make a clear distinction between individual technologies that are unrelated with each other and relatedness density; however, these individual technologies may nevertheless be related to the technological portfolio which is proxied by the relatedness density variable.

2.4 Relatedness and agents of technological novelty

In this last section, we move one step forward and ask whether the effect of relatedness varies across agents. We investigate who are the actors that introduce technological novelty within regions and whether they build on relatedness to achieve this goal. The heart of the discussion is whether new activities are introduced by new players and whether they build on local resources to do so. In order to distinguish between agents, we rely on inventive activity and distinguish 1) organizations with large inventive activity that have the characteristics of incumbent firms; 2) organizations that patent for the first time and have the characteristics of start-ups, 3) organizations that patent in the region for the first time and may potentially bring extra-regional knowledge and 4) universities that are located in the region but may benefit from external linkages to renew their knowledge base.

Because large incumbents are strongly embedded in their region, it is expected that they develop inventions that reuse local capabilities and resources that reinforce local specialization (Neffke et al., 2018). In contrast, novelty is most likely to be introduced by smaller and novel players that are less constrained by regional specialization and structure. The question is then whether they build on local experience to develop innovations that are related to the local knowledge base (Klepper, 2007) or rather unrelated. Actors that enter the region for the first time have a higher likelihood of producing novelty but may not be able to benefit from the local market due to a lack of anchoring. This is probably the reason why Neffke *et al.*, (2018) find that new firms introduce novel and unrelated activities when they come from outside the region. Finally, we consider the specific role of universities. When universities collaborate intensively with local actors they probably reinforce local specialization, however as they are embedded in extra-regional networks, they may have wider access to potentially unrelated technologies. In sum, through their degree of embeddedness and local networks, large incumbents and universities will tend to reinforce current specialization and decrease the degree of novelty. We propose to test the following hypotheses.

Hypothesis 3. Large incumbents and universities are less likely to rely on local relatedness to produce novelty.

Hypothesis 4. Small and novel players are more likely to rely on local relatedness to produce novelty.

3. Data, variables and estimation strategy

3.1. Data and indicators

The empirical analysis uses the OECD REGPAT database (March 2018 Edition) and the EPO and PCT patent applications at the NUTS 2 level and refers to a sample of 22 French regions with priority year between 1990-2010. A patent is assigned to a region if at least one of the inventors listed on the patent document reports an address located in that region. In case where two inventors of the same patent report location in different regions, patents are affected to each region.

Our objective is to estimate whether a new technological combination occurring in a region is related to the technology portfolio of that region. Following the literature on new technological combinations (Fleming, 2001; Verhoeven et al., 2016; Arts and Fleming, 2018) and applying it at the regional level, we count the number of times a subclass pair combining two IPC codes (International Patent Classification) at the 4 digit level has already been used in a given year and region. Henceforth, a combination is considered as new if it appears for the first time in a given year and region, that is, the count is equal to one. On this basis, we identify each new technological combination entering the region. The sample includes all patents with two or more IPC combinations and amounts to 50% of all patents (46,596 out of 92,000). Of those 50%, there are 70% with inventors from only one region.

We consider *two dependent variables*. The first is *New technological combination*; it takes the value one if the subclass combination appears in the region for the first time in history² and zero otherwise. In order to contrast the impact of relatedness on inventions that are not novel, we compute also the variable *Reuse of technological combination* which counts the number of times a given combination, composing the patent, has already been used in the region.

The main explanatory variable is the *relatedness density* of each IPC subclass. For each subclass pair, we use the *mean of the relatedness density* of both subclasses. Relatedness density is computed following the method proposed by Boschma et al., (2015). To capture the relatedness of technologies available in regions, we use the OECD REGPAT 2018 database. All individual patents are allocated to one of the 245 NUTS 2 regions³ based on inventors' location and one of the 633 IPC codes at the 4-digit level (IPC4) using fractional count (Kogler et al., 2017). Thus, if a patent has been invented by inventors located in three NUTS 2 regions and if the patent is characterized by 2 IPC codes, a patent is allocated at a NUTS 2 region and IPC code with a weight of $1/3 * 1/2 = 1/6$. The advantage is that it avoids double counting of patents.

² Over the whole history of the EPO patents

³ Based on REGPAT regions for 18 countries – EU 15 plus Norway and Switzerland and the US ("DE","BE","FR","IT","LU","NL","DK","IR","UK","GR","ES","PT","AT","FI","SE","CH","NO","GB")

Based on fractional count, we compute the revealed comparative advantage for each region and technology, which is the relative frequency of a technology in a region compared to the frequency of the same technology over all regions. If the value is larger than 1, the region is considered as having a revealed comparative advantage (0/1). More precisely, a region r has a revealed comparative advantage (RCA) in technological class i if the ratio below is higher than 1:

$$\frac{patents_{i,r} / \sum_i patents_{i,r}}{\sum_r patents_{i,r} / \sum_i \sum_r patents_{i,r}} \quad (1)$$

The next step is to compute the conditional probability of having a comparative advantage in two technologies i and j . We compute the conditional probability of having comparative advantage in a technology i conditional on having comparative advantage in technology j and vice versa. To avoid the issue when a region is the only one who has a specialization in one specific technological class, we keep the minimum value of those two conditional probabilities. Hence, to measure the relatedness between each two pair of technological classes i and j , we build variable $\varphi_{i,j}$ following Hidalgo et al. (2007) and Boschma et al. (2015):

$$\varphi_{i,j} = \min\{P(RCA_i|RCA_j), P(RCA_j|RCA_i)\} \quad (2)$$

$\varphi_{i,j}$, relatedness, is the minimum of these two pairwise conditional probabilities. $P(RCA_i|RCA_j)$ is the number of regions with a RCA in both i and j divided by the number of regions with RCA in j , and $P(RCA_j|RCA_i)$ is the number of regions with a RCA in both j and i divided by the number of regions with RCA in i . We get a relatedness matrix with conditional probabilities [0,1]. Finally, we consider only the highest conditional probabilities. Thus, we consider a 5% threshold meaning that only the top 5% of all technology pairs that have the highest relatedness are considered as related, while the remaining 95% are considered as unrelated. After calculating the relatedness $\varphi_{i,j}$, we follow Boschma et al. (2015) and calculate the relatedness density for French regions, at NUTS level 2, specific to each class i and region r . This is calculated as the weighted i of $\varphi_{i,j}$. More formally:

$$\text{Relatedness Density}_{i,r} = \frac{\sum_{j \in r, j \neq i} \varphi_{i,j}}{\sum_{j \neq i} \varphi_{i,j}} \quad (3)$$

The relatedness density variable captures how each technological class i is connected to the technological portfolio of the region r . This is a combination of two factors: how connected is the technological class i to the rest of the technological space, and the portfolio of technological specialization of each specific region r . Formally, in equation 3 the numerator measures the sum of relatedness from technology class i to all technological classes j that are part of the region portfolio ($RCA_{j,r} > 1$), and the denominator is the sum of the relatedness to all technological classes present in the technological space. All in all, a high level of our main

explanatory variable of interest, relatedness density, reveals a strong connection to the regional technological portfolio, while a low level of relatedness density shows that the technological class i is sparsely embedded in the knowledge portfolio of region r . In order to cope with the fact that relatedness density may have a higher impact for lower values and lower impact for higher values as we expect that high specialization may rather limit novelty (see Hypothesis 1), we also estimate the quadratic form of relatedness density.

In addition, we characterize each technological subclass pair used in a patent in order to identify whether they reinforce or rather depart from the preexisting technological set of capabilities in the region. *New subclass* takes value one if the combination includes at least one technology that is new to the region; *No subclass has RCA >1* takes value one if none of the combined subclasses have comparative advantage within the region, *One subclass has RCA >1* takes value one if one of the combined technologies has comparative advantage and zero otherwise; both variables are included in the regression and the base reference is when both subclasses have RCA >1. *Unrelated technologies* takes value one if two technologies are usually not related within a region. It is based on the conditional probability, $\varphi_{i,j}$, computed in the previous section and independent of the specific regional context.

In order to investigate the impact of relatedness density by type of agents, we distinguish five categories of applicants. *Entry* takes value one if the patent is invented by an applicant that appears for the first time in the OECD HAN database, March 2018. *Entry in the region* takes value one if the applicant applies for the first time for a patent with inventors located in the region. In this case, it may be a large incumbent or a subsidiary that has already patented with an inventor team in other regions. *University* takes the value one if the applicant is a French university. *Local applicant* takes value one if the applicant has inventors all located in a single region over the past five years as an indicator of the applicant's location of its inventive activity. *Experience (# patents)* is the applicant's number of patents accumulated over the past five years as provided by the HAN OECD database.

We do also include regional variables to control for the potential determinants of novel combinations at the technological and economic level. *Number of patents in the subclass* is the average number of patents applied in the region in each of the subclasses combined during the five previous years. *Number of associated subclasses* is the average number of subclasses to which our focal subclass pair is combined in the region during the five previous years. *Technological specialization* of the region is measured by the average location quotient weighted by the number of patents computed at the region and subclass level (Boschma et al. 2015). *Technological generality* is the average generality index of all patents in the region (Trajtenberg et al., 1997; Squicciarini et al., 2013) as provided by OECD Patent Quality Indicators database, March 2018. The generality index measures to what extent citations come from distinct technological domains and thus the higher potential for cross-technological recombination within the region. *Inventors with external collaborations* is the average number of inventors within the region with extra-regional collaborations and computed for each subclass combined in the patent.

Finally, we include a number of time-varying region characteristics such as the number of employees in the region (*Employment*), the number of inhabitants by square meters (*Population density*), the economic wealth of a region (*Income per employee*) (Boschma et al. 2015) from INSEE.

3.2. Empirical strategy: estimated model, estimator choice and endogeneity issues

Our empirical strategy is to decompose each patent in a number of subclass pairs and apply a patent fixed effect estimation procedure. Through these fixed effects, we capture all effects that are common to a patent and as all patent-specific variables are perfectly collinear with the patent fixed effects, they are not included⁴ and we focus our estimations on the impact of technology-region variables on novelty. Furthermore, year fixed effects are not needed as a patent is invariant in time. We include subclass fixed effects (two for each combination): they capture all factors specific and common to each technology subclass that are common to all patents based on these codes. At last, note that everything that is region specific is only estimated when using the sample of patents with inventors from more than one region. When including all patents, even those reported in two or more regions, we do also include region fixed effects. We end-up with 46,596 patents and 264,948 observations.

