Quality Dimensions of Knowledge and Regional Development

Relatedness, Complexity, Novelty, and Impact of Knowledge

Lars Mewes

Quality Dimensions of Knowledge and Regional Development

Relatedness, Complexity, Novelty, and Impact of Knowledge

Kwaliteitsdimensies van Kennis en Regionale Ontwikkeling

Verwantschap, Complexiteit, Nieuwheid, en Invloed van Kennis

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

donderdag 14 november 2019 des middags te 4.15 uur

Lars Mewes

geboren op 22 oktober 1988
te Perleberg, Duitsland

Promotor:	Prof. dr. R.A. Boschma
Copromotoren:	Dr. T. Brökel
	Dr. P.M.A. Balland

This thesis was (partly) accomplished with financial support from Lower Saxony's Ministry of Science and Culture (MWK).
Dit proefschrift werd mede mogelijk gemaakt met financiële steun van Nedersaksen Ministerie van Wetenschap en Cultuur (MWK).

Committee: Prof. dr. U. Cantner
Prof. dr. K. Frenken
Prof. dr. J.H. Garretsen
Prof. dr. F. van Oort
Prof. dr. R. Veugelers

Contents

1 Introduction

1.1 Motivation – technological knowledge, human development, and geography

On 17 December 1903, four miles south of Kitty Hawk, North Carolina, on a windy and cold winter day, a powered airplane took off and flew for 12 seconds - wobbly and brief. Onboard the small plane: Orville Wright. Shortly afterwards, his brother Wilbur repeated the attempts and even lifted the Wright airplane off the ground for 59 seconds. On this historic day, the Wright brothers completed the first powered manned air flight, a technological milestone in human aviation history (Mohler 2004). In June 1919, sixteen years after the Wright brothers, John Alcock and Arthur W. Brown landed in Clifdon, Ireland, sixteen hours after they took off in St. John, Newfoundland. They accomplished the first non-stop flight across the Atlantic Ocean (New York Times 1919). The blink of an eye later in terms of human history, in 1961, the cosmonaut Yuri Alekseyevich Gagarin entered outer space. His journey was part of the moon race between the Soviet Union and the United States, which ended on 21 July 1969. 357,923 kilometers from Houston mission control center, Apollo 11 touched down on an unknown place on which no human had ever before set foot – the moon. Neil Armstrong and Buzz Aldrin were to be the first humans to walk on the moon 66 years after the Wright brothers flew 260 meters above the ground.

Scientific discoveries and technological progress are cornerstones of human development and deeply shape our society. We cannot imagine our world without technology and without its advancement. The speed at which technological improvements occur has significantly increased since the industrial revolution. For centuries of human history, horses and horse carriages were the dominant transport vehicle - from the Roman Empire in the first century AD to the British Empire 1,900 years later. More than 300,000 horses in London and more than 100,000 in New York worked for the cities' transportation systems around 1900. The dependence on horses for the transportation of people and goods even caused a manure problem that began to poison the local population (Johnson 2015). But in the space of just a few years, horses were replaced by cars, Gagarin entered the cosmos, and Armstrong and Aldrin walked on the moon.

Groundbreaking milestones such as the first powered flight and the moon landing stand out in the history of technology and they help to balance the bigger picture. The rapid succession of such breakthroughs in the blink of an eye in terms of human history suggests that the speed of technological progress has dramatically increased. The acceleration of human development is not restricted to such illustrative examples, but concerns technological development in general. One way to quantify

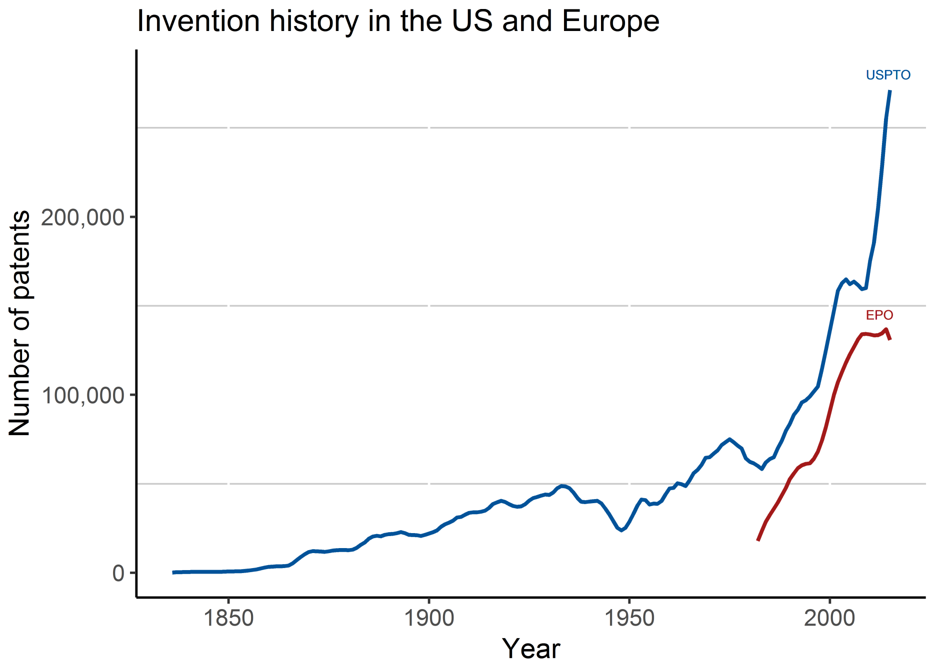

Figure 1.1: Number of inventions at the USPTO since 1836 and EPO since 1975 (data source: USPTO and OECD REGPAT Database for EPO patents)

technological development is to use data on patented inventions (Griliches 1990). Figure 1.1 visualizes the development of technological knowledge from 1836 to 2017 using patents from the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO). Patents document the what, when, and where of technological development and are widely used in empirical research (Griliches 1990; Feldman 1994; Acs et al. 2002; Verspagen and Schoenmakers 2004). The figure not only indicates that the annual number of inventions has continuously increased over the last century, but that patent growth shows exponential features. Accordingly, technological progress has significantly accelerated since the 1950s and technological knowledge plays an important role in our society.

Technological development, more generally, influences our daily lives and well-being. Technical achievements in agriculture allow an increasing world population to be supplied with food (Godfray et al. 2010) and advances in the life sciences have increased average life expectancy in most parts of the world in the last two centuries (Cervellati and Sunde 2005). Technological knowledge and its development certainly also has negative effects, such as increasing carbon dioxide emissions since the industrial revolution (Lamarque et al. 2010), but nevertheless generally contributes to a better life.

The effects of technological knowledge also include economic development. Abramovitz (1956) and Solow (1957) showed that conventional factors such as capital and labor cannot explain 90 percent of economic growth observed in developed, industrialized countries. Instead of capital and labor accumulation, it was argued that the unexplained residual must be due to other factors such as productivity growth. In his growth model, Solow (1957) attributed productivity growth to technical change, which allowed capital and labor to be more productive. Solow's model is called the exogenous growth model, as technological change is not included directly, but was present exogenously. Since then, much research has been devoted

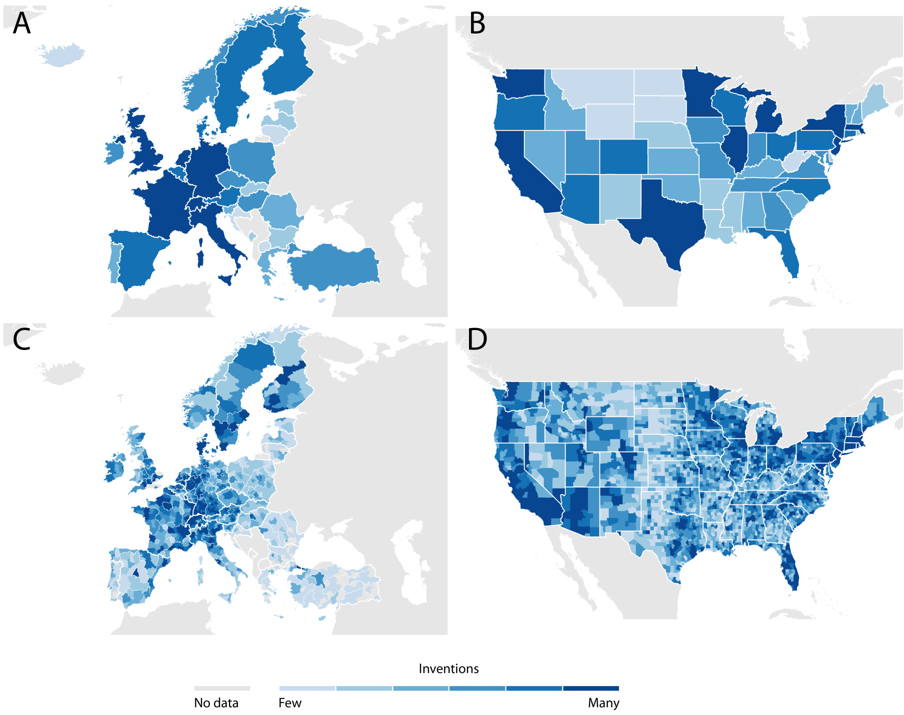

Figure 1.2: Invention maps of **A** European countries, **B** US States, **C** European NUTS 3 regions, and **D** US counties showing the number of inventions between 2011 and 2015 (data source: OECD REGPAT Database for EPO patents and USPTO)

to including technological knowledge and progress into economic growth theories and to explaining technological development (Nelson and Winter 1982; Romer 1990; Grossman and Helpman 1991; Aghion and Howitt 1998).

Economic growth, however, is not equally distributed across countries, and even regions within countries show substantial variation (Fagerberg et al. 1997). One fundamental reason for different levels of economic growth is the uneven distribution of knowledge creation in regions (Glaeser et al. 1992; Henderson et al. 2001). Mapping invention activities, for example, in Europe and the USA (see Figure 1.2) reveals that regions within countries differ greatly in their ability to produce new knowledge. Hence, countries represent relatively large-scale spatial units that hide much of the prevalent regional variation of invention activities. Figure 1.2, secondly, highlights that invention activities concentrate in particular hotspots of knowledge creation such as Silicon Valley in California. The spatial concentration of knowledge has been documented in several empirical studies (Feldman 1994; Acs et al. 2002; Verspagen and Schoenmakers 2004; Moreno et al. 2005). In models of regional growth, knowledge, for example, indicated by the number of inventions, number of R&D employees, or total R&D expenditures, is therefore often considered as an important variable with which to explain regional economic success (e.g. Barrell and Pain 1997; Rodríguez-Pose 1999; Crescenzi 2005; Audretsch and Keilbach 2008; Parent and LeSage 2012; Piergiovanni et al. 2012). Hence, explaining the spatial concentration of knowledge and thus differences of knowledge production in regions informs our understanding of the uneven economic development in regions.

But why does knowledge so pervasively concentrate in space? The spatial concentration of knowledge and the importance of location seems extremely paradoxical in an era in which modern information and communication technology (ICT) and global airline networks increasingly reshape our perception of distance. Amsterdam, for example, might be more tightly connected to New York than to Groningen (Brockmann and Helbing 2013). Although advances in ICT and low-price air travel have moved geographically dispersed places closer together, firms still tend to colocate in close geographic proximity (Rodríguez-Pose and Crescenzi 2008).

To understand the benefits of co-location and therefore the spatial clustering of innovation, it is necessary to understand fundamental elements of knowledge and knowledge production. New knowledge is rarely entirely new, but rather builds on existing knowledge. When the Wright brothers conducted their first flight attempts, they could build on previous discoveries and developments of aviation pioneers such as George Cayley, Percy Pilcher, Otto Lilienthal, or Hiram Maxim (Mohler 2004). The metaphor of dwarfs standing on the shoulders of giants nicely expresses what theories of innovation and technological change teach us: new knowledge is often the result of the combination of existing pieces of knowledge (Usher 1954; Kuznets 1962; Nelson and Winter 1982; Utterback 1996). Deconstructing the Wright airplane reveals several essential building blocks: the wing design, the propellers, the combustion engine, and the control system. Many of these components were improved by the Wright brothers, but their technological origins date back to foregone inventors (Mohler 2004). Hence, existing knowledge is a key resource for new knowledge and innovation.

Combining knowledge components requires interaction with other people in order to solve complex problems (Wuchty et al. 2007; Aunger 2010). The Wright brothers did not work on their plane in lonesome ingenuity, but collaborated with specialized engineers to advance their dream of powered aviation (Mohler 2004). However, not all knowledge is alike, and some knowledge is more difficult to communicate and share than other forms of knowledge. The concept of tacitness (Polanyi 1966) describes knowledge that is immensely difficult if not impossible to codify, as it is deeply embedded in individual memories and routines. For example, the Wright brothers conducted a series of 700 to 1,000 flights between 1902 and 1903, equipping them with extraordinal piloting skills particularly suited for their plane (Mohler 2004). Teaching others how to fly their aircraft might have been impossible given their experience. Other prominent examples of tacit knowledge are swimming or cycling. It is impossible to teach someone how to cycle by only using words, but it is possible to cook a delicious meal by following a recipe, which represents codified knowledge about cooking the meal. Thinking of tacitness as a degree rather than a category allows the conceptualization of knowledge ranging from fully tacit (e.g. cycling) to fully codified (e.g. recipe) (Cowan et al. 2000). Codifying knowledge thereby increases the speed and decreases the cost of transaction (Zander and Kogut 1995). More tacit knowledge, in contrast, remains highly exclusive to its owner (Maskell and Malmberg 1999).

The tacit dimension of knowledge therefore constrains knowledge flows in space and provides a clear link to economic geography. Exchanging and combining tacit knowledge requires face-to-face interaction between individuals, which is difficult over long distances (Gertler 2003). Urban regions with their agglomerations of skilled people in close geographic proximity provide a fruitful environment for intense knowledge interaction (Feldman and Audretsch 1999). Although knowledge flows are subtle and thus often elude a clear observation, empirical research shows that they are more likely to occur in close geographic proximity (Jaffe et al. 1993). The geographic stickiness of knowledge hinders rapid dissemination and binds knowledge to specific places (Hippel 1994).

Knowledge can spill over not only via intended knowledge exchange (e.g. formal collaboration), but also due to unintended knowledge flows. Another crucial attribute of knowledge therefore concerns its excludability or non-excludability respectively. If knowledge is excludable, firms investing in knowledge creation and innovation exclusively benefit from their outcomes or restrict knowledge flows to partners. In contrast, if knowledge is non-excludable, it can spill over to third parties, intended or unintended, that did not pay for the knowledge creation process (Grossman and Helpman 1991). Firms can therefore benefit from investments in innovation on the part of others by co-locating to them in close geographic proximity. Knowledge thereby flows along specific channels, as firms and their employees are, for example, linked to other firms in the region via formal (e.g. business partnerships, research collaborations) and informal (e.g. friendships, occasional interactions) relations, increasing the likelihood of knowledge exchange in close geographic proximity (Granovetter 1985; Uzzi 1997; Bathelt et al. 2004).

Hence, location is important for knowledge production! Places, however, are not equal, differing substantially in important prerequisites of knowledge generation. Population size is one fundamental characteristic in which places are different and which has concrete implications for knowledge production. New knowledge might emerge in many places, but is produced at faster rates and in larger quantities in cities. The productivity of larger cities has been elaborated upon and documented in the early work of Kuznets (1960). The core argument is that larger cities are more productive than smaller towns regarding a variety of socio-economic phenomena due to the density of population and skills. In a recent approach, Bettencourt et al. (2007a) borrows the concept of scaling from biology and explores how socio-economic outcomes scale with cities' population size. Scaling analyses suggest that invention activities and the number of inventors grow disproportionately with city size. More precisely, doubling the population of a city results in more than twice as many inventions and inventors, indicating a super-linear relationship and illustrating the productivity of cities in purely quantitative terms.

The economic reasoning for the productivity of cities is often associated with different types of externalities in urban environments. Urbanization externalities, for example, describe the advantages that arise from the pure size of cities such as the infrastructure including transportation networks, education facilities, financial institutions, or the large local labor market allowing firms to seek qualified employees (Mills 1967; Henderson 1974; Sveikauskas 1975). In addition, the central place theory (Christaller 1933) locates cities within a hierarchical city network according to their functionalities. All the functions of smaller towns can be found in larger ones but not vice versa. Larger cities such as New York, London and Singapore, with their rich infrastructure of financial institutions, top universities, and research facilities, attract highly educated and skilled people. The density of heterogeneous people in general and the diversity of skilled individuals more specifically increases the likelihood of interaction of diverse knowledge in large cities (Jacobs 1969).

Chinitz (1961) argues that beyond cities' pure size, their economic structure is particularly important to explain the benefits of co-localization. Therefore, two other types of externalities describe how the benefits of co-location are linked to knowledge spillover and the economic structure of places. Firstly, Marshall-Arrow-Romer externalities describe that the spatial concentration of firms within the same or similar industry provides a variety of advantages for co-located firms (Marshall 1890; Arrow 1962; Romer 1986). Besides labor market pooling and input-output linkages, this concerns knowledge spillovers. The main argument is that valuable knowledge primarily spills over between firms within the same industry. Hence, firms benefit from the co-location of firms of the same industry. Secondly, Jacobs externalities state that crucial sources of knowledge spillovers are external to the industry (Jacobs 1969). Accordingly, the advantage of co-location is not so much related to specialization, but to local diversity, which is highly related to the population size of cities. A diverse pool of knowledge in cities provides opportunities for cross-fertilization of knowledge between different industries and stimulates the emergence of entirely new ideas. Current empirical research is not in favor of one or the other externality (Beaudry and Schiffauerova 2009).

The concepts of relatedness and related variety build a bridge between Marshall and Jacobs externalities (Frenken et al. 2007; Boschma and Frenken 2011; Neffke et al. 2011). Due to bounded rationality and absorptive capacity, actors cannot use any knowledge that spills over. Recipients need a cognitive base to communicate, understand, and process new knowledge effectively (Cohen and Levinthal 1990). According to Nooteboom et al. (2007), actors benefit most from knowledge exchange if their knowledge bases are neither too similar nor too distant, but related to each other. The so-called optimal cognitive distance describes the cognitive proximity when learning is most effective. These findings elevated the concept of cognitive proximity from the micro scale of individual learning of firms (Nooteboom et al. 2007) to the macro scale of collective learning of regions and countries (Frenken et al. 2007; Hidalgo et al. 2007). Accordingly, an economic structure characterized by related variety should induce the fruitful exchange of ideas and stimulate regional development (Frenken et al. 2007).

Economic structures are, however, not static, but change over time. While some regions undergo severe structural changes and lose their former leading role in innovation (e.g. Detroit), others prosper and become epicenters of knowledge creation (e.g. Silicon Valley). An evolutionary perspective explicitly considers the heterogeneity of places with all their economic and socio-institutional characteristics that have evolved over time and that are fundamental to understanding current economic structures in regions (Boschma and Frenken 2006; Martin and Sunley 2006). Acquiring new knowledge depends strongly on existing knowledge, its combination, and invested resources. Economic actors develop specific routines and heuristics to manage the riskiness and uncertainty of knowledge creation (Nelson and Winter 1982). Knowledge production therefore becomes strongly path-dependent and accumulates over time. As a consequence, economic actors cannot jump between any activity, but only gradually adjust or change their current activities (Neffke et al. 2011).

At the regional level, path dependency expresses that past developments shape today's structures. Hence, understanding current regional developments is only possible by understanding the shadow of the past. In economic geography, a path often describes the development of a particular industry or technology at a specific place with all its institutional endowments. In the beginning, small historic events initialize self-reinforcing processes that lead to the creation and establishment of new development paths in regions (Maskell and Malmberg 1999; Martin and Sunley 2006). Path dependency also explains the downfall of old industrial regions. Regional lock-ins occur when regions are dependent on and trapped in maturing industries. The Ruhr Area, for example, was strongly dependent on coal mining and heavy machinery, industries without much of a future in Germany, but most of the local knowledge and institutional setting was tailored to this specific industry (Grabher 1993). A deviation from such developments by creating and establishing new sustaining paths is difficult and protracted. Regional competences therefore do not change fast, but evolve over time.

Pairing our understanding of path dependency and tacitness with the advantage of agglomeration helps to understand why knowledge activities concentrate in specific places. Over time, regions not only accumulate knowledge and skills, but also infrastructure and institutions that sustain their regional development paths. Maskell and Malmberg (1999) used the term localized capabilities to emphasize that many of these competences cannot simply be transferred to other places due to their path-dependent and tacit nature. Present capabilities therefore shape future opportunities by attracting specific sectors for which they are particularly suited or by providing ideal conditions for new path creation. As a result, certain knowledge concentrates at specific places in which agglomeration externalities exhibit their forces and thereby contribute to a spatial clustering of industries (Malmberg et al. 1996; Maskell and Malmberg 1999). Hence, places develop certain capabilities, specialize in specific industries, and produce different knowledge outcomes.

1.2 Qualitative differences of technological knowledge

Although existing approaches often highlight the heterogeneity of places and knowledge creation, the empirical literature tends to focus on knowledge creation as either a pure quantitative outcome or input. Precisely, knowledge usually enters models of knowledge production and diffusion or economic development as a pure quantity without considering differences in the quality of knowledge (O'hUallichain 1999; Rodríguez-Pose 1999; Acs et al. 2002; Fritsch 2002; Crescenzi 2005; O'hUallichain and Leslie 2005; Buesa et al. 2006; Bettencourt et al. 2007a; Bettencourt et al. 2007b). Research from various disciplines, however, emphasizes that knowledge is not alike, showing substantial differences regarding quality (Trajtenberg 1990; Chandy and Tellis 1998; Hargadon 2003).

The Wright brothers' achievement in powered flying reset the prevalent imagination and opportunities of mobility around 1900. Shortly after their first powered flight, airlines began to connect the world, reshaped the perception of geographic distances, and gave birth to the modern aviation industry (Pirie 2009). Some domains are inherently complex to understand and involve problems that are only solved by tremendous efforts in science and technology. For example, no private businessman ever tried to land on the moon. The USA, as a nation, entered the space race when the Soviets launched Gagarin as the first man into space in 1961. Only ten years later, NASA put Armstrong and Aldrin on the moon. The moon landing is certainly a milestone in the human history of space and aviation technology. Its realization in such a short amount of time definitely represents a complex endeavor involving numerous scientists, engineers, and large amounts of public funding in order to master the grand technical challenges (Stine 2009). The Apollo mission substantially advanced countless technologies (e.g. information and communication technology) and reshaped existing understandings of what is technologically feasible (Mazzucato 2014).

Qualitative differences of new knowledge might be related to the heterogeneity of places. Places differ in their ability to produce outstanding technological improvements, and these differences potentially impact regional economic development. For example, Carl Benz, Gottlieb Daimler and Wilhelm Maybach advanced automobile technology in the 1880s and 1890s. Only 70 kilometers apart from each other, Benz in Mannheim and Daimler together with Maybach in Stuttgart, developed the first automobiles. Benz' patent of the first automobile can be considered as the birth of the global automotive industry. Shortly after, Daimler and Maybach founded the Daimler Motoren Gesellschaft in Stuttgart and sold their first automobile in 1892. Later, the company became the world-renowned Daimler AG. Today, the headquarters of Daimler, including its large production site, are still located in the region of Stuttgart. The regions now hosts more than 3,000 specialized suppliers with approximately 200,000 employees working in the regional automotive industry (Strambach and Klement 2013).

These illustrations show that knowledge and innovation greatly differ in quality, and they emphasize that quality is not a one-dimensional concept. In fact, the literature provides a range of dimensions according to which knowledge quality can be differentiated. For instance, new knowledge production relies on the combination of existing knowledge. Relatedness highlights that not every technology, or more generally, any piece of knowledge can be combined with the same effort and with the same success. Relatedness between technologies facilitates effective communication and learning and reduces uncertainties and risks (Nooteboom et al. 2007; Frenken et al. 2007; Neffke et al. 2011). Alternatively, knowledge differs in its underlying complexity. Advancing and connecting the multitude of different technologies required for the moon landing might have been more complex than developing the PageRank algorithm for Google's search engine (Fleming and Sorenson 2001). Knowledge also varies in its degree of novelty. Some inventions introduce radically new products or processes such as Gutenberg's printing press, whereas others add marginal improvements or represent incremental modifications (Chandy and Tellis 1998). Knowledge is also heterogeneous in its impact on the economy, society and technology (Trajtenberg 1990). The invention of the first automobile or the first powered airplane led to the emergence of new industries and reshaped existing technological paradigms. Other inventions do not even make it to the market. Hence, *(a) relatedness*, *(b) complexity*, *(c) degree of novelty*, and *(d) impact* represent four important dimensions of knowledge quality, on which the present thesis centers. More precisely, each of the four dimensions represents the core of a chapter, introduced in more detail in the following section.

1.2.1 Relatedness

Regions undergo continuous change and relatedness plays an important role in the evolution of regional economic structures (Boschma and Frenken 2011; Neffke et al. 2011). New industries emerge, while others diminish. This continuous change is linked to technological progress. New technological knowledge gives birth to new industries, such as the invention of the automobile around 1900 in Southwest Germany, and makes other sectors obsolete, such as the coach making industry. Boschma and Wenting (2007) studied how and where the British automobile industry emerged. According to their findings, regions were more likely to acquire new competences in the production of automobiles when they could build on strong competences in the manufacturing of bicycles and coaches. Regions equipped with related competences (e.g. bicycle and coach making) had more favorable starting conditions than regions missing such knowledge.

The concept of knowledge relatedness is therefore important for understanding the collective learning and evolution of regional capabilities as a path-dependent process. Regions are unlikely to acquire new knowledge independent of existing competences. Hidalgo et al. (2007) show that countries are more likely to export new products if they are related to their existing export portfolio. Beyond product diversification of countries, relatedness also shows its relevance for the regional level and a variety of different indicators (Neffke et al. 2011; Boschma et al. 2015; Klement and Strambach 2019). Hence, the empirical literature presents ample evidence that related diversification is the norm rather than the exception (Hidalgo et al. 2018).

The notion of relatedness and related diversification provides solid arguments for place-based policies and tailored policy schemes in favor of related diversification. For instance, the Smart Specialization policy of the EU requires regions to identify their existing strength and future opportunities in order to receive public support (Foray et al. 2011). In case of Smart Specialization, policy supports the path-dependent process of related diversification for which there might be solid arguments, as regions specialize and build competitive advantages in specific domains while policy reduces risks of failed investments (Martin and Sunley 2006). However, related diversification as one expression of path dependency might not be able to prevent potential regional cognitive-lock ins. Rather than supporting existing strengths, as evident in the current EU strategy, regional policy might also be well-advised to support unrelated instead of related diversification to increase regional resilience against cognitive lock-ins and external shocks (Frenken et al. 2007). Irrespective of the diversification type, i.e. related or unrelated, it still remains an open question as to what extent policy can intervene in the path-dependent process of regional diversification. Boschma and Gianelle (2014, p. 6) phrased this as the "one-million-dollar" question.

Chapter 2 takes this research gap as a starting point and asks: *Do publicly funded R&D projects break the path dependency of collective learning in regions?* The empirical approach in Chapter 2 relies on patents from the OECD REGPAT Database as an indicator of technological knowledge and on information on subsidized R&D projects from the German Federal Ministry of Education and Research (BMBF). The dataset has been used in a series of previous works studying the relationship between government support and innovation in regions (Fornahl et al. 2011; Broekel and Graf 2012; Broekel 2015). The BMBF data entails information about the recipients, purposes, locations and duration of the funded projects. A self-constructed concordance based on the matched-patent-subsidies-firm database of the Halle Institute of Economic Research links the data of subsidized R&D projects to patented inventions. With this data at hand, the link between publicly funded R&D projects and regional diversification is investigated for 141 German labor market regions between 1991 and 2010. The results confirm prior studies: relatedness is an important explanation for diversification in German labor market regions. Regarding innovation policy, R&D subsidies are more likely to be allocated to related activities. In addition, subsidized R&D projects are found to be positively associated with regional diversification. Previous research emphasized the different effects of R&D subsidies when granted to individual and collaborative research projects (Broekel and Graf 2012; Broekel et al. 2017). The BMBF data allowed such a distinction and the empirical results suggest that collaborative R&D more strongly influence regional diversification than individual R&D projects. Moreover, collaborative R&D can to some extent compensate for missing relatedness by increasing the likelihood of successful entries when relatedness density is low. Although policy is part of the path dependency of collective learning in regions, as it is more likely to allocate resources to related activities, joint research projects can nevertheless facilitate diversification into unrelated activities.

1.2.2 Complexity

Classical models of endogenous growth suggest that an increase in R&D resources should increase the growth rate of an economy (Romer 1990; Aghion and Howitt 1998). Jones (1995), however, observed that US growth rates experienced no significant increase although research intensity and educational attainment has grown substantially over the long term. Pintea and Thompson (2007) relate this paradox to the rising complexity of technology and technological development. Technologies are said to be complex if they consist of a large number of components and require large amounts of information for their reproduction (Simon 1962; Winter 1987; Zander and Kogut 1995). Complex technologies therefore require the functioning of multiple, interdependent components, and small errors can cause large problems (Sorenson et al. 2006). The learning model of Jovanovic and Nyarko (1995) indicates that learning in complex domains is therefore more difficult, slower, and increasingly demands more R&D resources.

In today's knowledge economy, knowledge represents a key resource and knowledge complexity is argued to have concrete economic consequences. Knowledge that is simple and easy to copy is less likely to equip economic actors with profound growth potentials. Complex knowledge, in contrast, is less likely to spill over to competitors, as it exhibits stronger tendencies to resist diffusion and is therefore likely to represent a valuable resource (Sorenson et al. 2006). Economic actors equipped with competitive advantage in complex domains are therefore likely to earn the economic benefits (Kogut and Zander 1992; Zander and Kogut 1995).

Although the economic consequences of technological complexity have been well discussed, the empirical evidence, however, is still scarce and unsatisfying for two primary reasons. Firstly, empirical approaches investigating the impact of technological complexity on regional economic development are restricted to indirect evidence either by relying on economic complexity as an implicit measure of technological complexity (Hidalgo and Hausmann 2009; Hausmann et al. 2013; Bahar et al. 2014) or by indirectly relating technological development with corresponding economic growth (Petralia et al. 2017; Balland et al. 2019). Secondly, the empirical evidence is primarily restricted to the country level (Hidalgo and Hausmann 2009; Hausmann et al. 2013; Bahar et al. 2014; Petralia et al. 2017). Regions within countries, however, show substantial differences in their capability to produce complex technologies (Balland and Rigby 2017). Countries therefore represent rather crude spatial units, which neglect substantial regional variation within countries. Hence, to what extent technological complexity is linked to the economic development of regions remains unexplored.

The empirical investigation in Chapter 3 is motivated by this research gap and particularly asks: *Are complex technologies important for regional economic development?* The empirical approach relies on GDP per capita as an indicator of economic growth in 166 European NUTS 2 regions between 2000 and 2015. The OECD REGPAT Database provides information on patented inventions and serves as the indicator of technological knowledge. Technological complexity is calculated based on the complexity index developed by Broekel (2019) called structural diversity. The complexity values are linked to regional invention activities to assess a region's capability to produce complex knowledge. The findings of the empirical analysis suggest that European regions substantially differ in their ability to produce complex technologies. Although complex knowledge concentrates in some large urban areas such as Paris, Madrid and Munich, the results indicate that it is not exclusively an urban phenomenon. In addition, it is shown that regional variances in knowledge complexity are linked to economic growth in regions. More precisely, a 10 percent increase in complexity is associated with a corresponding increase in regional economic growth of about 0.28 percent.

1.2.3 Degree of novelty

It was in Mainz sometime in 1454 or 1455 when Johannes Gutenberg printed the last page of the bible, later known as the Gutenberg Bible. This moment might represent the start of modern book printing and revolutionized the production and diffusion of books fundamentally. Before the Gutenberg revolution, the production of written material was time and cost-intensive because it relied on manual reproduction, limiting production and diffusion of books to a small share of people such as monks and nuns. Gutenberg's development of the printing press, however, fundamentally reshaped the art of printing and represents a turning point in human history. Clearly, he relied on existing technologies, for example presses used in wine production, modifying them in a revolutionary way and putting them into a new context. His printing press reduced the time and cost for book reproduction dramatically and thus allowed a faster dissemination of books to a larger share of the population (Rees 2005).

Gaining a better understanding of novelty and its geographic patterns is therefore taken up in Chapter 4. It builds on the theoretical as well as empirical insights of scaling analyses, which show that the geographic concentration of knowledge production is not random, but in favor of larger cities (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007a; Bettencourt et al. 2007b). These authors estimate a scaling model with the reported coefficients, however only reflecting the productivity of cities in pure quantitative terms with respect to innovation without considering different degrees of novelty in technological knowledge. Cities as hotspots of innovation, however, also concentrate essential functionalities and fundamental resources that might influence novelty creation in qualitative terms. In particular, Jacobs externalities state that innovation activities benefit from the diversity in larger cities as it facilitates cross-fertilization of unrelated knowledge rather than being locked in industry-internal thought patterns. Empirical evidence supports this argument and even goes a step further. Atypical combinations, which bridge unrelated knowledge areas, have the potential to create radically new outcomes, whereas typical combinations, in contrast, continue to link related knowledge leading to rather incremental novelties (Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016).

Chapter 4 examines this research gap and asks: *Are cities hotspots of truly novel ideas?* Historical patent documents covering the years 1836 to 2010 serve as the empirical basis to investigate the research question (Petralia et al. 2016). A time span of 174 years allows the unraveling of long-term trends in technological development and acknowledges the path dependency in knowledge creation that leads to persistent geographic patterns. The empirical approach relies on the theoretical conceptualization of inventions as the result of knowledge combinations. Deconstructing inventions into their combinations allows the application of z-scores to distinguish between atypical (i.e. truly novel) and typical (i.e. incremental) combinations (Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016). The scaling analysis in Chapter 4 first reveals that novelty has increasingly concentrated in large metropolitan areas in the last 174 years of US invention history. Hence, cities' productivity in knowledge production not only refers to pure quantity, but also includes novelty. This observation is connected to the second finding in this exercise that technological diversity in cities increases linearly with their size. Inventors in larger cities have more local opportunities to explore new combinations than inventors in smaller towns.

1.2.4 Impact

New knowledge varies greatly according to its impact on subsequent innovation (Trajtenberg 1990). Researchers assess impact on subsequent knowledge creation processes, for instance by using citation data in scientific publications or patents (Garfield 1970; Trajtenberg 1990). Citation data reveals how often a certain scientific publication or patent is used as an input for subsequent ones and therefore allows a distinction between highly impactful and less impactful knowledge outcomes. Empirical research indicates that highly cited outcomes, e.g. impactful innovations, are also likely to generate economic value(Harhoff et al. 1999; Hall et al. 2005).

A famous example in this context is the following: in 1998, the journal Computer Networks and ISDN Systems published the paper *"The anatomy of a large-scale hypertextual Web search engine"*. In the paper, the authors first present the prototype of a large-scale algorithm that searches the web. Their algorithm uses the link structure of web pages to identify important pages and to create a quality ranking. The authors were Sergey Brin and Lawrence Page (1998) (1998). They called their algorithm PageRank and their prototype of a search engine Google. Shortly after the publication, Brin and Page founded Google Inc. located in the garage of their friend Susan Wojcicki in Menlo Park, California (Vise and Malseed 2005). Now, Google employs around 20,000 people in Mountain View, 10 kilometers from its birth location in Menlo Park. Just as other technological breakthroughs deeply shaped the economic structure in San Francisco's Bay Area long before, Google has done so again in the last twenty years (Storper et al. 2015). Thus far, Brin and Page's paper has received nearly 4,000 citations in the Web of Science (July 2019). The fact that approximately 0.03% of all publications in the Web of Science collection receive more than 1,000 citations illustrates the impact of this paper[1] (Van Noorden et al. 2014).

Although impactful innovation can have substantial regional consequences, as illustrated by the example of Google, variations in the impact of innovations are rarely considered in economic geography research. However, it should be important whether a region produces large amounts of knowledge with marginal impact or if it produces impactful outcomes that substantially affect regional development by offering new growth paths. Among the few studies in economic geography that consider technological impact in their empirical analyses, the works of Ejermo (2009) as well as Castaldi and Los (2017) highlight that impactful innovations concentrate more strongly in space than conventional innovations. The strong concentration of impactful innovations, however, raises more questions. Why are some regions more capable of producing impactful innovations than others or, asked differently, what explains the geographic concentration of impactful innovations? Castaldi et al. (2015) find that an economic structure characterized by unrelated variety is linked to the emergence of high-impact innovations in regions. Beyond this, however, research in economic geography is (still) astonishingly silent about the underlying reasons behind regions' ability to produce impactful outcomes.

Chapter 5 approaches this research gap and presents the prevailing regional culture regarding openness to innovation as one possible explanation for the observed regional variation in creating impactful innovations. In particular, a regional social climate that cultivates new ideas and values creative thinking is argued to play a crucial role for the emergence of impactful innovations. Chapter 5 thereby builds on existing theoretical frameworks in economic geography that – sometimes explicitly and sometimes implicitly – discuss the role of regional openness for innovation without considering its role for the creation of impactful innovation (Saxenian 1994; Rodríguez-Pose 1999; Florida 2002). However, existing works rely on rather crude and indirect proxies to measure regional openness, such as the share of homosexuals (Florida 2003). To overcome these shortcomings, Chapter 5 builds on recent insights in the field of *geographical psychology*. Psychologists have made substantial advances in conceptualizing and measuring individuals' personality via personality traits. The so-called Big Five summarize five personality traits (agreeableness, extraversion, conscientiousness, neuroticism and openness) (John and Srivastava 1999). The trait openness, in particular, is associated with innovation, as it describes individuals' inventiveness, creativity, originality and curiosity (McCrae 1987; King et al. 1996; McCrae 1996; John and Srivastava 1999). Chapter 5 follows such a macro-psychological approach and asks: *Does regional openness influence the impactfulness of regional innovation activities?*

The empirical approach relies on patent data from the USPTO as a proxy for innovation activities in 382 metropolitan statistical areas (MSAs) in the US between 2000 and 2010. Following previous approaches, innovations' impact is assessed using class and cohort corrected patent citation counts as a reliable measure of impact (Trajtenberg 1990; Hall et al. 2005). Specifically, regional innovation activities are ranked according to their received number of citations, which allows a calculation

[1] The corresponding USPTO patent US6285999 received 923 citations (July 2019), which is also considered a high number. Citation frequencies between scholarly publications and patents are not comparable.

of the impact of regional innovation activities in different percentiles of the regional impact distribution. To capture regional personality differences in openness, the responses of approximately 1.26 million participants from the internet personality project (Gosling et al. 2004) are aggregated to the MSA level. The empirical results support the substantial differences between highly impactful innovations and the average innovation quality (Castaldi et al. 2015). Regional openness is associated with a stronger output of highly impactful innovation in regions, but not with the average innovation quality. Macro-psychological openness is tested against indirect indicators of openness applied in previous research, such as Florida's *"Gay Index"*. The obtained results are robust, as macro-psychological openness either outperforms indirect measures of open-mindedness or explains an additional share of the regional variation of highly impactful innovations alongside open-mindedness.

1.3 Outline of the thesis

In summary, scientific discoveries and technological milestones are important characteristics of human development. The curiosity and passion of bright minds such as Gutenberg or the Wright brothers and the considerable efforts made by humans to accomplish outstanding achievements (e.g. moon landing) highlights the fascination for technological development and the importance of new knowledge for our society. The importance of new knowledge for economic growth paired with the spatial concentration of innovation activities has motivated previous research and has put the topic high on the research agenda in economic geography.

Many concepts in economic geography highlight the heterogeneity of places and knowledge. The focus in previous research, however, has been more on knowledge quantity and less on differences in knowledge quality across places. This dissertation therefore emphasizes that quality is an important feature of new knowledge and innovation varying substantially between places. Assessing knowledge quality in places has the potential to improve our understanding of knowledge production and regional development. To acknowledge that differences in knowledge quality manifest themselves along multiple dimensions, this thesis assesses quality by relying on relatedness, complexity, degree of novelty and impact as important quality characteristics. Each quality dimension represents the core of one chapter.

Hence, this dissertation begins by asking if policy can intervene in the path-dependent process of regional diversification by focusing on the role of public subsidies for R&D projects that may influence (related) diversification of regions (Chapter 2). Subsequently, Chapter 3 places knowledge complexity at the center of empirical analysis and studies the role of complexity for regional economic development. Chapter 4 unravels US invention history by quantifying novelty in knowledge combinations and studies their relation to city size. Finally, Chapter 5 shifts the attention to impactful innovations regions. It presents differences in regions' openness to innovation as a hidden cultural trait to explain regional variations in impactful innovations. The concluding Chapter 6 summarizes the contribution of all empirical chapters, discusses limitations of this research and presents future research opportunities.

2 Subsidized to Change? Analyzing the Impact of R&D Policy on Regional Diversification

Abstract: Previous research shows ample evidence that regional diversification is strongly path-dependent, as regions are more likely to diversify into related than unrelated activities. Although related diversification strengthens regions' existing capabilities, it also can lead to cognitive lock-ins. In this paper, we ask whether innovation policy in terms of R&D subsidies can intervene in regional diversification. Can R&D subsidies even break the path-dependency by facilitating unrelated diversification? We answer this question by linking information on R&D subsidies with patent data and analyzing the technological diversification of 141 German labor-market regions between 1991 and 2010. Our findings suggest that R&D subsidies positively influence regional diversification. In addition, we find significant differences between types of subsidy. Subsidized joint R&D projects have a larger effect on entry probabilities than subsidized R&D projects conducted by single organizations. To some extent, collaborative R&D can even compensate for missing relatedness by facilitating diversification into unrelated activities.

This chapter is co-authored with Tom Broekel. The PhD candidate is the first author of the article. The paper has been revised and resubmitted to The Annals of Regional Science.

2.1 Introduction

R&D subsidies are a crucial tool of modern (regional) innovation policy. Prominently, the EU-Framework Programmes are based on direct, project-based R&D subsidization, and so are many national as well as regional initiatives (Dohse 2000; Breschi and Cusmano 2004; Defazio et al. 2009; Barajas et al. 2012; Broekel 2015). Existing investigations of such programs' effects tend to focus on their contribution to innovation processes whereby particular attention is being paid to the quantity of additional R&D efforts invested and alternatively the numbers of innovations being created (Czarnitzki et al. 2007; Zúñiga-Vicente et al. 2014).

The present paper extends the analysis of R&D support programs by investigating their impact on regional (technological) diversification. For regions, diversifying into new activities has been shown to be key for their long-term economic development (Frenken et al. 2007; Pinheiro et al. 2018). In recognition of this, the European Union supports diversification with its Smart Specialization strategy implemented in its current cohesion policy (McCann and Ortega-Argiles 2013; Foray et al. 2011).

We argue that R&D support programs that are not intended to contribute to diversification processes may nevertheless be effective in this direction. Most importantly, this is because many of these programs facilitate R&D and stimulate inter-organizational collaboration. These two activities are at the heart of (technological) diversification, and organizations are highly interested in obtaining public support for both. Moreover, the current allocation of R&D subsidization has the potential of (unintentionally) steering parts of the support toward regions and technologies with larger potentials of successful diversification. The evaluation of the potentials utilizes the idea of *related diversification* (Boschma and Frenken 2011; Boschma et al. 2017), according to which diversification processes are more likely to occur and be successful when the new activity is related to existing regional competences.

Our paper thereby fills a gap in the existing literature, as, so far, few efforts have been made to assess systematically the contribution of R&D policy to regional diversification (Boschma and Gianelle 2014). Moreover, most investigations of R&D support are restricted to the firm level (Czarnitzki et al. 2007; Czarnitzki and Lopes-Bento 2013), while attention has only recently been drawn to the regional level (Maggioni et al. 2014; Broekel 2015; Broekel et al. 2017).

We support our theoretical arguments with an empirical investigation on the contribution of project-based R&D subsidization by the Federal Government of Germany to regional technological diversification processes. Firstly, we explore the extent to which the allocation of R&D subsidies facilitates regional diversification. Secondly, we test if these R&D subsidies increase the chances of successful diversification in general and if they are rather conducive for regions gaining competences in related or unrelated technological fields. Thirdly, we differentiate between subsidies for individual and for joint projects.

Our empirical study builds on a panel regression approach utilizing data on 141 German labor-market regions covering the period from 1991 to 2010. Patent information is used as an indicator for technology-oriented R&D activities in regions and matched with subsidized R&D projects. Our empirical results confirm the path-dependent nature of regional technological diversification, which is driven by technological relatedness. In addition, R&D subsidies are more likely allocated to related capabilities in regions, indicating the tendency of policy to be part of the path dependency in regional diversification. Our study confirms that R&D subsidies stimulate technological diversification in regions. The identified positive effects are particularly pronounced and robust in the case of subsidized joint R&D projects. Lastly, we find that R&D subsidies are an appropriate policy to help regions broadening their technological portfolio by (partially) compensating for lacking relatedness.

The remainder of the study is organized as followss Section 2.2 provides an overview of the existing literature on regional diversification and R&D policy. We describe our data and empirical approach in Section 2.3. The empirical results are part of Section 2.4. The paper concludes with a discussion of our results in Section 2.5.

2.2 R&D subsidies and regional diversification

2.2.1 R&D subsidies and diversification

R&D policy programs are justified by knowledge creation and innovation being important production factors for economic growth. Nevertheless, knowledge creation suffers from significant market failures (Nelson 1959; Arrow 1962; McCann and Ortega-Argiles 2013). For instance, firms cannot fully benefit from their R&D investments, as new knowledge might lack appropriability and spills over to third parties, giving rise to positive externalities. Similarly, R&D projects are characterized by significant uncertainty making, ex-ante calculations of investments into R&D a difficult task. Increasing complexity of technologies also requires efforts exceeding individual firms' capabilities. Accordingly, collaboration with other organizations becomes a necessity, which raises the danger of moral hazard and unintended knowledge spillover (Hagedoorn 2002; Cassiman and Veugelers 2002; Broekel 2015). In sum, private R&D investments are likely to fall short of a social optimum. This motivates and justifies public intervention, which seeks to close the gap between actual and socially desired levels of knowledge creation by supporting R&D activities.

There are numerous instruments policy may use to achieve this goal. Among the most prominent and frequently used are project-based R&D subsidies (Aschhoff 2008). These are intended to increase R&D activities of organizations regarding innovation input and output. Concerning the input, one major question is whether firms use public subsidies as a complementary and additional financial source to realize R&D projects or if they "crowd out" private investments. The large body of empirical research finds mixed results. Although a general crowding-out effect cannot be ruled out and depends largely on firm characteristics, the majority of studies find evidence for additionality effects (Busom 2000; Czarnitzki and Hussinger 2004; Zúñiga-Vicente et al. 2014). Regarding innovation output, public subsidies seem to stimulate R&D activities. A number of studies show the positive effect of R&D subsidies on firms' innovativeness (Czarnitzki et al. 2007; Czarnitzki and Hussinger 2018; Ebersberger and Lehtoranta 2008). That is, significant parts of private R&D activities would not have been realized without subsidization, implying that public subsidies seem to complement private R&D.

Yet the design of R&D subsidization programs offers a lot of flexibility, which allows for substantial "fine-tuning" of initiatives. For instance, subsidization can be restricted to specific organizations (location, size, industry), to selected fields (technologies, sectors), or to particular modes of R&D (individual or joint). Policy can also decide about starting dates and time periods of support. Usually, R&D subsidies are granted through competitive bidding procedures (Aschhoff 2008), and they are targeted at innovative self-discovery processes (Hausmann and Rodrik 2003) with the stimulation of inter-organizational knowledge exchange becoming an increasingly important feature (Broekel and Graf 2012).

All of these features are used in contemporary policies to varying degrees. For instance, the EU-Framework programmes (EU-FRP) are focused on supporting R&D and on stimulating interregional as well as international knowledge diffusion by exclusively supporting collaborative projects (Scherngell and Barber 2009; Maggioni et al. 2014). Another example of R&D subsidization with specific features is the German BioRegio contest. This initiative focused on advancing one particular technology (biotechnology) and rewarded proposals building on and stimulating intra-regional collaboration (Dohse 2000).

While most empirical studies have examined at the effects of R&D subsidies at the firm level evaluating their allocation and impact, we seek to extend this perspective in this study. More precisely, we argue that project-based R&D subsidization may play a role in regional diversification processes. Interestingly, linking policy to regional diversification has rarely been done in the literature. An exception concerns the case study by Coenen et al. (2015) that investigates opportunities, barriers, and limits of regional innovation policy aiming at the renewal of mature industries. The authors show, for the case of the forest industry in North Sweden, that regional innovation policy can accompany the process of regional diversification by supporting the adoption and creation of related technologies. Our study complements this approach by focusing on a particular policy, namely, R&D subsidies and their effects on regional diversification.

2.2.2 Regional diversification and relatedness

Regional diversification is in the focus of contemporary innovation policy. For instance, the EU's Smart Specialization strategy aims at fostering (technological) diversification around regions' core activities (Foray et al. 2011). Thereby, policy seeks to exploit the benefits associated with diversification. For instance, diversification positively relates to the level of income, allowing regions to climb the ladder of economic development (Imbs and Wacziarg 2003). Diversified regions are, moreover, less likely to run into the trap of cognitive lock-ins (Grabher 1993) and are less prone to suffer from exogenous shocks because of portfolio effects (Frenken et al. 2007). Regional R&D competences in multiple fields also give rise to synergies increasing the exploitation and experimentation of technological opportunities (Foray et al. 2011).

A large stream of literature increasingly devotes its research to the path-dependent feature of regional diversification expressed by the crucial role of relatedness (Hidalgo et al. 2007; Boschma and Frenken 2011; Neffke et al. 2011; Hidalgo et al. 2018). Concepts such as *related diversification* and *regional branching* (Boschma and Frenken 2011) highlight that regional diversification is not a random process but that existing capabilities influence the development of future capabilities. The so-called *"principle of relatedness"* (Hidalgo et al. 2018) is not only working at the individual level of firms (Teece et al. 1994; Breschi et al. 2003) but shows its importance at different spatial scales. For example, Hidalgo et al. (2007) show that nations are more likely to diversify into new export products that are related to their existing product portfolio. Neffke et al. (2011) transferred this approach to the regional level. By relying on information about products of Swedish manufacturing firms, they show that new industries do not emerge randomly across space. Rather, they are more likely to emerge in regions where related capabilities already exist. Essletzbichler (2015) confirms this finding for industrial diversification in US metropolitan areas. Similar results are obtained by Boschma et al. (2013) for the export profile of Spanish regions. By comparing the impact of relatedness for different spatial levels, the authors also show related industries to play a more crucial role at the regional compared to the national level. (Rigby 2015) and (Boschma et al. 2015) analyze regional diversification in US metropolitan areas. Both find that technology entries are positively, and exits are negatively, correlated with their relatedness to regions' technology portfolios.

The ample empirical evidence for related diversification being the norm rather than the exception reveals the dominant role of path dependency in diversification processes. By building on related capabilities, economic actors follow existing technological trajectories, rely on established routines, and build on familiar knowledge (Nelson and Winter 1982; Dosi 1988). Building on existing capabilities rather than exploring completely new ones reduces uncertainties and risks while increasing the likelihood of successful diversification.

The path dependency in regional diversification certainly has substantial advantages. For instance, regions can specialize and build competitive advantages in certain activities providing them with important growth opportunities (Martin and Sunley 2006; Boschma and Frenken 2006). The continuous specialization of the Silicon Valley into information and communication technologies is a prominent example of successful related diversification along a promising path (Storper et al. 2015). Nevertheless, related diversification can also lead to regional lock-ins by following mature paths with little future prospects, such as in the German Ruhr-Area (Grabher 1993). Diversification into unrelated activities can prevent such lock-ins by broadening the set of regional capabilities. In addition, it increases regional resilience toward external shocks (Frenken et al. 2007). Yet unrelated diversification requires the exploration of new knowledge, which is uncertain, risky, and less promising.

