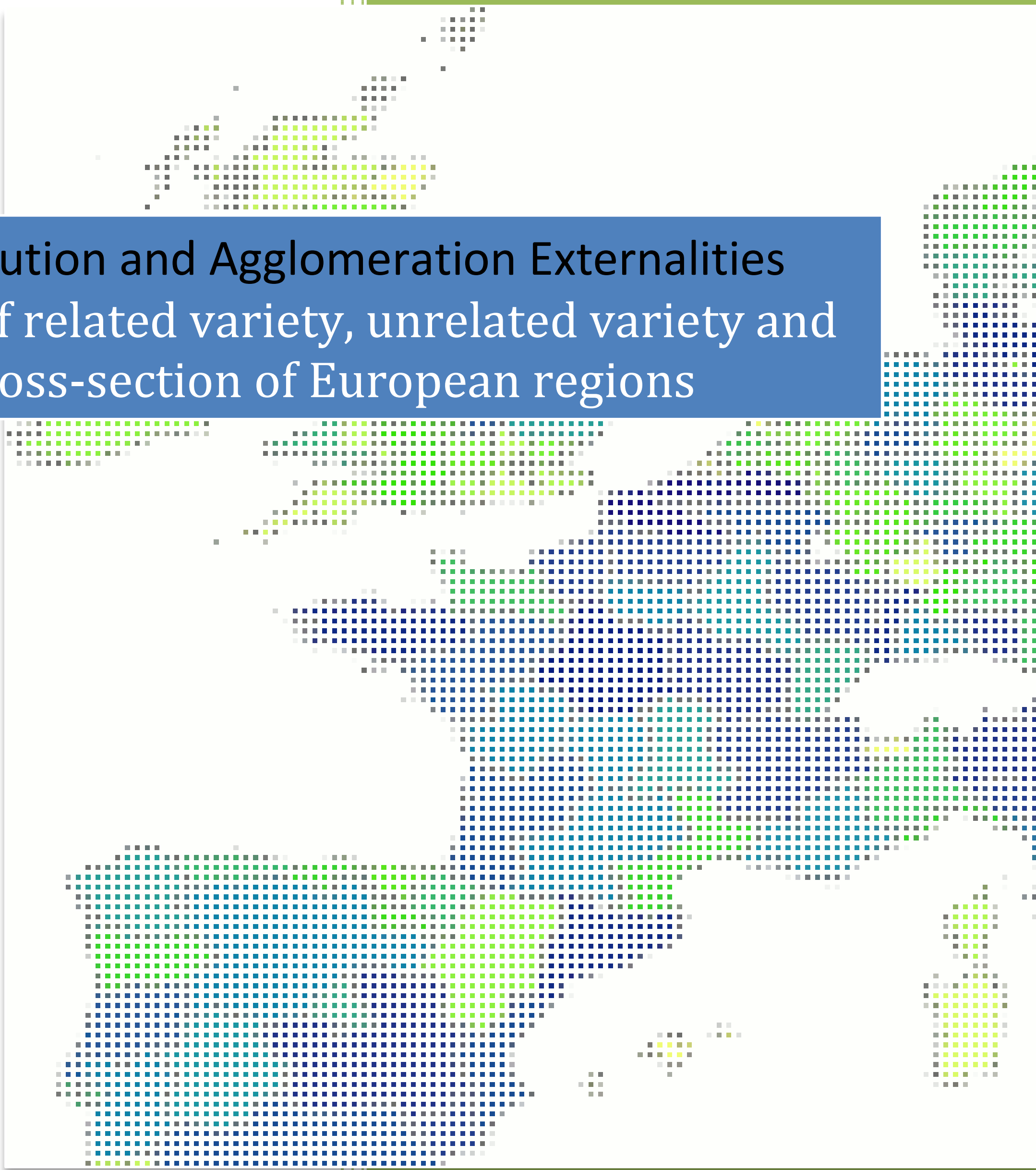
Sector structure, Evolution and Agglomeration Externalities
Testing the impact of related variety, unrelated variety and specialization in a cross-section of European regions



Master Thesis Human Geography and Planning

Utrecht University, Faculty of Geosciences

Jan-Jelle Witte

Supervisors:

- Prof. Dr. F. G. van Oort (Utrecht University)*
- Prof. Dr. R. A. Boschma (Utrecht University)

* main supervisor

September 2011

Universiteit Utrecht

Sector structure, Evolution and Agglomeration Externalities

Testing the impact of related variety, unrelated variety and specialization in a cross-section of European regions

Master Thesis Human Geography and Planning

Utrecht University

Faculty of Geosciences

Jan-Jelle Witte

Supervisors:

- Prof. Dr. F. G. van Oort (Utrecht University)*
- Prof. Dr. R. A. Boschma (Utrecht University)

* main supervisor

September 2011

Acknowledgements

This report is the product of just over a year of theoretical and empirical research. When Frank van Oort offered me the opportunity to use a large database to put a key model in Economic Geography to the test I grasped it, little suspecting of the challenges that lay ahead. With great patience he has guided me through the process of trial and error, at the same time teaching me things about science that no textbook can teach. What I am most thankful for is that at an early stage of my career Frank gave me the tools and the extensive guidance needed to do real scientific research. He made me feel more like a colleague than a student, even though at the start of the research I still had much to learn.

I also want to thank my second supervisor Ron Boschma, who thoroughly immersed me into the paradigm of Evolutionary Economic Geography and gave me valuable feedback throughout the research project. He especially had a strong influence on my theoretical understanding of regional agglomeration externalities and Economic Geography in general, and by now it is hard for me to think about these topics without path dependence, selection and cognitive proximity coming to mind.

Moreover this research would not have been possible without the large and valuable database which the Netherlands Environmental Assessment Agency has generously shared with me, and in particular I would like to mention Dario Diodato who also gave me important feedback on the technicalities of measuring regional sector structure. Finally I thank my family and friends for their support and patience during this long research project.

As I am completing this research project it feels like a beginning rather than an end. While it has answered a few of the questions I started out with its most important outcome is to give me a big new puzzle to grapple with, and I am looking forward to taking up this new challenge.

Jan-Jelle Witte

Utrecht, September 14th 2011

Summary

The relation between regional sector structures and their resulting agglomeration externalities is one of the most thoroughly studied research topics in Economic Geography and Urban and Regional Economics. Recently however this research area has faced conflicting empirical findings, while insufficiently detailed theoretical models make it hard to find the cause of these disagreements. This study builds on the model of agglomeration externalities as proposed by Frenken et al. (2007), which inspired by Evolutionary Economic Geography distinguishes related variety and unrelated variety besides specialization and urbanization.

A cross-section of 234 NUTS2 regions in 19 European countries is used to put the model to the test. Since this database is based on firm-level data, two types of sector structure variables are calculated which are rare in previous research: one based on the number of firms per sector, and the other weighted by the operating revenue of firms per sector. While several country-level tests have previously confirmed predictions based on this model, the empirical analysis in this research cannot find clear evidence in favor of the model.

When variables based on absolute numbers of firms are used limited evidence for the employment enhancing effect of related variety can be found, but the positive relation between specialization and labor productivity is too weak to be confirmed. The expected protection against unemployment growth in regions with a structure of unrelated variety could not be found, and when variables are weighted by operating revenue none of the sector structure variables show a clear effect. Controlling for the number of hours worked (for employment growth and labor productivity growth) and for specialization calculated at different levels of sector detail does not yield significantly different effects.

The paucity of confirmatory evidence may be caused by interference of the modifiable areal unit problem, besides the possibility that the expected effects of the model are only visible at different geographical or temporal scales. Further research is called for to disentangle these possible explanations of the research findings.

Contents

List of Figures and Tables

Figures

Chapter 3

Appendix A

Tables

Chapter 2

Chapter 3

Appendix B

1. Introduction

Ever since the seminal contribution of Marshall (1890) the twin concepts of sector structure and agglomeration externalities have been at the heart of Economic Geography. After extensive empirical testing and theoretical debate a set of tools has emerged for the study of the long-term growth prospects of regions, and this process of gradual refinement has recently been accelerated by new research findings. While Marshall already understood that agglomeration externalities are composed of very different elements with their own dynamics, ranging from pecuniary externalities to the spill-over of knowledge, insights from the endogenous growth theory and the evolutionary perspective on economic growth have recently shifted research focus more strongly towards knowledge externalities. At the same time our understanding of the role of different sector structures in enhancing or impeding the generation of agglomeration externalities has been deepened by studies that followed Glaeser (1992) in comparing the different effects of regional specialization and diversity.

The dichotomy of regional specialization and diversity employed following Glaeser (e.g. Henderson et al., 1995, Feldman & Audretsch, 1999; see de Groot et al. (2009) for a list of more recent studies) is however running into both theoretical and empirical limitations. Empirical outcomes are conflicting (Beaudry & Schiffaurova, 2009) and theoretical models are not detailed enough to guide empirical studies in the selection of the most suitable spatial scale, time scale and level of sector detail at which to measure sector structure. At a time of protracted economic uncertainties models of agglomeration externalities and sector structure could provide policymakers with much-needed guidance on how to promote sustainable regional growth and employment in the long run. This adds a sense of urgency to answering the open questions of which sector structures promote regional economic development, and at what geographical and temporal scale.

The literature has recently made progress towards building up a more coherent theoretical framework for agglomeration externalities by linking this concept with the evolutionary approach to Economic Geography. Frenken et al. (2007) have disentangled the concept of diversity into related variety and unrelated variety, and argue that based on insights from the evolutionary perspective related variety should be especially important in generating knowledge externalities. While significant empirical support has been provided for this hypothesis (besides Frenken et al. 2007 also Essletzbichler 2007, Bishop & Gripaios 2010, Boschma & Iammarino 2009 and Neffke et al., 2011), this support has so far been limited to country studies only. Based on a European database covering 19 EU countries this study puts the hypotheses on agglomeration externalities stemming from the evolutionary perspective more thoroughly to the test.

While significant progress has been made in improving our theoretical understanding of agglomeration externalities and sector structure, recent studies still leave many questions unanswered. This study explores theoretical reasons for the selection of among others the geographical and temporal scale of analysis and the level of sector-detail used in calculating measures of specialization and variety. The available data allows a partial test of these theoretical guidelines, leaving others for further research.

The following chapter will review the literature on agglomeration externalities with a focus on the role of regional sector structure, and explores ways in which the evolutionary perspective on

Economic Geography can help clarify a number of open questions: which forms of externalities (pecuniary or knowledge-related) to expect to matter at which time-scales; the selection of the outcome variable of the model; the geographical scale of analysis and the level of sector detail most relevant to the calculation of specialization and diversity measures. Chapter 3 explains the empirical model and data sources used to test the theoretical model, and chapter 4 presents the results of these tests. Chapter 5 concludes and suggests several directions for further research.

2. Literature review

In the quest to understand uneven economic development a key question is why some regions are able to maintain high levels of economic growth and employment, while other regions either do not reach this goal or cannot sustain it in the long run. It is a fascinating scientific question that also has great implications for policy making. Over the past decades scholars of this question have moved towards a consensus that whatever the precise answer may be, innovation is likely to play the key role in it. While natural resources, an attractive location and other factors that mostly lead to cost savings have often been invoked as explanations for regional competitiveness, they are usually temporary factors that only sustain a specific type of economic activity. Technological change continuously redefines which resources and location factors can support competitive industries, and when technology shifts natural resources cannot be adapted to fit with the new requirements. Knowledge is the key production factor that can be adapted to keep up with technological change, while knowledge is at the same time the key force driving technological change forward. Knowledge is likely to be the most complex, but also the most important factor in explaining why some regions are able to sustain high levels of economic growth and employment in the long run.

Sustainable innovative regional economies differ from other regions not in their level of productivity at any point in time, but by their ability to regain competitiveness any time they are hit by a crisis or technological change has obsoleted their key industries. In other words despite the fact that they face crises just like any other region, they can still maintain a meta-stability in which one cutting-edge industry after another carry the economy forward. One of the scholars of this question who continues to inspire Economic Geographers is Jane Jacobs. Throughout human history she finds examples of cities and regions which did not rely on one successful product to carry their economy forward, and therefore did not sink back towards lower standards of living when a key industry lost its edge. Instead such regions enjoy a meta-stability in which firms and workers continuously find ways to do their work in a new way and produce new products. While many of these innovations may hardly have been noticeable at the time since they more often arise from normal worker's experience and people's daily lives rather than from state-of-the-art laboratories, they can set in motion development paths with the potential to over time transform the local economy (Jacobs, 1969). The vision arises of regional economies being able to almost naturally and spontaneously update themselves, unless their creativity is stifled by unhelpful economic structures or rules. This is a very different starting point than to exclusively focus on investments in R&D, paying out in terms of stocks of patents that lead to a predictable improvement in firm productivity, as has been very common in Economics and Economic Geography. Instead the focus is on regions facing the need to create a continuing stream of new products and technologies, even of entire new industrial sectors, to keep a meta-stable level of high incomes and high economic growth. This goes beyond common measures of innovation as patents or other 'lumps' of R&D output. Instead it is innovation in the sense of being able to completely renew the economic base of a region. So how can regional differences in the ability to innovate and create new products and industrial sectors be explained? A rich diversity of theoretical approaches has undertaken to answer this question, mostly revolving around benefits stemming from agglomerations of firms in innovative regions. This chapter first offers a critical survey of the literature and argues that there is a lack of understanding of the mechanisms of these various forms of agglomeration externalities. The second half of the chapter first discusses whether the evolutionary perspective on economic geography is compatible with the concept of agglomeration externalities, and then uses evolutionary concepts as a starting point for building a more detailed theoretical model of agglomeration externalities which will be empirically tested in the next chapter.

2.1: Clusters and institutions

In the New Industrial Geography or Institutional Economic Geography a rich literature has developed from around the 1990s onwards on the mechanisms that drive high rates of innovation in some regions but not in others. In this literature formal (e.g. legal rules) and informal institutions (social norms and shared expectations) shared by entrepreneurs and their workers and tied to a specific regional context lead to cost savings (e.g. by reduced transaction costs) and learning effects (e.g. by knowledge sharing and R&D cooperation) which firms within that region can enjoy. Institutional Economic Geography has provided detailed accounts of the mechanisms that may lead to differences of regional levels of innovation, and the learning effects described in the literature seem to fit especially well in a small number of specific regional that display exceptional and sustained high levels of innovation. The precise mechanisms associated with regional competitiveness are often unique to the region in which they developed, ranging from local cultures of risk-taking and openness to outsiders with good entrepreneurial ideas (as in Silicon Valley, e.g. Saxenian, 1996) to local networks of co-operation and mutual support between co-located firms that built up strong bonds of trust over long periods of time (as in the Third Italy region, e.g. Markusen, 1996). Recently through social network analysis great progress has been made in making these mechanisms concrete and testable (e.g. Boschma & Ter Wal, 2007).

But overall many of the mechanisms based on institutions and clusters are still hard to quantify, measure and test. Arguments that are essentially based on the most successful cases of innovative economies allow very limited generalization to the economic landscape in general, and these case studies often result in over-fitted models in which it is unclear which variable is more important than another. Moreover most researchers in the institutional approach would deny that formal modeling of their theoretical ideas is possible at all, since they argue that the mechanisms they describe can only be understood within their respective regional contexts and cannot be abstracted into more general, universal hypotheses. This resistance to generalization and modeling has led to a great diversity (some would say chaos) of often inconsistent concepts (Martin & Sunley, 2003), and theories in which it is unclear how exactly institutions are tied to territories and improve the competitiveness of this entire territory. (Boschma & Frenken, 2006: 287). If territorial institutions cannot be precisely defined and measured then this seems to limit the scope for comparative research. Notwithstanding these problems or inherent limitations, elements of these institutional models are already widely used for the design of regional development policy far beyond the cases on which these models were originally based, which adds to the urgency of getting a better understanding of the mechanisms behind regional differences in innovation levels (Martin & Sunley, 2003).

2.2: Agglomeration externalities

Many scholars in Economic Geography and the closely related research fields of Geographical Economics, Regional Science and Urban and Regional Economics (here they are grouped together as Geographical Economics) try to answer the same question as Institutional Economic Geography, but with a methodology of formal modeling. Rather than attempting to capture the full richness of possible factors that may promote or impede innovation within a specific region or cluster, they build up a model starting from the most basic hypotheses on agglomeration externalities and then adding additional potential factors one by one. The goal is to explain not only the extraordinary success of some regions, but differences in the economic landscape as a whole. Based on endogenous growth theory in Economics, Geographical Economists focus on finding the places where knowledge is created and study how knowledge diffuses from there as a key question for understanding economic growth. They find knowledge creation in dense concentrations of economic activities (which usually, but not necessarily, coincides with cities, Burger et al., 2011), with knowledge spill-over being geographically limited (Jaffe et al., 1993) and therefore also occurring mostly within the regions or cities where knowledge is created.

A key concept arising from these interlinked research fields is that of agglomeration externalities. When firms co-locate they influence each other both positively (through cost savings and knowledge spillovers) and negatively (e.g. through crowding effects, leading to higher land prices and congestion). In other words economic activity bunched together within a city or region can make every firm within that city or region more competitive (receiving more knowledge spill-overs, benefiting from cost savings) or less competitive (having to pay higher factor prices, being hurt by crowding effects). Rather than studying specific cases of agglomeration, or regional clusters, this strand of research tries to understand agglomeration or clustering in general. For if clustering as a general process does not produce the expected agglomeration externalities, then the study of examples of regional clusters also seems fruitless (Weterings et al. 2007).

Recently in Geographical Economics the study of agglomeration externalities has taken the form of models that focus on the sector composition of firms in agglomerations, rather than on agglomeration of economic activity per se. They juxtapose regional specialization and regional diversity in sector distribution to find which is associated with productivity growth and employment growth. Specialization is associated with localization economies (firms only benefit from co-location with firms in the same industrial sector), which usually takes the form of cost savings for co-located firms who use shared pools of specialized labor and shared intermediary suppliers and service-providers. Besides cost savings specialization has also been associated with knowledge spill-overs between firms, assuming that firms in the same sector face similar challenges and opportunities and can therefore learn from each other. Diversity on the other hand is commonly associated with urbanization economies (firms benefit from being located in big, densely populated cities) and with so-called Jacobs' externalities (firms benefit from co-located firms in any sector). Big and dense cities can by virtue of their population size or level of economic activity sustain shared facilities such as harbors, airports and universities, which reduce operating costs (e.g. lower transportation costs for intermediary and final products in the vicinity of a harbor or airport; lower search costs for highly educated workers in the proximity of a university). While such urbanization economies are basically about cost savings, the concept of Jacobs' externalities attempts to capture the knowledge spill-overs firms may enjoy in diversified cities. Although the precise mechanisms are rarely worked out in a systematic way, the main idea is that as firms producing different products and harboring different knowledge and ideas are co-located new ideas and products spring up from their interactions. So both specialization and diversity are associated with both pecuniary externalities and knowledge externalities.

However if in a knowledge-based economy knowledge spill-overs and learning effects rather than cost savings are the key factor in understanding long-term regional productivity and employment growth, as suggested by endogenous growth and the evolutionary perspective, then it becomes essential to separate these kinds of agglomeration externalities. A comparison of two sector structures which can both be associated with knowledge dynamics seems unlikely to result in a clear outcome. Moreover Frenken et al. 2007 suggest that long-term gains in economic performance of regions due to a diverse sector structure can also be explained by a mechanism not related to agglomeration externalities, namely the portfolio effect. If part of the cyclical shocks to a regional economy are asymmetric (affecting one or a few sectors rather than all sectors equally), then the more sectors a region harbors the higher the chance that some of the region's firms are not or only to a limited extent affected by any one cyclical shock. In other words the more diverse a region's sector structure is, the more stable its overall employment level will be. Diverse regions are likely to face many small shocks to their economy rather than fewer but heavier shocks, and the former may be less damaging in the long run.

