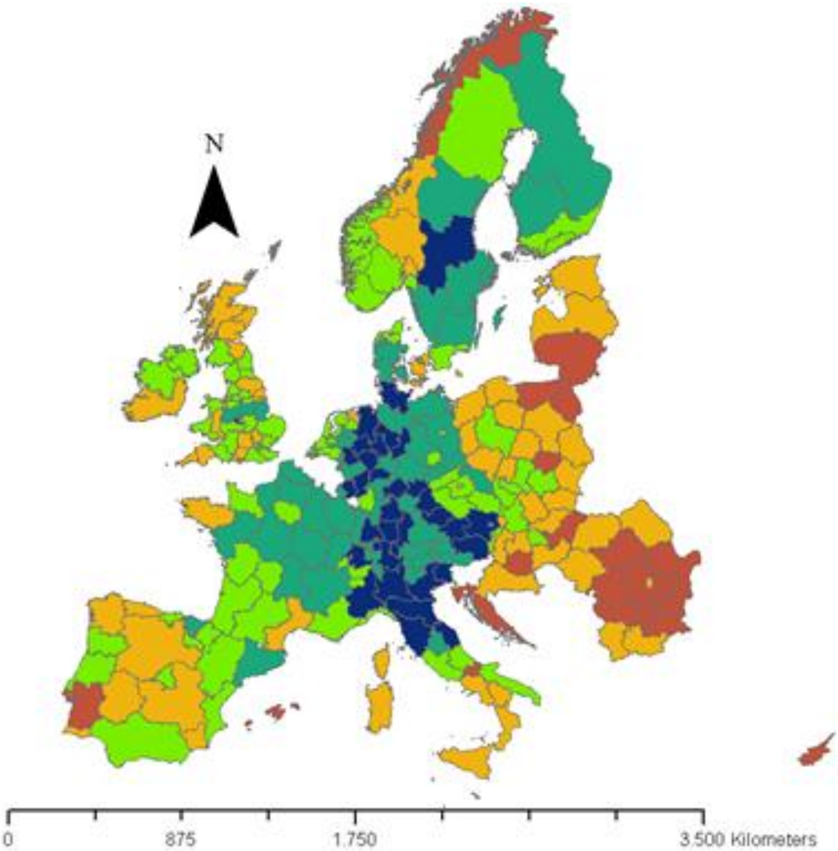
The effect of regional variety on complexity in Europe: Unraveling complexity

A research on the influence of regional technological variety on the complexity of European NUTS-2 regions between 2002 and 2016.



Date: 8 November 2019

Name: Friso Pietersen – 6307256

Utrecht University

Contact: F.m.pietersen@students.uu.nl

Coordinator: Ron Boschma

Foreword

The master thesis marks the end of a challenging and very exciting period in my academic career. It has been a challenge and not always a linear process. However, I'm very delightful with the choice of economic geography. I have been able develop many skills and recombine these for this research. I have learnt many academic skills, such as academic writing and conducting statistical analyses in the program 'R'. I would like to thank Teresa Farinha with helping to pick a very interesting topic in the field of Economic Geography and supporting me. Furthermore, I would like to give a special thanks to Pierre-Alexandre Balland for providing me a large database of patent claims. I'm also very thankful of all the guidance offered by prof.dr. Boschma with finalizing my master thesis.

Friso Pietersen

Abstract

As to this date, the influence of regional variety on regional complexity has not been investigated, whereas these concepts play a central role in the economic geography literature. This study proposes a new method that is based on the influence of the regional technological composition expressed in related, semi-related and unrelated variety on regional complexity. The results show that regions in general lose complexity. Mainly, high complex regions 'lose' technological specializations and face difficulties with maintaining the level of complexity. Furthermore, it is shown that geographical relatedness density and semi-related variety have a strong and positive effect on the level of regional complexity.

Table of contents

1. Introduction

How can a region effectively move from being dependent on resources such as oil and gas, to produce high-tech products? If we would be able to identify such path, it eventually leads to competitive regions on the long-run. More precisely, it would be possible to know when – and how much to invest to create a sustainable, resilient and competitive region. However, successful investments are not guaranteed. As such, the Dutch newspaper 'NRC' stated that European member states are not able to spend their European investment funds. In fact, the total amount of non-spent money increased from 267,3€ billion euro to 281,2 billion euro in 2019 (NRC, 2019). According to European court of auditors a main problem is that the majority of the countries do not know on what they can spend their fund on (NRC, 2019).

The European Union has implemented the smart specialization framework to tackle challenges like these. The main goal of this framework is to identify existing strengths, and look for hidden and novel opportunities on which regions can build competitive advantage (Balland, Boschma, Crespo & Rigby, 2019; Foray, David & Hall, 2009). This is crucial for regions because continuous shifts in the technological landscape are required to maintain regional competitive advantage. However, as illustrated by the previous case, there is not always a suitable destination for EU funds, whereas those regions seek to develop and expand their economic activities (Balland et al., 2019).

Three major concepts in the economic geography literature contribute to the understanding of regional development: *relatedness, economic complexity,* and *technological variety*. Altogether, they bring insight into how such funds could be spent.

Firstly, Hidalgo, Klinger, Barabási, & Hausmann (2007) took a new approach to look at a country's pattern of development. They mapped the allocation of products within a country into a *product-space*. They found that *related* products that are not yet produced, are more likely to be produced when they are related to the existing products in a country. Building upon this research, various studies showed that regions often diversify into related products, industries, skills and knowledge (Hidalgo et al., 2007; Neffke et al., 2011; Boschma, Balland & Kogler, 2014; Balland et al., 2019, Hidalgo et al., 2018).

Secondly, as strong regional differences exist, all products, skills, technologies, and knowledge have a spatially concentrated character (Hidalgo & Hausmann, 2009). This is caused by the nature of tacit knowledge. Tacit knowledge is embodied in networks of human capital – and human capital is only exclusive to few places (Balland et al., 2019). This implies that some forms of knowledge are spatially concentrated and exclusive to very few areas (Balland & Rigby, 2017).

This creates the foundation of complexity and competitive advantage (Balland et al., 2019). As knowledge is a recombination of various fields of technologies (Weitzman, 1998), the *diversity* of technologies within a region fosters new *unique* recombination opportunities. The more recombination opportunities in a small area exist, the more complex knowledge can be developed. For instance, it is similar to building a tower of Lego bricks, the more bricks are added to the tower, the more complex the building becomes – and thus harder to imitate for other people (Hidalgo & Hausmann, 2009). Therefore, complexity is regarded as a source of competitive advantage (Balland & Rigby, 2017).

Besides the principle of relatedness and economic complexity one can understand regional development by looking at the variety of technologies (Frenken et al., 2007). The sectoral and technological composition of a region influence how regions diversify. As such, a smart phone is constructed by a combination of technologies that is related to batteries, chips, video, antennas, audio and the internet (Castaldi et al., 2015). This means that a related sectoral composition, *related variety,* is associated with more recombinant innovation opportunities (Castaldi et al., 2015; Frenken et al., 2007).

The opposite of related variety, *unrelated variety,* has a positive effect on breakthrough innovations, which on its place leads to new emerging industries and employment growth. In this sense, related variety is regarded as a facilitator of the bulk of innovations, while unrelated variety facilitates breakthrough innovations (Castadli et al., 2015).

However, there has not been any study that has investigated the influence of the regional technological composition on complexity, whereas the probability that the technological composition of a region influences the level of complex activities within a region is likely. As of yet, there is no study that has investigated if regional complexity builds upon a related technological structure, or that the influence of semi-related and unrelated activities is more important to foster regional complexity.

This gap of knowledge brings pressure on the competitive advantage of European regions as it is required to adapt to a continuously changing technological landscape (Balland et al., 2019). Furthermore, the mismatch of the EU fund destinations illustrates that identifying suitable destinations for regional investments is very important. Moreover, economic complexity is found to be a source of long-term economic growth (Hidalgo & Hausmann, 2009; Balland & Rigby, 2017; Balland et al., 2019). Besides that, it benefits inhabitants of regions as the complexity of economic activities has a positive effect on income equality (Hartmann, Guevara & Jara-Figueroa, 2017). Consequently, it is imperative to better understand how regions develop complex activities.

Hence, investigating how technological variety influences the complexity of a region is an important contribution to the economic geography literature. Moreover, it brings a stronger foundation for the smart specialization framework as it may provide more insight into how regions have developed complex capabilities. This research takes a new approach to analyze complexity. It proposes a new method that is based on the variety of technologies embedded in a region. It argues that, if related variety as well as unrelated variety support innovation (Castaldi et al., 2015), a composition of both may facilitate new complex activities. Consequently, the aim of this study is to answer the following research question:

"To what extent does regional technological variety, among other factors, influence the level of regional complexity in the EU in the period of 2006-2016?"

The main research question can be disaggregated into two sub-questions:

1. *"To what extent does regional technological variety influence the level of regional complexity?"*

2. *"To what extent does regional complexity, among other factors, influence entry and exit-rates of technological specializations?"*

The second sub-question involves the influence of regional complexity on the introduction of new technological specializations. As argued previously, regions often diversify into related products, industries, skills and knowledge that is related to the existing activities (Hidalgo et al., 2018). On the other hand, it has not been investigated if already complex regions are able to attract new technologies more easily. According to Hidalgo & Hausmann (2009) complex regions are diverse and specialized because of its ubiquity of activities. This would imply that those regions have a high relatedness with technologies that are not yet present in that region. Furthermore, Balland et al. (2019) found that complex technologies are not likely to enter a region, but the probability increases when they are related to the already existing technologies.

However, there has not been any study that investigated to what extent complex regions influence the entry-rates of new technological specializations. This, while Balland & Rigby (2017) found that complexity in various US cities has decreased. It could be possible that complex regions lose momentum as they are not able to adapt to a rapid changing technological landscape. Therefore, the second question involves the influence of regional complexity on the entry-rates of technologies:

This paper will be structured as follows. In the following chapter, the theoretical framework is discussed and elaborates on the concepts of economic complexity, the principle of relatedness and technological variety. The final part of the theoretical framework provides a synthesis of all discussed theories and formulates hypotheses for the analysis. The third chapter discusses the data and methodology. In this section the concepts of knowledge complexity, relatedness (density), and technological variety are operationalized. The fourth chapter presents the analysis in which the results are discussed. Furthermore, the hypotheses and research questions will be answered. The fifth and final chapter will conclude this paper. It also discusses shortcomings and provides new subjects for future research with policy recommendations.

2. Theoretical Framework

In this chapter the concepts of relatedness, complexity and regional variety will be described according to the economic geography literature. These concepts will be discussed to provide a broad, yet narrowed context for the analysis. Firstly, economic complexity will be discussed. Then, the principle of (un)relatedness will be described and its implication on regional diversification. The next paragraph discusses regional variety. Finally, the main theories will be synthesized and used as a foundation for three hypotheses.

2.1, Economic complexity

For many firms and regions competitive advantage correlates with the production of non-ubiquitous activities (Balland & Rigby, 2017; Asheim & Gertler, 2005). Economic success depends on knowledge production and organization, national innovation systems and its diffusion. Therefore, most of the wealth within the boundary of a nation is intangible (Pugliese et al., 2017). Consequently, the exclusiveness of knowledge correlates with long-run economic performance (Balland & Rigby, 2017; Maskell & Malmberg, 1999) and economic growth (Hidalgo and Hausmann, 2009, Balland et al., 2019). Therefore, the existing regional capabilities are associated with the potential approachable level of income (Hidalgo & Hausmann, 2009).

A method to empirically measure complex capabilities of a region is to look at the complexity of products (Hidalgo & Hausmann, 2009) or knowledge (Balland et al., 2019; Balland & Rigby, 2017). The most complex activities are exclusive to relatively few regions, because only very few regions are able to recombine those activities into more complex activities (Balland & Rigby, 2017; Fleming & Sorenson, 2001). That naturally implies that complex activities are associated with the diversity of activities because more combinations of interactions take place to increase efficiency and specialization (Hidalgo & Hausmann, 2009).

Complexity can be further understood by dividing knowledge into two forms. On the one hand, explicit *knowledge* involves the flow of knowledge that is transmittable in formal language or blueprints (Hausmann et al., 2014). This type of knowledge is easier to imitate and does not have a competitive nature. On the other hand, *tacit knowledge* requires time and direct experience of the knowledge in order to transmit and embed the knowledge (Howells, 2002; Pérez-Luño, Medina, Lavado, Rodriguez, 2011). Long-term informal actions, such as face-to-face meetings, informal rules and local knowhow require the exchange of tacit knowledge between actors. As a result, tacit knowledge requires a form of (spatial) proximity and leads to the development of competitive regional capabilities (Hausmann et al., 2014). Hence, it explains why knowledge can be sticky and excluded to only very few regions (Lorentzen, 2008).

Hidalgo & Hausmann (2009) argue that the underlying mechanism of economic complexity is associated with the theorem of division of labor by Adam Smith. It is assumed that wealth is associated with the division of labor. As regions and people tend to specialize, the economic efficiency increases. This results in the idea that the productivity of a region resides on embedded regional capabilities that are non-tradable (Hidalgo & Hausmann, 2009). In this embedded network of regional capabilities, the inhabitants represent a society that reflects the level knowledge. This is formed by the diversity and ubiquity of the knowledge across every individual and their ability to recombine that knowledge into new knowledge (Hausmann et al., 2014). This, however, depends on the size of the market as more participants allow to specialize further.

As such, the level of complex knowledge is determined by the diversity of knowledge across individuals and the possibility to recombine this into new knowledge (Balland & Rigby, 2017; Haussmann et al., 2014). Thus, efficiency-seeking or complex activities are determined by the inhabitants of a region (Hausmann et al., 2014). Consequently, economic differences can be understood by knowledge intensive places, because knowledge intensive activities take place in knowledge intensive places (Balland & Rigby, 2017; Hidalgo et al., 2014). Therefore, one can see regional complexity as a tool that enables us to understand the nature of regional capabilities that explain differences in levels of wealth across the world.

The construct of complexity is very important and crucial for understanding regional competitive advantage. Currently, it was found that regional complexity acts as a source of competitive advantage (Balland & Rigby, 2017; Balland et al., 2019) and income equality (Hartmann, 2017). Furthermore, Balland et al. (2019) showed that regional technological complexity increases if those technologies were related to the existing capabilities. However, relatedness does not necessarily mean that regions naturally become more complex. So far, there is still little known under what circumstances relatedness matters for regions to develop complex capabilities (Boschma, 2017). The following section provides more insight into this by discussing the concept of relatedness.

2.2, Relatedness

In the previous section it has been argued that economic complexity offers insight into how regional differences in terms of wealth and capabilities exist. A central aspect in the field of economic geography is to look at the history of a region in order to explain its current development trajectory (Kogler, 2015). As showed in the previous section, a fundamental aspect behind regional development is the theory of division of labor by Adam Smith. The theorem naturally implied that processes of efficiency (i.e. innovative activities to create comparative advantage) between firms, government and institutions are place-specific (Kogler, 2015; Hidalgo & Hausmann, 2009) as regional differences exist in an open market. Hence, the question arose why and how some regions became locked into development paths that lose momentum, while other regions were able to reinvent themselves through opening new paths of development (Kogler, 2015; Hidalgo et al., 2018).

A prevalent approach in the economic geography literature is the theory of *recombinant innovation* (Weitzman, 1998). The theory of *recombinant innovation* is drawn from the assumption that innovation is a process of recombined search over different technology landscapes (Fleming & Sorenson, 2001; Martin, 2010). Hence, it has been argued that, as innovation is a recombination of activities among different actors, a common base of technological, cognitive, social and geographical capabilities enhances mutual understanding between actors, thus improving the creation of knowledge and innovation (Kogler 2015). This naturally implies that innovation draws on previous knowledge (Frenken et al., 2007) and has a path-dependent nature.

To further understand how economics or regions diversify overtime, the economic geography literature has looked into the field of evolutionary economic geography. A starting point for evolutionary economic geographers is the assumption of creative destruction described by Schumpter (1942, p:82). Essentially, creative destruction involves the ongoing process in which firms strive to achieve competitive advantage through novelty and innovative activities (Kogler, 2015). Consequently, disruptive change and innovative activities cause the economic landscape to change over time.

The principle of relatedness is crucial for understanding how the economic landscape has changed over time (Hidalgo et al., 2018). Relatedness within the economic geography literature is the proximity between products, industries or research areas that require similar knowledge or input (Hidalgo et al., 2018). Asheim, Boschma & Cooke (2011) found that regional development is presumably fostered, if knowledge is able spillover between local sectors. This suggests that the level of relatedness between capabilities (e.g. in terms of an entire industry or a technology field) enables one to understand how processes of (regional) diversification occurred (Hidalgo et al., 2018). This has also been empirically shown in different fields of analysis.

As argued previously, innovation occurs in regions by recombining the pre-existing regional capabilities (Boschma, 2017; Weitzman, 1998). Capabilities can be products (Hidalgo et al., 2007), complexity of knowledge (Balland & Rigby, 2017), professions (Neffke & Henning, 2009; Farinha et al., 2019), or industries (Neffke et al., 2011).

It has been shown that regions branch into industries that are closely related to existing portfolio of industries (Hidalgo et al., 2018). Furthermore, Neffke et al. (2011) found empirical evidence that rise and fall of industries is strongly conditioned by the level of relatedness between industries at the regional level. Moreover, the likelihood of diversifying into previously non-existing capabilities that are related to the existing set of regional capabilities is higher than without such a relation (Neffke & Henning, 2009; Essletzbichler, 2015; Boschma, Balland & Kogler, 2014; Petralia, Balland, & Morrison, 2017).

Consequently, relatedness has shown that the existing experiences, competencies of individuals and entities (often described as capabilities) offers insight into present economic structures (Neffke et al., 2011) as well as future regional trajectories (Kogler, 2015; Boschma, 2017). Therefore, regional capabilities have a path-dependent nature in which the existing capabilities embedded within a region shape and reshape this path-dependent trajectory (Pinheiro et al., 2018; Hidalgo et al., 2018). Within this path-dependent trajectory, the principle of relatedness is the underlying mechanism of (related) diversification that offers insight into why and how some regions became locked into development paths that lose moment, while other regions were able to thrive (Hidalgo et al., 2018).