2.2.3 R&D subsidies and regional diversification

Can project-based R&D subsidies impact regional diversification? If so, how? Firstly, diversification requires organizations to leave existing routines by exploring new activities involving novel (at least to the organization) knowledge and technologies. It further implies less foresight on potential outcomes and lower abilities to plan R&D processes as well as commercialization possibilities. Existing routines are less helpful in designing financial plans, selecting appropriate suppliers, or buying needed equipment. Consequently, diversification-oriented R&D can be expected to represent a risky and uncertain undertaking. Economic actors therefore show a tendency to avoid diversification into completely new activities. As subsidization is more strongly utilized for riskier projects (Fier et al. 2006), we argue that organizations are highly likely to make use of subsidization for (risky) diversification activities.

Secondly, the effects of project-based R&D subsidies unfold beyond the individual firm (Broekel 2015; Maggioni et al. 2014). Organizations are embedded into regional economies through labor mobility, collaboration, social networks, input-output linkages, and other types of interactions. This is highlighted in various approaches, including regional innovation systems, learning regions, and clusters (Cooke 1998; Florida 1995; Porter 2000). Accordingly, knowledge and competences that are acquired in subsidized projects are more likely to be picked up and utilized by other regional actors. In this sense, R&D subsidies present a resource inflow into the region's innovation system, which supports its general innovation activities, including those oriented toward diversification.

Thirdly, regional diversification frequently takes place through spin-off and start-up processes (Boschma and Wenting 2007; Boschma and Frenken 2011). At the same time, spin-offs in particular have been identified as frequent and above-average recipients of R&D subsidies (Cantner and Kösters 2012). The added value of the support thereby exceeds what has been discussed above. (Fier et al. 2006) identify subsidies to support university spin-outs by adding credibility and strengthening public relations. Under the assumption that there is no discrimination against spin-offs active in technologies new to a region, R&D subsidies thereby directly support regional diversification.

Fourthly, many R&D subsidization initiatives seek to advance particular technologies (e.g., biotechnology). Announcing such initiatives signals to economic actors that these technologies are (at least in the eyes of policymakers) promising and may offer economic potential. If effective, this is likely to stimulate actors to expand already-existing activities in these technologies or diversify into these activities.

In sum, R&D subsidies encourage riskier research, expand R&D resources, and exert particular benefits for spin-offs as well as spin-outs. In turn, all these contribute to regional diversification. Notably, the discussed effects are largely independent of the policy being designed to support diversification. Naturally, such diversification-enhancing effects are amplified when R&D subsidization policies aim to support diversification, as was the case in the BioRegio contest (Dohse 2000).

On this basis, we further argue that subsidies do not equally impact all diversification processes. We particularly expect them to matter more for regions diversifying along existing technological trajectories (related diversification). The primary reason for this is that the subsidies are more likely to be received by projects building on existing regional competences. Innovation policy does not allocate R&D subsidies randomly. Applications need to pass a review process, which usually aims at selecting those with the highest chances of being successful (Aubert et al. 2011). This applies to applications with applicants' competences meeting those necessary for the successful completion of projects. In addition, organizations usually require technological expertise, prior experiences, infrastructure, and matching qualifications to write convincing applications. Such is more likely given when organizations are active in similar or related activities (Blanes and Busom 2004; Aschhoff 2008).

This is not restricted to the organizational level. For instance, (Broekel et al. 2015b) show that even when controlling for organizational characteristics, being located in a regional cluster (of related activities) increases the chances of receiving R&D subsidies (at least in the case of EU-FRP). One of the reasons for this is that organizations located within clusters "*are more likely to learn about subsidization programs, which is probable to translate into higher application rates*" (Broekel et al. 2015b, p. 1433). It seems reasonable to assume that this especially applies to policy initiatives related to activities of the organizations within the cluster. Consequently, we expect that R&D policy plays a role in the path dependency in regional diversification by preferentially allocating public resources to related, rather than to unrelated, capabilities in regions. Our first hypothesis reads as follows:

H1a: Project-based subsidization of R&D positively influences technological diversification in regions.

H1b: Project-based subsidization of R&D is more likely to to contribute to related diversification.

While these arguments refer to R&D subsidies in general, we argue that the influence of R&D policy depends on its specific mode. Previous research has shown that the effects of R&D subsidization differ between subsidies granted to individual- and joint-research projects (Broekel and Graf 2012; Broekel 2015). In contrast to subsidies for individual projects, supporting joint R&D projects has a greater potential for stimulating the exploration of new knowledge and activities, as these require organizations to collaborate. Consequently, such support is likely to alter organizations' and regions' embeddedness into intra-regional and inter-regional knowledge networks (Fier et al. 2006; Wanzenböck et al. 2013; Broekel 2015; Töpfer et al. 2017). For instance, (Broekel et al. 2017) measure the technological similarity of partners in subsidized projects and find these to be rather heterogeneous. Firms are also shown to particularly add science organizations to their portfolio of collaboration partners when participating in subsidized R&D projects (Fier et al. 2006).

The utilization of subsidies to explore new knowledge is further highlighted by the location of collaboration partners. In Germany, only 12% of collaborations established by joint projects subsidized by the federal government connect partners within the same region (Broekel and Mueller 2018). In the case of the EU-FRP for biotechnology, this figure is as small as one percent (Broekel et al. 2015b). Accordingly, project-based subsidies are frequently employed to establish or strengthen relations with dissimilar actors from different regions, which is crucial and typical for diversification activities (Hagedoorn 1993; Boschma and Frenken 2011; Oort et al. 2015). We therefore expect subsidies for joint (collaborative) research to have stronger effects than individual grants, due to their impact on collaboration and knowledge networks. As collaborative R&D subsidies facilitate knowledge exchange between new and heterogeneous actors, we particularly expect joint-research projects to increase the likelihood of unrelated diversification in regions. Our second hypothesis summarizes these arguments as follows:

H2a: Subsidized joint R&D projects contribute to a larger extent to technological diversification in regions than do individual R&D projects.

H2b: Subsidized joint R&D projects facilitate regional diversification into unrelated activities.

2.3 Data and methods

2.3.1 Measuring regional diversification

To study the relationship between R&D subsidies and regional diversification, we focus on 141 German labor-market regions (LMR), as defined by Kosfeld and Werner (2012). Moreover, our data cover the years from 1991 to 2010. In a common manner, we use patent data to approximate technological activities (Boschma et al. 2015; Rigby 2015; Balland et al. 2019). Despite well-discussed drawbacks (Griliches 1990; Cohen et al. 2000), patents entail detailed information about the invention process, such as the date, location, and technology, all of which are fundamental for our empirical analysis. We extract patent information from the *OECD REGPAT Database*, which covers patent applications at the *European Patent Office* (EPO). Based on inventors' residences, we assign patents to the corresponding LMR. For smaller regions in particular, annual patent counts are known to fluctuate, strongly challenging robust estimations. We therefore aggregate our data into four 5-year periods (1991-1995, 1996-2000, 2001-2005, 2006-2010).

Technologies are classified according to the *International Patent Classification* (IPC). The IPC summarizes hierarchically eight classes at the highest and more than 71,000 classes at the lowest level. We aggregate the data to the four-digit IPC level, which differentiates between 630 distinct technology classes. The four-digit level represents the best trade-off between a maximum number of technologies and sufficiently large patent counts in each of these classes.

Previous studies relied on the location quotient (LQ), also called reveal technological advantage (RTA), to identify diversification processes. For example, LQ values larger than one signal the existence of technological competences in a region, and values below signal their absence. Successful diversification is then identified when the LQ grows from below one to above one between two periods (Boschma et al. 2015; Rigby 2015; Cortinovis et al. 2017; Balland et al. 2019). We refrain from this approach for two important reasons. Firstly, being a relative measure, the LQ approach allows technologies to "artificially" emerge in regions simply by decreasing patent numbers in other regions. Secondly, the LQ is normalized at the regional and technology levels, which can interfere with the inclusion of regional and technology fixed effects in panel regressions.

We therefore rely on an alternative and more direct approach to assess diversification processes by concentrating on absolute changes in regional patent numbers. More precisely, we create the binary dependent variable *Entry* with a value of 1 if we do not observe any patents in technology k in region r and period t, and a positive value in the subsequent period $t + 1$. We intensively checked the data for random fluctuations between subsequent periods, which can inflate the number of observed entries. The aggregation of regional patent information into 5-year periods, however, eliminated such cases almost completely.

2.3.2 Information on R&D subsidies

Our main explanatory variable, *Subsidies*, represents the sum of R&D projects in technology class k and region r at time t. The so-called *Foerderkatalog* of the German Federal Ministry of Education and Research (BMBF) serves as our data source. The BMBF data cover the largest parts of project-based R&D support at the national level in Germany (Czarnitzki et al. 2007; Broekel and Graf 2012) and have been used in a number of previous studies (Broekel and Graf 2012; Broekel et al. 2015a; Broekel et al. 2015b; Cantner and Kösters 2012; Fornahl et al. 2011). The data provide detailed information on granted individual and joint R&D projects, such as the starting and ending dates, the location of the executing organization, and a technological classification called *Leistungsplansystematik* (LPS).

The LPS is a classification scheme developed by the BMBF and consists of 47 main classes. The main classes are, similarly to the IPC, disaggregated into more fine-grained subclasses, which comprise 1,395 unique classes at the most detailed level. To create the variable *Subsidies*, we need to match the information on R&D subsidies with the patent data. Both are based on different classification schemes (IPC and LPS), which prevents a direct matching. Moreover, there is no existing concordance of the two classifications.

We therefore develop such a concordance. For this, we reduce the information contained in the Foerderkatalog by excluding classes that are irrelevant for patent-based innovation activities. This primarily refers to subsidies in the fields of social sciences, general support for higher education, gender support, and labor conditions. Next, we utilize a matched-patent-subsidies-firm database created by the Halle Institute of Economic Research. This database includes 325,497 patent applications by 5,398 German applicants between 1999 and 2017. It also contains information on 64,156 grants of the Foerderkatalog with 10,624 uniquely identified beneficiaries. In this case, beneficiaries represent so-called executive units (*"Ausführende Stelle"*) (see Broekel and Graf 2012).

In this database, grant beneficiaries and patent applicants are linked by name-matching. Hence, the IPC classes of their patents can be linked to the LPS classes of their grants. In principle, this information allows for a matching of the most fine-grained level of the IPC and LPS. In this case, however, the majority of links are established by a single incidence of IPC classes coinciding with LPS classes, i.e., there is only one organization with a patent in IPC class k and a grant in LPS class l. Moreover, the concordance is characterized by an excessive number of zeros, as only few matches of the $71,000$ (IPC) $*$ $1,395$ (LPS) cases are realized.

To render the concordance more robust, we therefore establish the link on a more aggregated level, which also makes the concordance correspond to the data employed in this study. More precisely, we aggregate the IPC classes to the four-digit level and the LPC to the 47 *main classes* defined in (BMBF 2014). It is important to note that, not all LPS main classes are relevant for patent-based innovation (e.g., arts and humanities). We eliminate such classes and eventually obtain 30 LPS main classes that are matched to 617 out of 630 empirically observed IPC classes. For these, we calculate the share of organizations $S_{l,k}$ with grants in LPS l that also patent in IPC k:

$$S_{l,k} = \frac{n_{l,k}}{\sum_{x=1}^{X_l} n_x} \tag{2.1}$$

with $n_{l,k}$ being the number of organizations with at least one patent in k and grant in l. X_l is the total number of organizations with grants in l. On this basis, we calculate the number of subsidized projects, $Subsidies_{l,k}$, assigned to region r and technology k by multiplying the number of grants in l acquired by regional organizations with patents in k with $S_{l,k}$.

Following the discussion in Section 2.2, we calculate *Subsidies* in three versions: on the basis of all subsidized projects (*Subsidies*), for individual projects ($Subsidies^{\text{Single}}$), and considering only joint projects ($Subsidies^{\text{Joint}}$).

2.3.3 Relatedness density

Our second most important explanatory variable is relatedness. We follow the literature in constructing this variable as a density measure (Hidalgo et al. 2007; Rigby 2015; Boschma et al. 2015). More precisely, relatedness density reveals how well technologies fit to the regional technology landscape. It is constructed in two steps.

Firstly, we measure technological relatedness between each pair of technologies. The literature suggests four major approaches: (i) entropy-based (Frenken et al. 2007), (ii) input-output linkages (Essletzbichler 2015), (iii) spatial co-occurrence (Hidalgo et al. 2007), and (iv) co-classification (Engelsman and Raan 1994). We follow the fourth approach and calculate technological relatedness between two technologies (four-digit patent classes) based on their co-classification pattern (co-occurrence of patent classes on patents). The cosine similarity gives us a measure of technological relatedness between each technology pair (Breschi et al. 2003).

Secondly, we determine which technologies belong to regions' technology portfolios at a given time. Straightforwardly, we use patent counts with positive numbers indicating the presence of a technology in a region. Following Hidalgo et al. (2007), we measure relatedness density on this basis as:

$$Density_{k,r} = \frac{\sum_m x_m \, \rho_{k,m}}{\sum_m \rho_{k,m}} * 100 \tag{2.2}$$

where *Density* stands for relatedness density. ρ indicates the technological relatedness between technology k and m, while x_m is equal to 1 if technology m is part of the regional portfolio ($Patents > 0$) and 0 otherwise ($Patents = 0$). Consequently, we obtain a 141 x 630 matrix including the relatedness density for each of the 630 IPC classes in all 141 LMRs indicating their respective relatedness to the existing technology portfolio of regions.

2.3.4 Control variables

In addition to R&D subsidies and relatedness density, the empirical literature has identified a number of other determinants of regional technological diversification. Firstly, knowledge spillover from adjacent regions can potentially impact regional diversification processes (Boschma et al. 2013). We account for these potential spatial spillovers and include technological activities in neighboring regions (*Neighbor Patents*) as a spatially lagged variable. The variable counts the number of patents in technology k of all neighboring regions s of region r. Regions s and r are neighbors if they share a common border.

Hidalgo et al. (2007) demonstrate that diverse regions with larger sets of capabilities have more opportunities to move into new fields than regions with narrow sets. The regional diversity (*Diversity*) variable detects this. It is defined as the number of technologies with positive patent counts in a region.

We also consider the number of regional patents ($Regional\ Patents_i$) to control for the size of the regional patent stock. Lastly, the size of technologies is controlled for by considering the number of patents in a given technology ($Technology\ Size_k$). Descriptive statistics and correlations for all variables are reported in Table 2.1.

Table 2.1: Summary statistics and correlation matrix

	Variables	Min	Max	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Entry	0	1	0.13	0.33								
2	Subsidies	0	7.89	0.08	0.21	0.31							
3	Subsidies$^{\text{Single}}$	0	7.03	0.05	0.16	0.28	0.97						
4	Subsidies$^{\text{Joint}}$	0	2.30	0.03	0.07	0.29	0.86	0.72					
5	Density	0	100	8.55	16.14	0.15	0.13	0.06	0.27				
6	Neighbor Patents	0	419.66	2.64	8.96	0.19	0.03	0.01	0.09	0.14			
7	Regional Patents	0	13,144.90	451.05	952.56	0.11	0.12	0.03	0.28	0.56	0.05		
8	Diversity	0	469	77.04	86.10	0.15	0.14	0.05	0.31	0.64	0.11	0.83	
9	Technology Size	0	5,829.40	93.24	229.64	0.15	0.07	0.03	0.15	0.07	0.54	-0.02	0.02

All correlations are significant with $p < 0.001$

2.3.5 Empirical model

We follow an established approach in the literature on regional diversification to set up our empirical model (Boschma et al. 2015; Balland et al. 2019). More precisely, we rely on panel regressions to explain the status of technological diversification in a region. Our basic model is specified as follows:

$$Entry_{k,r,t} = \beta_1 Subsidies_{k,r,t-1} + \beta_2 Density_{k,r,t-1} + X_{k,r,t-1} + \tau_k + \pi_r + \omega_t + \epsilon_{k,r,t} \quad (2.3)$$

$Entry$ indicates the status of diversification into technology k of region r at time t. Accordingly, all estimations are based at the region-technology level. $Subsidies$ summarizes the number of subsidized R&D projects. In alternative models, it is replaced with the number of individual ($Subsidies^{\text{Single}}$) and joint projects ($Subsidies^{\text{Joint}}$). $Density$ is the relatedness density, and X is the vector of control variables. All estimations include technology (τ), region (π), and time (ω) fixed effects capturing time-invariant, unobserved, heterogeneity. We assume a time delay with which our dependent variable responds to variation in the explanatory variables. R&D subsidies, for example, are unlikely to cause immediate effects visible in innovation activities as approximated by patents. Rather, they unfold their influence in subsequent years. Consequently, we lag the explanatory variables by one time period, which corresponds to 5 years.

As $Entry$ is a binary variable, a logit regression is applicable. Nevertheless, logit regressions with many fixed effects and few time periods can lead to the prominent incidental parameters problem causing biased results (Neyman and Scott 1948). Therefore, we rather rely on a linear probability model (LPM) to assess the probability that technology k emerges in region r. We nevertheless, report the results of the three-way fixed effects logit regression in our robustness checks (see table 2.7 in Appendix section 2.A). An entry model implies restricting the observations to those cases in which an entry is possible. Accordingly, we reduce the sample to all potential cases of entry, which corresponds to technology k being absent from the regional technology portfolio in $t-1$ (zero patents).

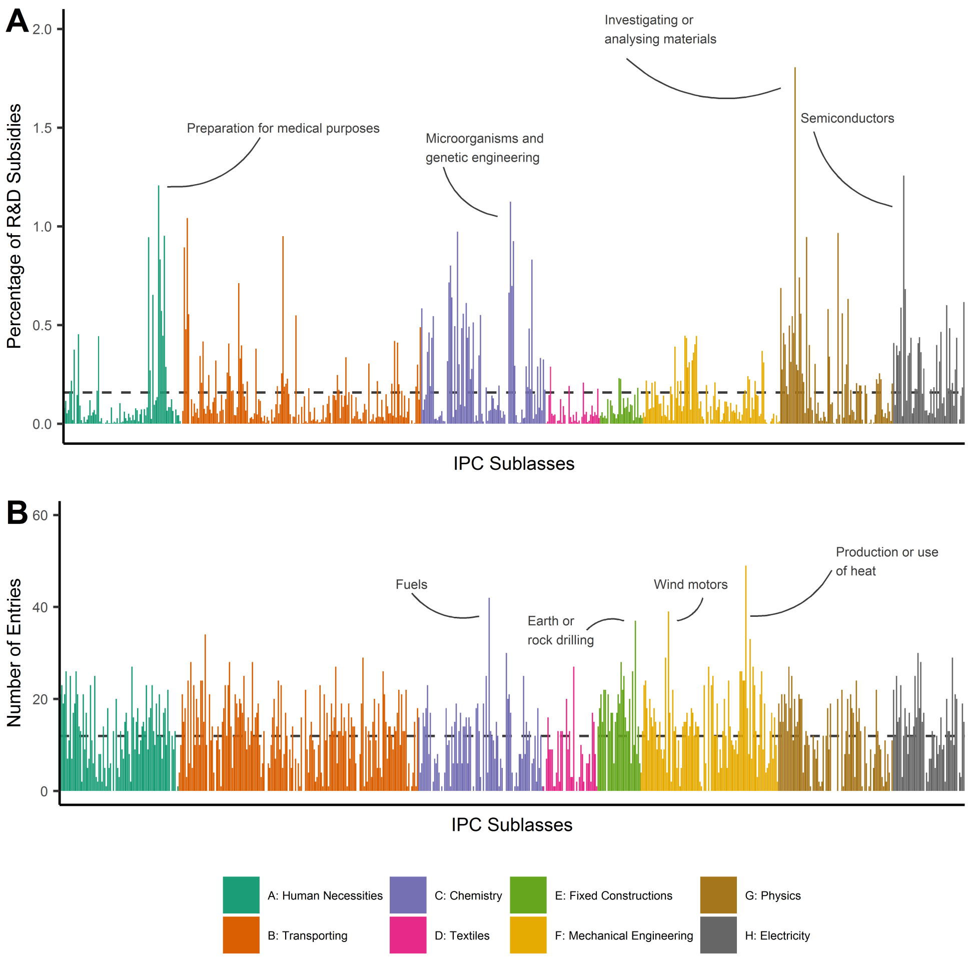

Figure 2.1: Distribution of **A** R&D subsidies and **B** percentage of entries across IPC subclasses between 2006 and 2010. Colors indicate the eight IPC main sections. The dashed horizontal lines represent the sample mean.

2.4 Results

2.4.1 The allocation of R&D subsidies

We start with the exploration of R&D subsidies' allocation. Panel A in figure 2.1 reveals the distribution of R&D subsidies across the 630 IPC subclasses between 2006 and 2010. The colors indicate the eight main sections of the IPC. Panel A shows that subsidies are not widely scattered across all main sections but rather concentrate in specific domains. A large portion of subsidies flows into technologies belonging to physics, chemistry, electricity, and human necessities. In contrast, textiles, mechanical engineering, and construction technologies receive considerably less subsidies. IPC subclasses, such as G01N (Investigating or Analysing Material), H01L (Semiconductors), A61K (Preparation for Medical Purposes), and C12N (Microorganisms and Genetic engineering) are among the most strongly subsidized technologies.

Panel B of Figure 2.1 shows how frequently technologies emerge in regions. Larger entry numbers indicate that many regions diversified into the according technologies. This reflects the spatial diffusion of these technologies within Germany. Entry numbers vary considerably between technologies, with each IPC subsection being characterized by low- and high-entry technologies. The visual inspection of Figure 2.1 reveals that subsidies are not necessarily allocated to technologies with the highest numbers of entries. For example, technologies in mechanical engineering and fixed construction show large numbers of entries and receive comparatively few subsidies. In other cases, there seems to be some alignment. For instance, the top four technologies with the highest entry numbers (F24J = Production of use of heat, C10L = Fuels, F03D = Wind motors, and E21B = Earth and rock drilling) represent technological fields related to renewable energy production or energy usage. Renewable energies have become very popular in Germany and are still strongly subsidized to support the transition from fossil energy sources to renewables (Jacobsson and Lauber 2006). This is also reflected in our data, as in this case, subsidization seems to correspond to technological entry.

Another interesting aspect to look at is the relationship between subsidy allocation and relatedness density. Figure 2.2 visualizes relatedness density differentiated by subsidized and non-subsidized projects over all four time periods (panel A to D). It is striking that relatedness density substantially differs between subsidized and non-subsidized technologies. Subsidized technologies are on average characterized by higher relatedness densities than the non-subsidized ones. Notably, this difference has grown over time. This suggests that R&D policy is increasingly subsidizing related technologies in regions.

We expand the visual inspection of the relationship between subsidy allocation and relatedness density with a linear panel regression. $Subsidies$ (and its disaggregation into $Subsidies^{Single}$ and $Subsidies^{Joint}$) serves as the dependent variable and $Density$ as the main explanatory variable. Fixed effects and additional control variables capture potential confounders. Table 2.2 reports the results. They clearly support the previous visual interpretation. Technologies in regions are more likely to receive R&D subsidies when they are related to existing regional capabilities.

In sum, the results for the allocation of subsidies in Germany suggest that contemporary project-based R&D subsidization has a tendency to support path-dependent, related diversification in regions.

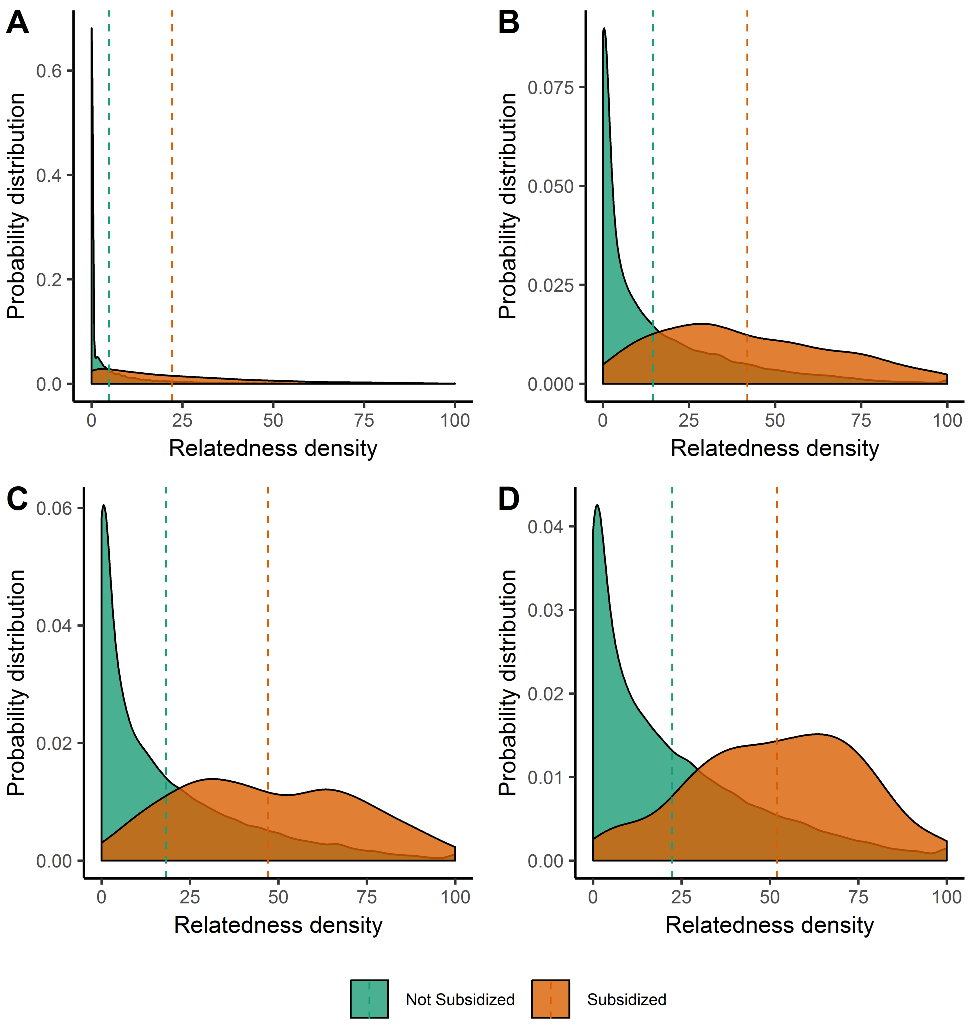

Figure 2.2: Relationship between relatedness density and R&D subsidies in different time periods with **A** 1991-1995, **B** 1996-2000, **C** 2001-2005, and **D** 2006-2010.

Table 2.2: Regression results for the allocation of subsidies

	Y = Subsidies		
	Subsidies	$\text{Subsidies}^{\text{Single}}$	$\text{Subsidies}^{\text{Joint}}$
	(1a)	(1b)	(1c)
Density	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0002)	(0.0001)	(0.0001)
Neighbor Patents	-0.0004	-0.0002^{*}	-0.0002
	(0.0002)	(0.0001)	(0.0002)
Regional Patents	-0.00003^{***}	-0.00003^{***}	-0.00001^{*}
	(0.00000)	(0.00000)	(0.00000)
Diversity	-0.0003^{**}	-0.0004^{***}	0.00005
	(0.0001)	(0.0001)	(0.00004)
Technology Size	-0.0001^{***}	-0.0001^{***}	0.00000
	(0.00003)	(0.00002)	(0.00001)
Time FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Observations	280,392	280,392	280,392
Adjusted R^2	0.471	0.424	0.475

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors were clustered at the regional and technology level.

2.4.2 The relationship between R&D subsidies and technological diversification in regions

The link between R&D subsidies and technological diversification in regions is at the center of the present paper. Figure 2.3 maps entry rates[1](panel A), the average relatedness density (panel B), the spatial allocation of R&D subsidies (panel C), and the number of patents (panel D) across the 141 German regions. The maps highlight a number of interesting spatial patterns. Firstly, entry rates tend to be larger in regions with higher patenting activities. For example, South Germany, with Munich and Stuttgart as innovative regions, is characterized by particularly high entry rates. Similar patterns are also observed for the West of Germany with Cologne and North Germany with Hamburg and Hanover as centers of innovation and technological entries. Nevertheless, some regions experience high entry rates while being only moderately successful in patenting (e.g., Chemnitz and Dresden in Saxony).

Secondly, higher entry rates seem to strongly correlate with the average relatedness density in regions. That is, regions characterized by higher relatedness densities

[1]Entry rates correspond to the number of realized entries divided by the number of potential entries.

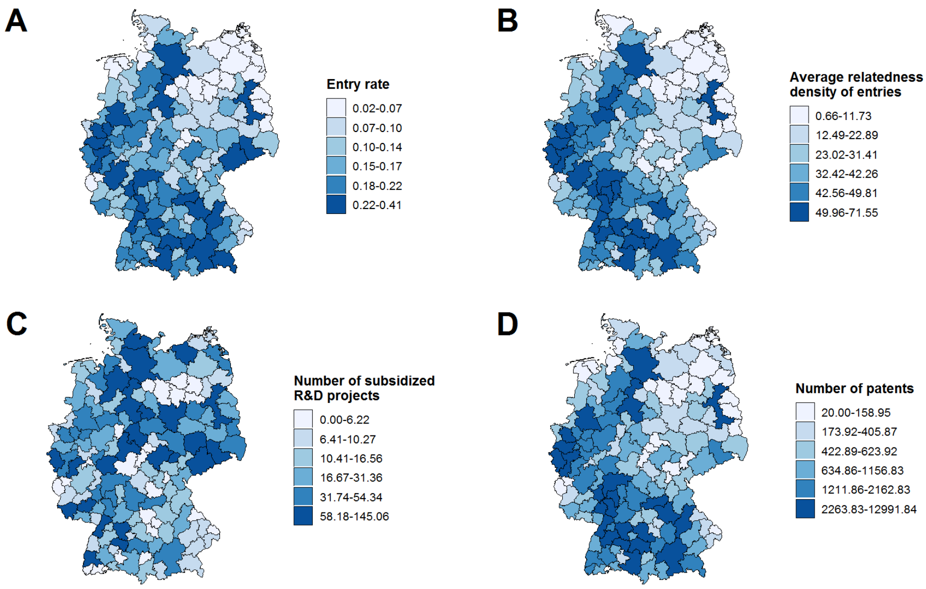

Figure 2.3: **A** Entry rates, i.e., realized entries divided by possible entries, **B** average relatedness density of realized entries, **C** number of subsidized R&D projects, and **D** number of patents in German LMRs between 2006-2010.

also realize a larger share of their entries. This visual observation corresponds to the ample empirical evidence that related activities are more likely to emerge in regions than unrelated activities (Neffke et al. 2011; Boschma et al. 2013; Boschma et al. 2015; Rigby 2015; Balland et al. 2019).

Thirdly, regions with lower patenting activities and lower entry rates (e.g., North-Eastern regions) receive more R&D subsidies than innovative regions with higher entry rates. More precisely, 9 out of the top 10, and 12 of the top 20 regions with the most subsidized R&D projects are located in the North and East of Germany. Accordingly, the allocation of R&D subsidies seems to follow a convergence strategy by favoring regions with fewer technological activities.

Our central results of the regression analysis linking subsidies to entries are reported in Table 2.3. Regarding the control variables (see Model 2d, 2e, and 2f), we find patenting activities in neighboring regions (*Neighbor Patents*) to be positively associated with regional technological diversification, which is indicated by the significantly positive coefficients for this variable in all models. Accordingly, being in spatial proximity to regions already successful in a particular technology, renders diversification into this technology more likely. The positive link between activities in neighboring regions and regional diversification supports the idea of spatial knowledge spillovers, which are intensified by geographic proximity (Jaffe et al. 1993).

In addition, our models suggest that entries are less likely to occur in regions with large knowledge stocks. The corresponding coefficient of *Regional Patents* is significantly negative. Most likely, this is the outcome of a level effect: regions with strong inventive activities are already well diversified and successful and, hence, there are fewer opportunities for further diversification (see for example, Imbs and Wacziarg 2003). A similar argument applies to the size of technologies, *Technology Size*. Its coefficient is significantly negative, indicating that large technologies are less likely to emerge in regions. This is likely driven by large technologies already being well diffused in space and, hence, they have fewer (remaining) opportunities to emerge. *Diversity* remains insignificant, which is most likely due to its effect being captured by *Regional Patents* or by the fixed effects.

Table 2.3: Regression results of linear probability model for entries

	Y = Entry					
	(2a)	(2b)	(2c)	(2d)	(2e)	(2f)
Subsidies	0.288***		0.288***		0.284***	0.274***
	(0.030)		(0.030)		(0.030)	(0.028)
Density		0.001***	0.001***	0.002***	0.001***	0.001***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Neighbor Patents				0.004***	0.004***	0.004***
				(0.001)	(0.0005)	(0.0005)
Diversity				−0.0001	−0.0001	−0.0001
				(0.0001)	(0.0001)	(0.0001)
Regional Patents				−0.00003***	−0.00001**	−0.00001**
				(0.00001)	(0.00000)	(0.00000)
Technology Size				−0.0001***	−0.0001***	−0.0001***
				(0.00002)	(0.00002)	(0.00002)
Subsidies x Density						0.001
						(0.001)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	273,825	288,543	273,825	288,543	273,825	273,825
Adjusted R^2	0.212	0.192	0.213	0.202	0.222	0.222

Note: **p<0.05; **p<0.01; ***p<0.001.*
Robust standard errors were clustered at the regional and technology level.

In all models, relatedness density is significantly positive. Technologies are more likely to emerge in regions that are related to existing regional capabilities, which confirms the path dependency of regional diversification and the idea of regional branching. Hence, our results confirm the numerous empirical studies on this matter (Boschma et al. 2013; Boschma et al. 2015; Rigby 2015; Balland et al. 2019).

We now turn toward the heart of our analysis. *Subsidies* is included into the base Model 2b without any additional variables. Its coefficient becomes significantly positive. The variable remains significant when including relatedness density (Model 1c) and further control variables (Model 1e). Accordingly, we confirm our hypothesis *H1a*, as the relationship between subsidized R&D projects and regional diversification is positive.

To approach our hypothesis *H1b* regarding a potential interplay between subsidies and relatedness, we included an interaction term of *Density* and *Subsidies* in Model 2f. Nevertheless, the corresponding coefficient remains insignificant. Accordingly, entries are not more likely to occur when the underlying technologies are related to the regional technology portfolio and receive R&D subsidies. Based on this finding, we reject our hypothesis *H1b*.

Besides the significance of the coefficient, it is usually also interesting to discuss the effect strength. Our matching of subsidies to patent data has severe implications for the interpretation of effect sizes of *Subsidies*, however. Most subsidized R&D projects are allocated (i.e., divided) to multiple technologies (IPC subclasses). This results in a fractional counting of projects, such that for each observation (technology-region combination), the absolute numbers of assigned projects do not reflect full projects but rather the corresponding shares of a project assigned to this technology by the matching procedure presented in Section 2.3.2. Accordingly, the obtained coefficient of *Subsidies* does not correspond to full projects but to fractionally allocated project numbers. With this in mind, we suggest the following interpretation: Increasing the numbers of fractionally allocated subsidized R&D projects by 0.012 will increase the probability of entries by approximately 0.35%.[2]. Accordingly, subsidies' effects appear to be relatively small.

We hypothesized that subsidies for single and joint projects are likely to have distinct effects on regional diversification (*H2a*). Table 2.4 reports the corresponding results of this differentiation. We include both subsidy types in different models. The results are robust throughout all specifications (Models 3a to 3f).

We observe substantial differences between the two subsidy variables. Both variables' coefficients are significantly positive, which confirms the previously identified positive relation of subsidies and diversification. In line with previous studies (Fornahl et al. 2011; Broekel et al. 2015a), however, the coefficient of $Subsidies^{Joint}$ [lower bound = 0.69, upper bound = 1.06], as reported in Model 3b, is significantly larger than $Subsidies^{Single}$ [lower bound = 0.2, upper bound = 0.40], as reported in Model 3a. This suggests that subsidies for joint R&D projects increase the likelihood of entries to a larger extent than do subsidies for individual projects, which confirms our hypothesis *H2a*. Expanding the numbers of joint projects by the average change between two consecutive time periods of 0.015 increases the entry probability by approximately 1.31%[3]. We also test for potential interaction effects of the differentiated versions of subsidies and relatedness to investigate hypothesis *H2b*. Interestingly, and in contrast to the findings for all subsidies, we find a significantly negative coefficient for the interaction of $Subsidies^{Joint}$ and *Density*. This finding suggests that subsidizing joint projects can compensate for a lack of relatedness to some extent.

We investigate the interaction of *Subsidies* and *Density* in more detail by grouping our observations into three sub-samples. The sub-samples represent different parts of the distribution of relatedness density values, namely, low, mid, and higher relatedness values[4]. Models 4a and 4b in Table 2.5 report the results for the subsample with *low* relatedness density. *Density* is found to be insignificant, while

[2]Increasing the average numbers of subsidized projects in a technology and region by one unit (the standard way of interpretation) equals an increase of about 91% in the numbers of projects. Due to the fractionally allocated project numbers, this is, however incorrect. Rather, the coefficient of *Subsidies* in Model 2b (0.288) and the average change of subsidized projects between t and $t-1$ in our entry sample, which equals 0.012, correspond to the following effect sizes: 0.012∗0.288∗100 = 0.35%.

[3]0.015 (average change in number of joint projects) * 0.875 (coefficient of $Subsidies^{Joint}$ in Model 3b) * 100.

[4]The three groups are defined by observations belonging to the 5-25% lowest relatedness density values (low), to the highest 75-95% (high), and those falling in between, i.e., 40-60% of relatedness density values (mid).

Table 2.4: Regression results of linear probability model for entries and subsidies for individual and joint projects

	Y = Entry					
	(3a)	(3b)	(3c)	(3d)	(3e)	(3f)
$\text{Subsidies}^{\text{Single}}$	0.333***		0.169***	0.151***	0.316***	
	(0.035)		(0.033)	(0.036)	(0.032)	
$\text{Subsidies}^{\text{Joint}}$		0.875***	0.630***	0.642***		0.957***
		(0.094)	(0.088)	(0.086)		(0.122)
Density				0.001***	0.001***	0.001***
				(0.0001)	(0.0001)	(0.0001)
Neighbor Patents				0.004***	0.004***	0.004***
				(0.0005)	(0.0005)	(0.0005)
Diversity				−0.0002**	−0.00004	−0.0003***
				(0.0001)	(0.0001)	(0.0001)
Regional Patents				−0.00002***	−0.00001**	−0.00002***
				(0.00000)	(0.00000)	(0.00000)
Technology Size				−0.0001***	−0.0001***	−0.0001***
				(0.00002)	(0.00002)	(0.00002)
$\text{Subsidies}^{\text{Single}}$ x Density					0.001	
					(0.001)	
$\text{Subsidies}^{\text{Joint}}$ x Density						−0.003*
						(0.002)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	273,825	273,825	273,825	273,825	273,825	273,825
Adjusted R^2	0.207	0.211	0.214	0.224	0.218	0.222

Note: **p<0.05; **p<0.01; ***p<0.001.*
Robust standard errors were clustered at the regional and technology level.

the estimated coefficient of *Subsidies* is significantly positive. Again, our results suggest that it is important to consider the subsidy mode, as $Subsidies^{Single}$ [lower bound = -0.068, upper bound = 0.161] is insignificant and $Subsidies^{Joint}$ [lower bound = 0.111, upper bound = 0.897] is significantly positive. This suggests that R&D subsidies for collaborative projects can compensate for missing relatedness, as there are no instances of high density in this sample and, hence, they cannot drive entry probabilities. The results change for larger relatedness values. Now *Density* becomes significant as well, while the coefficient of $Subsidies^{Joint}$ [lower bound = 0.234, upper bound = 0.512] decreases in size (Model 4f). Accordingly, these results confirm our hypothesis *H2b*: Subsidies for joint projects are able to facilitate unrelated diversification, while this is not the case for subsidized individual projects.

Table 2.5: Regression results of linear probability model for three different levels of relatedness density

	Y = Entry					
	Low	Low	Mid	Mid	High	High
	(4a)	(4b)	(4c)	(4d)	(4e)	(4f)
Subsidies	0.146^{***}		0.133^{***}		0.132^{***}	
	(0.025)		(0.030)		(0.024)	
$\text{Subsidies}^{\text{Single}}$		0.047		0.043		0.069
		(0.059)		(0.038)		(0.037)
$\text{Subsidies}^{\text{Joint}}$		0.504^{*}		0.449^{***}		0.373^{***}
		(0.201)		(0.078)		(0.071)
Density	−0.0001	−0.0002	0.002^{*}	0.003^{*}	0.002^{***}	0.002^{***}
	(0.002)	(0.002)	(0.001)	(0.001)	(0.0003)	(0.0003)
Neighbor Patents	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Regional Patents	−0.00002	−0.00001	−0.00001	−0.00001	-0.00001^{*}	-0.00001^{**}
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Diversity	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.0005^{**}	-0.001^{***}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Technology Size	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}	-0.0001^{***}	-0.0001^{***}
	(0.00004)	(0.00004)	(0.00003)	(0.00003)	(0.00003)	(0.00003)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	25,394	25,115	25,466	25,240	26,109	26,041
Adjusted R^2	0.218	0.221	0.226	0.229	0.250	0.252

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors were clustered at the regional and technology level.

2.4.3 Robustness analyses

When evaluating the effects of R&D subsidies on regional diversification, endogeneity of subsidies represents a crucial concern. In our case, endogeneity can occur if technology entries in regions impact subsidy allocation. The use of time lags of 5 years implies that technology entries would need to influence the allocation of subsidies to that same technology in the region 5 years before (when it was not existent there). While this is a highly unlikely scenario, there might be effects at work that operate over long time periods.

Another reason for endogeneity to arise in our context is the non-random selection of recipients (Busom 2000; David et al. 2000; Aubert et al. 2011). R&D policy is more likely to reward projects with higher likelihoods of success. Such is probable when recipients have been successful in acquiring projects in previous periods. For instance, subsidy recipients could have accomplished entries of technologies in regions in previous time periods, which, in turn, positively influenced the likelihood of receiving grants in subsequent projects in related technologies. Addressing this endogeneity problem is not straightforward. One possibility is to apply instrumental variables regressions (IV). Such requires a valid instrument at the level of technology-region pairs that additionally varies over time, however. We follow (Koski and Pajarinen 2015) and use the total numbers of subsidized projects (across all regions) in each technology to instrument the potentially endogenous subsidy variables at the region-technology level. The underlying rationale is that an increase in the total numbers of subsidized projects generally increases a specific regions' probability to acquire a subsidized project in this technology.

The instrument fulfills the exclusion restriction, as the number of supported projects at the national level in a specific technology 5 years prior, has no effect on the entry probability of this technology in a particular region, other than through their direct allocation to this region. This argumentation would not hold in the case of individual regions dominating the receiving of subsidies and the entry of related technologies. Our data clearly show, however, that this is not the case. Another challenge could be that our dependent variable, $Entry$, again, has an effect on the allocation of federal subsidies to technologies 5 years before. We believe this to be highly unlikely, as the emergence of single technologies in some regions does not influence the allocation of subsidies by the federal government five years before. In principle, this concerns influential recipients of R&D grants, such as large universities, research institutions, and firms. Although these actors receive most subsidies (Broekel and Brachert 2015), the average share of subsidized projects received by an individual region of all subsidized projects in one technology is 0.6% (median share equals 0.22%). Accordingly, the influence of large regional actors on the general allocation seems to be rather marginal. Consequently, we are confident that our instrument, the total number of subsidized projects in a technology for the region-specific numbers, is suitable to address potential endogeneity concerns. To achieve a closer link to the instrumented variables, we differentiate between individual ($Total^{Single}$) and joint projects ($Total^{Joint}$) in the construction of the instruments.

The IV regressions emphasize that the distinction between individual and collaborative subsidies is fundamentally important. Table 2.6 reports the results of the first- and second-stage regressions. The first-stage regression confirms that $Total^{Single}$ (Model 5a) and $Total^{Joint}$ (Model 5c) are positively related to the number of subsidized projects at the regional level. Nevertheless, the previously observed (weak) effect of individual projects on entry probably disappears in the second-stage regression. The coefficient remains positive but is insignificant (p-value = 0.36) (Model 5b). In contrast, the instrumental variable regression in model 5d confirms our results for the subsidization of joint projects. The obtained coefficient of $Subsidies^{Joint}$ remains significantly positive (p-value = 0.01) in the IV specification.

Consequently, the IV regressions substantiate our previous finding of a positive effect of collaborative R&D subsidies on regional technological entry.

Table 2.6: Results of instrumental variables regression

	1st Stage (Y = $\text{Subsidies}^{\text{Single}}$)	2nd Stage (Y = Entry)	1st Stage (Y = $\text{Subsidies}^{\text{Joint}}$)	2nd Stage (Y = Entry)
	(5a)	(5b)	(5c)	(5d)
$Total^{Single}$	0.004***			
	(0.001)			
$\text{Subsidies}^{\text{Single}}$		0.148		
		(0.161)		
$Total^{Joint}$			0.001***	
			(0.0004)	
$\text{Subsidies}^{\text{Joint}}$				1.031**
				(0.400)
Density	0.001***	0.001***	0.0004***	0.001***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Neighbor Patents	−0.0004**	0.004***	−0.0002	0.004***
	(0.0001)	(0.0005)	(0.0001)	(0.0005)
Regional Patents	−0.00003***	−0.00002**	−0.00001*	−0.00002**
	(0.00000)	(0.00001)	(0.00000)	(0.00001)
Diversity	−0.0004***	−0.0001	0.0001	−0.0003***
	(0.0001)	(0.0001)	(0.00004)	(0.0001)
Technology Size	−0.0001***	−0.0001**	−0.00001*	−0.0001***
	(0.00002)	(0.00003)	(0.00001)	(0.00002)
Time FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes
Observations	273,790	273,790	273,790	273,790
Adjusted R^2	0.454	0.214	0.441	0.222

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors were clustered at the regional and technology level.

2.5 Discussion and conclusion

Our study discusses and empirically tests the relationship between project-based R&D subsidies and regional technological diversification. It thereby contributes to two literature streams: the assessment of R&D subsidies' effects and the literature on regional diversification. Existing studies on the effects of R&D subsidies primarily focus on their general contribution to innovation activities and their potential stimulation of R&D efforts, efficiency, and outputs. In this study, we argue that they may also support technological diversification, despite not necessarily being intended to do so. Accordingly, R&D subsidies may induce additional (positive) effects that have not yet been considered in existing evaluations. With respect to the literature on regional diversification, our study adds a crucial perspective that remains underdeveloped. While (related) diversification is empirically well investigated (Hidalgo et al. 2007; Rigby 2015; Boschma et al. 2015; Essletzbichler 2015), little attention has been paid to the role of R&D policy in this context. Although we are not evaluating contemporary R&D policies' general support for regional diversification, our study draws attention toward potential side effects of other, not directly diversification-related R&D policies.

We complement our arguments with an empirical study on the technological diversification of German regions and project-based R&D subsidization of the federal government. Our empirical results for the allocation of these R&D subsidies show their allocation tends to be positively biased toward regions offering related competences. Accordingly, R&D policy seems to be part of the path dependency in regional diversification, as it manifests related activities. This suggests a rather risk-averse allocation strategy. As related activities have greater chances of becoming successful than other activities (Neffke et al. 2011; Boschma et al. 2015; Rigby 2015), supporting such minimizes the chances of failure (see discussions in Dohse (2000), Cantner and Kösters (2012), and Aubert et al. (2011)). Most likely, it is the competitive character of the allocation process through which this risk aversion is implemented. When evaluating applications, applicants' and applications' quality are relatively easy to assess and evaluate. Therefore, they are likely to be weighted more strongly than less "objective" aspects, such as novelty and future development potentials.

From the perspective of the literature on related variety (Frenken et al. 2007; Neffke et al. 2011) and the Smart Specialization strategy of the EU (Foray et al. 2011), our findings have to be evaluated as evidence for a positive contribution of the R&D subsidization policy to regions' future growth and prosperity. By allocating subsidies to technologies related to regions' existing portfolios, R&D subsidies support the emergence and growth of *related variety*. This has been argued and empirically shown to stimulate regional (related) technological diversification, which, in turn, has been confirmed to matter for regions' long-term economic growth (Frenken et al. 2007; Neffke et al. 2011; Kogler et al. 2013).

However, our study raises a crucial question rarely discussed in this context: Should policy, in fact, try to (directly or indirectly) facilitate related diversification? The regional branching mechanism suggests that related technologies are the most likely to emerge in regions (Boschma and Frenken 2010). Put differently, is related diversification truly troubled by market failures justifying policy intervention? In addition, one may argue that regional branching implies that diversification is a path-dependent process that eventually leads to a thinning out of regional knowledge diversity. This in turn makes lock-in scenarios more likely, which are to be avoided due to their negative impact on growth and future developments.

In contrast, from a market-failure perspective, it can be argued that stimulating unrelated diversification should be the focus of R&D policy, to break the constraints of existing path dependencies. Supporting unrelated diversification policy increases regional knowledge diversity. Through a portfolio effect, diversity will render regions more resilient to external shocks, which is proposed as one of the main goals of innovation policy (Martin 2012). In addition, regional technological diversity lays the foundation for unexpected and uncommon knowledge recombination, which frequently forms the basis for breakthrough inventions (Uzzi et al. 2013; Kim et al. 2016).

In accordance to this perspective, our empirical results do not hint at a multiplicative effect of R&D subsidies and relatedness. In contrast, our findings suggest the existence of a substitutional relationship between relatedness and R&D subsidies at the regional level. Hence, R&D subsidies contribute to regions diversifying into unrelated activities to some extent.

In addition, our results reveal the importance of differentiating between subsidies for individual- and joint-research projects (Broekel 2015). Subsidies for joint R&D projects exert a much stronger effect on regional technological diversification than those for individual projects. The difference becomes even more pronounced when applying instrumental variable regressions. In particular, subsidies for joint R&D projects are also able to compensate for missing relatedness to some extent. Similar is not observed for individual R&D subsidies. Most likely, it is their stimulation of interactions between new and heterogeneous actors from different regions facilitating inter-organizational learning that explains their advantage in this context. This adds to existing research showing their higher effectiveness for stimulating innovation activities in general (Fornahl et al. 2011; Broekel 2015; Broekel et al. 2017). It also begs the question of why the majority of projects subsidized by the German federal government do not yet involve inter-organizational collaboration (Broekel and Graf 2012).

Our paper opens a number of avenues for future research. The scope of our study is limited to technological diversification in regions, approximated by patent data. Although patent data have their justification and are often used in this context (Boschma et al. 2015; Rigby 2015; Balland et al. 2019), they also limit our analysis to technologies that can be patented. It is therefore important to study the link between subsidies and other forms of diversification to improve our understanding of policy impact on regional diversification. For instance, this concerns sectoral diversification measured with information on the occupational composition in regions, representing a crucial next step for future research.

Additionally, R&D policy still lacks the appropriate tools to identify promising but underdeveloped technologies and for evaluating the spatial context in which they (best) evolve. We believe that our paper takes a step in that direction by showing that regional branching helps in understanding the economic transformation of regions. Moreover, we provide an empirical set-up for evaluating the role of a specific policy tool (R&D subsidies) in this context.

2.A Robustness analyses

Table 2.7: Regression results of logit model for entries

	Y = Entry					
	(6a)	(6b)	(6c)	(6d)	(6e)	(6f)
Subsidies	1.059***		1.064***		0.969***	0.892***
	(0.125)		(0.124)		(0.118)	(0.132)
Density		0.008***	0.008***	0.012***	0.011***	0.010***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Neighbor Patents				0.015***	0.016***	0.016***
				(0.002)	(0.002)	(0.002)
Diversity				−0.005***	−0.004***	−0.004***
				(0.0005)	(0.001)	(0.001)
Regional Patents				−0.00003	−0.00002	−0.00003
				(0.00003)	(0.00002)	(0.00002)
Technology Size				−0.0004***	−0.0003***	−0.0003***
				(0.0001)	(0.0001)	(0.0001)
Subsidies x Density						0.005*
						(0.002)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	273,825	288,543	273,825	288,543	273,825	273,825

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors were clustered at the regional and technology level.