Not surprisingly the empirical results from studies comparing specialization and diversity are ambiguous, sometimes specialization turns out to be the most important source of agglomeration externalities (e.g. Henderson et al., 1995) and sometimes it is diversity (e.g. Glaeser et al., 1992). In fact the number of studies that find positive results for specialization is roughly equal to the number of studies that find benefits of diversity (Beaudry & Schiffaurova, 2009). With such great empirical disagreement on the effects of regional specialization and regional diversity this basic model of agglomeration externalities remains a weak basis for further theory building. It may be possible to explain these different outcomes by adding additional variables to the model, for example by saying that specialization benefits mature firms while diversity benefits young and growing firms (as in the industrial life-cycle) (Duranton & Puga, 2000). However a more fruitful way to tackle these issues may be to address the limited understanding of the mechanisms behind agglomeration externalities. Frenken et al. (2007) propose to disentangle the sector structure of diversity by distinguishing urbanization externalities, Jacobs externalities and the portfolio effect and testing the effect of each element of diversity separately. However they do not propose a way to disentangle the effects of specialization, which can still consist of both pecuniary and knowledge externalities.

Even after the contribution of Frenken et al. (2007) we are still left with a model of agglomeration externalities that offers very little detail in its understanding of its mechanisms. And while the models so far do often place a strong emphasis on knowledge dynamics they are usually based on a simplistic conceptualization of knowledge incompatible with for example Jacobs' (1969) work, since they tend to think in terms of accumulating 'lumps' of knowledge rather than innovation in terms of continually renewing your economic base (Neffke et al., 2011, compare a quantitative and a qualitative conception of knowledge). As a result of these shortcomings the models give very little guidance to empirical studies as to the appropriate geographical scale on which agglomeration externalities are to be expected to work. Empirical studies so far have ranged from regions as large as US states and as small as separate districts within cities, and their findings seem to be determined partly by the geographical scale chosen (Beaudry & Schiffaurova, 2009). Similarly it is unclear at what time scale the effects on the economic performance of regions should take place, and whether the effects are expected to be similar or different on all time scales. Only mechanisms that can be expected to work on the long run would provide possible answers to the question of why some regions reach high and meta-stable economic performance while other regions do not. This study proposes an enhanced model of agglomeration externalities in order to solve these limitations, and finds inspiration in Evolutionary Economic Geography.

2.3: Evolutionary Economic Geography

The evolutionary perspective on Economic Geography both aims to solve the limited understanding of the mechanisms of agglomeration externalities, and offers a much more elaborate understanding of the meaning and significance of innovation. While it started as basically a spatial version of the evolutionary approach to Economics, it has recently grown into a distinct and coherent theoretical framework in Economic Geography. However there is as of yet no fully developed evolutionary model of agglomeration externalities. Also there are some theoretical contradictions between evolutionary thinking and the literature on agglomeration externalities, which will be discussed after a brief overview of the core ideas of the evolutionary perspective.

The key idea in the evolutionary approach is that evolutionary change is not limited to the mutation of and selection on the genes of plants and animals in Biology, but that this is a fundamental feature of many other aspects of life as well (Essletzbichler & Rigby, 2010:44). Whenever there is a population that consists of entities (be it animals, people, organizations, languages or other agents or structures) that differ from each other and that exists in a context of scarcity (be it of food, money, votes, members or other resources) then it is likely that selection processes occur. Entities that possess traits that, under the specific circumstances of the place and time in which they exist, increase their access to the scarce resource they depend on are likely to survive while other entities may cease to exist. The traits (be it the ability of an animal to cope with low temperatures or the ability of a firm to minimize its operation costs, or any physical or behavioral trait) carried by the surviving entities will become more common in the population while traits that did not give a benefit under the precise circumstances of that time and place will become less common. While in Biological evolution animals cannot change their genes during their life-time, entities beyond Biology can often adapt their traits to become better adapted to the circumstances of their time and place (Boschma & Frenken, 2006:78). For example a firm can learn to become more cost efficient, or a political party can learn to use new media to attract a new generation of voters. So well-adapted traits can become more common not only because they are being carried by entities which have a higher chance of survival, but also because entities try to unlearn ill-adapted traits and learn the well-adapted trait. And just as genes can change spontaneously through mutation leading to new variety within the population, the traits of entities beyond Biology can also change in spontaneous ways (or in ways so complex that they appear as spontaneous mutation). Selection environments continuously change, thereby continuously redefining which resources are scarce and which are abundant and which traits increase or decrease access to these resources. Changing selection environments combined with the spontaneous mutation of traits result in always changing populations, always adapting towards the moving target of the selection environment and never reaching a stable equilibrium. While the causal link between selection environments and firm populations is very complex, it is possible that characteristics of the selection environment influence economic performance (Essletzbichler & Rigby 2010:54).

While it is not difficult to conceive of firms as entities competing for scarce resources, the application of evolutionary change to social systems also requires a concept equivalent to genes in Biology. The concept of routines has been proposed to describe how firms acquire and retain the traits that make them more or less fit within their specific selection environment. Routines are organizational skills that guide the action of a firm, and that are more than the sum of individual skills of the firms' workers. They consist of tacit knowledge and learning-by-doing, so they are hard to imitate. Differences in routines are a key aspect of what makes the individual firms in a population heterogeneous, and the tacitness of the knowledge embodied in routines causes the differences among firms to remain rather than quickly diffuse. While animals cannot change their genes to increase their fitness, firms can consciously or unconsciously change their routines to adapt to their changing environment. Since routines are tacit and deeply ingrained into the minds of the actors that make up a firm it is a costly, slow and painful process to change them, and firms are likely to avoid doing so if they can. If firms do end up changing their routines they are likely to face limitations as to the scope of these changes, as will be elaborated below. By studying the routines of heterogeneous firms, and the changing selection environments in which these firms operate, Evolutionary Economic Geography aims to understand "*the spatial distribution of routines over time*" and especially "[..] *the creation and diffusion of new routines in space, and the mechanisms through which the diffusion of 'fitter' routines occurs*". And as a consequence "[..] *the emergence of spatial agglomerations is to be analysed* [..] *in terms of the historically grown spatial concentration of knowledge residing in organizational routines*" (Boschma & Frenken, 2006:277-279).

2.4: Technological evolution and innovation

The evolutionary perspective on regional development has fundamental implications for our understanding of how innovation works and why it is so crucial for long-term economic growth. At the firm level innovation cannot be understood as firms 'purchasing lumps of knowledge' through R&D investment, buying the 'lumps' with the highest expected value until the marginal return of more innovation is zero and further investment is inefficient (a similar argument is made in Neffke et al., 2011). Instead firms act according to behavioral rules (in other words: actors are boundedly rational rather than perfectly rational) which they developed over time and which may be well-adapted or out-of-tune with the current market environment, and it is the change of these routines themselves which constitutes innovation.

When a car manufacturer successfully upgrades its production from petrol cars to electric cars this innovation should not be understood as if the firm simply acquired a new machine that enables it to produce a new product, but rather as the firm changing its set of routines to become able to produce, market and compete in a new product sector. This might for example include changing the behavioral routine "we produce what the majority of customers wants to buy at this moment" to "we produce a car which customers can afford to use in the near future, actively building a consumer market for it rather than responding only to the currently existing consumer market". Many of such changes of routines would be necessary before this particular innovation track can be successfully completed. A factory stocked with updated machines producing a new product is then only the most visible aspect of an innovation track carried by change in routines. And while it is relatively easy to buy or create the new machinery, many of the essential routine changes cannot be brought about by investing in R&D and instead have to be learned through experience or the imitation of the experience of other firms.

Besides innovation being a more complicated process for firms to carry out, a large part of the innovations that occur are likely to be the result of serendipitous and unplanned sparks that occur when different ideas are combined as 'Neue Kombinationen'. While investing in R&D can help in bringing ideas together, it is equally important for firms to have unwritten business rules that promote creativity, risk taking and trial and error. And without any investment in R&D, or even without a conscious intention to innovate, it should still be possible for innovation to occur from the interaction between firms and between workers within a firm. At the regional scale an increase in the "aggregate regional techniques of production" cannot be understood as the aggregate of the process innovations carried out by firms in that region, but rather as the change in routines of incumbent firms, the differential growth of incumbent firms, and evolutionary selection of some routines at the cost of other routines through firm exit and firms unlearning unfit routines (Essletzbichler & Rigby, 2010:50).

Besides offering a different understanding of how innovation works, the evolutionary approach also provides an answer to why it is so important. As selection environments evolve they constantly change the definition of what routines are fit and which are unfit, so continuous innovation is needed for firms to be able to keep up with the constantly changing rules of the game. This implies that in a comparison of the potential benefits from agglomeration externalities, knowledge externalities contributing to the ability of firms to innovate is expected to strongly outweigh cost saving externalities from for example shared infrastructures and shared pools of specialized labor which allow firms to produce the same products more cheaply.

The argument so far suggests some reasons why the controversy between the benefits of specialized regions and diverse regions has not moved much towards an outcome. Firstly it is crucial that innovation is defined in the right way. Using R&D investments or hours worked (FTE) by knowledge workers is unlikely to capture innovation as the change of routines allowing new products to be successfully developed. Patents and new product announcements are perhaps the closest we can get to measuring innovation in a way that is consistent with an evolutionary perspective, but these are still only indicators and not measures of innovation as change of routines. If innovation is not measured in an appropriate way then its importance, and thereby the importance of those agglomeration externalities leading to innovation, is likely to be underestimated.

Secondly cost savings and knowledge spill-overs must be clearly distinguished, since agglomeration externalities that do not lead to knowledge spill-overs and innovation are not expected to contribute significantly to long-term growth. As discussed in paragraph 2.2 the concepts of specialized and diverse regional industrial structures do not adequately make this distinction. Thirdly any model of long-term regional economic development should take a dynamic rather than a static perspective. In a static model actors make one-off decisions in response to the environment at that time, and regional structures and factor endowment deliver a one-off benefit to those firms that can access them. This may suffice for understanding pecuniary externalities resulting from specialization, since it is reasonable to assume that some forms of cost saving more or less automatically occur when firms become co-located. If a time lag is included in such a static model of agglomeration externalities then co-located firms in the same industry sector will also receive a one-off benefit from the infrastructure that is created for them once a critical mass of agglomeration is achieved, and also from the shared pool of labor this firms build up over time and from other shared factor endowments that develop. Perhaps most of the benefits of specialization can be captured in this static analysis.

However the importance of knowledge spill-overs and collaborated learning processes, mostly associated with regional diversity, are likely to be strongly underestimated by such models. The decision to trust another firm enough to engage in collaborated innovation projects is not a one-off decision but rather one that depends on the gradual creation of trust between firms, which may be a complex process that not only takes time but also takes many steps to reach its conclusion. Likewise a knowledge spill-over may give some direct benefit to the receiving firm, but the real significance of knowledge spill-overs is that the innovation it triggers can lead to new knowledge spill-overs, possibly triggering another innovation and so forth. Such indirect effects, feedback effects and chain reactions are completely lost in a static model with or without a time lag, and require instead a truly dynamic understanding of agglomeration externalities that is more challenging to capture in a formal model. Assuming that agglomeration externalities from diversity are mostly related to knowledge spill-overs and learning effects and specialization mostly to pecuniary externalities, the static models used so far to compare specialization and diversity will underestimate the benefits of diversity. The longer the time scale of the analysis, the stronger the impact of feedback effects and chain reactions will be on the outcome of economic growth and therefore the more strongly knowledge-related agglomeration externalities will be under-estimated. The discussion of the evolutionary approach to Economic Geography so far has made it clear that evolutionary models are by definition dynamic, and pay close attention to complex feedback mechanisms and the ways in which decisions, through chain reactions, can have strong consequences on the long run (Boschma & Frenken 2006: 284-285).

2.5: Evolution and Path Dependence

So far the idea that agglomeration externalities can be beneficial to a region’s long term growth prospects seems compatible with the evolutionary approach, provided that dynamic models with suitable measures of innovation and a clear distinction of knowledge-based vis-à-vis pecuniary externalities are used. However before an understanding of agglomeration externalities compatible with Evolutionary Economic Geography can be proposed, there is another challenge to be met first. This challenge is that the idea that the characteristics and factor endowments of a region can strongly promote its economic development is much more problematic in Evolutionary Economic Geography than it is in Geographical Economics. In many studies in Evolutionary Economic Geography it is even problematic to assume any influence at all from existing structures and factor endowments. The problem arises when technological change and product development is analyzed in a more systematic way, as is done through the product life-cycle model.

Scholars of the product life-cycle contend that innovation is not a continuous and unbroken process in which firms can make small improvements to their production on a year by year basis, without necessarily ever running out of new potential improvements to make. Rather innovation consists of distinct development paths of products that are introduced, refined and finally become first mature products and then obsolete products. For example in an influential study of the evolution of the British automobile industry (Boschma & Wenting 2007) the development trajectory of car production follows a distinct pattern. First many firms attempt to produce cars according to their own designs. After some time car designs become more stable and similar to one another and a ‘shake-out’ process starts in which most of the initial car producers are driven out of the market by a smaller and smaller number of growing firms whose designs become more dominant in the industry. Finally only one or a few firms are still in the market, competing mainly on efficiently producing a standardized design rather than competing on innovative new designs. Based on this model it is commonly argued that agglomeration externalities matter only in later stages of the product life-cycle when product designs have started to stabilize, while in the early phase of the product life-cycle firms only suffer from the competition of co-located firms rather than benefit from agglomeration externalities. Since in the mature phase of the product life-cycle firms compete on efficiency rather than innovative designs, only agglomeration externalities that provide cost savings would be expected to matter (Boschma & Wenting 2007:218). If all product development trajectories follow the pattern of this ideal product life-cycle then the influence of agglomeration externalities on the aggregate of technological change would be limited to extending the lives of mature industries rather than fostering innovation and technological change. The causal relationship could even be reversed in this standard model, with exogenous technological change leading first to populations of firms that only experience agglomeration diseconomies from crowding and fierce local competition and only later (and due only to the internal logic of the technological trajectory) to a situation in which firms can benefit from agglomeration economies. As a consequence models of regional economic development could ignore any pre-existing structures and factor endowments (in other words, the models can assume neutral space), and then allow processes other than agglomeration externalities (such as spin-off) to create factor endowments and regional structures of specialization and diversity. Once created, these factor endowments and regional structures may or may not lead to agglomeration externalities, but in either case these externalities would merely re-enforce the pattern that was already developing (Boschma & Frenken 2006: 289-290). Can the concept of agglomeration externalities, whether from specialization or diversity, be applied in an evolutionary framework at all, and if so would it be a key factor or a marginal one?

The question of whether existing structures (e.g. sector structure, factor endowments) influence current and future development, and if so in what way, can be systematically explored using the concept of path dependency. This concept did not emerge from the evolutionary perspective itself, but has over time proven to be very compatible with concepts of evolving routines and selection environments (Martin & Sunley 2010). The concept of path dependence seems to overlap with the concept of product life-cycles, except that it is more broad and allows technological trajectories to develop according to very different patterns. Moreover, industries, products and firms can evolve in continuously changing ways rather than being tied to any fixed pattern that eventually leads to some form of equilibrium outcome.

The key idea is that regional development proceeds in paths in which one phase builds on the previous one. At the heart of these paths are technologies, which gradually emerge and unfold their impact on the economy. The development of a new technology (for example the computer) is a long and gradual process, and even when it is introduced to the economy in the form of a new product or production process it still only gradually makes its impact. Firms and other actors do not immediately respond to the new challenges and opportunities as if they made an over-night shift in their demand and supply curves, but rather they gradually adapt their routines to fit with the new situation. Even when computers were already affordable to smaller firms and normal households people continued to use old methods, and they took time to learn the new skills necessary for unlocking the full potential of the computer. Decades later many of the new opportunities brought about by the invention of the computer have yet to be implemented on a large scale, or have not even been discovered yet. Since technological change is a gradual process many things can change while the path is still unfolding itself. At each stage in the development process there are several directions the development path can take, some directions being more likely and others less likely but still possible. So the development path up to the present, in other words the ‘history’ in the model, provides constraints and opportunities to current development but does not determine it (Martin & Sunley 2006, Martin & Sunley 2010).

If history does not determine development then paths are not fully predictable, and the assumption of neutral space is unwarranted. However in early conceptions of path dependency, paths did exhibit certain characteristics that largely coincide with the different phases of the product life-cyle model described above. First many unfinished designs are proposed, then some designs emerge and push out the rest. Innovation is first focused on new product development, but becomes limited to process innovation in the mature phase of the product life-cycle. In the original conception of path dependence (Martin & Sunley, 2010 call this the 'David conception' after its key proponent) these typical paths were conceived as slowly moving towards an equilibrium, and once enough momentum is built up in one direction then it becomes more and more inescapable that this equilibrium will be the final outcome. In the final stage one dominant design dominates all others, further change becomes nearly impossible and the economy is in equilibrium until an exogenous shock breaks it open again. Often the example of the QWERTY keyboard design is used as an illustration of this typical technological trajectory. The QWERTY design first came to dominate the production of the first typing machines since it solved a problem that had plagued earlier designs. When keys located next to each other on the keyboard were typed in rapid succession the machine could jam, which was solved by the QWERTY design by locating common combinations of letters away from each other and thereby forcing the typist to slow down. By the time the electronic type-writer and the computer replaced earlier type-writers QWERTY had become an inferior design that needlessly slows down the typist. But because it had effectively pushed other designs out of the market and whole generations of typists had been trained using this design, it had become impossible for a more optimal design to replace QWERTY. Only a very powerful exogenous shock could still unlock keyboard production from the QWERTY design (e.g. Martin & Sunley 2010). If this kind of path dependence were characteristic for economic development then knowledge externalities would not have any impact at all on mature industries, since new knowledge could not lead to any change in product design.

Later Setterfield proposed a conception of path dependence that retains thinking in terms of development stages and equilibria, but that allows the economy to endogenously break out of equilibrium when a new technology obsoletes an earlier technology, making the old dominant design irrelevant and turning the equilibrium outcome into only a temporary one (Setterfield 1997, in Martin & Sunley 2010:70-72). But as Martin and Sunley (2010) argue such a conception of path dependence is still limited in several ways. Firstly it conceives paths as either unrelated to one another, or related only in the limited sense of a successor obsoleting a predecessor. So for example the development of the computer and the mobile phone are two technological paths that follow their own internal logic and do not influence each other. This conflicts with the idea (e.g. in Jacobs 1969) that new technologies can be the product of the interaction among different technological paths. Secondly, and related to the latter point, its prediction that in the long run space is neutral and structures and factor endowments do not exert a significant influence on the development of technologies and industries begs the question why in the real world innovation and production tend to be concentrated in some regions over long periods of time. Based on equilibrium conceptions of path dependence one can reject agglomeration externalities as a sufficient explanation for this, but it is unclear what the alternative explanations might be. Thirdly the use of equilibrium thinking conflicts with the idea of open-ended and continuously ongoing change which is at the heart of evolutionary thinking (whether in Economics or Geography). If spontaneous mutation and ongoing change of selection environments is a typical feature of economies it is hard to see how any equilibrium could maintain itself for a long time. And finally there are empirical reasons to doubt whether the typical paths, coinciding with the product life-cycle, actually describe how technologies develop in the real world, as will be elaborated below. To address these limitations Martin and Sunley (2010) distinguish a third type of path dependency which they go on to develop, namely a non-equilibrium conception.

2.6: Non-equilibrium Path Dependence

In the non-equilibrium conception of path dependence the ideal path from a diversity of unfinished designs, through build-up of critical mass behind one design, and to lock-in and stability, mirroring the ideal product life-cycle pattern, is only one possibility and an unlikely one. This is how a path will develop if no mutations and adaptations take place along the way. Examples that do seem to follow this ideal pattern, as the QWERTY keyboard design mentioned above, are the exception rather than the rule (Martin & Sunley, 2010:77). For example the development path of the computer was fundamentally impacted when the internet emerged, and now it is again moving into unexpected directions because of influences of mobile phone development. For illustration, the development of the internet has led to the concept of ‘cloud computer’ (software runs on external servers, which computers can access through the internet), potentially reversing the course of computers becoming ever more advanced and instead cause computers to lose functionalities which can better be run on an external server. The rise of the mobile phone is already causing the basic design of a computer – the once so dominant twin of desktop and laptop - to split into many grades of portability and at the same time many levels of functionality, with portable but ‘stupid’ computers (with a relatively low processing capacity) becoming highly competitive on the market. In terms of path dependency one path has encountered two different paths, and their interaction led to adaptation, mutation and co-evolution. Also mature products can resurge and demature, rather than forever stay locked-in or passively wait to be obsoleted by a new product (Martin & Sunley 2010:77).