The estimated model is:

New technological combination _{r,i,j,t}

$$= \alpha_p + \beta_0 + \beta_1 \text{Relatedness Density}_{r,i,j,t} + \beta_2 X'_{r,t-1} + \beta_3 Z_{f,t} + \phi_r + \psi_f + \epsilon_{i,c,t}$$

Reuse of technological combination _{r,i,j,t}

$$= \alpha_p + \beta_0 + \beta_1 \text{Relatedness Density}_{r,i,j,t} + \beta_2 X'_{r,t-1} + \beta_3 Z_{f,t} + \phi_r + \psi_f + \epsilon_{i,c,t}$$

As we keep only patents with at least three pairs of combinations (three IPC subclasses), the patent fixed effects α_p captures all common feature from a single patent p . Both our dependent variables are measured at the region, subclass pair combination i, j and time t level. Our explanatory variable of interest, Relatedness Density, is calculated at this same level, and measured as explained above. The explanatory variables X' are listed above, such as population density: they are variables at the regional level, that can be estimated as many patents in our sample have inventors located in different regions. The conditioning set Z refers to variables at the same region-technology level such as our variable of interest: the number of patents in the subclass and the number of associated subclasses. At last, we include region

⁴ Such as the number of inventors, the characteristics of patents such as family size, technological scope, or firm characteristics that are usually controlled for in a patent analysis.

fixed effects ϕ and technological (subclass) fixed effects ψ . Our standard errors are clustered at the patent level⁵.

We estimate the above models using a traditional linear panel data fixed effect estimator, also known as least square dummy variable (LSDV). To our knowledge, to use a linear probability model (LPM, or applying an OLS) in a panel data setting with a binary dependent variable is the most appropriate empirical method (Boschma et al., 2015). Our identification strategy relies on the variation of combinations of technological classes within a patent: that is, we estimate the effect of relatedness density on novelty controlling for patent fixed effects. In contrast, including a fixed effect in a nonlinear model, such as logit or probit, leads to an incidental parameter problem. Nonetheless, in Table A2 in the appendix, we estimate our benchmark specification (Column 6 of Table 3) using a logit and a probit, without patent fixed effects, although keeping region and technological class fixed effects. Our main results are robust to these alternative estimation methods.

The endogeneity of our independent variable of interest (Relatedness Density) could arise from two sources: reverse causality and omitted variables. We explain in detail below how we tackle these issues.

Regarding reverse causality, endogeneity is very limited. At first, Relatedness Density is primarily computed using data from 245 NUTS 2 regions of 18 countries. This is the step explained in equation (2), in which we compute the relatedness for each subclass pair. Therefore, it is very unlikely that a French region may influence this calculation. Next, the numerator of Relatedness Density (equation 3) is calculated as the sum j of technological subclasses present in region r . Straightforward reverse causality for the technological class i can then be excluded, as the variable that measures a new combination of a pair of IPC classes does not enter directly the calculation of relatedness density. Nonetheless, as in our regressions we use the average relatedness density of both IPC's, the relatedness of the IPC i is used in the calculation of the relatedness density of the IPC class j . Once again, note that there are 633 IPC codes in this level of disaggregation: a reverse causality effect is again very low. At last, the RCA dummy in equation 3 ($RCA_{i,j,r} > 1$) incorporates information on the technological classes of patents in the region. An issue would arise if a specific patent drove the number of the other technological classes j having a revealed comparative advantage in the region. A concern would be that patents with many IPC codes develop some kind of technological spillovers in the region. In that sense, it could lead to a situation where having many IPC codes in a patent draws a higher count in equation 1, generating higher observations of $RCA > 1$. Therefore, to address this issue, we estimate again our model by dropping patents with a high number of IPC classes. We discuss these results in section 5 below.

Regarding the omitted variable bias, the relationship between relatedness density and recombinant novelty could catch unobserved determinants of novelty. However, our

⁵ We test two different levels of clustering to check if our results are robust. In Table A2 in the appendix, we present results from our benchmark specification (column 6 of Table 3) clustering our standard errors either at the region (Column 1) or IPC4-region combination (Column 2) level. Our main results are robust to this change.

empirical strategy includes patent, technology class and region fixed effects. These capture common features for all combinations of IPCs in a patent, for every single IPC in the combination which is common to all patents and regions, and for characteristics of patents and IPCs located in a single region. For example, the patent fixed effects capture variables such as patent scope, patent quality and backward or forward citations. The technology class fixed effects captures intrinsic and invariant characteristics of each technology. At last, the regional fixed effects capture the state of technological development of a region, such as being a lead or a laggard. Summed up, these set of variables control a great share of the determinants of recombinant novelty. However, note that our variable is at the technology-regional level: it is specific for each technology in each region for a given year. For example, the density of links to other technologies for IPC subclass H04N in region Bretagne in a given year. In that direction, we control for two variables (our Z variables above) that can be specific to a technology in a given region: the average number of patents applied in the region in each of the subclasses combined during the five previous years (nb of patents in subclass), and the average number of subclasses to which our focal subclass pair is combined in the region during the five previous years (nb of associated subclasses). These variables control for knowledge and spillovers related to the subclasses in the same region. In the end, our estimated model covers the different levels potentially determining recombinant novelty.

4. Descriptive statistics

Descriptive statistics and correlation for the variables used in the regressions are provided in the appendix in Table A1. In order to better understand the regional process of novelty generation, this section provides descriptive statistics on the percentage of patent novelty, new combinations, relatedness density and specialization averaged at the region level. As relatedness density relies on regional specialization, we do also explore whether the technologies that are combined within a patent produced in a region relies on local specialization (i.e. $RCA > 1$). More specifically, we consider for each patent combination whether it includes none, one or two subclasses with a revealed comparative advantage larger than one and compute the *mean RCA*. We do also consider whether a given technology pair combines subclasses from different fields or different sectors based on WIPO classification (Schmoch, 2008).

----- Table 1 here -----

Table 1 describes the spatial characteristics of patent novelty and their related new combinations for each region. The table shows a certain variability in the percentage of novel patents (resp. new combinations) across regions ranging from 17.3 % (resp. 7,7%) in Ile de France which is a leading region with many patent applications and a large technological portfolio to 67.7% (resp. 62,5%) in Limousin which has the opposite characteristics.

Relatedness density is ranging from 0.23 for Limousin to 0.54 for Languedoc-Roussillon whereas Ile de France has a rather high value of 0.37 as well as Rhône-Alpes (0.42) which characterizes portfolio for which technologies are diversified and rather related. Regarding regional specialization, most regions combine technologies for which only one out of the two combined technologies are based on regional specialization (Mean RCA <1) and only seven regions including Ile de France combine technologies for which both are built on local specialization (Mean RCA >1).

Table 1 also shows that new combinations associate technologies from different fields or different sectors confirming previous findings from the literature (Nemet and Johnson, 2012; Keijl et al., 2016) : across regions, 65.9 % to 76.5 % of all combinations are cross-fields and 25% to 40% are cross-sector. Finally, the last column indicates the number of subclasses (out of 633) for which a region has a revealed comparative advantage larger than one. Rhône-Alpes and Île de France are the regions with the largest number of specializations.

Table 2 explores the relationship between novel combinations and the characteristics of combined subclasses. It shows descriptive statistics on the characteristics of novel versus already used technological combinations and t-tests for the comparison of means. The table summarizes the 264,948 technological combinations (50,960 new to the region versus 213,988 already used combinations) characterizing the 46,596 patents in the final sample. Regarding already used technologies, the table indicates that on average a combination is used 153.54 times in a region and associates individual subclasses that are used on average between 356.25 and 539.36 times. In comparison, new combinations rely on subclasses used eight times less⁶. Thus, it is not surprising that *new combinations build on technologies in which the region is not specialized*: 39% of new combinations are based on technologies for which the region is not specialized (RCA >1), 44% rely on a combination for which only one subclass has RCA larger than one and only 17% associated subclasses for which both have a revealed comparative advantage. The differences are statistically significant when compared to reused combinations. An interesting point to note here is that 44% of new combinations combine RCA with non-RCA technologies suggesting that local organizations explore new combinations by relying on local expertise combined with more exploratory or even technologies that are new to the region.

----- Table 2 here -----

Table 2 also indicates that the mean relatedness density is lower and significant for new combinations but the difference is small. Finally, 84% of new combinations combine technologies from different fields (compared to 69% for reused combinations) and 52 % from different sectors (compared to 26%). And 15% (resp. 12 %) of new combinations occur in Ile de France (resp. Rhône-Alpes) region.

⁶ Table 2 shows that IPC codes are used on average 60.81 and 57.91 times in new technological combinations compared to 539.36 and 356.25 for technological reuse.

5. Estimation results

5.1 Main results

In this section, we estimate the conditions under which technological combinations occurring in a region are related to the technology portfolio of that region. In order to do so, we contrast the impact of relatedness density on the number of reused combinations and the likelihood of introducing a new combination in order to explore whether the impact changes given the degree of novelty as measured by the number of times a combination is used in the region.

Table 3 presents our main results. The first three columns test the impact of relatedness density on the *Reuse of a technological combination*, that is, the number of times a given combination has been used in the region. The variable is introduced in a linear and in a quadratic form, and the impact is clearly linear and positive meaning that relatedness density favors the combination of technologies that are already used and related to the regional technological portfolio. This result is depicted in Figure 1a. Results hold in column 3 when we introduce control variables concerning the patent subclass recombination characteristics. When the combined subclasses have no comparative advantage ($RCA > 1$) or only one, the number of reused combinations is lower than when both are specialized, as the coefficients associated with these variables are negative and significant. In sum, these results confirm the hypothesis 1a that relatedness and local specialization increase the likelihood of reusing existing technological combinations. Regarding our control variables, the level of activity in the region such as the income per employee, the number of patents or inventors with external collaborations in the subclass has a positive impact on the number of recombinant reuses. New subclass has also a positive impact on the number of already used combinations⁷.