Table 2.8: Regression results for specialization (LQ greater 1 as entry threshold)

	Y = Entry					
	(7a)	(7b)	(7c)	(7d)	(7e)	(7f)
Subsidies	−0.013		−0.014*		−0.016*	0.002
	(0.007)		(0.007)		(0.007)	(0.012)
Density		0.0002	0.0002*	0.001***	0.001***	0.001***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Neighbor Patents				0.0004**	0.0004**	0.0004**
				(0.0001)	(0.0001)	(0.0001)
Diversity				−0.001***	−0.001***	−0.001***
				(0.0001)	(0.0001)	(0.0001)
Regional Patents				−0.00000	−0.00000	0.00000
				(0.00000)	(0.00000)	(0.00000)
Technology Size				−0.0001***	−0.0001***	−0.0001***
				(0.00001)	(0.00001)	(0.00001)
Subsidies x Density						−0.0003*
						(0.0002)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277,135	291,650	277,135	291,650	277,135	277,135
Adjusted R^2	0.056	0.058	0.056	0.065	0.062	0.062

Note: **p<0.05; **p<0.01; ***p<0.001*
Robust standard errors were clustered at the regional and technology level.

3 Technological Complexity and Economic Growth of Regions

Abstract: The effect of technological complexity on regional economic growth has yet not been investigated. We assess the complexity of technological activities in 166 European NUTS 2 regions using the new measure of Structural Diversity. Panel regressions covering the years 2000 to 2015 suggest that technological complexity is a positive predictor of economic growth at the regional level. A ten percent increase in regional complexity is associated with a corresponding GDP per capita growth by about 0.28 percent. The conducive role of knowledge complexity for economic growth is therefore not only evident for countries, but also for regions revealing a relationship, which is systematic at different spatial scales and for different types of knowledge complexity.

This chapter is co-authored with Tom Broekel. The PhD candidate is the first author of the article. The paper is currently in a revise-and-resubmit stage in Research Policy.

3.1 Introduction

"I remember thinking how comfortable it was, this division of labor which made it unnecessary for me to study fogs, winds, tides, and navigation, in order to visit my friend who lived across an arm of the sea. It was good that men should be specialists [. . .]. The peculiar knowledge of the pilot and captain sufficed for many thousands of people who knew no more of the sea and navigation than I knew" (London 1904).

Over a hundred years ago, Humphrey van Weyden praised the benefits of specialization and division of labor aboard a small vessel in Jack London's famous novel "The Sea Wolf". They allowed him to concentrate on the things that caught his interest and talents more than others. An implicit consequence of specialization and division of labor is the constantly increasing complexity of the world's knowledge (Aunger 2010). Another one is increasing productivity, which makes greater economic surpluses possible and allows for sustaining larger population sizes (Smith 1776). The division of labor, however, requires the coordination and cooperation of specialists to utilize large amounts of diverse knowledge, which is easier in larger and more densely populated areas (Becker and Murphy 1992). In turn, such larger and more densely connected populations fuel further specialization and division of labor (Sveikauskas 1975). This self-reinforcing process accelerated the richness and complexity of knowledge production over time (Kremer 1993; Henrich 2004).

Knowledge, in general, represents a critical resource in today's knowledge economy (Lucas 1988; Romer 1990). However, not all knowledge is equally valuable. More complex knowledge is argued to be a fundamental building block of competitive advantage and economic growth (Kogut and Zander 1992). Its economic relevance rests on the idea that complex knowledge is difficult to imitate and only few economic actors have the capabilities to produce it (Storper 2010). Accordingly, firms and economies with complex knowledge are likely to earn rents in form of higher growth and wealth (Kogut and Zander 1992; Teece et al. 1997; Hidalgo and Hausmann 2009).

Until now, empirical evidence is scarce and restricted to economic complexity as measured by the product portfolio of an economy (Hidalgo and Hausmann 2009; Hausmann et al. 2013; Bahar et al. 2014). Production, however, is only one dimension of knowledge complexity in which economies compete. Technological knowledge production is complementary and similarly vital for economies' competitiveness and growth (Nelson and Winter 1982; Lucas 1988; Romer 1990). Yet, the empirical relation between technological complexity and economic growth is still unexplored.

In this article, we seek to close this research gap by explicitly studying this relation at the regional level, i.e. for European *NUTS 2* regions between 2000 and 2016. Following previous studies, we approximate the complexity of technologies by relying on patent documents (Fleming and Sorenson 2001). In particular, we assess technological complexity using the recently developed measure of *Structural Diversity* (Broekel 2019).

The panel regression results confirm that technological complexity is a positive and significant predictor of economic growth in European regions. More precise, we find that a 10 percent increase in technological complexity relates to a 0.28 percent increase in regional GDP. Accordingly, the paper complements previous studies highlighting a positive link between economic complexity and economic growth by presenting quantitative evidence of the economic relevance of technological complexity.

Our study is structured as follows. Section 3.2 provides an overview of the theoretical and empirical literature on knowledge complexity. Section 3.3 presents the empirical data and our estimation approach. The empirical results are presented in Section 3.4. Section 3.5 concludes the paper.

3.2 Theoretical background and literature overview

Knowledge production is a fundamental source of long-term economic growth (Kuznets 1962; Nelson and Winter 1982; Romer 1990; Grossman and Helpman 1991; Aghion and Howitt 1998) and helps to understand the uneven growth patterns of regions (Glaeser et al. 1992; Fagerberg et al. 1997; Henderson et al. 2001). Knowledge accumulates over time in and adheres to certain locations, which leads to a strong spatial concentration of knowledge in regions (Feldman 1994). One important reason for the spatial concentration of knowledge is due to the sensitivity of knowledge spillovers to geographic distance limiting the spatial diffusion of knowledge and contributing to its geographic concentration (Jaffe et al. 1993; Markusen 1996). Crucially, the degree of spatial concentration varies significantly between knowledge domains (Breschi and Malerba 1997). While in the past, researchers highlighted the role of tacit knowledge in this context, knowledge complexity has increasingly been in the focus as one crucial dimension that explains the varying spatial concentration of knowledge domains.

However, in contrast to the intensity of the discussion on knowledge complexity and its economic relevance in the literature, there is (still) no common definition of knowledge complexity. Yet, consensus seems to exist on a number of basic features. To Winter (1987, p. 177), complexity is *"the amount of information required to characterize the item of knowledge in question"*. Zander and Kogut (1995) rely on a similar understanding of knowledge complexity, which focuses on the diversity of knowledge combination. Accordingly, knowledge *"is more complex when it draws upon distinct and multiple kinds of components"* (Zander and Kogut 1995, p. 79). Kauffman (1993) defines complexity in a related manner, as the interaction between size and interdependence of components. This builds on Simon's (1962) description of complex systems. For him, complexity is *"made up of a large number of parts that interact in a nonsimple way"* (Simon 1962, p. 468). Interestingly, there is a some similarity to Polanyi's (1966) notion of tacitness. The more information, e.g. diverse range of combinations, interdependencies, and competences, a system entails, the more difficult becomes communication and codification. Similar arguments are found in the literature on technological complexity. Here, technologies are described as compositions of multiple components that are combined to fulfill a specific purpose (Usher 1954; Hargadon 2003; Arthur 2009). The number of components, their intensity of combination, and the ways how they are combined, are here seen as primary determinants of technologies' complexity (Fleming and Sorenson 2001; Broekel 2019).

Complexity is one important dimension determining the costs and time of knowledge imitation. Particular capabilities are required to absorb and successfully integrate more complex types of knowledge (Teece 1977; Rogers 1983; Winter 1987; Kogut and Zander 1992). In addition, errors in imitation tend to become more frequent with growing complexity, which impacts performance negatively and suggests that imitation is not a promising strategy in complex knowledge domains (Rivkin 2000). Hence, complex knowledge is less likely to spillover to competitors. In addition, Yayavaram and Chen (2015) demonstrate that complexity not only challenges learning inputs, but also influences innovation output. The acquisition of new and complex knowledge in innovation processes impedes learning and hurts innovation outcomes. Consequently, complex knowledge represents an entry barrier, as it is more difficult to learn and to copy.

The ability to learn and acquire complex knowledge is therefore argued to be more valuable and to translate into higher economic rents than knowledge that can be easily acquired (Winter 1987; Kogut and Zander 1992; Zander and Kogut 1995; Teece et al. 1997; Storper 2010). As complex knowledge represents a critical resource, economic actors can build competitive advantage based on complex knowledge providing them with profound growth potentials and access to quasi-monopolistic rents (Teece 1977; Kogut and Zander 1992; Zander and Kogut 1995; Teece et al. 1997; Rivkin 2000; McEvily and Chakravarthy 2002; Sorenson et al. 2006). Empirical insights back this argument, as Fleming and Sorenson (2001) show that more complex inventions receive more citations (as an indicator of impact and value) indicating that complex inventions are more valuable than technologically simpler inventions. Their benefits are also shown to stay with the inventor (Sorenson et al. 2006).

Crucially, geographic proximity plays an important role in the creation and diffusion of complex knowledge. It is widely accepted and empirically confirmed that geographic proximity facilitates interactions and engagement in networks (Becker et al. 1999; Boschma 2005; Breschi and Lissoni 2009). Thereby, geographic proximity stimulates interactive learning required for the creation of complex knowledge. In addition, it eases its exchange by allowing for easier and quicker feedback, spontaneous interactions of heterogeneous actors, and more efficient communication (Malmberg and Power 2005). Empirical confirmation for these arguments are delivered by the study of Balland and Rigby (2017). These authors find that complex technologies diffuse slower in space than simple ones. More indirectly, Broekel (2019) studies the relation between technological complexity and spatial concentration by using information on patented inventions. His empirical evidence suggests a positive relationship between technological complexity and geographic concentration. Importantly, complex technologies do not randomly concentrate in space, but rather seem to follow a distinct spatial pattern as shown by Balland et al. (2018). Their results highlight that complex technologies concentrate in large urban agglomerations. Moreover, this concentration in urban areas has steadily increased over the last 150 years.

In light of the previous discussion, this also suggest that the economic benefits of complex knowledge are unevenly distributed in space with a tendency to concentrate in urban agglomerations. The question at the heart of the present paper is therefore, if regional differences in complex knowledge explain the uneven economic growth of regions?

Existing empirical research only provides indirect evidence for the economic relevance of technological complexity at the country level. In their seminal paper, Hidalgo and Hausmann (2009) introduce the *Economic Complexity Index* (ECI) to approximate the economic complexity of countries based on their production capabilities. The ECI builds on the spatial distribution of export products across countries. In this framework, products (and the knowledge underlying their production) exported by few and most diversified economies are assumed to be more complex. On this basis, the authors show empirically that countries with greater economic complexity are characterized by higher levels of GDP per capita and experience higher short-term GDP growth. Subsequent studies have supported the findings of Hidalgo and Hausmann (2009) that economic complexity matters for countries' economic growth (Ferrarini and Scaramozzino 2016; Stojkoski et al. 2016). However, the effects of technological complexity on economic growth have not been studied at the regional level so far. This contrasts the crucial role of geographic proximity for the production and diffusion of complex knowledge, which is more likely captured at the regional than at the national level (Balland and Rigby 2017; Balland et al. 2018; Broekel 2019). The present paper seeks to fill this gap and the argumentation above leads to the following hypothesis:

H1: Higher levels of technological complexity are beneficial for regional economic growth.

3.3 Materials and methods

Our unit of analysis are NUTS 2 regions in Europe for which we collect a rich set of variables for all years between 2000 and 2015. We choose NUTS 2 regions primarily for reasons of data availability. Clearly, labor market regions would be more appropriate to capture the regional dimension of innovation processes. However, there is no common definition for all EU member states and many empirical variables are not available at other levels. The final sample size for the empirical estimations is 2,656 observations composed of 166 unique regions observed in 16 years. Figure 3.1 maps the regions in our sample. We only consider regions for which we have the full set of information for all variables. In a common manner, we approximate economic growth by change in GDP. Information on GDP in NUTS 2 regions comes from *Eurostat*, which provides regional time-series data. On this basis we define our dependent variable as GDP per capita at Power Purchasing Standards in year t.

3.3.1 Regional technological complexity

Our central independent variable is technological complexity at the regional level. The construction of this variable relies on patent information. We use patent data of the *OECD REGPAT Database* (March 2018 version), which covers annual patent applications to the *European Patent Office* (EPO). Although patents come with several disadvantages, they are nevertheless widely-used in empirical analyses to study technological activities (Griliches 1990). This is mainly because patents are the only large-scale data source providing such detailed information about regional knowledge stocks.

Calculating technological complexity is not a straightforward task, as there is no established method so far. The Economic Complexity Indicator by Hidalgo and Hausmann (2009) seems to be the most prominent approach in today's literature. However, it was developed to assess the economic complexity of countries based on their export portfolios. While the ECI has been used to approximate technological complexity using information on patent activities of countries and regions (Balland and Rigby 2017; Petralia et al. 2017), such applications face a number of issues. For instance, the ECI is based on the spatial distribution of technologies, which may create endogeneity issues in spatial research. Moreover, the spatial distribution of technologies is shaped by many factors of which complexity is but one. The character of an index also makes the comparison of the ECI over time and its inclusion in panel regressions problematic. Moreover, many of its empirical characteristics do not reflect with what one would expect of technological complexity (see Broekel 2019 for a discussion). We therefore rely on the measure of *Structural Diversity* that was recently developed by Broekel (2019) and is more closely attached to the notion of technological complexity as presented in the previous section[1].

Structural Diversity relies on information theory and assesses the diversity of knowledge combinations of a technology. It rests on the idea of technologies consisting of components that are combined with each other (Hargadon 2003; Arthur 2009). Consequently, they can be represented as networks with components as nodes and their combinations as links (so-called combinatorial networks). For instance, a table can be seen as a combination of four poles and one table plate, i.e. its components. The idea of structural diversity is to measure the diversity of how these components are combined with each other. In case of a table, all four poles are directly "combined" with the table plate but not among each other. Accordingly, the combinatorial network of a table corresponds to a star-like network composed of one central and four peripheral components. Since little information is required for its description, the star-like network of a table represents a relatively simple network structure. In contrast, some components of a car might also be related in a star-like manner (front, back and side windows with the car body), while others may rather be connected in form of a "line": steering wheel to steering column to steering gear[2]. The diversity of these combinatorial structures (topologies), in addition to their size and interdependency, determines the amount of information required for their description.

Complex networks entail more information than simpler networks (Emmert-Streib and Dehmer 2012; Broekel 2019). Accordingly, the complexity of a technology increases with the information required to describe its combinatorial network. Structural Diversity exactly measures this diversity of combinatorial structures that determines the amount of information required for their description. Since complex technologies entail more information, they are in turn more difficult to learn and to copy limiting their diffusion, which represents one essential characteristic of complexity (Kogut and Zander 1992).

In practice, Structural Diversity is approximated with the *Network Diversity-Score* (NDS) developed by Emmert-Streib and Dehmer (2012). The NDS uses network topologies (e.g. size, modules, graphlets) to assess the complexity of network structures by distinguishing between simple/ordered, complex, and random networks. Complex networks are characterized by larger topological diversity than simple/ordered ones, and random networks show even higher levels of topological diversity than complex ones[3]. The NDS measure captures and quantifies these differences on a continuous scale with small values indicating networks with high degrees

[1]We also used the ECI (Hidalgo and Hausmann 2009) to estimate regional complexity scores (see Appendix 3.B for more details).

[2]This is a very simplified illustration to highlight the central idea and not an actually representation of the "car" technology.

[3]This is not to say that there are actually fully random combinatorial networks. Rather, networks with less ordered structures tend to have higher levels topological diversity.

of randomness and larger values signaling the extent to which ordered, i.e. simpler, topologies shape these networks' structures. However, when considering knowledge production, higher complexity is generally associated with higher values. Hence, Broekel (2019) suggests to invert and log-scale the measure such that larger values indicate higher complexity values. This transformation simplifies the interpretation.

We calculate technological complexity for all 655 4-digit technology classes of the *Cooperative Patent Classification* (CPC) using this measure of structural diversity. For each technology class c ($c = 1, \ldots, n$), we build its corresponding (binarized) network composed of the co-occurrence of the most fine-grained classes (10-digit) appearing on patents p classified into c. In this way, the combinatorial network entails all components that constitute technology c as well as all components they are combined with (Broekel 2019). The NDS is applied to each network resulting in the complexity measure cpx_c. Note, to make the networks more stable, we use a three-year moving window approach, i.e. the value cpx_c in year t is based on all patents in c in the years t to $t-2$.

The 655 individual complexity scores are subsequently assigned to the corresponding CPC classes appearing on patent documents. The mean complexity of all classes c on patent i represents the average patent complexity $pcpx$:

$$pcpx_i = \frac{1}{n} \sum_{c=1}^{n} p_{i,c} \tag{3.1}$$

The literature does not provide a common approach of how to adequately aggregate a patent level (complexity) measure to the regional level. Put differently, what is the appropriate way to approximate the technological complexity of a regions' patent portfolio when only the complexity of each patent is known? Inspired by the work of Balland et al. (2018), we therefore propose a regional complexity measure based on the regional distribution of patents' complexity values. More precisely, we define regional complexity as the average complexity of all patents that are above any x^{th} percentile of the regional complexity distribution. This is motivated by the idea that the existence of relatively less complex patents in a region, i.e. simple activities, does not provide any information about a region's capability to develop and manage complex technological activities. The raw mean, for example, would discriminate against complex technologies by including the simplest ones. More formally, let P^{th} denote the set of patents' complexity scores $pcpx$ ($pcpx = 1, \ldots, n$) belonging to the x^{th} percentile of the regional complexity distribution. Regional complexity $rcpx$ in region r and time t can then be defined as follows:

$$pcpx_{r,t} = \frac{1}{n} \sum_{i \in P_{th}}^{n} pcpx_{i,r,t} \tag{3.2}$$

where th defines the percentile threshold. For example, if th takes the value 10, the regional complexity represents the average of the top 10% of the most complex patents in region r at time t. Obviously, the obtained regional complexity value depends on the rather arbitrary definition of th. We assume that the relationship between regional complexity and economic growth increases by shifting the threshold to more complex activities (e.g. from top 50% to the top 10%), as, in light of the above discussion, the top percentiles represent the (unobserved) regional ability to produce and utilize complex technologies. We present the threshold sensitivity along with our results in Section 3.4. The chosen approach is similar to the one applied in Balland et al. (2018), who chose a threshold of the top 25 percent most complex activities in order to identify scaling relations.

3.3.2 Control variables

In addition to complexity, the literature has identified other determinants of regional growth and potential confounders of complexity, which are important to control for. We distinguish between two sets of control variables that approximate regional technological capabilities and those providing information on the local economic structure of regions.

Technological Capabilities

We consider the number of regional patents to control for the size of the local knowledge stock and thus for the cumulative character of knowledge. The long discussion about specialization and diversity indicates that not only size effects, but also the local technology structure plays a fundamental role for regional growth. This debate has not come to a final conclusion yet and it rather seems that both, specialization and diversity, can be beneficial (Beaudry and Schiffauerova 2009). The distribution of patents across technologies gives information about regional specialization and diversity respectively. In a common manner, we measure specialization as the average location quotient. To approximate regional diversity, we rely on the Shannon entropy. The exponential of the individual entropy scores gives a diversity score, which is comparable across regions (Jost 2006).

Lastly, complexity is sometimes associated with high-tech activities (Eurostat 2016). For example, Eurostat defines high-tech as a predetermined set of patent classes. To test complexity against this exogenous definition of high-tech activities, we include the regional share of patents in high-technologies (as defined by 2016 2016) as an explanatory variable into the analysis.

Regional Economic Structure

We complement our patent-based indicators with economic variables at the regional level, which we all collected from Eurostat. The literature on urban scaling has shown that populated places are more productive with respect to socio-economic outcomes such as GDP and innovation (Bettencourt et al. 2007b). To control for these urbanization effects, we include population density as an additional explanatory variable. The availability of human capital in form of highly educated people is also beneficial for regional growth (Lucas 1988). Additionally, the increasing complexity of technologies requires better skilled labor. In line with previous studies, we use the share of people with a tertiary education as a proxy for human capital (Broekel 2012). We also control for local unemployment rates, as higher rates are negatively associated with economic growth. Lastly, we include the share of employees in manufacturing. Table 3.1 summarizes all variables, their empirical definition, and their data sources. Basic descriptive statistics and correlations between these variables are reported in Table 3.2.

Table 3.1: List of variables with their definitions and data sources

Variable	Definition	Data Source
rcpx	Average regional complexity (in ln): [Average of the top % most complex patents of the regional complexity distribution]	OECD REGPAT Database, own calculation
lgdp	Gross domestic product per inhabitant (in ln): [Total gross domestic product (purchasing power standards) divided by total population]	Eurostat
lpat	Total number of regional patents (in ln)	OECD REGPAT Database, own calculation
lq	Average regional location quotient	OECD REGPAT Database, own calculation
div	Regional diversity measured as the exponential of the Shannon entropy of the regional patent distribution (Jost 2006).	OECD REGPAT Database, own calculation
htec-pat	Share of patents in high-tech (in %): [Total regional patents in high-tech classes divided by total number of regional patents * 100]	OECD REGPAT Database, own calculation. Note: High-tech classification is based on predefined technology classes considered as high-tech by Eurostat (2016).
lpopdens	Population density (in ln): [Economically active population aged 15-64 / Land area in square km]	Eurostat
hc	Human capital (in %) defined as: [Persons with tertiary education aged 25-64 divided by total population aged 25-64]	Eurostat
unemp	Regional unemployment rate (in %) defined as: [Unemployed persons divided by economically active population * 100]	Eurostat
share manufac	Employees in manufacturing (in %) defined as: [Employees in manufacturing divided by total number of employees * 100]	Eurostat

Table 3.2: Descriptives and Correlations

	Variables	Min	Max	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	lgdp	9.43	11.26	10.17	0.25									
2	rcpx	11.14	13.15	12.17	0.27	0.12								
3	lpat	3.91	9.35	6.53	1.11	0.42	0.28							
4	lq	1.31	46.06	4.75	4.25	-0.06	-0.33	-0.52						
5	div	10.31	264.32	126.51	43.67	0.24	-0.02	0.71	-0.57					
6	htec-pat	0.31	16	4.13	2.60	0.02	0.66	0.12	-0.21	-0.20				
7	lpopdens	0.38	7.67	4.44	1.17	0.30	0.09	0.33	-0.33	0.23	0.05			
8	share manufac	4.70	44.88	22.12	8.16	-0.24	-0.12	0.14	0.01	0.24	-0.20	-0.15		
9	hc	7	52.20	26.16	8.14	0.35	0.23	0.16	-0.11	-0.09	0.28	0.16	-0.51	
10	unemp	1.20	36.20	7.36	4.08	-0.35	0.09	-0.14	-0.02	-0.09	0.06	-0.05	-0.15	0.01

3.3.3 Methodology

We estimate a dynamic panel model with region and time fixed effects to identify the relationship between regional complexity and economic development in the following form:

$$lgdp_{r,t+x} = \tau lgdp_{r,t} + \beta_1 rcpx_{r,t} + \gamma X_{r,t} + \phi_r + v_t + \mu_{r,t} \tag{3.3}$$

where $lgdp_{r,t+x}$ denotes our dependent variable GDP per capita for every spatial unit r in the sample at time $t+x$ and $rcpx_{r,t}$ represents regional technological complexity. The subscript $t+x$ denotes that GDP on the left-hand side is serially leading. The dynamic nature of our panel model assumes that past values of $lgdp_t$ influence subsequent ones in $t+x$. Since the nature of the underlying time lag structure is unknown, we explore time lags of 1 to 6 years. The scalar τ is the response parameter of our dependent variable in its lagged version in time $lgdp_t$. $X_{r,t}$ is a $N \times K$ matrix of control variables. The corresponding $K \times 1$ vector γ contains the response parameters of our control variables. As mentioned above, we include regional and time fixed effects as denoted by ϕ_r and v_t respectively, to control for unobserved heterogeneity that is constant over time. $\mu_{r,t}$ denotes the error term, which is assumed to be spatially as well as serially autocorrelated. Therefore, we calculate robust standard errors clustered at the regional and time level (Cameron et al. 2011).

3.4 Results

3.4.1 Complexity and regional growth

All results in this section rely on the chosen threshold of the 10^{th} percentile in the calculations of regional complexities. That is, a region's technological complexity corresponds to the mean structural diversity of patents with the 10% highest values[4]. We also restricted our analysis to regions with at least 50 patents per year. This threshold is necessary to provide reliable results for all variables that are based on patent data[5].

Before we turn to our estimation outcomes, we present descriptive results regarding regional complexity in Europe. Figure 3.1 maps regional complexity across our sample of regions for the time period 2000-2015 (panel A). Values are grouped from low to high complexity using quantiles of the cross-regional complexity distribution. The map visualizes some interesting spatial patterns. In general, high complexities are relatively scattered across the continent. Almost every country has at least one region in the highest complexity group and in many cases, this is the capital city or the region with the largest population. The south of Germany and large parts of Scandinavia, generally considered as highly R&D intensive with many technological leaders, represent agglomerations of high complexity regions. Figure 3.1 panel B shows the distribution of regional complexity in our sample. Regional

[4]The results are very robust with respect to the choice of the percentile, which will be further discussed at the end of this section.

[5]Due to the threshold of 50, 59 regions were removed from the sample. Our results, however, are not sensitive to the chosen patent threshold as indicated by our robustness check reported in Section 3.C

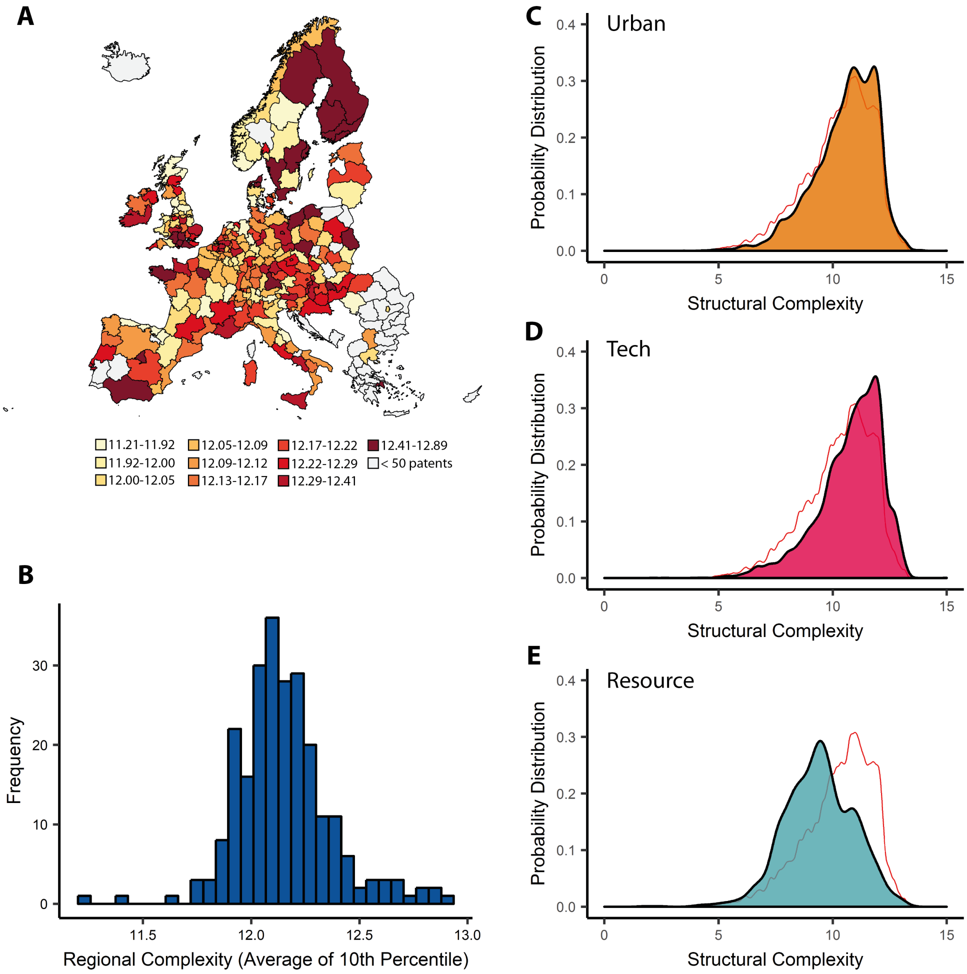

Figure 3.1: Technological complexity in Europe between 2000 and 2015. **A** Map and **B** distribution of regional complexity scores across all regions. Patents' complexity distribution in all regions (red line) compared with three selected regions **C** urban region (Berlin, Germany), **D** technology-intensive region (Oxfordshire, United Kingdom), and **E** resource-intensive region (Agder og Rogaland, Norway).

complexity scores are rather normally distributed with some outliers at both ends of the distribution.

Figures 3.1 C-E visualizes intra-regional complexity distributions in three sample regions in relation to the regional average. Berlin represents a large metropolitan area considered as a growing tech-region (panel C, *"Urban"*), Oxfordshire is a well-known R&D intensive region (panel D, *"Tech"*), and the economy of Agder og Rogaland in South-West Norway is focused on extracting technologies in the oil and gas industry (panel E, *"Resource"*). The complexity distribution of their patenting activities visualizes these structural differences. The distribution of Agder og Rogaland has a wide range and is centered at relatively low complexity values compared with the European average. Oxfordshire's distribution is characterized by a rather narrow range concentrated at the top end of the complexity distribution supporting the region's image as a R&D hub. Berlin's (urban) distribution of patents' complexity is similar to that of Oxfordshire signaling its strength in highly complex technologies. However, its distribution is, in contrast, much wider, as the city is also producing substantial numbers of patents with rather medium levels of complexity.

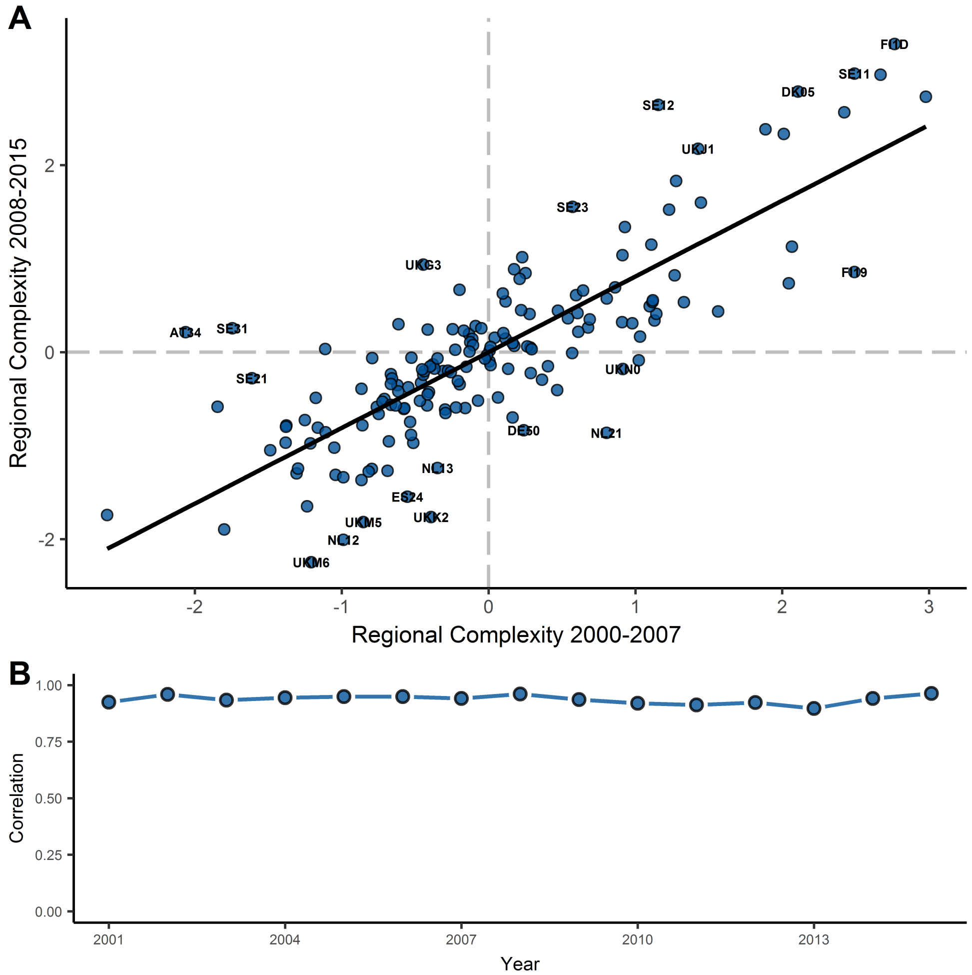

Figure 3.2: Cross-regional dynamics of complexity over time. **A** Regional complexity in two consecutive time periods. The corresponding correlation coefficient is 0.79. **B** Annual Spearman correlation of regional complexity between 2001 and 2015.

Similar distributions can be observed for other large urban regions in Europe such as Paris, Munich or Madrid. Accordingly, the regional complexity distribution illustrates structural differences in regions' technology profiles that correspond to their general technological and economic structures.

Figure 3.2 displays cross-regional dynamics of complexity over time. Panel A compares the complexity of regions in two consecutive time periods. Both values are highly correlated with, as indicated by the Spearman correlation coefficient of 0.79. This suggests that regional complexity is relatively persistent over time. Panel B in Figure 3.2 further supports this observations by showing the Spearman correlation coefficient of consecutive complexity values, i.e. $rcpx_t$ and $rcpx_{t+1}$. Over a 15-year time period, the coefficient lies in the range of 0.90 and 0.96 indicating that regional complexity changed relatively slowly between 2000 and 2015. Nevertheless, regional technological complexity is not time-invariant as illustrated by panel A in Figure 3.2. Some regions managed to substantially increase their ability to produce complex technologies. This particularly applies to regions in North and West Europe such as Stockholm (SE11) or Oxfordshire (UKJ1). The strongest increase is observed in Vorarlberg (AT34) with about 4.5%.

To answer our main research question, we employ growth regressions to understand the underlying time dynamics between complexity and GDP growth. For these, it is essential to consider the potential of a time lag in complexity's effect on regional growth. Table 3.3 summarizes the results for the full regression model considering all control variables and columns 1 to 6 correspond to different time lags in years (i.e. 1-6 years).

All models are relatively robust regarding the results of the control variables. With three exceptions, all control variables behave in line with our expectations. Past values of GDP (*lgdp*) are a strong and positive predictor for subsequent GDP. Its explanatory power also declines with increasing time lags until it finally yields insignificant estimates in Model 6. This finds reflection in the constantly decreasing goodness of fit from Model 1 to 6.The local knowledge stock (*lpat*) and specialization (*lq*) are significantly positive indicating the importance of the local knowledge stock in terms of size and structure for economic growth (Model 1 to 3). However, the local knowledge stock becomes insignificant when considering time lags exceeding three years (Models 4-6). Regional unemployment (*unemp*) is significantly negative when lagged by one or two years.

Three findings do not correspond to our expectations. The share of high-tech patents ($htec-pat$) does not predict economic growth in any Model specification. To a certain extent, this is due to the consideration of technological complexity, which outperforms the share of high-tech patents (see also Table 3.4). The coefficient of population density (*lpopdens*) is significantly negative in most models (Model 4 to 6). It implies that urban regions grow less than rural regions. Similarly human capital (*hc*) is not a robust predictor of economic growth, as the variable yields insignificant and in later models significantly negative results. We believe that the consideration of patents and GDP already capture most of differences population densities and in the quality of regions' human capital explaining the results for both variables.

Using this empirical set-up, we explore the relationship between complexity and economic growth. The coefficient of complexity (*rcpx*) is insignificant when time lags of one and two years are applied (Model 1 and 2). This changes for time lags of three and more years (Model 3 to Model 6). Complexity becomes a positive and significant predictor of economic growth. The dependence of complexity's coefficient on the applied time-lag is visualized in Figure 3.3. The coefficient increases from 0.028 (Model 3) to a maximum of 0.045 (Model 5). The difference, however, is not statistically significant (see Figure 3.3). The insignificance of *rcpx* for small time lags is very reasonable because technological capabilities do not unfold a direct and immediate effect on economic growth but rather seem more important for economic growth in longer time periods.

Complexity and GDP per capita are measured on a logarithmic scale (see Table 3.1 for an overview). Therefore, we can interpret the coefficients as elasticities. That is to say, a one percent increase in average regional complexity yields a 0.027% increase in GDP per capita three years later (Model 3). To put this into perspective, the average growth rate of complexity between 2000 and 2015 is 1.3%. Accordingly, a one percent increase in regional complexity represents a change, which is close to the average growth of complexity over 15 years.

As an additional benchmark, a one percent increase in patents translates into a growth in GDP per capita by about 0.031% (see Model 3). The effect size of *lpat* is 14% larger compared with *rcpx*. Accordingly, technological complexity ap-

Table 3.3: Panel regressions with different time lags ranging from 1 to 6 years

	Y: GDP per capita (in ln)					
	(1)	(2)	(3)	(4)	(5)	(6)
lgdp	0.809***	0.615***	0.445***	0.298***	0.161	0.049
	(0.041)	(0.066)	(0.084)	(0.089)	(0.085)	(0.058)
rcpx	0.008	0.014	0.028*	0.039**	0.045**	0.041**
	(0.005)	(0.009)	(0.012)	(0.014)	(0.015)	(0.014)
lpat	0.014**	0.027**	0.032**	0.028	0.012	−0.005
	(0.005)	(0.008)	(0.010)	(0.015)	(0.017)	(0.020)
lq	0.001	0.002*	0.002**	0.002	0.002	0.0005
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
div	0.00003	0.0001	0.0001	0.0001	0.0002	0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
htec-pat	0.001	0.002	0.002	0.002	0.001	0.0002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
lpopdens	−0.041	−0.075	−0.119	−0.159*	−0.193**	−0.222**
	(0.029)	(0.054)	(0.063)	(0.071)	(0.075)	(0.080)
share manufac	0.0003	0.002	0.003	0.003	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
hc	0.0004	0.0004	−0.0001	−0.002	−0.005***	−0.006***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
unemp	−0.002**	−0.003**	−0.003	−0.002	0.002	0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	2,375	2,209	2,043	1,877	1,711	1,545
R^2	0.748	0.521	0.328	0.180	0.110	0.134
Adjusted R^2	0.726	0.477	0.260	0.089	0.003	0.019

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
All independent variables are lagged from 1 to 6 years.
The corresponding time lags in years are represented by the column numbers.
Models are calculated using regional and time fixed effects.
Robust standard errors are clustered at the regional and time level.

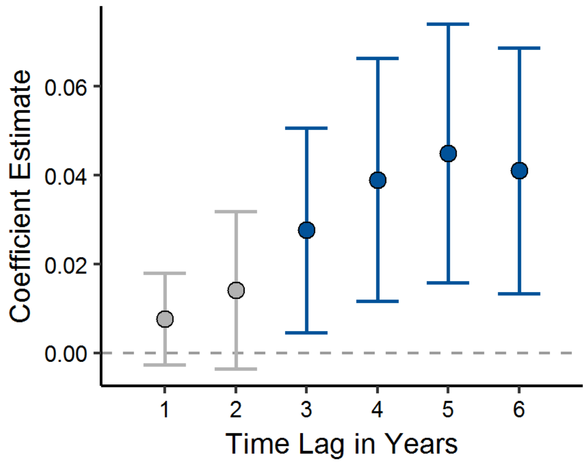

Figure 3.3: Coefficients of regional complexity and corresponding confidence intervals estimated in different time lags ranging from 1 to 6 years. The results correspond with the those of $rcpx$ obtained in Models 1 to 6 in Table 3.3.

pears to be a significant but rather weak factor for explaining regional variations in economic growth. It should be pointed out that when excluding $htec - pat$ from the estimations, the coefficient of regional complexity increases from 0.028 to 0.034. That means, it reaches about the same magnitude as the coefficient for the number of regional patents (see Table 3.4).

3.4.2 Robustness analysis

Considering a time lag of three years between complexity and economic growth yields the most robust results. We therefore use this set-up to further explore the robustness of our findings for alternative model specification. For instance, we change the set of control variables (Table 3.4). Data on population density, human capital, manufacturing share, and unemployment was not equally available for every region throughout all time periods. This explains the changing numbers of observations between the models. The results in general and the significantly positive coefficient of complexity in particular, are robust to changing sets of control variables and with respect to small variations in sample sizes.

As mentioned in Section 3.3.1, we expect the relationship between regional complexity and economic growth to be conditional on the chosen x_{th} percentile threshold. To test this, we re-calculate the full model including all variables and successively increase the percentile of patents considered as the basis for regional complexity. We shift the threshold to less complex activities and keep the time-lag constant at three years. Figure 3.4 displays the estimated coefficients of regional complexity for these different thresholds. Up to the 70_{th} percentile (i.e. top 70% of the regional complexity distribution,), $rcpx$ remains a significant and positive predictor of regional growth excluding the fifth percentile. Thresholds above the 70_{th} percentile result in insignificant coefficients. Shifting percentiles to the left discriminates against the ability to produce complex technologies by successively including simpler ones. As these are less relevant for complexity-based regional competitive advantage, the indicator loses its explanatory power for regional growth.

Important to note, our results provide evidence of a statistically positive asso-

Table 3.4: Panel regressions with constant time lags of 3 years and different sets of variables

	Y: GDP per capita (in ln)				
	Base	Controls	Full	Without lpat	Without htec-pat
	(1)	(2)	(3)	(4)	(5)
lgdp	0.552***	0.450***	0.445***	0.479***	0.440***
	(0.072)	(0.087)	(0.084)	(0.082)	(0.086)
rcpx	0.022*		0.028*	0.026*	0.034**
	(0.010)		(0.012)	(0.011)	(0.011)
lpat		0.031**	0.032**		0.034**
		(0.011)	(0.010)		(0.011)
lq		0.002*	0.002**	0.002*	0.002**
		(0.001)	(0.001)	(0.001)	(0.001)
div		0.0001	0.0001	0.0001	0.0001
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
htec-pat		0.003*	0.002	0.003	
		(0.002)	(0.002)	(0.002)	
lpopdens		−0.116	−0.119	−0.103	−0.119
		(0.066)	(0.063)	(0.060)	(0.063)
share manufac		0.003	0.003	0.003	0.003
		(0.002)	(0.002)	(0.002)	(0.002)
hc		−0.0002	−0.0001	−0.0001	0.00001
		(0.002)	(0.002)	(0.002)	(0.002)
unemp		−0.003	−0.003	−0.003	−0.003
		(0.002)	(0.002)	(0.002)	(0.002)
Observations	2,158	2,043	2,043	2,043	2,043
R^2	0.271	0.323	0.328	0.319	0.326
Adjusted R^2	0.205	0.255	0.260	0.251	0.259

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
All independent variables are lagged from 1 to 6 years.
Time lags in years are represented by the column numbers.
Models are calculated using regional and time fixed effects.
Robust standard errors are clustered at the regional and time level.

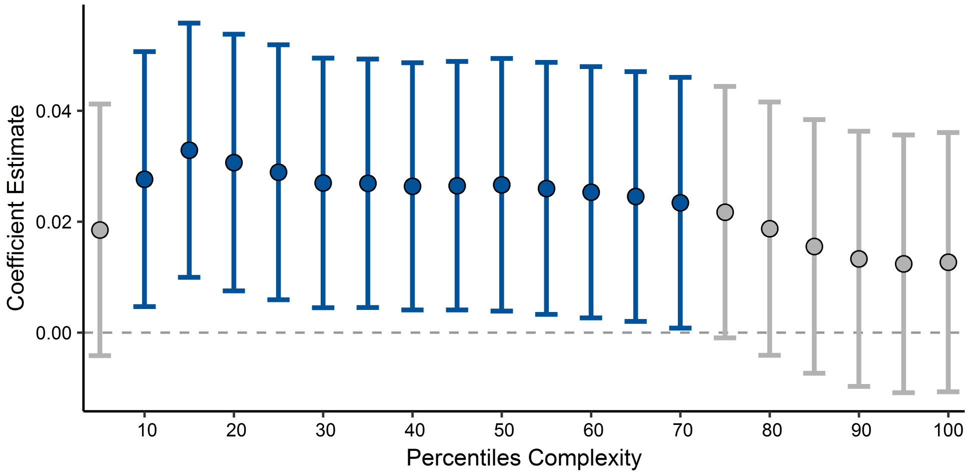

Figure 3.4: Effects of threshold choice on the relationship between regional complexity and economic growth.

ciation, i.e. correlation, between knowledge complexity and economic growth. In our estimations, we include complexity in different time lags, considered time and region fixed effects as well as a rich set of potential confounders. However, this might not be enough to control for all potential endogeneity issues. That is to say, higher rates of economic growth could, in principle, attract and enable regions to engage in complex activities. To get a first insight into this, we repeat the analysis but this time with complexity as the dependent variable. The results show that GDP is insignificant in all estimations suggesting that it is complexity, which drives growth and not vice versa (see Appendix 3.A). In the future, researchers may find ways to instrument regional complexity and thereby substantiate the statistical evidence such that causal inference becomes possible.

3.5 Conclusion

In this article, we analyzed the economic benefits of technological complexity for regional economic growth. Our results suggest that the relationship between technological complexity and economic growth in regions is positive. Regional differences in the capability to produce and exploit complex knowledge can therefore explain differences in the economic growth of regions. These findings therefore underpin the conceptualization of knowledge complexity as a competitive advantage and complements existing studies at the firm and country level (Kogut and Zander 1992; Hidalgo and Hausmann 2009).

However, there are a number of aspects that need to be taken into consideration when interpreting the results. We approximate technological complexity with information on patented inventions in Europe and calculate regional complexity scores for European NUTS 2 regions using the measure of structural diversity. Applying a dynamic panel-regression, we identify the relationship between regional complexity and economic growth to be positive and significant. Moreover, its strength is similar to that between patents and economic growth. However, our results suggest the existence of a statistically positive association, i.e. correlation, between knowledge complexity and economic growth. While we do not find any indication of a significant relationship the other way around, lacking an instrument for complexity, our study design does not allow for causal inference.

Due to data limitations, our analysis was restricted to a specific time horizon of 15 years. Considering the longevity of economic development, 15 years captures primarily short-term dynamics. It is likely that complexity unfolds its effects on economic growth over even longer time periods (Fink et al. 2017). In addition, our analysis was based on NUTS 2 regions in Europe. Although NUTS 2 regions are important entities for regional policy decisions, they represent administrative rather than functional units. Functional regions in terms of metropolitan areas or labor market regions are often used in empirical analyses to limit spatial biases, for instance, due to commuting patterns. Future research should replicate our study using functional regions to ensure the robustness of our results for different spatial units and scales.

Although our robustness checks underline the importance of technological complexity for regional economic growth, our findings only apply for technological complexity measured with Structural Diversity. Using two alternative complexity indicators, Hidalgo and Hausmann (2009) and Fleming and Sorenson (2001) yielded different results emphasizing a crucial problem in complexity research. Current empirical investigations use a large variety of complexity measures (Hidalgo and Hausmann 2009; Balland and Rigby 2017; Balland et al. 2018) impeding the comparison of empirical results across studies. The study of Broekel (2019) was a first attempt to compare different measures of technological complexity suggesting crucial differences between them, which explains why different complexity measures yield different results. Clearly, more methodological research is necessary to improve our understanding about the appropriate application of the complexity measures.

Nevertheless, our results fuel a number of important discussions. By showing that knowledge complexity has economic implications, our findings highlight the importance of building competitive advantage in complex activities. As knowledge complexity grows over time and demands more qualified individuals and more intensive collaboration (Powell et al. 1996; Pintea and Thompson 2007; Wuchty et al. 2007; Broekel 2019), places that attract qualified individuals and that are embedded in inter-regional knowledge networks will further benefit in this regard. As suggested by Balland et al. (2018), this is likely to amplify the geographic concentration of complex innovation activities even more and might be one of the reasons why urban agglomerations are increasingly becoming the epicenters of innovation. Although our results imply that knowledge complexity shows a tendency to concentrate in large metropolitan areas (e.g. Paris, Madrid, Berlin, Stockholm, Munich), complexity is restricted to urban agglomerations. This finding stays in contrast to the findings of Balland et al. (2018), who demonstrate that large cities have increasingly become hotspots of complex knowledge production in the USA. The cross-country case of European regions in this study suggests that complexity might have different implications for regions depending on the geographic context. Future research should investigate regional characteristics that influence regions' capability to produce complex technologies in more detail.

Knowledge complexity has also entered current policy debates (Balland et al. 2019). In this vein, knowledge complexity represents an ambivalent concept for policy makers. Nowadays, policy requires regions to invest in promising diversification strategies as evident in the smart specialization strategy of the EU to facilitate regional development (Foray et al. 2011). As argued by Balland et al. (2019), the combination of complexity and relatedness therefore provides a promising concept to derive such "smart" diversification strategies. Accordingly, building regional competitive advantage in new activities should be beneficial (i.e. complex activities) and feasible (i.e. related activities) for regions. Besides its positive impact on technological growth (Balland et al. 2019), our results indicate that such a policy might also directly facilitate economic growth.

However, it remains unknown how regions can exactly build competitive advantage in complex activities and if this strategy is suited and desirable for every region. Of similar relevance is the question if and how policy can influence the level of regional complexity. As the increasing complexity of knowledge production demands better qualified individuals and more collaboration, programs targeting these are promising candidates in this context. The EU *Framework Programme* (FP) might be important in this respect, as it is explicitly designed to facilitate knowledge and expertise exchange between regions. Their monetary incentives may help in overcoming barriers of knowledge diffusion that are particularly pronounced in case of complex knowledge (Balland and Rigby 2017).

3.A Reverse causality

As mentioned in the conclusion (see Section 3.5), our results reveal statistical associations and should not be interpreted as causal relationships. One potential source of endogeneity represents reverse causality. That is, complex technological activities might be attracted or made possible by higher growth rates of GDP. To assess this issue in more detail, we ran regressions with regional complexity *rcpx* as dependent variable and GDP per capita as the main explanatory variable. Table 3.5 reports the corresponding results. *lgdp* stays insignificant in all other models. Accordingly, these results suggest that lagged values of GDP do not predict regional complexity.

Table 3.5: Panel regressions with different time lags ranging from 1 to 6 years and rcpx as the dependent variable

	Y: Regional Complexity					
	(1)	(2)	(3)	(4)	(5)	(6)
lgdp	−0.057	−0.086	−0.047	−0.044	−0.030	−0.060
	(0.061)	(0.116)	(0.159)	(0.178)	(0.172)	(0.191)
rcpx	0.776***	0.469***	0.098	−0.056	−0.167***	−0.199***
	(0.020)	(0.048)	(0.064)	(0.065)	(0.050)	(0.043)
lpat	0.014	0.021	0.025	0.031	0.051	0.060
	(0.021)	(0.039)	(0.052)	(0.057)	(0.053)	(0.051)
lq	−0.002	−0.003	−0.003	−0.001	−0.002	−0.003
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
div	−0.0002	−0.0003	−0.0003	−0.0005	−0.001*	−0.002***
	(0.0002)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.001)
htec_pat	0.003	0.008	0.013*	0.010	0.004	−0.006
	(0.003)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
lpopdens	−0.077	−0.045	−0.085	−0.070	−0.029	0.262
	(0.060)	(0.132)	(0.165)	(0.165)	(0.196)	(0.174)
share_manufac	0.001	0.0004	0.001	−0.0001	−0.001	−0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
hc	0.0002	0.002	0.002	0.003	0.002	−0.002
	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
unemp	0.002	0.005*	0.009***	0.011***	0.011***	0.007*
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	2,375	2,209	2,043	1,877	1,711	1,545
R^2	0.640	0.284	0.067	0.045	0.064	0.090
Adjusted R^2	0.609	0.217	−0.027	−0.060	−0.050	−0.031

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
All independent variables are lagged from 1 to 6 years.
The corresponding time lags in years are represented by the column numbers.
Models are calculated using regional and time fixed effects.
Robust standard errors are clustered at the regional and time level.