Martin & Sunley (2010) argue that while knowledge spill-overs and other agglomeration externalities can be the critical mass that causes a path to lock in, these same forces can also lead to adaptation and the creation of new variety (p.75) as in the case of the technological trajectory of the computer splitting into many sub-trajectories under the influence of the internet and the mobile phone. Knowledge spill-overs as forces of lock-in are not much more than cost savings in collaborated learning in that they facilitate the production of existing products, while knowledge spill-overs as forces that keep development paths away from lock-in are much more like the encounters of ideas leading to Neue Kombinationen that are so fundamental to Evolutionary Economic Geography. This makes the basic model of the product life-cycle, as in the David and Setterfield conceptions of path dependency, much more complex since many interactions (between firms, between firms and universities, between firms and other institutions, etc; for example see Ponds et al. (2010) for an elaboration of possible interactions, also showing the additional factors that further complicate the model) can cause a path to move in new and unexpected directions , keeping it far from anything resembling an equilibrium at all except a few odd cases like the QWERTY keyboard. In other words knowledge spill-overs from existing structures and factor endowments are back in the model, not as a marginal factor but as the driving force behind evolutionary change. At the same time the role of neutral space in the model is diminished, and perhaps even eliminated in all but the odd case. If most paths stay out of lock-in altogether then the model never reaches the point where neutral space can be assumed since paths become long and unbroken rather than distinct cycles with a clear beginning and end. While paths may adapt and mutate beyond recognition (e.g. compare the first designs of the telephone with the current use of internet telephony, which is not much more than a piece of software combined with the built-in sound devices of a computer) they can still be conceived of as the same long-winding path. If this point is accepted then the basic idea of path dependence, namely that earlier stages of the same path influence its own future development, means that nearly all paths we observe in the economic landscape are continuously being influenced by their own and each other's history. Paths with a distinct beginning (that is, not resulting from the gradual mutation of an earlier design) or a distinct ending (the path being wiped out rather than being able to adapt into new forms) become only one possibility. And whether a path is a mutated version of an earlier path or a newly emerging path, in both cases structures and factor endowments will significantly influence their development.

2.7: Path dependence in Economic Geography: WLO and WTO

While path dependence has mostly been applied in the field of Evolutionary Economics, it has already exerted some influence on Economic Geography as well in the form of the windows of locational opportunity (WLO) model (Storper & Walker, 1989, Boschma, 1997). The WLO model allows the spatial implications of technological paths to be modeled and has so far been tied closely to the product life-cycle. Here it will be argued that the WLO model is also compatible with non-equilibrium conceptions of path dependence, and therefore compatible with models of agglomeration externalities.

In the most basic form of the WLO model a new technology first gives rise to many competing designs being developed in many different locations, and since in the next stage of the product life cycle a dominant design is selected more or less at random the location where the new technology is developed into a full-fledged industry is also selected more or less at random. Using the WLO concept the new technology has created a window of locational opportunity, allowing many regions (except those that lack the most basic generic infrastructure of industrial production and product development) the opportunity to enter the competition for hosting the resulting industry. As the product reaches a mature phase and innovation becomes limited to process innovation the windows of locational opportunity gradually close as the region producing the dominant design builds up a critical mass through agglomeration externalities, until the windows have closed altogether and the system is in lock-in and equilibrium. Only the development of a new technology can break the lock-in and re-open the windows of locational opportunity (Martin & Sunley 2010: 66-67).

While Martin and Sunley reject the WLO model as being incompatible with a non-equilibrium conception of path dependence this is an overly limiting interpretation of the model. Boschma (2007) argues that depending on the technological path in question the windows can be less than fully opened in the initial stage if that path can benefit more strongly from factor endowments than the 'standard' path described in the product life-cycle. For the same reason some paths can be more strongly influenced by factor endowments than others in the mature stage of agglomeration and lock-in. This moves the model closer to being compatible with non-equilibrium path dependence, and the WLO model can be made fully compatible if it is tied not to the standard product life-cycle but to a more general conception of technological paths: windows of technological opportunity (WTO).

While the windows of locational opportunity model describes the amount of freedom of a path to locate in geographical space, an additional concept of windows of technological opportunity (WTO) can be added that describes the freedom of a path to locate in technological space. Technological space basically maps the proximity between technologies in terms of the level of similarity of their products and of the routines and equipment required to produce these products. To return to the example of the QWERTY keyboard, this path is clearly fixed in technological space with very little room to change its design. In other words its windows of technological opportunity are closed. As an illustration one could imagine a hypothetical computer keyboard industry dominated by a firm which patented the QWERTY design. On the one hand this firm would be unable to change its keyboard design even if it wanted to (since the WTO are closed) and on the other hand this industry would be completely concentrated in one geographical location since new entry to the market is impossible hence its windows of locational opportunity are also closed.

At this point the relationship between the product life-cycle, equilibrium and non-equilibrium path dependence and the WLO will be much more clear. According to the equilibrium conceptions of path dependence the WTO of technological paths always follows the familiar pattern of the product life-cycle, being completely open in the early phase and gradually closing towards the mature phase of the path after which the path is locked in with completely closed WTO. The windows of locational opportunity would precisely follow the pattern of the WTO and hence also move from completely open through a gradual closure and finally completely closed windows of locational opportunity. However the non-equilibrium conception of path dependence would predict that the WTO are never completely opened since previous paths create constraints on the development of new paths. Similarly while the WTO might go through periods of closure (e.g. the telephone converging towards a standard design in a period of little innovation), the WTO are likely to be thrown open again whenever interactions with new paths create new opportunities for innovation (the availability of wireless networks affording mobile phones to replace the standard fixed-line phone; the internet allowing the creation of internet telephony). The WLO are again likely to follow the patterns of the WTO, which in the case of non-equilibrium path dependence means that they can go through periods of closure and opening. The application of the WLO model to non-equilibrium path dependence creates many more questions, which will be left for further research.

2.8: Product differentiation and related variety

Recent developments in Evolutionary Economic Geography have created more opportunities for moving Economic Geography in the direction of non-equilibrium path dependence, conceiving of paths as being influenced by co-located paths and historical paths. In Frenken & Boschma (2007), Boschma & Frenken (2009) and Neffke et al. (2011) a model of regional growth as a branching process is proposed. The core of this model is that firms have to continuously diversify their production in order to keep growing, and that this growth proceeds as a branching out towards products related to what the firm is already producing. In the model continuous diversification is necessary because products are characterized by a limited marginal growth of consumer demand and efficient production scale, causing firm growth to level off when the growth potential of their product is reached. The only way to avoid stagnation of firm growth is for that firm to diversify to new products that still have growth potential and may offer temporary monopoly profits. In Evolutionary Economic Geography production of any good requires the firm to acquire the necessary skills in the form of routines. Some of these routines are firm specific and can be applied to production in general (e.g. being skilled at human resource management, or knowing how to talk with customers), but other routines are only useful for the production of a specific good (e.g. understanding how to use child-friendly materials is vital for a toy manufacturer but will have very little value if that firm diversifies into producing cars). Diversification is therefore costly in terms of learning new product specific routines and losing the sunk investments in the product specific routines that are abandoned in the process of creative destruction. Besides being costly learning new routines is also likely to involve the uncertainty of possible failure, and some new routines may be so costly and uncertain that it is practically impossible to learn them (e.g. a toy factory learning to produce computer chips). However diversifying into products related to the products the firm can already produce is likely to be less costly and uncertain, and may involve less destruction of sunk investments. These are all reasons for rational actors to choose diversification to a related product. Under the assumption of bounded rationality firms can also have non-rational reasons for making this choice. For example firms will be more aware of market opportunities for related products, and are likely to have more confidence that these products can be profitable. And even in the face of promising options in different product markets, unwritten rules and business cultures can make it seem completely obvious for firm managers that they should only consider producing products as close as possible to what they have produced in the past (the adage of 'staying close to your core competences' comes to mind) even though they may not be able to explain why exactly this is the best firm strategy. So far the argument for diversification to related products has been mostly about branching to related products being more likely to be attempted and more likely to succeed because it is less costly and uncertain. Besides this the new product a firm wants to diversify to may itself be the result of a recombination created when two related ideas came together through knowledge spill-overs (Neffke et al., 2011), for example in the form of an unexpected serendipitous outcome of bringing two ideas together as a Neue Kombination as in Schumpeter's work. The combination of two unrelated ideas can also lead to innovative ideas, and the less related the ideas are the more radical the resulting idea is likely to be. However at some point new ideas will end up being too radical, requiring too costly and uncertain investments in learning totally new skills, and simply appearing too far-fetched to the actors involved. The number of ideas for new product development that remain after discarding (what appear to the actors involved) the impractical ones is expected to be higher in the case of recombinations from related variety. In short the technological relatedness between products is the missing link in building an evolutionary model of agglomeration externalities.

The concept of relatedness can be explained very concisely using the related concept of cognitive proximity, developed by Nooteboom (e.g. Nooteboom, 2000). Two persons or firms can learn new things from each other provided that they do not have an identical stock of knowledge, that is if there is some cognitive distance between them. But if the cognitive distance is too big they are likely to use different concepts and methods, and overcoming these differences may be so time- and energy-consuming that it is practically impossible. So whenever actors communicate in order to learn something there is an optimal level of cognitive proximity at which most learning can take place (Boschma & Frenken, 2009). For example a bicycle producer and the first producers of automobiles created different products, using different product-specific routines. But some product-specific routines were useful in both production processes, for example the earliest cars may have had tires, gears and breaks quite similar to that of a bicycle. And the first automobile producers may also have found much to learn in the firm-specific routines of bicycle producers as for example they could learn how to create a market for a new mode of transportation, since both products were probably first perceived as luxury items and only later became daily necessities for the mass consumer (for a study applying the concept of relatedness on inter-industry interactions between the UK bicycle industry and the emerging automobile industry see Boschma & Wenting, 2007). This example also illustrates another point, namely that relatedness has to be studied in a dynamic way (Boschma & Frenken, 2009:12). Since the early days the automobile industry has changed dramatically, since new materials are used and production has ofcourse become much more specialized and sophisticated. In other words the cognitive distance between the bicycle and automobile industries has increased so much that the same industries that used to be related probably cannot be regarded as related industries anymore, and learning opportunities between firms in these industries have probably become very limited.

In terms of path dependence, what is emerging is an Evolutionary Economic Geography model of regional factor endowments laid down by previous paths and by co-located contemporary paths consisting of a related variety increasing the number of new paths that are conceived of, the number of paths that will actually be attempted by the firms involved and the number of paths that successfully develop (Neffke et al., 2011). The implication of this is that an evolutionary model can depart from the assumption of neutral space as the starting point of a model's dynamics and accept agglomeration externalities as the key connection between successive paths and as the engine of new path emergence.

And while growth as a branching process as proposed in Frenken & Boschma 2007 may originally have been conceived as one Setterfield-type path coming to an end and a new one (in a related industrial sector) being started by the same firm, it can also be conceived of in terms of continuously ongoing adaptation and non-equilibrium path dependency. The new combinations resulting from knowledge spill-overs bringing together related but not yet connected ideas, such as the internet combining with the computer to lead to the concept of cloud computing, can be sufficient to cause a technological trajectory to continue to evolve and stay out of equilibrium. So factor endowments as related variety serve three interlinked purposes, namely to make path branching to new sectors (product diversification) more likely to be conceived of because of knowledge spill-overs from related variety, secondly causing these new paths (products) to be less costly and therefore more likely to succeed because of related variety as factor endowments supporting the growth of related paths, and thirdly to increase the chance that paths can adapt and mutate and thereby reduce the chance of the path ever coming to an end altogether. When these three roles of related variety are combined it can form the core of a non-equilibrium Evolutionary Economic Geography model of path dependence. Aggregated to the regional level related variety should allow a region's population of firms to keep growing in the long run, resulting in higher long-term growth and employment in that region.

At this point a more complete answer can be proposed to the challenge of finding the industrial structure (specialization or diversity) which promotes long-run innovation and growth. Specialized factor endowments should not make any significant contribution to the learning processes and recombination of ideas that drives regional economies, and may only lead to cost savings that become insignificant in models in which innovation is the driving force. Regional diversity is more compatible with a growth model based on non-equilibrium path dependency, but it should be outperformed by a special type of diversity namely related variety. Agglomeration externalities from a structure of related variety should promote not only higher growth rates over periods of a few years, but also in the very long-term of decades and more.

2.9: Mechanisms of agglomeration externalities

So in the model of firm growth as a branching process firms continually have to diversify in order to keep growing. Firms are most likely to diversify into related products rather than products totally new to them since it will be easier for them to bridge the congitive distance and acquire the necessary product-specific routines to be able to produce the new product. Using these evolutionary concepts, and the insights on pecuniary externalities and the porftolio effect discussed in section 2.2 we can now draw a more complete picture of the precise transfer mechanisms for bringing together related ideas leading to recombinations, and for firms accessing the benefits of related variety when they diversify their products.

Regional economic performance can be divided into high productivity (associated with profitable and competitive firms, and high wages), high employment and low unemployment, each being sustainable in the long run (which are also the key targets of the *Europe 2020* strategy, European Commission, 2010). Agglomeration externalities can contribute to these aims of regional economic performance in three ways. Firstly pecuniary externalities can help firms save costs, leading to higher productivity in the short term but not in the long term when structural change becomes a dominant factor. Secondly knowledge externalities can lead to process innovation (more efficient production of the same product) and product innovation (the development of enhanced or new products). Process innovation is associated with productivity growth in the short term but not in the long term. It is not associated with employment growth, and may even lead to a reduction in employment if new production techniques reduce the demand for labor. Product innovation is associated with employment growth, since new products tend to be more labor intensive than mature products (which firms can afford since new products are associated with temporary monopoly profits). Product innovation is also associated with productivity growth, but more in the long term (it prepares firms to deal with structural change) than in the short term (when the new products are still produced inefficiently and new markets are yet to be fully exploited). Thirdly portfolio effects can reduce the damage dealt by cyclical shocks to a regional economy, helping to avoid long-term unemployment. The argument here is that while short, shallow recessions (e.g. a diverse region faces a downturn in one of its industries) tend to cause only temporary unemployment, longer and deeper recessions (e.g. a specialized region losing its key industry) are more likely to cause workers to become long-term unemployed and perhaps unemployable by the time new jobs are available in the region. Since this research is interested in long-term economic performance, those agglomeration externalities associated only with short-term effects (pecuniary externalities, portfolio effect and knowledge externalities leading to process innovation) are not elaborated further. As described in section 2.2 urbanization externalities are also regarded as nothing more than pecuniary externalities when the effects of related and unrelated variety are disentangled from it. Moreover since density of economic activity leads to both cost savings (e.g. shared infrastructure) and congestion, it is unclear whether urbanization externalities lead to net cost savings or higher costs when the effects of related and unrelated variety have already been accounted for. The remainder of this section will elaborate the mechanisms of knowledge spill-overs leading to product innovation.

Product innovation (leading to product differentiation) on the level of regions can take place both through product differentiation carried out by existing firms or through new firm start-ups which enter the market with a new product. In the case of existing firms product differentiation can take place in two ways: either through radical innovation or through a series of incremental innovations sustained over a period of time, gradually moving the firm's production to a distinctly new product. high levels of either radical or sustained incremental innovation (or both) can lead to the process of product differentiation which in the evolutionary perspective is expected to lead to employment and long-term productivity gains. In the former case this takes place by the starting up of new product life-cycles, while in the case of sustained incremental innovation it takes place through keeping existing product life-cycles out of lock-in ("out of equilibrium").

Both product innovation by a new firm start-up and radical innovation by existing firms requires a recombination of ideas, a Neue Kombination. It also requires firms to be able to acquire the required product-specific routines to produce and market the new product. In the case of a new firm start-up it also requires an entrepreneur and a regional economy with the necessary incentives and facilities (or at least the absence of barriers to entry) to make new firm start-up feasible. Product differentiation through sustained incremental innovation only requires firms to be able to acquire new product-specific routines as they start to face new challenges of production and marketing of a gradually changing product.

Regions with a sector structure of related variety are most likely to possess all 'inputs' for the process of product differentiation. Regions with a structure of unrelated variety are less likely to offer an environment favorable to product differentiation, while regions with a strong specialization in one or a few products are unlikely to possess any of the 'inputs' needed for product differentiation to occur, as will be described below.

Neue Kombinationen can occur either serendipitously through informal contacts among workers of different firms, or through formal contacts among firms engaged in projects aimed at collaborated knowledge production. However the more radical a recombination of ideas is the less likely it is that firms can conceive of the recombination in advance, plan the recombination to happen and develop it from scratch in a collaborated innovation project, which makes informal contacts a more likely source of Neue Kombinationen than formal contacts. Both formal and informal contacts are more likely to result in fruitful new combinations if there is some but not too much cognitive distance between the interacting workers (see Neffke et al. 2011:12 for a discussion of evidence), which is most likely to be true if the firms involved are located in a region with a structure of related variety. In regions with a structure of unrelated variety it is more likely that actors with a large cognitive distance interact with each other. If recombinations result from these interactions they are likely to be much more radical than those resulting from actors with a moderate cognitive distance (e.g. if the interaction between a car manufacturer and a toy producer leads to any new product concepts at all, it is likely to be a rather radical and far-fetched idea). The number of recombinations perceived as feasible, the number that is actually attempted to be carried out, and the number that eventually result in a succesful product are likely to be much lower in the case of unrelated variety than for related variety. Finally regions with a specialized sector structure are only likely to facilitate interactions among actors with little or no cognitive distance, which is unlikely to result in any Neue Kombinationen.

The second input for product differentiation is the ability of firms to acquire product-specific routines. As described in section 2.3 routines are a form of tacit knowledge which is embodied in people and cannot be acquired by observing somebody's behavior or learned from a book. So when firms want to produce a new product they have to gain access to workers who already possess the same or similar routines as the ones required for the production of this new product. Firms can do this by either hiring workers with relevant product-specific routines, or merging with or acquiring a firm which already has access to workers with these routines. Since firms are more likely to differentiate towards a product related to their current production, the workers they require to produce their new product are more likely to be those with related product-specific routines (Boschma et al. 2009 provides evidence that hiring workers with related skills contributes more to firm performance than hiring workers with very similar or unrelated skills; Neffke et al. 2011:10 discusses evidence that mergers and acquisitions between technologically related firms results in higher innovative performance than M&A's between unrelated or technologically similar firms). These are again most likely to be available in regions with a sector structure of related variety, less likely in regions with unrelated variety and least likely in specialized regions (in the extreme case of a region fully specialized in one product it is by definition impossible for firms to have access to workers with the product-specific routines to produce a new product).

Finally the 'inputs' related to entrepreneurship are relevant in the case of new products introduced by new firm start-ups. These factors are largely not an issue of regional sector structure but more of having local formal and informal institutions which encourage and facilitate entrepreneurship, e.g. norms which do not discourage risk taking and the availability of investment capital for risky start-ups. But a sector structure of related variety can also make a contribution to this 'input' of product differentiation: if the entrepreneur of a start-up firm has gained experience by working in a firm that produces a product related to the one produced by the new firm, this can increase the quality of the start-up firm and increase its chance of survival (Klepper, 2007; Boschma & Wenting, 2007). The previous experience of entrepreneurs is less likely to be helpful in the case of unrelated products.