As argued in the latter section, the process of diversification is determined by the existing capabilities of a region. It is shown that relatedness is a strong predictor of branching into related activities and contributes to the understanding of regional development paths (Boschma & Gianelle, 2013). However, Boschma (2015) argued that the principle of regional resilience should be redefined and understood via an evolutionary perspective. For instance, by understanding what impact a shock (i.e. a collapse of an industry) has on the capacity of a region to branch into a new growth trajectory. This perspective linked regional resilience with the ability to branch into a new growing path. Then, the purpose behind the principle of relatedness is to seek a strategy that leads to new development paths, building on the existing regional capabilities.

This, however, cannot be solely done by looking at relatedness. The importance of the principle of relatedness relies in aborting the impediment of path dependency, i.e. moving away from a resource-dependent economy (Hidalgo et al., 2018). Therefore, if a region were to attract only related activities, it may specialize, but does that contribute to resilience in the long-run and does it contribute to preventing lock-in effect? (Boschma, 2017; Boschma, 2015). Furthermore, if a region were to specialize into only high value industries, there is a high chance that investments fail, because it can create a *cathedral in the desert* situation. Similarly, if a region were to invest into too low value industries, the region risks limiting its further development (Farinha et al., 2019). Therefore, relatedness does not mean that a region should look for over-specialization, but it rather means that a region should understand its unique development path that leads to diversification (Boschma, 2015; hidalgo et al., 2018). The following section discusses (un)relatedness in terms of a technological portfolio.

2.3, Related and unrelated variety

It has been argued that in economic theory one can distinguish between two types of variety. One functions as a source of knowledge spillovers, *Jacobs externalities*, and one that functions as the portfolio to protect regions from external shocks (Frenken et al., 2007). Frenken et al., (2007) distinguished between *related variety* (relatedness), the portfolio of closely related industries that enable local knowledge spillovers, and *unrelated variety*, the present portfolio of unrelated industries that protest a region against external shocks.

As knowledge is a construct of a combination of technologies (Weitzman, 1998), the sectoral composition of a region determines its future opportunities (Ejermo 2005; Castaldi et al., 2015). Hence, it has been argued that related variety facilitates knowledge spillovers because closely related sectors offer more easily recombinant opportunities (Frenken et al., 2007).

Indeed, empirical evidence was found that related variety stimulates employment growth (Frenken et al., 2007) and regional growth (Boschma, Minondo, Navarro, 2012). However, Frenken et al. (2007) found that unrelated variety lowers unemployment growth, confirming the hypothesis that unrelated variety protects regions from external shocks. Furthermore, the findings of Frenken et al. (2007) opened the question whether related variety supports innovative output (Tavasolli & Carbonara, 2014; Castaldi et al., 2015). Moreover, it allowed one to better understand diversification processes and regional shock resilience (Hidalgo et al., 2018; Boschma, 20117).

Recently, more studies made an argument to better understand the function of unrelated activities as in most cases relatedness has been attributed as the main component of regional diversification, while the role of unrelated variety in the process of diversification remains rather unclear (Tanner, 2014; Boschma, 2017). The question arose if unrelated variety may also be an important driver of long-term economic development as it may provide a solution to regional shock resilience, i.e. such as falling industries in regions that were dependent on that industry (Boschma, 2017). For instance, firm innovative performance is optimal when firms are located at the perfect distance (Fitjar, Huber & Rodríguez-Pose, 2016). Furthermore, it has been argued that diversifying into solely related activities may increase the risks of lock-in (Saviotti & Frenken, 2008).

Evidence of the importance of regional variety has been provided by Castaldi et al. (2015). They distinguished between related variety, semi-related variety and unrelated variety. It was argued that technological breakthroughs stem from combinations of unrelated technologies and foster radical innovation, while related variety enhances the bulk of innovations that are incrementally built on related technologies. In their study on technological breakthroughs it was found that related variety as well as unrelated variety play a significant role on innovative outputs of a region, while semi-related variety has a negative influence. The difference is that related variety accounts for the majority of innovative outputs while unrelated variety increased the likelihood of breakthrough innovations (Castaldi et al., 2015).

Similarly, Boschma & Capone (2015) found that unrelated diversification is associated with large jumps in the development of industries and shifts in local capabilities. More studies about unrelatedness found that the entrance of unrelated activities into an economy have a significant effect on future economic growth, depending on regional stage of economic development (Pinheiro et al., 2018). The results showed that attracting unrelated activities mostly occurs in countries that are at an intermediate or advanced state of development. It has also been shown that, as a country has stronger innovative capabilities, the importance of attracting related activities decreases. Therefore, relatedness is less important for regions with stronger knowledge and innovative capabilities while it is more important for developing regions (Xiao, Boschma, & Andersson, 2018).

Although relatedness offers opportunities for diversification strategies, branching into unrelated activities also can be brought up as a diversification strategy. It can be beneficial for some regions that already have a strong existing base of innovative capabilities. Furthermore, it has been suggested that diversification is not only about understanding what to target next, but also knowing when to target a specific product, industry or technology (Alshamsi, Pinheiro & Hidalgo, 2017).

Moreover, Alshamsi, Pinheiro & Hidalgo (2018) provided evidence that targeting highly connected activities and research areas are more beneficial at an intermediate and relatively low level of diversification. It has also been suggested that policy should not always be directed at targeting the most related product, as unrelated activities may also facilitate innovative breakthrough activities. Therefore, diversification strategies should not only be involved in targeting the most related activity, but also focus on when and how a relatively unrelated activity should be attracted (Hidalgo et al., 2018).

2.4, Synthesis of theories and hypotheses

Innovation enables a firm or a region to develop competitive advantage (Schumpeter, 1934). Fleming & Sorenson (2001) see innovation as a continuous recombination of different technological landscapes. Without innovation, the complexity of economic activities cannot be achieved. At the same time, complex activities are more difficult to produce (Hausmann et al., 2014). A general characteristic of complexity is that the theory of labor suggests that complexity is based around place-based activities (Hidalgo & Hausmann, 2009), and thus difficult to replicate if it were to be reproduced in another location (Balland & Rigby, 2017). This is fostered by an eco-system of tacit knowledge, indicating that complex activities are sticky in space (Dougherty & Dunne, 2011).

Methods of Hidalgo & Hausmann (2009), Balland and Rigby (2017) and Farinha et al. (2019) allowed one to empirically measure complexity in terms of products, knowledge, technologies and jobs. However, little is known about the influence of technological variety on regional complexity. So far, it has been investigated that relatedness of technologies has a positive relation with the introduction of more complex activities (Balland et al., 2019).

When the influence of regional variety on regional complexity is discussed it is only known that both related variety and unrelated variety lead to innovation (Castaldi et al., 2015). Yet, it is not known how complexity is influenced by regional variety. Tavassoli & Carbonara (2014) found that related variety enhances innovative input in terms of patent applications. This does, however, tell us nothing about the quality of the output. Questions such as the influence of variety on the quality of innovation, expressed in complexity, should be answered. Does related variety yield a positive effect on complexity? Or are semi-related variety and unrelated variety more important to foster regional complexity? And at what level of complexity does technological variety matter most?

These questions are important to answer as it brings more insight into a diversification dilemma (Balland et al., 2019) of when, and how to introduce new technologies. Perhaps it is possible that regions may benefit more from the introduction of semi-related or unrelated activities than related activities (Boschma, 2017). This study, therefore, focuses to bring more insight into this discourse by analyzing how regional variety influences complexity at the regional level.

Firstly, it is shown that the concept of relatedness predicts entry -and exits of economic activities, resilience and future development trajectories (Neffke & Henning, 2009; Essletzbichler, 2015; Boschma et al., 2014; Petralia et al., 2017). This idea is derived from the recombinant innovation theory (Weitzman, 1998). Furthermore, it has been shown that related activities are more likely to be introduced than unrelated activities. For instance, a high level of related activities with a non-existing activity increases the likelihood that the new activity will enter the regions. Therefore, *related variety has a positive effect on the introduction of new technological specializations* (1).

Besides the positive effect of related technologies on new technological specializations, it is shown that relatedness has a positive effect on the introduction of complex technologies (Balland et al., 2019). Hidalgo and Hausmann (2009) argue that complex countries possess diverse and ubiquitous technologies. Consequently, regions that are diverse, are easier capable of recombining their existing capabilities into new technologies. Thus, the more technologies a region possesses, the more branching opportunities exist. The quotient of branching opportunities can be understood as relatedness density (Balland et al., 2019). Therefore, r*elatedness density has a positive effect on regional complexity* (2).

Derived from the recombinant innovation theory (Weitzman, 1998), it is expected that related variety enhance innovative outputs as related technologies facilitate more technological recombination opportunities. On the other hand, unrelated variety shed another light on the development trajectory of regions. Castaldi et al. (2015) found that unrelated variety and related variety foster innovative output. Furthermore, Boschma & Capone found that unrelated diversification is associated with large jumps in the development of industries and local capabilities. Alshami et al. (2018) showed that regions should not always be directed at targeting the most related activity, as this is more important at a lower state of development. Moreover, the importance of attracting related activities to the existing capabilities depends on the stage of development and the innovative capabilities (Alshami et al., 2018; Xiao et al., 2018).

Yet, the effect of related of unrelated variety on regional complexity has not been investigated before. Whereas, both related variety as unrelated variety account for innovation, which changes the perspective of recombinant innovation (Castaldi et al., 2015). Mainly unrelated variety was the driver of radical breakthroughs, while related variety accounted for the bulk of innovations. Therefore, it is expected that a synthesis of related variety, semi-related variety and unrelated variety is an important driver of regional complexity as each form of variety enhances innovation, and thus complexity. This leads us to the final hypothesis of this study, *technological variety, both related, semi-related as well as unrelated, have a positive effect on regional complexity* (3).

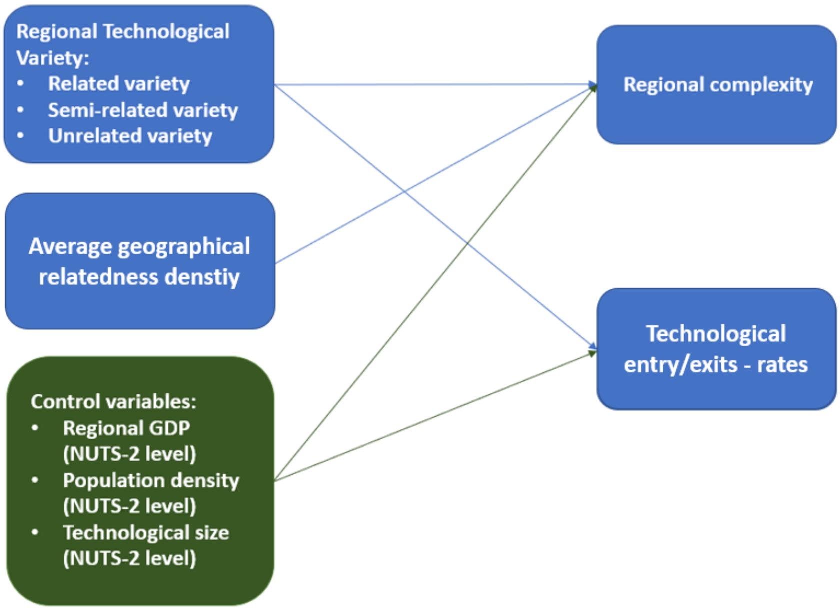


Figure 1: Conceptual model. (Source: Author)

As illustrated by the sub-questions, the two independent variables in this model are regional complexity and technological entries/exits. Four explanatory variables are used to analyze the influence on both independent variables. Firstly, regional technological variety is used and split-up between related variety, semi-related variety and unrelated variety. A similar distinction has been made in Castaldi et al. (2015). Furthermore, the average geographical relatedness density, based on geographical co-occurrences has been used. Finally, regional GDP, population density and technological size have been used as control effects. The summation of every hypothesis is as follows:

Hypothesis 1

Technological related variety has a positive effect on the entrance of new technological specializations at the regional level.

Hypothesis 2

Average geographical relatedness density has a positive effect on regional complexity.

Hypothesis 3

Technological variety, both related as well as unrelated, have a positive effect on regional complexity at the regional level.

3. Data

Smart specialization is aimed to identify technological opportunities and is built on the regional capabilities of the region and its potential to upgrade technologies. In the economic geography literature patent data has often been used to measure regional capabilities empirically (Balland et al., 2019). Therefore, in this research, the main source of data is the OECD-REGPAT database. The OECD-REGPAT derives from PATSTAT which entails all patent applications filed to the EPO between 1977 and 2016 (Balland et al., 2019). The data include all countries of the OECD. Although the geographical coverage of the data accounts for all OECD countries, I will solely investigate the 28 member-states of the EU. This is done to bring more focus to the research. The dataset contains the tracking of all relevant stages of the patent life-cycle for this research, including location and dates.

The unit of analysis will be the NUTS[1]-2 level regions of the 28 European OECD countries. The NUTS-2 level is a geographical nomenclature that subdivides the European Union (EU) into regions at various levels of aggregation (Eurostat, n.d.). Although NUTS-2 regions do not account for an exact economic unit of analysis, it has been used more often in the economic geography literature (Balland et al., 2019) as it accounts for basic regions for the application of regional policies (Eurostat, n.d.). Furthermore, the dataset entails 290 NUTS-2 regions which are not equally divided across all European countries, i.e. Germany contains 39 NUTS-2 regions while Denmark has 6 NUTS-2 regions.

Another feature of the database is that the patent database distinguishes between 655 patent classes, which is aggregated at three levels[2]. Patents at the one-digit level show large differences, while patents at the three-digit level show more similarities because they share common knowledge input (Hall et al., 2003). The ability to measure distances between technology patents enables worldwide patent offices to group patents systematically under groups, classes and subclasses and further aggregates over space and time (Balland & Rigby, 2017). This offers insight into the history, geography and technological characteristics of invention (Petralia, Balland & Rigby, 2016) and allows one to represent complexity (Hidalgo & Hausmann, 2009; Balland and Rigby, 2017), relatedness (Boschma et al., 2014) and variety of the technological landscape (Castaldi et al., 2015).

To compare entry rates of technologies (expressed in patents) over a period of time, a selection has been made of three non-overlapping periods of time, 2002-2006(period 1), 2007-2011(period 2) and 2012-2016(period 3). The data accounts for the total number of patents granted in each 5-year period, i.e. all patents granted in technology *i* in period 1[3]. This provides insight in the evolution of technology entrances over a long time-span and reduces potential noise of outliers.

This research has two dependent variables, the *average regional complexity level* and *technological entrances*. Both dependent variables offer insight into the geography of complex capabilities and the history of regional diversification processes. Furthermore, it enables one to bring more depth to the smart specialization framework proposed by Balland et al. (2019).

Average relatedness density and technological variety are used as explanatory variables - and will be discussed profoundly in the following sections. Three types of variety can be distinguished: Related Variety (RV), Semi-related Variety (SRV), and Unrelated Variety (UV). By unpacking the regional composition in its three types of variety, it may predict that a certain level of variety is optimal for attracting new technologies into the respective region. This has never been researched before and may lead to new profound knowledge on regional complexity. The main goal is to find what level of technological variety and relatedness density is ideal to limit the risks of failed investments.

This chapter consists of three parts. The first part entails operationalizing the independent variables, the entry -and exit-rates of technologies for each region. Then, I will proceed to identify geographical relatedness (density), technological complexity and computing the level three levels of regional

---

[1] *Nomenclature of territorial units for statistics.*

[2] *An in-depth explanation will be given in section 3.4*

[3] *Thus, there is no distinction between the year of 2002 and 2003, but only between the period as a whole, 2002-2006 and the following period.*

variety. The final part involves the analysis of various regression models to check for significant effects on regional complexity and technological growth.

3.2, Explanatory variables

The creation of a technology space is similar to the method of reflections by Hidalgo & Hausmann (2007), in which products are connected with countries. In this research the technology space connects technology classes to regions, in which each set of technologies represents a field of specialized knowledge in a bipartite network. Hence, the geography of technology production, expressed in granted patents in technologies can be made. The connection of technology classes to regions, represented in a network, show what type of technology (i) is present in region (r) in time (t). A very simplified network representation will be shown in figure 2, in which T accounts for technologies and R accounts for regions.

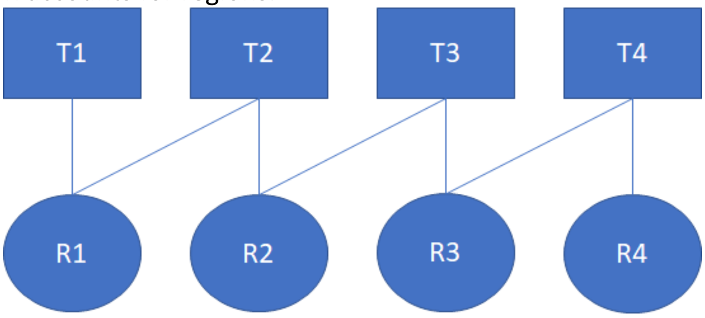


Figure 2: Simplified bipartite network of technologies and regions

The first variable that will be explained is technological complexity. Following the method of reflections by Hidalgo & Hausmann (2009), a region possesses complex products if it produces products that relatively few other countries are able to reproduce. In the case of this research, the first step is to construct an n*k matrix of 290 regions and 655 CPC classes, in which each cell C_{ij} represents the number of patents in technology *i* in region *r* in time *t*[4]. To reduce noise and solely select relevant technologies, only regions with revealed comparative advantage (RCA) in terms of patenting activity are selected (Balland et al., 2019; Boschma, Balland & Kogler, 2014; Hidalgo et al., 2007). The RCA measure is an indication of revealed comparative advantage from a region – technology matrix (Balland, 2017). To capture the RCA, the share of patents in a given technology *i* in region *r* in time *t,* is divided by the total number of patents granted in technology *i for* all regions in time *t*:

$$\text{RCA} = 1 \text{ if; } \frac{patents_{r,t}(i) / \sum_i patents_{r,t}(i)}{\sum_r patents_{r,t}(i) / \sum_r \sum_i patents_{r,t}(i)} > 1$$

Although, Hidalgo & Hausmann (2009) offer insight in measuring complexity of products in their method of reflections, a similar method can be applied for computing the technological complexity index (TCI), based on the *diversity* of regions and the *ubiquity* of technologies. It is given that the higher the amount of RCA's in a region is, the more diverse the region is. The other variable, ubiquity, is given by the number of regions that exhibit RCA in a particular technology. This is the sum of technologies embedded in a region weighted by their ubiquity (Balland & Rigby, 2017). In the following formulas $M_{r,i}$ is a representation of the regional-technology matrix. Based on the number of RCA's, $K_{r,0}$ is a representation of the degree of centrality of the region in the region-technology matrix in region *r*. $K_{i,0}$ is the degree of centrality given by the number of regions that possess and RCA in a particular technology:

[4] *In the case of this research time t refers to one of the three periods, i.e. 2002-2006.*

$$Diversity = K_{r,0} = \sum_{i} M_{r,i}$$

$$Ubiquity = K_{i,0} = \sum_{r} M_{r,i}$$

However, to construct the knowledge complexity index the eigenvector method proposed by Balland & Rigby (2017) is applied[5]. The starting point method is to compute a binary-valued matrix that connects a region to a technology class in which the region exhibits an RCA. For instance, if the region were to have a specialization in technology *i,* it is given the value 1 – and 0 otherwise. This matrix (M) has the dimensions of *n*= 290 regions by *k*=655 CPC-classes. As mentioned previously, complex technologies are relatively low represented in all regions (ubiquity) and are often found in regions that possess a high number of RCA's (diverse regions). In the case of the matrix M, diversity is the number of columns in which region *r* possesses an RCA in a given technology *i* and ubiquity is captured by measuring how often RCA technology *i* is possessed by all regions (Balland and Rigby, 2017).