3.B Alternative complexity measures

There is no standard way of calculating regional complexity. In the main part of our study, we use Structural Diversity as complexity indicator (see Section 3.3.1 for more details). As Broekel (2019) shows, this measure captures important properties of technological complexity: It grows over time, it is larger for collaborative and younger technologies, and it is positively associated to R&D and spatial concentration. Nevertheless, to explore the sensitivity of the results with respect to the choice of the applied complexity measure, we replicated the regressions using alternative measures. We follow Balland and Rigby (2017) and estimate regional complexity using the Economic Complexity Indicator (ECI) originally developed by Hidalgo and Hausmann (2009). The ECI builds on the idea that complex activities are geographically concentrated in the most diverse places. Although the measure has primarily been developed to assess a country's economic complexity based on its export portfolio, Balland and Rigby (2017) transfer the measure to calculate regional technological complexity. We estimate the regional complexity kci based on the distribution of 655 different technologies at the four-digit CPC level across all NUTS 2 regions. Table 3.6 reports the corresponding regression results. The measure is not significant in Models 1 to 5. In Model 6, kci is even significantly negative.

A second alternative is the NK measure introduced by Fleming and Sorenson (2001). Similarly to structural diversity, NK was constructed to fit patent data. NK approximates the interdependence of knowledge components within technologies by quantifying the co-occurrence frequency of technology classes on patents. For this, we use the most fine-grained level of the CPC. The resulting measure of interdependence can be interpreted as technological complexity. We calculate nk for every patent and, as for Structural Diversity, we aggregate the patent level indicator to the regional level by taking the average of nk in the x^{th} percentile of the regional complexity distribution. Again, we use the 10^{th} percentile to calculate regional complexity scores, as these provided robust results in case of structural diversity. As reported in Table 3.7, nk is insignificant in Models 2 to 6 and significantly negative in Model 1. Accordingly, our results remain conditional on the use of the measure of structural diversity.

3.C Sensitivity to sample selection

We measure regional complexity values using the average of xth percentile complexity distribution in each region. To ensure robust estimates, a considerable amount of regional number of patents is necessary. In the estimations reported in Section 3.4, we therefore remove observations, which have less than 50 patents in a given time period. Certainly, this threshold is arbitrary and conditions the general sample on existing patent activities. We therefore rerun our regression considering the full set of control variables using different patent thresholds. Figure 3.5 reports the sensitivity of our results to the chosen threshold. All independent variables are lagged three years. The results are robust for different patent thresholds between excluding the a threshold of ten, which seems to be an outlier.

Table 3.6: Panel regressions with different time lags ranging from 1 to 6 years and kci as the regional complexity indicator

	Y: GDP per capita (in ln)					
	(1)	(2)	(3)	(4)	(5)	(6)
lgdp	0.807***	0.617***	0.448***	0.300***	0.172*	0.062
	(0.042)	(0.066)	(0.084)	(0.090)	(0.086)	(0.059)
kci	0.0002	−0.0001	0.0002	0.0003	−0.001	−0.002*
	(0.0002)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)
lpat	0.014**	0.027**	0.030**	0.025	0.012	−0.002
	(0.005)	(0.009)	(0.011)	(0.015)	(0.018)	(0.020)
lq	0.001	0.001*	0.002*	0.002	0.002	0.0005
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
div	0.00002	0.0001	0.0001	0.0001	0.0002	0.0002
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
htec-pat	0.001	0.002	0.003*	0.004	0.003	0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
lpopdens	−0.041	−0.073	−0.116	−0.155*	−0.187*	−0.216**
	(0.030)	(0.056)	(0.067)	(0.076)	(0.079)	(0.082)
share manufac	0.0003	0.002	0.003	0.002	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
hc	0.0005	0.0003	−0.0002	−0.002	−0.006***	−0.006***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
unemp	−0.002**	−0.003*	−0.003	−0.001	0.002	0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	2,375	2,209	2,043	1,877	1,711	1,545
R^2	0.747	0.520	0.323	0.170	0.101	0.137
Adjusted R^2	0.726	0.475	0.255	0.078	−0.009	0.022

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
All independent variables are lagged from 1 to 6 years.
The corresponding time lags in years are represented by the column numbers.
Models are calculated using regional and time fixed effects.
Robust standard errors are clustered at the regional and time level.

Table 3.7: Panel regressions with different time lags ranging from 1 to 6 years and nk as the regional complexity indicator

	Y: GDP per capita (in ln)					
	(1)	(2)	(3)	(4)	(5)	(6)
lgdp	0.808***	0.616***	0.449***	0.302**	0.165	0.052
	(0.042)	(0.068)	(0.087)	(0.092)	(0.087)	(0.059)
nk	−0.005*	−0.005	−0.006	−0.009	−0.003	0.0004
	(0.002)	(0.004)	(0.006)	(0.007)	(0.007)	(0.006)
lpat	0.015**	0.028**	0.032**	0.028	0.010	−0.008
	(0.005)	(0.009)	(0.011)	(0.015)	(0.018)	(0.020)
lq	0.001	0.001*	0.002*	0.002	0.001	0.0002
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
div	0.00002	0.0001	0.0001	0.0001	0.0002	0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
htec-pat	0.001	0.002	0.003*	0.004	0.003	0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
lpopdens	−0.041	−0.074	−0.117	−0.157*	−0.188*	−0.216**
	(0.030)	(0.055)	(0.066)	(0.075)	(0.080)	(0.080)
share manufac	0.0003	0.002	0.003	0.003	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
hc	0.0004	0.0003	−0.0002	−0.002	−0.005***	−0.006***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
unemp	−0.002**	−0.003*	−0.003	−0.001	0.002	0.006*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	2,375	2,209	2,043	1,877	1,711	1,545
R^2	0.748	0.521	0.324	0.171	0.096	0.122
Adjusted R^2	0.726	0.476	0.255	0.080	−0.013	0.004

Note: **p<0.05; **p<0.01; ***p<0.001.*
All independent variables are lagged from 1 to 6 years.
The corresponding time lags in years are represented by the column numbers.
Models are calculated using regional and time fixed effects.
Robust standard errors are clustered at the regional and time level.

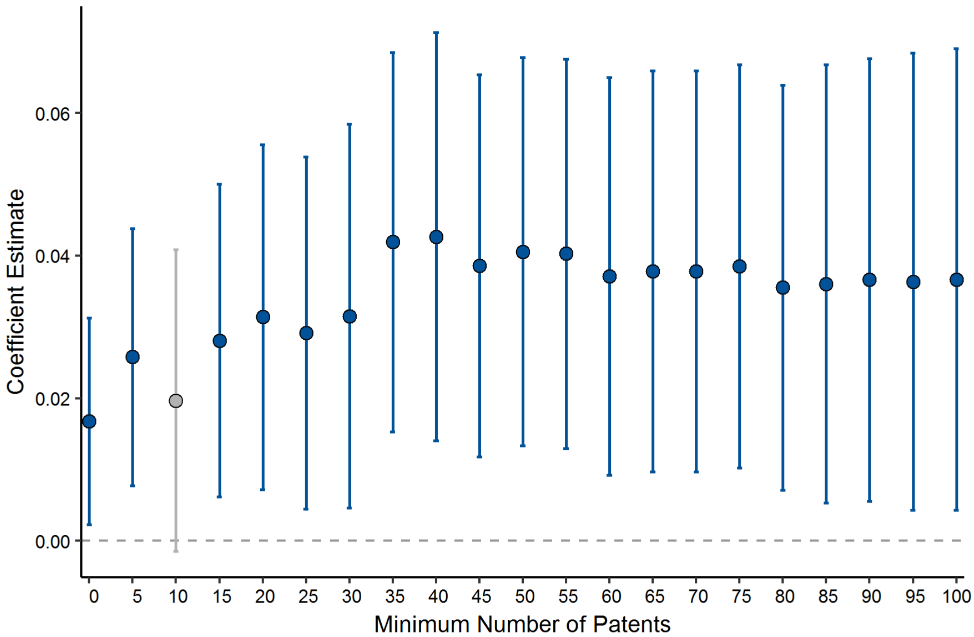

Figure 3.5: Sensitivity of estimations to the chosen patent threshold.

4 Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836 to 2010

Abstract: Cities are epicenters for invention. Scaling analyses have verified the productivity of cities and demonstrate a superlinear relationship between cities' population size and invention performance. However, little is known about what kinds of inventions correlate with city size. Is the productivity of cities only limited to invention quantity? I shift the focus on the quality of idea creation by investigating how cities influence the art of knowledge combinations. Atypical combinations introduce novel and unexpected linkages between knowledge domains. They express creativity in inventions and are particularly important for technological breakthroughs. My study of 174 years of invention history in metropolitan areas in the US reveals a superlinear scaling of atypical combinations with population size. The observed scaling grows over time indicating a geographic shift toward cities since the early twentieth century. The productivity of large cities is thus not only restricted to quantity but also includes quality in invention processes.

This chapter is a single authored paper published in 2019 as ***"Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836 to 2010"*** *in* ***Economic Geography, 95(4), 341-361****.*

4.1 Introduction

It is well known that invention activities are spatially concentrated (Audretsch and Feldman 1996) and primarily an urban phenomenon (Bettencourt et al. 2007b). Empirically, scaling analyses demonstrate the predominance of cities and reveal a superlinear scaling of inventors and inventions with respect to city size. That is, a disproportionate number of inventors and inventions concentrate in large cities indicating increasing returns to urbanization (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007b; Bettencourt et al. 2007a).

The productivity of cities rests on the idea of inventions being the outcome of knowledge combinations. This requires people to interact as knowledge is distributed across individuals, organizations, and institutions (Usher 1954; Nelson and Winter 1982; Utterback 1996; Hargadon 2003). Large cities provide more opportunities for knowledge combinations due to the concentration of critical requirements such as people, diversity, creativity, skills, infrastructure, and financial resources (Kuznets 1960; Jacobs 1969; Florida 2002; Glaeser 2011). The compactness of these factors in cities facilitates information flows among actors, stimulating knowledge combinations and in turn inventive outcomes (Bettencourt et al. 2007b). But how urban environments influence the art of knowledge combinations remains unexplored.

The large and diverse pool of existing knowledge provides large cities with more opportunities to explore atypical combinations than their nonurban counterparts. Atypical combinations introduce novel and unfamiliar linkages between less connected knowledge domains. They are an essential feature of creativity and a fundamental building block of high-impact science and technological breakthroughs (Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016). The exclusive focus on invention quantity in existing scaling analyses, however, overlooks such differences in quality (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007b; Bettencourt et al. 2007a).

In this article, I address the lacuna in scaling analysis by studying knowledge combinations with respect to city size and particularly ask the following: How does urban knowledge diversity relate to knowledge combinations? How do atypical knowledge combinations scale with city size? Are cities more explorative because their diversity allows them to be?

Empirically, I rely on scaling analysis to study how knowledge combinations relate to cities' population size and technological diversity. Following Uzzi et al. (2013), I distinguish between atypical and typical knowledge combinations based on z-score measures to proxy knowledge exploration and exploitation, respectively. This empirical approach relies on historic patent data from 1836 to 2010, which enables me to study the geography of knowledge combinations over 174 years of US invention history (Petralia et al. 2016). Studying almost two centuries allows me to reveal true long-term dynamics of knowledge combinations.

My main findings suggest that large cities increasingly concentrate atypical combinations and thus have become crucially important for knowledge exploration in the long run. I associate this development to the systematic relationship between knowledge diversity and city size. The knowledge diversity in large cities provides more opportunities for distinct knowledge combinations and to explore new combinations. Thus, large cities drive technological progress not only in quantitative but also in qualitative terms. The increasing concentration in large cities, however, reinforces a widening between urban centers and the rest of the country.

The article is organized as follows. The literature on the geography of invention and knowledge combinations is presented in the next section. I describe the data and empirical methods in the two sections following that. The results are presented and discussed in the penultimate section. The final section concludes the article.

4.2 Theoretical underpinnings

4.2.1 The geography of invention

The notion of the death of distance has culminated in Friedman's (2005) claim of the *flat world*. This stream of research argues that technological change erodes the obstacles (e.g., physical barriers, travel time, sociocultural differences) that once limited the exchange of labor, goods, and knowledge (O'Brien 1992; Castells 1996; Cairncross 1997). In particular, innovation in telecommunication and computing technologies unfasten the mobility of production factors and detach economic activity from its territorial and socioeconomic context (O'Brien 1992; Castells 1996). Accordingly, technological progress spreads economic activities to every part of the world and enhances the global diffusion of knowledge. In such a scenario, location becomes less relevant, reducing the geographic concentration of economic activities of all kinds and eventually diminishing spatial inequalities over time.

Friedman's thesis has revitalized an active debate about the role of geography for economic activities (Christopherson et al. 2008; Florida et al. 2008b; Rodríguez-Pose and Crescenzi 2008). The spatial distribution of the world economy doubts a flattening of the world, as economic activities and wealth are increasingly concentrated in space. More precisely, overwhelming empirical evidence is pointing in the exact opposite direction to what was proclaimed by Friedman and others. Scott (1993) and Saxenian (1994), for example, analyzed the prevailing concentration of certain industries (i.e. semiconductors and aerospace) in California and Massachusetts showing that geographic clustering is a common phenomenon. Most paradoxically, the digital industry - believed to be the driver that flattens the world - is itself highly clustered (Zook 2000). Beyond single case studies, it has been shown that economic activities, more generally, concentrate in specific locations and that the concentration tends to grow over time (Marshall 1890; Hall and Markusen 1985; Ellison and Glaeser 1999; Dumais et al. 2002; Ellison et al. 2010). Geography therefore represents an important determinant in order to understand economic development and inequalities between cities.

Of all economic activities, the tendency toward spatial concentration is even stronger for invention activities. Spatial patterns of invention have been the subject in a growing body of empirical studies, showing that invention activities are not equally distributed across regions, but rather occur highly concentrated in space (Feldman 1994; O'hUallichain 1999; Acs et al. 2002; Dumais et al. 2002; Sonn and Storper 2008; Feldman and Kogler 2010; Castaldi and Los 2017). Most striking, the spatial concentration is relatively persistent and, more importantly, increases over time (Varga 1999; Co 2002; O'hUallichain and Leslie 2005; Sonn and Park 2011) challenging the death of distance argument. Since knowledge is a crucial source for economic growth (Lucas 1988; Romer 1990), regions more capable of creating new knowledge possess an economic advantage over less inventive regions (Feldman and

Florida 1994).

The observed concentration is systematic, since a large body of empirical research suggests invention is primarily an urban phenomenon. In particular, the inventive performance of metropolitan areas grows disproportionately with population size, indicating increasing returns to urbanization (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007b; Bettencourt et al. 2007a). These findings indicate a spatial concentration of invention activities in larger metropolitan areas.Kuznets (1960) elaborated how a larger population size is associated with a greater productivity of new knowledge. However, not explicitly referring to Kuznets' work, more recent contributions rely on his thoughts about cities as centers for knowledge creation. Bettencourt et al. (2007a) adopted a theoretical and methodological framework called scaling, which stems from biology (Schmidt-Nielsen 1984; West 1997) and quantifies the relation between size (e.g., body size, population size) and aggregated outcomes (e.g., metabolism rate, wealth, and inventions). The dependence of invention activity Y on population size N can be expressed as a scaling law of the following form (Bettencourt et al. 2007b):

$$Y = Y_0 N^{\beta} \tag{4.1}$$

where β is the scaling exponent, which falls into three broad categories revealing three different scaling mechanisms. First, β smaller than one expresses a sublinear relationship implying economies of scale. Second, a linear relationship is evident if β equals one. Third, if β is greater than one, the relation between population size and inventive performance of a city is superlinear, revealing increasing returns to urbanization, as reported, for example, in Bettencourt et al. (2007b). That is, if a city doubles its population size, it increases its inventive output more than twice as much. The empirically confirmed superlinear scaling of invention activities (Carlino et al. 2007; Arbesman et al. 2009) reveals the dominant role of large metropolitan areas for invention, at least, in quantitative terms.

But why are cities so remarkably productive with respect to inventions? The literature on urban scaling attributes the productivity to two major interdependent factors: population size and knowledge diversity (Kuznets 1960; Jacobs 1969; Bettencourt et al. 2007b). Highly skilled and creative minds increasingly concentrate in urban areas stimulating creative processes such as invention activities. Inventors in cities thus have access to a larger and also more diverse pool of knowledge than inventors living outside of cities. This is crucially important, since inventions often build on the combination of existing knowledge (Usher 1954; Nelson and Winter 1982; Utterback 1996) and thus on interpersonal interactions that are facilitated by geographic proximity (Liben-Nowell et al. 2005). Urban environments provide more opportunities for knowledge exchange between actors and thus facilitate knowledge combinations (Bettencourt et al. 2007b). Nevertheless, existing scaling analyses do not ask how cities influence knowledge combinations. Therefore, they disregard qualitative differences of knowledge combinations and treat inventions as a homogeneous quantity (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007b; Bettencourt et al. 2007a; Carlino et al. 2007; Sonn and Park 2011). By analyzing and evaluating the novelty of knowledge combinations, I particularly shift the focus from quantity to quality and extend existing approaches. In the next section, I argue that knowledge combinations are heterogeneous and that cities concentrate essential factors, which affect knowledge combinations in their quality.

4.2.2 Geography of knowledge combinations

Knowledge combinations represent an important mechanism of idea creation (Usher 1954; Nelson and Winter 1982; Utterback 1996; Hargadon 2003; Arthur 2009). Inventions consist of multiple components that are put together in a novel way to fulfill a specific purpose. The components themselves are rarely completely new; rather, they typically represent existing pieces of knowledge (Arthur 2009). Crucially, the art of creatively combining different knowledge domains is one important source for different degrees of novelty across inventions (Ahuja and Lampert 2001).

Exploration and exploitation are two important search processes in research and development (R&D), which differ significantly in their underlying combinatorial characteristics (March 1991). Exploitation thereby refers to the reuse and refinement of existing combinations, whereas exploration describes the search for and development of new combinations. Exploring new combinations implies higher costs and risks than reusing proven combinations. Due to these characteristics, combinations identifying exploitation occur more frequently and hence represent *typical* combinations. In contrast, combinations resulting from exploration are rather rare and *atypical* among observed combinations. In line with previous studies, I rely on the terminology of atypical (typical) combinations as proxies for exploration (exploitation) (Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016). Combinatorial characteristics are a strong predictor for the impact of inventions. It is the combination of previously disconnected components, i.e. exploration, that leads to novel ideas and high impact results (Fleming 2001; Dahlin and Behrens 2005; Schoenmakers and Duysters 2010; Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016; Verhoeven et al. 2016).

Consequently, spatial variance of exploration and exploitation will affect the regional outcome of invention quality. That is, places of knowledge exploration (i.e., regions that are more capable of combining knowledge in an explorative fashion) are more likely to produce atypical inventions. However, no study exists that seeks to identify such interregional variations. The literature on knowledge combinations is silent about possible geographic patterns and offers little insight into the geography of invention. Although the combinatorial character of knowledge is embedded in contemporary concepts of economic geography, for example, related variety (Frenken et al. 2007), differences between regions have not been taken into account to explain spatial inequalities. I therefore shift the focus explicitly to knowledge combinations as the research object in order to disentangle the geography of invention in qualitative terms. But why should places differ regarding the intensity of exploration and exploitation?

The literature on urbanization externalities suggests regional diversity playing a major role for knowledge combinations. The argument harkens back to Jacobs (1969), who described the benefits of large and diverse cities for socioeconomic interactions. Firms, for example, benefit from a cross-fertilization of ideas between industries, rather than being stuck in industry-internal thought patterns. Hence, diversity increases the likelihood of knowledge spillovers between heterogeneous actors (Bettencourt et al. 2008; Arbesman et al. 2009). Regions with large (knowledge) diversity, in particular, provide more opportunities for knowledge combinations than less diverse cities where such diversity is missing. Being located in diverse environments allows drawing from larger pools of distinct knowledge pieces (e.g., technologies, sectors, industries), which in turn increases the opportunities for atypical combinations.

The geographic nature of knowledge spillovers reinforces the importance of regional diversity for combination processes. An exhaustive literature demonstrates that knowledge, in general, does not travel easily over long geographic distances. More precisely, knowledge tends to stay in the same region where it was once created, although the effect diminishes as technologies mature (Jaffe 1989; Jaffe et al. 1993; Anselin et al. 1997; Varga 2000). It is often argued that codified knowledge travels more easily than tacit knowledge, while tacit knowledge is more likely to adhere to specific places (Hippel 1994; Maskell and Malmberg 1999; Gertler 2003). Yet, it is difficult to assess the difference between codified and tacit knowledge empirically. Balland and Rigby (2017) disentangled the two knowledge types by arguing that tacitness can at least partially be captured by the complexity of what is known. Their findings suggested that knowledge complexity limits the geographic distance of knowledge spillovers even more. However, in most instances, codified and tacit knowledge are complements, and, hence, the geographic stickiness of the latter will also reduce the mobility of the former (Cowan and Foray 1997). Accordingly, the local knowledge base represents a crucial determinant of regional knowledge combination processes. Consequently, more diverse cities have access to a larger variety of local knowledge, enabling them to realize more distinct combinations than less diverse cities.

Diversity is critically linked to urbanization. Larger cities, usually, host more different industries than smaller towns. Recently, Youn et al. (2016) analyzed how diversity of business activities relates to city size in US metropolitan areas. They found a linear relationship between city size and business diversity. In an earlier work, Mori et al. (2008) observed a similar relationship between industrial activities and the population size of metropolitan areas in Japan. Clearly, this pattern is not limited to a single nation. The theoretical logic behind the observed linear scaling of population and diversity rests on the notion of the urban hierarchy (Christaller 1933). The central idea is that activities found in the largest cities include those located in the smallest towns, but not vice versa. Larger cities (i.e. central places) provide more sophisticated products, services, and technologies for their less populated surroundings. New York, for example, has a larger potential to explore new knowledge combinations than Branson, Missouri.

Regional diversity, however, is not sufficient to actually explore new combinations. It rather indicates the potential that could be explored. Importantly, exploration requires certain skills and actors to use the given potential, which are not equally distributed across space (Glaeser and Maré 2001; Florida 2002; Bettencourt et al. 2007a; Combes et al. 2008; Bacolod et al. 2009; Storper and Scott 2008; Lee et al. 2016). Spatial wage disparities (i.e. the urban wage premium) indicate that people living in larger cities earn more then their nonurban counterparts (Weber 1899; Glaeser and Maré 2001). Combes et al. (2008)) attributed this observation to the spatial sorting of skills. Up to half of the wage disparities is explained by differences of the local workforce composition. Relatedly,Bettencourt et al. (2007a) observed a superlinear scaling for both creative employment, as defined by Florida (2002), and R&D employment. That is, individuals with better qualifications for exploring and exploiting knowledge combinations tend to concentrate in larger and more densely populated cities. It follows that cities not only have the larger potential for atypical knowledge combinations but also have a higher capacity (due to the urban concentration of the skills and talents needed for this task) to exploit these potentials. Based on this, I expect atypical combinations to concentrate in large cities. I hypothesize this relationship as follows:

Atypical and typical knowledge combinations scale superlinearly with city size. However, atypical combinations scale to a larger extent with city size than typical combinations.

4.3 Data

In line with previous studies, I rely on patent data to analyze invention activities as results of combinatorial search processes (Fleming 2001; Dahlin and Behrens 2005; Schoenmakers and Duysters 2010; Arts and Veugelers 2015; Kim et al. 2016). Patent data has some peculiarities, which affect the results. Patent activities are not equally distributed across firms, technologies, and sectors. Most importantly, the tendency to patent an invention is biased in favor of manufacturing activities (Griliches 1990). Thus, patents underestimate the inventive outcome in less manufacturing-intensive regions. Eventually, the decision to patent rests on strategic judgment. Not every invention results in a patent for various reasons, for example, information disclosure, the ease of circumventing patent claims, and application costs (Cohen et al. 2000). Acs et al. (2002), however, found that patents are a reliable indicator for measuring invention activities at the regional level.

I draw the patent data from three different data sources. The first source is *HistPat*, which was recently generated by Petralia et al. (2016) and is publicly available. This data set contains geographic information on patents from the *United States Patent and Trademark Office* (USPTO) ranging from 1836 to 1975. I complement HistPat by using the data set from Li et al. (2014), which covered the years 1975 to 2010 and contained geographic information as well. Third, I used the *Master Classification File* of the USPTO Bulk Storage System, which provides information on technology classes for the whole time span. The data sets were matched by using patent numbers as unique identifiers. With this data in hand, I was able to analyze the geography of knowledge combinations for granted US patents over the last 174 years.

Patent data reveal how knowledge is combined, as each invention is classified into at least one technology class. In many cases, one single invention is grouped into more than one class. This information has been used to study the knowledge combination process (Fleming 2001; Dahlin and Behrens 2005; Schoenmakers and Duysters 2010; Kim et al. 2016). The underlying classification scheme is the *Cooperative Patent Classification* (CPC). The CPC has been established to harmonize individual classification systems between the USPTO and the European Patent Office. Using the CPC thus allows for cross-country comparison of empirical results.

Scholars have long debated how to define a city theoretically and for the purposes of quantitative research (Arcaute et al. 2014; Louf and Barthelemy 2014). HistPat locates patents not to American cities, but rather to counties. This signifier of invention location does not suffice. The county level represents a narrow administrative boundary; it does not take into account interregional dependencies crossing county boundaries. Focusing on county boundaries can therefore lead to spatial bias, since inventors living in one region could potentially generate their invention in neighboring ones. To capture such interregional interdependencies and to minimize spatial bias, most geographic analyses use functional units (Bettencourt et al. 2007a; Youn et al. 2016). In this study, I use 171 *Combined Statistical Areas* (CSA), which is the largest unit of the Metropolitan Statistical Areas in the United States.

I gathered population data of US counties back to the first documented entries, which were in New York County in 1656. I used *Wikipedia* as a data source to obtain the information for every US county, then aggregated the population size to the CSA level[1]. The population data are only available for ten-year periods. However, these data allow for constructing a panel covering a long time period.

4.4 Methods

4.4.1 Z-scores approach

Following Uzzi et al. (2013) and Kim et al. (2016), I investigate the combinatorial nature of invention by applying z-score measures at the subclass level of the CPC[2]. Teece et al. (1994) introduced z-scores for estimating the relatedness between industries. Z-scores compare the observed combinations of technology classes to what would be expected under the assumption that combinations are random. More formally, the z-score is expressed as follows:

$$z_{i,j} = \frac{o_{i,j} - u_{i,j}}{\sigma_{i,j}} \tag{4.2}$$

where $o_{i,j}$, j is the empirically observed co-occurrence count of technology classes i and j. The expected co-occurrence and standard deviation are $u_{i,j}$ and $\sigma_{i,j}$, respectively. A high value for $o_{i,j}$ can be driven by the combination of i and j or by a high number of patents n for both classes. If n_i and n_j are large, one can expect to observe a fair amount of combinations, even if there is little synergy between them. By contrast, a small n_i and n_j result in a relative small number of combinations. To control for this effect, I compare the observed co-occurrence $o_{i,j}$ to what can be expected given n_i and n_j, if knowledge combinations were random (Teece et al. 1994).

The expected co-occurrence, $u_{i,j}$, represents a hypergeometric distribution and is thus given by the product of the number of patents in both technology classes n_i and n_j divided by the total number of patents N:

$$u_{i,j} = \frac{n_i n_j}{N} \tag{4.3}$$

and its standard deviation $\sigma_{i,j}$ is given by:

$$\sigma^2_{i,j} = u_{i,j}\left(1 - \frac{n_i}{N}\right)\left(\frac{N - n_j}{N - 1}\right) \tag{4.4}$$

[1]I used Wikipedia because it offers data for the entire 174 years of observation. I compared the population size for the most recent years with official data sources, such as *www.census.gov*, finding no differences.

[2]As a robustness check, I also used the CPC class level (three digits) showing that results are independent of the technological resolution (see Appendix 4.B.

If i and j were combined more often than expected, equation 4.2 produces a positive value. A positive z-score indicates a typical class combination and, relatedly, an invention that recombines known elements. Conversely, if the two classes i and j are rarely paired together relative to their expected occurrence, equation 4.2 produces a negative number. This indicates an atypical knowledge combination and, relatedly, an innovative invention.

I can only consider patents that were assigned to at least two technology classes when discussing knowledge combinations because z-scores measure the typicality of combinations between technology pairs. Single class patents shed no light on the combination process. This provides a total sample of 1,706,499 patents granted to inventors living in US metropolitan areas.

4.4.2 Cumulative knowledge combinations

Knowledge accumulates over time, giving rise to the emergence of technological trajectories (Dosi 1982; Nelson and Winter 1982). However, the characteristics of knowledge combinations can vary over time. An atypical combination, for example, can diffuse in the knowledge space, if it is repeated in subsequent inventions. Atypical then becomes typical, under the right circumstances and on a long enough time line. Conversely, a certain combination can lose its typicality over time if it is superseded by newer knowledge combinations. To capture this temporal evolution, I rely on an approach similar to the one applied by Kim et al. (2016). For example, if t is 1950, I consider all patents from the beginning of the observation in 1836 to 1950 to calculate $o_{i,j}$[3]. This approach takes into account the cumulative nature of knowledge production and allows the z-scores to evolve over time.

4.4.3 Scaling analysis

Urban scaling analyses express the dependency of a certain quantity Y (e.g., air pollution, bike thefts, inventions) on cities' population size N as a power-law relation (Bettencourt et al. 2007a):

$$Y = Y_0 N^{\beta} \tag{4.5}$$

or its linear transformation

$$\log(Y) = \log(Y_0) + \beta \log(N) \tag{4.6}$$

with Y_0 representing a normalization constant. I estimated β by using an ordinary least squares estimation. Thus, β can be interpreted as the exponent of population size N, with β falling into one of three categories: $\beta = 1$ (linear), $\beta < 1$ (sublinear), and $\beta > 1$ (superlinear) (see Section 4.2.1). I use 95 percent confidence intervals to test the significance of the exponents falling into one of the three categories. A superlinear relation, for example, is often associated with increasing returns to urbanization. When N doubles in size, Y increases more than twice as much.

[3] I also applied a twenty-year rolling window approach in which the history of knowledge combination washes out over time; see Figure 4.6 in Appendix 4.A. The cumulative and the rolling window approach correlated on average at a high level, with $0.9 < R < 1$.

4.5 Results

In a first step, I analyzed the scaling relation between technological diversity and the population size of cities. One simple measure of diversity is the number of distinct technologies D in a city. A given technology class belongs to the local portfolio if at least one corresponding patent is filed. The hierarchical nature of the CPC allows the number of distinct technologies at a more granular level to be analyzed. Youn et al. (2016) showed that the resolution by which technologies are considered *distinct* clearly affects the results. I control for this observation by using three different levels of technological resolution as defined by the CPC: *subclasses* ($D_{max} = 654$), *groups* ($D_{max} = 10,154$), and *subgroups* ($D_{max} = 218,570$).

Figure 4.1 illustrates D as a function of population size at different levels of technological resolution for the whole time span. D is normalized by D_{max} to ensure comparability between resolution levels (Figure 4.1, panel A). Diversity at the subclass (red dots) and group level (green dots) strongly follows a logarithmic law. The corresponding exponents $\beta_{subclass} = 0.22 \pm 0.02$ and $\beta_{group} = 0.56 \pm 0.03$ imply that diversity relates sublinearly to population size as $\beta < 1$. This finding suggests that larger cities are more diverse but that diversity does not increase disproportionately with city size.

When using the most fine-grained level of distinction, subgroups (blue dots in panels A and D of Figure 4.1), the exponent changes to $\beta_{subgroup} = 0.95 \pm 0.04$. The corresponding 95 percent confidence interval ranges from 0.86 to 1.03. Hence, the range includes $\beta \approx 1$, which corresponds to a linear relation of diversity and city size. The result is similar to that of Youn et al. (2016), who observed an exponent of $\beta = 0.98 \pm 0.02$ for the relation between diversity of business activities and city size. Accordingly, technological diversity is also strongly related to city size in a linear fashion. This relationship, however, is very sensitive to the level of technological resolution.

Next, I analyzed how the US cities' local diversity relates to knowledge combinations. As was explained in "Theoretical Underpinnings," a proportional increase of diversity shifts knowledge combination opportunities (distinct knowledge combinations). The CPC distinguishes 654 different subclasses (D_{max}), enabling 213,531 distinct class combinations. Using subclasses is sufficient to study knowledge combinations, as cities realize only a small fraction of what is theoretically feasible. The average share of realized combinations across all cities is 3 percent. The most diverse city is New York, with patents in 630 different technologies between 1990 and 2010. New York's knowledge base allows for 198,135 distinct combinations, of which 17,182 were realized (9 percent). Local diversity can be seen as the endogenous potential for knowledge combinations.

Figure 4.2 plots the relationship between diversity and distinct class combinations at the city level at four different time periods. In 1850, the relationship was almost linear. Over the years, the curve became steeper, as cities' technology portfolios grew more diverse.

I investigated the relationship between diversity and knowledge combinations once more by employing the scaling approach. Figure 4.3 visualizes the development of the scaling exponent, β, over time. The scaling exponent of diversity is larger than one, indicating an overproportionate increase of distinct class combinations with cities' diversity.

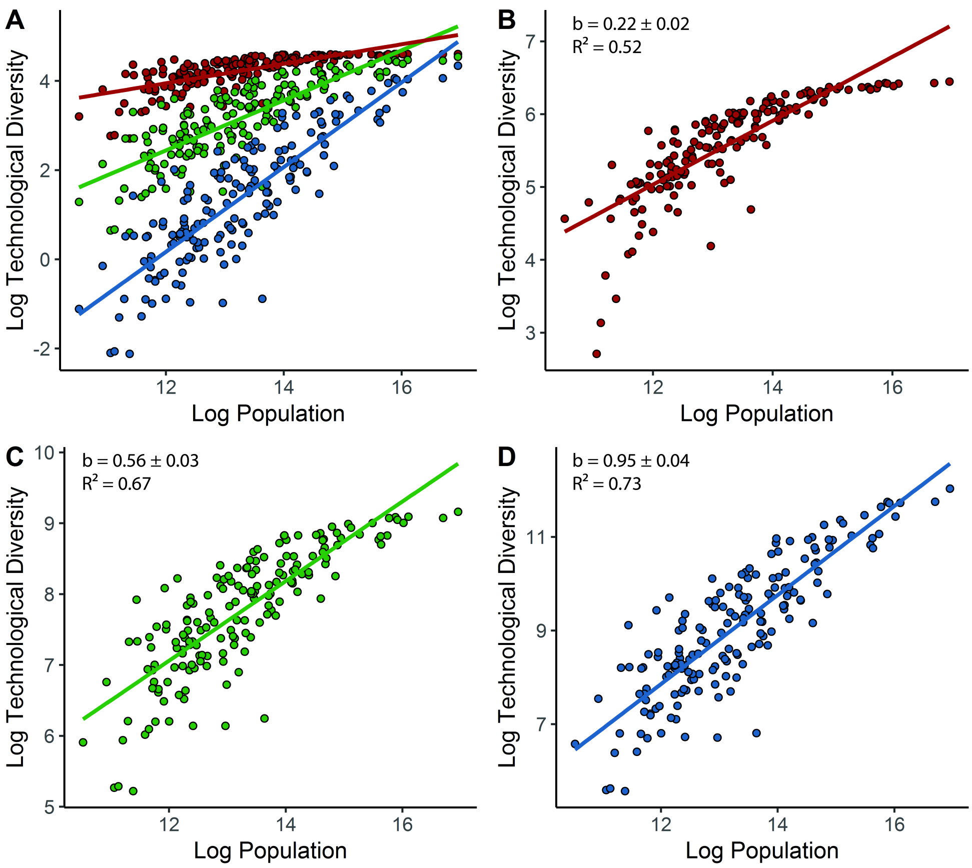

Figure 4.1: Technological diversity as a function of population size at three different levels of technological resolution. Diversity is normalized in **A** by D_{max} for comparability reasons. Scaling relations between population and technological diversity **B** at the subclass level ($D_{max} = 654$), **C** at the group level ($D_{max} = 10,154$), and **D** at the subgroup level ($D_{max} = 218,570$).

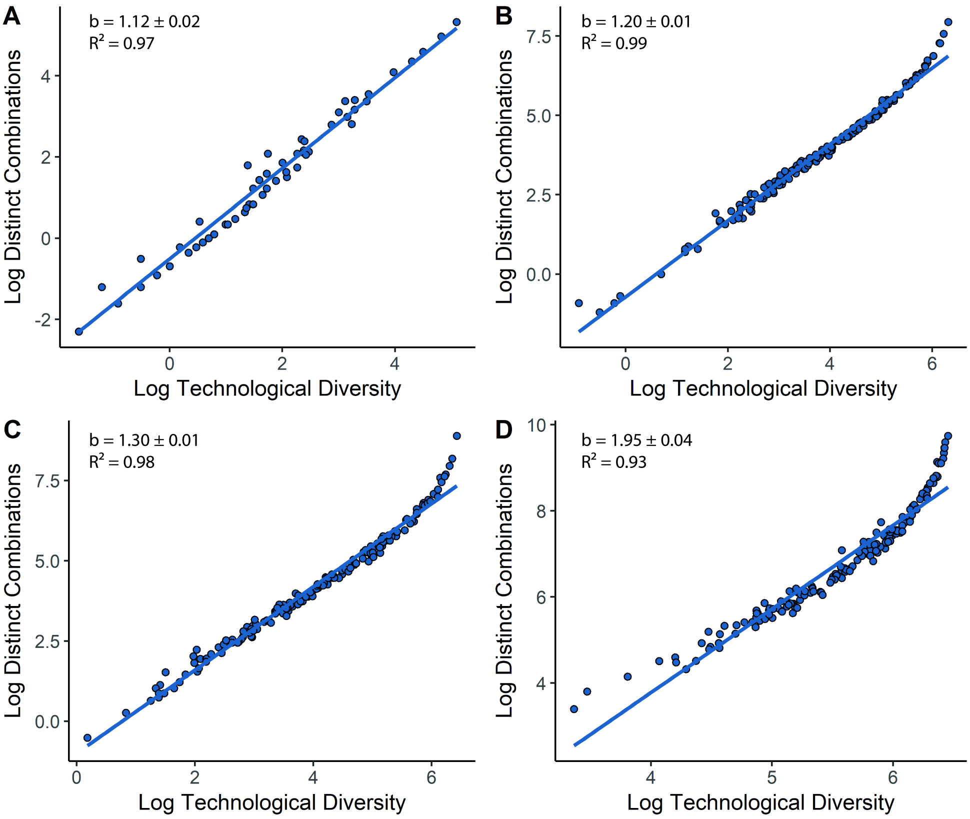

Figure 4.2: Scaling relationship between technological diversity and the total number of distinct class combinations **A** in 1850, **B** in 1900, **C** in 1950, and **D** in 2010 in US metropolitan areas.

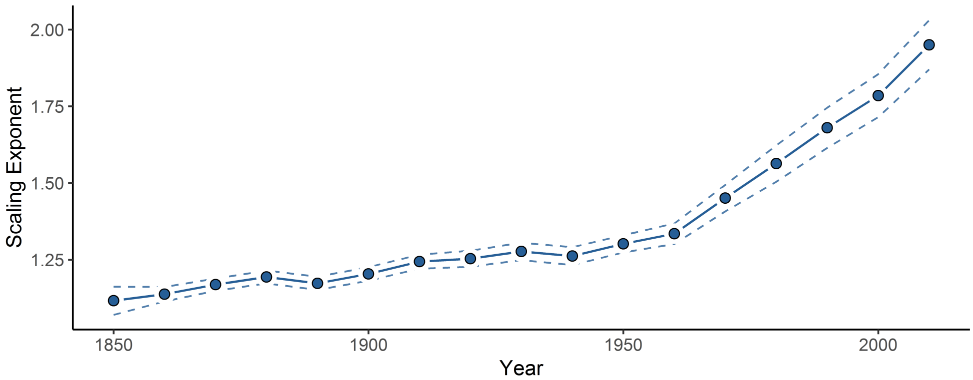

Figure 4.3: Scaling exponent of diversity with respect to the number of distinct combinations over time. Dashed lines indicate the 95 percent confidence interval.

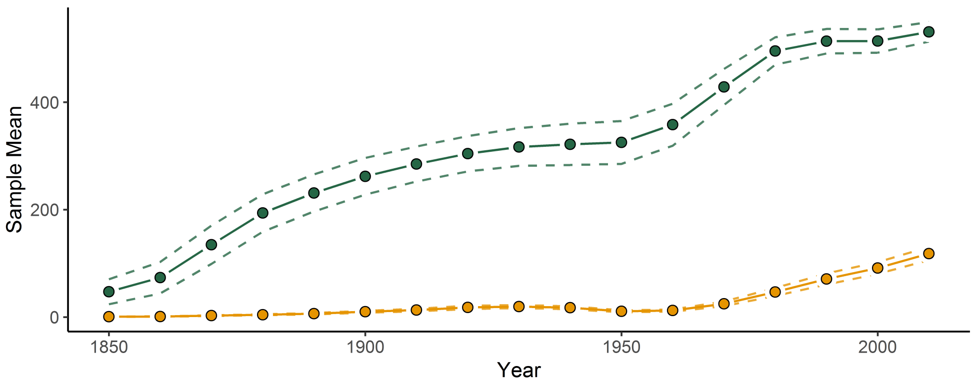

Figure 4.4: Average number of technologies in the most diversified (green line) and least diversified cities (orange line). Dashed lines indicate the 95 percent confidence intervals.

In addition, Figure 4.3 shows that scaling increases over time. I interpret this finding as evidence for growing disparities between the least and the most diverse cities. To understand this finding in greater depth, I divided the sample into two subsamples based on each city's diversity in each year. The most diverse cities belong to the upper quartile, and the least diverse cities to the lower quartile. I compared both groups' sample means and corresponding 95 percent confidence intervals based on the one sample t-test. Figure 4.4 visualizes the result. The difference between both sample means is significant and clearly increases over time, emphasizing the increasing disparity between the groups. This disparity is largely driven by the increasing diversity of the most diverse cities, such as New York, Greater Boston, Los Angeles, Chicago, and the Bay Area (in San Francisco).

In a further analysis, I examined the correlation between knowledge combinations typicality and population size. My hypothesis claims that the resources needed for expanding the set of knowledge combinations are especially concentrated in large cities, such that larger cities have more atypical knowledge combinations (see Section 4.2.2).

Figure 4.5 illustrates β of atypical (red line) and typical (blue line) combinations in relation to population size over time. Typical combinations serve as the baseline scaling of knowledge combinations against which atypical combinations are tested

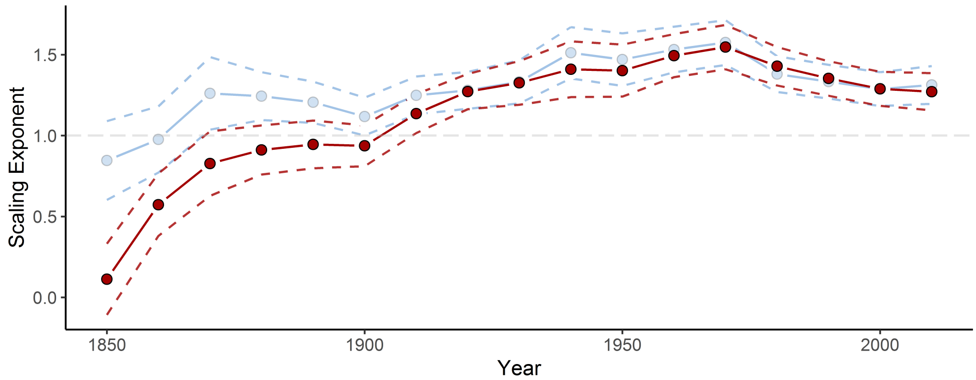

Figure 4.5: Scaling exponent of population size over time for atypical (red line) and typical combinations (blue line). Dashed lines indicate the 95 percent confidence interval.

and put into relation. The scaling exponent of atypical combinations has increased over the last 174 years. Until 1900, the exponent was smaller than one; this indicates there were no particular benefits of city size at that time. Since then, atypical combinations appear to become an urban phenomenon, with $\beta > 1$ and a maximum of 1.54 ± 0.07 in 1970. Between 1970 and 2010, the scaling exponent, however, has slightly decreased.

Interestingly, urbanization is not just favorable for atypical but also for typical combinations. For most years since 1836, the scaling exponent of typical combinations has been greater than one and larger than the exponent of atypical combinations. That is, cities have been more successful at knowledge exploitation than exploration. In the last decade, both exponents have converged to almost the same value. Based on this finding, I may only partially confirm my hypothesis: both atypical and typical combinations scale superlinearly with city size, but atypical combinations do not scale to a larger extent than typical combinations.

4.6 Conclusion

The increasing availability of large and historic data sets opens new possibilities for empirical research. This study is among the first analyzing the geography of invention over almost two centuries. My analysis of American invention history reveals that knowledge exploration clearly concentrates in large cities. That is, atypical combinations scale superlinearly with cities' population size. The scaling exponent significantly increased over the last 174 years, which suggests that large cities drive technological progress not only in quantitative but also in qualitative terms. This finding challenges the prominent death of distance thesis in almost all regards (Friedman 2005).

I attribute the growing importance to the opportunities given in large cities. In particular, knowledge diversity in large cities provides opportunities for knowledge combinations not found in smaller and less diverse towns. Beyond diversity, larger cities also concentrate the skills to exploit the given diversity. Inventors in large cities realize a disproportionate number of distinct knowledge combinations, which also affects the exploration of new combinations. Given the cumulative nature of knowledge, wealth, innovation, and human skill, my results suggest a self-reinforcing process that favors metropolitan centers for knowledge creation. Thus, knowledge creation plays a major role for creating and maintaining spatial inequalities.

Increasing spatial inequalities have profound implications for regional development and policy making. Inequalities unfold in the form of invention activities, as one crucial economic activity that transforms our economy and society. The benefits of knowledge creation in large cities are not shared by all regions and reinforces a widening divergence between large cities - as centers of knowledge exploration - and smaller towns. Given the importance of geography for knowledge generation, it is unlikely that spatial concentration of invention activities will stop. Earlier research, moreover, observes a decreasing productivity of R&D and highlights that more resources and capabilities are necessary to yield useful R&D outcomes (Lanjouw and Schankerman 2004; Wuchty et al. 2007; Jones et al. 2008). Large cities provide the required resources and capabilities in close geographic proximity. Smaller towns lack the requirements to compete, get disconnected, and fall behind. It should be, furthermore, in the interest of policy makers that all places benefit from urban externalities. That is, policy has to consider how to distribute the novelty created in the centers down the urban hierarchy to smaller towns and lagging regions.

However, much research remains to be done. Why did it take longer for atypical combinations to scale that strongly with city size? Has this process stopped, or will it continue? Moreover, atypical knowledge combinations do not automatically imply a high technological impact or economic value. Thus, it remains unclear precisely how (a)typical combinations relate to the economic performance of cities and how they explain local stories of success and failure.

4.A Moving window vs. cumulative approach

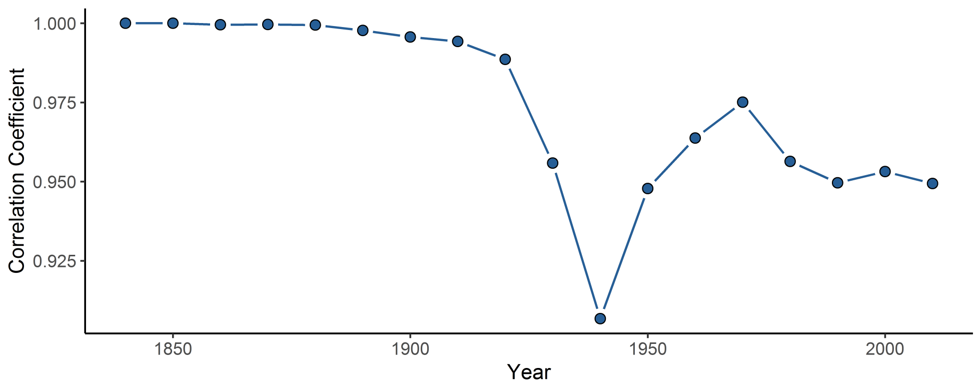

Figure 4.6: Correlation coefficient between z-scores calculated in a rolling window (twenty years) and a cumulative approach (see Section 4.5).

4.B Robustness analysis

To check if the results described are not affected by the choice to use the four-digit CPC level (CPC4), I repeated the analysis by using a different level of technological aggregation, that is, three-digit CPC (CPC3). The CPC3 distinguishes between 127 different technologies. The figures clearly show that my results are relatively robust using the CPC3. As the CPC4 reveals more technological details than CPC3, I decided to use the CPC4 as the main level for my analysis.

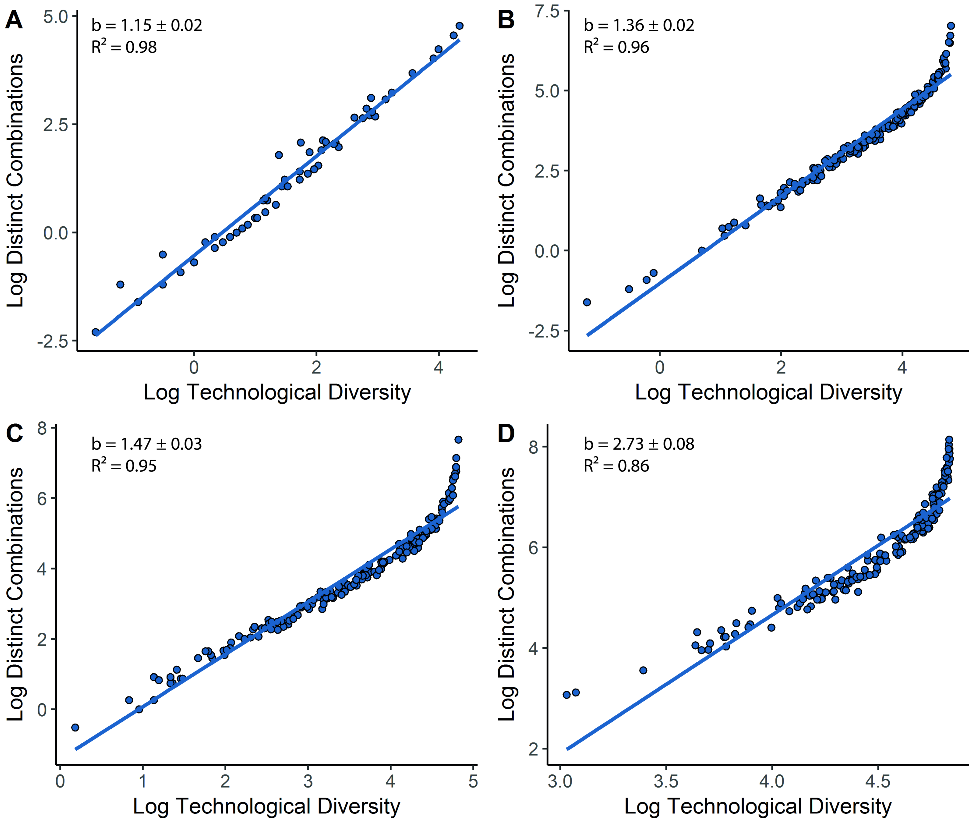

Figure 4.7: Scaling relationship between technological diversity and the total number of distinct class combinations **A** in 1850, **B** in 1900, **C** in 1950, and **D** in 2010 in US metropolitan areas using the CPC3.