2.10: The geographical scales of agglomeration externalities

Based on these detailed mechanisms of knowledge externalities we can also be more specific on the geographical scale on which these mechanisms are expected to work. The mechanisms described above require either the ability for workers belonging to different firms to have regular meetings (formal and informal contacts leading to recombinations of ideas; a merger or acquisition in which workers are not forced to relocate to the new head-office), or the ability to hire workers or make the workers from a merged or acquired firm relocate to the new head-office. Both are very likely to be facilitated by geographical proximity, but to a different extent. Regular informal meetings are likely to require the highest degree of geographical proximity since they depend on regular meetings that occur by chance. The probability for this to occur can be expected to decline rapidly with distance (but other forms of proximity may compensate for a lack of geographical proximity, see Boschma 2005). Being able to hire workers with relevant skills is also expected to require a fairly high degree of geographical proximity. If workers already live within commuting distance from the firm that wants to hire them then this should make it much easier to hire these workers. However if workers live beyond commuting distance they could only be hired if they are willing to move house, which the workers may be unwilling to do (or only if they receive a compensation for it). And finally the likelihood that a firm is able to hire workers at a very high distance, and / or beyond a national border, is very low since workers will be unlikely to be willing to move house. The necessity for geographical proximity may be lowest in the case of regular formal meetings. Since this kind of meetings does not depend on chance but are planned and organized to occur, the probability that they occur does not decline with distance. However regular meetings at long distances still involve high travel costs. On the other hand, as described above formal meetings are expected to be the least important factor in facilitating product differentiation, so the most important factors are also those that require the highest level of geographical proximity.

While the portfolio effect and pecuniary externalities are not the main focus of this research, they should also be associated with a geographical scale in order to test them against knowledge externalities in a single model. Pecuniary externalities may work at a lower level of geographical proximity than knowledge externalities. Of the two Marshallian mechanisms of pecuniary externalities only the shared pool of skilled labor requires a high level of geographical proximity since it depends on firms being able to hire workers, which should to a large extent depend on whether they live within commuting distance of the firm. The other mechanism, namely shared intermediate producers and service providers, should mainly be limited by transportation costs, which can be expected to impose a less severe need for geographical proximity than the mechanisms of knowledge externalities. The portfolio effect, which connects a diverse economy to a reduced risk of high unemployment due to asymmetrical cyclical shocks, should be expected to work mostly through the labor market. In a diverse economy the unemployment caused by a shock to one economic sector can be compensated by continued hiring of workers by unaffected sectors. However if firms are not located in the same labor market area it becomes less likely that unemployed workers in one sector and employment opportunities in another sector would become connected, while at a high geographical distance there should not be any noticeable portfolio effect. In a Keynesian framework the portfolio effect can also work through consumer demand. If an asymmetrical cyclical shock reduces the employment and thereby the consumption by workers in one sector, the maintained employment and therefore consumption by workers in an unaffected sector should keep consumer demand at a normal level and prevent increased unemployment in local retail and service providers due to decreased local demand. Again this mechanism requires geographical proximity to the extent that workers in the affected and unaffected sector should visit the same local shops and service providers. Further research could explore the mechanisms of the portfolio effect and pecuniary externalities, and the geographical scale at which these mechanisms are expected to work in much more detail.

2.11: Hypotheses

Having proposed a model of the mechanisms of agglomeration externalities from an evolutionary perspective with their associated temporal and spatial scales, and having connected this with regional sector structure and the resulting economic performance, a number of hypotheses can now be proposed:

> Hypothesis 1: *Regions with a sector structure of related variety experience an increased rate of product innovation, which leads to higher employment on the short run and to both higher employment and higher productivity in the long run*
>
> Hypothesis 2: *Regions with a sector structure of unrelated variety experience less job losses from asymmetric shocks which leads to lower unemployment, more so in the long run than in the short run*
>
> Hypothesis 3: *Regions with a sector structure of specialization experience an increased rate of process innovation and reduced production costs which leads to higher productivity, more so in the short run than in the long run. To the extent that process innovation is labor saving, it will lead to lower employment in both the short and long run.*

When all these expected effects of sector structure are accounted for it is expected that the density of economic activity will still have an independent effect on economic performance. However since it is unclear from the model whether these urbanization externalities would be a net positive or negative effect, and which aspects of economic performance would be affected by it, no hypothesis will be tested for urbanization externalities in this research.

As Frenken et al (2007) suggested there is a composition effect between the sector structure variables. While a perfectly specialized region by definition cannot have any variety, a region specialized in a few large sectors can experience some of the effects of related variety or unrelated variety, depending on the cognitive proximity between these large sectors. In the same way a region with a structure of related or unrelated variety may also harbor some relatively large sectors which experience to some extent the benefits and costs stemming from specialization.

The next chapter introduces the empirical model and database used to test the theoretical model proposed above.

3. Empirical model

In the previous chapter a theoretical model of agglomeration externalities was described and summarized in three key hypotheses. This chapter first describes the empirical model and then the data used to put these hypotheses to the test.

3.1: Operationalized hypotheses

The hypotheses described in section 2.11 connect the sector structure of regions with the economic performance of those regions in terms of employment and productivity. For unrelated variety (hypothesis 2) this is a direct link: it is expected to prevent high unemployment due to asymmetrical shocks to the regional economy. In the case of related variety (hypothesis 1) and specialization (hypothesis 3) the link is indirect: these types of sector structure are expected to lead to higher rates of product innovation and process innovation respectively. However as described in section 2.4 it is very difficult to measure innovation in a way that is consistent with the evolutionary perspective on economic development. Therefore in the empirical model only the expected indirect effects of related variety and specialization will be measured. Earlier research (Frenken et al. 2007) has taken the same approach.

The hypotheses distinguish different effects at different time-scales, but in this research it will not yet be possible to make this distinction empirically. Related variety is expected to lead to higher employment in the short term, and to both higher employment and higher productivity when structural change has taken place in the economy and the full importance of product differentiation becomes visible. Specialization on the other hand is expected to lead to higher productivity and lower employment in the short run, but lower employment and lower overall levels of production in the long run when structural change makes specialization in a limited range of products a problematic strategy. The data available allows the dependent variables to be measured over a period of ten years, which is arguably not long enough to fully notice the effects of structural change. Therefore the empirical model is expected to mostly capture the short-run effects of regional sector structure, which are higher employment in the case of related variety and higher productivity in the case of specialization.

Since earlier studies suggest a composition effect between the different forms of sector structure, this study will follow the established approach of testing related variety, unrelated variety and specialization simultaneously. The operationalized hypotheses are therefore as follows:

> Hypothesis 4: *In the short run employment growth is positively related to related variety, negatively related to specialization*
>
> Hypothesis 5: *In the short run labor productivity is positively related to specialization*
>
> Hypothesis 6: *In the short run unemployment growth is negatively related to unrelated variety*

3.2: Data overview

The data used for this research stems from two sources. For the dependent variables and most control variables a database compiled by the consultancy firm Cambridge Econometrics is used. This data largely coincides with the data available at Eurostat, but missing values and breaks in time series that plague the Eurostat data are absent in the Cambridge Econometrics data. For the main predictors, namely related variety, unrelated variety and specialization, the Amadeus database compiled by the consultancy Bureau van Dijk was used, as will be described in detail below.

After selecting those countries and regions for which data for all independents and dependents is available, the resulting database covers a total of 19 EU countries at the regional scale (table 1). As explained below two versions of the main predictors (related variety, unrelated variety, specialization) have been calculated, one version based on the absolute number of firms per sector per region and one version weighted by the operating revenue of firms. While the unweighted variables can be calculated for all 19 countries, the weighted variables can only be calculated for 16 countries due to issues of data quality. Of the 19 countries included in this research 14 are EU15 countries ('old' member states) while 5 are New Member States in East-Central Europe. In other words for the EU15 the database is complete except for Luxembourg, while half of the New Member States are included. Of the countries missing in the weighted analysis two are EU15 countries (Austria and Greece) and one is a New Member State (Slovenia).

Table 1: countries included in the database, unweighted and weighted main predictor variables

	Unweighted	*Weighted by revenue*
EU15:	14	12
Austria	x	
Belgium	x	x
Denmark	x	x
Finland	x	x
France	x	x
Germany	x	x
Greece	x	
Ireland	x	x
Italy	x	x
Portugal	x	x
The Netherlands	x	x
Spain	x	x
Sweden	x	x
United Kingdom	x	x
New Member States:	5	4
Czech Republic	x	x
Hungary	x	x
Poland	x	x
Slovenia	x	
Slowakia	x	x

The database is limited to regions in continental Europe. This means that French, Portuguese and Spanish overseas territories are not included due to data unavailability.

3.3: Geographical scale

As discussed in section 2.10 there are theoretical reasons to expect that the different types of sector structure work best at different spatial scales. So while related variety is expected to work at very short geographical scales (close enough for frequent chance interactions to occur), unrelated variety and specialization is expected to also function at larger distances. As described in more detail in section 2.10 unrelated variety (based on the portfolio effect) is linked to the ability of unemployed workers to find new jobs within commuting distance, while specialization is about firms being located close enough to each other to use shared inputs and resources. So while the scale of NUTS3 regions may be most suitable for detecting the effects of related variety, the distances within NUTS2 regions may be more suitable in the case of unrelated variety and specialization.

Data unavailability makes an analysis at the NUTS3 scale impossible, so whenever possible NUTS2 regions are used. The only exception are the regions in Belgium for which only data at the scale of NUTS1 regions is available, producing a total of 234 regions (210 regions when weighted sector structure variables are used). As a consequence of the choice of geographical scale the analysis may be expected to underestimate the effects of related variety while unrelated variety and specialization is expected to be measured to a more complete extent in this database.

3.4: Dependents

All dependent variables are supplied by the Cambridge Econometrics database. First data for the period 1998 to 2008 is tested, and then several alternative time spans are used to test the robustness of the models. The period is broken up in two intervals (1998-2003 and 2003-2008) to see if particular circumstances in these time periods have a strong influence on the models. And besides the 2003-2008 period the dependents are also calculated for 2003-2007 to test for the influence of the 2008 financial crisis and following recession

A cursory glance on the data for employment growth (figure 1 and 2), labor productivity growth (figure 3 and 4) and unemployment growth (figure 7 and 8) reveals that there are great differences between regions in the EU15 countries and regions in the New Member States. In 1998-2003 these regions in East-Central Europe – with a few exceptions – score very low on employment growth, and very high on labor productivity growth and unemployment growth. In 2003-2008 the picture is reversed for employment growth and unemployment growth from very low to very high growth figures, while labor productivity growth remains exceptionally high in most of these regions.

The most likely explanation for this is that the dependent variables are impacted strongly by the transformations which the New Member States have gone through in the period of study. The often painful process of post-communist transition in the 1990s, characterized by the privatization or closure of state-owned enterprises and an overhaul of trade and industrial specialization patterns, is likely to have left a mark on the regional economic performance in the 1998-2003 period. Subsequently the 2003-2008 period includes the EU accession and economic integration of the New Member States with the EU.

The hypotheses on the impact of sector structure on regional economic performance are about structural differences between regions, not about explaining particular processes taking place in one specific period of time. The events that happened in the New Member States between 1998 and 2008 are precisely such unique historical processes. While they are mostly unrelated to structural characteristics of regions themselves, they are likely to have a strong but temporary impact on the economic performance of these regions.

Two methods are tried to correct for this strong influence on the analysis. Firstly the initial level of the dependent variable is added to all dependents that measure growth. So in the case of employment growth it becomes a factor of initial employment besides the other independent variables. This has the additional advantage that the strength of convergence or divergence of economic performance can be measured. Secondly a dummy for belonging to the New Member States is added to the analysis to see if this controls for the special circumstances faced by the regions in these countries. This results in the following dependent variables:

- *EMPLOYMENT GROWTH*
- *LABOR PRODUCTIVITY GROWTH*
- *UNEMPLOYMENT GROWTH*

In the case of employment and labor productivity both a standard measure and a measure corrected for differences in the average number of hours worked are calculated. The reason is that there are significant differences in the number of hours in the average working week, partly due to the fact that part-time work is more common in some countries than in others. Over the 1998-2008 period values range from about 31 hours per week in Dutch regions, to about 41 to 45 hours in Greek regions. According to Gardiner et al. (2004) this may have a significant impact on comparative studies at the EU level.

Figures 1 through 10 show the dependent variables over the periods 1998-2003 and 2003-2008. Maps of alternative measures and time periods of the dependent variables are provided in appendix 1. Note that the maps show relative differences between regions in five equal categories of 20% of cases each. So for example the highest category on every map contains the 20% highest scoring regions for that variable. The absolute values associated with these categories change over time, so a region can have the same absolute score in the two time periods but a different relative position compared to other regions.

The maps of employment growth (figure 1 and 2) show great differences in the EU. All values are calculated as “new divided by old”, so values above 1 indicate growth while values below 1 show decline. In the period 1998-2003 much of East-Central Europe was still characterized by a large contraction of employment and the same was true for regions in the former DDR. Most regions in the Netherlands and many in the UK also saw declines in employment, perhaps due to the recession that hit parts of Europe in the early 2000s. In the 2003-2008 period however most of Europe saw gains in employment. This time East-Central Europe (with the exception of Hungary) showed the biggest gains, while regions with declines in employment were mostly limited to parts of Portugal, France and Greece.

Figure 1: Employment Growth 1998-2003

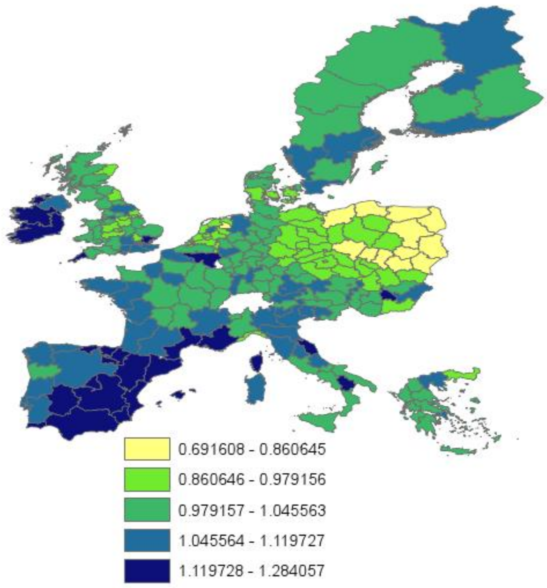
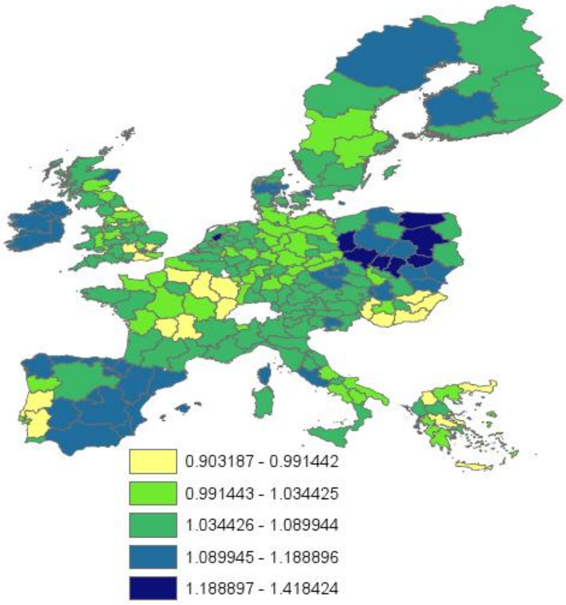


Figure 2: Employment Growth 2003-2008



These differences are not due to the 2008 financial crisis since the differences between 2003-2008 and 2003-2007 are very small (see figure A1 through A3 in appendix A for maps of the dependent variables in 2003-2007).

In both 1998-2003 and 2003-2008 Irish and Spanish regions showed the greatest gains in employment. With hindsight this was probably partly due to the real estate bubbles that took place most strongly in these countries, which led to large but temporary rises in employment in sectors such as construction.

Figure 3: Labor Productivity Growth 1998-2003

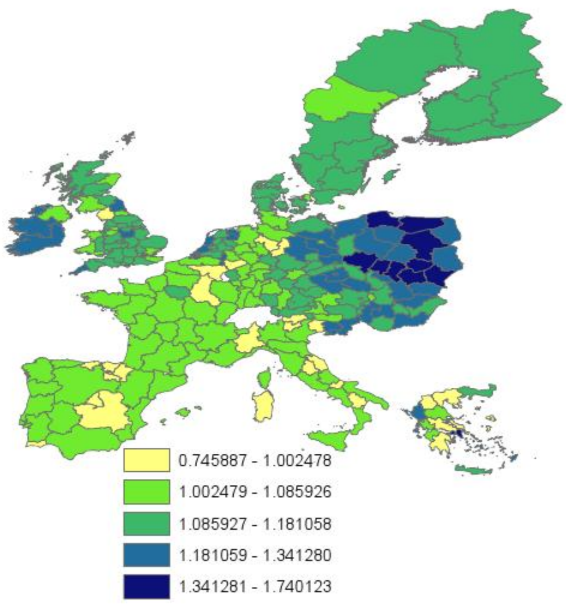


Figure 4: Labor Productivity Growth 2003-2008

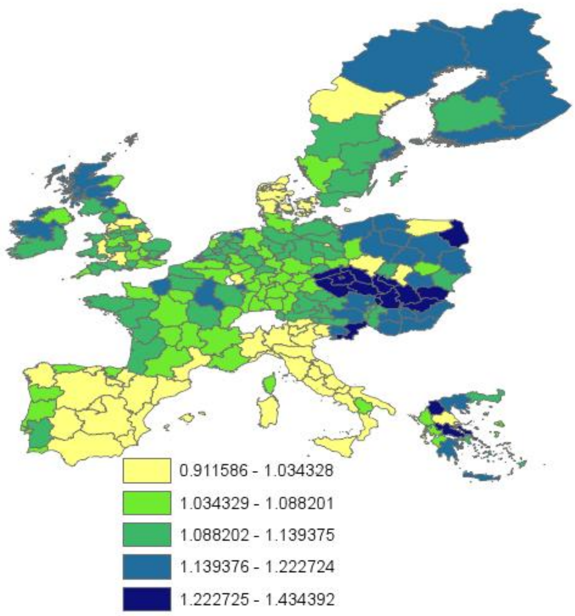


Figure 5: Average Labor Productivity 1998-2003

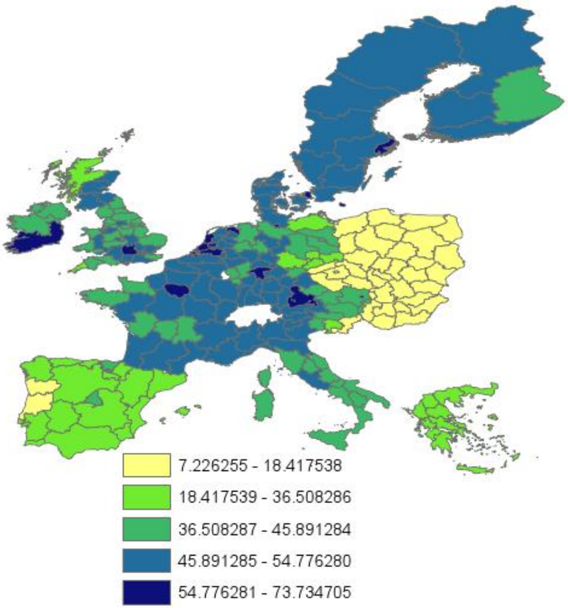
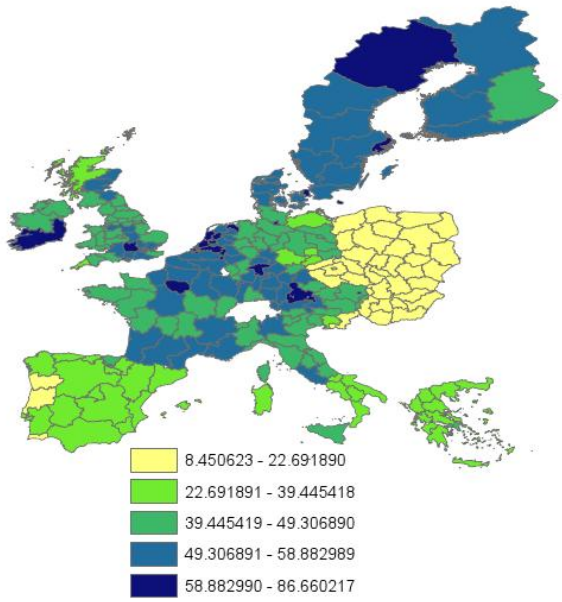


Figure 6: Average Labor Productivity 2003-2008



A final issue is that employment growth is not only determined by the ability of the regional economy to create jobs, but also by population growth. This will be controlled for by adding population growth as a control variable for all models using employment growth as dependent variable.