The remainder of Table 3 tests the impact of relatedness density on the occurrence of a new recombination. In column 4 relatedness density is introduced linearly and the impact is negative and strongly significant. Relatedness density appears to reduce the likelihood of introducing a new combination in a region. In column 5 and 6, we test the quadratic form which is significant and the curve has an inverted u-shape. The average marginal effect shows that the slope of Relatedness density is positive only for small values of relatedness up to .30 and then becomes negative and decreases sharply as Relatedness density increases. The maximum probability of novelty corresponds to a relatedness density of around 0.3⁸ as illustrated by Fig. 1b. This result holds throughout the paper and the interpretation is the

⁷ This result is explained because the dependent variable also includes the cases in which the combination occurs for the first time. If we exclude the “novel” combinations, the impact becomes negative of course.

⁸ The inflexion point of the u-shape curve is $-b_1/(2*b_2) = -0.104/(2*0.173) = 0,3$ (column 6 of table 3)
The inflexion point is within the set of possible values as the median of Relatedness density is 0.36 and the first quartile is .24 and the third quartile is .51.

following⁹: as relatedness to the regional portfolio becomes stronger, the probability of introducing a new combination of subclass diminishes, supporting hypothesis 1b. Relatedness and local specialization reduce the likelihood of producing new combinations. Columns 7 to 10 explore the impact of relatedness density given the characteristics of the combined technologies in three situations through interactions. The results are easier to interpret based on graphical representations. Figure 1c illustrates the interaction between Relatedness density and new subclass (column 7) showing that the introduction of a new subclass in the region increases the likelihood of having a novel combination as the degree of relatedness density increases. This confirms hypothesis 2a as well as the results found by Boschma et al. (2015) that relatedness density eases the introduction of new technologies in a region. If the combinations rely on technologies that are already present, the impact is decreasing as the degree of relatedness density increases. Figure 1d illustrates the impact when the combined technologies are unrelated to each other. The inverted u-shaped curve moves upwards indicating that the likelihood of introducing a new combination is significantly higher when subclasses are unrelated given a level of relatedness density meaning that relatedness facilitates the introduction of unrelated and uncommon combinations confirming hypothesis 2b. However, as before the impact reduces as relatedness density increases. Finally, Figure 1e illustrates the interaction between relatedness density and whether the combined subclasses belong or not to the local specialization ($RCA > 1$). Results support hypothesis 2c and show that novelty is favored when subclasses are not or only partly specialized as compared to the case where both subclasses have revealed technological advantage and the relationship with relatedness density has an inverted-u shape. Compared to figure 1b the decreasing part is less sharp for a given level of relatedness density. These results support hypothesis 2 (a,b,c) that relatedness increases the likelihood of novelty when the combined technologies are new to the region, have no comparative advantage or are unrelated.

----- Table 3 here -----

Table 4 tests the impact of relatedness density on the likelihood of a new technological combination given the type of agents. Again, the strategy is based on interactions and the results are the following. Table 4 provides partial support for hypothesis 3 as universities do not seem to rely on relatedness density to develop new technological combinations (column 1). The interaction is negative and not significant. In contrast, large incumbents approximated by their level of patenting over the past five years do not rely on the local technological

⁹ This result also holds when we test novelty at different geographical levels. Table A3 shows a similar quadratic relationship between relatedness density and novelty when the new combination occurs for the first time at the world level (i.e. it is a new combination pair for the whole EPO database history disregarding location) and at the level of France (i.e. it is a new combination pair for patents applied by French resident inventors disregarding the region in which it occurs). These two variables differ from our main dependent variable which considers a new combination occurring for the first time within a French region.

portfolio to produce novelty (column 2). The impact is clearly negative and significant. Figure 2a shows that relatedness density benefits marginally more to small and medium agents which provides support to hypothesis 4. The agents that benefit most from relatedness density are applicants that are located in a single region (column 3). They clearly rely on this local environment to produce novelty. Finally, column 4 and 5 test the impact of agents' entry on novelty and in relationship with relatedness density and results are remarkably similar to those obtained by Neffke *et al.*, (2018). Those actors that benefit from local specialization are not start-ups that patent for the first time but rather agents that enter the region for the first time but have already a patenting record in other regions. This result, illustrated by Fig 2c., provides partial support to hypothesis 4 that novel players are more constrained by their local environment and more likely to rely on local relatedness to produce novelty.

The next section is devoted to robustness checks.

----- Table 4 here -----

5.2 Robustness Checks

Table 5 presents some robustness checks using our benchmark specification (column 6 of Table 3) which uses *new technological combination* as the dependent variable. The objective is to verify if our main result, the quadratic relationship between relatedness density and novelty, is robust to different samples that could potentially explain its effect.

The first four columns focus on inventor location. In the first column, we only estimate our relationship using patents that have inventors located in a single region. Our results remain similar. In column 2 we use only patents for which inventors are located in at least two different NUTS 2 regions: relatedness density remains negatively and significantly associated with novel recombination. Then, we distinguish estimates for inventors located in the largest French innovative regions (namely, Ile-de-France and Rhône-Alpes in column 3) and outside those regions (column 4). Results are similar in both specifications. The following columns focus on the technological characteristics of combinations. Column 5 and 6 test the effect when the combined technologies are cross-field or from similar fields using WIPO classification (Schmoch, 2008). Our main result remains constant. Next, we estimate our relationship using a sample of patents that have at least one new recombination among all combined pairs: relatedness density still has a quadratic impact associated with patent novelty. At last, we verify if relatedness density is robust to the estimation without extreme values for this variable, namely the top and bottom 1% of the distribution. Results remain unchanged.

To further test the robustness of our results, we run our benchmark specification on the relative density measure instead of the relatedness density. Table 1 shows that regions have different technological portfolios with an average relatedness density ranging from 0.23 to 0.56. These different absolute values are not comparable from one region to another if technologies are not in the same interval throughout all regions. More importantly, if regions with a high-density on average are also those that do not combine new subclass pairs, our

results could be driven only by these observations. Note that, although our estimations already include patent and fixed effects, we still decide to be conservative and test the robustness of our results to the estimation using a relative density measure following Pinheiro *et al.*, (2018). Relative density enables to compare the density of technologies not present in the region with the density of a region's option set (OS). The option set of a region comprises all the technologies that are not yet in that region. Relative density is computed as:

$$Relative\ Density_{i,r} = \frac{RD_{i,r} - mean(RD_{r,OS})}{\sigma_O(RD_{r,OS})}$$

where $RD_{i,r}$ is relatedness density of a technology in a region as in equation (3), $mean(RD_{r,OS})$ is the average density of all technological subclasses in the option set, that is, technologies in which the region has no revealed comparative advantage and the denominator is the standard deviation of the density of the technologies in the same group OS (Pinheiro et al., 2018). This measure centers around zero and negative values indicate unrelated technologies. Thus, we test the impact of a variable "Unrelatedness" which takes value 1 when relative density is negative and zero otherwise. Results in column 9 confirm that new technological combinations rely on technologies that are mainly unrelated to the local portfolio as the associated coefficient is positive and significant.

----- Table 5 here -----

Table 6 tests the robustness of our results for other potential biases. In the first five columns we verify how the dependent variable evolves through the period of study. As novelty is any combination that has never occurred before, regions with a larger knowledge base in the end of the period will find it more difficult to generate new combinations¹⁰. We tackle these potential biases in two ways. First, in columns 1 to 3, we split the sample period in three: 1990-1996, 1997-2003 and 2004-2010. Our main result is robust for these separate periods. Second, in columns 4 and 5 we split the sample in two, above and below the median, and the criteria we use is the share of new combinations of the total of combinations of a region for the year in the middle of our period (2000). Results are robust for both samples. In the last two columns, we check for potential endogeneity from reverse causality as explained in section 3.2. This could be an issue if a patent with many IPC codes drive the number of other technological classes in the same region, generating technological spillovers that could raise the potential number of RCA and thus raising relatedness density. In that direction, to curb a possible indirect effect of reverse causality from IPC codes to relatedness density, we estimate our benchmark specification using only patents with a few numbers of IPC codes. At first, in column 6 we present results from an estimation only with patents having no more than 6 IPC

¹⁰ We thank one referee for pointing out this potential bias.

codes (15 combinations), which account for 86% of our observations. In column 7 we show estimation results for patents with no more than 8 IPC codes (28 combinations), which account for 95% of our observations. Our main result remains unchanged.

----- Table 6 here -----

6. Conclusion

In this paper, we contribute to explain the emergence of technological novelty for individual innovations in a given region, following the recombinant literature (Fleming 2007; Verhoeven et al., 2016). Using patent data for French regions over the period 1990-2010, we contrast events in which technological subclass pairs are combined for the first time with events in which they have been reused already many times. We use a representative sample of patents combining at least three subclass pairs in order to exploit variations of technological characteristics at the regional level while controlling for patent fixed effects amongst other covariates. Our indicator of recombinant novelty is regressed on the average relatedness density of the combined subclasses indicating whether they are related to the technological portfolio of the region. To measure our explained variable of interest, we follow the relatedness literature (Hidalgo et al. 2007; Hidalgo et al., 2018) which shows that the emergence of new activities in a region is driven by the number of related activities present in that location. Yet, as highlighted by Pinheiro et al. (2018, p.2), “what is true on average is not true for every instance. While countries and regions are more likely to enter related economic activities, sometimes they deviate from this behavior and enter unrelated activities...”.

Our findings shed light on the aforementioned instances: relatedness density is associated with an increase in the number of times a combination has been already used in a region. More importantly, relatedness density is mostly negative associated with recombinant novelty. To be precise, relatedness density has an inverted quadratic relationship with novelty: for small levels of relatedness density, it increases the likelihood of a new pair of subclass technologies to be introduced in a region. This positive effect tapers quickly and with an increasing level of relatedness density, novelty emerges less often. This negative effect of relatedness density decreases when the subclass pairs are not both part of the set of technologies in which the region has a revealed comparative advantage, or when they are unrelated to one another. We have also checked if different actors react similarly to relatedness density when producing novelty in their innovations. Our results show that while large incumbents are less dependent on relatedness, small and novel players build on the local technological portfolio to create novelty, and relatedness density does not facilitate universities to engage in recombinant novelty.