Following the eigenvector method, the next step is to row standardize matrix M and its transpose M^t. Then, product matrix (B) = $(M * M^T)$ is computed by multiplying the transposed ubiquity of technologies with the diversity of regions. The result is a square matrix of 290 regions * 290 regions. If the order of the multiplication is reversed it results in a 655 technologies * 655 technologies matrix, indicating the technological complexity for each technology class *i* (Balland et al., 2019; Balland and Rigby, 2017). Thus, it possible to calculate the complexity of a region and a specific technology. Finally, by capturing the second eigenvector of the squared matrix B, the order of the matrix is normalized and transformed into a list of rankings with regional complexity. This is standardized as follows:

$$KCI_i = \frac{\vec{Q} - \langle\vec{Q}\rangle}{stdev(Q)}$$

A representation of the most complex regions in Europe in the third period is illustrated in the map below. The average regional complexity is scaled between 0 and 100, in which 100 means the most complex region.

[5] *This method is based on the method of reflection and has been applied in Balland et al. (2018) and Farinha et al. (2019) to calculate a complexity index.*

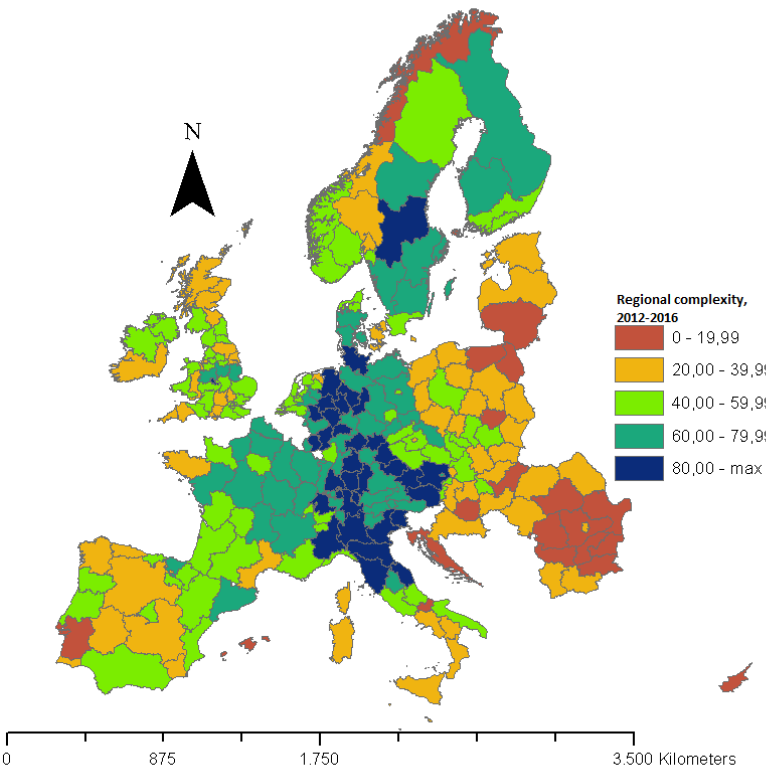


Figure 3: Average regional complexity in period 3 (2012-2016), (Source: Author).'

3.3, Geographical relatedness and relatedness density

Relatedness is computed as the frequency that two patent classes co-occur based on their geographical co-occurrence. Similar to measuring technological complexity, the first step is to identify for every technological class *i,* its revealed comparative advantage (RCA)[6]. Thus, if region *r* has more than an average number of patents granted in technology *i*, technology *i* is an RCA for region *r.* As a result, the RCA is a region-technology matrix in which 1 indicates that a region is specialized in technology *i*, while everything <1 indicates that there is no specialization in technology *i*. The RCA matrix is represented in a 290 regions * 655 technologies binary matrix. Thus, in a similar fashion with technological complexity, geographical relatedness is based on regional specialization.

Then, the geographical measure of relatedness is captured by measuring for each technology *i,* the number of other technologies *j* that co-occur within region *r* in period *t*. This is done by capturing the transpose of the RCA matrix, resulting in a technology-technology matrix that accounts for the relatedness between each 655 technology classes, based on their co-occurrences.

Furthermore, the standardization method by Van Eck & Waltman (2009) has been used. This standardization method is implemented in the relatedness function of the EconGeo R package (Balland, 2017) and involves a conditional-probability-based measurement, which evaluates whether the observed co-occurrences are higher than the expected values based on a probability calculus. The conditional relatedness formula is represented as follows:

$$GeoRelatedness(M_{ji}, S_j, S_i, T) = M_{ji}/(N_{co} * ((S_j/T) * S_i/(T - S_j) + (S_i/T) * (S_j/T - S_i)))$$

In this formula M_{ji} represents the number of co-occurrences of technology classes *i* and *j.* S_j, S_i are respectively the number of co-occurrences of technology class *j* and *i*. Co-occurrences are based on the RCA, as co-occurrences are computed as a function of specialization in region *r*. T is the sum of

[6] *This is the same formula described in the previous section.*

every regional technology-class specialisation and M is the sum of all co-occurrences. The outcomes of technological relatedness are lower-bounded by zero, while they are not upper bounded, hence there is no maximum value. Therefore, if the outcome of the formula were to be zero, technology class *i* and *j* never co-locate within each of the 290 regions, while an outcome of higher than one indicates that two technology classes co-locate more often than is expected by chance (Farinha et al., 2019).

An interesting feature of geographical relatedness is that it can be illustrated in a tree-span network to provide a clear view on the technological structure in the EU. This network is built upon a Minimum Spanning Tree network (Farinha et al., 2019) that allows one to create an overview of the main links that connect the technological structure of Europe. In the case of the technology space, the nodes are the technologies and the ties represent their respective levels of relatedness. Hence, this provides an overview of the core technologies and its structure within the EU. The geographical relatedness of each technology class is illustrated in figure 4. As shown in figure 4, technologies have a tendency to cluster. For instance, various highlighted technologies show that the technology is connected with a large number of other technologies and is positioned as a central node.

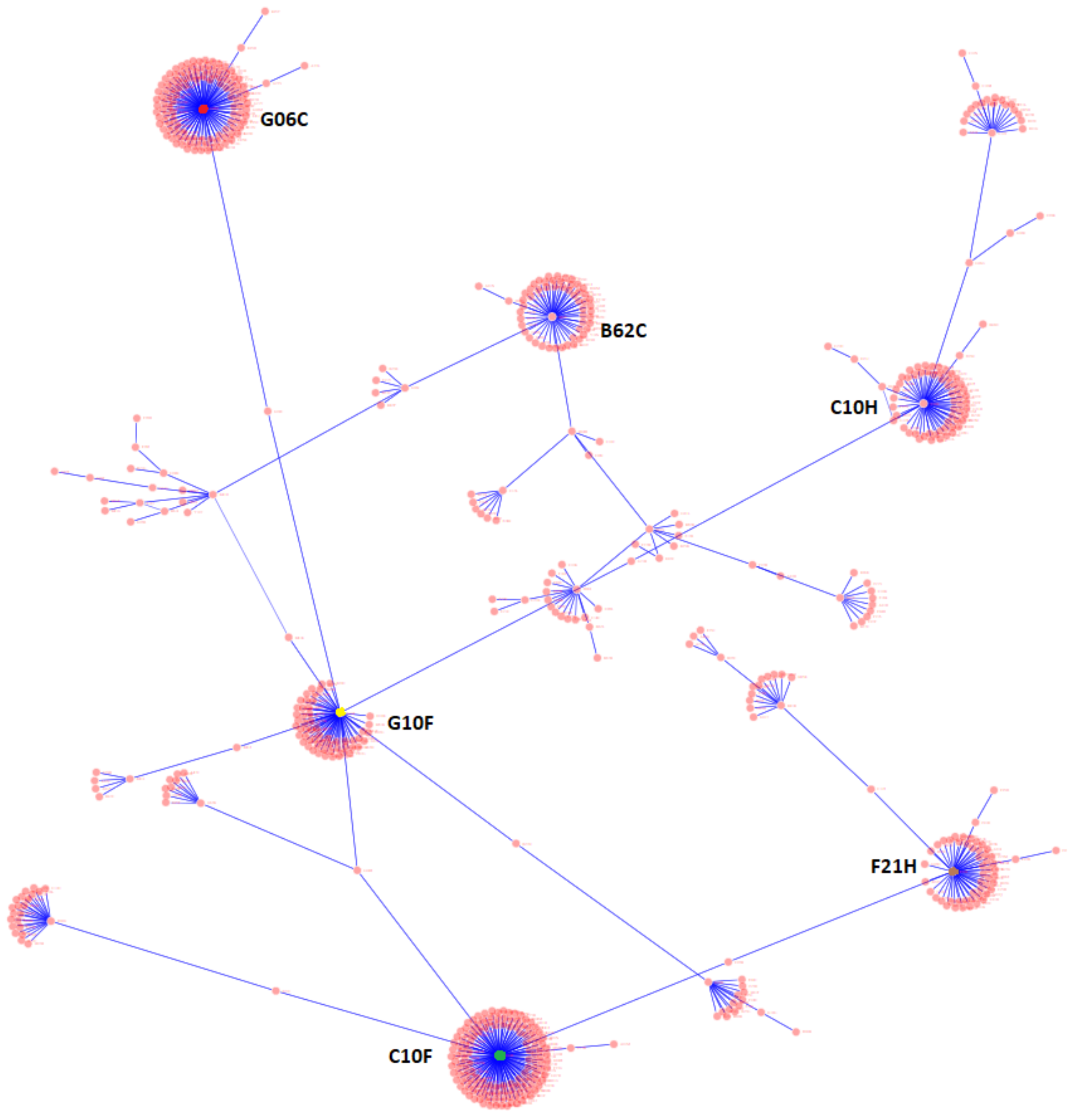


Figure 4, Geographical relatedness of technologies in period 1, 2002-2006.

The final step is to calculate the geographical relatedness density. This is done to analyze the relatedness between technological specializations in a regional portfolio to another, previously non-existing, technological specialization (Balland et al., 2019). Relatedness density is thus an overall average score that indicates branching opportunities for regions to develop new technologies. A high level of relatedness density indicates more opportunities for branching into previously non-existing technologies, while a low level of relatedness density indicates the opposite. The relatedness density measure is derived from the relatedness of technology class *i* to all other technology classes *j* in region *r*, divided by the sum of relatedness of technology class *i* to all other technology classes *j* in all 290 regions. Multiplying the outcome by 100 gives a percentage of related technology-class specializations within a region.

$$GeoRelDensity_{i,r,t} = \frac{\sum_{j \epsilon r \neq i} \emptyset ij}{\sum_{j \neq i} \emptyset ij} * 100$$

For instance, if RCA technology class *i* is related (in terms of geographical co-occurrences) to 5 other RCA technology-classes *j*, and if region *r* were to have 1 RCA technology in its portfolio, then the relatedness density of technology class *i* and region *r* will be 20%, (1/5)*100= 20%. Hence, a relatedness density of 0% is an indication that no other technology is related to technology *i* in region *r*, or the other way around if the relatedness density were to be 100% (Boschma et al., 2014).

3.4, Capturing regional technological variety in all forms

Regional variety can be considered as a form of technological distance (or proximity) between patents by using the information of references, classifications, and inventor identities in patent documents (Yan & Luo, 2017). Every technology class belongs to a certain class in which each technology field has a different distance to another field (Breschi, Lissoni, & Malerba, 2003; Yan & Luo, 2017). The ability to measure distances between technology patents enables worldwide patent offices to group patents systematically under groups, classes and subclasses and further aggregates over space and time (Balland & Rigby, 2017).

Essentially, the characteristics of every granted patent identifies to which distinct technology class the patent belongs. Hence, the technology code of a patent represents its distance/proximity with every other technology. Regarding this research there are 655 technology classes, which belong to 126 subgroups and 9 main technology groups. Thus, representing technology classes at the 3-digit level, the 2-digit level, and the 1-digit level. Regional variety can be found by computing the relationship between each technology class' relationship for every digit-level.

In Frenken et al. (2007) and Castalid et al. (2015) regional variety has been measured with entropy measures. Entropy is derived from information theory and enables one to capture the uncertainty of probability distributions based on the amount of information that is available. Essentially, the more information there is available, the more likely the probability is that a certain event occurs. For instance, if there were to be a region with only 1 patent granted in one of the 655 patent classes, the uncertainty of where the patent belongs will be zero, as the patent belongs to one class. However, if a region were to have a total of 10.000 patents equally distributed over each of the 655 patent classes, the uncertainty of where the patent belongs will be very high as the probability that it belongs to technological class *i* will be much lower. Therefore, entropy in terms of granted patents, tells us how diversified a region is. The entropy level of H is given by:

$$H = \sum_{i=1}^{n} p_i ln(\frac{1}{p_i})$$

with:

$$p_i ln(\frac{1}{p_i}) = 0 \; if \; p_i = 0$$

$E_i$is the event that region *r* is patenting in any technological class *i*, and $p_i$is the probability of event $E_i$occurs with *i* depending on the number of *n* total technology classes. As illustrated by the previous example, the entropy level is bounded below by zero because an entropy level of 0 indicates that there will be no uncertainty that event $E_i$occurs.

Furthermore, entropy is decomposable at different levels of aggregation (Castaldi et al., 2015; Frenken et al., 2007). This feature allows one to compute the entropy level H for each finer-grained level of technological class. In the case of the data, all 655 patent classes can be grouped into 126 groups, which on their place can be grouped into 9 technological categories. Each level of aggregation represents a finer-grained description of the patent itself.

For instance, technology class *A01B*, "*Soil working in agriculture or forestry*", can be grouped under subcategory *A01*, *agriculture,* and agriculture can be placed under category *A, human necessities*. Thus, the sum of all technology classes represents one subcategory, and the sum of all subcategories represent one category. In terms of variety, the 3-digit level patent classes A01B and A01C have more in common than A01 and A21 as they are different subgroups, -and A01 has more in common with A21 than with B01 as they are different categories at the 1-digit level. Therefore, each level of aggregation represents a type of variety. Similar to Castaldi et al. (2015), entropy has been distinguished into three levels of regional variety: *related variety (3-digit patent class), semi-related variety (2-digit patent class) and unrelated variety (1-digit patent class)*.

The aggregation of patents in smaller subgroups is discussed in the decomposition theory of Theil (1972). This theorem entails the relationship between the between-group entropy $H_0$and the entropy level *H* at the level of events. For this research, it is suggested that technological variety (at the three-digit level) is given by the sum of technological variety at the two-digit level and one-digit level (Castaldi et al., 2015). Consequently, the first measured level of variety is *Unrelated Variety*. Unrelated Variety is given by the entropy at the one-digit level, which entails the sum of all patents for each of the major 9 different patent categories (i.e. the sum of all technology classes in group A, human necessities in a specific region):

$$UV_{rt} = \sum_{k=1}^{9} \quad s_{k,rt} ln(\frac{1}{s_{k,rt}})$$

$S_{k,rt}$is derived from the share of patents in technology group *K* in region *r* in time *t*. Following the decomposition theory, *Semi-related variety* is derived from the entropy at the two-digit level minus the entropy at the one-digit level. In other words, SRV is derived from the variety at the subgroup level minus the variety at the group level:

$$SRV_{rt} = \sum_{l=1}^{126} \quad s_{l,rt} ln(\frac{1}{s_{l,rt}}) - \sum_{k=1}^{9} \quad s_{k,rt} ln(\frac{1}{s_{k,rt}})$$

where *l* represents the technological subgroup. Consequently, *Related Variety* is derived from the entropy at the three-digit level minus the entropy at the two-digit level. Similarly, related variety is the variety of all technological classes minus the variety at the subgroup level:

$$RV_{rt} = \sum_{i=1}^{655} \quad s_{i,rt} ln(\frac{1}{s_{i,rt}}) - \sum_{l=1}^{126} \quad s_{l,rt} ln(\frac{1}{s_{l,rt}})$$

in which *i* represents all technological classes.

It should be noted that the level of variety is not mutually exclusive. For instance, a region can be characterized by a high level of related variety and a high level of unrelated variety (Castaldi et al., 2015; Frenken et al., 2007). These regions might be specialized, and therefore possess a large number of related technological classes, while the technological portfolio also includes a large number of unrelated industries. This brings difficulties when it comes to comparing a region that is characterized with a high level of related variety with similarities. In the case of this research there is no single variable computed as technological similarity. Thus, for the final regression model each measurement of variety will be used in the model to see what the effects are on the entrance of technologies.

3.5, Econometric models

It has already been investigated that relatedness density and technological complexity influence the probability that a region specializes in previously non-existing technologies (Balland et al., 2019; Essletzbichler, 2015; Boschma et al., 2014). In the case of this research the dependent variables are *regional complexity* and *technological entrances.* However, regional complexity has never been used as a dependent variable before, while the entry of technologies has been used as a dependent variable before.