Differences between regions in labor productivity growth (figure 3 and 4) are quite pronounced, although again this seems partly explained by national-scale processes. The New Member States again show significantly higher growth rates in most regions, with growth in Poland appearing to set in earlier than growth in the Czech Republic and Slowakia. Spanish, Italian and Danish labor productivity growth is relatively limited in the 2003-2008 period, although again it should be noted that this does not necessarily mean that their growth in absolute values was lower since the maps are only intended to show relative differences.

For labor productivity growth it is meaningful to include both growth rates and average labor productivity values (figure 5 and 6) in the analysis. A comparison of the maps shows the importance of testing both measures of labor productivity. It turns out high growth regions are mostly growing from a low level of labor productivity, while low growth regions were mostly already at a high level. Any growth that can be attributed to a temporary process of catch-up should be treated as noise rather than information in this analysis, since it cannot be explained by the structural characteristics of regions. The data in figure 5 and 6 is in thousands of euros of added value per employee, so while on average workers in the New Member States regions included in this analysis created less than 23.000 euro of value in 2003-2008 per worker, workers in the best performing EU15 regions produced up to about 86.600 euro in value per worker. While these absolute values are different in the two time periods, the relative differences have stayed nearly the same regardless of the above average catch-up growth in New Member States regions.

Finally the data on unemployment (figure 7 through 10) follows by now familiar patterns. Unemployment growth (figure 7 and 8) in the New Member States starts high in 1998-2003 but changes into a decline of unemployment in 2003-2008 (consistent with the pattern of employment growth). In most EU15 regions unemployment levels are stagnant or decreasing in 1998-2003 (notable exceptions are Portugal and large parts of Germany), while in 2003-2008 Ireland, Spain, large parts of the UK and Italy show strong increases in unemployment. A comparison with figure 9 and 10 shows that in most cases it is again a case of high growth from a low base (unemployment growth in regions with low unemployment rates), and low growth or decline of unemployment from a high unemployment level. Such developments suggest a process of inter-regional convergence, while exceptions to this rule are more likely to be due to the structural characteristics of regions which are the focus of this research.

It may be noticed that the lowest value for unemployment recorded on the maps is 0.0%, which may appear to be an error. This zero value turns out to be due to the Finnish autonomous region of Åland, which according to Cambridge Econometrics had zero unemployed persons throughout the period of study. A report by the Nordic National Statistical Institutes (2010) reports 0.3% unemployment for this region in 2009, which drops to 0.0% for some age cohorts. The region has a prosperous tourist economy and a tiny population of about 28,000 in 2010, while job seekers are deterred by strict controls on immigration to the autonomous region. While definitely an outlier, this region's value does not seem to be a data input error.

An issue that arises from the exploration of the data so far is that some processes seem to be mostly national rather than regional scale developments. In as far as they are caused by national-scale differences they are considered noise that should be controlled for in order to focus on regional-scale dynamics. In order to address this issue the models will also be run with country fixed effects dummies added to the model, which should allow the models to detect what differences between regions are left to be explained when national-scale processes are controlled for.

Figure 7: Unemployment Growth 1998-2003

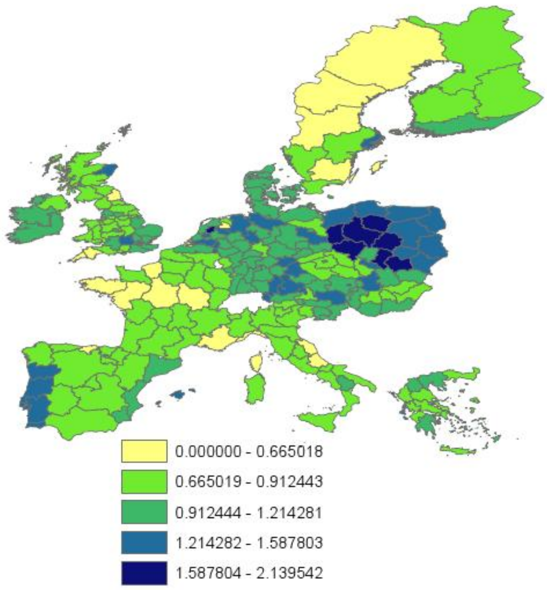


Figure 8: Unemployment Growth 2003-2008

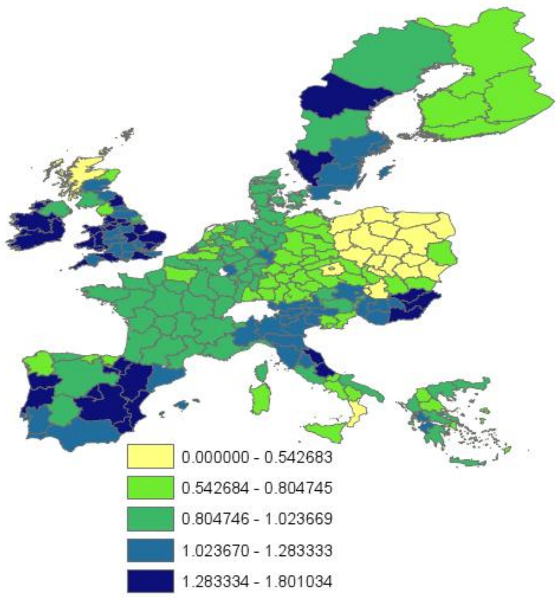
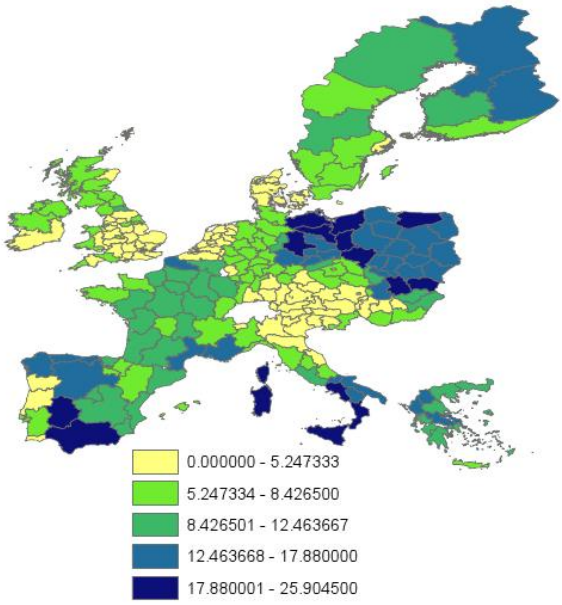
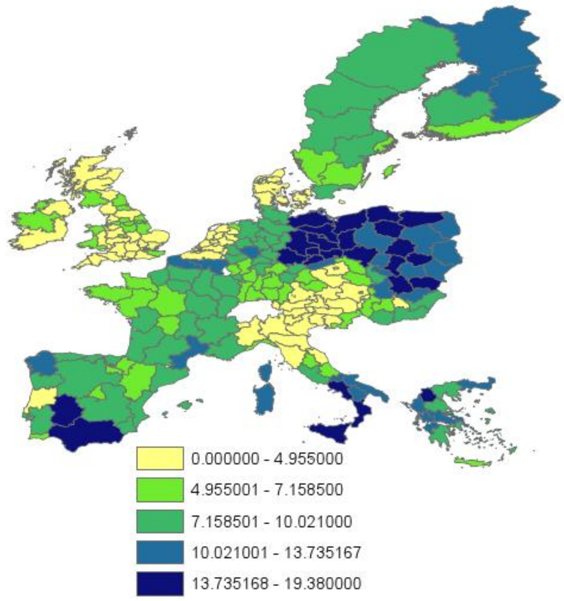


Figure 9: Average Unemployment Rate 1998-2003 *Figure 10: Average Unemployment Rate 2003-2008*





3.5: Independents - main predictors

The main predictors of interest in this research are the sector structure variables related variety, unrelated variety and specialization.

RELATED VARIETY

UNRELATED VARIETY

Since Frenken et al. (2007) is the seminal study on sector structure analysis distinguishing related variety and unrelated variety, their measures for the main predictors using the entropy measure are used. Entropy measures the amount of diversity among the cases in a population. Applied to the sector composition of firms, it measures to what extent the population of firms deviates from a perfectly homogeneous population. If all firms in a region operate in the same sector there is no variety and entropy is zero. Entropy takes the maximum value if a region's firms are spread equally over all sectors. This maximum value depends on the number of sectors included in the analysis, with each additional sector increasing the maximum possible entropy value. In this research the number of sectors included in the calculations of regional entropy equals the total of all sectors present in the EU, making entropy values comparable between regions. Equation 1 shows how entropy E for region J is calculated (adapted from Frenken et al. 2007:689, equation 1):

$$E_j = \sum_{k=1}^{n} \frac{firms_{jk}}{firms_j} \log_2 \left(\frac{1}{firms_{jk} / firms_j} \right) \quad (1)$$

So to calculate entropy for each sector the share of sector K in the number of firms in region J is multiplied by the log base 2 of 1 over the share of sector K in the number of firms in region J, and the resulting sector-scores are summed for region J.

The key advantage of the entropy measure of sectoral variety is that entropy can be decomposed at different levels of sector detail. The data available makes it possible to calculate entropy at 4 levels of sector detail (1 through 4 digit level of the NACE rev.2 industry classification). Because of the decomposable nature of the entropy measure it is possible to calculate how much additional variety is added when entropy is measured at a higher level of sector detail. For example a region may have a lot of variety in 2 digit sectors (e.g. manufacture of food products, manufacture of textiles, computer programming and consultancy) but within these broad sectors it may have very little variety in 4 digit sectors (e.g. among others processing of fish versus processing of meat, potatoes or vegetables within the manufacture of food products sector). In this case the region has a lot of variety of technologically unrelated sectors, but little variety of technologically closely related sectors. In other words the region has a high level of unrelated variety but a low level of related variety, which is expected to protect the region from cyclical unemployment but not expected to lead to higher levels of product innovation. Conversely a region can be found to have little unrelated variety (e.g. only a few broad sectors like manufacture of food products), but a lot of variety of related sectors within these broad sectors. This should allow higher levels of knowledge sharing and recombinations to take place leading to product differentiation, while a portfolio effect protecting against cyclical shocks would be very limited).

Following convention, unrelated variety is measured as entropy at the 2 digit level (broad sectors), while related variety is the difference between 4 digit level (detailed sectors) and 2 digit level entropy (but Frenken at al. 2007 use 5 digit level instead of 4 digit, which is not available in the EU-wide database). A more elaborate discussion of the entropy measure can be found in Frenken et al. (2004).

SPECIALIZATION

Because of data limitations this study cannot follow Frenken et al. (2007) in using the LOS-index for the calculation of specialization. Instead, the Theil index is combined with Location Quotients, to arrive at a measure of specialization which compares regional sector structures with the EU average as proposed by Cutrini (2006). Location Quotients show to what extent a region's sector structure differs from in this case the sector structure of the EU as a whole. If a region has any sectors that are significantly over-represented in that region relative to the EU average this region can be said to be specialized in those sectors. While LQ's are calculated seperately for each sector, the Theil index can be used to arrive at an aggregated score of specialization at the regional level. The result is a score that is the weighted sum of the log of the LQ's of all sectors of a region. It takes the value of 0 when a region's sector structure exactly mirrors the EU average, indicating that the region cannot be said to be specialized in any sector. The more the region's sector structure deviates from the EU average, the higher the value of the Theil index (Cutrini, 2006:8).

Equation 2 shows the calculation of the regional Theil index based on the Location Quotient for region J, where K stands for sector and EU for the EU total (adapted from Citrini, 2006:8, equation 13):

$$T_j = \sum_{k=1}^{n} \frac{firms_{j\,k}}{firms_j} \ln\left(\frac{firms_{j\,k} / firms_j}{firms_{EU\,k} / firms_{EU}} \right) \qquad (2)$$

In other words for each sector the share of sector K in the number of firms in region J is multiplied by the natural log of the Location Quotient (share of sector K in the number of firms in region J, divided by the share of sector K in the number of firms in the EU), and the resulting sector-scores are summed for region J.

Data on sector structure

All main predictors (related variety, unrelated variety and specialization) are calculated from the sector structure of regions in the period 1999-2009. The database used (called Amadeus) is a collection of financial statements made by Bureau van Dijk of all individual firms in most of the EU27 countries (see table 1 for a list of the countries included in this research). The financial statements included in the database for each firm form a sample of all financial statements published by the firm during this ten year period. So for example for firm A the financial statements for the years 1999, 2002, 2005 and 2008 have been included, for firm B the statements for the years 2000, 2003, 2006 and 2009 and so forth. This sample of years was made by the collector of the original database and appears to be a random draw. To avoid double-counting, for this study for each firm only the most recent financial statement was taken from the original database. The result is a complete list of firms which have published at least one financial statement during the 1999-2009 period (n= 9,837,479).

The sector structure variables can be calculated on the absolute number of firms per sector, but can also be weighted by the size of firms. This is likely to give a more accurate picture of a region's sector structure. If the measure does not include weights a region with one very large firm in sector K and several tiny firms in sector L may not seem specialized in sector K, while in reality a majority of the workers and capital in that region are devoted to sector K. Frenken et al. (2007) use sectoral shares of employment, but employment data in Amadeus is very incomplete. Instead operating revenue is used to calculate weighted sector structure variables.

While operating revenue may fluctuate from year to year it can still be expected to give an accurate overall picture of the relative size of firms, and therefore be a good basis for the calculation of regional sector structure. Amadeus reports the operating revenue of all firms which according to their national legislation have to report this to their governments, which is likely to exclude non-profit organizations and very small firms. The resulting database is much smaller (n= 244,670) than the one based on absolute numbers of firms but the nature of the selection may improve rather than reduce the quality of the data, for two reasons. Firstly this is because the model on agglomeration externalities was built with for-profit firms in mind and may not be applicable to non-profit organizations. And secondly very small firms may exist only for a short time, or may be registered but never reach the phase of actually creating revenue. In both cases these small firms would have a negligible impact on a region's sector structure and should ideally be excluded from the calculation of related variety, unrelated variety and specialization. However in the case of three countries the data on operating revenue is too incomplete to be used. Since using absolute numbers of firms and using a measure weighted by revenue both have advantages and drawbacks, both methods will be used in order to compare their merit.

Since the use of a database of individual firms is relatively new in the literature on agglomeration externalities, some further discussion is in order. The key question is to what extent the method used here results in an accurate depiction of the sector structure throughout the period of study. A potential problem arises when the dynamics of firm entry and exit are taken into account.

Besides firms that have existed throughout the 1999-2009 period, this database also includes firms that were founded after 1999 but before 2009, and firms that have existed at least in 1999 but later went out of business. In both cases it is likely that part of the firms included only existed during a short interval within the 1999-2009 period, and should have less impact on the economic performance of their region. It is not possible to exclude firms that did not exist throughout the full ten-year period since the original database does not make a distinction between firms for which in any specific year a financial statement is missing because the firm went out of business, or firms which did not go out of business but whose financial statement for this specific year was not included in the database. However assuming that sector structures is a stable property of regional economies with little year-by-year change, firm entry and exit dynamics is not expected to undermine the reliability of the analysis in this research. If any problem remains it would be expected to be caused by very small firms which only exist for a short period of time. These firms are excluded from the weighted measures of sector structure, making these potentially more reliable.

Another issue is related to the expected time lag between the dependent variable and its effect on the independent variables. Earlier studies on regional specialization and diversity usually compare the regional sector structure at T_0 to the employment, unemployment and or productivity growth at $T_{0 \text{ through } 1}$. In this way if a time lag exists before the impact of sector structure takes effect it would not be problematic. This study compares the sector structure at $T_{0 \text{ through } 1}$ to the employment, unemployment and productivity growth at $T_{0 \text{ through } 1}$ (here: 1999-2009 for the main predictors and 1998-2008 for the dependents, the difference arising from data availability issues). In other words while studies commonly lag the dependent variable behind the independent variables, this study compares the dependent and independent variables at the same time period. The alternative would have been to calculate the sector structure in 1999 (T_0) or for the period 1989-1999 ($T_{-1 \text{ through } 0}$), but neither are possible due to data limitations.

Again this should not be a major issue if it can be assumed that sector structures are stable over time, because in that case the sector structures in $T_{-1 \text{ through } 0}$ should be very similar to the ones in $T_{0 \text{ through } 1.}$ Since the dependent variables are split into two time periods this should to some extent provide a way to check to what extent the assumption of stable sector structures is reasonable. If the sector structure in $T_{-1 \text{ through } 0}$ was in fact very different than the one measured in $T_{0 \text{ through } 1}$ then the sector structure variables should fit the later time period (2003-2008) much better than the earlier time period (1998-2003), which would be more strongly impacted by the situation in $T_{-1 \text{ through } 0}$.

Having discussed the methods of calculating the sector structure variables, the resulting variables can now be inspected (figure 11 through 16).