These main results are robust to different covariates and a series of fixed effects. Moreover, our results are also robust to different samples of regions, subclasses within and

across fields, to samples of patents with at least one novelty in their combinations, extreme values for relatedness density, as well as the introduction of relative density measure. We also check that our results are robust to the evolution of our dependent variable, endogeneity, other estimation methods and the level of clustering of our standard-errors.

Our results shed light on a few important policy implications. First, in order to foster regional economic growth and development, relatedness (Balland et al., 2018) plays an important role. We contribute to this literature by showing that the novelty of these activities does not follow the same pattern. In order to set foot on novel innovations, regions also need to explore more distant and less mastered pieces of knowledge, breaking apart from path and place dependence. This implies that the policy recommendations for exploring into distant and unknown knowledge should be adapted to the local environment. Second, these policy recommendations should also consider the type of agents present in the region when trying to foster novelty. Our empirical exploration suggests that the local technological environment heterogeneously constraints the production of novelty depending on the type of actor introducing novelty. While some actors are not affected, some are more constrained while others rely more on relatedness density. Policy design needs then to incorporate the novelty profile of agents in a given region when developing tools to promote novelty. As in the case of structural change in Neffke *et al.*, (2018) we find that local agents are more embedded in the local technological space and novelty is sparked by agents producing already in other regions. In sum, the interaction between local relatedness and the profile of agents is crucial to regional policy design.

Nonetheless, our results and methodology have some limitations which pave the way to future improvements. First, further microeconomic analysis could empirically test specific underlying mechanisms that spark the emergence of new combinations. Second, the field could also benefit from deepening the knowledge associated with the agency dimension: are the new combinations driven by multinationals or other type of companies, and what are they previous experience in introducing novelty elsewhere? Third, an important path is to explore in which type of the technologies are those new combinations intrinsically more complex or not (in the sense of Hidalgo, 2021), and do they pave the way to robust and durable new trajectories? Finally, focusing on specific improvements, we could, for example, find a new way to aggregate the relatedness density of two subclass pairs rather than the average of both single measures, such as a weighted average. We could also interact our variable of interest with other potential determinants of recombinant novelty, especially those that allow regions to bring knowledge and technologies from outside its territory, such as migrant inventors.

7. References

- Arts, S., Fleming, L., 2018. Paradise of Novelty—Or Loss of Human Capital? Exploring New Fields and Inventive Output. *Organ. Sci.* 29, 1074–1092. doi:10.1287/orsc.2018.1216
- Arts, S., Veugelers, R., 2015. Technology familiarity, recombinant novelty, and breakthrough invention. *Ind. Corp. Chang.* 24, 1215–1246. doi:10.1093/icc/dtu029

- Balland, P.A., Boschma, R., Crespo, J., Rigby, D.L., 2018. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Reg. Stud.* 1–17. doi:10.1080/00343404.2018.1437900
- Boschma, R., 2017. Relatedness as driver of regional diversification: a research agenda. *Reg. Stud.* 51, 351–364. doi:10.1080/00343404.2016.1254767
- 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. *Ind. Corp. Chang.* 24, 223–250. doi:10.1093/icc/dtu012
- Boschma, R., Heimeriks, G., Balland, P.-A., 2014. Scientific knowledge dynamics and relatedness in biotech cities. *Res. Policy* 43, 107–114. doi:10.1016/j.respol.2013.07.009
- Carnabuci, G., Operti, E., 2013. Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. *Strateg. Manag. J.* 34, 1591–1613. doi:10.1002/smj.2084
- Castaldi, C., Frenken, K., Los, B., 2015. Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. *Reg. Stud.* 49, 767–781. doi:10.1080/00343404.2014.940305
- Caviggioli, F., 2016. Technology fusion: Identification and analysis of the drivers of technology convergence using patent data. *Technovation* 55–56, 22–32. doi:10.1016/J.TECHNOVATION.2016.04.003
- Crespo, J., Balland, P.-A., Boschma, R., Rigby, D., 2017. Regional diversification opportunities and smart specialization strategies - Publications Office of the EU. doi:10.2777/133737
- Fleming, L., 2007. Breakthroughs and the 'Long Tail' of Innovation. *MIT Sloan Manag. Rev.* 49, 69–74.
- Fleming, L., 2001. Recombinant Uncertainty in Technological Search. *Manage. Sci.* 47, 117–132. doi:10.1287/mnsc.47.1.117.10671
- Heimeriks, G., Boschma, R., 2014. The path- and place-dependent nature of scientific knowledge production in biotech 1986-2008. *J. Econ. Geogr.* 14, 339–364. doi:10.1093/jeg/lbs052
- Henning, M., Stam, E., Wenting, R., 2013. Path Dependence Research in Regional Economic Development: Cacophony or Knowledge Accumulation? *Reg. Stud.* 47, 1348–1362. doi:10.1080/00343404.2012.750422
- Hesse, K., Fornahl, D., 2020. Essential ingredients for radical innovations? The role of (un-)related variety and external linkages in Germany. *Pap. Reg. Sci. pirs.12527.* doi:10.1111/pirs.12527
- Hidalgo, C.A., 2021. Economic complexity theory and applications. *Nat. Rev. Phys.* 3, 92–113. doi:10.1038/s42254-020-00275-1
- 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. pp. 451–457. doi:10.1007/978-3-319-96661-8_46
- Hidalgo, C.A., Klinger, B., Barabási, A.-L., Hausmann, R., 2007. The Product Space Conditions the Development of Nations. *Science (80-)*. 317, 482 LP – 487.
- Janssen, M.J., Frenken, K., 2019. Cross-specialisation policy: rationales and options for linking unrelated industries. *Cambridge J. Reg. Econ. Soc.* doi:10.1093/cjres/rsz001
- Keijl, S., Gilsing, V.A., Knobens, J., Duysters, G., 2016. The two faces of inventions: The relationship between recombination and impact in pharmaceutical biotechnology. *Res. Policy* 45, 1061–1074. doi:10.1016/j.respol.2016.02.008
- Klepper, S., 2007. Disagreements, Spinoffs, and the Evolution of Detroit as the Capital of the

- U.S. Automobile Industry. *Manage. Sci.* 53, 616–631. doi:10.1287/mnsc.1060.0683
- Kogler, D.F., Essletzbichler, J., Rigby, D.L., 2017. The evolution of specialization in the EU15 knowledge space. *J. Econ. Geogr.* 17, 345–373.
- Kogler, D.F., Rigby, D.L., Tucker, I., 2013. Mapping Knowledge Space and Technological Relatedness in US Cities. *Eur. Plan. Stud.* 21, 1374–1391. doi:10.1080/09654313.2012.755832
- Li, D., Heimeriks, G., Alkemade, F., 2021. Recombinant invention in solar photovoltaic technology: can geographical proximity bridge technological distance? *Reg. Stud.* 55, 605–616. doi:10.1080/00343404.2020.1839639
- March, J.G., 1991. *Exploration and Exploitation in Organizational Learning*, Organization Science.
- Martin, R., Sunley, P., 2006. Path dependence and regional economic evolution. *J. Econ. Geogr.* 6, 395–437. doi:10.1093/jeg/lbl012
- Montesor, S., Orsatti, G., Quatraro, F., 2020. Technological novelty and Key Enabling Technologies: Evidence from European regions (No. 2020–05).
- Neffke, F., Hartog, M., Boschma, R., Henning, M., 2018. Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification. *Econ. Geogr.* 94, 23–48. doi:10.1080/00130095.2017.1391691
- Nelson, R.R., Winter, S.G., 1982. *An Evolutionary Theory of Economic Change*. Cambridge University Press.
- Nemet, G.F., Johnson, E., 2012. Do important inventions benefit from knowledge originating in other technological domains? *Res. Policy* 41, 190–200. doi:10.1016/J.RESPOL.2011.08.009
- Pinheiro, F.L., Alshamsi, A., Hartmann, D., Boschma, R., Hidalgo, C.A., 2018. Shooting High or Low: Do Countries Benefit from Entering Unrelated Activities?
- Rigby, D.L., Essletzbichler, J., 2008. Evolution, Process Variety, and Regional Trajectories of Technological Change in U.S. Manufacturing. *Econ. Geogr.* 73, 269–284. doi:10.1111/j.1944-8287.1997.tb00089.x
- Rosenkopf, L., Nerkar, A., 2001. Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strateg. Manag. J.* 22, 287–306. doi:10.1002/smj.160
- Schmoch, U., 2008. *Concept of a Technology Classification for Country Comparisons*, Final Report to the World Intellectual Property Organisation (WIPO).
- Schoenmakers, W., Duysters, G., 2010. The technological origins of radical inventions. *Res. Policy* 39, 1051–1059. doi:10.1016/j.respol.2010.05.013
- Squicciarini, M., Dernis, H., Criscuolo, C., 2013. *Measuring Patent Quality* (No. 2013/03), OECD Science, Technology and Industry Working Papers. OECD Publishing, Paris. doi:http://dx.doi.org/10.1787/5k4522wkw1r8-en
- Trajtenberg, M., Henderson, R., Jaffe, A., 1997. University Versus Corporate Patents: A Window On The Basicness Of Invention. *Econ. Innov. New Technol.* 5, 19–50. doi:10.1080/10438599700000006
- Verhoeven, D., Bakker, J., Veugelers, R., 2016. Measuring technological novelty with patent-based indicators. *Res. Policy* 45, 707–723. doi:10.1016/j.respol.2015.11.010
- Weitzman, M.L., 1998. Recombinant growth. *Q. J. Econ.* 113, 331–360. doi:10.1162/003355398555595

TABLE 1. Spatial characteristics of novel patents, new combinations and specialization