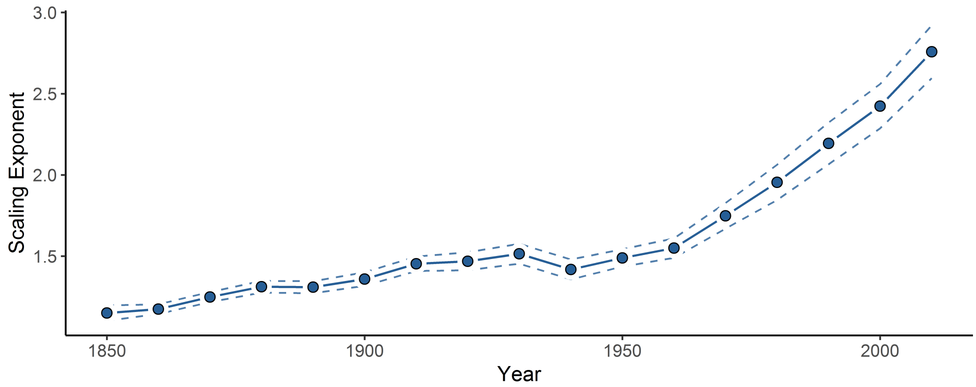

Figure 4.8: Scaling exponent of diversity with respect to the number of distinct combinations over time using CPC3. Dashed lines indicate the 95 percent confidence interval.

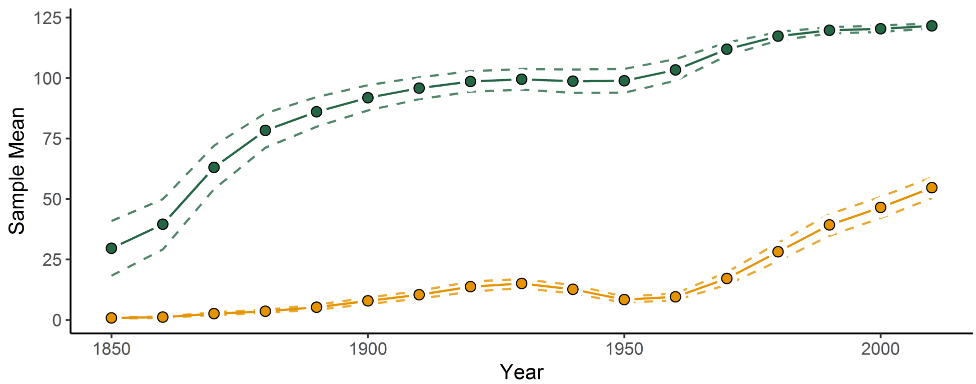

Figure 4.9: Average number of technologies in the most diversified (green line) and least diversified cities (orange line) using CPC3. Dashed lines indicate the 95 percent confidence intervals.

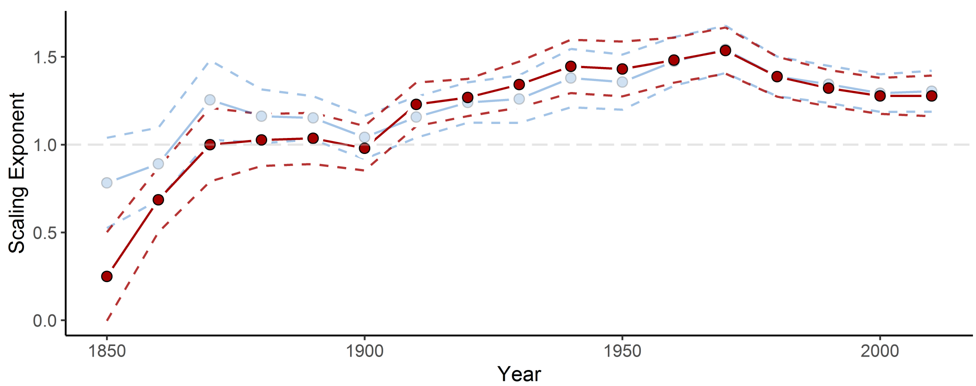

Figure 4.10: Scaling exponent of population size over time for atypical (red line) and typical combinations (blue line) using CPC3. Dashed lines indicate the 95 percent confidence interval.

5 The Effect of Macro-Psychological Openness on Impactful Innovation in US Metropolitan Areas

Abstract: High-impact innovations yield important benefits for the economic well-being of regions. While previous research has shown that regions substantially differ regarding their innovation quality, the sources of these regional differences remain largely unexplored. Here we argue, and test the assumption, that cultural differences between regions in terms of population-level trait openness help explaining the observed regional differences in innovation quality. Although previous research studied the effect of regional openness on innovation, the empirical evidence is limited to innovation quantity and relies on indirect statistical indicators of regional openness. To overcome the shortcomings, we directly measure peoples' openness in regions based on personality information of more than 1.26 million individuals from 382 Metropolitan Statistical Areas in the United States and test the macro-psychological approach against the indirect open-mindedness indicators applied in previous research (Florida's *"Creative Class"*, *"Gay Index"*, and *"Bohemian Index"*). We assess innovation quality in regions by relying on patent data and patent citations. Our results show that the effect of macro-psychological openness on innovation quality in regions is substantial for highly impactful innovations and nearly absent for the average innovation quality. These unique effects of macro-psychological openness persist even when controlling for indirect open-mindedness indicators, educational attainment, and the economic structure of regions.

This chapter is co-authored with Tobias Ebert, Martin Obschonka, P. Jason Rentfrow, Sam Gosling, and Jeff Potter. The PhD candidate is the first author of the article.

5.1 Introduction

On July 12, 2019, researchers from San Francisco proclaimed the rejuvenation of nine elderly men by turning back the men's biological clocks two years (Bahnsen 2019). Although their research has not gone through a peer-review process yet, the expected economic value, the impact on society and subsequent innovation could be tremendous. Was it just a coincidence that such an innovation was developed in San Francisco? Previous research suggests that it was not. Places substantially differ in their ability to produce high-impact innovations and California is one of them (Ejermo 2009; Castaldi and Los 2017). Little is known about the places that produce innovations of outstanding impact, as much research has been devoted to innovation quantity and less to quality (see Malecki 2010, for a review). Because innovations vary enormously in their importance, simple innovation counts are less informative and yield indistinct conclusions, as, for example, the number of innovations grows, while their quality declines (Hall and Ziedonis 2001). The larger technological and economic gains associated with high-impact innovations (Trajtenberg 1990) suggest that more innovations must not necessarily imply increasing benefits. But why are some places more successful in producing impactful innovations?

It might need a specific regional culture for impactful ideas to rise and flourish (Boytchev 2019). In fact, the importance of regional cultures for innovation is – sometimes explicitly and implicitly – discussed in previous research (Saxenian 1994; Rodríguez-Pose 1999; Florida 2002; Sandberg and Aarikka-Stenroos 2014). In her case study,Saxenian (1994) compared regional characteristics to explain the success in Silicon Valley and the decline along Route 128 in Massachusetts. Much of her work emphasizes cultural differences with people in the Valley being more open and flexible towards innovation than along Route 128. More general, Rodríguez-Pose (1999), distinguishes between innovation prone and innovation averse societies, which translate the regional potential into actual innovation behavior. Florida (2002) argues that regions' level of tolerance, which includes open-mindedness, plays an important role for innovation. In contrast, restrictive cultures represent an essential barrier for innovation. Sandberg and Aarikka-Stenroos (2014, p. 1298) define a restrictive culture as "shared values and beliefs that characterize groups of people in a particular place and orient their resistance to innovations." Such restrictive cultures may not only prevent single minds from having great ideas, but also create an environment in which novel ideas lack necessary appreciation and support (Riffai et al. 2012).

Regional culture, however, remains an elusive concept. While all the previously described contributions assume a prominent role of regional openness, a common understanding of regional openness is missing and its intangible nature greatly challenged empirical investigations. Lacking direct measurements, existing approaches therefore relied on indirect proxies. Prominently, Florida (2002) used the share of homosexuals, bohemians or creative occupations as indicators of tolerance. Although these measures surely have their justification given the available data and have greatly advanced our understanding of the importance of regional cultures to explain differences in innovation across regions, their interpretation is far from being clear and contain strong assumptions.

A way to overcome many shortcomings of previous research comes from psychology and uses personality characteristics of individuals (e.g. traits) living in a region to assess local cultures (Rentfrow et al. 2013). By using large-scale data sets with personality information of millions of people, psychologists reveal regional differences in the prevalence of personality traits (Rentfrow et al. 2008; Rentfrow et al. 2013; Jokela et al. 2015). Recent empirical research suggests that regional personality differences have crucial implications for regional development, as they impact economic growth and entrepreneurship in regions (Stuetzer et al. 2018; Garretsen et al. 2018b; Obschonka et al. 2015; Stuetzer et al. 2016).

Here we assess regional openness based on the personality information of more than 1.26 million participants from an online survey (Gosling et al. 2004). We aggregated the trait openness of the people in 382 *Metropolitan Statistical Areas* (MSAs) in the United States (US) and study the relationship between regional openness on innovation quality in regions. As a proxy for regional innovation, we rely on patent data from the *United States Patent and Trademark Office* (USPTO). Rather than assume that all innovation are equally important, we distinguish innovations‘ impact based on citation data (Trajtenberg 1990; Harhoff et al. 1999; Hall et al. 2005). We test our personality-based approach against the widely-used tolerance indicators of Florida (2003) to differentiate macro-psychological openness from the open-mindedness included in tolerance.

Our results indicate that a psychological regional climate of high openness is associated with a stronger output of *impactful* innovations, but not with a stronger output of *average* innovations. Regional openness as measured with these psychological data explains a substantial share of the regional variation of impactful innovations beyond potential confounders and the tolerance indicators used in previous research (Florida 2002). Hence, such an open psychological climate might indeed function as a hotbed for particularly creative and ambitious innovations, whereas it seems not necessary for the production of average, less impactful innovations.

The remainder of our article is structured as follows: We present and discuss relevant theoretical underpinnings as well as empirical works and derive our research hypothesis in Section 5.2. Section 3.3 entails the data and our methodological approach. We present our main results and robustness checks in Section 5.4 Section 5.5 concludes the paper.

5.2 Theoretical and empirical literature overview

5.2.1 Impactful innovations

Although previous research highlights the heterogeneity of knowledge (Polanyi 1966; Nelson and Winter 1982) and places (Malmberg and Power 2005; Maskell and Malmberg 1999; Martin and Sunley 2006; Boschma and Frenken 2011), the primary focus of empirical research is still on innovation quantity (Malecki 2010). Innovations are not a homogeneous quantity but substantially differ regarding qualitative features of which impact represents a crucial one. The differences in innovations’ impact emphasizes that innovations vary in their importance for subsequent innovations, their implications for society, and their economic value (Trajtenberg 1990; Harhoff et al. 1999; Hall et al. 2005).

The literature uses terminologies such as radical innovations (Chandy and Tellis 1998), disruptive innovations (Christensen 1997), or technological breakthroughs (Hargadon 2003) to indicate valuable innovations. Such terminologies, however, imply a specific consequence, for example, a breakthrough or the introduction of a radically new product, which is often not represented by the empirical measures. As we measure the impact of innovations in our empirical investigations, we use impactful innovations throughout the article to ensure a close bonding of theory and empirical analysis and to avoid misleading interpretations.

Innovation activities are generally strongly clustered in specific regions (Feldman 1994; Acs et al. 2002; Verspagen and Schoenmakers 2004). Considering innovations' impact reveals that impactful innovations are even more concentrated in space (Ejermo 2009; Castaldi and Los 2017). In other words, only very few places are able to create impactful innovations - these high-performing regions appear to benefit from *"special regional features"* that enables these impactful innovations. As impactful innovations are associated with higher economic returns (Trajtenberg 1990), it is important for regional economies if they produce highly impactful or less-impactful innovations. For instance, the Haber-Bosch innovation at BASF secured global food production and created hundreds of jobs in the region (Bosch 1932). Brin's and Page's innovation of an algorithm that searches the web lead to the foundation of Google, which now employs more than 20,000 people in the region (Brin and Page 1998).

However, we still know very little why some places produce more impactful innovations than others - what are these special regional features that enable some regions to produce innovations of outstanding quality? Among the few existing studies in this field, Castaldi et al. (2015) analyzed the influence of regional capabilities on highly impactful innovations. Their findings suggest that unrelated variety enhances the emergence of impactful innovations in regions and, hence, highlights the importance of the local economic structure for regional innovation quality. Except these insights, empirical research is (still) astonishingly silent about the underlying reasons of regions' ability to produce innovations of outstanding quality.

Beyond conventional, *"hard"*, economic factors, such as the local economic structure, other, more *"soft"* factors might play a crucial role to explain the regional variance in innovation quality. These soft, or intangible, factors emphasize that regional innovation is not only dependent on the people directly involved in R&D processes but requires trust (Fukuyama 1995), social capital (Putnam 2001), informal institutions (Rodríguez-Pose 1999), or a specific regional culture (Huggins and Thompson 2017). Regional culture as one important regional feature expresses the collective behavior of the local population determined by shared beliefs, social values and norms (Hofstede 1980; Huggins and Thompson 2017). One reason that people think and act differently in different places is due to differences in regional cultures (Oyserman 2017). Regional culture, however, remains elusive and its intangible characteristics emphasize that *"something is in the air"* (Marshall 1890, p. 198) that is difficult to capture empirically but that impacts regional innovation.

The framework of Florida (2002) highlights the importance of regional culture for regional development, including innovation. More precisely, tolerance, as a measure of openness to innovation, of the local population is an important building block in Florida's framework to explain regional differences in innovation. Open-mindedness indicates a regional social climate that cultivates new ideas and enhances knowledge flows (Florida 2002). Florida's work has influenced a large stream of research, investigating the role openness plays for regional innovation (Knudsen et al. 2008; Lee et al. 2010; Qian 2013; Sleuwaegen and Boiardi 2014). These empirical studies are, however, limited to the effect of regional openness on innovation quantity.

We argue that differences found in regional openness also play a crucial role in explaining the differences in regional innovation quality observed by Ejermo (2009) and Castaldi and Los (2017). A second limitation of previous research concerns the use of indirect openness indicators that are based on strong assumptions, such as the regional share of creative, bohemian, and homosexual people (Florida 2002; Florida 2003; Knudsen et al. 2008; Florida et al. 2008a; Acs and Megyesi 2009; Qian 2013). Although there might be a statistical link between sexual orientation or occupational choices and personal mindset, it is clear that not all homosexual people, bohemians or creative workers are necessarily truly tolerant or open-minded. In addition, high regional shares of, for example, homosexuals might be due to regions' openness, but other, not observed, factors could also play an important role. We therefore present and test a new approach from psychology to assess regional cultures, i.e. regional openness in this case, to overcome the shortcomings of previous empirical research by using the personality of the local population as a direct and robust measure of regional openness.

5.2.2 Regional personality differences

The emergence of the so-called Five Factor Model (Big Five) of personality (John and Srivastava 1999) has enabled psychologists to reliably and robustly assess personality differences between individuals. The Big Five are an exhaustive taxonomy of personality consisting of five broad dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness). These five personality dimensions are partially rooted in biology (Jang et al. 1998), culturally universal (Benet-Martínez and John 2000) and fairly stable across the lifespan (Roberts et al. 2006). Given their robustness and universality, the Big Five form a valid instrument to measure cultural differences (McCrae 2001; Hofstede and McCrae 2004; Rentfrow et al. 2008).

Large-scale personality data allowed studying more fine-grained cultural differences between regions. In their pioneering paper,Rentfrow et al. (2008) examined personality traits across US states based on data of more than half a million residents. Their empirical analysis of aggregated personalities suggests that US states show significant differences for all big five personality traits. Openness, for example, is highest in Washington DC, New York, Oregon, Massachusetts, Washington State, and California. In contrast, states low on openness are Wyoming, Nebraska, Iowa, Kentucky, and Alabama. As US states still represent a relatively large regional entity, studies have analyzed the granularity of personality differences for different definitions of regions such as metropolitan areas (Ebert et al. 2019), counties (Obschonka et al. 2018), and city districts (Jokela et al. 2015). These studies demonstrate the persistence of personality differences for a multitude of spatial scales. Systematic differences between regions are not restricted to a single nation, but have also been documented for various national and cultural contexts such as Great Britain (Rentfrow et al. 2015), Switzerland (Götz et al. 2018), Russia (Allik et al. 2009), and China (Wei et al. 2017) indicating regional personality differences as a universal phenomenon.

Crucially, these differences of personality across regions are linked to corresponding socio-economic behavior in regions. Spanning various national contexts, regional personality differences have been linked to critical macro-level outcomes such as voting behaviors (Obschonka et al. 2018; Garretsen et al. 2018a), emotional health (Mc-

Cann 2011), drug use (Harrington and Gelfand 2014), personal well-being (Rentfrow et al. 2009), entrepreneurial activities (Obschonka et al. 2015; Stuetzer et al. 2018), economic resilience (Obschonka et al. 2016), or economic growth (Stuetzer et al. 2018; Garretsen et al. 2018b).

5.2.3 Macro-psychological openness and impactful innovations

Among the Big Five traits, openness is the dimension most closely associated with innovation (George and Zhou 2001). Openness describes people's inventiveness, creativity, originality, and curiosity (McCrae 1987; King et al. 1996; McCrae 1996). Open individuals are able to create and tolerate new ideas, while less open individuals, are less likely to try something new and tend to be less flexible to new ideas (McCrae 1987; Gurtman 1995) . A positive effect of openness on innovation is highly intuitive and backed by empirical research at the country level (Steel et al. 2012). However, the relationship has not been confirmed at the regional level. Specifically, Lee (2017) studied the relationship between personality differences and innovation quantity in UK regions and did not find an effect of regional openness on innovation quantity. Instead, his results report a positive association between conscientiousness and innovation quantity. Such a finding suggests that not an open regional culture, but rather the prevalence of hard-working mindsets with focus on self-discipline and task completion are conducive to innovation.

There might be good reasons for Lee's (2017) finding and a clear explanation why openness is less relevant for a region's quantitative innovation output. Importantly, there are strong theoretical arguments why regional openness might be more relevant for innovation quality. Impactful innovations meaningfully differ from incremental innovations regarding their characteristics and prerequisites. Incremental innovations represent minor improvements of existing technologies and are considered technologically as well as economically less valuable (Trajtenberg 1990). For these types of innovations, hard-work, commitment, and discipline, i.e. attributes associated with conscientiousness, might be more important than inventive and creative thinking (Drucker 1985).

In contrast, high-impact innovations can break with main-stream conventions, leave existing trajectories, and create new technological paths (Hargadon 2003). They explicitly involve exploration and unconventionality (Schilling and Green 2011; Uzzi et al. 2013; Kaplan and Vakili 2015; Kim et al. 2016). Such an exploration of new ideas requires (a) creativity and action outside core competences and (b) overcoming rigidities and uncertainties (Leonard-Barton 1992; O'Connor 1998). Accordingly, for these types of innovations, out-of-the-box thinking, risk taking, and tolerance of mistakes, i.e. attributes associated with openness, might be key.

This view on a particular relevance of openness for impactful innovations is in line with research in psychology and management showing that open cultures are particularly relevant for the qualitative output of innovation processes. Open cultures welcome new ideas, creativity, and unconventional thinking. In other words, open cultures provide an environment that is conducive to the emergence and subsequent acceptance of novel and creative ideas (Reilly et al. 2002; Aronson et al. 2008; Tellis et al. 2009; Sandberg and Aarikka-Stenroos 2014; Fitjar and Rodríguez-Pose 2011). In contrast, strong barriers preventing highly impactful innovations are rigid thinking (Leonard-Barton 1992) and restrictive mindsets including a "fear of change, fear of failure, conservative decision-making, and restrictive organizational culture" (Sandberg and Aarikka-Stenroos 2014, p. 1298).

5.2.4 How can regional openness influence impactful innovations in regions?

Innovations are produced by only a tiny fraction of the regional population: innovators. Focusing on innovation impact, yields an even smaller subsample of innovators. So, how are regional levels of openness in local populations linked to a regional outcome produced by only a tiny group?

The conceptual framework of Rentfrow et al. (2008) provides two mechanisms through which regional levels of openness manifest in aggregate innovation performance. First, the observed outcome at the regional level might result from a clustering of persons expressing their individual disposition (Roberts et al. 2007; Huggins and Thompson 2017). Individual level research shows that the personality of inventors is linked to their success. Precisely, inventors high on openness produce more impactful inventions than inventors low on openness, which is more likely in more open regions (Zwick et al. 2017). Accordingly, regions higher on openness feature a higher probability that open people are directly involved in the innovation process.

The second mechanism focuses on social influence. If a region features a disproportionate share of people with a specific disposition (openness in our case), the behavioral tendencies associated with that disposition occur more often, become accepted, and socially valued (i.e. a social norm) (Rentfrow et al. 2008). Living in a region where a large share of people is open, could produce a social norm, which welcomes new ideas and values originality. Via social influences, the regional social norm eventually affects peoples' behavior and attitudes, irrespective of their natural disposition (Latané 1981; Huggins and Thompson 2017). Regional social norms can influence innovators via different channels. Through imprinting, for example, firms can take on elements of their local environment (Stinchcombe 1965). Accordingly, firms adopt persistent characteristics that reflect prominent features of the local environment such as the prevalent social norm (Marquis and Tilcsik 2013). Hence, the regional social norm becomes apparent in firm attributes and influence the behavior of their employees. Another channel is based on the embeddedness of people and firms in networks (Granovetter 1985). Individuals constantly interact with their social environment outside the firm including formal (e.g. business relations) as well as informal links (e.g. friendships, occasional relationships) (Granovetter 1985; Uzzi 1997). Firms are also linked to other organizations or institutions (e.g. inter-firm or university-firm collaborations, spin-off activities, buyer-supplier relations). Bathelt et al. (2004) call the sum of social and economic relations within a region local buzz, which facilitates the diffusion of the prevalent social norm among economic actors. Consequently, even if innovators are not open themselves, they can be influenced in their behavior by the accepted social norm in the region (Sandberg and Aarikka-Stenroos 2014). Our theoretical reflections lead to the following hypothesis:

H1: Regional openness has a positive effect on highly impactful innovations in regions, but not on the average innovation quality.

5.3 Materials and methods

Defining regions is not a trivial task and can have significant implications for regional analyses (Manley 2014). We choose Metropolitan Statistical Areas (MSAs) in the US as our spatial unit of analysis, as they are functional regions - consisting of an urban core and adjacent hinterland with strong social and economic ties. MSAs represent a standard set of spatial entities, which reveal pronounced differences in innovations (Boschma et al. 2015) and macro-psychological characteristics (Obschonka et al. 2015). Hence, all variables represent regional aggregations on the level of 382 MSAs.

5.3.1 Impactful innovations in regions

Patents are the most widely established indicator of innovation activities used in countless studies analyzing innovation activities. Although patents as a measure of innovation certainly have well-known limitations (Griliches 1990; Cohen et al. 2000), they nevertheless are one of most established and best-validated indicator of innovations (Acs et al. 2002). We use patent data from the USPTO as our primary data source. Additionally, we rely on the location information by (Li et al. 2014) to assign patents to corresponding MSAs based on inventors' residential information.

Patents vary substantially regarding their socio-economic and technological impact (Trajtenberg 1990; Harhoff et al. 1999; Hall et al. 2005). These differences in patents quality show significant geographic variation highlighting the importance to consider patents' impact in geographic analyses (Ejermo 2009; Castaldi and Los 2017). Following a common approach, we weight patents according to their received number of citations as a measure of impact. In analogy to citations in academia, patents also refer to preceding patents to indicate the relevance of the cited patent for subsequent ones. Hence, forward citations are a suitable way to measure a patent's impact (Trajtenberg 1990; Hall et al. 2001; Hall et al. 2005).

Citation counts, however, have some peculiarities, which are important to control for. First, proximity influences citation behavior (Hall et al. 2001). Citations are, therefore, more likely when citing and cited patent occur close to each other. That is to say, patents cite others not only because of their importance, but because they appeared in close geographic proximity (Jaffe et al. 1993). To control for geographic proximity, we applied a rather conservative approach and excluded all intra-regional citations from our sample.

Second, cohort specific citation patterns can distort citation counts. Citation patterns vary significantly between technology classes. Some classes cite more and faster than others (Hall et al. 2001). These patterns can have severe consequences for geographic analyses, as some technology classes concentrate in specific regions (Breschi and Malerba 1997). Relying on simple citations counts can therefore distort the results and their interpretation, because differences in citations counts are biased by location decisions. Some regions might have higher citation counts, simply because they are home to many citation-intensive technology classes. Hall et al. (2001) therefore suggest class and cohort corrected citation counts by dividing the received number of citations with the average number of citations in a given class and year. This correction eliminates any variance in citations counts due to location decisions and reveals true impact differences. Theoretically, the corrected citation count runs from zero to infinity. Counts greater than one indicate that patents received more citations than the average patent in the same class and year.

But when is a patent truly impactful or even highly impactful? Clearly, there is no correct answer to this question. Studies usually rely on a fixed and rather arbitrary threshold to identify high-impact such as the top 1, 5, and 10 percent cited patents (Uzzi et al. 2013; Kim et al. 2016). Few exceptions identify high-impact based on endogenous selection procedures (Castaldi et al. 2015; Castaldi and Los 2017). Both approaches, however, have three disadvantages, which are crucial in the context of our analysis. Firstly, they dichotomize a continuous outcome, i.e. either high-impact or not. This disregards information and treats all impactful innovations as a homogeneous group. Secondly, grouping invention in either high-impact or not unnecessarily results in a count variable with many zeros at the regional level, which again removes important information and makes impactful innovations a rare event. Third, bucketing patents into two groups seems also inflexible, as it only allows to compare high-impact with the general population of patents. The link between openness and innovation, however, might have more fine-grained nuances.

We therefore present an alternative approach based on each regions' impact distribution. The regional impact distributions represents the ranking of patents according to their impact in decreasing order for every regions. We calculate the mean of the corresponding regional impact distribution in different x^{th} percentiles of the distribution. For example, if x takes the value 10, the corresponding mean in the 10^{th} percentile indicates the average number of citations received by the top 10 percent of the cited patents in a region. The 100^{th} percentile obviously gives the mean impact of all cited patents, i.e. the average innovation quality in regions. This approach allows us to gradually shift percentiles from high-impact closer to the average and analyze the relationship between openness and impact more flexibly for all percentiles of the impact distribution. Similar approaches, for example, have been applied to measure the complexity of economic and technological activities in a region (Balland et al. 2018).

5.3.2 Macro-psychological openness in regions

Our main independent variable is regional openness based on personality information of the local population in regions. To measure regional levels of openness across MSAs, we used data collected between 2005 and 2015 via the Gosling-Potter Internet Personality project – a website that gives participants customized feedback on their personality after completion of an online survey (Gosling et al. 2004). The Gosling-Potter data is the largest self-report database in psychology and has proven its demographic and psychometric suitability for cross-regional research in a wide variety of previous studies (Rentfrow et al. 2008; Rentfrow et al. 2013). Within this data, personality is assessed via the Big Five Inventory (BFI), a widely established measure of the Big Five (John and Srivastava 1999). The BFI consists of 44 items that contain short phrases of prototypical markers of each of the five dimensions: extraversion, agreeableness, conscientiousness, emotional stability and openness. Participants reported the degree to which they agreed with each statement using a 5-point rating scale (ranging from 1 [*Disagree strongly*] to 5 [*Agree strongly*]).

For the present investigation, we only included participants who reported living within one of the 382 US metropolitan areas, were between 18 and 90 years of age, and completed the personality measure. Regional sample sizes range from 149 (The Villages, Florida) to 63,877 (New York-New Jersey). Overall, regional sample sizes almost perfectly correlated with the size of the actual regional population ($r = 0.96$, $p < 0.000$). We therefore considered all of the 382 MSAs in our empirical analysis. Our final sample comprised 1,269,225 participants. However, as is typical for online studies (Gosling et al. 2004), the demographic composition of our sample is skewed, with females (64.82%) and younger people (mean age: 31.07, SD: 11.86) being overrepresented . To derive regional openness scores, we followed the standard approach in geographical psychology (Rentfrow et al. 2008) and averaged individual openness scores for each region.

Previous economic research on regional openness typically considers indirect open-mindedness indicators such the share of creatives, bohemians, and homosexuals in a region (Florida 2003; Lee et al. 2004; Florida et al. 2008a; Knudsen et al. 2008; Acs and Megyesi 2009; Qian 2013; Sleuwaegen and Boiardi 2014). We thus explore to what extent trait openness differentiates or goes beyond these open-mindedness indicators applied in previous research. To create the open-mindedness variables we rely on data from the American Community Survey (U.S. Census Bureau 2010) and calculate the share of people employed in creative industries (*"creative"*) defined as arts, design, entertainment, sports, and media occupations. A subsample of this group are, what Florida (2003) calls bohemians, people working as actors, dancers, designers, directors, musicians, photographers, producers, and writers. Using these occupations, we calculate the regional share of bohemians (*"bohemians"*). Finally, we consider the share of same-sex couples in a region (*"homosexuals"*) to approximate the number of homosexuals in a region.

5.3.3 Control variables

We try to carve out the unique effect of regional openness on impactful innovations. Therefore, we applied a conservative analysis and included a rich set of potential confounders and standard control variables. Firstly, the regional knowledge intensity can have an influence on innovation quality in regions. Human capital is a crucial variable that explains regional differences in innovation (Pater and Lewandowska 2015) and better qualified workers are also more likely to produce more impactful innovations (Zwick et al. 2017). Relying on data from the American Community Survey (U.S. Census Bureau 2010), we constructed two different variables to indicate human capital, which are the share of the regional population with a bachelor's degree (*"bachelor"*) and the share of the regional population working in science related occupations (*"science"*) defined as computer, engineering, and science occupations. To control for the regional presence of high-quality research, we also collected data about so-called star scientists (*"stars"*) from *Clarivate Analytics'* list of *Highly-Cited Researches* (hcr.clarivate.com) (Zucker and Darby 1996). Additionally, we consider the number of patents per capita (*"patents"*) as an indicator of the regional knowledge stock, which we also draw from the USPTO.

Secondly, innovation activities are dependent on the regional economic structure. A large body of research in economic geography discusses the impact of specialization and diversity on local outcomes (Beaudry and Schiffauerova 2009). We therefore calculated the average location quotient (Balland 2017) as a standard indicator of technological specialization (*"lq"*) and the exponentiated Shannon entropy

(Jost 2006) as a measure of technological diversity (*"diversity"*) based on regional patenting activities. Patents are also likely biased towards innovation activities in manufacturing sectors. We therefore calculated the share of the local labor force employed in manufacturing (*"manufacturing"*) based on US census data (U.S. Census Bureau 2010).

Finally, we control for the regional population density (*"popdens"*), because urban areas are disproportionately more innovative than smaller towns (O'hUallichain 1999; Bettencourt et al. 2007b). Additionally, population density can also be regarded as a catch-all variable to proxy a large array of regional attributes such as such as land prices, size of the labor market, and availability of infrastructure. Again, the US Census (U.S. Census Bureau 2010) serves as our data source for the regional population. Table 5.1 summarizes all variables used in our analyses, their definitions, time coverage, and data sources. Table 5.2 provides summary statistics a correlation matrix of all variables.

5.4 Empirical results

5.4.1 Mapping regional openness and impactful innovations in regions

Figure 5.1 panel A maps z-standardized openness scores for 382 MSAs and shows substantial regional variation. Openness concentrates along the coastal regions, in urban areas, and university towns (Madison, Wisconsin and Boulder, Colorado). MSAs with a pronounced open regional culture are Santa Fe (New Mexico), Savannah (Georgia), Santa Cruz-Watsonville (California), Santa Maria-Santa Barbara (California), and San Francisco-Oakland-Hayward (California). In contrast, MSAs low on openness are Grand Island (Nebraska), Chambersburg-Waynesboro (Pennsylvania), Rome (Georgia), Danville (Illinois), and Lima (Ohio). Substantial regional variation of openness has been shown for US states in previous studies (Rentfrow et al. 2008; Rentfrow et al. 2013). Panel A also emphasizes that US states represent relatively large regional entities that hide much of the regional heterogeneity. MSAs within the same state show substantial differences regarding openness. For example, Texas congregates MSAs ranging from very low (San Angelo) to very high (Austin-Round Rock) openness values. This is not restricted to relatively large states such as Texas, but also observed for smaller states such as Washington, Indiana or Louisiana.

As reported in many studies, innovation activities strongly concentrate in space (Feldman 1994; Acs et al. 2002; Verspagen and Schoenmakers 2004). Mapping the average innovation quality in panel B of Figure 5.1, confirms these strong geographic patterns for MSAs in the US. Panel C visualizes highly impactful innovations, i.e. the average in the 10^{th} percentile of the regional impact distribution. Interestingly, some regions that produce highly impactful innovations (panel C) have on average a rather medium innovation quality (panel B) such as Seattle-Tacoma-Bellevue (Washington State), Boise-City (Idaho), and Sacramento-Roseville (California). The different geographic patterns in panel B and C suggest that the average innovation quality in regions does not automatically imply that regions also produce highly impactful innovations.

By comparing the maps in panel A and panel C of Figure 5.1, we observe con-

Table 5.1: List of variables with their definitions, time coverage, and data sources

Variable	**Definition**	**Time Coverage**	**Data Source**
average innovation impact	Average number of citations received by patents in a region	2001-2010	USPTO for raw patent data and Li et al. (2014) for geographic information
impactful innovations	Average number of citations in x^{th} of the regional impact distribution [For example: $x = 10$, indicates the average number of citations received by the top 10 percent of the cited patents in a region]	2001-2010	USPTO for raw patent data and Li et al. (2014) for geographic information
openness	Average trait openness of individuals in regions	2004-2014	Gosling-Potter Internet Personality Project (Gosling et al. 2004)
bachelor	Percentage of the regional population (25 to 64) years with bachelor degree or higher	2006-2010	U.S. Census Bureau (2010)
stars	Regional number of highly-cited researchers (i.e. star scientists) per 1,000 college students	2001	Clarivate Analytics for highly-cited researchers and U.S. Census Bureau (2000) for college undergraduates
patents	Regional number of patents per capita in log	2001-2010	USPTO for raw patent data and Li et al. (2014) for geographic informatio
science	Percentage of the regional labor force employed in science-related (computer, engineering, and science) occupations	2006-2010	U.S. Census Bureau (2010)
lq	Average location quotient	2001-2010	USPTO for raw patent data and Li et al. (2014) for geographic information
diversity	Exponentiated Shannon entropy (Jost 2006)	2001-2010	USPTO for raw patent data and Li et al. (2014) for geographic information
manufacturing	Percentage of the regional labor force employed in manufacturing	2006-2010	U.S. Census Bureau (2010)
popdens	Regional population density calculated as the total population per square km in log	2006-2010	U.S. Census Bureau (2010)
Open-mindedness, i.e. indirect tolerance indicators			
creative	Percentage of the regional labor force employed in creative (arts, design, entertainment, sports, and media) occupations	2006-2010	U.S. Census Bureau (2010)
bohemians	Regional number of bohemian occupations per 1,000 occupations According to Florida (2003), bohemian occupations are actors, dancers, designers, directors, musicians, photographers, producers, and writers	2006-2010	U.S. Census Bureau (2010)
homosexuals	Regional number of same-sex couples per 1,000 households	2006-2010	U.S. Census Bureau (2010)

Table 5.2: Descriptives and correlations

	Variables	Min	Max	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1	average impact	0.45	1.30	0.98	0.08																	
2	impactful innovations[$x^{th} = 10^{th}$]	0.45	3.88	2.47	0.50	0.73																
3	openness	3.49	3.90	3.67	0.06	0.15	0.37															
4	bachelor	11.68	58.74	26.57	8.17	0.27	0.58	0.32														
5	stars	0	2.75	0.11	0.30	0.21	0.39	0.26	0.54													
6	patents	1.28	8.66	5.17	1.07	0.33	0.59	0.20	0.68	0.42												
7	science	1.21	16.26	4.67	2.22	0.27	0.58	0.26	0.80	0.52	0.73											
8	lq	1.23	360.09	12.18	21.84	-0.07	-0.36	-0.25	-0.42	-0.23	-0.40	-0.38										
9	diversity	3	204.77	91.81	41.15	0.15	0.28	0.26	0.36	0.15	0.35	0.29	-0.60									
10	manufacturing	1.97	39.14	11.54	5.66	0.08	-0.12	-0.40	-0.22	-0.14	0.24	-0.04	0.10	-0.04								
11	popdens	1.01	6.92	4.22	0.93	0.19	0.31	0.12	0.29	0.25	0.40	0.28	-0.30	0.51	0.21							
12	creative	0.34	4.36	1.64	0.56	0.25	0.53	0.49	0.80	0.39	0.54	0.59	-0.41	0.36	-0.31	0.22						
13	bohemians	1.45	19.68	8.07	2.86	0.30	0.54	0.48	0.74	0.37	0.58	0.54	-0.40	0.40	-0.21	0.31	0.88					
14	homosexuals	0.99	8.09	3.42	1.03	0.22	0.33	0.36	0.32	0.23	0.21	0.17	-0.16	0.12	-0.26	0.09	0.39	0.40				
15	conscientiousness	3.35	3.84	3.55	0.06	-0.14	-0.21	-0.22	-0.08	-0.12	-0.16	-0.15	0.12	-0.04	0.04	-0.06	-0.10	-0.11	-0.08			
16	extraversion	3.16	3.53	3.32	0.06	0	-0.12	-0.44	-0.02	-0.11	-0.07	-0.12	0.12	0.05	0.19	0.13	-0.08	-0.07	-0.18	0.18		
17	agreeableness	3.58	4	3.75	0.05	-0.10	-0.18	-0.40	-0.02	-0.07	-0.10	-0.11	0.06	0.01	0.16	0.04	-0.09	-0.10	-0.13	0.66	0.49	
18	neuroticism	2.70	3.12	2.92	0.06	-0.07	-0.12	0	-0.25	-0.04	-0.07	-0.16	0.13	-0.18	0.14	0.03	-0.21	-0.22	0	-0.36	-0.44	-0.50

siderable overlaps between regions that are high on openness and produce very impactful innovations. For example, among these regions are Santa Cruz-Watsonville (California) and San Francisco-Oakland-Hayward (California). Both regions appear in the top five MSAs regarding openness and high-impact. To go beyond a visual interpretation, we analyze the relationship between openness and innovation activities more systematically in the next subsection.

5.4.2 The relationship between regional openness and impactful innovations

Our hypothesis states that regional openness and is more important for highly impactful innovations in regions. To investigate our hypothesis more systematically, we calculated impact in different percentiles along the regional impact distribution (see Section 5.3.1). In a first approach, we correlate regional openness with the average impact in different percentiles by shifting x^{th} from 1 to 100, i.e. from the most impactful innovations in regions to the average innovation quality. Figure 5.2 displays the correlation coefficient of regional openness and innovation quality with corresponding confidence intervals. The correlation between regional openness and highly impactful innovations is 0.35 (lower = 0.26, upper = 0.44) and considerably decreases as we consider less impactful innovations. This result suggests that the link between regional openness and innovation quality is stronger for highly impactful innovations.

To examine the link more systematically, we investigate the relationship between regional openness and innovation in several multivariate settings relying on linear regressions. That is, we analyze the extent to which regional openness explains additional variance of innovation activities' impact beyond a conservative set of economic control variables. Note, that we cannot include all control variables in the same estimations as some variables are highly correlated. For example, the correlation between the share of individuals with bachelor degree and science related occupations is 0.78 with $p < 0.000$. We therefore restrict our analysis in this section to a selected set of control variables and explore the sensitivity of the results with alternative controls in our robustness checks in the next subsection. The variance inflation vector is below 3 in all estimated models suggesting that multicollinearity is not an issue in our main estimations. As our analysis is based on spatial data, we test the presence of spatial dependencies of the residuals by applying a Lagrange Multiplier Test (LMT). The LMT's p-values are reported in every results table. Spatial autocorrelation is present in a number of models as indicated by a LMT p-value below 0.05. To test the effect of spatial dependency on our results, we estimated a spatial lag model in our robustness checks (see Section 5.4.4).

In a first set of models, we regress the average innovation quality on regional openness and our selected set of control variables. Table 5.3 reports the results with standardized coefficients. Regarding the control variables, human capital is a positive and significant predictor of average innovation quality in regions as indicated by the general level of educational attainment (*"bachelor"*, b = 0.015, lower = 0.003, upper = 0.027, p = 0.015) as well as the presence of top researchers (*"stars"*, b = 0.007, lower = 0.001, upper = 0.013, p = 0.025) in Model 1b and also in Model 1c when we include openness. The average quality is also higher in regions with a stronger focus on manufacturing industries, as the share of manufacturing activities

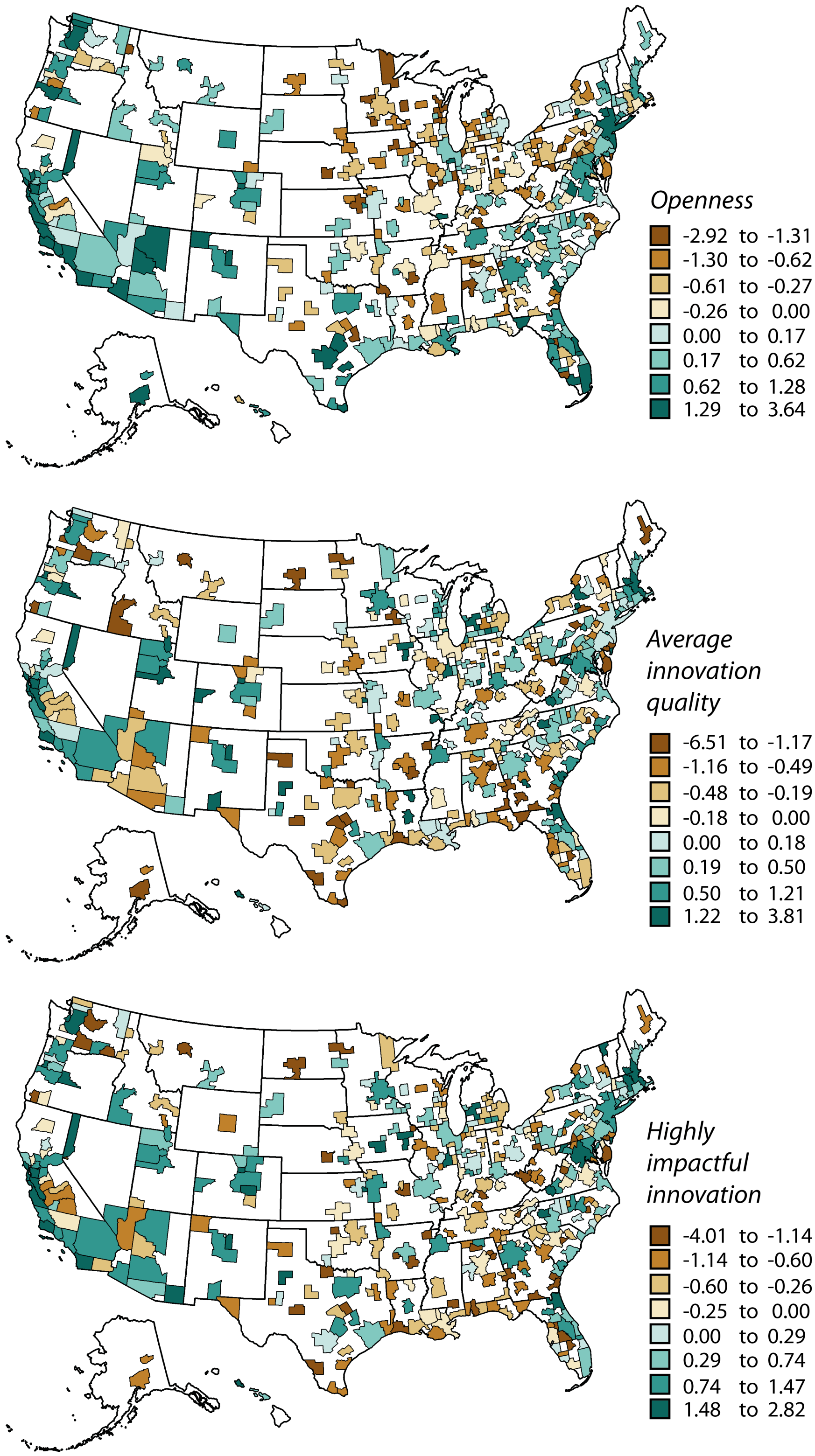

Figure 5.1: Maps of **A** regional openness, **B** average innovation quality (average innovation impact), and **C** highly impactful innovations (average of the 10% most impactful innovations in regions). All three variables have been standardized.

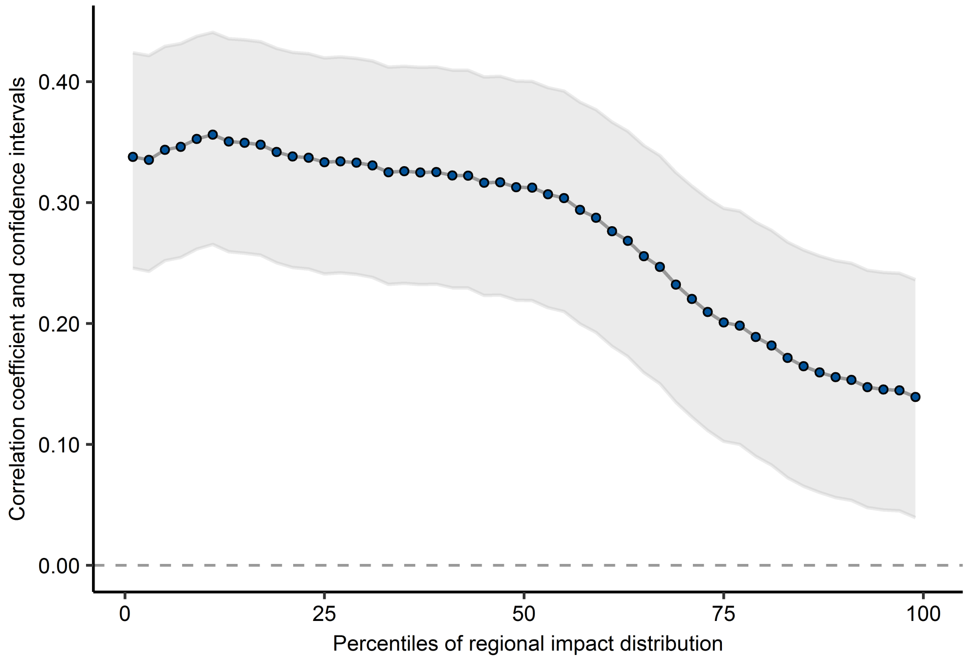

Figure 5.2: Pearson correlation coefficient and corresponding confidence intervals between innovation quality calculated in different percentiles of the regional impact distribution and regional openness.

is positive and significant (*"manufacturing"*, b = 0.01, lower = 0.002, upper = 0.018, p = 0.019).

Model 1a demonstrates that *openness* is positively related to the average quality in regions (b = 0.011, lower = 0.001, upper = 0.020, p = 0.029). If we consider the set of control variables, the coefficient of openness stays positive, but becomes insignificant (b = 0.007, lower = -0.003, upper = 0.018, p = 0.172). In general, the goodness of fit is rather low if we only consider openness in the base Model 1a (R^2 = 0.01) and also if we consider openness together with our control variables in Model 1c (R^2 = 0.167). By comparing the R^2 in Model 1b and 1c, we can calculate the additional gain of considering openness beyond the standard set of economic control variables. Accordingly, including regional openness results in a 1.83 percent increase in R^2.

In a second set of models, we analyze the relationship between regional openness and highly impactful innovations measured as the average impact in the 10th percentile of the regional impact distribution. Table 5.4 reports the corresponding results. Again, human capital is positive and significant for the share of people with bachelor's degree (*"bachelor"*, b = 0.209, lower = 0.152, upper = 0.265, p $<$ 0.000) and the presence of top researchers (*"stars"*, b = 0.048, lower = 0.006, upper = 0.090, p = 0.026). Interestingly, population density was insignificant in Model 1c and 1b and becomes significant in Models 2b (*"popdens"*, b = 0.068, lower = 0.016, upper = 0.119, p = 0.01) and 2c (*"popdens"*, b = 0.063, lower = 0.012, upper = 0.114, p = 0.016). This result suggests that urban environments are more important for the emergence of highly impactful innovations in MSAs than for the average innovation quality.

Regarding openness, the coefficient is positively significant in the base Model 2a

Table 5.3: Regression results for the average innovation quality

Y = Average innovation quality in regions ($x^{th} = 100^{th}$)			
	Openness	Controls	Openness + Controls
	(1a)	(1b)	(1c)
openness	0.011*		0.007
	(0.005)		(0.005)
bachelor		0.015*	0.014*
		(0.006)	(0.006)
stars		0.007*	0.006*
		(0.003)	(0.003)
lq		−0.023	−0.023
		(0.032)	(0.034)
diversity		−0.002	−0.003
		(0.009)	(0.010)
manufacturing		0.010*	0.012**
		(0.004)	(0.005)
popdens		0.004	0.004
		(0.005)	(0.005)
constant	0.984***	0.984***	0.984***
	(0.004)	(0.004)	(0.004)
LMT p-value	0.117	0.045	0.079
Observations	382	382	382
Adjusted R^2	0.014	0.164	0.167

Note: **p<0.05; **p<0.01; ***p<0.001.*
Robust standard errors in parentheses.
All independent variables are standardized.

Table 5.4: Regression results for highly impactful innovations

Y = Highly impactful innovations in regions ($x^{th} = 10^{th}$)			
	Openness	Controls	Openness + Controls
	(2a)	(2b)	(2c)
openness	0.177^{***}		0.097^{***}
	(0.027)		(0.024)
bachelor		0.209^{***}	0.196^{***}
		(0.029)	(0.028)
stars		0.048^{*}	0.037
		(0.021)	(0.019)
lq		-0.129^{***}	-0.136^{***}
		(0.028)	(0.013)
diversity		−0.006	−0.025
		(0.027)	(0.026)
manufacturing		−0.009	0.023
		(0.021)	(0.023)
popdens		0.068^{*}	0.063^{*}
		(0.026)	(0.026)
constant	2.468^{***}	2.468^{***}	2.468^{***}
	(0.024)	(0.020)	(0.020)
LMT p-value	0.002	0	0.004
Observations	382	382	382
Adjusted R^2	0.122	0.406	0.434

Note: *$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors in parentheses.
All independent variables are standardized.

without any control variables (b = 0.177, lower = 0.124, upper = 0.230, $p < 0.000$). The coefficient size decreases in Model 2c with control variables, but keeps its sign and significance (b = 0.097, lower = 0.051, upper = 0.144, $p < 0.000$) suggesting a robust link between highly impactful innovations and regional openness. Ceteris paribus, a one standard deviation increase of openness (which is comparable with the difference of openness between Miami and San Francisco) increases the average impact of highly impactful innovations in regions by 0.097. Miami's innovation impact would grow by 3.63 percent. Regarding the explanatory power of openness, its consideration results in a 6.9 percent growth in R^2 (comparison of Model 2b and 2c).

To analyze the sensitivity of our results with regard to the chosen percentile x^{th}, we rerun the full Model (2c) for different definitions of impact. Figure 5.3 summarizes the sensitivity analysis and has two important implications. Firstly, panel A shows the coefficients of openness and its corresponding confidence intervals. As the dependent variable was calculated in different percentiles of the regional impact distribution in which citation scores significantly differ, the coefficients cannot be compared directly. Higher percentiles automatically have higher values and thus gain higher coefficient estimates. To allow for direct comparison and to ease the interpretation, we standardize the variables on both sides of the equation. As shown by panel A, the results are robust regarding the chosen threshold, since regional openness is positive and significant in high-impact percentiles, for example, ranging from 1 to 20 and turns insignificant in upper percentiles ranging from 76 to 100. Additionally, the size of the coefficient constantly decreases.

Secondly, panel B depicts the goodness of fit as measured by the adjusted R^2 again for all percentiles of the regional impact distribution. The adjusted R^2 constantly decreases. By comparing the R^2 in models with and without openness, we can calculate the growth in R^2, which is attributed to openness. Panel C visualizes the growth in R^2 and shows that openness produces the highest growth in the high-impact percentiles. The additional gain of openness constantly decreases by shifting the percentiles towards the average innovation quality in regions. Based on the results presented in this subsection, we conclude that considering innovations' impact has crucial effects on the results. In sum, we interpret our findings as a support for our hypothesis that the link between regional openness and innovation is strongest for highly impactful innovations and less important for the average innovation quality in regions.