Various control variables will be used for the final econometric model. These are included to control for the explanatory variables as it is possible that the significance or the effect of the variables decreases after including these control variables. There are three variables at the regional level, GDP per capita (1), population density (2) and number of regional technological claims (3).

First of all, GDP[7] per capita is used as it was found that it is an important driver of technological diversification (Balland et al., 2019; Petralia et al., 2017). Furthermore, population density accounts for agglomeration effects[8] (Balland et al., 2019; Boschma et al., 2014). Finally, technological size reflects the potential of a region to recombine knowledge into new knowledge (Balland et al., 2019). As the main source of patent data spans a timeframe of 5 non-overlapping years, and the fact that population density changes slightly over time, the average regional GDP and population density for every 5 years will be taken. Finally, technological claims accounts for the possibility to recombine knowledge, thus, enabling a region to branch into new technologies.

However, as explained previously, an RCA specialization in a region is divided by the total sum of patents granted in technology *i*. Thus, the calculation of RCA's implies that a regional specialization is caused by two types of relationships. On the one hand, it implies that region *r* is specialized due to its high level of specialization in technology *i*. On the other hand, it implies that region *r* is specialized in technology *i* because of its relative number of granted patents in technology *i* with respect to other regions. Therefore, being specialized in technology *i* may be caused by the average level of specialization in the EU as a whole. In other words, an exit of technology *i* in region *r* can be caused by an increase in the total number of granted patents in technology *i* across all regions, while the share of patents did not decrease in region *r*.

To overcome this double relationship a time-lag model for both periods has been included. Hence, the influence of all explanatory variables on the dependent variables in T_{+1} has been investigated. Furthermore, the dataset has been split between one time-lagged model for the second period and one time-lagged model for the third period[9]. This means that the models investigate in two periods what the influence of the explanatory models is on complexity in T_{+1}. The econometric equation for the first and second model is written in the following way:

[7] *The GDP per capita is derived from Eurostat, https://ec.europa.eu/.*

[8] *The population density is derived from Eurostat, https://ec.europa.eu. It accounts for the population density of all regions besides regions UKI1 and UKI2.*

[9] *This selection has been made because the overall complexity levels decreased largely between the second and third period. (See section 4.1 for more information).*

$$Y_{r,t} = [Complexity_{r.t}]$$

$$Y_{r,t} = \beta_1 RV_{r,t-1} + \beta_2 SRV_{r,t-1} + \beta_3 UV_{r,t-1} + \beta_4 GDP_{r,t-1} + \beta_5 GeoRel.Density_{r,t-1} + \beta_6 PopDensity_{r,t-1} + \beta_7 Technologicalclaims_{r,t-1}$$

To bring more insight into whether differences exist by the level of complexity, two models are added. One model accounts for the 50% lowest complex regions – and the other model accounts for the 50% highest complex regions. Both models are time-lagged and cover two periods. The third and the fourth model are respectively represented in the following manner:

$Y_{r,t} = [50\%\ low\ Complexity_{r.t}]$ with N = 290

$$Y_{r,t} = \beta_1 RV_{r,t-1} + \beta_2 SRV_{r,t-1} + \beta_3 UV_{r,t-1} + \beta_4 GDP_{r,t-1} + \beta_5 GeoRel.Density_{r,t-1} + \beta_6 PopDensity_{r,t-1} + \beta_7 Technologicalclaims_{r,t-1}$$

and:

$Y_{r,t} = [50\%\ high\ Complexity_{r.t}]$ with N = 290

$$Y_{r,t} = \beta_1 RV_{r,t-1} + \beta_2 SRV_{r,t-1} + \beta_3 UV_{r,t-1} + \beta_4 GDP_{r,t-1} + \beta_5 GeoRel.Density_{r,t-1} + \beta_6 PopDensity_{r,t-1} + \beta_7 Technologicalclaims_{r,t-1}$$

Furthermore, a technological entry model has been used to test hypothesis 1. These models unravel the impact of all explanatory variables on technological entrances. The starting point is to compute an entry-quotient of technologies. That is the change in entries between period T and T_{+1}, divided by the maximum possible entries. Consequently, a region that already possesses a large number of technologies will have a higher quotient than a region with a small technological portfolio. Technological entry is measured as the change of the sum of RCA technology classes in region *r* that previously did not belong to the technology space of a region in time T_{-1}, but enters in time t. This is illustrated by the following equation:

$$Y_{i,r,t} = [\frac{Entry_{i.r.t}}{\sum_{i=1}^{655} \quad - \ I_{r,t}}]$$

In which $\sum_{i=1}^{655}$ is the total number of technologies and $I_{r,t}$ the current number of technologies in region r in time t. Consequently, the technological entry model is written as follows:

$$Y_{r,t} = \beta_1 RV_{r,t-1} + \beta_2 SRV_{r,t-1} + \beta_3 UV_{r,t-1} + \beta_3 GDP_{r,t-1} + \beta_4 GeoRel.Density_{r,t-1}\ \beta_6 Techcomplexity_{i,t-1} + $$

$$+\beta_7 PopDensity_{r,t-1} + \beta_8 Technologicalclaims_{r,t-1}$$

Finally, a division is made between two models. This is done in a similar way as the first and second model because the total number of technological claims decreased largely between the second and third period (33%). Hence, the fifth and sixth model are split between period 2 and 3. The fixed effect for time accounts for all independent variables of the previous timespan (t_{-1}). T_{-1} does not refer to the previous year, but to the previous time period, so the entry-quotient in time t changes because of the explanatory variables in time t_{-1}.

4. Results

In this part of the research the descriptive statistics all variables will be discussed. This part entails an in-depth view of the geography of regional variety, complexity and average geographical relatedness density over time. The second part outlines the results of the econometric models.

4.1 Descriptive statistics

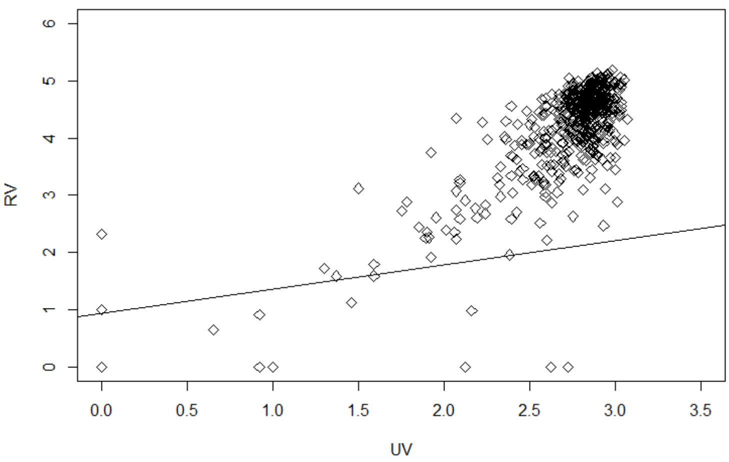
This part starts with the summary statistics of the variables. The data overlaps with 3 timeframes, hence, the total number of N = 870 cases is three times 290 regions. In the analysis the word technology is always regarded as a technology specialization (RCA technology) of a region as this is the starting point of measurement in the whole analysis, for more detail see the previous chapter.

Statistic	N	Mean	St.Dev.	Min	Pctl(25)	Pctl(75)	Max
Avg.Rel.Dens	870	21.24	12.4	0	10.75	30.88	54.19
RCI	870	67.96	17.59	0	58.06	80.10	100
UV	870	2.695	0.51	0	2.679	2.893	3.068
SRV	870	2.148	0.65	0	1.94	2.572	3.078
RV	870	4.16	0.95	0	3.952	4.701	5.189
Reg.GDP	870	50059	4174	0	18619	60556	623650
Pop.Density	870	281.02	493.20	0	74.05	257	4253.9
Tech.size	870	3221	6029.92	0	136.5	3611.2	56862

Table 1: descriptive of each variable, including control variables. (Source: Author)

First of all, each type of variety has a different mean and max. As expected, the level of related variety is higher than the level of variety of SRV and UV. This is due to the fact that it would be more uncertain to what technology class a certain patent would belong as it accounts for the most fine-grained level of technology class. However, it stands out that the level of UV, on average, is higher than that of SRV, which is quite unexpected. This would implicate that regions tend to be characterized by relative higher levels of related variety and unrelated variety, while fewer regions possess a portfolio of SRV. However, the variation of the standard deviation of each type of variety has a hierarchical order which was also found in Castaldi et al (2015).

It should be noted that, although entropy enables one to measure within-group entropy, related variety correlates with unrelated variety (Castaldi et al., 2015; Frenken et al., 2007). This correlation is also represented in plot 1. The higher the level of UV, the higher the level of RV (see diagonal).



Plot 1, Unrelated Variety (UV) versus Related Variety (RV). (Source: Author)

The following two maps show the level of variety (figure 5) and unrelated variety (figure 6) based on entropy levels in Europe in the period 2002-2006. Semi-related variety is left out because the purpose of this section is to give a rather simple overview of the composition of variety in Europe. Figure 5 shows that generally most regions have a relative high level of related variety, besides Eastern-European countries. Mainly central and Northern-European countries possess the highest level of related variety. This indicates that those regions have a portfolio of relatively higher related technologies. Figure 6 shows the level of unrelated variety in the period 2002-2006. Similar to the level of related variety, most regions tend to have a relative high level of unrelated variety (2,5 - max). However, the highest levels of unrelated variety are more exclusive to fewer regions. What sparks out is that indeed some regions[10] possess a high level of related variety as well as a high level of unrelated variety, indicating that the type of variety is not mutually exclusive.

[10] *I.e. the Dutch region Gelderland has both a high level of unrelated variety as well as related variety.*

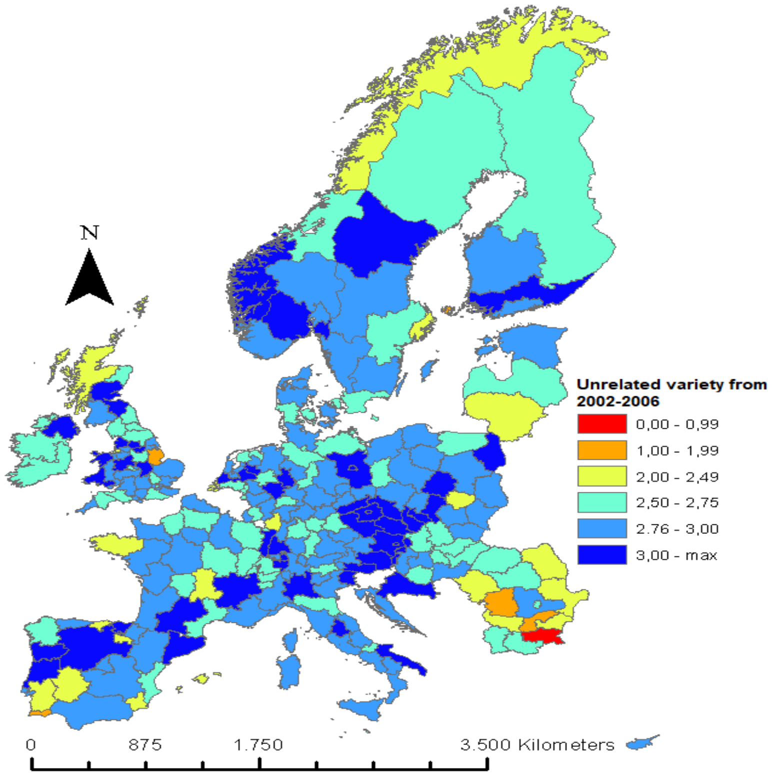


Figure 5, Related variety from 2002-2006. (Source: Author)

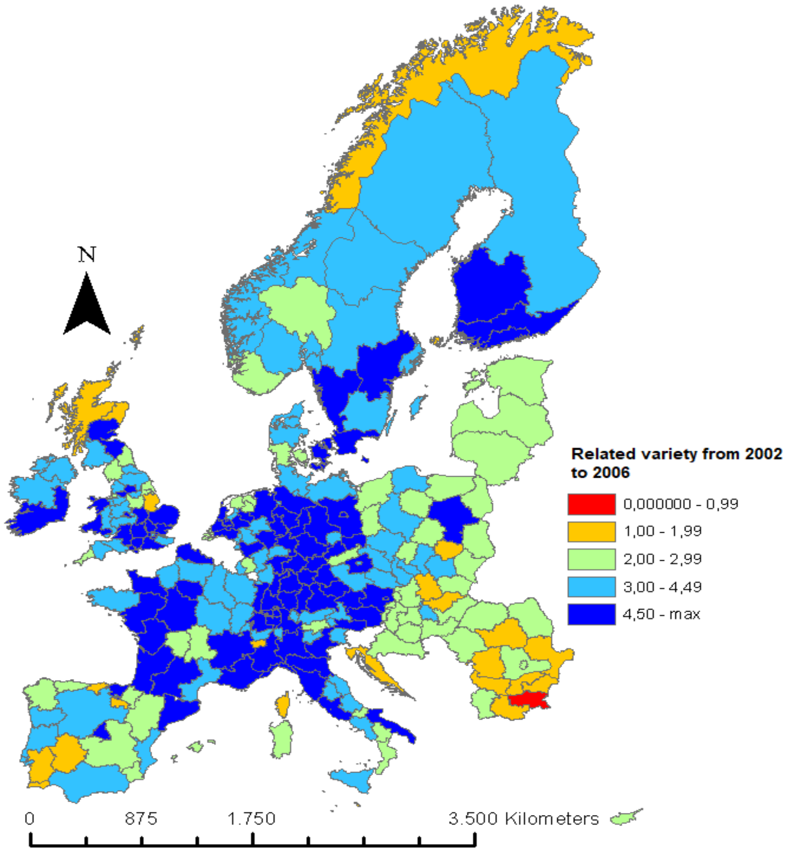


Figure 6, Unrelated variety from 2002 to 2006. (Source: Author)

Secondly, as shown in table 1, the average relatedness density is 21.24, which means that on average 21.24% of the technological specializations within a region are related. The most diverse region possesses an average relatedness diversity of 54.19. The standard deviation of 12.4 indicates that large differences exists between the regions. *Figure 7* illustrates branching opportunities in Europe based on the average relatedness density in 2002-2006. It can be derived from the map that some regions possess a much higher relatedness density. For instance, eastern and south-eastern European regions show a much lower relatedness density in comparison with the central-northern part of Europe.

Thirdly, table 1 shows that the regional complexity index is bound between zero and hundred. The higher the score, the higher the level of average regional complexity. The average level of complexity is 67.96 and the 25% highest ranked regions have a complexity score of 80.1 or more. The dispersion of regional complexity in the period 2002-2006 is shown in figure 7.

Similar to the level of relatedness density, the map of complexity shows that Central-Northern European countries possess the highest level of complexity while the Eastern -Southeastern European countries possess the lowest levels of complexity. A comparison with figure 8 shows that the level of complexity is more exclusive to few regions between 2002 and 2006. Mainly the central part of Europe, South Germany, Switzerland, Austria and Northern Italy possess a complexity level of 80 or higher. The high level of complexity also shows a correlation with relatedness density as the average relatedness density is over 40 in those areas. This correlation is also shown in *plot 2*. The higher the level of relatedness density, the higher the level of average regional complexity[11].

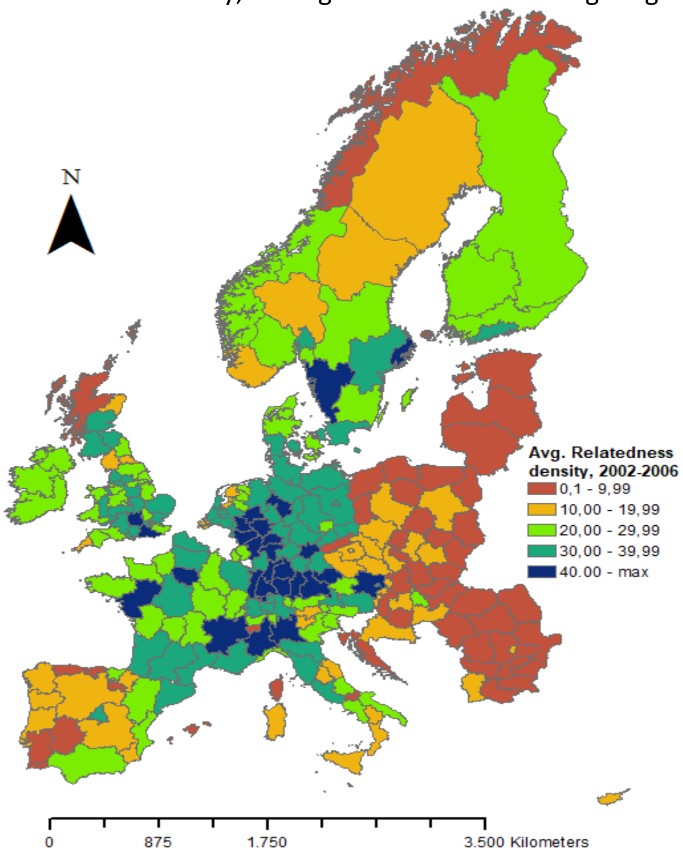
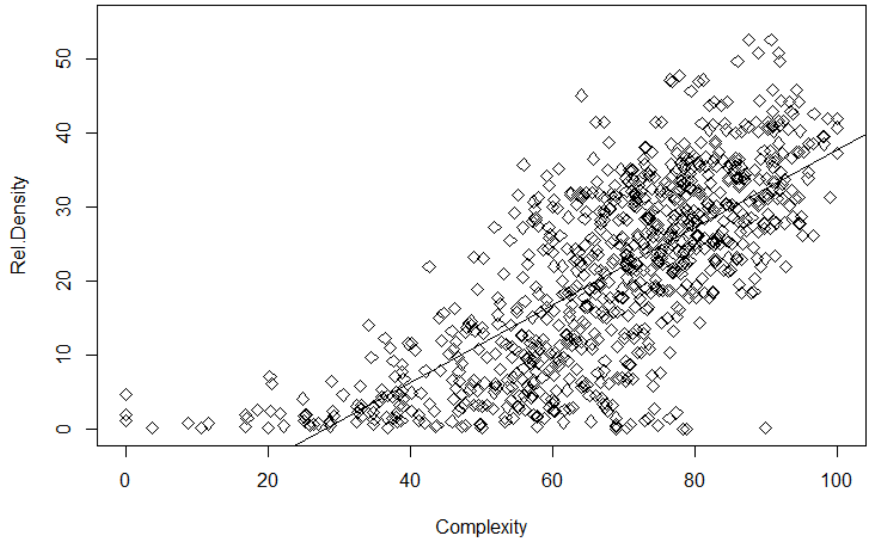