NUTS2	Regions	% Patent Novelty ¹		% New Combinations ²		Relatedness Density ³		Mean RCA ⁴	Cross Field	Cross Sector	Region # RCA ⁵
		Mean	SD	Mean	SD	Mean	SD	Mean	Mean	Mean	Mean
FR10	ÎLE DE FRANCE	0.173	0.378	0.077	0.268	.371	.221	1.03	.72	.297	218
FR21	CHAMPAGNE-ARDENNE	0.543	0.499	0.454	0.498	.293	.151	.774	.692	.29	134
FR22	PICARDIE	0.441	0.497	0.355	0.479	.393	.176	.984	.751	.364	162
FR23	HAUTE-NORMANDIE	0.378	0.485	0.26	0.439	.383	.147	1.2	.732	.367	137
FR24	CENTRE	0.457	0.498	0.356	0.479	.264	.171	.608	.731	.308	166
FR25	BASSE-NORMANDIE	0.623	0.485	0.493	0.5	.324	.117	.938	.765	.409	123
FR26	BOURGOGNE	0.572	0.495	0.46	0.498	.345	.173	.883	.707	.388	168
FR30	NORD - PAS-DE-CALAIS	0.436	0.496	0.282	0.45	.484	.135	.986	.716	.299	192
FR41	LORRAINE	0.511	0.5	0.397	0.489	.36	.171	.961	.659	.353	160
FR42	ALSACE	0.261	0.439	0.158	0.364	.465	.14	1.44	.735	.256	159
FR43	FRANCHE-COMTÉ	0.662	0.473	0.56	0.496	.288	.178	.768	.685	.385	130
FR51	PAYS DE LA LOIRE	0.546	0.498	0.427	0.495	.382	.193	.672	.705	.335	187
FR52	BRETAGNE	0.359	0.48	0.287	0.453	.397	.25	.905	.733	.27	124
FR53	POITOU-CHARENTES	0.548	0.498	0.452	0.498	.296	.122	.752	.683	.258	141
FR61	AQUITAINE	0.464	0.499	0.351	0.477	.455	.143	1.13	.744	.367	184
FR62	MIDI-PYRÉNÉES	0.403	0.491	0.286	0.452	.36	.152	1.01	.698	.271	164
FR63	LIMOUSIN	0.677	0.469	0.625	0.485	.232	.0939	.784	.731	.387	88.9
FR71	RHÔNE-ALPES	0.288	0.453	0.151	0.358	.42	.139	.814	.742	.338	227
FR72	AUVERGNE	0.385	0.487	0.268	0.443	.258	.116	1.18	.673	.392	101
FR81	LANGUEDOC-ROUSSILLON	0.341	0.474	0.252	0.434	.546	.183	1.51	.725	.253	148
FR82	PROVENCE-ALPES-CÔTE D'AZUR	0.355	0.478	0.231	0.422	.38	.153	.899	.722	.311	183

¹ % Patent Novelty is the percentage of patents which include at least one technological combination new to the region

² % New Combinations is the percentage of combinations that are new to the region

³ Relatedness density is the average relatedness density of the combined technologies

⁴ For each patent combination, we consider whether it combines none, one or two subclasses belonging to the regional specialization and averaged At the regional level

⁵ Total number of subclasses which represent a revealed comparative advantage (RCA)

TABLE 2. Ttest of novel versus reuse of technological combinations

	Reuse of technological combinations				New technological combinations				Ttes	
	mean	sd	min	max	mean	sd	min	max	Mean	sd
Recombinant novelty (Region)	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	-1.00	0.00
Recombinant reuse ¹	153.54	369.78	2.00	3089.00	1.00	0.00	1.00	1.00	152.54***	1.64
# of times a subclass IPC1 has been used ²	539.36	822.82	2.00	4623.00	60.81	175.33	1.00	4619.00	478.55***	3.66
# of times a subclass IPC2 has been used ²	356.25	494.92	2.00	4618.00	57.91	144.51	1.00	3366.00	298.34***	2.21
Relatedness Density IPC1	0.39	0.21	0.00	1.00	0.36	0.21	0.00	1.00	0.03***	0.00
Relatedness Density IPC2	0.40	0.23	0.00	1.00	0.36	0.22	0.00	1.00	0.04***	0.00
Relatedness density	0.39	0.20	0.00	1.00	0.36	0.17	0.00	1.00	0.03***	0.00
New subclass	0.00	0.00	0.00	0.00	0.13	0.34	0.00	1.00	-0.13***	0.00
Both subclasses have RCA >1	0.35	0.48	0.00	1.00	0.17	0.38	0.00	1.00	0.18***	0.00
One subclass has RCA >1	0.36	0.48	0.00	1.00	0.44	0.50	0.00	1.00	-0.09***	0.00
No subclass has RCA > 1	0.29	0.46	0.00	1.00	0.39	0.49	0.00	1.00	-0.09***	0.00
Unrelated techn. (subclass)	0.34	0.47	0.00	1.00	0.72	0.45	0.00	1.00	-0.39***	0.00
Different fields	0.69	0.46	0.00	1.00	0.84	0.36	0.00	1.00	-0.15***	0.00
Different sectors	0.26	0.44	0.00	1.00	0.52	0.50	0.00	1.00	-0.26***	0.00
IDF – Ile de France	0.43	0.50	0.00	1.00	0.15	0.36	0.00	1.00	0.28***	0.00
Rhône-Alpes	0.17	0.37	0.00	1.00	0.12	0.33	0.00	1.00	0.04***	0.00
Observations	213988				50960				264948	

¹ # of times a combination has been used in the region

² Each invention is decomposed in subclass pairs (i.e. combination IPC1-IPC2) – The table characterizes these combinations.

TABLE 3. Impact of relatedness density on the number of reused combinations and the likelihood of a new combination (Patent level with patent fixed effects)

Regional characteristics	# of recomb. Reuse			New recombination (0/1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Relatedness density	0.653*** [0.027]	-0.061 [0.078]	0.154** [0.072]	-0.091*** [0.010]	0.154*** [0.026]	0.104*** [0.026]	0.102*** [0.027]	0.036 [0.030]	-0.127*** [0.033]	-0.161*** [0.038]
Relatedness density sq		0.840*** [0.086]	0.063 [0.079]		-0.288*** [0.028]	-0.173*** [0.027]	-0.173*** [0.028]	-0.063** [0.031]	0.051 [0.034]	0.119*** [0.038]
New subclass	0.175*** [0.018]	0.171*** [0.018]	0.234*** [0.018]	0.283*** [0.008]	0.284*** [0.008]	0.279*** [0.008]	0.283*** [0.015]	0.278*** [0.008]	0.279*** [0.008]	0.296*** [0.016]
New subclass x Relatedness density							-0.130 [0.095]			-0.236** [0.100]
New subclass x Relatedness density sq							0.326** [0.158]			0.479*** [0.163]
Patent recombination										
No subclass has RCA >1			-0.445*** [0.011]			0.036*** [0.004]	0.035*** [0.004]	0.037*** [0.004]	-0.029*** [0.010]	-0.023** [0.011]
One subclass has RCA >1			-0.363*** [0.007]			0.037*** [0.002]	0.036*** [0.002]	0.037*** [0.002]	-0.038*** [0.008]	-0.031*** [0.008]
Unrelated subclasses			-0.716*** [0.008]			0.119*** [0.003]	0.119*** [0.003]	0.105*** [0.007]	0.119*** [0.003]	0.108*** [0.008]
Unrelated subclasses x Relatedness density								0.151*** [0.034]		0.129*** [0.036]
Unrelated subclasses x Relatedness density sq								-0.238*** [0.037]		-0.207*** [0.038]
No subclass has RCA >1 x Relatedness density									0.324*** [0.058]	0.293*** [0.063]
One subclass has RCA >1 x Relatedness density									0.370*** [0.037]	0.329*** [0.038]
No subclass has RCA >1 x Relatedness density sq									-0.367*** [0.081]	-0.328*** [0.086]
One subclass has RCA >1 x Relatedness density sq									-0.399*** [0.041]	-0.347*** [0.042]
Regional Controls										
Population density	0.371 [0.591]	0.327 [0.593]	0.668 [0.577]	-0.187 [0.192]	-0.172 [0.192]	-0.192 [0.191]	-0.190 [0.191]	-0.193 [0.191]	-0.201 [0.191]	-0.198 [0.191]
Income per employee	0.775*** [0.213]	0.756*** [0.214]	0.454** [0.208]	0.016 [0.063]	0.022 [0.063]	0.045 [0.063]	0.044 [0.063]	0.041 [0.063]	0.052 [0.062]	0.047 [0.062]
Employment	-0.190 [0.441]	-0.317 [0.443]	-0.233 [0.431]	0.293** [0.141]	0.337** [0.141]	0.326** [0.140]	0.327** [0.140]	0.316** [0.140]	0.333** [0.140]	0.325** [0.140]
Technological generality	0.035 [0.326]	0.110 [0.328]	0.110 [0.321]	0.214** [0.105]	0.188+ [0.105]	0.191+ [0.105]	0.192+ [0.105]	0.190+ [0.105]	0.187+ [0.104]	0.186+ [0.105]
Technological specialization	-0.063 [0.046]	-0.076 [0.046]	-0.076+ [0.045]	-0.027+ [0.016]	-0.023 [0.016]	-0.023 [0.016]	-0.023 [0.016]	-0.022 [0.016]	-0.022 [0.016]	-0.021 [0.016]

Nb of patents in the subclass	0.325***	0.329***	0.229***	0.015***	0.013***	0.018***	0.018***	0.019***	0.018***	0.018***
	[0.011]	[0.011]	[0.011]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Number of associated subclasses	-0.128***	-0.128***	-0.091***	-0.160***	-0.159***	-0.163***	-0.163***	-0.163***	-0.162***	-0.163***
	[0.011]	[0.011]	[0.011]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Inventors with external collaborations	0.234***	0.235***	0.251***	0.024***	0.023***	0.021***	0.021***	0.020***	0.021***	0.020***
	[0.008]	[0.008]	[0.008]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Constant	-7.988	-5.603	-4.111	-2.491	-3.307+	-3.429+	-3.451**	-3.245+	-3.512**	-3.366+
	[5.630]	[5.657]	[5.489]	[1.763]	[1.764]	[1.760]	[1.761]	[1.759]	[1.757]	[1.756]
Observations	264948	264948	264948	264948	264948	264948	264948	264948	264948	264948
R-Squared	.6	.6	.64	.22	.22	.23	.23	.23	.23	.23