5.4.3 Macro-psychological openness against open-mindedness

The intangible nature of openness challenged empirical investigations. Previous approaches therefore relied on indirect measures. Prominently,Florida (2002) argues that technology, talent, and tolerance are three important factors of regional development. He defines tolerance as "openness, inclusiveness, and diversity to all ethnicities, races, and walks of life" (Florida 2003, p. 10). Lacking direct measurements, he relied on indirect proxies such as the share of creative occupations, bohemians, and homosexuals to approximate regional open-mindedness (Florida 2003). Instead, geographical psychologists argue that openness measured as a personality trait is a more robust and more direct way to assess the prevailing culture in a region via the

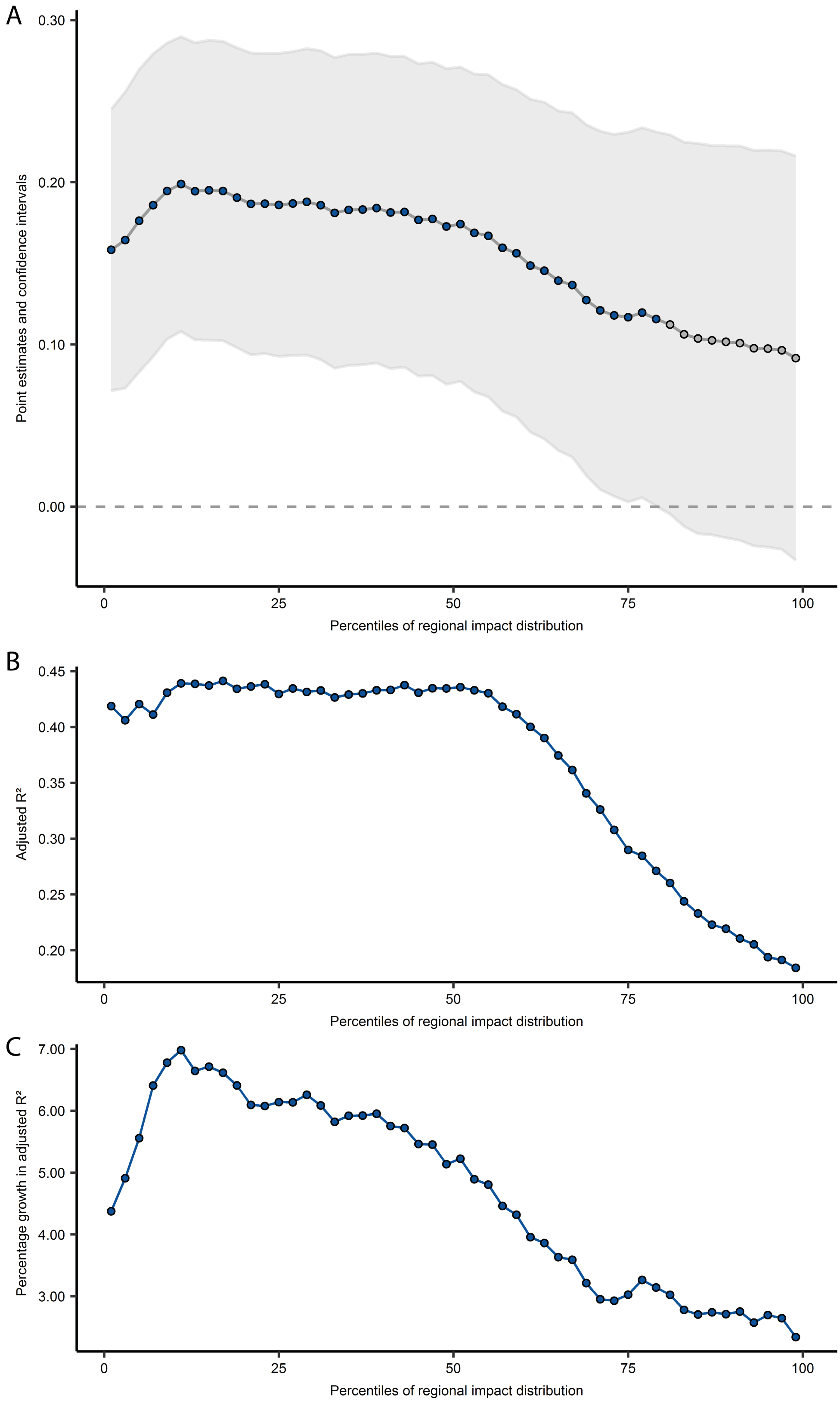

Figure 5.3: Sensitivity of regression results to the chosen percentiles of the regional impact distribution. **A** Coefficient and corresponding confidence intervals of openness, **B** adjusted R^2 for the estimated models, and **C** the growth in R^2 due to openness.

personality of its individuals (Rentfrow et al. 2008). We explore how trait openness explains regional variation of highly impactful innovations in comparison with Florida's indirect indicators of open-mindedness. Table 5.5 documents our results. Note that the number of observations varies in Models 3a to 3d, since information about the relevant indicators were not existing for all regions.

Table 5.5: Regression results for macro-psychological openness against open-mindedness

Y = Highly impactful innovations in regions ($x^{th} = 10^{th}$)				
	Openness + Creative	Openness + Bohemians	Openness + Homosexuals	Full Model
	(3a)	(3b)	(3c)	(3d)
openness	0.083**	0.086**	0.088***	0.078**
	(0.026)	(0.026)	(0.026)	(0.028)
creative	0.054			−0.033
	(0.037)			(0.058)
bohemians		0.068*		0.073
		(0.034)		(0.049)
homosexuals			0.060**	0.049*
			(0.022)	(0.024)
bachelor	0.153***	0.159***	0.191***	0.168***
	(0.038)	(0.038)	(0.029)	(0.042)
stars	0.041*	0.033	0.026	0.030
	(0.019)	(0.020)	(0.021)	(0.022)
lq	−0.127*	−0.141***	−0.145***	−0.129*
	(0.050)	(0.012)	(0.012)	(0.062)
diversity	−0.036	−0.045	−0.038	−0.047
	(0.027)	(0.027)	(0.026)	(0.027)
manufacturing	0.018	0.026	0.032	0.024
	(0.026)	(0.024)	(0.023)	(0.027)
popdens	0.063*	0.063*	0.065*	0.058*
	(0.026)	(0.026)	(0.025)	(0.026)
constant	2.481***	2.472***	2.471***	2.482***
	(0.020)	(0.020)	(0.020)	(0.022)
LMT p-value	0.008	0.019	0.027	0.033
Observations	366	352	352	341
Adjusted R^2	0.396	0.445	0.450	0.409

Note: *$^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.*
Robust standard errors in parentheses.
All independent variables are standardized.

In all models, trait openness is a robust and significant predictor of impactful innovations beyond or alongside indirect measures of open-mindedness used in previous studies. In Model 3a, we test trait openness against the share of creative occupations, which turns out to be insignificant (*"creative"*, b = 0.054, lower = -0.02, upper = 0.127, p = 0.151). The share of bohemians is positively significant (*"bohemians"*, b = 0.086, lower = 0.001, upper = 0.135, p = 0.048) alongside trait openness in Model 3b. Both coefficients show a similar magnitude suggesting that openness and the share of bohemians capture different aspects. The same applies for the share of homosexuals (*"homosexuals"*, b = 0.06, lower = 0.017, upper = 0.103, p = 0.007) in Model 3c, which is also positively significant alongside regional openness. However, openness is in all models the strongest predictor among the proxies suggested by Florida. These results (a) underline the robustness of our results with regards to alternative measures and (b) highlight that open-mindedness approximated by the share of bohemians and homosexuals measures something different than trait openness.

5.4.4 Robustness analysis

To address the direction of effects and to rule out alternative explanations, we applied a number of robustness checks. First, we test additional control variables and the effect of spatial dependencies. Second, we test if selection-bias of the online survey drive the results. Finally, we discuss and analyze endogeneity concerns due to selective migration. We restrict our robustness check to highly impactful innovations as measured by the 10^{th} percentile of the regional impact distribution, since these represent the focus of our study and delivered robust results

Additional control variables and spatial autocorrelation

Some variables are highly correlated and their inclusion in the model would result in multicollinearity. This particularly concerns variables for human capital such as patents and science related occupations (r = 0.71, p < 0.000). We include these alternative variables for human capital and exclude the share of individuals with bachelor degree. Table 5.6 reports the corresponding results. As expected, the number of patents per capita (*"patents"*, b = 0.246, lower = 0.198, upper = 0.293, p < 0.000) in Model 4a as well as the share of science related occupations (*"science"*, b = 0.179, lower = 0.122, upper = 0.237, p < 0.000) in Model 4b are positive and significant. Regions that produce more innovations, i.e. have a larger knowledge stock, are also more likely to produce more impactful ones. Correspondingly, a larger share of scientific occupations is also beneficial for more impactful innovations. The link between openness and high-impact is not affected by the two variables indicating robust results in this regard.

Openness is only one of the Big Five personality traits. We therefore tested, whether our results might be driven by a different personality dimension that partially overlaps with openness (e.g. extraversion). Model 4c, therefore, reports the remaining four Big Five traits alongside openness. Again, the point estimate of openness is smaller in size, but stays significant (b = 0.067, lower = 0.009, upper = 0.124, p = 0.024). The smaller point estimate is not surprising, since the personality traits show substantial correlations and are not orthogonal. For example, regional openness and extraversion are correlated with r = -0.409 (p < 0.000). Among the four remaining traits, only neuroticism (b = -0.052, lower = -0.105, upper = 0.000, p = 0.052) and conscientiousness (b = -0.055, lower = -0.115, upper = 0.004, p = 0.07) show p-values around a significance level of 5 percent. Both traits are negatively associated with innovation activities in regions. Particularly the result of conscientiousness crucially qualifies Lee's (2017) finding, who reported a positive

Table 5.6: Regression results for additional control variables

Y = Highly impactful innovations in regions ($x^{th} = 10^{th}$)			
	Openness + Patents	Openness + Science	Openness + Traits
	(4a)	(4b)	(4c)
openness	0.076^{**}	0.093^{***}	0.067^{*}
	(0.023)	(0.026)	(0.029)
patents	0.246^{***}		
	(0.024)		
science		0.179^{***}	
		(0.029)	
conscientiousness			-0.055
			(0.031)
extraversion			-0.019
			(0.029)
agreeableness			-0.031
			(0.037)
neuroticism			-0.052
			(0.027)
bachelor			0.188^{***}
			(0.029)
stars	0.037^{*}	0.041^{*}	0.033
	(0.018)	(0.020)	(0.018)
lq	-0.091^{***}	-0.117^{*}	-0.134^{***}
	(0.023)	(0.050)	(0.019)
diversity	-0.020	-0.013	-0.033
	(0.024)	(0.027)	(0.026)
manufacturing	-0.088^{***}	-0.023	0.021
	(0.026)	(0.027)	(0.022)
popdens	0.057^{*}	0.069^{*}	0.076^{**}
	(0.026)	(0.028)	(0.027)
constant	2.468^{***}	2.486^{***}	2.468^{***}
	(0.019)	(0.021)	(0.019)
LMT p-value	0.058	0.003	0.016
Observations	382	366	382
Adjusted R^2	0.477	0.395	0.447

Note: **p<0.05; **p<0.01; ***p<0.001.*
Robust standard errors in parentheses.
All independent variables are standardized.

link with pure innovation quantity. Our results thus clearly suggest that openness is most important for the creation of highly impactful innovations. Looking beyond innovation quantity by conditioning on innovations' impact reveals a different picture and highlight that different regional cultures are relevant for different types of innovation outcomes depending on innovations' quality.

Spatial autocorrelation can affect regression results (Anselin 1988). Spatial dependency was an issue in some model specifications as indicated by significant LMT p-values. To explore whether spatial dependencies truly affect our results, we estimate a spatial lag model in which we include the spatially lagged dependent variable as an additional control variable on the right hand side of the equation. That is, innovation activities in neighboring regions can potentially affect innovation activities in the focal region. Table 5.7 reports the corresponding results. The spatial lag (*"rho"*, b = 0.068, lower = 0.025, upper = 0.110, p = 0.002) is positive and significant suggesting that a spatial lag model is the correct choice to address spatial dependency. The significant spatial lag variable also indicates that being located in a cluster of regions with highly impactful innovation outcomes is beneficial. The LMT for spatial autocorrelation also suggests that the spatial lag model reduces spatial dependencies as the corresponding p-value increases to 0.114 compared with a value of 0.004 in Model 2c in Table 5.4. Again, neither the magnitude nor the p-value of regional openness is affected by the spatial lag highlighting the robustness of our results.

Sample skewness and endogeneity

To address the skewness of our sample, we follow previous approaches (Ebert et al. 2019; Stuetzer et al. 2018) and measure openness with sampling weights. To do so, we formed three age groups (18 to 24, 25 to 34; and greater than 34) for both genders. For each region, we then calculated (a) the share of participants in our data and (b) the share of the actual regional population that falls into each of the six category (3 age groups and 2 genders). We divided the actual shares by the shares in our data and then used this ratio as sampling weight when aggregating our data to the regional level. For example, when a participant belongs to a group that is undersampled in our data (e.g. old males) than this observation will receive a weight greater than 1, while a participant belonging to a group that is oversampled (e.g. young females) will receive a weight smaller than 1. Model 6a in Table 5.8 reports the results for the sample-corrected measure of openness. Although the coefficient is smaller in size, openness stays significant (b = 0.063, lower = 0.014, upper = 0.112, p = 0.012). This result suggests that sample bias does not significantly affect our results.

Endogeneity can arise as people with open personalities migrate to innovative MSAs. Innovative regions might attract open people, because they are more likely to meet their personal preferences for novelty and innovation. To account for this endogeneity issue, we follow previous research (Stuetzer et al. 2018) and compute regional openness based on the respondents' residences in their youth before any occupational and migration choices were made. Model 6b documents the corresponding results. The point estimate of openness (b = 0.094, lower = 0.041, upper = 0.148, $p < 0.000$) is slightly smaller compared with the main Model 2c in Table 5.4, but stays significant. This finding suggests that selective migration does not influence our results.

Table 5.7: Regression results for spatial lag model

	Y = Highly impactful innovations in regions ($x^{th} = 10^{th}$)
	Spatial Lag (5a)
openness	0.095***
	(0.022)
bachelor	0.200***
	(0.025)
stars	0.032
	(0.023)
lq	−0.138***
	(0.021)
diversity	−0.028
	(0.026)
manufacturing	0.018
	(0.022)
popdens	0.045
	(0.025)
rho	0.068***
	(0.022)
constant	2.317***
	(0.052)
LMT p-value	0.114
Observations	382

Note: **p<0.05; **p<0.01; ***p<0.001.*
All independent variables are standardized.

Table 5.8: Regression results for sample skewness and selective migration

Y = Highly impactful innovations in regions ($x^{th} = 10^{th}$)		
	Openness Sample Weights (6a)	Openness Youth Residence (6b)
openness	0.063*	0.094***
	(0.025)	(0.027)
bachelor	0.192***	0.189***
	(0.030)	(0.029)
stars	0.044*	0.037
	(0.021)	(0.021)
lq	−0.128***	−0.137***
	(0.019)	(0.015)
diversity	−0.016	−0.018
	(0.026)	(0.026)
manufacturing	0.002	0.013
	(0.022)	(0.021)
popdens	0.064*	0.060*
	(0.026)	(0.026)
constant	2.468***	2.468***
	(0.020)	(0.020)
LMT p-value	0.002	0.009
Observations	382	382
Adjusted R^2	0.417	0.433

Note: **p<0.05; **p<0.01; ***p<0.001.*
Robust standard errors in parentheses.
All independent variables are standardized.

5.5 Discussion and conclusions

In this article, we investigated the role of regional culture, the psychological climate characterized by a strong openness, to explain regional differences in innovation quality as measured with innovations' impact. We used a personality-based approach to overcome the shortcomings of previous research in approximating regional cultures. Our results suggest substantial difference between high-quality innovations, i.e. the most impactful innovations and the average innovation quality regarding the role of regional openness. Our empirical results show that the relationship between regional openness and innovation quality is strongest for highly impactful innovations and less pronounced for the average innovation quality in regions. Our results are robust to a multitude of robustness checks including alternative control variables, spatial modelling, and selective-migration strengthening the substance of our findings.

Our results complement previous findings demonstrating a positive link between openness and innovation at the national level (Steel et al. 2012) and extend these by considering innovations' impact. Considering the quality of innovation might explain previous findings at the regional level that regional openness has no effect on innovation quantity (Lee 2017). The sensitivity of highly impactful innovations to the regional innovation culture might be due to substantial differences between innovations of average and high quality. Highly impactful ideas can transform technological landscapes, give birth to new industries or are responsible for the decline of others with unknown socio-economic consequences. Curiosity and trust in innovation processes with hardly predictable outcomes is a beneficial prerequisite in this context. Impactful ideas might therefore depend more on a local culture, which is open with respect to innovation, than innovations with average impact.

Previous research primarily focused on innovation quantity and less on regional differences in innovation quality. Among the few contributions, Castaldi et al. (2015) reveal that regional capabilities are important to explain innovation quality in regions. Our study adds a different dimensions and shows that regional culture plays a curcial role in explaining the regional variation in innovation quality. Our investigation complements previous research that highlighted the role of regional cultures and regional openness (Florida 2002) but that focused on innovation quantity and used indirect indicators of regional culture. Based on our findings, the personality-based approach, provides a new and robust methodology to overcome the shortcomings of previous research in measuring intangible constructs such as culture and the local psychological climate.

However, there might be complex mechanisms at play – the "special regional features" – that together enable some regions to produce more high-quality innovations than others. Regional capabilities (Castaldi et al. 2015) and culture play an important role (as indicated in our study), but other determinants might be important as well. For example, Cortinovis et al. (2017) highlighted the importance of social capital for regional development. In particular, bridging social capital, which connects diverse groups of actors (Putnam 2001), is an important regional asset that might drive innovation quality. More specifically, regional openness paired with bridging social capital could increase the innovation quality in regions, as regional openness generally facilitates interaction (McCrae 1996), while bridging social capital enhances knowledge exchange between heterogeneous groups. Higher levels of regional openness coupled with higher levels of bridging social capital can facilitate knowledge exchange between diverse groups of actors, which is necessary for high-quality innovation (Schilling and Green 2011; Kim et al. 2016).

Although we applied several robustness checks, our study still has a number of limitations, which are important for future research. Foremost, this concerns the direction of effects. We addressed selective-migration as one crucial source of endogeneity and showed that it does not interfere our results. Besides selective-migration, endogeneity can also arise due to regional innovation quality influencing the institutional endowment of regions, which in turn increases regional openness. For instance, higher innovation quality might improve the regional endowment with education infrastructure, as firms need highly-skilled employees to produce top quality innovations. Educational attainment in turn is a strong predictor of openness (Rentfrow et al. 2008). Hence, it might be the longevity of innovation quality in regions that also impacts regional openness. We included the share of academics, star scientists, and science occupations in regions as control variables for the educational attainment in regions to rule out the potential confounding of education. Openness remained a strong predictor of innovation quality in regions. To address the issue of endogeneity more adequately, however, an instrumental variable regression (IV) represents one possible solution. Finding a valid instrument that fulfills all requirements is a necessary precondition, otherwise the cure can be worse than the disease (Bound et al. 1995). Although we could think of possible instruments, we were not confident that these fulfill the exclusion restriction as such that they improve the empirical investigation. Unraveling the true causal effects of regional openness on innovation quality is important in future research.

We discussed how regional openness can influence innovation activities by referring to the social-impact theory (Latané 1981), which highlights that individual innovators can be influenced in their behavior by the prevalent social norm that is part of the psychological climate in a region. As we analyzed regional aggregates, the interplay of micro scale and macro scale still represents a black-box. Future research could unravel the link between individual and regional level by using a multi-level research design. However, this requires personality data of individual innovators. Such individual-level data are difficult, if not impossible, to obtain, but new big data methods (e.g., social media analyses) might enable a new generation of research in this field (Obschonka and Audretsch 2019). Uncovering the impact of regional cultures on individual behavior could shed light on the social-impact theory at the macro-scale of regions and allows to ask which effect is more important – the individual or regional trait characteristics?

6 Conclusions

6.1 Main empirical findings

Knowledge and new knowledge creation are fundamental building blocks of our society and are linked to economic prosperity. Knowledge production concentrates in particular places, which explains why some regions are economically more successful than others. Consequently, identifying the causes that lead to and assessing the implications that follow from the spatial concentration of knowledge is key to understanding regional development. This dissertation has studied four dimensions of knowledge quality and has analyzed their implications for regional development: relatedness, complexity, novelty and impact. These four dimensions frame the thesis and are at the center of the four research questions that have been investigated in the four central chapters of this dissertation. The findings presented in this thesis for these four dimensions are briefly discussed in the following section, before central implications of the thesis for future research and policy are derived.

6.1.1 Relatedness

Existing research has shown the importance of knowledge relatedness to understand knowledge production as a path-dependent process (Neffke et al. 2011). Regions are unlikely to jump between any activity, but are more likely to develop new competences if these are related to existing ones. The empirical literature shows ample evidence that regions are more likely to diversify into related than into unrelated activities (Neffke et al. 2011; Boschma et al. 2015; Rigby 2015; Essletzbichler 2015). Overcoming a research gap on this issue, Chapter 2 investigated the role that policy plays for regional diversification and asked if policy can break the path dependency of regional diversification. To answer this question, the empirical analysis in Chapter 2 focused on subsidized R&D as one important form of policy intervention in knowledge production processes. The data on public R&D subsidies were obtained from the German Federal Ministry of Education and Research (BMBF) and were linked to technological activities in regions proxied by patent data.

The empirical findings in Chapter 2 indicate that R&D resources are more likely to be allocated to related activities in regions, suggesting that R&D policy plays a crucial role in the path dependency of regional diversification processes. The results further indicate that diversification into new activities is more likely when these activities have received R&D subsidies. Hence, R&D policy plays a crucial role for regional diversification. Previous research has argued that collaborative research has a larger potential to facilitate knowledge exchange between heterogeneous partners than individual projects, and therefore is more likely to lead to successful knowledge production (Broekel and Graf 2012; Broekel et al. 2017). The BMBF data allowed such a distinction to be made between collaborative and individual R&D projects. The results support this view and indicate that collaborative projects contribute to regional diversification to a larger extent than individual projects. A primary research goal was to investigate if policy can break the path dependency of regional diversification. Although policy is more likely to support related diversification, the results obtained in Chapter 2 suggest that the distinction between individual and joint projects is crucially important. R&D subsidies for collaborative projects show a tendency to compensate for missing relatedness and likely facilitate diversification into more unrelated activities, whereas individual projects do not show any effect on unrelated diversification.

The importance of relatedness can help regional policymakers to identify possible diversification options by identifying related activities in which regions have not yet built competitive advantages. Such an approach has entered current policy strategies, as evident in the Smart Specialization framework of the European Union in its 2020 strategy (Foray et al. 2011). However, it can also be questioned whether it is smart to publicly support diversification processes which are most likely to take place without policy intervention. The investigation in Chapter 2 indicates that R&D subsidies are more likely to be allocated to activities related to existing regional competences. It might also be "smart" from a regional policy perspective to support unrelated instead of related diversification. While related diversification rather supports specialization, unrelated diversification broadens the set of regional capabilities into different knowledge domains, which is argued to increase regional resilience against external shocks (Frenken et al. 2007). Collaborative research projects seem to represent a promising tool in this context.

6.1.2 Complexity

Previous research indicates that – in addition to relatedness – knowledge complexity has important implications for knowledge production in regions (Petralia et al. 2017; Balland and Rigby 2017; Balland et al. 2018; Balland et al. 2019). The increasing interest in knowledge complexity in recent years rests, among other factors, on the potential economic value attributed to complex knowledge and its relevance for regional development. However, there are few studies that relate knowledge complexity to regional development. It is this research gap that motivated Chapter 3.

Chapter 3 investigated the geographic patterns of knowledge complexity over time and its role for regional economic growth by building on a sample of 166 European NUTS 2 regions covering the years 2000 to 2015. Mapping regions' complexity of invention activities shows substantial variation between regions, suggesting differences in regions' capability to produce complex knowledge. Large quantities of knowledge in regions do not automatically translate into knowledge complexity. The relatively low correlation coefficient between patent counts and knowledge complexity in regions ($r = 0.27$) suggests some overlaps, but also highlights substantial differences between quantity and quality of knowledge production. The empirical analysis in Chapter 3 reveals that regional differences in knowledge complexity are relatively persistent over time.

The empirical investigation in Chapter 3 linked the spatial distribution of knowledge complexity with regional economic growth. In doing so, GDP per capita is regressed on the complexity of regional invention activities using fixed effects panel estimations. The regression results identify important time lag patterns. Knowledge complexity does not immediately link to economic growth, rather taking 3 to 6 years before gains in regional knowledge complexity translate into regional economic growth. More precisely, a 10 percent increase in regional complexity is associated with a corresponding increase in regional GDP per capita of about 0.28 percent. Taken together, the analysis confirms that regional variances in knowledge complexity are linked to economic growth of regions.

Accordingly, Chapter 3 complements existing works that – implicitly and explicitly – have studied the link between knowledge complexity and economic development. For example, Hidalgo and Hausmann (2009) provide evidence for the economic benefits of economic complexity at the country level. Taking the results at different spatial scales (e.g. countries, regions) for different indicators of knowledge complexity (e.g. economic complexity, technological complexity) together, the current empirical picture therefore suggests that knowledge complexity is associated with economic growth in a systematic way.

6.1.3 Degree of novelty

Previous research has highlighted that new knowledge varies greatly in its degree of novelty (Schilling and Green 2011; Uzzi et al. 2013; Kim et al. 2016). However, differences in novelty have rarely played a central role in economic geography. For instance, the prominent scaling analyses have previously demonstrated the productivity of cities regarding knowledge creation in purely quantitative terms (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007a; Bettencourt et al. 2007b). Chapter 4 shifted the focus of scaling from pure knowledge quantities to the degree of novelty and argued that cities concentrate important features that not only facilitate knowledge production in quantitative terms, but in particular involve the degree of novelty. For instance, the density of heterogeneous knowledge components in larger cities might provide more opportunities to explore new combinations than in smaller, less diverse towns (Youn et al. 2016).

By building on historic patent information of patented inventions in the US between 1836 and 2010, Chapter 4 provides a true long-term perspective on novelty creation in American cities with a particular focus on the relationship between city size and the degree of novelty. Z-score measures reveal the degree of novelty in knowledge combinations and treat novelty as a continuum ranging from very atypical (i.e. radically new combinations) to very typical (i.e. established combinations).

The empirical findings of Chapter 4 first reveal a linear relationship between technological diversity and city size as indicated by the obtained scaling coefficient. These results complement the study by Youn et al. (2016), who also find a linear relationship between city size and diversity of business activities. Accordingly, inventors in larger cities can draw locally from much more heterogeneous knowledge components than inventors in smaller towns, which allows for greater possibilities to explore and realize newer combinations. Importantly, Chapter 4 highlights that the linear scaling of technological diversity with city size is dependent on the depth of technological disaggregation. That is to say, higher levels of technological aggregation tend to hide much of the diversity existing in the largest cities. As shown in

Chapter 4, the 4-digit CPC level is not capable of representing the diversity found in the largest cities, as the scaling coefficient clearly suggests a sublinear scaling when depending on the 4-digit CPC. This result would suggest that larger cities are more diverse, but that diversity is not increasing disproportionately with city size. Using more fine-grained aggregate levels instead systematically shifts the relationship between technological diversity and city size towards linearity. In short, it is important for empirical research on which scale technological diversity is assessed.

The systematic relationship between technological diversity and city size might explain the main findings of Chapter 4 that atypical combinations increasingly concentrate in larger cities. Atypical combinations began to scale super-linearly with city size at the beginning of the 20th century and continued to increase until the 1970s. Since then, the scaling has remained on a constant level without showing any significant variation towards or away from further spatial concentration. Hence, Chapter 4 complements existing scaling analyses (O'hUallichain 1999; O'hUallichain and Leslie 2005; Bettencourt et al. 2007a; Bettencourt et al. 2007b) by indicating that cities' productivity in terms of knowledge production is not only restricted to pure quantity, but also includes novelty.

6.1.4 Impact

The last dimension studied in this thesis was impact. Previous research has shown that impactful innovations concentrate more strongly in space than conventional innovations (Ejermo 2009; Castaldi and Los 2017). Chapter 5 took the observed geographic variation as the primary motivation and asked for the underlying reasons why some regions are more capable of producing impactful knowledge outcomes than others. More precisely, Chapter 5 investigated the role of regional openness towards innovation for impactful innovations in regions and was particularly inspired by the study of Lee (2017) in which he relates the big five personality traits, which include openness, to regional innovation activities. Lee did not find any significant relationship between regional openness and innovation in the UK. It was argued that Lee's finding might be due to the exclusive focus on innovation production in purely quantitative terms.

Innovations are, however not a homogeneous quantity, but rather show substantial variation in several qualitative dimensions, of which impact is a crucial one. Impactful innovations deviate from less impactful innovations in a number of characteristics, which are fundamental. For example, impactful innovations require social climates that value creativity, originality, and out-of-the-box thinking (Sandberg and Aarikka-Stenroos 2014), which are attributes captured by openness. Chapter 5 thereby complements existing research in economic geography, in particular the work of Florida (2002) on the creative class, as open-mindedness is an explicit building-block of tolerance. However, due to a lack of systematic methodologies and data sources, Florida used rather indirect proxies such as the share of creatives, bohemians and homosexuals to assess regional levels of tolerance. It was therefore a second aim of Chapter 5 to overcome the indirect approaches of previous research by assessing regional openness using a personality-based approach.

Data on regional openness was obtained from the Internet-Personality Project using the Big-Five inventory (Gosling et al. 2004). The empirical analysis in Chapter 5 used patent citation data from the USPTO to approximate impactful innovations in metropolitan regions. In particular, impactful innovations were calculated on a continuous scale ranging from low-impact to high-impact based on the regional impact distribution. Assessing impactful innovations as a continuum rather than a homogeneous group as in previous approaches (Ejermo 2009; Castaldi et al. 2015; Castaldi and Los 2017) allowed an investigation of the relationship between regional openness and impactful innovations in more detail.

The results show that regional openness is particularly important for the most impactful innovations. The estimated effect sizes gradually decrease when shifting the impact distribution to less impactful innovations, indicating that openness is less important for average innovations. These results provide an explanation for the findings of Lee (2017) and therefore highlight that quantity and quality of knowledge production are not identical. It is therefore important to include impact in empirical analyses in order to study innovation beyond pure quantity. Secondly, the findings also emphasize the important role of regional openness to explain the geographic variation in impactful innovations observed by Castaldi and Los (2017). Including openness explained an additional 6.9 percent of the regional variance of impactful innovations beyond the standard set of economic control variables. In a robustness check, the empirical investigations tested regional openness measured with personality trait data against Florida's tolerance indicators, i.e. share of bohemians, homosexuals and creatives. The findings suggest that regional openness is related to the tolerance indicators, but it is not identical. Moreover, it was shown that regional openness either outperforms Florida's indicators in predicting impactful innovations or explains additional variance alongside tolerance. Hence, these results suggest that geographical psychology more generally and a personality-based approach more specifically can greatly inform research in economic geography by providing methodologies to capture intangible variables such as social norms, values and cultures (Obschonka 2017; Obschonka and Audretsch 2019).

6.2 Implications for future research

This dissertation highlighted that the consideration of knowledge quality in empirical research on knowledge production greatly improves our understanding of regional differences in innovative success, and that these differences are significant for their economic development. However, it is important to emphasize a number of limitations to acknowledge the boundaries of this dissertation and to indicate opportunities for future research.

Linking novelty and complexity – does complexity makes novelty creation more difficult?

Each empirical chapter was devoted to one of the four quality dimensions outlined in the introductory Chapter 1. Hence, the quality dimensions have been treated more in isolation and less in relation to each other. However, they are not necessarily independent of each other and show, conceptually as well as empirically, some significant overlaps. For example, coupling novelty and complexity has the potential to inform the decreasing research productivity observed by Jones (1995) and other empirical studies. Moore's law, for example, states that "the number of transistors packed onto a computer chip doubles approximately every two years" (Bloom et al. 2017, p. 2). However, the corresponding growth has only been reached by an increasing number of researchers that push Moore's law forward (Bloom et al. 2017). Larger team sizes are not restricted to semi-conductors, but are more generally observed in science and technology (Wuchty et al. 2007; Broekel 2019). Adding to this are the findings of Griliches (1994) that research spending per patent is increasing. Accordingly, the growth of inventive outcomes might be just the result of more R&D inputs, and the empirical evidence suggests that the efficiency of this process is decreasing.

Pintea and Thompson (2007) link this decreasing research productivity to increasing complexity over time, with the latter being empirically confirmed by Broekel (2019). So far, however, there is no empirical study that directly links the two phenomena. Accordingly, is research productivity decreasing because of complexity, or asked differently, are novel ideas becoming harder to find because of complexity?

If there is a positive relationship between complexity and research productivity, this might also suggest that the increasing complexity has slowed down novelty creation in cities since 1970, which provides an explanation for the empirical findings of Chapter 4. As with novelty, complexity also shows substantial variations between regions, with some regions producing more complex knowledge than others (see Chapter 3). Even more so, there is a trend that more complex knowledge increasingly concentrates in cities (Balland et al. 2018). This suggests that cities not only concentrate the functionalities and skills to produce more novelty, but also to advance complex knowledge domains. However, similarly to the decrease in scaling of novelty since the 1970s, as observed in the empirical analysis in Chapter 4, the scaling of knowledge complexity slows down around the same time period with only marginal increases since the 1970s (Balland et al. 2018). This raises the question of whether the two observations are linked or even consequential to each other. Has the increase in complexity slowed down novelty creation in cities? Or do the two dimensions evolve in a co-evolutionary process? This surely represents an interesting avenue for future research.

The role of formal institutions for knowledge quality - is governmental support required to manage ever-increasing complexity?

The increasing complexity of technologies paired with the difficulty to produce novelty constitutes a strong argument to call for more government support in knowledge production. In general, the state plays an important role for collective learning and innovation. The Apollo mission, for example, would not have been possible without government intervention. Although largely motivated by reasons of national prestige and interest, the research efforts during the Apollo mission fueled the development of crucial technologies that affected our society, economy and technological development far beyond the mission (Mazzucato 2014). Google's Page Rank algorithm, as another example, was partly financed by government support (Fleming et al. 2019). Government-funded patents receive on average more citations and tend to introduce greater novelty (Fleming et al. 2019). The results in Chapter 2 provide evidence that federally funded R&D projects can increase the success of regional diversification and that collaborative research, to some extent, can break the strong path dependencies of regional diversification. Hence, government support can play an important role for the quality of knowledge production.

The share of governmental funding in science and technology has steadily increased since World War II in the United States. In 1945, the share of public funding on all patented inventions in the US grew from 5 percent in 1945 to 30 percent in 2016 (Fleming et al. 2019). Knowledge production seems to increasingly rely on governmental support, either directly with the state as the primary actor (state conducts research) or indirectly as a crucial knowledge source (private research relies on public research). But why is this so? Complexity might be one of the explanations, as the increase of governmental support corresponds to an increase in technological complexity observed by Broekel (2019) for the same time period. As increasing complexity demands more research efforts, governmental support might be necessary to manage the increasing complexity and the corresponding difficulty to produce novel outcomes. Hence, the question remains as to whether government support of knowledge production is the answer to novelty creation having become more complex, and if so - to what end? Future research should start from these correlations and analyze the possible relationship between governmental support in certain knowledge domains and their complexity in more detail.

The role of informal institutions for knowledge quality and diversification

Governmental intervention in knowledge creation highlights the role of formal institutions. Besides formal institutions, informal institutions are also fundamental for knowledge creation (Amin 1999; Rodríguez-Pose and Storper 2009; Rodríguez-Pose and Di Cataldo 2015) and might impact the quality of created knowledge. However, the inherent fuzzyness of informal institutions represents an empirical challenge to investigate their role for innovation. As presented in Chapter 5, the emerging field of geographical psychology offers systematic and robust methodologies to assess the personalities of individuals in regions, representing one opportunity to capture informal institutions in the form of social norms and social values or cultures. In a first attempt, Obschonka et al. (2015) use the prevalence of personality traits in regions to assess entrepreneurial cultures in US metro areas and show their importance for explaining regional patterns of entrepreneurial activities. Chapter 5 represents another step in this direction by using regional openness to approximate the regional attitude towards (or against) innovation as a hidden informal institution. It was shown that regional openness is highly relevant for the emergence of impactful innovations in regions.

However, the underlying mechanisms driving the results in Chapter 5 are difficult to identify and rather indirect, leaving much room for future research on the role of informal institutions and knowledge quality. Conducting empirical analysis exclusively at the regional level has the disadvantage of hiding crucial mechanisms that either unfold at the individual level or are the result of the interplay between the individual and the regional level. As there is empirical evidence at the micro level of individuals and innovation performance (Zwick and Frosch 2017), the interaction between individuals and their socio-spatial environment still remains a black box. As argued in Chapter 5, one possible explanation for the importance of regional openness for impactful innovations might be that individuals are influenced in their own behavior by their local social environment irrespective of their natural disposition (Rentfrow et al. 2008). Hence, if less open innovators are surrounded by many open people, they are likely to adopt the prevalent social norms and values associated with openness in the region, which in turn might impact their innovation outcomes. Although this explanation is backed by the social-impact theory (Latané 1981), the interplay of individuals with their personality and their regional environment has not been studied with respect to regional innovation activities. One major obstacle in this context is data availability. The interplay between individuals and their social environment represents a multi-level problem that requires personality data at the individual level of innovators and the regional level. Such data is difficult but not impossible to obtain. Future research with access to big data at the individual level might be able to overcome this issue.

Besides innovation, informal institutions also impact regional diversification processes. For example, Cortinovis et al. (2017) analyze the role of bonding and bridging social capital as informal institutions in the context of regional diversification. Their findings suggest that bridging rather than bonding social capital influences the likelihood of a region diversifying into new activities. However, these authors also show that social capital cannot break the dominant role of relatedness (and thereby path dependency), although there are good arguments as to why bridging social capital, in particular, should be important for unrelated diversification. Bridging social capital shows the tendency to link heterogeneous groups of actors and facilitates information diffusion (Putnam 2001). This seems to be insufficient though to overcome path dependencies, and hence the question of how to accomplish unrelated diversification remains an important issue for future research.

Bridging heterogeneous groups of actors might also require openness of the local people to become engaged in such bridging interactions to facilitate unrelated diversification. Trait openness summarizes aspects that might not only be important for impactful innovations, but also for unrelated diversification processes. Regional openness might establish informal institutions in regions that allow them to jump in their technological evolution. In particular, an open regional culture values novelty, originality and creativity (McCrae 1996). These represent key characteristics that facilitate the combination of hitherto disconnected activities in regions. As openness also facilitates peoples' engagement in social relations and interaction (McCrae 1996), regional openness can also stimulate knowledge exchange between heterogeneous actors enhancing local knowledge diffusion. However, communication and particularly knowledge sharing requires common trust between different groups of actors, which is not guaranteed by openness alone. However, bridging social capital might be able to accomplish this (Putnam 2001). Hence, it might require both higher levels of regional openness and the existence of bridging social capital to overcome the forces of relatedness for regional diversification. Pairing social capital with a personality-based approach therefore offers a promising future research field to study unrelated diversification.

The interplay of openness and bridging social capital is not only important in the context of unrelated diversification, but also when it comes to the management of complex knowledge production. The Apollo program employed 400,000 people in over 20,000 private organizations, universities and research institutions (Gisler and Sornette 2009). Organizing and coordinating complexity thus demands the combination of countless heterogeneous components involving strong communication and interaction between various groups of actors from different organizations. Higher levels of bridging social capital in combination with openness is therefore likely to stimulate the production of complex knowledge by facilitating effective communication between heterogeneous actors. However, these relationships are still speculative and hence represent a promising venue for future research.

Impactful innovations as an outcome of relatedness, novelty and complexity in regions?

Chapter 5 placed impactful innovation at the center of attention and presented regional openness as an explanation of the observed variation in the creation of impactful innovations among regions. Until now, economic geography has not paid much attention to impactful innovations or related concepts such as superstar patents (Castaldi and Los 2017). Among the few existing studies, Castaldi et al. (2015) analyzed the importance of regional capabilities for the emergence of so-called superstar patents. Their study shows that an economic structure characterized by unrelated variety is particularly conducive for the development of high-impact innovations. This argument harkens back to the advantage of Jacobs externalities that local diversity facilitates cross-fertilization between different knowledge domains. However, even jointly, the results of Chapter 5 and the findings of Castaldi et al. (2015) present only initial insights into the subject of impactful innovations.

Impactful innovations might be an outcome which is characterized and driven by the other three dimensions of knowledge quality discussed in this dissertation. For instance, a large stream of inter-disciplinary research investigates the underlying mechanisms of impactful innovation from the perspective of knowledge recombination. Knowledge combinations play a particular role, as they are a strong predictor for the later impact of inventions. More precisely, new combinations connecting formerly disconnected components. i.e. atypical combinations, are a fundamental building block of high-impact outcomes in science and technology. In contrast, typical combinations linking related knowledge rather translate into less impactful outcomes (Schilling and Green 2011; Uzzi et al. 2013; Arts and Veugelers 2015; Kim et al. 2016). One can also establish a link between impactful innovation and knowledge complexity. Fleming and Sorenson (2001) demonstrate that complex patents represent a critical source for subsequent knowledge creation processes. Accordingly, more complex patents receive more citations, which signals their impact. Hence, the impact of knowledge production outcomes might be a consequence of the other three dimensions of knowledge quality and their interplay. In light of the results of the empirical analyses in this thesis, it seems plausible to argue that all the other three quality dimensions (relatedness, complexity, novelty) show substantial regional variation and might contribute to regions' ability to produce impactful knowledge. So far, however, each dimensions has been analyzed mostly in isolation from the others, calling for empirical studies considering the interaction of these dimensions.

Novelty and impact as drivers of economic development?

As previously mentioned, the degree of novelty is an essential feature of innovations' impact (Schilling and Green 2011; Uzzi et al. 2013; Arts and Veugelers 2015; Kim et al. 2016). However, both have been analyzed independently in this dissertation. Chapter 4 assessed the novelty in knowledge combinations without considering their impact on subsequent combinations or on regional development, and Chapter 5 studied impactful innovations without considering their degree of novelty. However, not every new combination is useful for subsequent developments and will translate into economic benefits. It could also be the case that some new combinations are in "advance of their time" and will only unfold their economic potential with a significant time delay (Raan 2004). Therefore, impact might be an important indicator for distinguishing less impactful from impactful combinations. Accordingly, measuring the impactfulness of knowledge combinations and linking it to regional economic development represents a crucial research opportunity for future investigations.

Similarly, the effects of impactful innovation for regional (economic) development are largely unexplored. In Chapter 5, it was implicitly assumed that more impactful innovations have a larger economic impact. The local success stories of impactful innovations, such as the invention of the automobile or Google, presented in the introduction to this thesis, provide anecdotal evidence that impactful ideas have the potential to stimulate subsequent regional development. However, we still know little about the extent to which novel and impactful outcomes may contribute to regional development.

Larger numbers of impactful innovations in regions might have a positive effect on the regional economy via several mechanisms. Firstly, ideas are one of the most crucial sources for entrepreneurial activities. The invention of the PageRank algorithm by Brin and Page (1998) led to the foundation of Google, which now employs thousands of people in the Bay Area. Higher levels of new firm creation in regions are, in turn, associated with higher levels of economic growth (Praag and Versloot 2008). Secondly, impactful innovations of existing firms in regions are positively associated with economic value and thus might induce regional growth. For example, Hall et al. (2005) showed that patent citations as a measure of patents' impact are positively associated with firm market values. Thirdly, impactful innovations carry the potential to offer new growth paths by creating new markets and industries (Chandy and Tellis 1998; Chandy and Tellis 2000). This dissertation, however, has not looked at the economic importance of impactful innovations. In light of this, studying the economic benefits of impactful innovations for regional development appears to be another promising topic for future research.

What makes complexity economically valuable?

The empirical analysis presented in Chapter 3 indicates that knowledge complexity is associated with economic growth in regions. The theoretical reasoning for this is based on the idea of complex knowledge's spatial stickiness (Kogut and Zander 1992). Complex knowledge is difficult to learn and copy, and is hence less likely to spill over to competitors. Therefore, complex knowledge usually gives its owners a temporary monopoly, which may translate into competitive advantages (Kogut and Zander 1992; Zander and Kogut 1995). How this actually happens and how the economic benefits of complexity unfold at the regional level, however, remains a black box. It is possible that complex knowledge gives firms in region A a competitive advantage over competitors in other regions. As a result, firms with complex knowledge in region A experience higher growth rates than firms in other regions. But is knowledge complexity only valuable because it resists a fast dissemination to other places, or because it provides more opportunities for future advancements (Hidalgo and

Hausmann 2009), or are there even other forces at play?

Another possible explanation for the superior economic value of complex knowledge may lie in its dependence on large investments and skilled individuals. Huntsville in Alabama, for example, has been endowed with NASA's Marshall Space Flight Center (MSFC) since 1960. It was set up to plan and construct the Saturn rocket that eventually brought Aldrin, Armstrong and Collins to the moon. Houston was equipped with the Lyndon B. Johnson Space Center in 1961 to conduct NASA's training, research and flight control for its human spaceflight programs, including Apollo. Such organizations represent large public investments with direct effects on local employment. During the Apollo program, about 7,500 people worked at MSFC's facilities in Huntsville. Many of the employees were not based in Huntsville, but moved there to work at the MSFC. Effects on employment were not exclusively restricted to the MSFC, but also involved contracting firms in the region. For example, Brown engineering grew from a small local firm to a prominent aerospace engineering corporation. In other cases, the new technical expertise in the region encouraged entrepreneurial activities. For instance, a computer scientist who came to Huntsville to work on the Saturn rocket founded Intergraph, which became a global computer and software firm (Dunar and Waring 1999). Hence, these stories might be related to the complexity of the knowledge involved in the projects, which required large investments that gave rise to employment and economic effects, which via strong multipliers stimulated local growth. While plausible, it is not clear whether similarly large investments focused on rather simple knowledge would have had the same type of effects. Accordingly, there are multiple alternative explanations for the observed empirical findings in Chapter 3. Each has distinct implications. Consequently, it is imperative to unravel the true working forces of complexity in more detail in future research.

How to measure knowledge complexity?

Complexity research in economic geography is still at an early stage, and there are a number of caveats which need to be addressed before this concept can be used to inform policymakers. Complexity still remains conceptually and empirically elusive. There are different measures which are based on different theoretical foundations, rely on different indicators, and use different data sources to capture knowledge complexity. The economic complexity indicator developed by Hidalgo and Hausmann (2009) relies on export products, whereas Balland et al. (2018) use the average age of education of employees in an industry to indicate economic complexity. In the same article, Balland et al. (2018) also measure complexity in scientific activities based on team sizes of authors in a scientific publication and technological complexity based on the vintage of knowledge combined in patents. Furthermore, Balland and Rigby (2017) deploy the method of Hidalgo and Hausmann (2009) to patent data to assess technological complexity with the so-called *Knowledge Complexity Indicator* (KCI). Fleming and Sorenson (2001) developed *Modular Complexity* and Broekel (2019) introduced *Structural Diversity* to indicate the complexity of patenting activities.

Clearly, it is principally fruitful for a research field when there are empirical opportunities to study a specific topic, but it also makes it difficult to compare and reproduce empirical results. For example, Chapter 3 relied on the measure of Structural Diversity to assess the link between technological complexity and economic growth, but also included robustness checks using alternative measures such as the KCI and Modular Complexity yielding different results. Critically reflecting on existing complexity measures is therefore crucial in future research, as the empirical results and derived implications depend on it. In addition, we still know little about the interplay between different complexity indicators, i.e. economic and technological complexity. For example, technological and economic complexity are often studied in isolation, but there might be strong interdependencies that could also inform regional policy and provide a more comprehensive picture of how complex technological and economic activities affect each other and influence regional development.

6.3 Policy implications

The complex system of regional policy, in the EU in particular, involves many spatial scales, numerous programs with a multitude of targets, and countless actors. Although the four empirical chapters in this thesis covered a relatively broad range of topics from regional diversification to technological complexity and economic growth, this dissertation still has a narrow scope in comparison to such policies. The focus was on knowledge creation as one crucial aspect, but clearly not the only aspect of regional development. It is therefore important to keep in mind that all policy implications put forward on the basis of this thesis only apply to regions that produce (technological) knowledge and for which this is an essential ingredient to their economic development. A second restriction is that this dissertation only dealt directly with policy in Chapter 2, while it was considered rather indirectly in Chapter 3. The two remaining Chapters 4 and 5 did not consider the policy dimension. Nevertheless, the theoretical arguments and empirical results of this dissertation provide some implications that support place-based approaches, which consider the heterogeneity of places and knowledge.

Chapter 3 has investigated the economic importance of technological complexity for regional economic development, which links it to Chapter 2 that analyzed the interplay of relatedness and public R&D support. Pairing complexity with relatedness (i.e. two dimensions of knowledge quality) has recently stimulated works in economic geography. Moreover, it represents a basis from which to derive policy implications. While relatedness reveals the ease of diversification by relying on existing capabilities, complexity gives the potential economic benefits of specific diversification directions. Accordingly, it can be argued that regions should diversify into complex but related technologies because this strategy offers the most promising opportunities (Balland et al. 2019). As shown in existing works and in this thesis, both dimensions can be measured using large-scale data, which then can be used to derive customized place-based strategies for regions. Despite a continued lack of empirical support, this strategy has already entered and shaped policy debates (Balland et al. 2019). In this context, Chapter 3 adds some much needed empirical evidence for the economic growth potentials of technological complexity.

However, such a strategy also relates to a number of central issues of regional policy. Firstly, when and how should policy intervene in the process of regional development? Learning complex activities is inherently slow and difficult and requires more resources than simple activities. Accordingly, they are rather costly but tend to be more risky. This may result in market failures and therefore justifies policy intervention. Given the need to approach complex problems in a collaborative fashion, supporting collaborative projects in complex areas seems to be an adequate policy recommendation. However, much of this is rather speculative and empirical research in this context is still sparse. The extent to which policy can facilitate complex knowledge production thus remains largely unknown.

A second central issue refers to related and unrelated diversification. As related diversification appears to be the norm rather than the exception, the question can be raised as to whether policy support should focus on unrelated instead of related activities in regions. Accordingly, policy can support activities that are rather unlikely to happen, i.e. unrelated activities. In this case, policy contributes to a broadening of the regional knowledge space. Such a policy has the potential to increase regional resilience against external shocks (Frenken et al. 2007) and to contribute to increasing the quality of regional knowledge production outcomes (Castaldi et al. 2015). Accordingly, such a policy requires the identification of the path-dependent developments in regions and the derivation of strategies that overcome these. Chapter 2 showed that R&D subsidies for collaborative research projects can to some extent compensate for missing relatedness, representing a promising policy tool to support unrelated activities. However, as was shown in previous research, R&D subsidies (at least in Germany) are primarily allocated to individual projects (Broekel and Graf 2012). Considering the effects of collaborative research revealed in Chapter 2 suggests to rethink the prevalent allocation strategies of R&D subsidies in Germany. In addition, Chapter 2 revealed that R&D policy primarily supports related activities in regions, suggesting that policy is rather part of the path-dependent development of regions for which there are good reasons. Supporting unrelated activities brings the danger of building cathedrals in the desert. Allocating public resources to related activities, in contrast, has a greater chance to be successful and hence justifies allocation decisions of policymakers.