Figure 7[12], relatedness density in the period 2002-2006. (Source: author)

[11] *The results of the regression analysis of relatedness density on regional complexity will be discussed in the following section.*

[12] *Due to changing NUTS-2 codes it was not possible the level of complexity of Belgium and a few other regions with the map. Therefore, these are left out. The same applies for all other maps.*



Plot 2, Relatedness density versus regional complexity. (Source: Author)

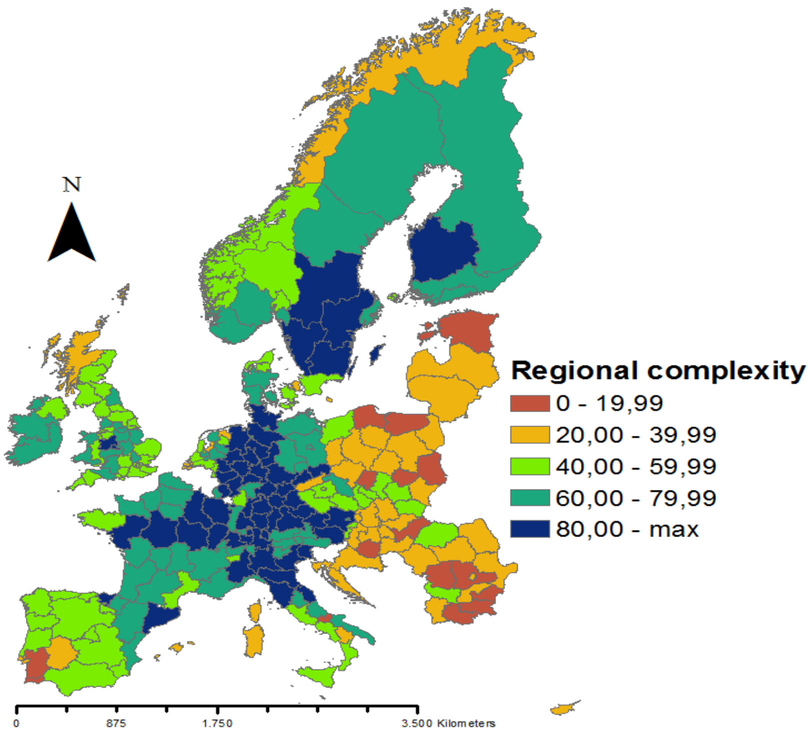


Figure 8, regional complexity in the period 2002-2006. (Source: Author)

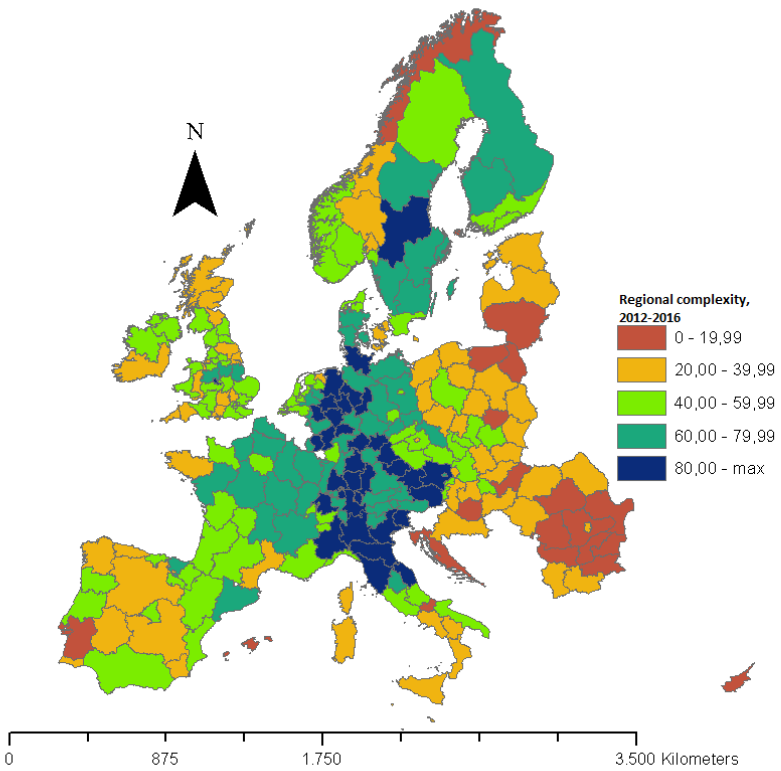


Figure 9, regional complexity in the period 2012-2016. (Source: Author)

Figure 9 shows the regional complexity in the final period, from 2012 to 2016. It can be derived from the map that the number of high complex regions has decreased significantly in comparison with figure 8. Furthermore, as indicated by the colors, the number of mid-complex regions[13] has increased in comparison with the first period. To show how the geography of complexity has changed over time a map is constructed with the percentual change of complexity between the first and final period.

Figure 10 represents the percentual change of complexity in Europe from 2002 to 2016. It is shown that the majority of the regions are subject to much change. Especially regions that possessed a relatively high level of complexity show a decline in their respective level of complexity. For instance, figure 8 shows that a large part of Germany, France and the Netherlands had a regional complexity level of 60 or higher. But, the majority of those regions in these countries showed a decline of 25% to 5% of their level of complexity (figure 10).

Furthermore, the figure shows that most regions do not increase their level of regional complexity by 5% or more. The majority of the regions show a decline. In fact, the mean level of complexity decreased from 69.61 to 62 between the first and third period. This implies that attracting complex technologies is not a straightforward process of simply recombining the existing portfolio of technologies.

[13] Mid complex regions are referred to as the regions with a light-green color.

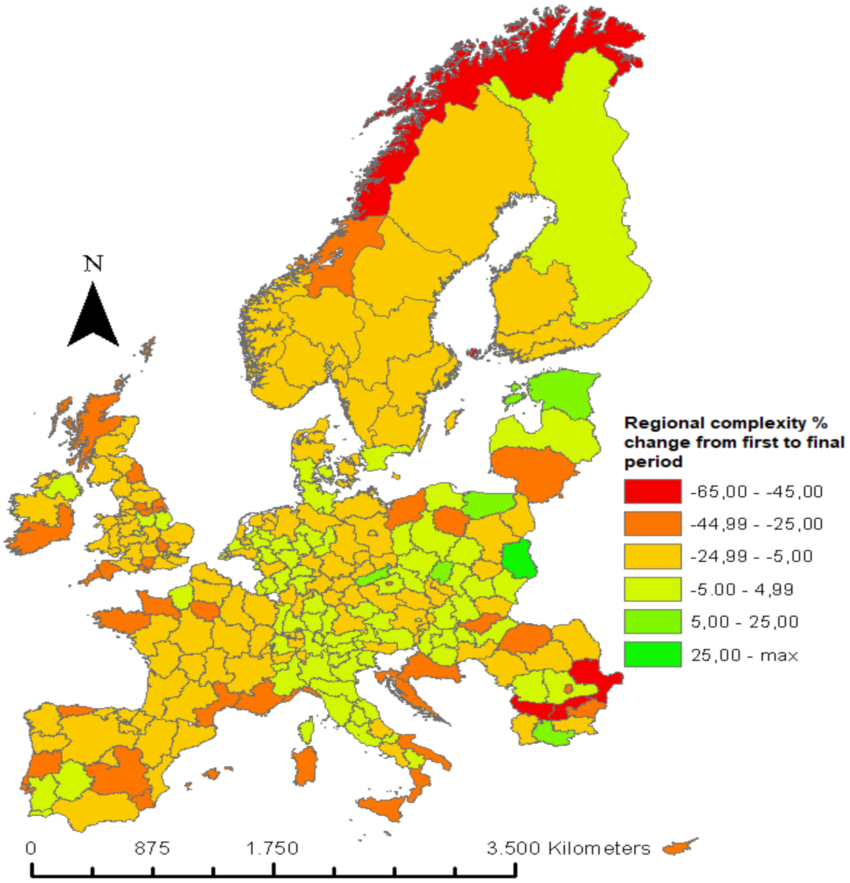


Figure 9, Regional complexity percentual change in the period 2002-2006 as to the period 2012-2016. (Source: Author)

Table 2 table gives a fine-grained representation of how complexity has changed over all three periods. Similar to figure 5, five levels of complexity are categorized*, very low-complexity* (0-20), *low complexity* (21-40*), mid complexity* (41-60), *high complexity* (61-80), *very high complexity* (81-100). The final column represents the percentual change between period 2 and 3, as the number of complex regions decreased largely in the final period. For each category of complexity (1-5), the number of regions has been counted. In the table, the categorical change is clearly visible. The number of lower-complex regions strongly increased between the second and third period, while the number of high-complex regions strongly decreased between the second and third period. It is shown that mainly the most complex regions have lost their position in the final period.

Complexity category	Total in Period 1	Total in Period 2	Total in Period 3
1	4	4	4
2	19	8	28
3	52	35	92
4	122	151	128
5	93	92	38

Table 2: Regional complexity in period 1 (2002-2006), period 2 (2007-2011) and period 3 (2012-2016) per level of complexity.

The regions in Europe show various trends. Firstly, it is shown that a large number of regions move from category 3 to category 4 between period 1 and period 2. As mentioned previously, in that period the total number of technological claims increased. However, it stands out that the number of very high complex regions declined. Although, the regions showed a general increase in their level of complexity, it is shown that regions were not able to reach the highest-complex levels. Secondly, *table 4* shows that especially mid- to high-complex regions show a large decline between the second and third period, with a respective decline of 15% and 59%. Thirdly, the number of lower-complex regions increased between period 2 and 3 due to regions losing their position (i.e., moving from category 4 to category 3).

This implies that regions in Europe show more equality in their level of complexity, while being a high-complex region is much more exclusive. In period 2, most regions sat in category 4 and 5, while the number of regions in category 2, 3 and 4 in the final period is much more common. Finally, it is visible that the number of regions represented in category 5 has only decreased between the first and final period, mostly in the final period. Therefore, it is shown in *table 2* that the process of increasing or maintaining the regional complexity level is a hard process and that the level of regional complexity is subject to change. A plausible explanation could be that the most complex technology patents are less likely to be cited (Balland & Rigby, 2017). Nevertheless, it cannot fully justify such a large decrease in regional complexity.

The following table reflects on the top 10 regions that lost the highest level of complexity. Each of these regions moved at least one category down. There is no clear geographical pattern in regions that lost a high number of complexity, as the regions vary from Eastern-European regions, to the UK – and Norway. All regions with a high number of complexity decrease had a complexity level between the 60-80 range (7) or the 4-60 range (3). This might indicate that those regions are more subject to change in terms of relative complexity loss.

Regions	Complexity lost	Initial complexity category
North central (BG)	-40,94	4
Sud-Vest Oltenia(RO)	-37,13	4
Jadranska Hrvatska(HR)	-29,19	4
East Yorkshire and Northern Lincolnshire (UK)	-28,55	4
South central(BG)	-28,26	4
Észak-Magyarország(HU)	-27,97	4
Nord-Norge(NO)	-27,4	3
Centru(RO)	-27,29	4
Extra-regio(DK)	-25,31	3
Extra-regio(NO)	-25,3	3

Table 3: top 10 highest losing complexity regions

In this part of the analysis, the change in entry of new technological specializations will be compared with regional complexity. Firstly, the technological size will be discussed. The 25% largest technological portfolios of a region have a size of 3269 and above (see table 1). This indicates that, with a standard deviation of 6029 and a range of 56862, the technological size per region differs largely. This is similar to the findings of Castaldi et al. (2015) – and is an indication that the output of patents strongly differs among European regions. Table 4 reflects on the regional composition in terms of patenting activity overtime. The number of patents claims firstly showed an increase in the mean, median and the max, while it dropped significantly (on average 33%) in the period afterwards. As shown previously, complexity has dropped significantly. Therefore, it is expected that the number of technological claims has an impact on regional complexity[14].

[14] *This will be further explained in the analysis in section 4.2*

Timeframe	Mean	Median	Max
2002-2006 (p1)	3052	832	51361
2007-2011 (p2)	3559	1104	56862
2012-2016 (p3)	2362	741	37728

Table 4: technological size over time, Source: Author

As patent applications and complexity have decreased, the relationship between complexity and the entry of new technological specializations is discussed in this part. The complexity changes between period 1 and period 2 - and the change in technological entry/exits between period 1 and period 2 is represented in *plot 3*. The blue dots represent the most complex regions, the green dots represent high complex regions, the orange dots represent the mid complex regions and the black dots all lower complex regions.

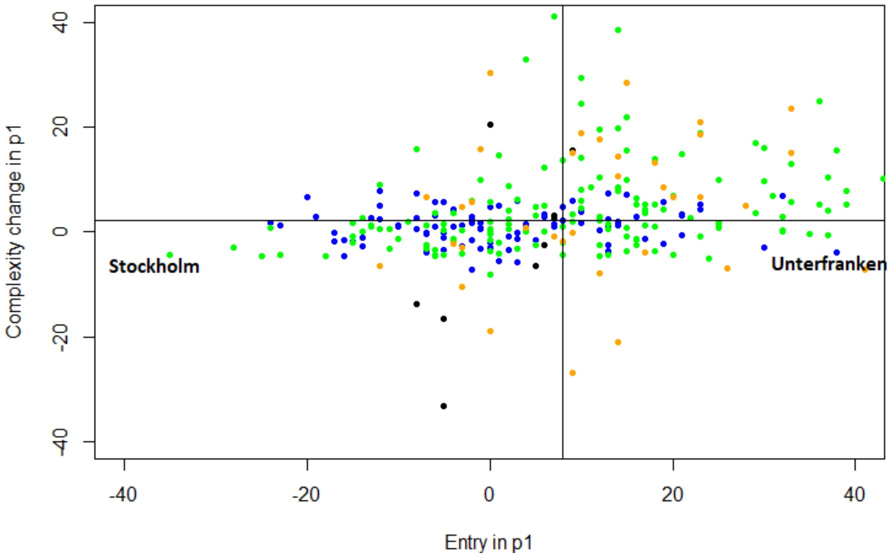
First of all, *plot 3* shows different trends. For instance, Unterfranken is a region with a positive balance of entrances in time *t*, but with a minor decline in its complexity level in time *t*. This may indicate that Unterfranken was a very diverse area and maintained it position. Yet, a large number of (non-ubiquitous) technologies entered the region[15]. Stockholm, on the other hand, is a region that showed a large decline in technological entrances but showed a similar decline in the level of complexity.

However, when the color of the dots is investigated there are some visible relationships. In general, the complexity change of green and blue dots lies beneath the average change in regional complexity, if those regions have technological entrances below average. Although it would be very hard for the most complex regions to improve their level of complexity (as complexity is upper-bounded by 100), they show very little increase and decrease in the level of complexity. The green dots, with a lower level of complexity, show a more scattered variation along the scatterplot. Moreover, the relationship between regional complexity and technological entrances in category 4 and 3 seems stronger in comparison with the blue dot.

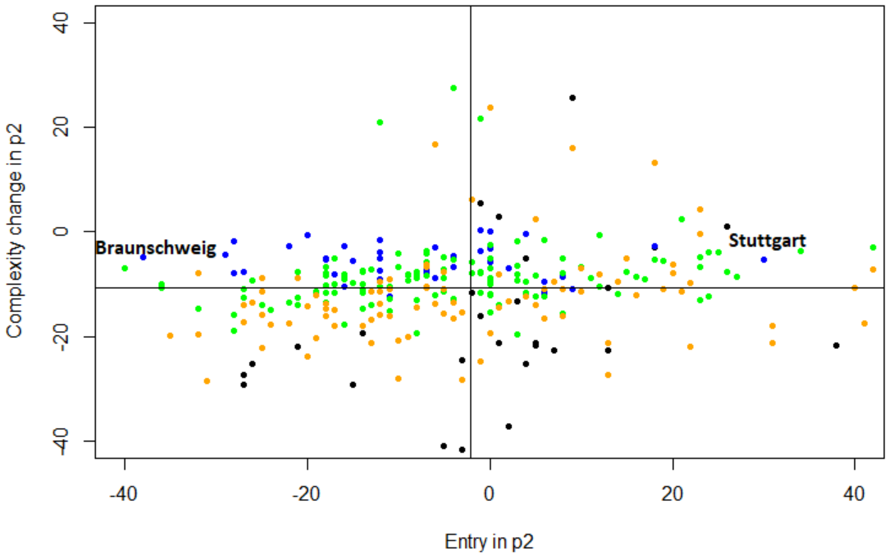
Plot 4 compares the change of entry in period 3 with the complexity change in period 3. This is done to see whether similar relationships exist because the total technological portfolio has decreased with 33% in the final period. Indeed, there is a shift visible between the average entry rates of technologies and the average decrease in regional complexity. Similar with the previous period, there are regions (Stuttgart) that show high entrances, yet with a minor decrease in complexity, and there are regions (Braunschweig) that show a large number of exits with a minor decrease in complexity.

However, in general, the number of high complex regions has visibly decreased. Moreover, a somewhat linear relationship is now visible: as the entry rates of technologies went down, the average complexity level has visibly decreased in that same period. Especially regions that have a mid-level of complexity are below the average decrease of -10,5%. Furthermore, it is visible that the number of blue-colored dots has decreased, and the orange dots have increased. This may imply that already complex regions face difficulties to maintain their (high) level of complexity.

[15] *These regions are used as an illustration. It is also possible that no new technological specializations have entered the region, but that other regions filed relatively fewer patent applications in that technology.*



Plot 3: scatterplot of entry-rates and regional complexity level in p2. (Source: Author).



Plot 4: scatterplot of entry-rates and regional complexity level in p3. (Source: Author).

4.2 Analysis I: regional complexity

Table 5 represents the analyses of how the explanatory variables influence the level of regional complexity. This section provides answer on hypothesis 2 and 3. Before presenting the analysis, various assumptions have been tested to check whether the results can be regarded as valid. The results of those assumptions can be found in the Appendices. The table starts with each variable regressed on the level of regional complexity in time t_{+1}. A distinction has been made between the second and third period. Hence, two models are represented in the table. One model accounts for the regional complexity in period 2 (2007-2011, model 1), and one accounts for the regional complexity level in period 3 (2012-2016, model 2). Furthermore, two more models have been added to appendix A[16].