Figure 11: Related Variety – unweighted

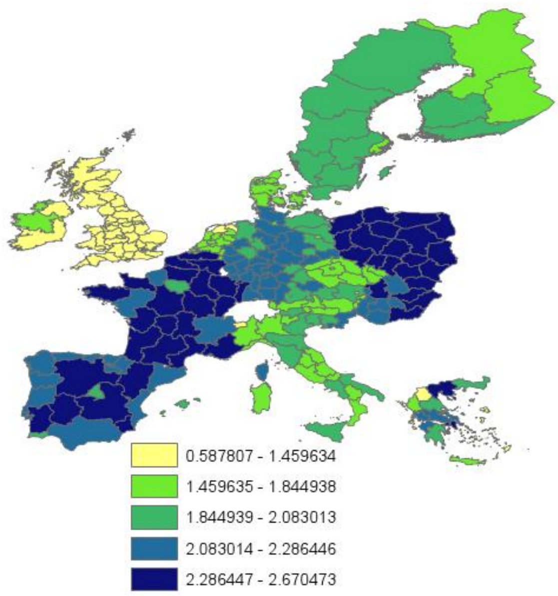


Figure 12: Related Variety – weighted by revenue

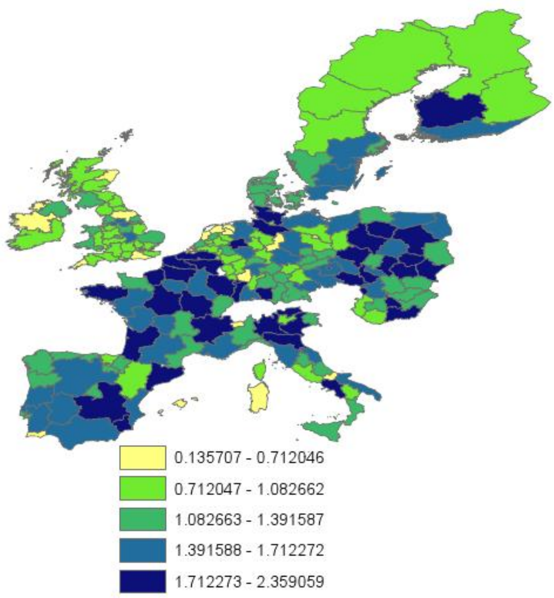


Figure 13: Unrelated Variety – unweighted

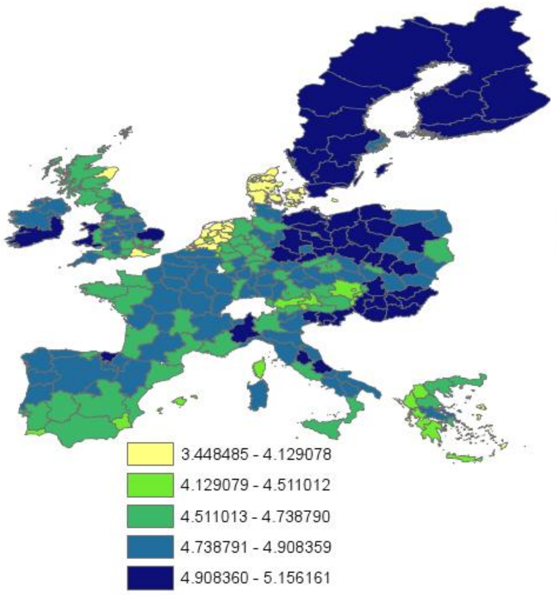


Figure 14: Unrelated Variety – weighted by revenue

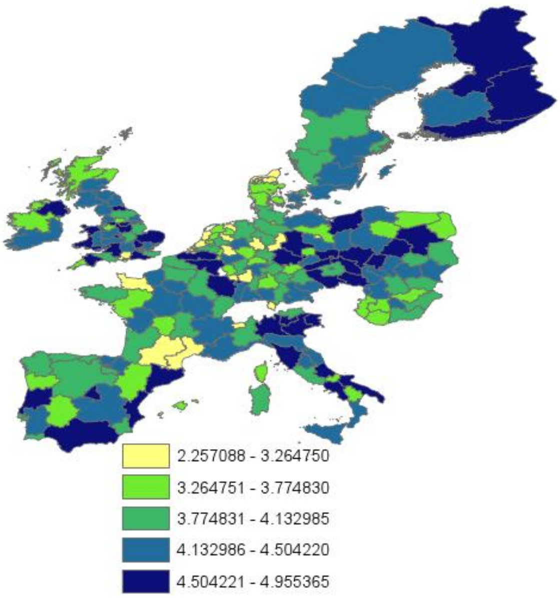


Figure 15: Specialization – unweighted, NACE 3 digit

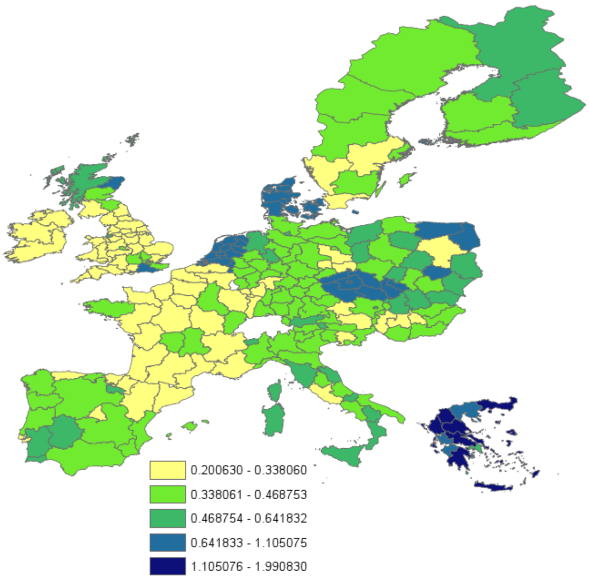
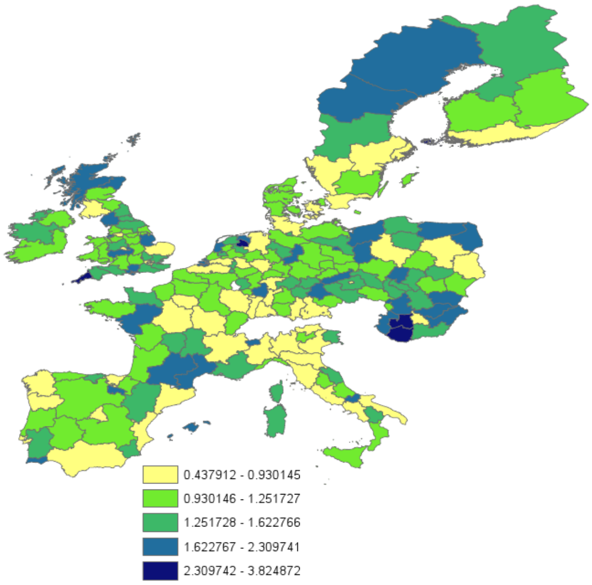


Figure 16: Specialization – weighted by revenue, NACE 3 digit



One difference is immediately apparent for all three variables. For the unweighted variables the level of the regional scores is quite similar within countries and strongly different in level between countries. The weighted variables show no such pattern. The country-level differences are strongest for the related variety levels of UK regions (which are all remarkably low) and the unrelated variety and specialization levels of the Netherlands and Denmark (all have very low unrelated variety and very high specialization). Moreover scores do not always match expectations. Large metropolitan areas such as Greater Londen, Ile de France, Madrid and the Randstad area in the Netherlands may be expected to show very high levels of variety and low levels of specialization. This is however only partly true, somewhat more so for the weighted than the unweighted variables. Other regions more clearly match expectations. For example regions dominated by natural resources extraction and refinement (e.g. natural gas in Groningen in the Netherlands and petrol in the Aberdeen region in Scotland) do show the expected high levels of specialization and low levels of variety.

The most likely source of bias in sector structure analysis is the modifiable areal unit problem (MAUP), which can affect measures sensitive to differences in the size of regions. Unfortunately this is the case for both the entropy measures of related and unrelated variety, and for the Theil index based on Location Quotients. If the size of regions is increased ceteris paribus more sectoral variety will be included in that region, resulting in higher values for related or unrelated variety and vice versa for a decrease in the size of regions. In the case of specialization the smaller the region the less variety will be included and the bigger the chance that some sectors will be over- or under-represented, giving rise to more extreme LQ's (further above and below average) and ultimately a higher score on specializaiton.

Since NUTS regions of the same level of aggregation (here NUTS2) can differ significantly in area it is likely that the MAUP is driving the unexpected outcomes shown in figures 11 through 16. In some cases NUTS regions are consistently bigger or smaller according to the member state they are located in (e.g. all NUTS2 regions of the Netherlands are relatively small while France's NUTS2 regions are large compared to the EU average), which can explain why some countries show consistently high or low values for the sector structure variables.

The problem is of course not the area of regions per se but the number of firms included in the region. To gain some insight into the extent of this problem the simple correlations between the sector structure variables and the number of firms per region can be calculated. If the suspected bias exists, the scores for related and unrelated variety should be positively related with the number of firms per region while the scores for specialization should be negatively related. First the correlations for unweighted sector structure (figure 17 through 19) and then for weighted sector structure (figure 20 through 22) are discussed.

Figure 17: Correlation between Related Variety (unweighted) and Number of Firms per Region

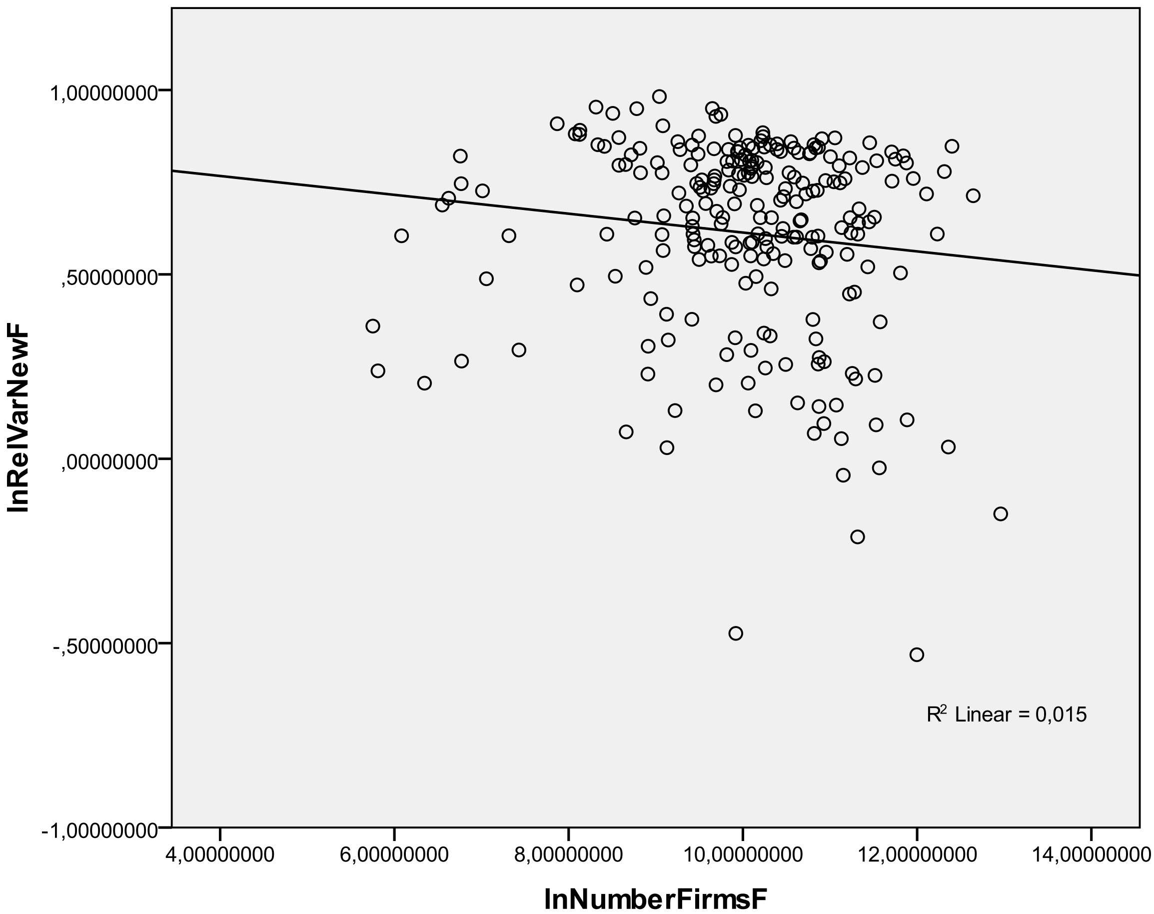


Figure 18: Correlation between Unrelated Variety (unweighted) and Number of Firms per Region

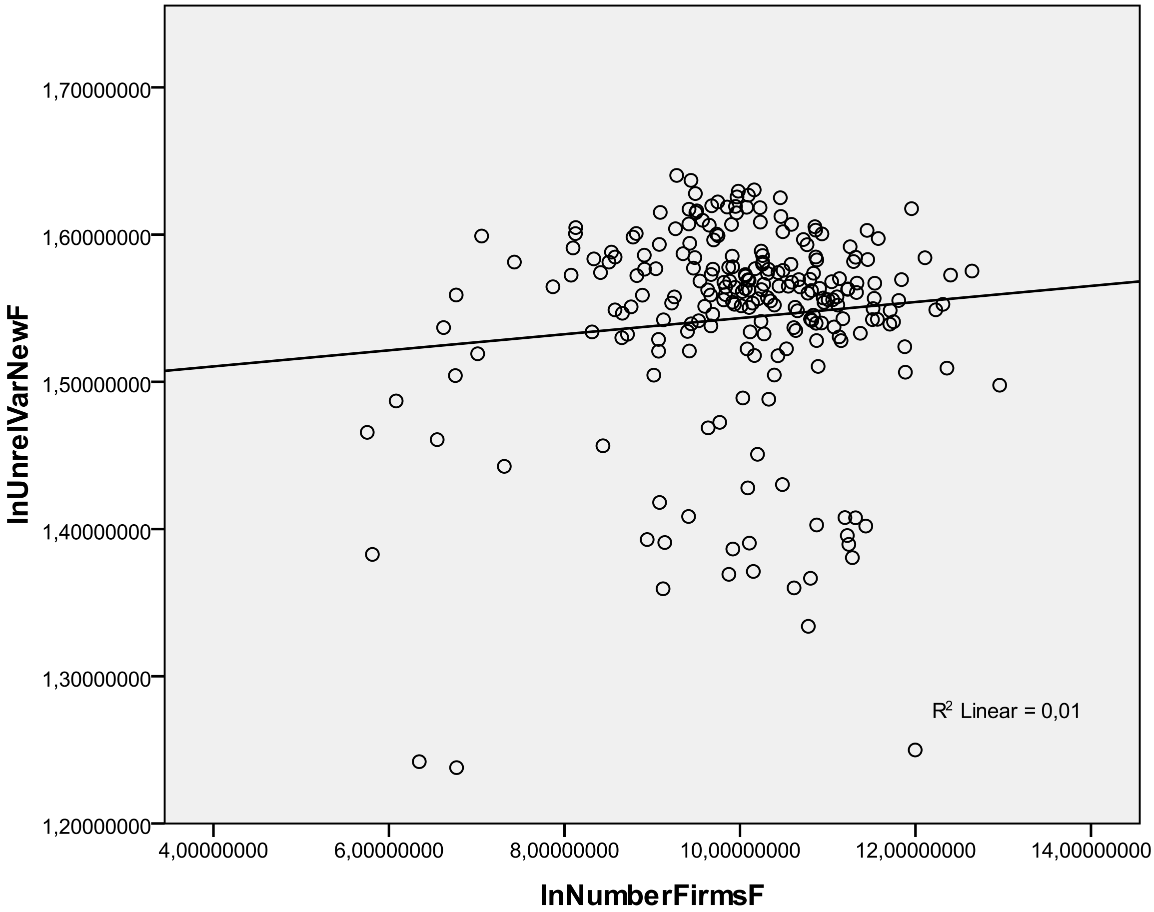
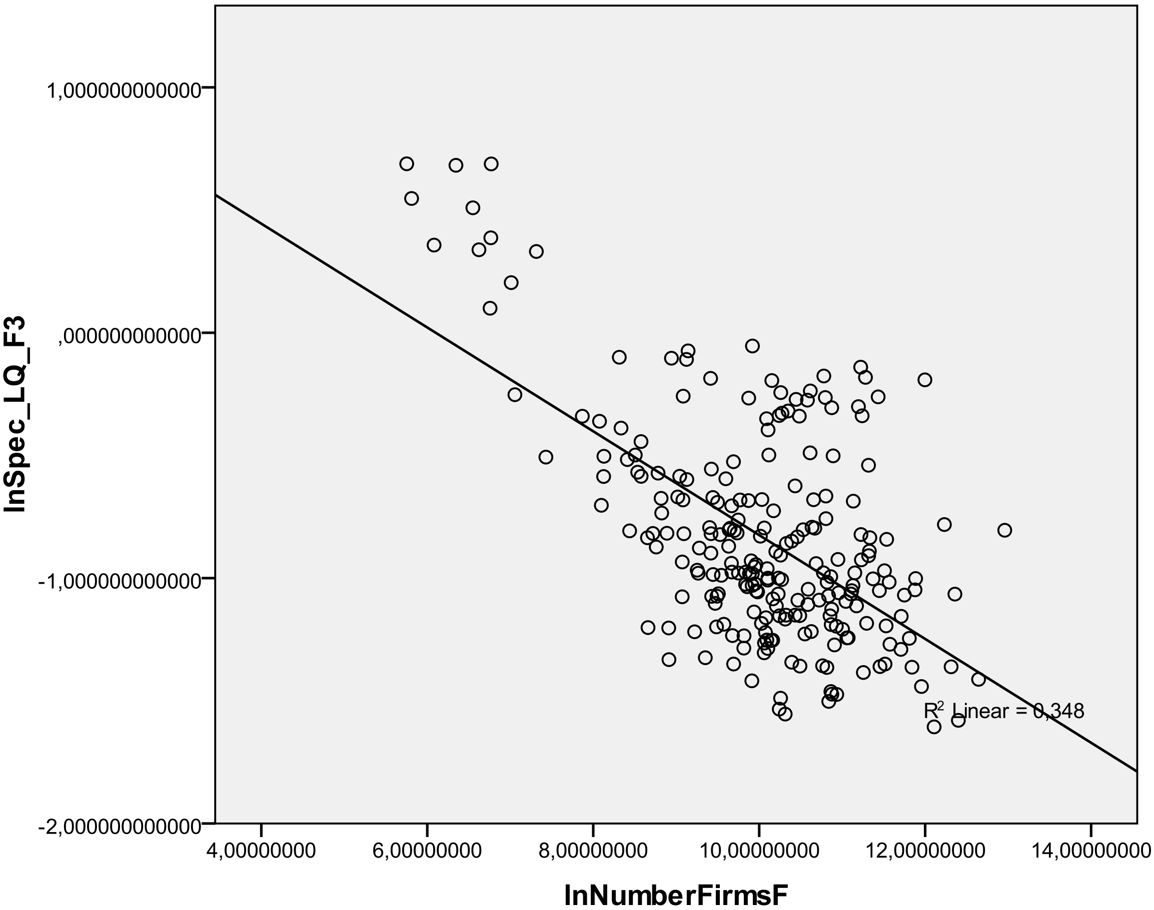


Figure 19: Correlation between Specialization (unweighted) and Number of Firms per Region



Unweighted related and unrelated variety seem hardly related to the number of firms per region, and related variety even shows a weak negative rather than the expected positive relation. Moreover a number of outliers may drive whatever weak correlation is found, so it should be concluded that there is no clear relation between unweighted related and unrelated variety and the number of firms. But unweighted specialization shows a much stronger negative relation with the number of firms per region, suggesting that there may be a significant bias caused by the modifiable areal unit problem.

Figure 20: Correlation between Related Variety and Number of Firms per Region (both weighted by revenue)

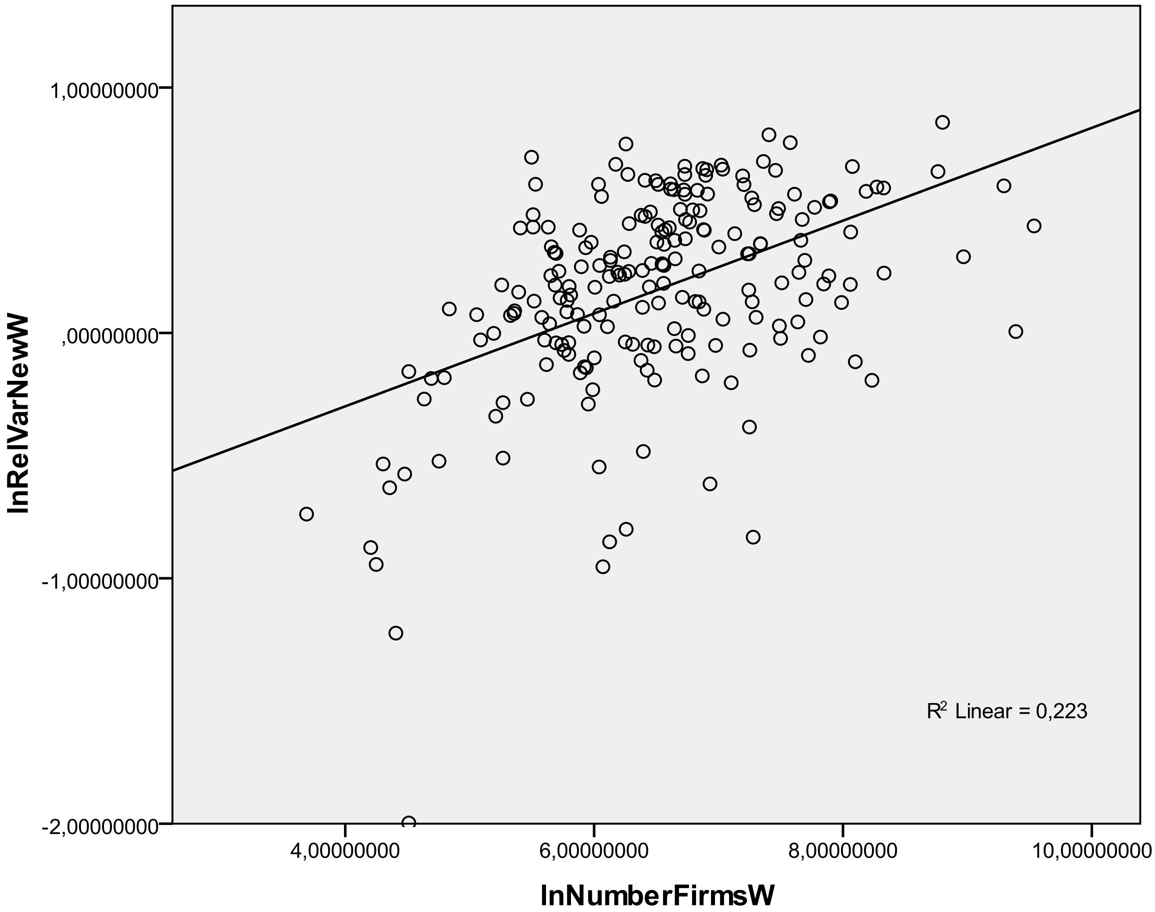


Figure 21: Correlation between Unrelated Variety and Number of Firms per Region (both weighted by revenue)