Although related diversification seems to be the norm, not all regions might be equally successful in diversifying into related activities. Related diversification rather supports the further specialization into related activities providing regions with potentials to build competitive advantages. In particular, regions with a narrow set of capabilities are likely to have disadvantages, as they have fewer opportunities for related diversification. In such cases, policy might be advised to support related instead of unrelated diversification. This debate shows that there is no "one-size-fits-all" policy (Tödtling and Trippl 2005). Unfortunately, few empirical works have disentangled the extent to which related and unrelated diversification contribute to regional development. One study by Pinheiro et al. (2018) is a first step in this direction. Their findings suggest that unrelated diversification is associated with higher levels of economic growth in regions. However, more empirical investigations are necessary in future research.

This dissertation marks an important step forward, as it has emphasized that knowledge quality is a key feature of new knowledge production often overlooked in previous research. The first printing press, the first automobile, the first powered flight and the first manned moon landing stand out in the history of technological development and illustrate that new knowledge substantially varies along multiple dimensions. The focus of this thesis was therefore on four important dimensions: relatedness, complexity, degree of novelty and impact, which acknowledge the multi-dimensionality of knowledge quality often overlooked in previous research. The heterogeneity of places is not only important for differences in the production of new knowledge in pure quantitative terms, as often highlighted in existing studies, but also crucially important to explain differences between places regarding the quality of new knowledge. It has been shown throughout this thesis that the four quality dimensions substantially vary between regions. Moreover, these qualitative differences of knowledge production between regions are crucial to explain the uneven development of places, for instance in terms of economic growth and collective learning. Policy should therefore increasingly include knowledge quality to design place-based development strategies. However, the empirical boundaries and the limitations of this thesis to derive concrete policy implications clearly show that much more research is needed. For instance, it was discussed that interdependencies between the quality dimensions can be expected and are likely to inform our understanding of knowledge quality and regional development. In addition, the four dimensions studied in this dissertation are not exclusive. There might be other dimensions of knowledge quality that have not been discussed yet and that might become important building blocks of future research. Further investigation is needed into how knowledge quality can improve our understanding of regional development, and better policies must be tailored to this.

Bibliography

Abramovitz, M. (1956). "Resource and Output Trends in the United States Since 1870". In: *The American Economic Review* 46.2, pp. 5–23.

Acs, Z. J., L. Anselin, and A. Varga (2002). "Patents and innovation counts as measures of regional production of new knowledge". In: *Research Policy* 31.7, pp. 1069–1085. DOI: 10.1016/S0048-7333(01)00184-6.

Acs, Z. J. and M. I. Megyesi (2009). "Creativity and industrial cities: A case study of Baltimore". In: *Entrepreneurship & Regional Development* 21.4, pp. 421–439. DOI: 10.1080/08985620903020086.

Aghion, P. and P. Howitt (1998). *Endogenous growth theory*. Cambridge, Mass.: MIT Press. 694 pp.

Ahuja, G. and C. M. Lampert (2001). "Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions". In: *Strategic Management Journal* 22.6, pp. 521–543. DOI: 10.1002/smj.176.

Allik, J., A. Realo, R. Mõttus, H. Pullmann, A. Trifonova, R. R. McCrae, and and 56 Members of the Russian Character and Personality Survey (2009). "Personality traits of Russians from the observer's perspective". In: *European Journal of Personality* 23.7, pp. 567–588. DOI: 10.1002/per.721.

Amin, A. (1999). "An Institutionalist Perspective on Regional Economic Development". In: *International Journal of Urban and Regional Research* 23.2, pp. 365–378. DOI: 10.1111/1468-2427.00201.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

Anselin, L., A. Varga, and Z. Acs (1997). "Local Geographic Spillovers between University Research and High Technology Innovations". In: *Journal of Urban Economics* 42.3, pp. 422–448. DOI: 10.1006/juec.1997.2032.

Arbesman, S., J. M. Kleinberg, and S. H. Strogatz (2009). "Superlinear scaling for innovation in cities". In: *Physical Review E* 79.1. DOI: 10.1103/PhysRevE.79.016115.

Arcaute, E., E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty (2014). "Constructing cities, deconstructing scaling laws". In: *Journal of The Royal Society Interface* 12.102, pp. 20140745–20140745. DOI: 10.1098/rsif.2014.0745.

Aronson, Z. H., R. R. Reilly, and G. S. Lynn (2008). "The role of leader personality in new product development success: an examination of teams developing radical and incremental innovations". In: *International Journal of Technology Management* 44.1, p. 5. DOI: 10.1504/IJTM.2008.020696.

Arrow, K. J. (1962). "The Economic Implications of Learning by Doing". In: *The Review of Economic Studies* 29.3, p. 155. DOI: 10.2307/2295952.

Arthur, W. B. (2009). *The nature of technology: what it is and how it evolves*. OCLC: 6ocm14315023. New York: Free Press. 246 pp.

Arts, S. and R. Veugelers (2015). “Technology familiarity, recombinant novelty, and breakthrough invention”. In: *Industrial and Corporate Change* 24.6, pp. 1215–1246. DOI: 10.1093/icc/dtu029.

Aschhoff, B. (2008). *Who Gets the Money? The Dynamics of R&D Project Subsidies in Germany*. 08-018. Centre for European Economic Research, pp. 1–37.

Aubert, C., O. Falck, and S. Heblich (2011). “Subsidizing National Champions: An Evolutionary Perspective”. In: Falck, O. *Industrial Policy for National Champions*. Ed. by C. Gollier and L. Woessmann. The MIT Press, pp. 63–88. DOI: 10.7551/mitpress/9780262016018.003.0004.

Audretsch, D. B. and M. P. Feldman (1996). “R&D spillovers and the geography of innovation and production”. In: *The American Economic Review* 86.3, pp. 630–640.

Audretsch, D. B. and M. Keilbach (2008). “Resolving the knowledge paradox: Knowledge-spillover entrepreneurship and economic growth”. In: *Research Policy* 37.10, pp. 1697–1705. DOI: 10.1016/j.respol.2008.08.008.

Aunger, R. (2010). “Types of technology”. In: *Technological Forecasting and Social Change* 77.5, pp. 762–782. DOI: 10.1016/j.techfore.2010.01.008.

Bacolod, M., B. S. Blum, and W. C. Strange (2009). “Skills in the city”. In: *Journal of Urban Economics* 65.2, pp. 136–153. DOI: 10.1016/j.jue.2008.09.003.

Bahar, D., R. Hausmann, and C. A. Hidalgo (2014). “Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?” In: *Journal of International Economics* 92.1, pp. 111–123. DOI: 10.1016/j.jinteco.2013.11.001.

Bahnsen, U. (2019). “Forscher wollen das Altern besiegt haben”. In: *Zeit Online*.

Balland, P.-A. (2017). *EconGeo: Computing Key Indicators of the Spatial Distribution of Economic Activities*.

Balland, P.-A., R. Boschma, J. Crespo, and D. L. Rigby (2019). “Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification”. In: *Regional Studies* 53.9, pp. 1252–1268. DOI: 10.1080/00343404.2018.1437900.

Balland, P.-A., C. Jara-Figueroa, S. Petralia, M. Steijn, D. Rigby, and C. A. Hidalgo (2018). “Complex Economic Activities Concentrate in Large Cities”. In: *Papers in Evolutionary Economic Geography* 18.29.

Balland, P.-A. and D. Rigby (2017). “The Geography of Complex Knowledge”. In: *Economic Geography* 93.1, pp. 1–23. DOI: 10.1080/00130095.2016.1205947.

Barajas, A., E. Huergo, and L. Moreno (2012). “Measuring the economic impact of research joint ventures supported by the EU Framework Programme”. In: *The Journal of Technology Transfer* 37.6, pp. 917–942. DOI: 10.1007/s10961-011-9222-y.

Barrell, R. and N. Pain (1997). “Foreign Direct Investment, Technological Change, and Economic Growth within Europe”. In: *The Economic Journal* 107.445, pp. 1770–1786. DOI: 10.1111/j.1468-0297.1997.tb00081.x.

Bathelt, H., A. Malmberg, and P. Maskell (2004). “Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation”. In: *Progress in Human Geography* 28.1, pp. 31–56. DOI: 10.1191/0309132504ph469oa.

Beaudry, C. and A. Schiffauerova (2009). “Who’s right, Marshall or Jacobs? The localization versus urbanization debate”. In: *Research Policy* 38.2, pp. 318–337. DOI: 10.1016/j.respol.2008.11.010.

Becker, G. S. and K. M. Murphy (1992). "The Division of Labor, Coordination Costs, and Knowledge". In: *The Quarterly Journal of Economics* 107.4, pp. 1137–1160. DOI: 10.2307/2118383.

Becker, G. S., E. L. Glaeser, and K. M. Murphy (1999). "Population and Economic Growth". In: *American Economic Review* 89.2, pp. 145–149. DOI: 10.1257/aer.89.2.145.

Benet-Martínez, V. and O. P. John (2000). "Toward the development of quasi-indigenous personality constructs - Measuring Los Cinco Grandes in Spain with indigenous Castilian markers". In: *American Behavioral Scientist* 44.1, pp. 141–157.

Bettencourt, L. M., J. Lobo, and G. B. West (2008). "Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics". In: *The European Physical Journal B* 63.3, pp. 285–293. DOI: 10.1140/epjb/e2008-00250-6.

Bettencourt, L. M., J. Lobo, D. Helbing, C. Kuhnert, and G. B. West (2007a). "Growth, innovation, scaling, and the pace of life in cities". In: *Proceedings of the National Academy of Sciences* 104.17, pp. 7301–7306. DOI: 10.1073/pnas.0610172104.

Bettencourt, L. M., J. Lobo, and D. Strumsky (2007b). "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size". In: *Research Policy* 36.1, pp. 107–120. DOI: 10.1016/j.respol.2006.09.026.

Blanes, J. V. and I. Busom (2004). "Who participates in R&D subsidy programs?" In: *Research Policy* 33.10, pp. 1459–1476. DOI: 10.1016/j.respol.2004.07.006.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2017). *Are Ideas Getting Harder to Find?* w23782. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w23782.

BMBF (2014). *Bundesbericht Forschung und Innovation 2014*. Bundesministerium für Bildung und Forschung (BMBF).

Bosch, C. (1932). "The Development of the Chemical High Pressure Method During the Establishment of the New Ammonia Industry". Nobel Lecture.

Boschma, R. (2005). "Proximity and Innovation: A Critical Assessment". In: *Regional Studies* 39.1, pp. 61–74. DOI: 10.1080/0034340052000320887.

Boschma, R., P.-A. Balland, and D. F. Kogler (2015). "Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010". In: *Industrial and Corporate Change* 24.1, pp. 223–250. DOI: 10.1093/icc/dtu012.

Boschma, R., L. Coenen, K. Frenken, and B. Truffer (2017). "Towards a theory of regional diversification: combining insights from Evolutionary Economic Geography and Transition Studies". In: *Regional Studies* 51.1, pp. 31–45. DOI: 10.1080/00343404.2016.1258460.

Boschma, R. and K. Frenken (2006). "Why is economic geography not an evolutionary science? Towards an evolutionary economic geography". In: *Journal of Economic Geography* 6.3, pp. 273–302. DOI: 10.1093/jeg/lbi022.

Boschma, R. and K. Frenken (2010). "The Spatial Evolution of Innovation Networks: A Proximity Perspective". In: *The Handbook of Evolutionary Economic Geography*. Ed. by R. Boschma and R. Martin. Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing.

Boschma, R. and K. Frenken (2011). “Technological relatedness and regional branching”. In: *Dynamic Geographies of Knowledge Creation, Diffusion and Innovation*. Ed. by H. Bathelt, M. P. Feldman, and D. F. Kogler. New York: Routledge, pp. 64–81.

Boschma, R. and C. Gianelle (2014). “Regional branching and smart specialisation policy”. In: OCLC: 1044411439. Luxembourg: Publications Office.

Boschma, R., A. Minondo, and M. Navarro (2013). “The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness”. In: *Economic Geography* 89.1, pp. 29–51. DOI: 10.1111/j.1944-8287.2012.01170.x.

Boschma, R. and R. Wenting (2007). “The spatial evolution of the British automobile industry: Does location matter?” In: *Industrial and Corporate Change* 16.2, pp. 213–238. DOI: 10.1093/icc/dtm004.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). “Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak”. In: *Journal of the American Statistical Association* 90.430, pp. 443–450. DOI: 10.1080/01621459.1995.10476536.

Boytchev, H. (2019). “British chemist battles xenophobia in Germany”. In: *Nature* 567.7749, S47–S47. DOI: 10.1038/d41586-019-00915-2.

Breschi, S. and F. Lissoni (2009). “Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows”. In: *Journal of Economic Geography* 9.4, pp. 439–468. DOI: 10.1093/jeg/lbp008.

Breschi, S. and L. Cusmano (2004). “Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU Framework Programmes”. In: *International Journal of Technology Management* 27.8, p. 747. DOI: 10.1504/IJTM.2004.004992.

Breschi, S., F. Lissoni, and F. Malerba (2003). “Knowledge-relatedness in firm technological diversification”. In: *Research Policy* 32.1, pp. 69–87. DOI: 10.1016/S0048-7333(02)00004-5.

Breschi, S. and F. Malerba (1997). “Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamic, and Spatial Boundaries”. In: *Systems of Innovation*. London, New York: Routledge, pp. 130–156.

Brin, S. and L. Page (1998). “The anatomy of a large-scale hypertextual Web search engine”. In: *Computer Networks and ISDN Systems* 30.1, pp. 107–117. DOI: 10.1016/S0169-7552(98)00110-X.

Brockmann, D. and D. Helbing (2013). “The Hidden Geometry of Complex, Network-Driven Contagion Phenomena”. In: *Science* 342.6164, pp. 1337–1342. DOI: 10.1126/science.1245200.

Broekel, T. (2012). “Collaboration Intensity and Regional Innovation Efficiency in Germany—A Conditional Efficiency Approach”. In: *Industry & Innovation* 19.2, pp. 155–179. DOI: 10.1080/13662716.2012.650884.

Broekel, T. (2015). “Do Cooperative Research and Development (R&D) Subsidies Stimulate Regional Innovation Efficiency? Evidence from Germany”. In: *Regional Studies* 49.7, pp. 1087–1110. DOI: 10.1080/00343404.2013.812781.

Broekel, T. (2019). “Using structural diversity to measure the complexity of technologies”. In: *PLOS ONE* 14.5. Ed. by H. Youn, e0216856. DOI: 10.1371/journal.pone.0216856.

Broekel, T. and M. Brachert (2015). "The structure and evolution of inter-sectoral technological complementarity in R&D in Germany from 1990 to 2011". In: *Journal of Evolutionary Economics* 25.4, pp. 755–785. DOI: 10.1007/s00191-015-0415-7.

Broekel, T., M. Brachert, M. Duschl, and T. Brenner (2017). "Joint R&D Subsidies, Related Variety, and Regional Innovation". In: *International Regional Science Review* 40.3, pp. 297–326. DOI: 10.1177/0160017615589007.

Broekel, T., T. Brenner, and M. Buerger (2015a). "An Investigation of the Relation between Cooperation Intensity and the Innovative Success of German Regions". In: *Spatial Economic Analysis* 10.1, pp. 52–78. DOI: 10.1080/17421772.2014.992359.

Broekel, T., D. Fornahl, and A. Morrison (2015b). "Another cluster premium: Innovation subsidies and R&D collaboration networks". In: *Research Policy* 44.8, pp. 1431–1444. DOI: 10.1016/j.respol.2015.05.002.

Broekel, T. and H. Graf (2012). "Public research intensity and the structure of German R&D networks: a comparison of 10 technologies". In: *Economics of Innovation and New Technology* 21.4, pp. 345–372. DOI: 10.1080/10438599.2011.582704.

Broekel, T. and W. Mueller (2018). "Critical links in knowledge networks – What about proximities and gatekeeper organisations?" In: *Industry and Innovation* 25.10, pp. 919–939. DOI: 10.1080/13662716.2017.1343130.

Buesa, M., J. Heijs, M. Martínez Pellitero, and T. Baumert (2006). "Regional systems of innovation and the knowledge production function: the Spanish case". In: *Technovation* 26.4, pp. 463–472. DOI: 10.1016/j.technovation.2004.11.007.

Busom, I. (2000). "An Empirical Evaluation of The Effects of R&D Subsidies". In: *Economics of Innovation and New Technology* 9.2, pp. 111–148. DOI: 10.1080/10438590000000006.

Cairncross, F. (1997). *The death of distance: how the communications revolution will change our lives*. Boston, Mass: Harvard Business School Press. 303 pp.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). "Robust Inference With Multiway Clustering". In: *Journal of Business & Economic Statistics* 29.2, pp. 238–249. DOI: 10.1198/jbes.2010.07136.

Cantner, U. and S. Kösters (2012). "Picking the winner? Empirical evidence on the targeting of R&D subsidies to start-ups". In: *Small Business Economics* 39.4, pp. 921–936. DOI: 10.1007/s11187-011-9340-9.

Carlino, G. A., S. Chatterjee, and R. M. Hunt (2007). "Urban density and the rate of invention". In: *Journal of Urban Economics* 61.3, pp. 389–419. DOI: 10.1016/j.jue.2006.08.003.

Cassiman, B. and R. Veugelers (2002). "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium". In: *The American Economic Review* 92.4, pp. 1169–1184.

Castaldi, C., K. Frenken, and B. Los (2015). "Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting". In: *Regional Studies* 49.5, pp. 767–781. DOI: 10.1080/00343404.2014.940305.

Castaldi, C. and B. Los (2017). "Geographical patterns in US inventive activity 1977–1998: The "regional inversion" was underestimated". In: *Research Policy* 46.7, pp. 1187–1197. DOI: 10.1016/j.respol.2017.04.005.

Castells, M. (1996). *The rise of the network society*. Information age. Oxford ; Malden, Mass: Blackwell Publishers.

Cervellati, M. and U. Sunde (2005). "Human Capital Formation, Life Expectancy, and the Process of Development". In: *American Economic Review* 95.5, pp. 1653–1672. DOI: 10.1257/000282805775014380.

Chandy, R. K. and G. J. Tellis (1998). "Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize". In: *Journal of Marketing Research* 35.4, pp. 474–487. DOI: 10.1177/002224379803500406.

Chandy, R. K. and G. J. Tellis (2000). "The Incumbent's Curse? Incumbency, Size, and Radical Product Innovation". In: *Journal of Marketing* 64.3, pp. 1–17. DOI: 10.1509/jmkg.64.3.1.18033.

Chinitz, B. (1961). "Contrasts in Agglomeration: New York and Pittsburgh". In: *The American Economic Review* 51.2, pp. 279–289.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. OCLC: 439706924. Darmstadt: Wissenschaftliche Buchgesellschaft. 331 pp.

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, Massachusetts, USA: Harvard Business School Press.

Christopherson, S., H. Garretsen, and R. Martin (2008). "The world is not flat: putting globalization in its place". In: *Cambridge Journal of Regions, Economy and Society* 1.3, pp. 343–349. DOI: 10.1093/cjres/rsn023.

Co, C. (2002). "Evolution of the Geography of Innovation: Evidence from Patent Data". In: *Growth and Change* 33.4, pp. 393–423. DOI: 10.1111/1468-2257.00204.

Coenen, L., J. Moodysson, and H. Martin (2015). "Path Renewal in Old Industrial Regions: Possibilities and Limitations for Regional Innovation Policy". In: *Regional Studies* 49.5, pp. 850–865. DOI: 10.1080/00343404.2014.979321.

Cohen, W. M. and D. A. Levinthal (1990). "Absorptive Capacity: A New Perspective on Learning and Innovation". In: *Administrative Science Quarterly* 35.1, p. 128. DOI: 10.2307/2393553.

Cohen, W., R. Nelson, and J. Walsh (2000). *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)*. w7552. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w7552.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). "Spatial wage disparities: Sorting matters!" In: *Journal of Urban Economics* 63.2, pp. 723–742. DOI: 10.1016/j.jue.2007.04.004.

Cooke, P. (1998). "Introduction: Origins of the Concept". In: *Regional Innovation Systems - The Role of Governances in a Globalized World*. Ed. by H.-J. Braczyk, P. Cooke, and M. Heidenreich. London: UCL Press, pp. 2–25.

Cortinovis, N., J. Xiao, R. Boschma, and F. G. van Oort (2017). "Quality of government and social capital as drivers of regional diversification in Europe". In: *Journal of Economic Geography* 17.6, pp. 1179–1208. DOI: 10.1093/jeg/lbx001.

Cowan, R., P. A. David, and D. Foray (2000). "The explicit economics of knowledge codification and tacitness". In: *Industrial and Corporate Change* 9.2, pp. 211–253. DOI: 10.1093/icc/9.2.211.

Cowan, R. and D. Foray (1997). "The Economics of Codification and the Diffusion of Knowledge". In: *Industrial and Corporate Change* 6.3, pp. 595–622. DOI: 10.1093/icc/6.3.595.

Crescenzi, R. (2005). "Innovation and Regional Growth in the Enlarged Europe: The Role of Local Innovative Capabilities, Peripherality, and Education". In: *Growth and Change* 36.4, pp. 471–507. DOI: 10.1111/j.1468-2257.2005.00291.x.

Czarnitzki, D., B. Ebersberger, and A. Fier (2007). "The relationship between R&D collaboration, subsidies and R&D performance: Empirical evidence from Finland and Germany". In: *Journal of Applied Econometrics* 22.7, pp. 1347–1366. DOI: 10.1002/jae.992.

Czarnitzki, D. and K. Hussinger (2004). *The Link Between R&D Subsidies, R&D Spending and Technological Performance*. 04-56. Centre for European Economic Research, pp. 1–32.

Czarnitzki, D. and K. Hussinger (2018). "Input and output additionality of R&D subsidies". In: *Applied Economics* 50.12, pp. 1324–1341. DOI: 10.1080/00036846.2017.1361010.

Czarnitzki, D. and C. Lopes-Bento (2013). "Value for money? New microeconometric evidence on public R&D grants in Flanders". In: *Research Policy* 42.1, pp. 76–89. DOI: 10.1016/j.respol.2012.04.008.

Dahlin, K. B. and D. M. Behrens (2005). "When is an invention really radical?" In: *Research Policy* 34.5, pp. 717–737. DOI: 10.1016/j.respol.2005.03.009.

David, P. A., B. H. Hall, and A. A. Toole (2000). "Is public R&D a complement or substitute for private R&D? A review of the econometric evidence". In: *Research Policy* 29.4, pp. 497–529. DOI: 10.1016/S0048-7333(99)00087-6.

Defazio, D., A. Lockett, and M. Wright (2009). "Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program". In: *Research Policy* 38.2, pp. 293–305. DOI: 10.1016/j.respol.2008.11.008.

Dohse, D. (2000). "Technology policy and the regions — the case of the BioRegio contest". In: *Research Policy* 29.9, pp. 1111–1133. DOI: 10.1016/S0048-7333(99)00077-3.

Dosi, G. (1982). "Technological paradigms and technological trajectories". In: *Research Policy* 11.3, pp. 147–162. DOI: 10.1016/0048-7333(82)90016-6.

Dosi, G. (1988). "Sources, Procedures, and Microeconornic Effects of Innovation". In: *Journal of Economic Literature* 26.3, pp. 1120–1171.

Drucker, P. F. (1985). "The discipline of innovation". In: *Harvard Business Review* 63.3, pp. 67–72.

Dumais, G., G. Ellison, and E. L. Glaeser (2002). "Geographic Concentration as a Dynamic Process". In: *Review of Economics and Statistics* 84.2, pp. 193–204. DOI: 10.1162/003465302317411479.

Dunar, A. J. and S. P. Waring (1999). *Power to explore: a history of Marshall Space Flight Center 1960-1990*. OCLC: 1046550321. Washington, DC: National Aeronautics and Space Administration, NASA History Office.

Ebersberger, B. and O. Lehtoranta (2008). *Effects of Public R&D Funding*. 100. VTT Technical Research Centre of Finland, pp. 1–27.

Ebert, T., F. M. Götz, M. Obschonka, L. Zmigrod, and P. J. Rentfrow (2019). "Regional variation in courage and entrepreneurship: The contrasting role of courage for the emergence and survival of start-ups in the United States". In: *Journal of Personality*. DOI: 10.1111/jopy.12454.

Ejermo, O. (2009). “Regional Innovation Measured by Patent Data—Does Quality Matter?: Research Paper”. In: *Industry & Innovation* 16.2, pp. 141–165. DOI: 10.1080/13662710902764246.

Ellison, G. and E. L. Glaeser (1999). “The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?” In: *American Economic Review* 89.2, pp. 311–316. DOI: 10.1257/aer.89.2.311.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010). “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns”. In: *American Economic Review* 100.3, pp. 1195–1213. DOI: 10.1257/aer.100.3.1195.

Emmert-Streib, F. and M. Dehmer (2012). “Exploring Statistical and Population Aspects of Network Complexity”. In: *PLoS ONE* 7.5. Ed. by A. J. Cannon, e34523. DOI: 10.1371/journal.pone.0034523.

Engelsman, E. and A. van Raan (1994). “A patent-based cartography of technology”. In: *Research Policy* 23.1, pp. 1–26. DOI: 10.1016/0048-7333(94)90024-8.

Essletzbichler, J. (2015). “Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas”. In: *Regional Studies* 49.5, pp. 752–766. DOI: 10.1080/00343404.2013.806793.

Eurostat (2016). *High-tech industry and knowledge-intensive services (htec)*.

Fagerberg, J., B. Verspagen, and M. Caniëls (1997). “Technology, Growth and Unemployment across European Regions”. In: *Regional Studies* 31.5, pp. 457–466. DOI: 10.1080/00343409750132252.

Feldman, M. P. (1994). *The Geography of Innovation*. Red. by C. Antonelli and B. Carlsson. Vol. 2. Economics of Science, Technology and Innovation. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-017-3333-5.

Feldman, M. P. and D. B. Audretsch (1999). “Innovation in cities: Science-based diversity, specialization and localized competition”. In: *European Economic Review* 43.2, pp. 409–429. DOI: 10.1016/S0014-2921(98)00047-6.

Feldman, M. P. and R. Florida (1994). “The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States”. In: *Annals of the Association of American Geographers* 84.2, pp. 210–229. DOI: 10.1111/j.1467-8306.1994.tb01735.x.

Feldman, M. P. and D. F. Kogler (2010). “Stylized Facts in the Geography of Innovation”. In: *Handbook of the Economics of Innovation*. Vol. 1. Elsevier, pp. 381–410. DOI: 10.1016/S0169-7218(10)01008-7.

Ferrarini, B. and P. Scaramozzino (2016). “Production complexity, adaptability and economic growth”. In: *Structural Change and Economic Dynamics* 37, pp. 52–61. DOI: 10.1016/j.strueco.2015.12.001.

Fier, A., B. Aschhoff, and H. Löhlein (2006). “Behavioural additionality of public R&D funding in Germany”. In: *OECD Government R&D Funding and Company Behaviour, Measuring Behavioural Additionality*, pp. 127–149.

Fink, T. M. A., M. Reeves, R. Palma, and R. S. Farr (2017). “Serendipity and strategy in rapid innovation”. In: *Nature Communications* 8.1. DOI: 10.1038/s41467-017-02042-w.

Fitjar, R. D. and A. Rodríguez-Pose (2011). “When Local Interaction Does Not Suffice: Sources of Firm Innovation in Urban Norway”. In: *Environment and Planning A: Economy and Space* 43.6, pp. 1248–1267. DOI: 10.1068/a43516.

Fleming, L., H. Greene, G. Li, M. Marx, and D. Yao (2019). “Government-funded research increasingly fuels innovation”. In: *Science* 364.6446, pp. 1139–1141. DOI: 10.1126/science.aaw2373.

Fleming, L. (2001). “Recombinant Uncertainty in Technological Search”. In: *Management Science* 47.1, pp. 117–132. DOI: 10.1287/mnsc.47.1.117.10671.

Fleming, L. and O. Sorenson (2001). “Technology as a complex adaptive system: evidence from patent data”. In: *Research Policy* 30.7, pp. 1019–1039. DOI: 10.1016/S0048-7333(00)00135-9.

Florida, R., C. Mellander, and K. Stolarick (2008a). “Inside the black box of regional development–human capital, the creative class and tolerance”. In: *Journal of Economic Geography* 8.5, pp. 615–649. DOI: 10.1093/jeg/lbn023.

Florida, R. (1995). “Toward the learning region”. In: *Futures* 27.5, pp. 527–536. DOI: 10.1016/0016-3287(95)00021-N.

Florida, R. (2002). *The rise of the creative class: and how it’s transforming work, leisure, community and everyday life*. OCLC: ocm54065465. New York, NY: Basic Books.

Florida, R. (2003). “Cities and the Creative Class”. In: *City and Community* 2.1, pp. 3–19. DOI: 10.1111/1540-6040.00034.

Florida, R., T. Gulden, and C. Mellander (2008b). “The rise of the mega-region”. In: *Cambridge Journal of Regions, Economy and Society* 1.3, pp. 459–476. DOI: 10.1093/cjres/rsn018.

Foray, D., P. A. David, and B. H. Hall (2011). *Smart specialization. From academic idea to political instrument , the surprising career of a concept and the difficulties involved in its implementation*. 001. Management of Technology and Entrepreneurship Institute, pp. 1–16.

Fornahl, D., T. Broekel, and R. Boschma (2011). “What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location: What drives patent performance of German biotech firms?” In: *Papers in Regional Science* 90.2, pp. 395–418. DOI: 10.1111/j.1435-5957.2011.00361.x.

Frenken, K., F. Van Oort, and T. Verburg (2007). “Related Variety, Unrelated Variety and Regional Economic Growth”. In: *Regional Studies* 41.5, pp. 685–697. DOI: 10.1080/00343400601120296.

Friedman, T. L. (2005). *The world is flat: a brief history of the twenty-first century*. OCLC: ocm64098349. New York: Farrar, Straus and Giroux.

Fritsch, M. (2002). “Measuring the Quality of Regional Innovation Systems: A Knowledge Production Function Approach”. In: *International Regional Science Review* 25.1, pp. 86–101. DOI: 10.1177/016001702762039394.

Fukuyama, F. (1995). *Trust: the social virtues and the creation of prosperity*. New York: Free Press. 457 pp.

Garfield, E. (1970). “Citation Indexing for Studying Science”. In: *Nature* 227.5259, pp. 669–671. DOI: 10.1038/227669a0.

Garretsen, H., J. I. Stoker, D. Soudis, R. L. Martin, and P. J. Rentfrow (2018a). “Brexit and the relevance of regional personality traits: more psychological Openness could have swung the regional vote”. In: *Cambridge Journal of Regions, Economy and Society* 11.1, pp. 165–175. DOI: 10.1093/cjres/rsx031.

Garretsen, H., J. I. Stoker, D. Soudis, R. Martin, and J. Rentfrow (2018b). "The relevance of personality traits for urban economic growth: making space for psychological factors". In: *Journal of Economic Geography*. DOI: 10.1093/jeg/lby025.

George, J. M. and J. Zhou (2001). "When openness to experience and conscientiousness are related to creative behavior: An interactional approach." In: *Journal of Applied Psychology* 86.3, pp. 513–524. DOI: 10.1037/0021-9010.86.3.513.

Gertler, M. S. (2003). "Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there)". In: *Journal of Economic Geography* 3.1, pp. 75–99. DOI: 10.1093/jeg/3.1.75.

Gisler, M. and D. Sornette (2009). "Exuberant Innovations: The Apollo Program". In: *Society* 46.1, pp. 55–68. DOI: 10.1007/s12115-008-9163-8.

Glaeser, E. L. (2011). *Triumph of the city: how our greatest invention makes us richer, smarter, greener, healthier, and happier*. OCLC: ocn650211168. New York: Penguin Press. 338 pp.

Glaeser, E. L., H. D. Kallal, and J. A. Scheinkman (1992). "Growth in Cities". In: *Journal of Political Economy* 100.6, pp. 1126–1152.

Glaeser, E. L. and D. C. Maré (2001). "Cities and Skills". In: *Journal of Labor Economics* 19.2, pp. 316–342. DOI: 10.1086/319563.

Godfray, H. C. J. et al. (2010). "Food Security: The Challenge of Feeding 9 Billion People". In: *Science* 327.5967, pp. 812–818. DOI: 10.1126/science.1185383.

Gosling, S. D., S. Vazire, S. Srivastava, and O. P. John (2004). "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires." In: *American Psychologist* 59.2, pp. 93–104. DOI: 10.1037/0003-066X.59.2.93.

Götz, F. M., T. Ebert, and P. J. Rentfrow (2018). "Regional Cultures and the Psychological Geography of Switzerland: Person–Environment–Fit in Personality Predicts Subjective Wellbeing". In: *Frontiers in Psychology* 9. DOI: 10.3389/fpsyg.2018.00517.

Grabher, G., ed. (1993). *The Embedded firm: on the socioeconomics of industrial networks*. London ; New York: Routledge. 306 pp.

Granovetter, M. (1985). "Economic Action and Social Structure: The Problem of Embeddedness". In: *American Journal of Sociology* 91.3, pp. 481–510. DOI: 10.1086/228311.

Griliches, Z. (1990). *Patent Statistics as Economic Indicators: A Survey*. 3301. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w3301.

Griliches, Z. (1994). "R&D and Productivity: The Econometric Evidence". In: *The American Economic Review* 84.1, pp. 1–23.

Grossman, G. M. and E. Helpman (1991). "Quality Ladders in the Theory of Growth". In: *The Review of Economic Studies* 58.1, p. 43. DOI: 10.2307/2298044.

Gurtman, M. B. (1995). "Personality Structure and Interpersonal Problems: A Theoretically-Guided Tem Analysis of the Inventory of Interpersonal Problems". In: *Assessment* 2.4, pp. 343–361. DOI: 10.1177/1073191195002004005.

Hagedoorn, J. (1993). "Understanding the rationale of strategic technology partnering: Nterorganizational modes of cooperation and sectoral differences". In: *Strategic Management Journal* 14.5, pp. 371–385. DOI: 10.1002/smj.4250140505.

Hagedoorn, J. (2002). “Inter-firm R&D partnerships: an overview of major trends and patterns since 1960”. In: *Research Policy* 31.4, pp. 477–492. DOI: 10.1016/S0048-7333(01)00120-2.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools”. In: *NBER Working Paper Series* 8498.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2005). “Market Value and Patent Citations”. In: *Rand Journal of Economics* 36.1, pp. 16–38.

Hall, P. and A. R. Markusen, eds. (1985). *Silicon landscapes.* Boston: Allen and Unwin. 160 pp.

Hargadon, A. (2003). *How breakthroughs happen: the surprising truth about how companies innovate.* Boston, Mass: Harvard Business School Press. 254 pp.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). “Citation Frequency and the Value of Patented Inventions”. In: *Review of Economics and Statistics* 81.3, pp. 511–515. DOI: 10.1162/003465399558265.

Harrington, J. R. and M. J. Gelfand (2014). “Tightness-looseness across the 50 united states”. In: *Proceedings of the National Academy of Sciences* 111.22, pp. 7990–7995. DOI: 10.1073/pnas.1317937111.

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jiminez, A. Simoes, and M. A. Yildirim (2013). *The Atlas of economic complexity: mapping paths to prosperity.* OCLC: 961922687. Cambridge: Center for International Development, Harvard University.

Hausmann, R. and D. Rodrik (2003). “Economic development as self-discovery”. In: *Journal of Development Economics* 72.2, pp. 603–633. DOI: 10.1016/S0304-3878(03)00124-X.

Henderson, J. V. (1974). “The Sizes and Types of Cities”. In: *The American Economic Review* 64.4, pp. 640–656.

Henderson, J. V., Z. Shalizi, and A. J. Venables (2001). “Geography and development”. In: *Journal of Economic Geography* 1.1, pp. 81–105. DOI: 10.1093/jeg/1.1.81.

Henrich, J. (2004). “Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses - The Tasmanian Case”. In: *American Antiquity* 69.2, pp. 197–214. DOI: 10.2307/4128416.

Hidalgo, C. A. and R. Hausmann (2009). “The building blocks of economic complexity”. In: *Proceedings of the National Academy of Sciences* 106.26, pp. 10570–10575. DOI: 10.1073/pnas.0900943106.

Hidalgo, C. A., B. Klinger, A.-L. Barabasi, and R. Hausmann (2007). “The Product Space Conditions the Development of Nations”. In: *Science* 317.5837, pp. 482–487. DOI: 10.1126/science.1144581.

Hidalgo, C. A. et al. (2018). “The Principle of Relatedness”. In: *Unifying Themes in Complex Systems IX.* Ed. by A. J. Morales, C. Gershenson, D. Braha, A. A. Minai, and Y. Bar-Yam. Cham: Springer International Publishing, pp. 451–457. DOI: 10.1007/978-3-319-96661-8_46.

Hippel, E. von (1994). ““Sticky Information” and the Locus of Problem Solving: Implications for Innovation”. In: *Management Science* 40.4, pp. 429–439. DOI: 10.1287/mnsc.40.4.429.

Hofstede, G. (1980). *Culture’s Consequences: Internal Differences in Work Related Values.* Beverly Hills California: Sage.

Hofstede, G. and R. R. McCrae (2004). “Personality and Culture Revisited: Linking Traits and Dimensions of Culture”. In: *Cross-Cultural Research* 38.1, pp. 52–88. DOI: 10.1177/1069397103259443.

Huggins, R. and P. Thompson (2017). “The behavioural foundations of urban and regional development: culture, psychology and agency”. In: *Journal of Economic Geography*. DOI: 10.1093/jeg/lbx040.

Imbs, J. and R. Wacziarg (2003). “Stages of Diversification”. In: *American Economic Review* 93.1, pp. 63–86. DOI: 10.1257/000282803321455160.

Jacobs, J. (1969). *The Economy of Cities*. Vintage books 584. OCLC: 222334636. New York: Vintage Books. 268 pp.

Jacobsson, S. and V. Lauber (2006). “The politics and policy of energy system transformation—explaining the German diffusion of renewable energy technology”. In: *Energy Policy* 34.3, pp. 256–276. DOI: 10.1016/j.enpol.2004.08.029.

Jaffe, A. B. (1989). “Real Effects of Academic Research”. In: *The American Economic Review* 79.5, pp. 957–970.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations”. In: *The Quarterly Journal of Economics* 108.3, pp. 577–598. DOI: 10.2307/2118401.

Jang, K. L., R. R. McCrae, A. Angleitner, R. Riemann, and W. J. Livesley (1998). “Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality.” In: *Journal of Personality and Social Psychology* 74.6, pp. 1556–1565. DOI: 10.1037/0022-3514.74.6.1556.

John, O. P. and S. Srivastava (1999). “The Big Five Trait taxonomy: History, measurement, and theoretical perspectives”. In: *Handbook of personality: Theory and research*. Ed. by A. Pervin and O. P. John. New York, NY: Guilford Press, pp. 102–138.

Johnson, B. (2015). “The Great Horse Manure Crisis of 1894”. In: *Historic UK*.

Jokela, M., W. Bleidorn, M. E. Lamb, S. D. Gosling, and P. J. Rentfrow (2015). “Geographically varying associations between personality and life satisfaction in the London metropolitan area”. In: *Proceedings of the National Academy of Sciences* 112.3, pp. 725–730. DOI: 10.1073/pnas.1415800112.

Jones, B. F., S. Wuchty, and B. Uzzi (2008). “Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science”. In: *Science* 322.5905, pp. 1259–1262. DOI: 10.1126/science.1158357.

Jones, C. I. (1995). “R&D-Based Models of Economic Growth”. In: *Journal of Political Economy* 103.4, pp. 759–784. DOI: 10.1086/262002.

Jost, L. (2006). “Entropy and Diversity”. In: *Oikos* 113.2, pp. 363–375. DOI: 10.1111/j.2006.0030-1299.14714.x.

Jovanovic, B. and Y. Nyarko (1995). “A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves”. In: *Brookings Papers on Economic Activity. Microeconomics* 1995, pp. 247–305. DOI: 10.2307/2534775.

Kaplan, S. and K. Vakili (2015). “The double-edged sword of recombination in breakthrough innovation”. In: *Strategic Management Journal* 36.10, pp. 1435–1457. DOI: 10.1002/smj.2294.

Kauffman, S. A. (1993). *The Origins of Order - Self-Organization and Selection in Evolution*. New York: Oxford University Press.

Kim, D., D. B. Cerigo, H. Jeong, and H. Youn (2016). "Technological novelty profile and invention's future impact". In: *EPJ Data Science* 5.1. DOI: 10.1140/epjds/s13688-016-0069-1.

King, L. A., L. M. Walker, and S. J. Broyles (1996). "Creativity and the Five-Factor Model". In: *Journal of Research in Personality* 30.2, pp. 189–203. DOI: 10.1006/jrpe.1996.0013.

Klement, B. and S. Strambach (2019). "Innovation in Creative Industries: Does (Related) Variety Matter for the Creativity of Urban Music Scenes?" In: *Economic Geography*, pp. 1–33. DOI: 10.1080/00130095.2018.1549944.

Knudsen, B., R. Florida, K. Stolarick, and G. Gates (2008). "Density and Creativity in U.S. Regions". In: *Annals of the Association of American Geographers* 98.2, pp. 461–478. DOI: 10.1080/00045600701851150.

Kogler, D. F., D. L. Rigby, and I. Tucker (2013). "Mapping Knowledge Space and Technological Relatedness in US Cities". In: *European Planning Studies* 21.9, pp. 1374–1391. DOI: 10.1080/09654313.2012.755832.

Kogut, B. and U. Zander (1992). "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology". In: *Organization Science* 3.3, pp. 383–397.

Kosfeld, R. and A. Werner (2012). "Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011". In: *Raumforschung und Raumordnung* 70.1, pp. 49–64. DOI: 10.1007/s13147-011-0137-8.

Koski, H. and M. Pajarinen (2015). "Subsidies, the Shadow of Death and Labor Productivity". In: *Journal of Industry, Competition and Trade* 15.2, pp. 189–204. DOI: 10.1007/s10842-014-0177-1.

Kremer, M. (1993). "Population Growth and Technological Change: One Million B.C. to 1990". In: *The Quarterly Journal of Economics* 108.3, pp. 681–716. DOI: 10.2307/2118405.

Kuznets, S. (1960). "Population Change and Aggregate Output". In: *Demographic and Economic Change in Developed Countries*. Princeton, NJ: Columbia University Press, pp. 324–351.

Kuznets, S. (1962). "Inventive Activity: Problems of Definition and Measurement". In: *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Ed. by National Bureau of Economic Research. Princeton University Press, pp. 19–52.

Lamarque, J.-F. et al. (2010). "Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application". In: *Atmospheric Chemistry and Physics* 10.15, pp. 7017–7039. DOI: 10.5194/acp-10-7017-2010.

Lanjouw, J. O. and M. Schankerman (2004). "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators*". In: *The Economic Journal* 114.495, pp. 441–465. DOI: 10.1111/j.1468-0297.2004.00216.x.

Latané, B. (1981). "The Psychology of Social Impact." In: *American Psychologist* 36.4, pp. 343–356. DOI: 10.1037/0003-066X.36.4.343.

Lee, N. (2017). "Psychology and the Geography of Innovation". In: *Economic Geography* 93.2, pp. 106–130. DOI: 10.1080/00130095.2016.1249845.

Lee, N., P. Sissons, and K. Jones (2016). "The Geography of Wage Inequality in British Cities". In: *Regional Studies* 50.10, pp. 1714–1727. DOI: 10.1080/00343404.2015.1053859.

Lee, S. Y., R. Florida, and Z. Acs (2004). “Creativity and Entrepreneurship: A Regional Analysis of New Firm Formation”. In: *Regional Studies* 38.8, pp. 879–891. DOI: 10.1080/0034340042000280910.

Lee, S. Y., R. Florida, and G. Gates (2010). “Innovation, Human Capital, and Creativity”. In: *International Review of Public Administration* 14.3, pp. 13–24. DOI: 10.1080/12294659.2010.10805158.

Leonard-Barton, D. (1992). “Core capabilities and core rigidities: A paradox in managing new product development”. In: *Strategic Management Journal* 13 (S1), pp. 111–125. DOI: 10.1002/smj.4250131009.

Li, G.-C., R. Lai, A. D’Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014). “Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010)”. In: *Research Policy* 43.6, pp. 941–955. DOI: 10.1016/j.respol.2014.01.012.

Liben-Nowell, D., J. Novak, R. Kumar, P. Raghavan, and A. Tomkins (2005). “Geographic routing in social networks”. In: *Proceedings of the National Academy of Sciences* 102.33, pp. 11623–11628. DOI: 10.1073/pnas.0503018102.

London, J. (1904). *The Sea Wolf*. Macmillan.

Louf, R. and M. Barthelemy (2014). “Scaling: Lost in the Smog”. In: *Environment and Planning B: Planning and Design* 41.5, pp. 767–769. DOI: 10.1068/b4105c.

Lucas, R. E. (1988). “On the mechanics of economic development”. In: *Journal of Monetary Economics* 22.1, pp. 3–42. DOI: 10.1016/0304-3932(88)90168-7.

Maggioni, M. A., T. E. Uberti, and M. Nosvelli (2014). “Does intentional mean hierarchical? Knowledge flows and innovative performance of European regions”. In: *The Annals of Regional Science* 53.2, pp. 453–485. DOI: 10.1007/s00168-014-0618-0.

Malecki, E. J. (2010). “Everywhere? The Geography of Knowledge”. In: *Journal of Regional Science* 50.1, pp. 493–513. DOI: 10.1111/j.1467-9787.2009.00640.x.

Malmberg, A. and D. Power (2005). “(How) Do (Firms in) Clusters Create Knowledge?” In: *Industry & Innovation* 12.4, pp. 409–431. DOI: 10.1080/13662710500381583.

Malmberg, A., O. Solvell, and I. Zander (1996). “Spatial Clustering, Local Accumulation of Knowledge and Firm Competitiveness”. In: *Geografiska Annaler. Series B, Human Geography* 78.2, p. 85. DOI: 10.2307/490807.

Manley, D. (2014). “Scale, Aggregation, and the Modifiable Areal Unit Problem”. In: *Handbook of Regional Science*. Ed. by M. M. Fischer and P. Nijkamp. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1157–1171. DOI: 10.1007/978-3-642-23430-9_69.

March, J. G. (1991). “Exploration and Exploitation in Organizational Learning”. In: *Organization Science* 2.1, pp. 71–87.

Markusen, A. (1996). “Sticky Places in Slippery Space: A Typology of Industrial Districts”. In: *Economic Geography* 72.3, p. 293. DOI: 10.2307/144402.

Marquis, C. and A. Tilcsik (2013). “Imprinting: Toward a Multilevel Theory”. In: *The Academy of Management Annals* 7.1, pp. 195–245. DOI: 10.1080/19416520.2013.766076.

Marshall, A. (1890). *Principles of economics*. OCLC: ocm36200952. London: McMillan.

Martin, R. (2012). “Regional economic resilience, hysteresis and recessionary shocks”. In: *Journal of Economic Geography* 12.1, pp. 1–32. DOI: 10.1093/jeg/lbr019.

Martin, R. and P. Sunley (2006). “Path dependence and regional economic evolution”. In: *Journal of Economic Geography* 6.4, pp. 395–437. DOI: 10.1093/jeg/lbl012.

Maskell, P. and A. Malmberg (1999). “Localised learning and industrial competitiveness”. In: *Cambridge Journal of Economics* 23.2, pp. 167–185. DOI: 10.1093/cje/23.2.167.

Mazzucato, M. (2014). *The entrepreneurial state: debunking public vs. private sector myths*. Rev. ed. OCLC: 889985243. London: Anthem Press. 237 pp.

McCann, P. and R. Ortega-Argiles (2013). “Modern regional innovation policy”. In: *Cambridge Journal of Regions, Economy and Society* 6.2, pp. 187–216. DOI: 10.1093/cjres/rst007.

McCann, S. J. H. (2011). “Emotional Health and the Big Five Personality Factors at the American State Level”. In: *Journal of Happiness Studies* 12.4, pp. 547–560. DOI: 10.1007/s10902-010-9215-9.

McCrae, R. R. (1987). “Creativity, divergent thinking, and openness to experience.” In: *Journal of Personality and Social Psychology* 52.6, pp. 1258–1265. DOI: 10.1037/0022-3514.52.6.1258.

McCrae, R. R. (1996). “Social consequences of experiential openness.” In: *Psychological Bulletin* 120.3, pp. 323–337. DOI: 10.1037/0033-2909.120.3.323.

McCrae, R. R. (2001). “Trait Psychology and Culture: Exploring Intercultural Comparisons”. In: *Journal of Personality* 69.6, pp. 819–846. DOI: 10.1111/1467-6494.696166.

McEvily, S. K. and B. Chakravarthy (2002). “The persistence of knowledge-based advantage: an empirical test for product performance and technological knowledge”. In: *Strategic Management Journal* 23.4, pp. 285–305. DOI: 10.1002/smj.223.

Mills, E. S. (1967). “An Aggregative Model of Resource Allocation in a Metropolitan Area”. In: *The American Economic Review* 57.2, pp. 197–210.

Mohler, S. R. (2004). “Human factors of powered flight: the Wright brothers’ contributions”. In: *Aviation, Space, and Environmental Medicine* 75.2, pp. 184–188.

Moreno, R., R. Paci, and S. Usai (2005). “Spatial Spillovers and Innovation Activity in European Regions”. In: *Environment and Planning A: Economy and Space* 37.10, pp. 1793–1812. DOI: 10.1068/a37341.

Mori, T., K. Nishikimi, and T. E. Smith (2008). “The Number-Average Size Rule: A New Empirical Relationship Between Industrial Location and City Size”. In: *Journal of Regional Science* 48.1, pp. 165–211. DOI: 10.1111/j.1467-9787.2008.00550.x.

Neffke, F., M. Henning, and R. Boschma (2011). “How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions”. In: *Economic Geography* 87.3, pp. 237–265. DOI: 10.1111/j.1944-8287.2011.01121.x.

Nelson, R. R. (1959). “The Simple Economics of Basic Scientific Research”. In: *Journal of Political Economy* 67.3, pp. 297–306. DOI: 10.1086/258177.

Nelson, R. R. and S. G. Winter (1982). *An evolutionary theory of economic change*. OCLC: 255191816. Cambridge, Mass.: The Belknap Press of Harvard Univ. Press. 437 pp.

New York Times (1919). “Alcock and Brown Fly Across Atlantic”. In: *The New York Times*.

Neyman, J. and E. L. Scott (1948). “Consistent Estimates Based on Partially Consistent Observations”. In: *Econometrica* 16.1, p. 1. DOI: 10.2307/1914288.

Nooteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing, and A. van den Oord (2007). “Optimal cognitive distance and absorptive capacity”. In: *Research Policy* 36.7, pp. 1016–1034. DOI: 10.1016/j.respol.2007.04.003.

O’Brien, R. (1992). *Global financial integration: the end of geography*. In collab. with R. I. of International Affairs. Chatham House papers. OCLC: 26315402. London: Pinter Publishers. 120 pp.

Obschonka, M. (2017). “The quest for the entrepreneurial culture: psychological Big Data in entrepreneurship research”. In: *Current Opinion in Behavioral Sciences* 18, pp. 69–74. DOI: 10.1016/j.cobeha.2017.07.014.

Obschonka, M. and D. B. Audretsch (2019). “Artificial intelligence and big data in entrepreneurship: a new era has begun”. In: *Small Business Economics*. DOI: 10.1007/s11187-019-00202-4.

Obschonka, M., M. Stuetzer, D. B. Audretsch, P. J. Rentfrow, J. Potter, and S. D. Gosling (2016). “Macropsychological Factors Predict Regional Economic Resilience During a Major Economic Crisis”. In: *Social Psychological and Personality Science* 7.2, pp. 95–104. DOI: 10.1177/1948550615608402.

Obschonka, M., M. Stuetzer, S. D. Gosling, P. J. Rentfrow, M. E. Lamb, J. Potter, and D. B. Audretsch (2015). “Entrepreneurial Regions: Do Macro-Psychological Cultural Characteristics of Regions Help Solve the “Knowledge Paradox” of Economics?” In: *PLOS ONE* 10.6. Ed. by L. J. Waldorp, e0129332. DOI: 10.1371/journal.pone.0129332.