	Dependent variable: regional complexity	
	Regional complexity T_{+1}. (model 1)	*Regional complexity $T2_{+1}$. (model 2)*
Relatedness density	0.36107682 *** (0.06904)	0.33520904*** (0.1184)
SRV	0.57204036*** (1.712)	0.45260445*** (3.610)
UV	-0.010124800 (2.192)	-0.06376551 (2.323)
RV	*Excluded*[17]	0.15178505* (1.364)
Population density	-0.04379048 (0.001003)	-0.05686903 (0.001201)
Regional GDP	-0.17202506*** (0.0001319)	-0.20600421*** (0.0001527)
Technological size	0.08795841 (0.0001122)	0.17583480** 0.0001300)
Constant	45.13*** (4.765)	18.20* (7.406)
Time lag	*yes*	*yes*
Observations	290	290
R^2	0.6923	0.6616
Adjusted R^2	0.6846	0.6516
Residual std. Error	8.143	9.374
F statistic (df = 240, 238)	90***	66,46***

*Table 5, multiple regression analysis of the explanatory variables on the dependent variables' Regional complexity in t+1 (1) and regional complexity in t2+1 (2). Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

[16] *One model with an entry-quotient is placed in the appendix (table 1 in appendix).*

[17] *A model with relatedness density (table 2 in appendix) is placed in the appendix because RV showed a large VIF and correlation with SRV and RV (figure 3 in appendix).*

Firstly, both model 1 and model 2 have a strong explanatory power, respectively 69,23% (R^2 = 0,6923) and 66,16% (R^2 = 0,6616) of the variance is explained by the explanatory variables. Mainly, relatedness density and semi-related variety (SRV) have a strong and positive effect on regional complexity. An increase of one-unit relatedness density increases the regional complexity by 36,1% (*ß* = 0,361) and 33,5% (*ß* = 0,335), respectively. Furthermore, an increase of one-unit SRV increases the regional complexity respectively by 57,2% (*ß* = 0,572) and 45,26% (*ß* =0,4526). What sparks out is that the effect of related variety (RV) is positive, but much smaller than SRV on regional complexity. This effect[18], in comparison with SRV, is three times smaller (*ß* = 0,152). Furthermore, unrelated variety (UV) does not have a significant effect on regional complexity.

Linking these results with the theory it can be argued that the results are somewhat in line with the discussed theorem. Firstly, Balland et al. (2019) found that complex technologies are more likely to be introduced when those are related to the existing portfolio. Although different dependent variables have been used, the underlining idea is similar. Regions that have a large number of branching opportunities have a higher regional complexity. Thus, it is likely that relatedness density has a positive effect on the introduction of complex activities. This confirms hypothesis 2.

Furthermore, Castaldi et al. (2015) found that RV and UV enhance invention output. Although the effect of RV is positive, it is three times smaller than SRV. However, the effect of unrelated variety (UV) was insignificant in both models. While regional complexity has never been used as a dependent variable in the economic geography literature, it was expected that RV, SRV as well as UV have a positive impact on the regional complexity.

In both models the population density has no significant effect on regional complexity while the effect of regional GDP is significant and negative. Thus, an increase in GDP, decreases the level of regional complexity. This may be caused by a relatively low number of cases (N=290) which makes the effect of residual outliers stronger. Besides, it is plausible that regional GDP does not reflect innovative output in terms of complex activities, but rather reflects robust outputs of regional GDP. Finally, the effect of technological size is positive and significant in the second model. This positive effect can be explained by the formula that calculates regional complexity, in which diversity (in terms of patenting activity) contributes to higher regional complexity.

Altogether, the results do not confirm *hypothesis 3*, which expected that technological variety has a positive impact on regional complexity. Only RV and SRV showed a significant and positive effect on regional complexity. On the other hand, it could be argued that SRV is an optimal technological composition of related -and unrelated technologies to foster complex outputs at the regional level. The distance between technologies in such a region is sufficient enough to foster recombinant opportunities, yet it is distant enough to foster novel and non-ubiquitous technologies. For instance, in a practical example regional complexity can be fostered by specializations in class A41 “Personal or domestic articles”, A42 “headwear” and A43 “footwear” as they relate in terms of the one-digit class “A” but differ in the three-digit category. In practice, the distance of semi-related technologies may facilitate relatedness in terms of complementarity effects (Farinha et al., 2019) as the cognitive distance is large enough – but not too little to foster innovative output, thus, upgrading regional complexity.

The second hypothesis, *relatedness density has a positive effect on the average regional complexity* has been confirmed. This is shown by the significant and strong positive coefficients in both models of relatedness density. The outcomes of relatedness density on average regional complexity are in line with previous findings. Balland et al. (2019) found that diversifying into complex technologies is easier when technologies are related to the existing knowledge core of a region.

---

[18] In the appendix another model has been included in which the effect of RV was very similar and significant.

As the descriptive analysis showed, complexity has decreased overtime, especially for regions with a relative higher level of complexity. Hence, the final part of the analysis on regional complexity is to investigate whether differences between high -and low complex regions[19] exist.

	Dependent variable: Low regional complexity / high regional complexity	
	Low complexity (model 3)	*High complexity (model 4)*
Relatedness density	0.17102852 * (0.0884)	0.37803708*** (0.06039)
SRV	0.40549819*** (1.503)	0.43537220 *** (1.312)
UV	0.07136746 (2.049)	-0.04550018 (2.615)
RV	0.07672354 (1.117)	-0.25997844** (1.141)
Population density	-0.01648865 (0.001278)	-0.05576878 (0.0005295)
Regional GDP	-0.08796393 (0.00001636)	-0.23516897*** (0.00000706)
Technological size	0.05450715 (0.0002031)	0.22889502*** (0.00005828)
Constant	24.33*** (5.633)	56.28*** (7.469)
Observations	290	290
R^2	0.3894	0.3996
Adjusted R^2	0.3703	0.378
Residual std. Error (df =223, 253)	9.267	6.038
F statistic (df = 223, 253)	20.32***	24.05***

*Table 6, effects of the explanatory variables on high – and low regional complexity. Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

Table 6 reflects on the effect of the explanatory variables on high -and low regional complexity. Firstly, in the third model 37,03% (R^2 = 0.3703) of the variance can be explained by the explanatory variables and in the fourth model 37,8% (R^2 = 0.378).

Secondly, the model shows differences in the significance of various variables. On the one hand, relatedness density has a stronger effect on high complex regions then on low complex regions, respectively 0.17 ($ß$ = 0,17) and 37,8 ($ß$ = 0,378). Similarly, SRV has a positive effect on both low – and high complex regions. The positive effect of SRV on regional complexity increases when regions become more complex, from ($ß$ = 0,405) to (ß = 0,435).

[19] *No distinction has been made between period 2 and 3 because more cases can be included. A model with all time-lagged periods individually has been added to the appendix (table 3). It is left out because it has less explanatory power and more room for outliers because it has a relative low number of cases.*

However, the effect of RV in model 4 was unforeseen. Model 3 shows that RV has no significant effect on lower complex regions, whereas model 4 shows a negative and significant effect for high complex regions. Thus, when technological related variety increases with one unit, the regional complexity decreases with almost 26%. In previous literature it was found that relatedness is of less importance for regions with a stronger and innovative capabilities (Xiao et al., 2018) or at a higher state of development (Pinheiro et al., 2018). This might imply that RV is more important at a lower or intermediate level of regional complexity.

On the other hand, the effect of UV is not significant. This suggest that mainly a SRV technological composition fosters complexity at the regional level. Furthermore, the effect of every control variable on low complex regions is insignificant. For high complex regions the effect of regional GDP is negative and significant (-0,235) which was similar in the full model. Technological size has a positive effect (*ß* = 0,229) on high complex regions, which was expected because high complexity is partly formed by the number of technological claims in a region.

4.3, Analysis II: Technological entrances

Balland et al. (2019) found that complex knowledge is more attractive for regions, but at the same time harder to produce. Hence, the relationship between technological entrances and complexity is not linear. Nevertheless, the influence of regional complexity on entry and exit-rates of technological specialization has not been investigated yet. This part of the analysis conducts the analysis of regional technological entrances[20]. It should be mentioned that entry or exit-rate is computed as the number of entries in t_{+1} divided by the total potential entries for a region in t_{+1}(see chapter 3). Furthermore, a distinction has been made between each period because period 2 is characterized by an average growth in technology entrances and period 3 is characterized by a strong decline in technology entrances.

	Dependent variable: Entry-quotient of potential entries	
	Entry-quotient T_{+1}. (model 5)	*Entry-quotient $T2_{+1}$. (model 6)*
Regional complexity	-0.59388009*** (0.0007684)	-0.007619744 (0.000251)
Relatedness density	*Excluded*[21]	*Excluded*
SRV	0.31563503 *** (0.006861)	-0.593566934*** (0.007049)
UV	-0.06222147 (0.009493)	-0.160657688* (0.01032)
RV	0.13163209 (0.005328)	0.589558408*** (0.003522)
Population density	-0.01730323 (0.000003786)	-0.04768003 (0.000003403)
Regional GDP	-0.09786021 (0.00000005079)	0.033022428 (0.00000004475)

[20] *Technological entrance is always regarded as a specialization of a technology in this section.*

[21] *In the appendix a model with the inclusion of relatedness density has been added (Appendix, table 4). Due to high correlation values and a high VIF, relatedness density has been excluded (see figure 14 in appendix). Furthermore, the explanatory power of the model was smaller with relatedness density included.*

Technological size	0.05327373 (0.0000004036)	0.048853500 (0.00000636)
Constant	0.04823 (0.01955)	0.02152 (0.02068)
Observations	290	290
R^2	0.2008	0.2864
Adjusted R^2	0.1773	0.2654
Residual std. Error (df = 238, 238)	0.02947	0.02678
F statistic (df = 238, 238)	8.543***	13.64***

*Table 7, effects of the explanatory variables on high – and low regional complexity. Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

Firstly, both models show a smaller explanatory power as to the first and second model, respectively (R^2 = 0,1773) and (R^2 = 0,2654). Furthermore, the fifth model shows that regional complexity has a negative effect on the entry-quotient in period 2. A one-unit increase of regional complexity decreases the entry-quotient by 59,38% (β = 0,5938). A plausible explanation could be that complex patents are less likely to be cited (Balland & Rigby, 2017), which decreases the number of granted complex patents.

Another explanation could be that there might be a limit to the diversity of complex regions. At a certain point to many technological specializations are embedded within a region, which does not arouse competitive advantage. This might imply that, tough it is easier for a complex region to attract new technologies, a region either does not benefit from new technological specializations – or they cannot keep up with a rapid changing technological landscape (Balland et al., 2019). Furthermore, with regards to patenting activity, Balland and Rigby (2017) found that regions with the highest number of technological claims are not the most complex regions. As mainly high complex regions lost complexity (see chapter 4.1), those regions could be more focused on specializing, rather than diversifying into less relevant technology fields. Nevertheless, this, should be further investigated in future research.

On the other hand, when the average decline in the entrance of technologies occurred in period 3, the level of regional complexity (in period 2) had no significant effect on the entry of technologies in period 3. This finding leaves room for more questions and requires more in-depth understanding of regional complexity. Therefore, a simple linear regression model has been made to understand the effect of regional complexity on entry-rates (in t_{+1}), without controlling for any variables. The effect for both periods was negative and significant with an average beta coefficient of -0.274 (β = -0,274) and an explanatory power of R^2 = 0.07548. Thus, as regional complexity increased, the entry-rates of technologies decreased the period afterwards. This might imply a relation between a loss of complexity and technologies (*see table 2 in paragraph 4.1*). Thus, in general complex regions face difficulties to maintain their complex position as they are characterized by a loss in complexity and technologies.

Secondly, the effect of SRV is contrasting. The fifth model shows that SRV has a positive impact on technological entrances in period 2, a one-unit increase of SRV increases technological entrances with 31,56% (0,3156). However, in the period afterwards, a one-unit increase of SRV decreases technology entrances in period 3 with 59,35% (β = 0,5935), which is a strong and negative relationship. Again, a single linear regression model has been applied to better understand the effect of SRV on the entry-quotient. It was found that SRV has no significant effect in the first period, while the effect in the second period was still negative but smaller, -0.258 (R^2 = 0.06707). This implies that regions with a relatively high semi-related technological portfolio have a negative influence on the entry-rates of technologies.

Thirdly, relatedness density is not included in this model but in table 4 in the appendix. The effect of relatedness density on the entry-quotient of technologies is negative. This is in contrast with earlier findings that relatedness density with previously non-existing technologies increases the likelihood that regions specialize in those technologies (Hidalgo et al., 2018).

An explanation could be that previous models suggested that relatedness density fosters regional complexity. Thus, if complex regions have a high relatedness density – and high complex regions are not likely to introduce new technologies, those regions are not very likely to introduce new technologies. This finding suggests a diversification paradox: Very diversified regions are more likely to introduce new technologies as they are more related to their existing portfolio of technologies. Yet, there might be a limit to diversification as complex regions with a higher level of relatedness density tend to have a negative entry-quotient. Similarly, the most complex regions are not necessarily regions with the highest number of technological claims (Balland & Rigby, 2017).

Furthermore, the effect of RV has a strong and positive effect on the entry-quotient according to model 6. An increase of a one-unit RV, increases technological entrances with 58,95% (ß = 0,5895). Although that is in contrast with the results of relatedness density, it confirms the hypothesis of Boschma & Gianelle (2013), in this study a related technological composition, has a positive effect on entries of technological specializations. These findings confirm hypothesis 1, related *technological variety has a positive impact on the entry of technologies in a region*.

Finally, the effect of UV is significant in the final model. UV decreases the entry-quotient of technologies (*ß* = -0,16), which indicates that regions that are characterized with high UV are less likely to introduce new technologies. Furthermore, all control variables have no significant effect in both models.

5. Conclusion & discussion

Previous research has shown that the concept of relatedness is associated with entry -and exits of economic activities, regional development processes and future development trajectories (Neffke & Henning, 2009; Essletzbichler, 2015; Boschma et al., 2014; Petralia et al., 2017). Furthermore, regional complexity is shown to be a source of comparative advantage and economic growth (Balland & Rigby, 2017). Related variety is regarded as a facilitator of the bulk of innovations and economic growth, while unrelated variety facilitates breakthrough innovations (Castadli et al., 2015). Although regional variety and complexity are two main concepts of the economic geography literature, the effect of regional variety on regional complexity has never been investigated before.

Nevertheless, it is of great importance to investigate the effect of regional technological variety on regional complexity. For policy matters this contributes to the understanding of how a certain set of technologies (i.e. semi-related as well as related) positively influence regional complexity. Furthermore, the insights have the potential to extent the smart specialization framework of Balland et al. (2019). This may lead to competitive regions on the long run (Hidalgo & Hausmann; Balland & Rigby, 2017) with a better income equality (Hartmann, 2017). Therefore, this research proposed a new method that analyses the effect of regional variety on regional complexity. Furthermore, it has been investigated how regional complexity influences the entry and exits of technological specializations at the regional level.

A main contribution of this research is that average regional complexity has a strong and negative influence on the entry-quotient of technologies. Mainly 'high' complex regions are the biggest "losers" in terms of complexity decrease. Those regions generally lost momentum in either technological entries, or maintaining their technological specializations. This might suggest a diversification paradox: complex, and thus diversified regions, are more likely to introduce new technologies. Yet, there might be a limit to diversification, as those regions with a higher level of relatedness density and complexity have a negative influence on the entry-quotient.

Similarly, the descriptive analysis has shown that regional complexity is difficult to maintain and has decreased largely between 2002 and 2016. As complexity has dropped in general, the results have shown that regions face difficulties with attracting or maintaining complex technologies, especially high complex regions. That, in combination with the likelihood that complex patents are less likely to be cited (Balland & Rigby, 2017), may cause a loss of regional complexity. Though, more research should investigate what has caused a decline in regional complexity.

Mainly related variety is an important driver of an increase in technological entrances, confirming *hypothesis 1*. This finding is in line with previous research in which was found that a higher level of related activities is positively associated with the introduction of previous non-existing activities (Hidalgo et al., 2018).

Another appealing results of this research is that semi-regional variety and relatedness density have a significant and strong effect on regional complexity. The effect of related variety was positive, but smaller. Especially higher complex regions benefit from a high semi-related technological structure and relatedness density. The latter variable is in line with the results of Balland et al. (2019), which suggested that relatedness density has a positive effect on the entry of previously non-existing complex technologies, thereby confirming *hypothesis 2*. An explanation for the positive and strong effect of relatedness density on regional complexity is that regions with a high level of relatedness density are diverse regions, which offer more unique technological recombination opportunities (Hidalgo & Hausmann., 2009; Weitzman, 1998).

On the other hand, unrelated variety has no influence on regional complexity, which rejects *hypothesis 3*. Yet, the results contribute to the hypothesis of Frenken et al. (2007) and Castaldi et al. (2015) in which was argued that related variety is found to be the main driver of recombinant innovative outputs and unrelated variety as the driver of breakthrough innovations. It is thus not solely a regional composition of related technologies that are of importance to foster regional complexity, but mostly semi-related technologies.

Moreover, when regions become more complex, a related technological composition of technologies seem to negatively influence regional complexity. This implies that too much relatedness may lead to a loss in complexity. Although this assumption has not been investigated previously, Boschma (2017) argued that too much related activities may harm regional resilience in the long-term as it is more subject to external shocks. Moreover, the importance of related capabilities depends on the stage of development and the innovative capabilities (Alshami et al., 2018; Xiao et al., 2018. Similarly, a too much related composition of technologies may harm the possibility to achieve breakthrough innovations that foster unique technological specializations, and thus increases the regional complexity.

Giving an answer to the main question: "*To what extent does regional technological variety influence the level of regional complexity*" it can be concluded that mainly a semi-related technological composition and relatedness density are beneficial for the average level of regional complexity. Moreover, the effect of relatedness density and semi-related variety is stronger when a region becomes more complex. Giving an answer to the sub-question: "*To what extent does regional complexity influence entry-rates of technological specializations?* It can be concluded that regional complexity negatively influences the entry of technological specializations.