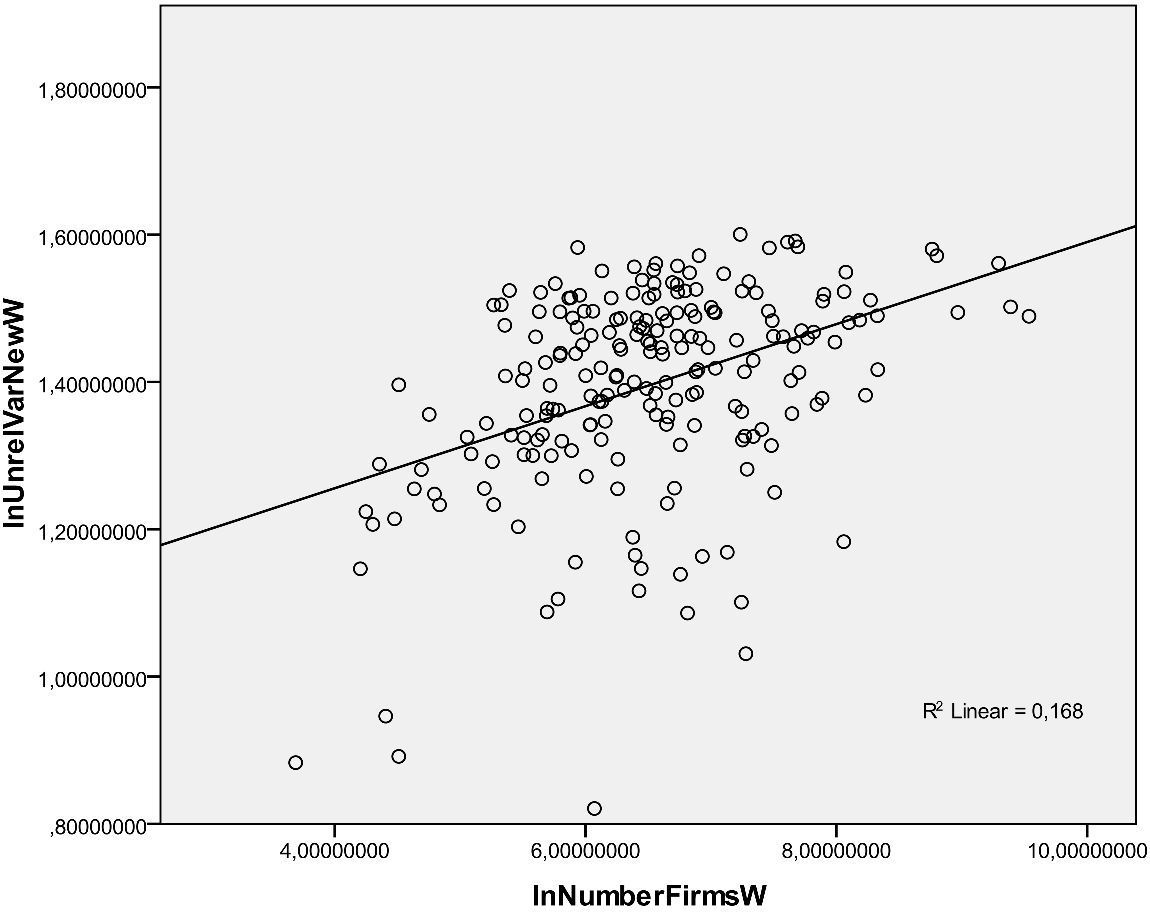
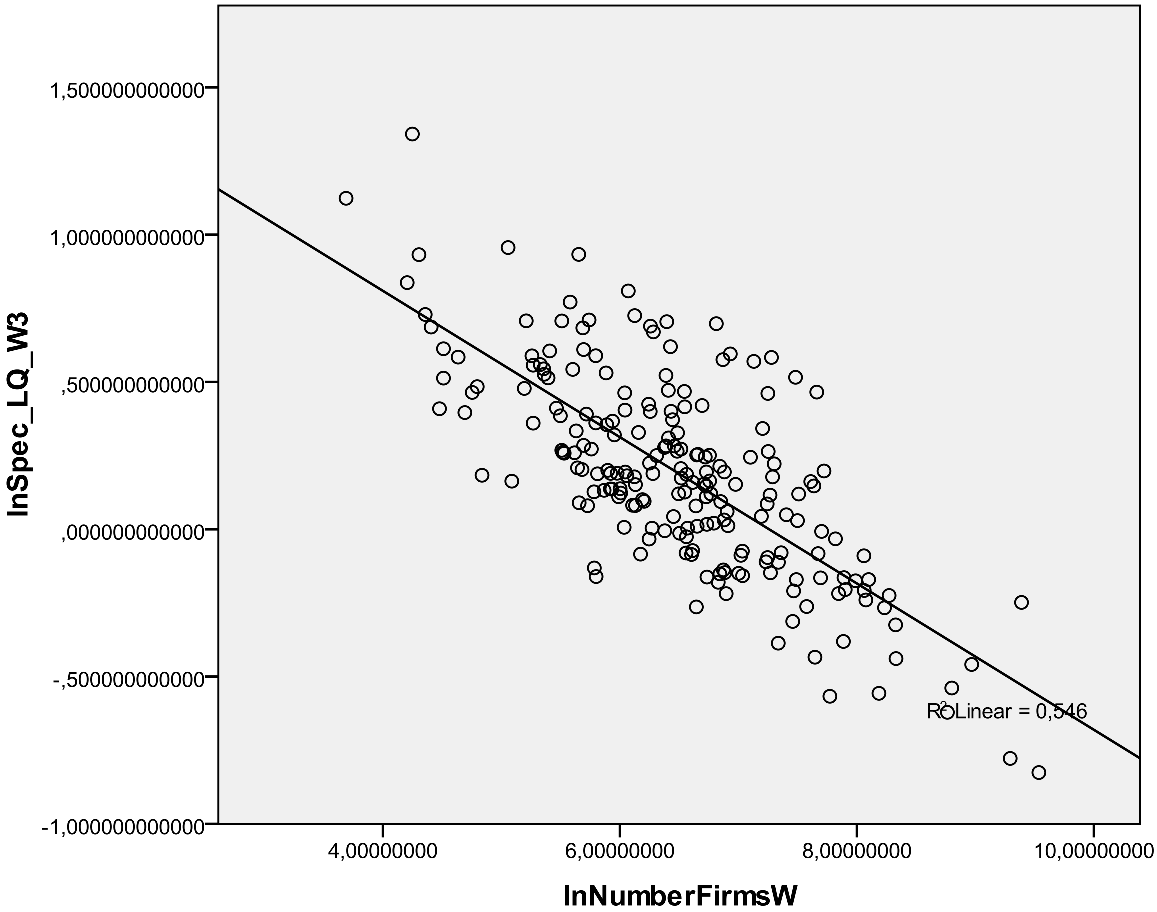


Figure 22: Correlation between Specialization and Number of Firms per Region (both weighted by revenue)



The weighted sector structure variables show much stronger signs of bias caused by differences in the size of regions. Related and unrelated variety are now quite clearly positively related to the number of firms (also weighted by revenue), while specialization shows the expected strong negative relation. Based on this reasonably strong evidence for a bias caused by the modifiable areal unit problem controls for the (weighted and unweighted) number of firms will be added to all regressions in this research.

3.6: Independents - controls

Besides the main predictors (the sector structure variables) a number of other independent variables will be added to the model. Some of them are expected to have a direct influence on the dependent variable, while others may have an indirect influence by strengthening the effect of the sector structure variables. For example for product innovation to take place besides a recombination of ideas (promoted by related variety) there is also a need for sufficient sources of R&D to turn the idea into a working product, and for workers with sufficient human capital to learn to use new production techniques and routines. Besides such direct and indirect influences on economic outcomes a number of dummy variables will be tried out to see which can take away the 'noise' caused by special circumstances that affect regional development in some countries but not in others, especially the catch-up of New Member States encountered in section 3.4. Maps for all control variables can be found in Appendix A.

INITIAL LEVEL OF THE INDEPENDENT

In all models the level of the independent variable at T_0 is added to control for different starting positions. So for the 1998-2003 analysis the 1998 value is included, and for the 2003-2008 analysis the 2003 value is used. This variable should correct for the expected pattern of high growth from a low initial level (mostly in the New Member States) versus low growth from an already high initial level. A negative sign for this variable would signal convergence.

POPULATION DENSITY

As discussed in chapter 2 population density is commonly regarded as a measure of urbanization economies. While urbanization economies have been a prime focus of prior agglomeration externalities research they are not part of the core model in this research. The reason is that while urbanization externalities originally included both pecuniary and knowledge externalities, they are reduced to only pecuniary externalities when related variety and unrelated variety have been accounted for. When pecuniary economies, pecuniary diseconomies and knowledge externalities are all lumped together urbanization externalities may still be expected to have a net positive influence on economic performance, but when only pecuniary economies and diseconomies are left it is hard to predict the net effect. Therefore population density as indicator for urbanization economies is still expected to be a significant variable in the model, but its expected sign is unclear. In all models population density is measured at T_0. The data on regional population are supplied by Cambridge Econometrics, while the area of regions in km^2 was supplied by Eurostat.

The NUTS classification of regions may cause some bias in the case of population density since in some countries large cities are classed as separate regions with little surrounding area included, while in other countries regions tend to include both a large city and a large area of surrounding rural or semi-rural land. In the former case population density will be much more pronounced than in the latter, even though actual population density (and hence the expected urbanization externalities) may be similar in the key cities in these regions.

A number of control variables are added following prior research, since they are both theoretically expected to influence the economic performance of regions and tend to be empirically confirmed as key factors.

HUMAN CAPITAL

This control variable is measured as the percentage of the regional labor force with a degree in higher (tertiary) education. Due to frequent breaks in the time series of Eurostat data only data for 2007 was complete for all regions, so this has been used in all models. It is assumed that education levels are a stable property of regions which does not radically change within the time frame of this research.

INVESTMENT

Total investment per region was supplied by Cambridge Econometrics, and measured at T_0 in all analyses. It has been changed into investment per worker in order to prevent strong collinearity between investment and the control for the initial level of the independent variable.

R&D

Data on total spending on intramural research and development was supplied by Eurostat, and has been distinguished into two components: R&D by private (for-profit) firms (Business R&D), and R&D by non-profit organizations (Non-Business R&D, the sum of R&D by academic institutions, governmens and NGOs). To prevent collinearity problems it was calculated as R&D per capita. The combination of breaks in time series and issues of data confidentiality means that only limited data were available to calculate these two variables. For most regions R&D data for 2005 was available, but for a number of regions data for 2004 or 2007 had to be used instead. For the region of Corsica only data for 2002 was available. Data for two German region was missing due to confidentiality and has been estimated based on patent data instead.

WAGE

The average wage of workers was supplied by Cambridge Econometrics, and is measured at T_0 in all models.

NUMBER OF FIRMS

Because of the potential bias of the sector structure variables caused by the modifiable areal unit problem, described in section 3.5, a control for the number of firms per region (weighted by revenue in the models using weighted sector structure variables) was added to every model. This data was calculated from the Amadeus database.

CAPITAL/LABOR RATIO

The Cambridge Econometrics database supplied data on total capital stock and total employment by region, which was used to calculate the capital/labor ratio. The variable is measured at T_0 in every model.

ACCESSIBILITY DUMMY

A dummy variable has been calculated to correct for differences in accessibility of regions. Baldwin & Wyplosz (2006) propose a classification of European regions into three levels of centrality: a European core region including the Benelux countries and part of France, Germany and the UK; an intermediate region including most of France, Germany, the UK, Italy and all of Denmark and Austria; and a European periphery including all new Member States, the Iberian peninsula, Ireland, Greece, most of Skandinavia, part of Schotland and part of Southern Italy. Their classification is adapted from a model of accessibility developed by Schürmann & Talaat (2000) which calculates the amount of economic activity (measured as GDP) within reach of any location within the EU by lorry under realistic assumptions of travel speed and maximum travel time. If the accessibility dummy is an effective control of relative centrality and peripherality of EU regions then the intermediate zone dummy should show a negative sign and the peripheral zone dummy a strongly negative sign in the regression models.

NEW MEMBER STATES DUMMY

In order to control for the marked differences between regions in EU15 countries and those in East-Central European countries a dummy is added for regions in the New Member States which joined the EU between 2004 and 2007.

3.7: Descriptives and measurement issues

Descriptives for all variables used in the analysis can be found in Appendix B (tables B1 through B3).

During data inspection it was found that a number of variables deviated substantially from the normal distribution. Transforming these variables to the natural log reduced deviance from normality to acceptable levels. Consequently all dependents and independents have been transformed to the natural log, so that the regression coefficients can be interpreted as a percentage change in the independent causing a certain percentage change in the dependent variable.

Finally a correlation matrix can be found in Appendix B (table B4). The most troubling simple correlations are those between Wage and Labor Productivity, Wage and C/L ratio and C/L ratio and Labor Productivity. Each of these correlations make theoretical sense. Firstly high labor productivity is expected to allow firms to pay high wages, but wages are not expected to have an impact on labor productivity. Since this study is interested in explaining labor productivity rather than wage level, Wage will not be used as an independent variable when Labor Productivity is the dependent. Secondly in economic growth theory the C/L ratio is the key driver of labor productivity, as ceteris paribus more and better tools and facilities make workers more productive. In a model estimating labor productivity the C/L ratio is likely to swamp the other predictors, making their coefficients less reliable and the model as a whole less informative. So while it is acknowledged that the C/L ratio is most likely the key factor driving labor productivity, it is left out of the model in order to study which other factors play a role besides the C/L ratio. And finally it makes sense that the C/L ratio is a strong predictor of the wage level, since Labor Productivity is a near perfect predictor of wage level and the C/L ratio shares much of its variance with Labor Productivity. Again it is opted to leave C/L ratio out of any model that includes Wage, in order to study whether wage level rather than the C/L ratio (appearing to be mostly an indicator of labor productivity) has a significant impact on the dependent variables.

4. Results

In this section a total of six models will be constructed to evaluate the relative importance of sector structure for explaining economic performance. For each of the three dependent variables a model with unweighted and one with weighted sector structure variables is used. As described in the previous chapter a control for the number of firms per region is added whenever sector structure variables are used, to correct for the expected bias caused by the modifiable areal unit problem. First a baseline model is built up which besides sector structure includes the initial level of the dependent variable, Population Density (a key variable in earlier research, but with an unclear expected sign in this research), Human Capital, and in the employment growth models Population Growth. Afterwards all other theoretically meaningful variables are added one by one to see if they have a significant impact on the model. Finally the robustness of the resulting full models is checked. Firstly all three dummy variables are added one by one (Accessibility, New Member States and Country Fixed Effects), and then the period of study is broken up in two periods to see if particular historical events have an impact on the model. Since z-values are used the relative importance of the independent variables can be directly compared.

Homoskedasticity is checked visually for all models from scatter plots. After transformation to the natural log there is no clear evidence for heterogeneous variances. Scatter plots for Labor Productivity show two distinct groups of cases, one with high and one with low growth levels. As expected inclusion of the New Member States dummy wipes away this pattern, suggesting that this dummy improves the quality of the models. However inclusion of this dummy also increases multicollinearity significantly, causing a trade-off. This study therefore reports the Labor Productivity models both with and without the New Member States dummy. Finally the scatter plots also show that most of the models have a number of significant outliers. The number of outliers is well within conventional rules of thumb, but the strength of the most extreme outliers is higher than expected. After having ruled out data entry errors and exploring possible causes for cases to deviate from the average (section 3.4 discusses the case of Åland, the most extreme outlier in the database) it is decided to keep the outliers in the database. The reason is firstly that this study is based on a census of regions rather than a sample, and secondly that the outliers do provide useful information on the actual situation in Europe.

5.1: Results for Employment Growth

The results for Employment Growth as dependent variable are provided in table 2 (unweighted sector structure) and table 3 (weighted). The baseline model (model 1) using unweighted sector structure variables shows no significant impact of sector structure on Employment Growth, while a significant impact of Related Variety was expected. Population Density shows a strong negative sign, suggesting that Employment Growth took place mostly outside of the most strongly urbanized regions. In other words it would be compatible with net urbanization diseconomies, for example caused by congestion. Human Capital has the expected positive sign, while Population Growth proves to be an essential control for a reliable model of Employment Growth (even though other studies on sector structure, e.g. Frenken et al. (2007) and Boschma & Iammarino (2009), tend to leave it out). When Investment and the two R&D variables are included the model is strengthened somewhat and Initial Employment changes sign (indicating convergence). Curiously enough Investment has the expected positive impact on Employment Growth, while Business R&D has an unexpected negative sign (Non-Business R&D is insignificant). Inclusion of Wage has some influence on the other predictors, but is insignificant itself.

Inclusion of the three dummy variables has a significant impact on the model. The Accessibility dummy (not shown) does not have the expected sign. The Intermediate Zone dummy performs better than the European core area rather than worse, while the periphery mirrors the effects of the New Member States dummy. Since it does not adequately measure accessibility and adds little information to what the New Member States dummy can tell us, the Accessibility dummy is not used. Its effect on the other models is found to be very similar, so it is not included in the remainder of the results. The New Member States dummy on the other hand is very informative and strongly significant. Initial Employment is strengthened, indicating that the convergence process is now more adequately measured by the model. Wage now gets the expected negative value, indicating that ceteris paribus higher wage levels hinder Employment Growth but only if the overall lower wage level in the New Member States is accounted for. While Population Density loses most of its significance, Related Variety becomes significant with the expected sign at the 0.05 level. When differences between the 'old' and 'new' EU member states are accounted for, regions with a sector structure of Related Variety have shown stronger Employment Growth in the 1998-2008 period.

When the model is corrected not for differences between New Member States and the rest of the EU, but for differences between individual countries (Country Fixed Effects) the model again shows some changes. Wage changes sign since the lower wage level in East-Central Europe is now not accounted for anymore, but Related Variety's positive sign is strengthened while Unrelated Variety now shows a clear negative impact on Employment Growth.

As a further robustness check the period of study is broken into two sub-periods, some interesting changes occur. Comparing with the full model without dummies (model 4), the sector structure variables hardly change (Unrelated Variety gets a marginally significant negative sign at the 0.10 level). But convergence (a negative sign for Initial Employment) appears to have been limited to the 1998-2003 period, while the significance of the effect of Population Density also seems to have disappeared by 2003-2008. All other controls change from significance to insignificance or vice versa over this time period. Since multicollinearity is limited (VIF scores are well below 10) this cannot explain the difference as an instability of the coefficients. Perhaps the 2003-2008 period, dominated by housing booms in much of Europe, was an atypical time period, also seeing that the adjusted R^2 for this period indicates that there is much more unexplained variance than for the 1998-2003 model.

The same analysis has been carried out with weighted rather than unweighted sector structure variables (table 3). While most variables show similar directions and strengths, the main difference is that when weighted sector structure is tested it does not show any significant outcomes except a marginally significant (0.10 level) negative sign for Specialization in the 2003-2008 period.

A final check of the models has been carried out by testing the sector structure models with Specialization not measured at the 3 digit level but at the 2 digit and 4 digit levels. These models yielded identical results with model 4, with none of the sector structure variables showing a significant impact.