Obschonka, M., M. Stuetzer, P. J. Rentfrow, N. Lee, J. Potter, and S. D. Gosling (2018). “Fear, Populism, and the Geopolitical Landscape: The “Sleeper Effect” of Neurotic Personality Traits on Regional Voting Behavior in the 2016 Brexit and Trump Elections”. In: *Social Psychological and Personality Science* 9.3, pp. 285–298. DOI: 10.1177/1948550618755874.

O’Connor, G. C. (1998). “Market Learning and Radical Innovation: A Cross Case Comparison of Eight Radical Innovation Projects”. In: *Journal of Product Innovation Management* 15.2, pp. 151–166. DOI: 10.1111/1540-5885.1520151.

O’hUallichain, B. (1999). “Patent Places: Size Matters”. In: *Journal of Regional Science* 39.4, pp. 613–636. DOI: 10.1111/0022-4146.00152.

O’hUallichain, B. and T. F. Leslie (2005). “Spatial Convergence and Spillovers in American Invention”. In: *Annals of the Association of American Geographers* 95.4, pp. 866–886. DOI: 10.1111/j.1467-8306.2005.00491.x.

Oort, F. van, S. de Geus, and T. Dogaru (2015). “Related Variety and Regional Economic Growth in a Cross-Section of European Urban Regions”. In: *European Planning Studies* 23.6, pp. 1110–1127. DOI: 10.1080/09654313.2014.905003.

Oyserman, D. (2017). “Culture Three Ways: Culture and Subcultures Within Countries”. In: *Annual Review of Psychology* 68.1, pp. 435–463. DOI: 10.1146/annurev-psych-122414-033617.

Parent, O. and J. P. LeSage (2012). “Determinants of Knowledge Production and Their Effects on Regional Economic Growth”. In: *Journal of Regional Science* 52.2, pp. 256–284. DOI: 10.1111/j.1467-9787.2011.00732.x.

Pater, R. and A. Lewandowska (2015). “Human capital and innovativeness of the European Union regions”. In: *Innovation: The European Journal of Social Science Research* 28.1, pp. 31–51. DOI: 10.1080/13511610.2014.962487.

Petralia, S., P.-A. Balland, and A. Morrison (2017). “Climbing the ladder of technological development”. In: *Research Policy* 46.5, pp. 956–969. DOI: 10.1016/j.respol.2017.03.012.

Petralia, S., P.-A. Balland, and D. L. Rigby (2016). “Unveiling the geography of historical patents in the United States from 1836 to 1975”. In: *Scientific Data* 3, p. 160074. DOI: 10.1038/sdata.2016.74.

Piergiovanni, R., M. A. Carree, and E. Santarelli (2012). “Creative industries, new business formation, and regional economic growth”. In: *Small Business Economics* 39.3, pp. 539–560. DOI: 10.1007/s11187-011-9329-4.

Pinheiro, F. L., A. Alshamsi, D. Hartmann, R. Boschma, and C. A. Hidalgo (2018). “Shooting High or Low: Do Countries Benefit from Entering Unrelated Activities?” In: *arXiv:1801.05352 [physics, q-fin]*.

Pintea, M. and P. Thompson (2007). “Technological complexity and economic growth”. In: *Review of Economic Dynamics* 10.2, pp. 276–293. DOI: 10.1016/j.red.2006.12.001.

Pirie, G. (2009). “Incidental tourism: British Imperial air travel in the 1930s”. In: *Journal of Tourism History* 1.1, pp. 49–66. DOI: 10.1080/17551820902742772.

Polanyi, M. (1966). *The Tacit Dimension.* New York: Anchor Day Books.

Porter, M. E. (2000). “Locations, clusters, and company strategy”. In: *The Oxford handbook of economic geography.* Ed. by G. L. Clark, M. P. Feldman, and M. S. Gertler. Oxford: Oxford University Press, pp. 253–274.

Powell, W. W., K. W. Koput, and L. Smith-Doerr (1996). “Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology”. In: *Administrative Science Quarterly* 41.1, p. 116. DOI: 10.2307/2393988.

Praag, C. M. van and P. H. Versloot (2008). “The Economic Benefits and Costs of Entrepreneurship: A Review of the Research”. In: *Foundations and Trends® in Entrepreneurship* 4.2, pp. 65–154. DOI: 10.1561/0300000012.

Putnam, R. D. (2001). *Bowling alone: the collapse and revival of American community.* 1. touchstone ed. OCLC: 248630671. New York, NY: Simon & Schuster. 541 pp.

Qian, H. (2013). “Diversity Versus Tolerance: The Social Drivers of Innovation and Entrepreneurship in US Cities”. In: *Urban Studies* 50.13, pp. 2718–2735. DOI: 10.1177/0042098013477703.

Raan, A. F. J. van (2004). “Sleeping Beauties in science”. In: *Scientometrics* 59.3, pp. 467–472. DOI: 10.1023/B:SCIE.0000018543.82441.f1.

Rees, F. (2005). *Johannes Gutenberg: Inventor of the Printing Press.* Minneapolis, MN: Compass Point Books.

Reilly, R. R., G. S. Lynn, and Z. H. Aronson (2002). “The role of personality in new product development team performance”. In: *Journal of Engineering and Technology Management* 19.1, pp. 39–58. DOI: 10.1016/S0923-4748(01)00045-5.

Rentfrow, P. J., S. D. Gosling, M. Jokela, D. J. Stillwell, M. Kosinski, and J. Potter (2013). “Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates.” In: *Journal of Personality and Social Psychology* 105.6, pp. 996–1012. DOI: 10.1037/a0034434.

Rentfrow, P. J., S. D. Gosling, and J. Potter (2008). “A Theory of the Emergence, Persistence, and Expression of Geographic Variation in Psychological Characteristics". In: *Perspectives On Psychological Science* 3.5, pp. 339–369. DOI: 10.1111/j.1745-6924.2008.00084.x.

Rentfrow, P. J., M. Jokela, and M. E. Lamb (2015). "Regional Personality Differences in Great Britain". In: *PLOS ONE* 10.3. Ed. by R. D. Latzman, e0122245. DOI: 10.1371/journal.pone.0122245.

Rentfrow, P. J., C. Mellander, and R. Florida (2009). "Happy States of America: A state-level analysis of psychological, economic, and social well-being". In: *Journal of Research in Personality* 43.6, pp. 1073–1082. DOI: 10.1016/j.jrp.2009.08.005.

Riffai, M., K. Grant, and D. Edgar (2012). "Big TAM in Oman: Exploring the promise of on-line banking, its adoption by customers and the challenges of banking in Oman". In: *International Journal of Information Management* 32.3, pp. 239–250. DOI: 10.1016/j.ijinfomgt.2011.11.007.

Rigby, D. L. (2015). "Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes". In: *Regional Studies* 49.11, pp. 1922–1937. DOI: 10.1080/00343404.2013.854878.

Rivkin, J. W. (2000). "Imitation of Complex Strategies". In: *Management Science* 46.6, pp. 824–844. DOI: 10.1287/mnsc.46.6.824.11940.

Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg (2007). "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes". In: *Perspectives on Psychological Science* 2.4, pp. 313–345. DOI: 10.1111/j.1745-6916.2007.00047.x.

Roberts, B. W., K. E. Walton, and W. Viechtbauer (2006). "Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies." In: *Psychological Bulletin* 132.1, pp. 1–25. DOI: 10.1037/0033-2909.132.1.1.

Rodríguez-Pose, A. (1999). "Innovation Prone and Innovation Averse Societies: Economic Performance in Europe". In: *Growth and Change* 30.1, pp. 75–105. DOI: 10.1111/0017-4815.00105.

Rodríguez-Pose, A. and R. Crescenzi (2008). "Mountains in a flat world: why proximity still matters for the location of economic activity". In: *Cambridge Journal of Regions, Economy and Society* 1.3, pp. 371–388. DOI: 10.1093/cjres/rsn011.

Rodríguez-Pose, A. and M. Di Cataldo (2015). "Quality of government and innovative performance in the regions of Europe". In: *Journal of Economic Geography* 15.4, pp. 673–706. DOI: 10.1093/jeg/lbu023.

Rodríguez-Pose, A. and M. Storper (2009). "Better Rules or Stronger Communities? On the Social Foundations of Institutional Change and Its Economic Effects". In: *Economic Geography* 82.1, pp. 1–25. DOI: 10.1111/j.1944-8287.2006.tb00286.x.

Rogers, E. M. (1983). *The Diffusion of Innovations.* 3rd ed. London ; New York: The Free Press.

Romer, P. M. (1986). "Increasing Returns and Long-Run Growth". In: *Journal of Political Economy* 94.5, pp. 1002–1037.

Romer, P. M. (1990). "Endogenous Technological Change". In: *Journal of Political Economy* 98.5, S71–S102. DOI: 10.1086/261725.

Sandberg, B. and L. Aarikka-Stenroos (2014). "What makes it so difficult? A systematic review on barriers to radical innovation". In: *Industrial Marketing Management* 43.8, pp. 1293–1305. DOI: 10.1016/j.indmarman.2014.08.003.

Saxenian, A. (1994). *Regional advantage: culture and competition in Silicon Valley and Route 128*. OCLC: 795833422. Cambridge, Mass.: Harvard Univ. Press. 226 pp.

Scherngell, T. and M. J. Barber (2009). "Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme". In: *Papers in Regional Science* 88.3, pp. 531–546. DOI: 10.1111/j.1435-5957.2008.00215.x.

Schilling, M. A. and E. Green (2011). "Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences". In: *Research Policy* 40.10, pp. 1321–1331. DOI: 10.1016/j.respol.2011.06.009.

Schmidt-Nielsen, K. (1984). *Scaling, why is animal size so important?* Cambridge ; New York: Cambridge University Press. 241 pp.

Schoenmakers, W. and G. Duysters (2010). "The technological origins of radical inventions". In: *Research Policy* 39.8, pp. 1051–1059. DOI: 10.1016/j.respol.2010.05.013.

Scott, A. J. (1993). *Technopolis: high-technology industry and regional development in southern California*. Berkeley: University of California Press. 322 pp.

Simon, H. A. (1962). "The Architecture of Complexity". In: *Proceedings of the American Philosophical Society* 106.6, pp. 467–482.

Sleuwaegen, L. and P. Boiardi (2014). "Creativity and regional innovation: Evidence from EU regions". In: *Research Policy* 43.9, pp. 1508–1522. DOI: 10.1016/j.respol.2014.03.014.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Solow, R. M. (1957). "Technical Change and the Aggregate Production Function". In: *The Review of Economics and Statistics* 39.3, p. 312. DOI: 10.2307/1926047.

Sonn, J. W. and I. K. Park (2011). "The Increasing Importance of Agglomeration Economies Hidden behind Convergence: Geography of Knowledge Production". In: *Urban Studies* 48.10, pp. 2180–2194. DOI: 10.1177/0042098010382679.

Sonn, J. W. and M. Storper (2008). "The Increasing Importance of Geographical Proximity in Knowledge Production: An Analysis of US Patent Citations, 1975–1997". In: *Environment and Planning A* 40.5, pp. 1020–1039. DOI: 10.1068/a3930.

Sorenson, O., J. W. Rivkin, and L. Fleming (2006). "Complexity, networks and knowledge flow". In: *Research Policy* 35.7, pp. 994–1017. DOI: 10.1016/j.respol.2006.05.002.

Steel, G. D., T. Rinne, and J. Fairweather (2012). "Personality, Nations, and Innovation: Relationships Between Personality Traits and National Innovation Scores". In: *Cross-Cultural Research* 46.1, pp. 3–30. DOI: 10.1177/1069397111409124.

Stinchcombe, A. L. (1965). "Social structure and organizations". In: *Handbook of organizations*. Ed. by J. G. March. Chicago: Rand McNally, pp. 142–193.

Stine, D. D. (2009). "The Manhattan Project, the Apollo Program, and Federal Energy Technology R&D Programs: A Comparative Analysis". In: *Congressional Research Service*, pp. 1–10.

Stojkoski, V., Z. Utkovski, and L. Kocarev (2016). “The Impact of Services on Economic Complexity: Service Sophistication as Route for Economic Growth”. In: *PLOS ONE* 11.8. Ed. by I. Sendiña-Nadal, pp. 1–29. DOI: 10.1371/journal.pone.0161633.

Storper, M. and A. J. Scott (2008). “Rethinking human capital, creativity and urban growth”. In: *Journal of Economic Geography* 9.2, pp. 147–167. DOI: 10.1093/jeg/lbn052.

Storper, M. (2010). “Why Does a City Grow? Specialisation, Human Capital or Institutions?” In: *Urban Studies* 47.10, pp. 2027–2050. DOI: 10.1177/0042098009359957.

Storper, M., T. Kemeny, N. P. Makarem, and T. Osman (2015). *The rise and fall of urban economies: lessons from San Francisco and Los Angeles*. Innovation and technology in the world economy. OCLC: 928689919. Stanford, California: Stanford Business Books. 305 pp.

Strambach, S. and B. Klement (2013). “Exploring plasticity in the development path of the automotive industry in Baden-Württemberg: the role of combinatorial knowledge dynamics”. In: *Zeitschrift für Wirtschaftsgeographie* 57.1. DOI: 10.1515/zfw.2013.0006.

Stuetzer, M., D. B. Balkin, M. Obschonka, S. D. Gosling, P. J. Rentfrow, and J. Potter (2018). “Entrepreneurship culture, knowledge spillovers and the growth of regions”. In: *Regional Studies* 52.5, pp. 608–618. DOI: 10.1080/00343404.2017.1294251.

Stuetzer, M., M. Obschonka, D. B. Audretsch, M. Wyrwich, P. J. Rentfrow, M. Coombes, L. Shaw-Taylor, and M. Satchell (2016). “Industry structure, entrepreneurship, and culture: An empirical analysis using historical coalfields”. In: *European Economic Review* 86, pp. 52–72. DOI: 10.1016/j.euroecorev.2015.08.012.

Sveikauskas, L. (1975). “The Productivity of Cities”. In: *The Quarterly Journal of Economics* 89.3, p. 393. DOI: 10.2307/1885259.

Teece, D. J. (1977). “Technology Transfer by Multinational Corporations: The Resource Cost of Transferring Technological Know-How”. In: *The Economic Journal* 87.346, pp. 242–261.

Teece, D. J., G. Pisano, and A. Shuen (1997). “Dynamic capabilities and strategic management”. In: *Strategic Management Journal* 18.7, pp. 509–533. DOI: 10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z.

Teece, D. J., R. Rumelt, G. Dosi, and S. Winter (1994). “Understanding corporate coherence”. In: *Journal of Economic Behavior & Organization* 23.1, pp. 1–30. DOI: 10.1016/0167-2681(94)90094-9.

Tellis, G. J., J. C. Prabhu, and R. K. Chandy (2009). “Radical Innovation Across Nations: The Preeminence of Corporate Culture”. In: *Journal of Marketing* 73.1, pp. 3–23. DOI: 10.1509/jmkg.73.1.3.

Tödtling, F. and M. Trippl (2005). “One size fits all?” In: *Research Policy* 34.8, pp. 1203–1219. DOI: 10.1016/j.respol.2005.01.018.

Töpfer, S., U. Cantner, and H. Graf (2017). “Structural dynamics of innovation networks in German Leading-Edge Clusters”. In: *The Journal of Technology Transfer*. DOI: 10.1007/s10961-017-9642-4.

Trajtenberg, M. (1990). “A Penny for Your Quotes: Patent Citations and the Value of Innovations”. In: *The RAND Journal of Economics* 21.1, p. 172. DOI: 10.2307/2555502.

U.S. Census Bureau (2000). *2000 Census*. U.S. Census Bureau.

U.S. Census Bureau (2010). *2006-2010 American Community Survey*. U.S. Census Bureau.

Usher, A. P. (1954). *A history of mechanical inventions*. New York: Dover. 450 pp.

Utterback, J. M. (1996). *Mastering the dynamics of innovation*. OCLC: 255818088. Boston, Mass: Harvard Business School. 253 pp.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013). “Atypical Combinations and Scientific Impact”. In: *Science* 342.6157, pp. 468–472. DOI: 10.1126/science.1240474.

Uzzi, B. (1997). “Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness”. In: *Administrative Science Quarterly* 42.1, p. 35. DOI: 10.2307/2393808.

Van Noorden, R., B. Maher, and R. Nuzzo (2014). “The top 100 papers”. In: *Nature* 514.7524, pp. 550–553. DOI: 10.1038/514550a.

Varga, A. (1999). “Time-Space Patterns of US Innovation: Stability or Change?” In: *Innovation, Networks and Localities*. Ed. by M. M. Fischer, L. Suarez-Villa, and M. Steiner. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 215–234. DOI: 10.1007/978-3-642-58524-1_10.

Varga, A. (2000). “Local Academic Knowledge Transfers and the Concentration of Economic Activity”. In: *Journal of Regional Science* 40.2, pp. 289–309. DOI: 10.1111/0022-4146.00175.

Verhoeven, D., J. Bakker, and R. Veugelers (2016). “Measuring technological novelty with patent-based indicators”. In: *Research Policy* 45.3, pp. 707–723. DOI: 10.1016/j.respol.2015.11.010.

Verspagen, B. and W. Schoenmakers (2004). “The spatial dimension of patenting by multinational firms in europe”. In: *Journal of Economic Geography* 4.1, pp. 23–42. DOI: 10.1093/jeg/4.1.23.

Vise, D. A. and M. Malseed (2005). *The Google story*. New York: Delacorte Press. 326 pp.

Wanzenböck, I., T. Scherngell, and M. M. Fischer (2013). “How do firm characteristics affect behavioural additionalities of public R&D subsidies? Evidence for the Austrian transport sector”. In: *Technovation* 33.2, pp. 66–77. DOI: 10.1016/j.technovation.2012.11.006.

Weber, A. F. (1899). *The growth of cities in the nineteenth century*. OCLC: 500351606. New York: Macmillan.

Wei, W. et al. (2017). “Regional ambient temperature is associated with human personality”. In: *Nature Human Behaviour* 1.12, pp. 890–895. DOI: 10.1038/s41562-017-0240-0.

West, G. B. (1997). “A General Model for the Origin of Allometric Scaling Laws in Biology”. In: *Science* 276.5309, pp. 122–126. DOI: 10.1126/science.276.5309.122.

Winter, S. G. (1987). “Knowledge and Competence as Strategic Assets”. In: *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*. New York: Harper & Row, Ballinger Division, pp. 165–187.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). "The Increasing Dominance of Teams in Production of Knowledge". In: *Science* 316.5827, pp. 1036–1039. DOI: 10.1126/science.1136099.

Yayavaram, S. and W.-R. Chen (2015). "Changes in firm knowledge couplings and firm innovation performance: The moderating role of technological complexity: Changes in Knowledge Couplings and Innovation Performance". In: *Strategic Management Journal* 36.3, pp. 377–396. DOI: 10.1002/smj.2218.

Youn, H., L. M. A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, and G. B. West (2016). "Scaling and universality in urban economic diversification". In: *Journal of The Royal Society Interface* 13.114, p. 20150937. DOI: 10.1098/rsif.2015.0937.

Zander, U. and B. Kogut (1995). "Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test". In: *Organization Science* 6.1, pp. 76–92. DOI: 10.1287/orsc.6.1.76.

Zook, M. A. (2000). "The Web of Production: The Economic Geography of Commercial Internet Content Production in the United States". In: *Environment and Planning A* 32.3, pp. 411–426. DOI: 10.1068/a32124.

Zucker, L. G. and M. R. Darby (1996). "Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry". In: *Proceedings of the National Academy of Sciences* 93.23, pp. 12709–12716. DOI: 10.1073/pnas.93.23.12709.

Zúñiga-Vicente, J. Á., C. Alonso-Borrego, F. J. Forcadell, and J. I. Galán (2014). "Assessing the Effect of Public Subsidies on Firm R&D Investment: A Survey". In: *Journal of Economic Surveys* 28.1, pp. 36–67. DOI: 10.1111/j.1467-6419.2012.00738.x.

Zwick, T. and K. Frosch (2017). "Attenuation bias when measuring inventive performance". In: *Economics of Innovation and New Technology* 26.3, pp. 195–201. DOI: 10.1080/10438599.2016.1155270.

Zwick, T., K. Frosch, K. Hoisl, and D. Harhoff (2017). "The power of individual-level drivers of inventive performance". In: *Research Policy* 46.1, pp. 121–137. DOI: 10.1016/j.respol.2016.10.007.

Nederlandse samenvatting

Baanbrekende mijlpalen als de eerste gemotoriseerde vlucht van de gebroeders Wright of de maanlanding zijn hoogtepunten in de geschiedenis van de technologie en dragen bij aan een evenwichtig beeld van het grotere geheel. Technologieën beïnvloeden ons dagelijks leven en welzijn en in het bijzonder beïnvloedt technologische kennis de economische ontwikkeling. Abramovitz (1956) en Solow (1957) toonden aan dat conventionele factoren zoals kapitaal en arbeid geen verklaring vormen voor 90 procent van de economische groei in ontwikkelde, geïndustrialiseerde landen. Zij betoogden dat niet de vermeerdering van kapitaal en arbeid, maar andere factoren zoals productiviteitsgroei verantwoordelijk moeten zijn voor het onverklaarde restant. In zijn groeimodel schrijft Solow (1957) productiviteitsgroei toe aan technische vooruitgang, waardoor de productiviteit van kapitaal en arbeid kon toenemen. Solows model wordt het exogene groeimodel genoemd, aangezien technologische verandering niet direct is opgenomen, maar als exogene factor wordt beschouwd. Sindsdien is er veel onderzoek verricht om technologische kennis en vooruitgang op te nemen in economische groeitheorieën en om technologische ontwikkeling te verklaren (Kuznets 1962, Nelson en Winter 1982, Romer 1990, Grossman en Helpman 1991, Aghion en Howitt 1998).

Economische groei is echter niet gelijk verdeeld over landen en zelfs regio's binnen een land vertonen grote verschillen (Fagerberg et al. 1997). Een belangrijke reden voor de uiteenlopende niveaus van economische groei is de ongelijke verdeling van kenniscreatie in regio's (Glaeser et al. 1992, Henderson et al. 2001). In regionale groeimodellen wordt kennis daarom vaak als belangrijke variabele beschouwd voor de verklaring van regionaal economisch succes (bijv. Barrel en Pain 1997, Rodriguez-Pose 1999, Crescenci 2005, Audretsch en Keilbach 2008, Parent en LeSage 2012, Piergiovanni et al. 2012). Door de geografische concentratie van kennis en bijgevolg de uiteenlopende niveaus van kennisproductie in regio's te verklaren, krijgen we een beter inzicht in de ongelijke economische ontwikkeling van regio's.

In de empirische literatuur wordt kenniscreatie echter vaak zuiver als kwantiteit gezien. In modellen voor kennisproductie of economische ontwikkeling wordt kennis vaak eenvoudigweg als kwantitatieve factor ingevoerd zonder dat rekening wordt gehouden met de verschillen in de kwaliteit van de kennis (zie Malecki 2010 voor een bespreking). Uit onderzoek op diverse vakgebieden blijkt echter dat niet alle kennis gelijk is, maar grote kwalitatieve verschillen vertoont (Trajtenberg 1990, Chandy en Tellis 1998, Hargadon 2003).

In de literatuur is een reeks aspecten terug te vinden aan de hand waarvan de kwaliteit van kennis kan worden gedifferentieerd. Zo is nieuwe kennisproductie afhankelijk van de combinatie van bestaande kennis. Verwantschap benadrukt dat niet elke technologie, of meer algemeen, niet alle kennis met dezelfde inspanningen en met hetzelfde succes kan worden gecombineerd. Verwantschap tussen technologieën vergemakkelijkt doeltreffende communicatie en leerprocessen en vermindert onzekerheden en risico's (Nooteboom et al. 2007, Frenken et al. 2007, Neffke et al. 2011). Daarnaast onderscheidt kennis zich wat betreft onderliggende complexiteit. Het ontwikkelen en verbinden van het grote aantal verschillende technologieën dat nodig was voor de maanlanding was wellicht complexer dan het ontwikkelen van het PageRank-algoritme voor de zoekmachine van Google (Fleming en Sorenson 2001). Ook is niet alle kennis even nieuw. Sommige uitvindingen leiden tot radicaal nieuwe producten of processen zoals de drukpers van Gutenberg, terwijl andere beperkte verbeteringen of incrementele wijzigingen vormen (Chandy en Tellis 1998). Kennis heeft bovendien een uiteenlopende impact op de economie, maatschappij en technologie (Trajtenberg 1990). De uitvinding van de eerste automobiel door Daimler en Benz of de eerste gemotoriseerde vlucht van de gebroeders Wright leidde tot de opkomst van nieuwe industrieën en hervormde bestaande technologische paradigma's. Andere uitvindingen worden zelfs geen innovaties. Kortom, (a) verwantschap, (b) complexiteit, (c) mate van nieuwigheid en (d) impact zijn vier belangrijke aspecten van de kwaliteit van kennis en vormen het middelpunt van deze dissertatie. Preciezer gezegd, aan elk van deze vier aspecten wordt een hoofdstuk gewijd. Deze hoofdstukken zullen hierna uitgebreider worden ingeleid.

Verwantschap

Regio's veranderen voortdurend en verwantschap speelt een belangrijke rol in de evolutie van regionale economische structuren (Boschma en Frenken 2011, Neffke et al. 2011). Nieuwe industrieën komen op, terwijl andere verdwijnen. Deze voortdurende verandering is verbonden aan technologische vooruitgang. Nieuwe technologische kennis leidt tot het ontstaan van nieuwe industrieën, zoals de uitvinding van de automobiel rond 1900 in Zuidwest-Duitsland en maakt andere sectoren achterhaald, zoals die van de koetsenmakers. Het concept van gerelateerde kennis geeft ons inzicht in het vermogen van regio's om collectief te leren en zich te ontwikkelen als padafhankelijk proces. Het is onwaarschijnlijk dat regio's los van bestaande competenties nieuwe kennis verwerven. Hidalgo en Hausmann (2007) laten zien dat er een grotere kans bestaat dat landen nieuwe producten exporteren als deze gerelateerd zijn aan het bestaande exportportfolio. Verwantschap heeft niet alleen betrekking op de productdiversificatie van landen, maar ook van regio's (Neffke et al. 2011, Boschma et al. 2015, Rigby 2015, Balland et al. 2018). Derhalve levert de empirische literatuur overvloedig bewijs voor de stelling dat gerelateerde diversificatie eerder regel dan uitzondering is (Hidalgo et al. 2018).

Het concept van verwantschap en gerelateerde diversificatie biedt goed onderbouwde argumenten voor regionaal beleid en toegesneden beleidsregelingen ten gunste van gerelateerde diversificatie. Zo schrijft de EU-strategie voor slimme specialisatie (Smart Specialization) voor dat regio's hun bestaande sterke punten en toekomstige kansen identificeren om subsidies te ontvangen (Foray et al. 2011). In het geval van slimme specialisatie steunt het beleid het padafhankelijke proces van gerelateerde diversificatie waarvoor degelijke argumenten kunnen bestaan, aangezien regio's zich specialiseren en concurrentievoordelen op specifieke gebieden creëren terwijl het beleid het risico op verkeerde investeringen beperkt (Martin en Sunley 2006). Gerelateerde diversificatie als een uiting van padafhankelijkheid is echter misschien niet in staat om mogelijke regionale cognitieve lock-ins te voorkomen. In plaats van bestaande sterke punten te ondersteunen, zoals gebeurt in de huidige EU-strategie, zou het regionale beleid er wellicht ook goed aan doen om niet-gerelateerde in plaats van gerelateerde diversificatie te steunen om cognitieve lock-ins en externe schokken in de regio beter te kunnen opvangen (Frenken et al. 2007). Ongeacht het type diversificatie – gerelateerd of niet-gerelateerd – blijft het nog steeds de vraag in welke mate beleid het padafhankelijke proces van regionale diversificatie kan beïnvloeden. Dit is volgens Boschma en Gianelle (2014, p. 6) de hamvraag.

In hoofdstuk 2 wordt deze leemte in het onderzoek als uitgangspunt genomen met de vraag: Verbreken door de overheid gefinancierde O&O-projecten de padafhankelijkheid van collectief leren in regio's? De empirische aanpak in hoofdstuk 2 gaat uit van octrooien uit de OESO REGPAT-database als indicator van technologische kennis en van informatie over gesubsidieerde O&O-projecten door het Duitse ministerie van Onderwijs en Onderzoek (BMBF). De dataset is gebruikt in een reeks eerdere projecten waarin de relatie tussen overheidssteun en innovatie in regio's werd onderzocht (Fornahl et al. 2011, Broekel en Graf 2012, Broekel 2015). De gegevens van het BMBF bevatten informatie over de begunstigden, doelstellingen, locaties en duur van de gefinancierde projecten. Een zelf vervaardigde concordantie op basis van de database met gegevens over octrooien en bijbehorende subsidies en bedrijven van het Halle Instituut voor Economisch Onderzoek koppelt de gegevens van gesubsidieerde O&O-projecten aan gepatenteerde uitvindingen. Met behulp van deze gegevens is de relatie onderzocht tussen door de overheid gefinancierde O&O-projecten en regionale diversificatie voor 141 Duitse arbeidsmarktregio's tussen 1991 en 2010. De resultaten bevestigen eerdere studies: verwantschap is een belangrijke verklaring voor diversificatie in Duitse arbeidsmarktregio's. Voor het innovatiebeleid geldt dat O&O-subsidies eerder worden toegekend aan gerelateerde activiteiten. Bovendien blijkt er een positief verband te zijn tussen door de overheid gefinancierde O&O-projecten en regionale diversificatie. Eerder onderzoek (Broekel en Graf 2012, Broekel et al. 2017) benadrukt dat O&O-subsidies die worden toegekend aan individuele en gezamenlijke onderzoeksprojecten verschillende effecten hebben. Op basis van de gegevens van het BMBF kan een dergelijk onderscheid worden gemaakt en de empirische resultaten suggereren dat gezamenlijke O&O-projecten de regionale diversificatie sterker beïnvloeden dan individuele. Bovendien kunnen gezamenlijke O&O-projecten tot op zekere hoogte een gebrek aan verwantschap compenseren door de kans op een succesvolle inschrijving te vergroten wanneer de concentratie van verwantschap laag is. Hoewel beleidsvorming deel uitmaakt van de padafhankelijkheid van collectief leren in regio's (aangezien de overheid meer geneigd is middelen toe te kennen aan gerelateerde activiteiten) kunnen gezamenlijke onderzoeksprojecten desalniettemin de diversificatie tot niet-gerelateerde activiteiten bevorderen.

Complexiteit

Klassieke modellen voor endogene groei suggereren dat een toename van O&O-middelen zou moeten leiden tot een sterkere groei van een economie (Romer 1990, Aghion en Howitt 1998). Jones (1995) merkte echter op dat de groei van de VS niet bijzonder toenam, ondanks het feit dat over een lange periode aanzienlijk meer was geïnvesteerd in onderzoek en onderwijs. Pintea en Thompson (2007) brengen deze paradox in verband met de toenemende complexiteit van technologieën en technologische ontwikkeling. Technologieën worden complex genoemd als ze uit een groot aantal onderdelen bestaan en voor hun reproductie grote hoeveelheden informatie vereist is (Simon 1962, Winter 1987, Zander en Kogut 1995). Voor complexe technologieën moeten meerdere, onderling afhankelijke componenten functioneren en kleine fouten kunnen grote problemen opleveren (Sorenson et al. 2006). Jovanovic en Nyarko's (1995) leertheorie geeft aan dat leerprocessen in complexe domeinen daarom moeizamer en trager verlopen en steeds meer O&O-middelen vergen.

In de huidige kenniseconomie is kennis een essentieel hulpmiddel en de complexiteit van kennis kan concrete economische gevolgen hebben. Simpele kennis die eenvoudig te kopiëren is, zal economische actoren waarschijnlijk geen groot groeipotentieel opleveren. Complexe kennis zal daarentegen niet zo snel in handen van concurrenten terechtkomen, aangezien deze moeilijker kan worden verspreid en vormt zo eerder een waardevolle hulpbron (Sorenson et al. 2006). Economische actoren met een concurrentievoorsprong in complexe domeinen zullen derhalve eerder de economische voordelen opstrijken (Kogut en Zander 1992, Zander en Kogut 1995).

Hoewel de economische gevolgen van technologische complexiteit veelvuldig zijn besproken, is er nog weinig empirisch bewijs. Het bestaande empirisch bewijs is om twee hoofdredenen onbevredigend. Ten eerste zijn de empirische benaderingen die worden gebruikt om de impact van technologische complexiteit op regionale economische ontwikkeling te onderzoeken beperkt tot indirect bewijs. Ze gaan ofwel uit van economische complexiteit als impliciete maatstaf voor technologische complexiteit (Hidalgo en Hausmann 2009, Hausmann et al. 2013, Bahar et al. 2014) of ze relateren technologische ontwikkeling indirect aan overeenkomstige economische groei (Petralia et al. 2017, Balland et al. 2018). Ten tweede is het empirisch bewijs hoofdzakelijk beperkt tot nationaal niveau (Hidalgo en Hausmann 2009, Hausmann et al. 2013, Bahar et al. 2014, Petralia et al. 2017). Regio's binnen landen kunnen complexe technologieën echter op heel verschillende manieren produceren (Balland en Rigby 2017). Landen zijn daarom nogal grove ruimtelijke eenheden die geen oog hebben voor de aanzienlijke regionale variatie binnen een land. De vraag in welke mate technologische complexiteit gerelateerd is aan de economische ontwikkeling van regio's blijft derhalve onbeantwoord.

Deze leemte in het onderzoek is de drijfveer voor het empirisch onderzoek in hoofdstuk 3. De hoofdvraag luidt: zijn complexe technologieën van belang voor regionale economische ontwikkeling? In de empirische aanpak wordt uitgegaan van bbp per capita als indicator voor economische groei in 166 Europese NUTS 2-regio's tussen 2000 en 2015. De OESO REGPAT-database biedt informatie over gepatenteerde uitvindingen en dient als indicator voor technologische kennis. Technologische complexiteit wordt gemeten op basis van de complexiteitsindex van Broekel (2019) die structurele diversiteit wordt genoemd. De complexiteitswaarden worden verbonden aan regionale uitvindingsactiviteiten om vast te stellen hoe goed regio's in staat zijn complexe kennis te produceren. De resultaten van de empirische analyse suggereren dat Europese regio's behoorlijk verschillen in hun vermogen om complexe technologieën te produceren. Hoewel complexe kennis zich concentreert in een aantal grote stedelijke gebieden zoals Parijs, Madrid en München, wijst de analyse uit dat het niet uitsluitend om een stedelijk fenomeen gaat. Daarnaast blijkt dat regionale verschillen in de complexiteit van kennis gerelateerd zijn aan economische groei in regio's. Om preciezer te zijn: een toename in complexiteit van tien procent wordt in verband gebracht met een overeenkomstige toename van de regionale economische groei van circa 0,28 procent.

Mate van nieuwigheid

Niet alle nieuwe kennis is even nieuw. In de economische geografie wordt nieuwigheid echter vaak als gegeven vereiste beschouwd voor technologische ontwikkeling, hoewel diverse concepten zoals radicale innovaties (Chandy en Tellis 1998), disruptieve innovaties (Christensen 1997) of technologische doorbraken (Hargadon 2003) substantiële verschillen tussen uitvindingen laten zien wat betreft nieuwigheid. Radicaal nieuwe uitvindingen kunnen bijvoorbeeld bestaande technologische paradigma's hervormen (Chandy en Tellis 1998) en kunnen bijgevolg de opkomst van nieuwe sectoren met aanmerkelijk groeipotentieel bevorderen terwijl andere sectoren achterhaald worden (Christensen 1997). De mate van nieuwigheid heeft daarom tamelijk sterke gevolgen voor de technologische ontwikkeling en heeft het potentieel om de socio-economische ontwikkeling van een regio te transformeren.

Hoofdstuk 4 heeft tot doel een beter inzicht te krijgen in het concept van nieuwigheid en zijn geografische patronen. Het maakt gebruik van de theoretische en empirische inzichten van schaalanalyses die aantonen dat de geografische concentratie van kennisproductie niet willekeurig is, maar grotere steden begunstigt (O'hUallichain 1999, O'hUallichain en Leslie 2005, Bettencourt et al. 2007a, Bettencourt et al. 2007b). Deze auteurs veronderstellen een schaalmodel met de gerapporteerde coëfficiënten, maar geven de productiviteit van steden ten aanzien van innovatie enkel in zuiver kwantitatieve begrippen weer, zonder de verschillende gradaties van nieuwigheid van technologische kennis in aanmerking te nemen. Steden als hotspots voor innovatie brengen echter ook essentiële functionaliteiten en fundamentele middelen samen die innovatie in kwalitatieve begrippen kunnen beïnvloeden.

In hoofdstuk 4 wordt deze onderzoeksleemte onderzocht en de vraag gesteld: zijn steden hotspots van werkelijk nieuwe ideeën? Historische octrooidocumenten van 1836 tot 2010 dienen als empirische grondslag om de onderzoeksvraag te beantwoorden (Petralia et al. 2016). Een tijdspanne van 174 jaar maakt het mogelijk om langetermijntendensen in technologische ontwikkeling te ontrafelen en bevestigt de padafhankelijkheid in kenniscreatie die leidt tot vaste geografische patronen. De empirische benadering in dit hoofdstuk gaat uit van de theoretische conceptualisering van uitvindingen als gevolg van kenniscombinaties. Door uitvindingen te ontleden in de combinaties waaruit ze bestaan, kunnen z scores worden toegepast om een onderscheid te maken tussen atypische (werkelijk nieuwe) en typische (incrementele) combinaties (Schilling en Green 2011, Uzzi et al. 2013, Kim et al. 2016). De schaalanalyse in hoofdstuk 4 laat ten eerste zien dat nieuwigheid zich de afgelopen 174 jaar van de Amerikaanse uitvindingsgeschiedenis steeds meer in grote metropolen concentreert. De productiviteit van steden wat betreft kennisproductie is dus niet zuiver kwantitatief van aard, maar heeft ook betrekking op de nieuwigheid.

Impact

De impact die nieuwe kennis op verdere innovatie heeft, loopt sterk uiteen (Trajtenberg 1990). Onderzoekers beoordelen de impact op verdere kenniscreatieprocessen bijvoorbeeld door gebruik te maken van citatiegegevens in wetenschappelijke publicaties of octrooien (Garfield 1970, Trajtenberg 1990). Citatiegegevens laten zien hoe vaak een bepaald wetenschappelijk artikel of octrooi is gebruikt als input voor latere artikelen of octrooien en maken het daardoor mogelijk een onderscheid te maken tussen kennisresultaten met een grote impact en die met minder impact. Empirisch onderzoek geeft aan dat veel geciteerde resultaten, zoals innovaties met grote impact, waarschijnlijk ook economische waarde opleveren (Trajtenberg 1990).

Hoewel innovatie met een grote impact aanzienlijke regionale gevolgen kan hebben, worden verschillen in de impact van innovaties zelden in aanmerking genomen bij onderzoek op het gebied van economische geografie. Toch zou het van belang moeten zijn of een regio grote hoeveelheden kennis met geringe impact produceert of kennis met veel impact die de regionale ontwikkeling aanmerkelijk kan beïnvloeden. Enkele van de weinige studies in economische geografie die technologische impact opnemen in hun empirische analyses zijn van Ejermo (2009) en Castaldi en Los (2017). Zij benadrukken dat er geografisch gezien een sterkere concentratie van innovaties met grote impact bestaat dan van conventionele innovaties. De sterke concentratie van invloedrijke innovaties werpt meer vragen op. Waarom zijn sommige regio's beter in staat om belangrijke innovaties te produceren dan andere? Castaldi et al. (2015) stellen vast dat een economische structuur die wordt gekenmerkt door niet-gerelateerde diversiteit verband houdt met het verschijnen van innovaties met een grote impact in regio's. Buiten dat wordt er in onderzoek naar economische geografie echter (nog) verbazingwekkend weinig gesproken over de onderliggende redenen waardoor regio's in staat zijn invloedrijke resultaten te produceren.

Hoofdstuk 5 behandelt deze leemte in het onderzoek en presenteert het heersende sociale klimaat ten aanzien van innovatie als een mogelijke verklaring voor de waargenomen regionale variatie in het creëren van innovaties met grote impact. Er wordt gesteld dat met name een regionaal sociaal klimaat dat open staat voor nieuwe ideeën en creatief denken waardeert, van doorslaggevend belang is voor het ontstaan van invloedrijke innovaties. In hoofdstuk 5 wordt daarom voortgebouwd op bestaand onderzoek naar de rol van het open karakter van regio's voor innovatie zonder rekening te houden met het belang ervan voor de creatie van invloedrijke innovaties (Rodríguez-Pose 1999, Florida 2002). Bestaande studies gaan echter uit van tamelijk grove en indirecte indicatoren om te meten hoe open regio's zijn, zoals het aandeel homoseksuelen (Florida 2003). Om deze tekortkomingen te verhelpen, wordt in hoofdstuk 5 gebruik gemaakt van psychologisch onderzoek. Psychologen hebben aanzienlijke vooruitgang geboekt in de conceptualisering en het meten van de persoonlijkheid van individuen via persoonlijkheidskenmerken. De zogenoemde 'grote vijf' zijn vijf persoonlijkheidskenmerken (John en Srivastava 1999). In het bijzonder het kenmerk openheid wordt geassocieerd met innovatie aangezien het beschrijft hoe vindingrijk, creatief, origineel en nieuwsgierig een persoon is (McCrae 1987, King et al. 1996, McCrae 1996, John en Srivastava 1999). In hoofdstuk 5 wordt een dergelijke macro-psychologische benadering gevolgd met de vraag: is een open karakter van regio's van invloed op de kwaliteit van innovatie in regio's?

De empirische aanpak in hoofdstuk 5 gaat uit van octrooigegevens van het Amerikaanse octrooi- en merkenbureau (USPTO) als indicator voor innovatie-activiteiten in 382 grootstedelijke statistische gebieden (MSA) in de Verenigde Staten tussen 2000 en 2010. Net als in voorgaande benaderingen wordt de impact van innovaties gemeten aan de hand van een telling van geciteerde octrooien, gecorrigeerd voor klasse en cohort (Trajtenberg 1990, Cohen et al 2002, Hall et al. 2005). Om precies te zijn, worden regionale innovatie-activiteiten gerangschikt volgens het aantal ontvangen citaten. Hiermee kan een berekening worden gemaakt van de impact van regionale innovatie-activiteiten in verschillende percentielen van de regionale impactverdeling. Om regionale persoonlijkheidsverschillen in openheid vast te leggen, worden de antwoorden van 1,27 miljoen deelnemers van het online persoonlijkheidsproject (Gosling et al. 2004) samengevoegd op MSA-niveau. Uit de empirische resultaten blijkt dat regio's onderling substantieel verschillen in hun vermogen om innovaties met grote impact te produceren. Het effect van een open karakter is het sterkst voor de meest invloedrijke innovaties en vrijwel niet bestaand voor de gemiddelde innovatiekwaliteit van regio's. Kortom, regio's met een open karakter produceren meer innovaties met grote impact.

Conclusies

Deze dissertatie vormt een belangrijke vooruitgang aangezien hierin, in tegenstelling tot eerder onderzoek, wordt benadrukt dat de kwaliteit van kennis een wezenlijke factor is voor de productie van nieuwe kennis. De eerste drukpers, de eerste auto, de eerste gemotoriseerde vlucht en de eerste man op de maan zijn hoogtepunten in de geschiedenis van de technologische ontwikkeling en illustreren dat nieuwe kennis op meerdere vlakken sterk uiteenloopt. Deze dissertatie is daarom gericht op vier belangrijke aspecten: verwantschap, complexiteit, mate van nieuwigheid en impact. Deze aspecten erkennen dat de kwaliteit van kennis meerdere dimensies heeft. In deze dissertatie is aangetoond dat deze kwaliteitsaspecten sterk uiteenlopen tussen regio's. De heterogeniteit van plaatsen is niet alleen van belang voor verschillen in de productie van nieuwe kennis in louter kwantitatieve begrippen, zoals vaak in eerder onderzoek is benadrukt, maar ook van fundamenteel belang om verschillen in kwaliteit van nieuwe kennis in verschillende plaatsen te verklaren. Daarnaast zijn deze kwalitatieve verschillen in kennisproductie tussen regio's essentieel om de ongelijke ontwikkeling van plaatsen te verklaren, bijvoorbeeld wat betreft economische groei en collectief leren. Bij beleid dient daarom de kwaliteit van kennis in aanmerking te worden genomen om regionale ontwikkelingsstrategieën te ontwerpen. De empirische grenzen en de beperkingen van deze dissertatie om concrete beleidsimplicaties af te leiden, laten echter duidelijk zien dat er nog veel meer onderzoek nodig is. Zo kan bijvoorbeeld worden aangenomen dat de kwaliteitsaspecten van elkaar afhankelijk zijn en dit kan bijdragen aan ons inzicht in de kwaliteit van kennis en regionale ontwikkeling. Bovendien zijn de vier aspecten die in deze dissertatie zijn bestudeerd niet exclusief. Er kunnen andere aspecten van kenniskwaliteit zijn die nog niet zijn behandeld en die belangrijke bouwstenen kunnen vormen voor toekomstig onderzoek. Er is meer onderzoek nodig naar hoe de kwaliteit van kennis ons begrip van regionale ontwikkeling kan verbeteren en om beter beleid uit te werken.

Acknowledgments

First of all, I am very grateful to the University of Utrecht and my three supervisors Ron Boschma, Tom Broekel, and Pierre-Alexandre Balland for giving me the opportunity to write this thesis and helping me to finish it. I am very thankful to have the opportunity to work with you, benefit from your experience, and to be inspired by your ideas and thoughts. In particular, I owe a great debt to Tom who convinced me to write this dissertation four years ago at a time when I was struggling and considering a job outside academia. Tom's scientific knowledge, inspiring ideas, and critical mind formed the basis of this dissertation. The exchange and critical discussions of existing research and new ideas is a crucial building block of every scientific work. In countless meetings with Tom, this dissertation has steadily improved.

This dissertation also has a second academic home. I spend most of the time as a PhD student in Hanover. The famous Schneiderberg 50, 3^{rd} floor, was my academic home for the last four years. To be honest, you are not a beauty at the first glance (I am sure you know), and some mornings your smell let your folks wish to turn around and run away. After all, you are warm, cosy, and make decent coffee. Therefore, I would like to thank the Leibniz Universiät of Hanover and the Institute of Economic and Cultural Geography for providing me with the perfect working facilities to conduct this research. As drinking coffee alone is not as tasty as sharing it with wonderful colleagues, I would like to thank all of them in Hanover (of course not only for drinking coffee). Sometimes it is good to have a walk after lunch, which always led us to the small island of Café Kopi. I would like to thank these moments of peace in this cozy coffee shop and the owners for remembering our usual orders. I am very happy that Kerstin Schäfer and Timo Kleiner came from Gießen to Hanover in the middle of my PhD program. Your interest in research, your ideas, and your critical minds have substantially improved my research. You are not only great colleagues, but also became very good friends.

Although a PhD is a lot of fun (besides all the work), the beginning is not always easy. Visiting the first conferences on your own can be tough and creates a feeling of being completely alone in the competitive world of academia. However, you are not. There are many other PhD students out there making exactly the same experience. During the last four years, I met so many other great PhD students at various conferences and workshops. In particular, the Young Economic Geographers Network (YEGN) is a great initiative that provides opportunities to meet friendly peers from around the world. Clearly, it is impossible to list all of you, but I am very thankful to all the PhD students I met in the past years and that somehow contributed to this work.

Research funding is an important cornerstone of our work. Without public funding this piece of work would not have been written. Therefore, I would like to acknowledge the steady support of the Ministry for Science and Culture of Lower

Saxony for funding my research project for the last four years.

To enumerate all the entire cast of people (students, teachers, professors, reviewers, friends, colleagues) who have contributed to my research (directly or indirectly) would amount to writing another book. Still, some people deserve special attention. I would like to thank my hometown crew (Chrischan, Christian, Hinrich, Jan, Jannis, Jos, Ole, Sascha) for all the fun we had and will have together, be that at various Christmas parties or on holidays. We know how to surprise each other. Special thanks goes also out to my dear colleague and good friend Moritz Breul. I always enjoy your company on our trips to the remote and forgotten places in the wild parts of Europe. Talking to you about research and all the other important and funny parts of life was and still is always enriching. Benjamin, my dear friend, thank you for all your inspiring thoughts and your loyalty since several years. Thanks to the Requefort group including Andi, Fabi, Fred, Johnny, Luca, and Tobi: Wir roquen! One should not underestimate the value of having a safe home with friendly people. Therefore, I am very thankful to the roof over my head and to the lovely people that share it with me. Jule, Leonie, Janna, Laila, and Nils, with all of you, it's more than a home. Thanks to Alex, Alix, Mo, and all my friends in Hanover. Your mental support cannot be overstated. A clear mind is also crucial, both for work and happiness generally. Without a clear cut between work and leisure time, I would still have survived the PhD, but with substantially less fun. One way to clear one's mind is climbing, at least for me. Thanks to my good friend Rike who not only belayed me during climbing, but also became a very good friend.

However, after all, I was a lucky man. In world that is as unequal as this one, I was able to write this thesis, because I was lucky in so many different ways. The world I was born into was a different one. Nearly 1,400 km of high metal fences and walls, watchtowers, and minefields divided Germany. In 28 years of German division, 790 people died as they tried to cross the Inner German border. Born in 1988 in East Germany, me and my family were part of this world. I was lucky that the Inner German border opened its gates one year later on November 9, in 1989. No, not because of David Hasselhoff. The German reunification gave me the chance to grow up in a free and democratic country. My life would have definitely been a different without that historic moment and I clearly doubt that this life would have been better. I was lucky that my parents stayed calm in this uncertain time of history and my parents managed the German reunification without losing orientation as so many others. Shortly after reunification, my parents were brave, as they quit their jobs, gave up their old life, and moved from a small village in East Germany, where there was not much of a hopeful future, to a small town near Hamburg called Buxtehude. It became our hometown and our parents guaranteed that we could grow up safely without much to care about. I am lucky that I had the opportunity to attend school, because countless children in this world cannot. Education is key to become a responsible person that cares about others. Education broadens our minds and gives us the ability to reflect. Without reflection, it would be difficult to be thankful.

Therefore, my deepest debt is owed to my family. I am deeply grateful to my parents that they gave me much more than a save home, but most of all patience, care, and trust. You opened my interests for so many inspiring things such as nature including plants and birds, sports, and geography in particular. Without having an Atlas always around, without quizzing German license plates on long car travels and without always asking *"where does this come from?"*, I would not have gained interest in the wonderful world of geography. Your trust gave me the chance to follow my own interest, always and at any time. I would like to thank you for not asking what I will do after I finish geography. I could not have provided you with an answer and I would doubt that a PhD program would have been in my top 10.

I am thankful to my mum who did not lose patience when I was not doing my homework, although I promised I will do them the next time (which I often did not), because playing football and going into the woods was much more interesting. I was lucky to have the best dad in the world, who never gave me strict rules, taught me playing football and how to build bows and arrows. You were always there when I needed you. Words can tell so much, but sometimes they cannot. Sometimes there are no words that can ever express how much I miss you, and I am deeply sad that you have not seen me finishing this little piece of work. Without my family and all the luck I had during my childhood, this thesis would not have been written. Without any doubt, I am most thankful to my family, to my mum and dad.

Curriculum vitae

Lars Mewes was born on October 14, 1988 in Perleberg, Germany. He holds a Bachelor degree in Geography from Philipps-Universität Marburg (2012) and a Master in Economic Geography from Leibniz Universität Hannover (2015). During his studies, he did an exchange semester at Utrecht University (2013). In 2015, he started working at the Institute of Economic and Cultural Geography of the Leibniz Universität Hannover as a research associate. He was awarded with the Best Paper Award of the PhD Workshop on Economic Geography at Utrecht University (2016).