Altogether, the findings have various implications for future research and policy matters. Policy directed at attracting semi-related sectors can be beneficial for regions to either maintain or increase the level of regional complexity. Too much related activities may harm complexity at a higher state of development. However, it should be mentioned that the levels of variety are not mutually exclusive. It rather indicates that a semi-related technological composition could be more relevant to attract complex activities. Furthermore, the findings of Castaldi et al. (2015) suggest that the importance of related and unrelated activities should not be neglected. In practice this may imply that a region should act as a facilitator that connects semi-related sectors, institutions, industries and knowledge hubs, without neglecting the importance of related and unrelated activities. Foremost, regional policy should not look for over-specialization, but rather look for diversifying into the right activities (in terms of relatedness) at the right time (Alshami et al., 2018; Hidalgo et al., 2018).

The results also bring some limitations of this research. Firstly, Farinha et al. (2019) showed how relatedness can be unfolded into three dimensions. This has brought more insight under what granularity of analysis relatedness truly matters. For instance, relatedness density is of great importance for regional complexity, but it should be investigated whether relatedness is based on complementarities or similarities (Boschma, 2017). Future research on regional complexity should thus unravel relatedness into the proposed dimensions (Farinha et al., 2019) to bring more insight into how relatedness truly matters for regional complexity.

Another limitation is that the analysis is conducted at the regional level and only used three timeframes. This allowed more room for unwanted noise in the data-analysis as the influence of yearly outliers is likely to be much higher. Moreover, the effect of the explanatory variables might have been relatively high because of the smaller data sample.

Furthermore, the analysis has taken into account each form of regional variety. Although, entropy allows one to decompose the data at different levels of aggregation, a relative high correlation (see appendices for more details) between the types of variety exists (Castaldi et al. 2015).

Finally, there is still a lack of scientific evidence what caused the large drop of patent applications and complexity levels, which may impede future patent research. Patent research is a plausible way to empirically measure the geography of knowledge, industries and skills in the field of economic geography. However, future researchers should be aware that, if a trend of decreasing patent claims were to continue in a larger scale, it might harm the robustness of the research.

References:

Alshamsi, A., Pinheiro, F. L., & Hidalgo, C. A. (2017). When to target hubs? Strategic diffusion in complex networks. arXiv preprint arXiv:1705.00232.

Alshamsi, A., Pinheiro, F. L., & Hidalgo, C. A. (2018). Optimal diversification strategies in the networks of related products and of related research areas. Nature communications, 9(1), 1328.

Asheim, B. T., & Gertler, M. S. (2005). The geography of innovation: regional innovation systems. In The Oxford handbook of innovation.

Asheim, B. T., Boschma, R., & Cooke, P. (2011). Constructing regional advantage: Platform policies based on related variety and differentiated knowledge bases. Regional studies, 45(7), 893-904.

Balland, P.A. (2017) Economic Geography in R: Introduction to the EconGeo Package, Papers in Evolutionary Economic Geography, 17 (09): 1-75

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2018). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. Regional Studies, 1-17.

Balland, P. A., & Rigby, D. (2017). The geography of complex knowledge. Economic Geography, 93(1), 1-23.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. Regional Studies, 51(3), 351-364.

Boschma, R., Balland, P. A., & Kogler, D. F. (2014). Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. Industrial and corporate change, 24(1), 223-250.

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. Research Policy, 44(10), 1902-1914.

Boschma, R., Coenen, L., Frenken, K., & Truffer, B. (2017). Towards a theory of regional diversification: combining insights from Evolutionary Economic Geography and Transition Studies. Regional Studies, 51(1), 31-45.

Boschma, R., Frenken, K., Bathelt, H., Feldman, M., & Kogler, D. (2012). Technological relatedness and regional branching. Beyond territory: Dynamic geographies of knowledge creation, diffusion and innovation, 64-81.

Boschma, R., & Gianelle, C. (2013). Regional branching and smart specialization policy. JRC technical reports, (06/2104).

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. Papers in Regional Science, 91(2), 241-256.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. Research policy, 32(1), 69-87.

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. Regional studies, 49(5), 767-781.

Ejermo, O. (2005). Technological diversity and Jacobs' externality hypothesis revisited. Growth and Change, 36(2), 167-195.

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas. Regional Studies, 49(5), 752-766.

Farinha Fernandes, T., Balland, P. A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. Industry and Innovation, 1-35.

Fitjar, R. D., Huber, F., & Rodríguez-Pose, A. (2016). Not too close, not too far: testing the Goldilocks principle of 'optimal'distance in innovation networks. Industry and Innovation, 23(6), 465-487

Foray, D., David, P. A., & Hall, B. (2009). Smart specialisation–the concept. Knowledge economists policy brief, 9(85), 100.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. Research policy, 30(7), 1019-1039.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. sage.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. Regional studies, 41(5), 685-697.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., & Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. World Development, 93, 75-93.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

Hausmann, R., & Klinger, B. (2007). The structure of the product space and the evolution of comparative advantage (No. 146). Center for International Development at Harvard University.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

Hidalgo, C. A., Balland, P. A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., ... & Neffke, F. (2018, July). The principle of relatedness. In International Conference on Complex Systems (pp. 451-457). Springer, Cham.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the national academy of sciences, 106(26), 10570-10575.

Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. Science, 317(5837), 482-487.

Kogler, D. F. (2015). Evolutionary economic geography–Theoretical and empirical progress.

Lorentzen, A. (2008). Knowledge networks in local and global space. Entrepreneurship and Regional Development, 20(6), 533-545.

Martin, R. (2010). Roepke Lecture in Economic Geography—Rethinking regional path dependence: beyond lock-in to evolution. Economic geography, 86(1), 1-27.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. Strategic Management Journal, 34(3), 297-316.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Economic geography, 87(3), 237-265.

Pérez-Luño, A., Medina, C. C., Lavado, A. C., & Rodríguez, G. C. (2011). How social capital and knowledge affect innovation. Journal of Business Research, 64(12), 1369-1376.

Petralia, S., Balland, P. A., & Rigby, D. L. (2016). Unveiling the geography of historical patents in the United States from 1836 to 1975. Scientific data, 3, 160074.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2018). Shooting High or Low: Do Countries Benefit from Entering Unrelated Activities?. arXiv preprint arXiv:1801.05352.

Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., & Gabrielli, A. (2017). Unfolding the innovation system for the development of countries: co-evolution of Science, Technology and Production. arXiv preprint arXiv:1707.05146.

Saviotti, P. P., & Frenken, K. (2008). Export variety and the economic performance of countries. Journal of Evolutionary Economics, 18(2), 201-218.

Sandu, S. (2012). Smart specialization concept and the status of its implementation in Romania. Procedia Economics and Finance, 3, 236-242.

Schumpeter, J. (1942). Creative destruction. Capitalism, socialism and democracy, 825, 82-85.

Petralia, S., Balland, P. A., & Morrison, A. (2017). Climbing the ladder of technological development. Research Policy, 46(5), 956-969.

Tanner, A. N. (2014). Regional branching reconsidered: Emergence of the fuel cell industry in European regions. Economic Geography, 90(4), 403-427.

Tavassoli, S., & Carbonara, N. (2014). The role of knowledge variety and intensity for regional innovation. Small Business Economics, 43(2), 493-509.

Theil, H. (1972). Statistical decomposition analysis; with applications in the social and administrative sciences (No. 04; HA33, T4.).

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

Yan, B., & Luo, J. (2017). Measuring technological distance for patent mapping. Journal of the Association for Information Science and Technology, 68(2), 423-437.

Appendices

Model assumptions:

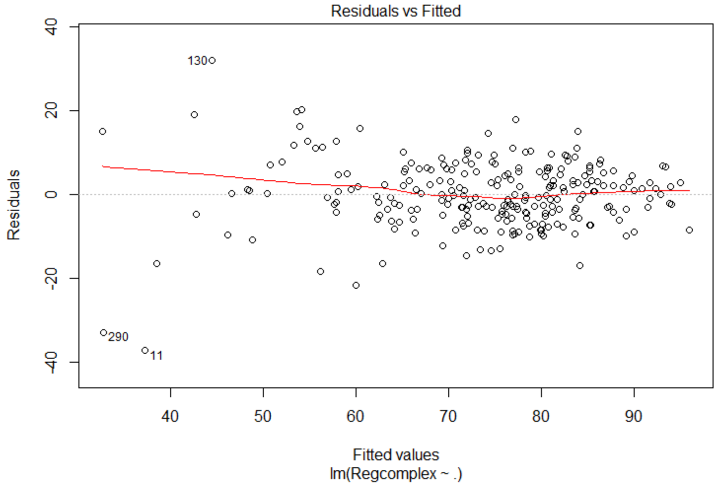


Figure 1, linearity of model 1.

It can be drawn from figure 1 that the assumption of linearity is met. This can be drawn from the fact that the dots follow a random pattern (Field, 2013).

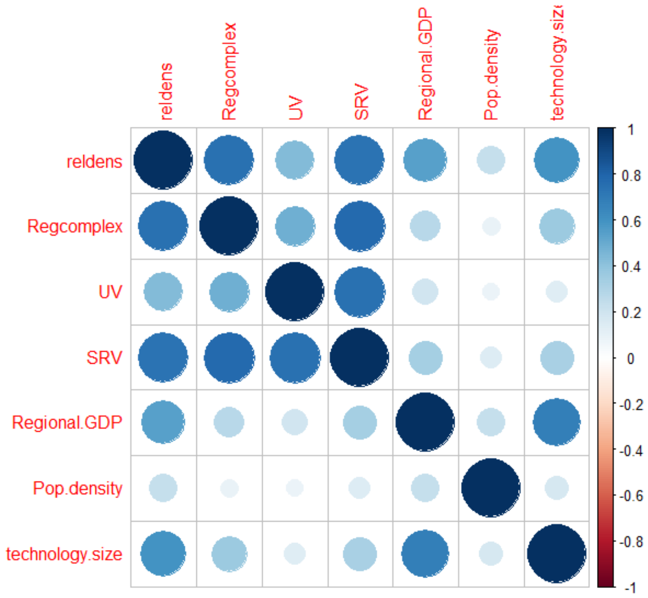


Figure 2, correlation plot of model 1

To check for correlation among the variables, the pearson's correlation has been used. When related variety was included in model 1, a correlation of 0,825 was found between UV and RV - and a correlation of 0,875 between RV and SRV. Furthermore, after checking in more detail, the VIF of RV exceeded 9. Hence, RV was left out. Figure 2 represents the Pearson correlation of the model, without RV. It was found that no variables exceed a boundary of 0,8 or -0,8 and the maximum VIF of model 1 is 3.2. Therefore, no correlation and multicollinearity exist (Field, 2013).

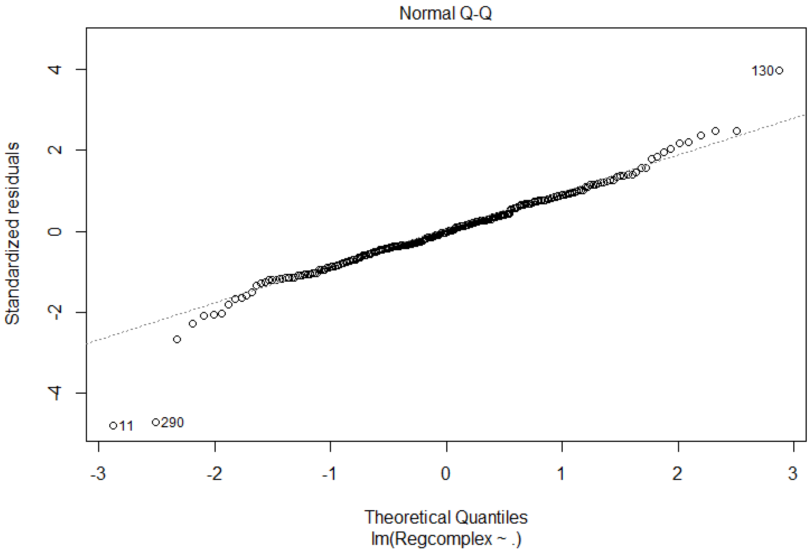


Figure 3, Q-Q plot model 1

The residuals follow the ab-line (dotted line) which indicates that the assumption of normal distribution is met (Field, 2013).

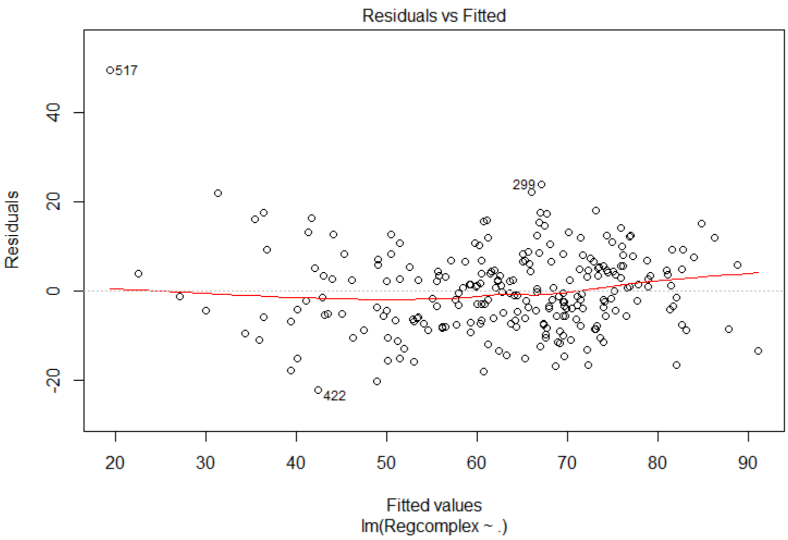


Figure 4, linearity of model 2.

The model meets the assumption of linearity as the residuals show a random pattern (Field, 2013).

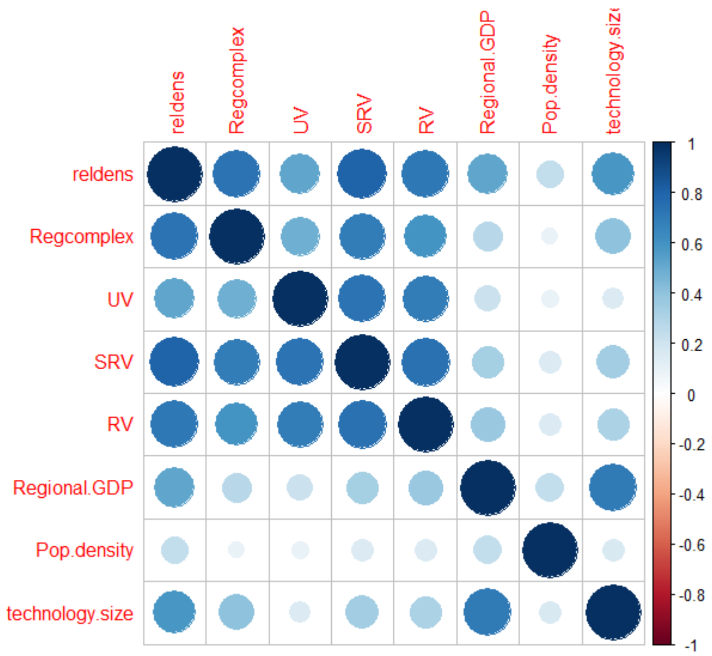


Figure 5, correlation plot of model 2.

Model 2 meets the requirement of the maximum Pearson correlation as no variable correlates with more than 0,8 (Field, 2013).

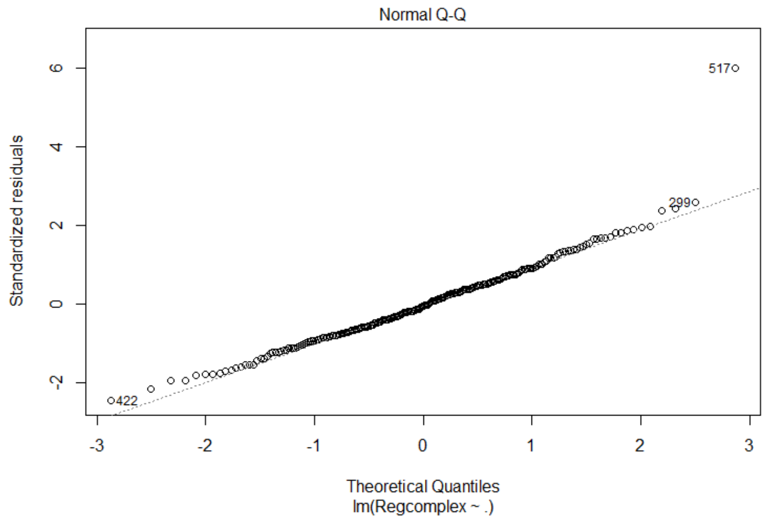


Figure 6, Q-Q plot of model 2.

The Q-Q plot meets the assumption of normal distribution as the residuals are near the diagonal that indicates a normal distribution (Field, 2013).

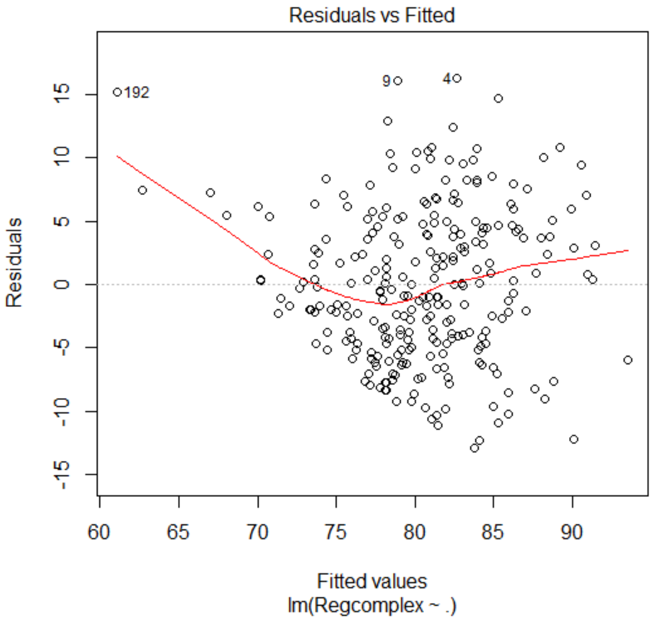


Figure 7, linearity of model 3.

It can be derived that the effect of a few outliers is large. Therefore, the outliers have been deleted and the assumption of linearity is met (de Vocht, 2016).