Robust standard errors clustered at the patent level in brackets, + 0.10 ** 0.05 ***0.01

Figure 1. Impact of relatedness density and interaction with combined technologies (subclasses)

Fig 1a Impact on recombinant reuse

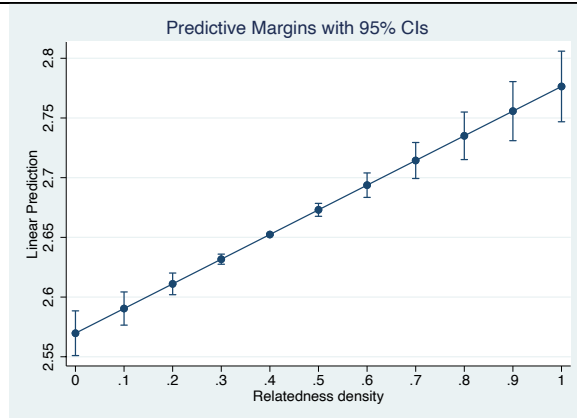
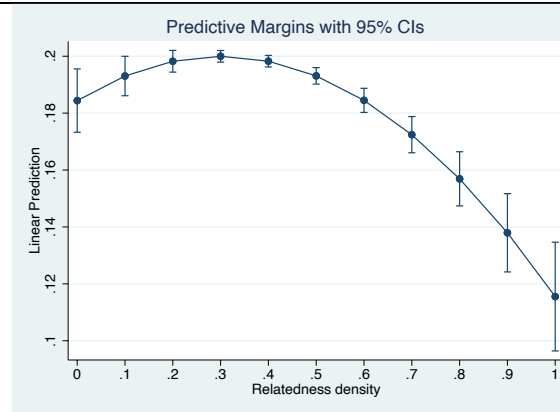


Fig 1b Impact on new recombination



Impact on recombinant novelty (New combination)

Fig 1c when the combination introduces a new subclass

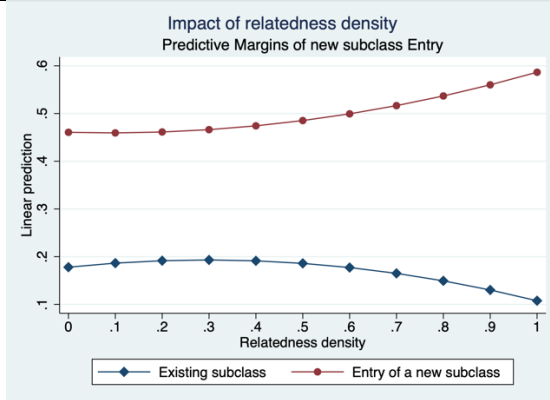


Fig 1d When the subclasses are unrelated

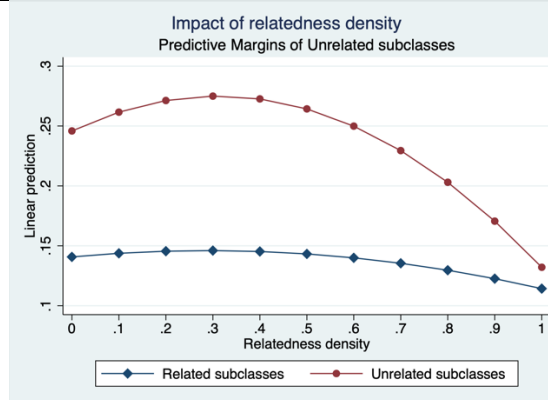


Fig 1e When the combined subclasses have not an RCA > 1

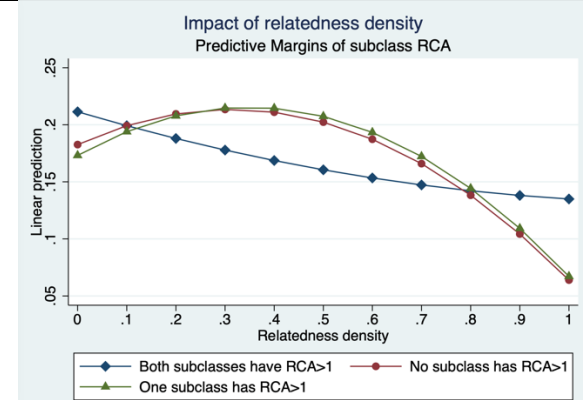


Table 4. Agency – Impact of relatedness density on new recombination.

	(1)	(2)	(3)	(4)	(5)
	b/se	b/se	b/se	b/se	b/se
Relatedness density	0.16*** [0.03]	0.24*** [0.05]	0.13*** [0.03]	0.14*** [0.03]	0.11*** [0.03]
Relatedness density # Relatedness density	-0.29*** [0.03]	-0.39*** [0.05]	-0.26*** [0.03]	-0.28*** [0.03]	-0.24*** [0.03]
New subclass	0.28*** [0.01]	0.28*** [0.01]	0.28*** [0.01]	0.28*** [0.01]	0.28*** [0.01]
University x Relatedness density	-0.05 [0.06]				
University x Relatedness density sq	0.01 [0.07]				
Experience (# patents) x Relatedness density		-0.02** [0.01]			
Experience (# patents) x Relatedness density sq		0.02** [0.01]			
Local applicant # Relatedness density			0.23*** [0.08]		
Local applicant # Relatedness density # Relatedness density			-0.24*** [0.08]		
Entry x Relatedness density				0.13 [0.09]	
Entry x Relatedness density sq				-0.12 [0.08]	
Entry in the region x Relatedness density					0.21*** [0.03]
Entry in the region x Relatedness density sq					-0.22*** [0.05]
Population density	-0.14 [0.19]	-0.17 [0.19]	-0.17 [0.19]	-0.17 [0.19]	-0.13 [0.19]
Income per employee	0.02 [0.06]	0.02 [0.06]	0.02 [0.06]	0.02 [0.06]	0.00 [0.06]
Employment	0.34** [0.14]	0.34** [0.14]	0.34** [0.14]	0.34** [0.14]	0.32** [0.14]
Technological generality	0.20+ [0.11]	0.19+ [0.10]	0.19+ [0.10]	0.19+ [0.10]	0.19+ [0.10]
Technological specialization	-0.02 [0.02]	-0.02 [0.02]	-0.02 [0.02]	-0.02 [0.02]	-0.02 [0.02]
Nb of patents in the subclass	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]	0.01*** [0.00]
Number of associated subclasses	-0.16*** [0.00]	-0.16*** [0.00]	-0.16*** [0.00]	-0.16*** [0.00]	-0.16*** [0.00]
Inventors with external collaborations	0.02*** [0.00]	0.02*** [0.00]	0.02*** [0.00]	0.02*** [0.00]	0.02*** [0.00]
Constant	-3.41+ [1.76]	-3.32+ [1.76]	-3.30+ [1.76]	-3.30+ [1.77]	-3.14+ [1.77]
Observations	264948	264948	264948	264948	264948
R-Squared	.22	.22	.22	.22	.22

Robust standard errors clustered at the patent level in brackets, + 0.10 ** 0.05 ***0.01

Patent, region fixed effects and combined subclass fixed effects

Figure 2. Impact of relatedness density on new recombination given the type of agents