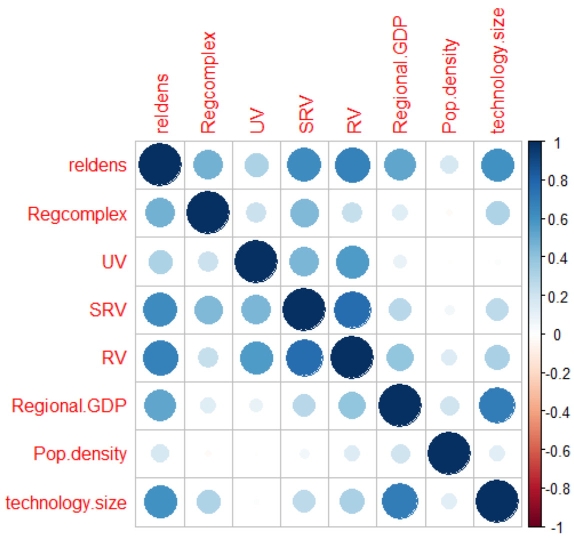


Figure 8, Pearson's correlation of model 3.

All values of the Pearson's correlation matrix do not exceed -0,8 and 0,8. Hence the assumption of correlation is met (Field, 2013).

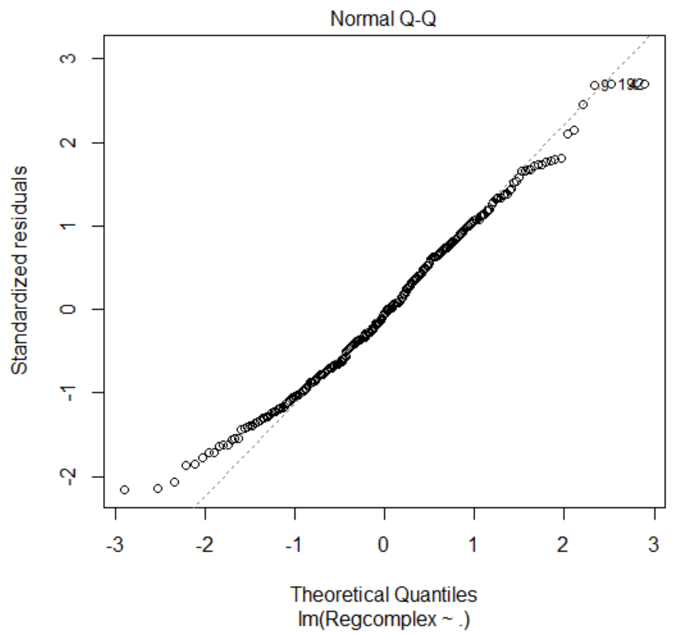


Figure 9, Q-Q plot of model 3.

The assumption of normal distribution is met. This is shown by the fact that the dotted lines roughly follow the diagonal (Field, 2013).

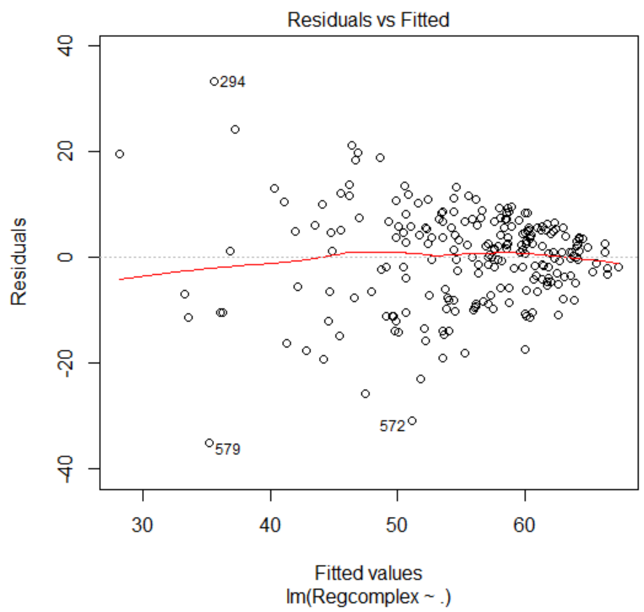


Figure 10, linearity of model 4

Although there is a minor curve, the pattern of the residuals is enough random to meet the requirement of linearity (de Vocht, 2016).

Figure 11, Pearson's correlation of model 4

It can be derived from figure 9 that the maximum Pearson's correlation is between -0,8 and 0,8. The assumption of correlation is met (Field, 2013).

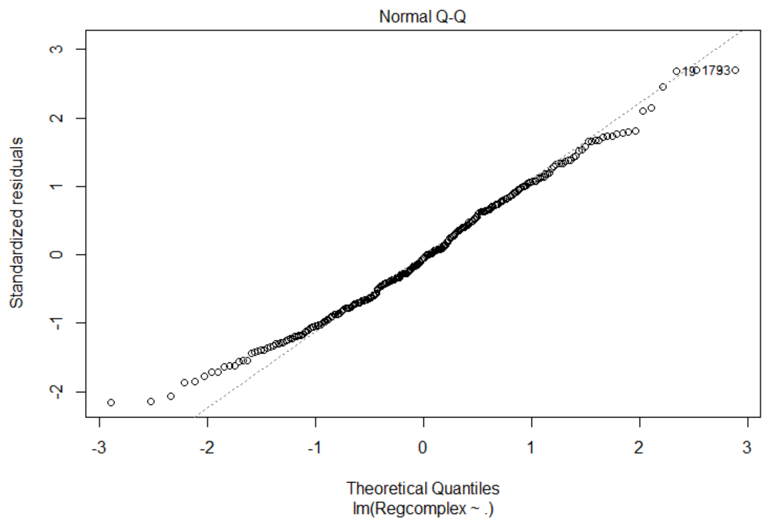


Figure 12, Q-Q plot of model 4

The assumption of normal distribution is met as all dots roughly follow the diagonal (Field, 2013).

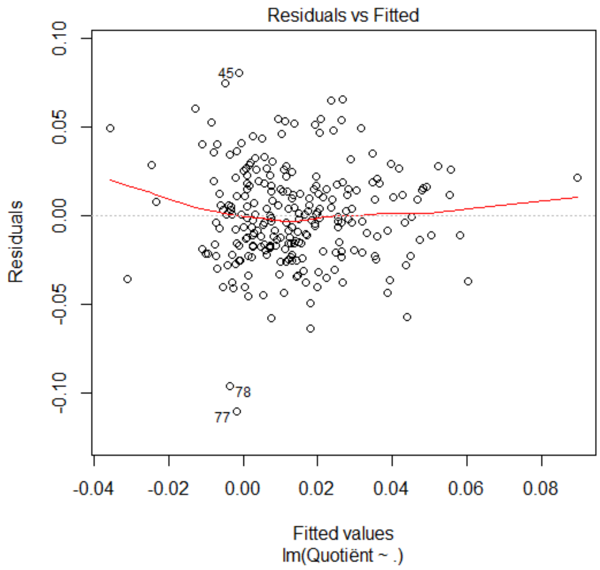


Figure 13, linearity of model 5.

Figure 13 shows the residuals vs fitted. It is shown that the dots have a random pattern. Therefore, the assumption of linearity is met (Field, 2013).

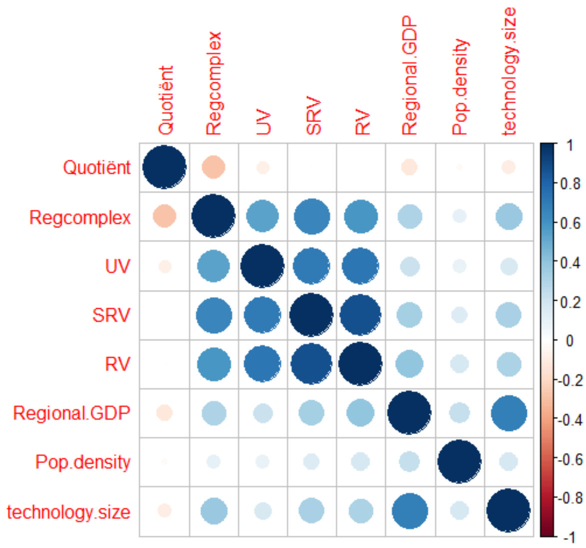


Figure 14, Pearson's correlation of model 5.

In figure 14 the correlation between all variables is visually represented. When Relatedness density was included in the model, a very high correlation between relatedness density and regional complexity existed (0,83). Furthermore, the VIF of relatedness density was 5,7. Therefore, relatedness density has been left-out.

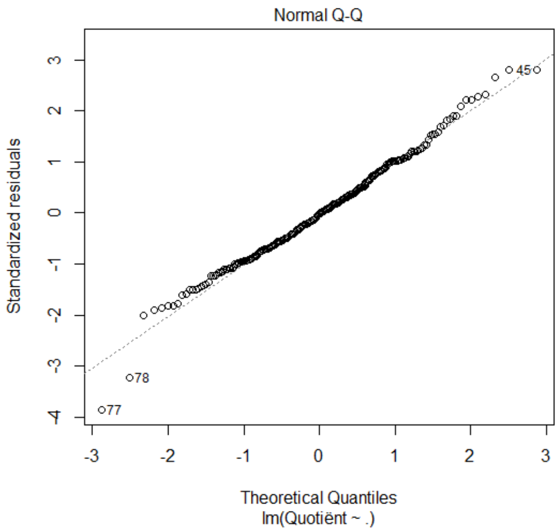


Figure 15, Q-Q plot of model 5.

Figure 15 shows the Q-Q plot of model 5. It is shown that the dots follow the diagonal. Therefore, the assumption of normal distribution is met (Field, 2013).

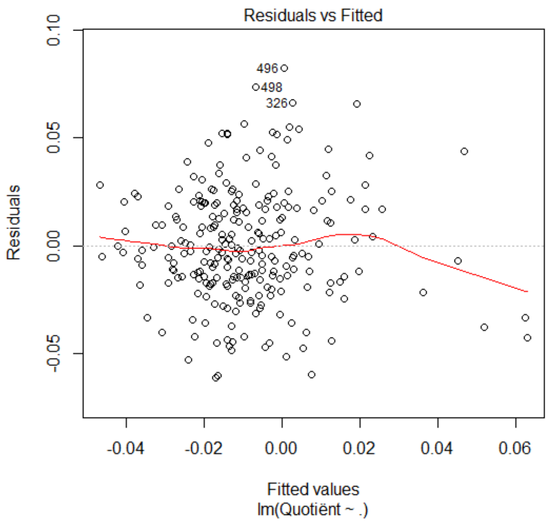


Figure 16, linearity of model 6.

As the dots show there is a random patter around the fitted line. The assumption of linearity is met (Field, 2013).

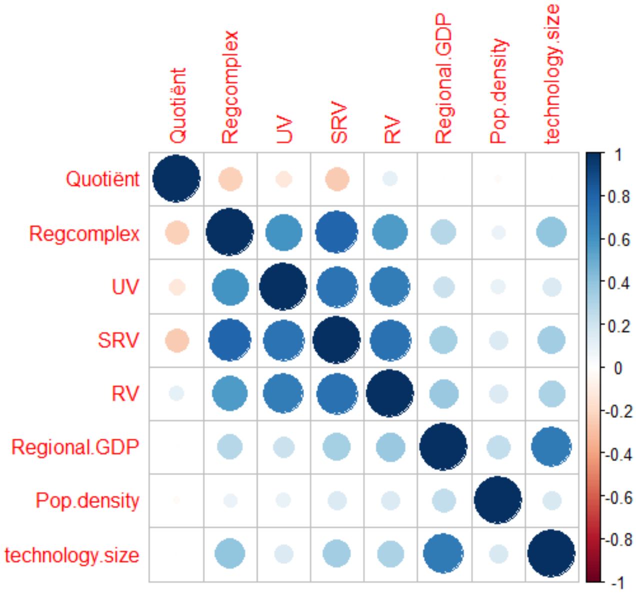


Figure 17, residuals vs fitted of model 6.

It is shown that no variable exceeds a Pearson's correlation of -0,8 and 0,8. Therefore, the assumption of correlation is met (Field, 2013).

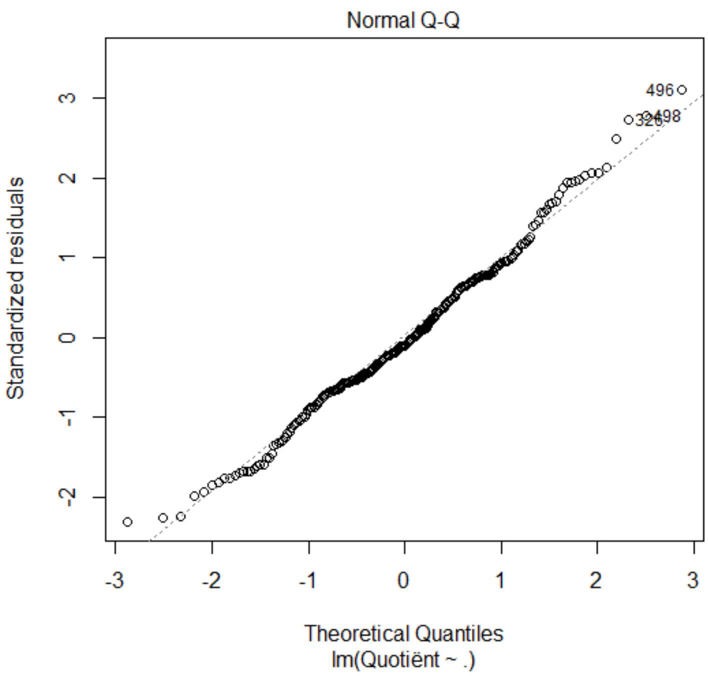


Figure 18, Q-Q plot of model 6.

The dots follow the diagonal line. Therefore, the assumption of normal distribution is met (Field, 2013).

Extra models

	Complexity model including entry-quotient
	Regional complexity T2+1 (model 2)
Relatedness density	0.24782938**' (0.1215)
SRV	0.47405713*** (2.281)
UV	-0.04081613 (3.555)
RV	0.17634721** (1.343)
Population density	-0.05142707 (0.00175)
Regional GDP	-0.20503094*** (0.00001493)
Technological size	0.18947725** (0.0001274)
Entry-quotient	-0.17274005*** (19.32)
Constant	13.81 (7.352)
Time lag	*yes*
Observations	290
R^2	0.6678
Adjusted R^2	0.6669
Residual std. Error	9.165
F statistic (df = 328)	62,32***

*Table 1, multiple regression analysis of the explanatory variables on the dependent variables' Regional complexity in t+2 with the entry-quotient included. Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

	Dependent variable: regional complexity
	Regional complexity T_{+1}. *(model 1)*
Relatedness density	0.59633754*** (0.08193)
SRV	*Excluded*
UV	*Excluded*
RV	0.29040796*** (1.283)
Population density	-0.05905396 (0.001216)
Regional GDP	-0.18299838*** (0.00001612)
Technological size	0.08795841 (0.0001365)
Constant	36.02*** (4.385)
Time lag	*yes*
Observations	290
R^2	0.5858
Adjusted R^2	0.5773
Residual std. Error	9.428
F statistic (df = 241)	68.18***

*Table 2, multiple regression analysis of the explanatory variables on the dependent variables' Regional complexity in t+1 with RV included and SRV – and UV excluded. Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

	High/low complexity model			
	Low complexity period 1	*Low complexity period 2*	*High complexity period 1*	*High complexity period 2*
Relatedness density	0.21752864 (0.1403484)	0.25193089 (0.1533)	0.2930191** (0.0791)	0.12786828 (0.1538)
SRV	0.60744381*** (2.4201228)	0.28126108* (2.628)	0.4635572*** (2.011)	0.49804957*** (2.906)
UV	0.05014753 (2.4992172)	0.03580327 (3.484)	-0.1559440* (3.669)	-0.08530624 (6.3)
RV	*Excluded*	0.38770270*** (1.379)	*Excluded*	-0.08005946 (1.987)
Population density	-0.02242856 (0.0019849)	-0.03221048 (0.001895)	-0.1006385 (0.0008608)	-0.07301566 (0.001099)
Regional GDP	-0.03384904 (0.0000306)	-0.10726566 (0.0002464)	-0.2698831** (0.0001066)	-0.30349044* (0.0001403)
Technological size	-0.04765866 (0.0006371)	-0.04590249 (0.0005868)	0.1998103 (0.00008974)	0.33213613 (0.0001224)
Constant	29.8312*** (4.7032887)	9.88 (6.808)	75.14 (9.438)	56.77*** (14.09)
Time lag	*yes*	*Yes*	*yes*	*Yes*
Observations	145	145	145	145
R^2	0.5648	0.5526	0.4374	0.329
Adjusted R^2	0.5406	0.5243	0.4104	0.2895
Residual std. Error	8.533	7.424	5.291	7.367
F statistic (df = 108, 111, 125, 119)	23.36***	19.58***	16.2***	8.334***

*Table 3, multiple regression analysis of the explanatory variables on high and low complexity divided for every period. Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*

	Dependent variable: Entry-quotient	
	Entry-quotient T_{+1}.	*Entry-quotient* $T2_{+1}$.
Relatedness density	-0.61929648*** (0.0003047)	-0.27965050* (0.0003425)
Regional complexity	-0.31004706** (0.0002002)	0.07028016 (0.0002523)
SRV	0.56744420*** (0.005692)	-0.29653522* (0.008317)
UV	-0.02128332 (0.008763)	0.06685847 (0.01047)
RV	*Excluded*	*Excluded*
Population density	0.03713555 (0.000003461)	0.01763733 (0.000003764)
Regional GDP	-0.02696060 (0.00000004808)	0.17153995 (0.00000004856)
Technological size	0.21708732* (0.0000004119)	0.14847960 (0.000003764)
Constant	0.01576 (0.01934)	0.008577 (0.02341)
Time lag	*yes*	*Yes*
Observations	290	290
R^2	0.2763	0.1473
Adjusted R^2	0.255	0.12222
Residual std. Error	0.02805	0.02928
F statistic (df = 238, 238)	12.98***	5.872***

*Table 4, multiple regression analysis of the explanatory variables on entry-quotient (including relatedness density). Coefficients are statistically significant at the *p < 0.05, **p < 0.01 and ***p < 0.001 level.*