Fig 2a. Size of applicants

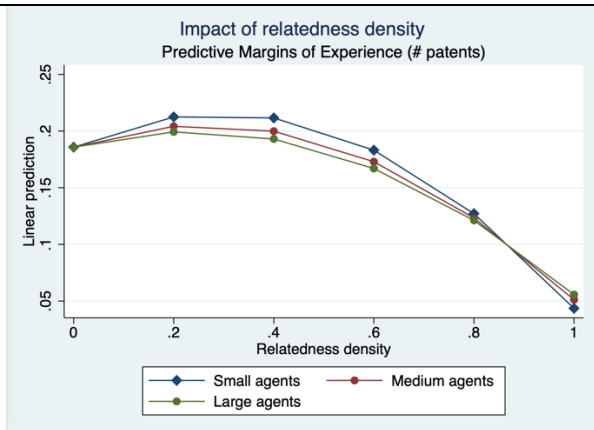


Fig 2b. Location of applicant

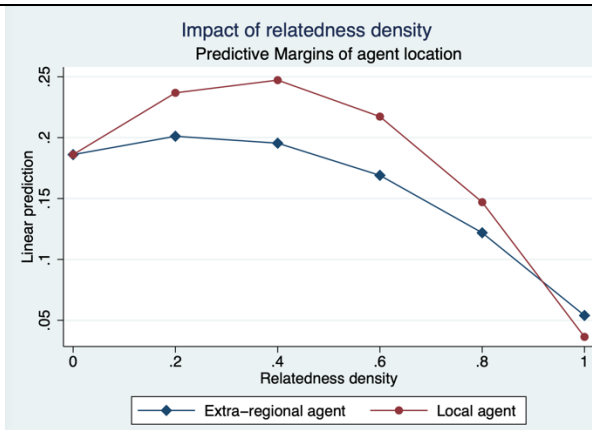


Fig 2c. Entry in the region

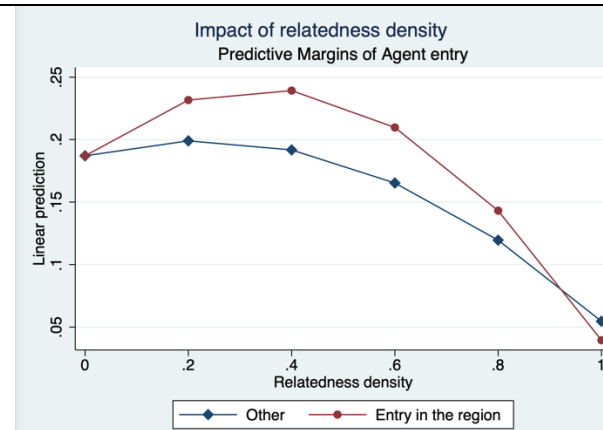


Table 5: Robustness Checks

	(1) Single Region ¹	(2) Multi Regions	(3) IDF and Rhône Alpes	(4) Outside IDF and Rhône Alpes	(5) Cross- field	(6) Similar Field	(7) Patent Novelty	(8) Without Extreme Values RD	(9) Relative Density <0
Regional characteristics									
Relatedness density	0.13*** [0.03]	0.07+ [0.04]	0.13*** [0.03]	0.09** [0.04]	0.08*** [0.03]	0.15*** [0.04]	0.11*** [0.04]	0.10*** [0.03]	
Relatedness density sq	-0.19*** [0.04]	-0.12*** [0.04]	-0.20*** [0.04]	-0.17*** [0.05]	-0.14*** [0.03]	-0.20*** [0.05]	-0.21*** [0.05]	-0.17*** [0.03]	
New subclass	0.26*** [0.01]	0.28*** [0.01]	0.36*** [0.04]	0.27*** [0.01]	0.26*** [0.01]	0.51*** [0.03]	0.21*** [0.01]	0.28*** [0.01]	0.28*** [0.01]
Unrelatedness (Relative density <0)									0.01*** [0.00]
Patent subclass recombination characteristics									
No subclass has RCA >1	0.04*** [0.00]	0.03*** [0.01]	0.02*** [0.01]	0.07*** [0.01]	0.04*** [0.00]	0.01 [0.01]	0.09*** [0.01]	0.04*** [0.00]	0.04*** [0.00]
One subclass has RCA >1	0.04*** [0.00]	0.04*** [0.00]	0.02*** [0.00]	0.06*** [0.00]	0.04*** [0.00]	0.01*** [0.00]	0.08*** [0.00]	0.04*** [0.00]	0.04*** [0.00]
Unrelated subclasses	0.11*** [0.00]	0.13*** [0.01]	0.08*** [0.00]	0.16*** [0.00]	0.12*** [0.00]	0.05*** [0.01]	0.19*** [0.00]	0.12*** [0.00]	0.12*** [0.00]
Regional characteristics - Controls									
Population density		-0.24 [0.19]	0.40 [2.05]	0.05 [0.30]	-0.07 [0.23]	-0.46** [0.22]	-0.21 [0.29]	-0.19 [0.19]	-0.20 [0.19]
Income per employee		0.04 [0.06]	-0.35 [0.41]	-0.60** [0.25]	0.05 [0.07]	0.02 [0.08]	-0.31*** [0.10]	0.04 [0.06]	0.05 [0.06]
Employment		0.35** [0.14]	-0.11 [0.95]	0.33 [0.23]	0.26 [0.17]	0.34** [0.16]	0.30 [0.22]	0.33** [0.14]	0.29** [0.14]
Technological generality		0.16 [0.10]	-0.09 [0.46]	0.02 [0.16]	0.23+ [0.12]	-0.01 [0.14]	0.25+ [0.15]	0.19+ [0.10]	0.18+ [0.10]
Technological specialization		-0.02 [0.02]	0.21 [0.19]	-0.01 [0.02]	-0.01 [0.02]	-0.05** [0.02]	0.01 [0.02]	-0.02 [0.02]	-0.02 [0.02]
Number of patents in the subclass	-0.01 [0.01]	0.03*** [0.00]	0.01 [0.01]	-0.01** [0.01]	0.01*** [0.00]	0.01 [0.01]	0.02*** [0.01]	0.02*** [0.00]	0.02*** [0.00]
Number of associated subclasses	-0.10*** [0.01]	-0.20*** [0.01]	-0.07*** [0.01]	-0.10*** [0.01]	-0.15*** [0.00]	-0.16*** [0.01]	-0.15*** [0.01]	-0.16*** [0.00]	-0.16*** [0.00]
Inventors with external collaborations	0.01 [0.00]	0.03*** [0.00]	0.01*** [0.00]	-0.00 [0.00]	0.02*** [0.00]	0.04*** [0.01]	-0.01 [0.00]	0.02*** [0.00]	0.02*** [0.00]
Constant	0.74*** [0.10]	-3.46+ [1.78]	2.58 [7.56]	2.35 [3.50]	-3.22 [2.06]	-1.88 [2.36]	0.81 [2.67]	-3.43+ [1.76]	-3.03+ [1.76]
Observations	168567	96381	142480	122468	223646	41302	124262	264948	264948
R-Squared	.19	.32	.19	.24	.24	.36	.3	.23	.23

Robust standard errors clustered at the patent level in brackets, + 0.10 ** 0.05 *** 0.01 -¹ In the single region specification, regional variables cannot be computed

Table 6: Robustness checks

	(1) Period 1	(2) Period 2	(3) Period 3	(4) Above_Median	(5) Below_Median	(6) nb comb <=15	(7) nb comb <=28
Regional characteristics							
Relatedness density	0.09+ [0.05]	0.06 [0.04]	0.08 [0.05]	0.09 [0.08]	0.10*** [0.03]	0.08*** [0.03]	0.10*** [0.02]
Relatedness density sq	-0.16*** [0.05]	-0.14*** [0.04]	-0.11** [0.05]	-0.21** [0.09]	-0.17*** [0.03]	-0.14*** [0.03]	-0.16*** [0.03]
New subclass	0.23*** [0.01]	0.31*** [0.01]	0.28*** [0.01]	0.25*** [0.01]	0.29*** [0.01]	0.28*** [0.01]	0.28*** [0.01]
Patent subclass recombination characteristics							
No subclass has RCA >1	0.04*** [0.01]	0.04*** [0.01]	0.03*** [0.01]	0.10*** [0.01]	0.03*** [0.00]	0.04*** [0.00]	0.04*** [0.00]
One subclass has RCA >1	0.04*** [0.01]	0.03*** [0.00]	0.03*** [0.00]	0.08*** [0.01]	0.03*** [0.00]	0.04*** [0.00]	0.04*** [0.00]
Unrelated subclasses	0.12*** [0.01]	0.10*** [0.00]	0.15*** [0.01]	0.16*** [0.01]	0.11*** [0.00]	0.12*** [0.00]	0.12*** [0.00]
Regional characteristics - Controls							
Population density	-1.31 [1.07]	-0.21 [0.72]	0.56 [0.70]	-1.36 [1.01]	0.00 [0.24]	-0.39** [0.19]	-0.31 [0.19]
Income per employee	-0.54 [0.56]	0.06 [0.26]	0.11 [0.13]	0.04 [0.59]	-0.01 [0.08]	0.05 [0.06]	0.05 [0.06]
Employment	0.19 [0.56]	1.12** [0.51]	0.51 [0.60]	0.42 [0.79]	0.30+ [0.17]	0.44*** [0.14]	0.34** [0.14]
Technological generality	-0.06 [0.32]	0.36 [0.30]	0.22 [0.24]	0.09 [0.31]	0.40*** [0.14]	0.17+ [0.10]	0.19+ [0.10]
Technological specialization	-0.12** [0.05]	0.05 [0.05]	-0.04 [0.04]	-0.07+ [0.04]	-0.00 [0.03]	-0.04** [0.02]	-0.03+ [0.02]
Number of patents in the subclass	0.01 [0.01]	0.02*** [0.01]	0.02*** [0.01]	-0.02** [0.01]	0.01*** [0.00]	0.02*** [0.00]	0.02*** [0.00]
Number of associated subclasses	-0.16*** [0.01]	-0.15*** [0.01]	-0.18*** [0.01]	-0.08*** [0.01]	-0.14*** [0.01]	-0.17*** [0.00]	-0.16*** [0.00]
Inventors with external collaborations	0.02*** [0.01]	0.03*** [0.00]	0.01*** [0.00]	-0.01 [0.01]	0.02*** [0.00]	0.02*** [0.00]	0.02*** [0.00]
Constant	10.01 [9.99]	-14.96** [7.46]	-10.88+ [5.90]	0.32 [8.39]	-3.47 [2.23]	-4.18** [1.72]	-3.09+ [1.74]
Observations	69985	116240	78723	48047	216901	230297	252431
R-Squared	.26	.26	.25	.27	.22	.22	.22
Robust standard errors clustered at the patent level in brackets, + 0.10 ** 0.05 ***0.01							

Table A1. Descriptive statistics and Correlation table

	mean	sd	min	max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1 Recombinant novelty (Region)	0.19	0.39	0.00	1.00	1.00																
2 Recombinant novelty (World)	0.01	0.09	0.00	1.00	0.19*	1.00															
3 Recombinant novelty (France)	0.06	0.23	0.00	1.00	0.50*	0.37*	1.00														
4 Recombinant reuse (#)	124.20	337.71	1.00	3089	-0.18*	-0.03*	-0.09*	1.00													
5 Recombinant reuse (log)	2.65	2.13	0.00	8.04	-0.61*	-0.11*	-0.31*	0.65*	1.00												
6 Relatedness density	0.39	0.19	0.00	1.00	-0.07*	-0.00*	-0.00	-0.16*	-0.06*	1.00											
7 Unrelated techn. (subclass)	0.41	0.49	0.00	1.00	0.31*	0.10*	0.23*	-0.20*	-0.43*	-0.01*	1.00										
8 New subclass	0.02	0.16	0.00	1.00	0.33*	0.09*	0.15*	-0.06*	-0.20*	-0.11*	0.10*	1.00									
9 Both specialized subclasses	0.31	0.46	0.00	1.00	-0.15*	-0.02*	-0.06*	0.13*	0.17*	0.50*	-0.07*	-0.10*	1.00								
10 No specialized subclass	0.31	0.46	0.00	1.00	0.08*	-0.00*	0.00	-0.12*	-0.09*	-0.44*	-0.09*	0.10*	-0.46*	1.00							
11 One specialized subclass	0.37	0.48	0.00	1.00	0.07*	0.02*	0.05*	-0.01*	-0.08*	-0.06*	0.15*	0.00	-0.52*	-0.52*	1.00						
12 Regional patent generality	0.36	0.04	0.21	0.45	-0.17*	-0.01*	-0.04*	0.12*	0.25*	0.14*	-0.05*	-0.09*	0.16*	-0.17*	0.01*	1.00					
13 Population density	5.54	1.08	3.76	6.89	-0.26*	-0.02*	-0.03*	0.31*	0.42*	-0.00*	-0.00	-0.16*	0.01*	-0.08*	0.07*	0.47*	1.00				
14 Income per employee	11.04	0.20	10.56	11.53	-0.21*	-0.03*	-0.05*	0.32*	0.37*	-0.05*	0.01*	-0.15*	0.01*	-0.08*	0.06*	0.05*	0.67*	1.00			
15 Employment	14.57	0.87	12.51	15.61	-0.27*	-0.01*	-0.02*	0.28*	0.40*	-0.02*	0.01*	-0.18*	-0.07*	-0.01*	0.07*	0.22*	0.87*	0.64*	1.00		
16 Subclass specialization	1.89	0.51	1.22	4.07	0.23*	0.01*	0.02*	-0.24*	-0.35*	-0.03*	-0.01*	0.17*	0.06*	0.00*	-0.06*	-0.15*	-0.75*	-0.59*	-0.88*	1.00	
17 Nb of patents in the subclass	4.31	1.74	0.00	7.89	-0.50*	-0.10*	-0.20*	0.52*	0.79*	0.03*	-0.18*	-0.28*	0.15*	-0.20*	0.05*	0.34*	0.63*	0.52*	0.63*	-0.56*	
18 Number of associated subclasses	3.85	1.20	0.00	5.77	-0.54*	-0.11*	-0.21*	0.36*	0.70*	0.07*	-0.16*	-0.34*	0.12*	-0.15*	0.03*	0.37*	0.62*	0.43*	0.64*	-0.57*	
19 Inventors with external collaborations	1.82	1.23	0.00	4.98	-0.42*	-0.09*	-0.19*	0.59*	0.80*	-0.10*	-0.22*	-0.19*	0.06*	-0.10*	0.04*	0.29*	0.54*	0.48*	0.53*	-0.47*	
20 Cross field	0.84	0.36	0.00	1.00	0.15*	0.03*	0.09*	-0.27*	-0.32*	0.08*	0.26*	0.03*	-0.03*	-0.02*	0.05*	-0.04*	-0.03*	-0.01*	-0.00*	0.01*	
21 Cross sector	0.59	0.49	0.00	1.00	0.16*	0.05*	0.12*	-0.17*	-0.26*	0.04*	0.19*	0.03*	-0.05*	0.02*	0.03*	-0.03*	-0.03*	0.03*	-0.03*	0.02*	
22 Entry	0.10	0.31	0.00	1.00	0.09*	0.02*	0.05*	-0.04*	-0.10*	-0.03*	0.02*	0.05*	-0.06*	0.06*	0.00*	-0.03*	-0.05*	-0.05*	-0.04*	0.04*	
23 Entry in the region	0.22	0.41	0.00	1.00	0.13*	0.01*	0.03*	-0.07*	-0.15*	-0.08*	-0.01*	0.09*	-0.11*	0.13*	-0.01*	-0.08*	-0.16*	-0.14*	-0.15*	0.13*	
24 Experience (# patents)	4.36	2.66	0.00	9.93	-0.14*	-0.04*	-0.07*	0.06*	0.15*	0.05*	0.00	-0.08*	0.09*	-0.08*	-0.01*	0.06*	0.11*	0.14*	0.09*	-0.08*	
25 Local applicant	0.13	0.33	0.00	1.00	0.11*	0.03*	0.07*	-0.06*	-0.13*	-0.01*	0.03*	0.06*	-0.05*	0.04*	0.01*	-0.04*	-0.05*	-0.06*	-0.03*	0.03*	
26 Local inventors	0.44	0.50	0.00	1.00	-0.00*	0.03*	0.05*	0.01*	0.01*	0.11*	0.08*	-0.02*	0.08*	-0.12*	0.03*	0.02*	0.17*	0.09*	0.22*	-0.15*	
27 University	0.18	0.38	0.00	1.00	-0.05*	-0.02*	-0.03*	0.04*	0.10*	-0.08*	-0.10*	-0.02*	-0.11*	0.11*	0.00	-0.04*	-0.01*	0.04*	0.05*	-0.07*	

* $p < 0.05$

Table A1. Descriptive statistics and Correlation table

	17	18	19	20	21	22	23	24	25	26	27
17 Nb of patents in the subclass	1.00										
18 Number of associated subclasses	0.93*	1.00									
19 Inventors with external collaborations	0.91*	0.79*	1.00								
20 Cross field	-0.17*	-0.09*	-0.21*	1.00							
21 Cross sector	-0.16*	-0.12*	-0.18*	0.52*	1.00						
22 Entry	-0.11*	-0.11*	-0.09*	0.03*	0.05*	1.00					
23 Entry in the region	-0.20*	-0.20*	-0.14*	0.01*	0.06*	0.65*	1.00				
24 Experience (# patents)	0.19*	0.20*	0.16*	-0.04*	-0.10*	-0.41*	-0.38*	1.00			
25 Local applicant	-0.13*	-0.14*	-0.12*	0.04*	0.05*	0.70*	0.46*	-0.54*	1.00		
26 Local inventors	0.08*	0.06*	-0.03*	0.06*	0.02*	-0.02*	-0.19*	-0.01*	0.07*	1.00	
27 University	0.09*	0.09*	0.13*	-0.01*	0.05*	0.00	0.04*	0.22*	-0.12*	-0.11*	1.00

Table A2

	Cluster 1	Cluster 2	Logit	Probit
	(1)	(2)	(3)	(4)
Regional characteristics				
Relatedness density	0.10*** [0.02]	0.10*** [0.03]	0.60*** [0.20]	0.35*** [0.10]
Relatedness density sq	-0.17*** [0.03]	-0.17*** [0.03]	-0.99*** [0.22]	-0.57*** [0.12]
New subclass	0.28*** [0.01]	0.28*** [0.01]	0.00 [.]	0.00 [.]
Patent subclass recombination characteristics				
No subclass has RCA >1	0.04*** [0.01]	0.04*** [0.00]	0.69*** [0.03]	0.37*** [0.02]
One subclass has RCA >1	0.04*** [0.01]	0.04*** [0.00]	0.59*** [0.02]	0.32*** [0.01]
Unrelated subclasses	0.12*** [0.02]	0.12*** [0.00]	1.39*** [0.02]	0.78*** [0.01]
Regional characteristics - Controls				
Population density	-0.19 [0.24]	-0.19 [0.18]	-1.04 [0.67]	-0.54 [0.37]
Income per employee	0.04 [0.06]	0.04 [0.06]	-1.64*** [0.11]	-0.94*** [0.06]
Employment	0.33*** [0.11]	0.33** [0.13]	1.26*** [0.35]	0.68*** [0.19]
Technological generality	0.19 [0.11]	0.19+ [0.10]	0.50 [0.38]	0.26 [0.21]
Technological specialization	-0.02+ [0.01]	-0.02 [0.01]	0.08 [0.05]	0.06+ [0.03]
Nb of patents in the subclass	0.02** [0.01]	0.02*** [0.00]	-0.10*** [0.02]	-0.03** [0.01]
Number of associated subclasses	-0.16*** [0.01]	-0.16*** [0.00]	-0.84*** [0.03]	-0.51*** [0.01]
Inventors with external collaborations	0.02*** [0.00]	0.02*** [0.00]	-0.16*** [0.02]	-0.08*** [0.01]
Constant	-3.34 [1130.22]	-3.34 [7286.47]	5.77+ [3.27]	3.52** [1.79]
Observations	264948	264948	258077	258077
R-Squared	.23	.23		

Robust standard errors clustered at the patent level in brackets, + 0.10 ** 0.05 ***0.01

Table A3. Robustness check – Impact of Relatedness density on combinations new to the world and New to France

	New to the world			New to France		
	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Relatedness density	0.004+	0.036***	0.030***	-0.005	0.140***	0.116***
	[0.002]	[0.008]	[0.008]	[0.005]	[0.016]	[0.016]
Relatedness density # Relatedness density		-0.038***	-0.032***		-0.171***	-0.130***
		[0.009]	[0.009]		[0.018]	[0.018]
New subclass	0.011***	0.012***	0.012***	0.047***	0.048***	0.048***
	[0.003]	[0.003]	[0.003]	[0.006]	[0.006]	[0.006]
No subclass has RCA >1			-0.002***			0.002
			[0.001]			[0.002]
One subclass has RCA >1			0.000			0.006***
			[0.001]			[0.002]
Unrelated subclasses			0.005***			0.051***
			[0.001]			[0.002]
Regional patent generality	0.002	-0.001	-0.001	0.014**	-0.002	0.000
	[0.003]	[0.003]	[0.003]	[0.007]	[0.008]	[0.008]
Population density	-0.005	-0.003	-0.000	-0.017+	-0.008	-0.005
	[0.004]	[0.004]	[0.004]	[0.009]	[0.010]	[0.011]
Income per employee	-0.009***	-0.009***	-0.011***	-0.027***	-0.023***	-0.023***
	[0.002]	[0.002]	[0.002]	[0.005]	[0.005]	[0.005]
Employment	0.004+	0.010***	0.010***	0.004	0.030***	0.026***
	[0.002]	[0.003]	[0.003]	[0.006]	[0.008]	[0.008]
Subclass specialization	0.000	0.001**	0.001**	0.000	0.003***	0.003***
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
Nb of patents in the subclass	-0.002**	-0.002**	-0.003***	-0.005**	-0.006***	-0.008***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Number of associated subclasses	-0.003**	-0.003**	-0.003**	-0.001	-0.001	-0.001
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Inventors with external collaborations	0.001	0.001	0.001	0.001	0.000	-0.001
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Constant	0.102***	-0.005	-0.000	0.385***	-0.102	-0.102
	[0.039]	[0.045]	[0.045]	[0.101]	[0.118]	[0.117]
Observations	264948	264948	264948	264948	264948	264948
R-Squared	.095	.095	.095	.14	.14	.15