Based on these models it is difficult to support *hypothesis 4*, which predicts a positive impact of Related Variety on Employment Growth and a negative impact of Specialization. When dummies are

Table 2: Dependent variable: EMPLOYMENT GROWTH*. Unweighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7a) OLS 1998-2003	(7b) OLS 2003-2008
CONSTANT	-0.066	-0.171	-0.242	-0.088	0.511*	-0.391	-0.291	-0.498***
	(-0.415)	(-1.027)	(-1.369)	(-0.377)	(1.924)	(-1.154)	(-1.346)	(2.879)
INITIAL EMPLOYMENT (LOG)	0.169**	-0.164*	-0.201**	-0.217**	-0.349***	-0.523***	-0.418***	0.122
	(1.995)	(-1.943)	(-2.406)	(-2.551)	(-3.975)	(-4.054)	(-4.537)	(1.104)
RELATED VARIETY (LOG)	0.060	0.060	0.061	0.061	0.120**	0.455***	0.013	0.095
	(1.130)	(1.140)	(1.175)	(1.167)	(2.297)	(3.513)	(0.233)	(1.461)
UNRELATED VARIETY (LOG)	0.008	0.052	0.040	0.021	0.024	-0.321***	0.064	-0.168*
	(0.108)	(0.672)	(0.529)	(0.261)	(0.315)	(-3.578)	(0.755)	(-1.696)
SPECIALIZATION (LOG)	-0.051	0.006	-0.032	-0.055	-0.081	-0.130	-0.070	-0.082
	(-0.616)	(0.072)	(-0.373)	(-0.613)	(-0.943)	(-1.218)	(-0.732)	(-0.746)
POPULATION DENSITY (LOG)	-0.159***	-0.156***	-0.161***	-0.162***	-0.099*	-0.253***	-0.159***	-0.073
	(-2.671)	(-2.640)	(-2.766)	(-2.783)	(-1.706)	(-4.161)	(-2.544)	(-0.999)
POPULATION GROWTH (LOG)	0.650***	0.650***	0.635***	0.636***	0.556***	0.398***	0.413***	0.431***
	(13.088)	(13.186)	(13.052)	(13.066)	(11.136)	(8.179)	(7.897)	(7.153)
NUMBER OF FIRMS (LOG)	0.168*	0.175*	0.280***	0.280***	0.434***	0.641***	0.542***	-0.195
	(1.827)	(1.908)	(2.793)	(2.791)	(4.195)	(3.754)	(5.040)	(-1.511)
HUMAN CAPITAL (LOG)	0.174***	0.156***	0.175***	0.179***	0.213***	0.170**	-0.055	0.438***
	(3.493)	(3.123)	(3.189)	(3.245)	(3.971)	(1.985)	(-0.920)	(6.313)
INVESTMENT (LOG)		0.104**	0.140**	0.162***	0.157***	0.029	0.173***	0.015
		(2.115)	(2.533)	(2.730)	(2.752)	(0.572)	(2.692)	(0.213)
BUSINESS R&D (LOG)			-0.214***	-0.180**	-0.058	-0.022	-0.246***	0.093
			(-3.111)	(-2.349)	(-0.738)	(-0.306)	(-2.984)	(0.957)
NON-BUSINESS R&D (LOG)			0.067	0.089	0.067	0.076	0.172**	-0.031
			(1.061)	(1.341)	(1.047)	(1.288)	(2.387)	(-0.369)
WAGE (LOG)				-0.089	-0.613***	0.624**	0.148	-0.578***
				(-1.017)	(-4.077)	(2.504)	(1.565)	(-5.088)
NEW MEMBER STATES DUMMY					-0.466***			
					(-4.204)			
COUNTRY FIXED EFFECTS						yes		
Adjusted R^2	0.569	0.576	0.590	0.590	0.619	0.769	0.520	0.355

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

Table 3: Dependent variable: EMPLOYMENT GROWTH. *Weighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7a) OLS 1998-2003	(7b) OLS 2003-2008
CONSTANT	0.000	-0.033	-0.102	-0.016	0.994***	-0.748*	-0.179	0.323***
	(0.005)	(-0.409)	(-1.113)	(-0.104)	(3.792)	(-1.775)	(-1.190)	(2.703)
INITIAL EMPLOYMENT (LOG)	-0.285***	-0.266***	-0.270***	-0.279***	-0.392***	-0.110	-0.453***	-0.062
	(-2.854)	(-2.659)	(-2.703)	(-2.763)	(-3.957)	(-0.840)	(-4.083)	(-0.474)
RELATED VARIETY (LOG)	-0.045	-0.036	-0.055	-0.066	-0.015	0.105	-0.058	-0.056
	(-0.702)	(-0.555)	(-0.839)	(-0.973)	(-0.224)	(1.618)	(-0.749)	(-0.662)
UNRELATED VARIETY (LOG)	0.072	0.087	0.082	0.080	0.022	-0.034	0.101	-0.019
	(1.287)	(1.559)	(1.472)	(1.434)	(0.395)	(-0.618)	(1.581)	(-0.271)
SPECIALIZATION (LOG)	-0.097	-0.068	-0.085	-0.094	-0.053	-0.03882	-0.003	-0.160*
	(-1.383)	(-0.950)	(-1.176)	(-1.286)	(-0.747)	(-0.60426)	(-0.034)	(-1.739)
POPULATION DENSITY (LOG)	-0.169***	-0.171***	-0.173***	-0.172***	-0.107**	-0.167**	-0.154**	-0.079
	(-3.078)	(-3.129)	(-3.199)	(-3.181)	(-1.995)	(-2.492)	(-2.515)	(-1.168)
POPULATION GROWTH (LOG)	0.627***	0.623***	0.626***	0.630***	0.539***	0.500***	0.410***	0.396***
	(12.427)	(12.408)	(12.537)	(12.524)	(10.423)	(9.411)	(7.426)	(6.283)
NUMBER OF FIRMS (LOG)	0.264**	0.248**	0.283**	0.282**	0.412***	0.016	0.498***	0.042
	(2.345)	(2.212)	(2.415)	(2.403)	(3.574)	(0.103)	(3.928)	(0.272)
HUMAN CAPITAL (LOG)	0.150***	0.129**	0.164***	0.167***	0.227***	0.116	-0.099	0.490***
	(2.989)	(2.506)	(2.967)	(3.008)	(4.177)	(1.226)	(-1.570)	(6.973)
INVESTMENT (LOG)		0.087*	0.125**	0.142**	0.135**	0.016	0.140**	0.084
		(1.739)	(2.279)	(2.364)	(2.362)	(0.284)	(2.062)	(1.119)
BUSINESS R&D (LOG)			-0.169***	-0.145**	-0.034	-0.037	-0.069	-0.101
			(-2.607)	(-1.985)	(-0.458)	(-0.505)	(-0.836)	(-1.062)
NON-BUSINESS R&D (LOG)			0.047	0.060	0.046	0.085	0.093	-0.041
			(0.680)	(0.831)	(0.687)	(1.272)	(1.161)	(-0.459)
WAGE (LOG)				-0.061	-0.813***	0.668**	0.234**	-0.61607***
				(-0.697)	(-4.465)	(2.128)	(2.338)	(-5.35811)
NEW MEMBER STATES DUMMY					-0.700***			
					(-4.647)			
COUNTRY FIXED EFFECTS						yes		
Adjusted R^2	0.599	0.603	0.612	0.611	0.648	0.749	0.502	0.387

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

used there is some support for the prediction for Related Variety in the case of unweighted models but not for weighted models, while the prediction for Specialization cannot be supported by any of the models.

5.2: Results for Labor Productivity Growth

Table 4 presents the results for Labor Productivity Growth as dependent variable (unweighted sector structure) and table 5 for weighted sector structure. The expectation is that Specialization will have a significant and positive impact on this dependent variable. In the baseline model (model 1) Specialization does in fact show a positive sign but it is only marginally significant. For the rest only Initial Labor Productivity (negative, indicating very strong convergence) and Human Capital (strongly positive) are significant predictors. While inclusion of Investment does not significantly change the model, Business R&D is a significant contributor to Labor Productivity Growth and also strengthens the effect of Specialization (now significant at the 0.05 level).

The New Member States dummy (model 4) is again strongly significant, and weakens the effect of Specialization. When the model is corrected for country effects rather than the New Member States all variables are weakened and the significance of all but Human Capital and Initial Labor Productivity is wiped out.

When the time period is split in two, Specialization becomes a strongly significant predictor in the 2003-2008 period, but not in 1998-2003. Human Capital on the other hand is now only significant in 1998-2003, while Investment changes signs between the two periods. Again it seems that great changes occured during the period of study, with again much greater unexplained variance in the later half of the period.

Using weighted instead of unweighted sector structure variables (table 5) weakens the conclusions drawn so far. Weighted Specialization does not show any significant role for Specialization, while Unrelated Variety becomes significantly negative in models without dummies added. The final check (not shown) replacing 3 digit level Specialization with 2 digit and 4 digit measures does not change any of the coefficients.

Hypothesis 5 cannot be unambiguously supported by the outcomes of the models. The positive impact it predicts of Specialization on Labor Productivity is only found in some very specific circumstances, namely unweighted sector structure variables without dummies added to the model, and especially during the 2003-2008 period.

Table 4: Dependent variable: LABOR PRODUCTIVITY GROWTH. *Unweighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6a) OLS 1998-2003	(6b) OLS 2003-2008
CONSTANT	0.311	0.310	0.926***	0.610**	0.479*	0.700***	0.236
	(1.359)	(1.344)	(3.661)	(2.422)	(1.749)	(3.471)	(1.545)
INITIAL LABOR PRODUCTIVITY	-0.843***	-0.844***	-1.051***	-0.522***	-0.805***	-1.073***	-0.671***
(LOG)	(-15.191)	(-12.526)	(-14.229)	(-3.874)	(-4.906)	(-11.999)	(-6.345)
RELATED VARIETY (LOG)	-0.036	-0.036	-0.041	-0.040	-0.085	-0.091	0.035
	(-0.710)	(-0.709)	(-0.880)	(-0.884)	(-0.705)	(-1.619)	(0.546)
UNRELATED VARIETY (LOG)	0.014	0.014	-0.023	-0.048	-0.027	-0.152*	0.173*
	(0.189)	(0.192)	(-0.337)	(-0.717)	(-0.324)	(-1.945)	(1.958)
SPECIALIZATION (LOG)	0.161*	0.162*	0.162**	0.135*	-0.087	-0.029	0.418***
	(1.970)	(1.937)	(2.019)	(1.752)	(-0.877)	(-0.321)	(4.170)
POPULATION DENSITY (LOG)	0.076	0.076	0.081	0.044	0.025	0.054	0.076
	(1.437)	(1.430)	(1.633)	(0.919)	(0.454)	(0.914)	(1.164)
NUMBER OF FIRMS (LOG)	0.038	0.038	-0.155**	-0.175***	0.092	-0.270***	0.080
	(0.588)	(0.587)	(-2.235)	(-2.641)	(1.414)	(-2.703)	(0.678)
HUMAN CAPITAL (LOG)	0.325***	0.325***	0.261***	0.249***	0.180**	0.329***	0.094
	(6.291)	(6.278)	(5.022)	(4.993)	(2.280)	(5.694)	(1.443)
INVESTMENT (LOG)		0.002	-0.036	-0.038	0.013	0.216**	-0.284**
		(0.039)	(-0.629)	(-0.697)	(0.265)	(2.118)	(-2.467)
BUSINESS R&D (LOG)			0.415***	0.299***	0.040	0.188**	0.588***
			(5.944)	(4.184)	(0.589)	(2.423)	(6.611)
NON-BUSINESS R&D (LOG)			0.056	0.076	0.065	0.075	-0.010
			(0.878)	(1.251)	(1.179)	(1.081)	(-0.126)
NEW MEMBER STATES DUMMY				0.501***			
				(4.612)			
COUNTRY FIXED EFFECTS					yes		
Adjusted R^2	0.572	0.570	0.628	0.660	0.793	0.539	0.412

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

Table 5: Dependent variable: LABOR PRODUCTIVITY GROWTH. *Weighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6a) OLS 1998-2003	(6b) OLS 2003-2008
CONSTANT	0.499*** (4.934)	0.503*** (4.977)	0.883*** (6.968)	0.252* (1.740)	0.304* (1.900)	0.238** (2.374)	0.646*** (7.724)
INITIAL LABOR PRODUCTIVITY (LOG)	-0.928*** (-16.248)	-0.886*** (-12.744)	-1.094*** (-14.042)	-0.226 (-1.595)	-0.548*** (-3.391)	-1.078*** (-12.099)	-0.813*** (-7.311)
RELATED VARIETY (LOG)	-0.017 (-0.256)	-0.007 (-0.098)	-0.010 (-0.159)	-0.024 (-0.404)	0.015 (0.294)	-0.036 (-0.513)	-0.009 (-0.107)
UNRELATED VARIETY (LOG)	-0.112** (-2.041)	-0.118** (-2.143)	-0.103* (-1.966)	-0.072 (-1.538)	-0.035 (-0.828)	-0.053 (-0.919)	-0.128* (-1.897)
SPECIALIZATION (LOG)	-0.005 (-0.069)	-0.009 (-0.122)	-0.010 (-0.142)	-0.053 (-0.847)	-0.014 (-0.277)	0.007 (0.088)	-0.015 (-0.163)
POPULATION DENSITY (LOG)	0.065 (1.261)	0.067 (1.303)	0.067 (1.385)	0.012 (0.280)	-0.038 (-0.751)	0.064 (1.178)	0.044 (0.685)
NUMBER OF FIRMS (LOG)	-0.008 (-0.108)	-0.012 (-0.171)	-0.106 (-1.474)	-0.150** (-2.326)	0.059 (1.044)	-0.188* (-1.840)	0.153 (1.188)
HUMAN CAPITAL (LOG)	0.351*** (6.810)	0.354*** (6.868)	0.277*** (5.343)	0.257*** (5.527)	0.206*** (2.854)	0.401*** (7.013)	0.039 (0.567)
INVESTMENT (LOG)		-0.063 (-1.057)	-0.088 (-1.549)	-0.075 (-1.487)	0.024 (0.537)	0.226** (2.066)	-0.423*** (-3.094)
BUSINESS R&D (LOG)			0.323*** (4.800)	0.224*** (3.614)	-0.008 (-0.150)	0.101 (1.360)	0.525*** (5.964)
NON-BUSINESS R&D (LOG)			0.069 (1.057)	0.075 (1.270)	0.064 (1.244)	0.029 (0.406)	0.071 (0.839)
NEW MEMBER STATES DUMMY				0.837*** (7.052)			
COUNTRY FIXED EFFECTS					yes		
Adjusted R^2	0.615	0.615	0.657	0.725	0.849	0.585	0.420

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

5.3: Results for Unemployment Growth

Finally the results for the third dependent variable, Unemployment Growth, are presented (unweighted sector structure in table 6, weighted sector structure in table 7). If the hypothesis is correct Unrelated Variety should protect against surges in Unemployment Growth, translating into a negative sign in the models tested.

The baseline model (model 1) however suggests Specialization rather than Unrelated Variety as a protection against growth of unemployment. Initial Unemployment is negative, again showing strong convergence among European regions. Population Density is strongly positively related to Unemployment Growth, which is compatible with its negative impact on Employment Growth found in section 5.1. Note however that the adjusted R^2 for this model is relatively low, suggesting that there are other strong factors impacting Unemployment Growth not included in the model. Wage and Investment themselves do not show a significant impact on Unemployment Growth, although they do reduce the impact of Specialization somewhat. Only Business R&D is significantly and negatively related to Unemployment Growth.

Adding the New Member States dummy in model 5 strengthens the model considerably, bringing the adjusted R^2 to a more acceptable level. The dummy is highly significant itself and also make Related Variety and Wage significant predictors. Both of these variables gain an unexpected sign however, since Related Variety is now associated with higher unemployment while higher wages are associated with lower unemployment (the latter suggesting reverse causality, the outcome variable predicting the predictor variable). Country Fixed Effects wipes out the sector structure variables again, while giving Human Capital its expected sign (protecting against growth of unemployment).

Splitting the period of study in two reveals very strong differences in the model. While in the first period Related Variety and Population Density are associated with growth of unemployment and Wage with lower unemployment, during 2003-2008 the pattern is reversed with Specialization and Business R&D being significantly associated with lower unemployment while Wage and Investment predict higher unemployment. Note however that the 1998-2003 model explains only a fraction of all variance of the dependent variable, while the 2003-2008 model is reasonably strong.

Using weighted instead of unweighted sector structure variables (table 7) changes the patterns observed so far. Related Variety no longer has a significant impact on the model, while this time Specialization is only significant in the 1998-2003 period. Unrelated Variety gets the expected negative sign in two models, significantly in the 1998-2003 period but marginally significantly (0.10 level) in the full model with the New Member States dummy added. However in 2003-2008 it changes signs to become positively associated with Unemployment Growth. As usual there is no change in the model if Specialization is calculated at the 2 digit or 4 digit level rather than the 3 digit level in the models so far.

Hypothesis 6, which predicts that Unrelated Variety protects against growth of unemployment, is again very hard to support. Of all models tested only one, with weighted sector structure variables in the 1998-2003 period, clearly shows this effect. But in the 2003-2008 period the sign for Unrelated Variety is reversed, while over the entire period of study (1998-2008) there is hardly any evidence for a significant impact of Unrelated Variety on Unemployment Growth.

Table 6: Dependent variable: UNEMPLOYMENT GROWTH. *Unweighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7a) OLS 1998-2003	(7b) OLS 2003-2008
CONSTANT	-0.093	-0.180	-0.141	-1.268	2.544**	5.138***	0.858	-3.961***
	(-0.116)	(-0.191)	(-0.150)	(-1.033)	(2.132)	(3.333)	(0.760)	(-3.522)
INITIAL UNEMPLOYMENT (LOG)	-0.58062***	-0.57777***	-0.57998***	-0.65149***	-0.79518***	-0.72999***	-0.26689***	-0.54996***
	(-7.58983)	(-7.38092)	(-7.39764)	(-7.89554)	(-10.5361)	(-8.15203)	(-2.92141)	(-7.64233)
RELATED VARIETY (LOG)	0.080313	0.080266	0.078957	0.100365	0.203903***	-0.12529	0.279932***	-0.04387
	(1.182489)	(1.17925)	(1.158619)	(1.477338)	(3.304528)	(-0.80901)	(3.721655)	(-0.74537)
UNRELATED VARIETY (LOG)	-0.08076	-0.07542	-0.06809	-0.03029	0.051906	0.172991	-0.01929	0.088085
	(-0.88589)	(-0.78535)	(-0.70497)	(-0.30607)	(0.587181)	(1.599071)	(-0.17608)	(1.059397)
SPECIALIZATION (LOG)	-0.21631**	-0.20984*	-0.19808*	-0.20216*	-0.21317**	-0.0285	0.124549	-0.21116**
	(-2.14926)	(-1.95997)	(-1.83023)	(-1.8045)	(-2.14503)	(-0.22395)	(1.004154)	(-2.28766)
POPULATION DENSITY (LOG)	0.205828***	0.206755***	0.205337***	0.214316***	0.326143***	0.429483***	0.372825***	0.006153
	(3.016595)	(3.015131)	(2.990708)	(3.146668)	(5.252744)	(5.86856)	(4.944114)	(0.103453)
NUMBER OF FIRMS (LOG)	0.357743***	0.35492***	0.365738***	0.484772***	0.648347***	0.740162***	0.085129	0.487765***
	(4.015436)	(3.915355)	(3.984168)	(4.438763)	(6.5413)	(5.866038)	(0.704029)	(5.464654)
HUMAN CAPITAL (LOG)	-0.05289	-0.05708	-0.05536	-0.04146	0.033945	-0.44248***	0.082525	-0.07963
	(-0.84546)	(-0.8536)	(-0.82665)	(-0.57946)	(0.528728)	(-4.29763)	(1.041672)	(-1.3122)
WAGE (LOG)		0.013616	-0.02557	0.067663	-0.93949***	-1.33101***	-0.34452***	0.37638***
		(0.180166)	(-0.28102)	(0.624009)	(-5.83652)	(-4.42854)	(-2.86971)	(3.960951)
INVESTMENT (LOG)			0.057751	0.058351	0.039879	0.054172	-0.04374	0.147611**
			(0.77462)	(0.777105)	(0.598376)	(0.887666)	(-0.52618)	(2.43799)
BUSINESS R&D (LOG)				-0.2696***	-0.05918	0.05889	0.016229	-0.36148***
				(-2.62362)	(-0.62264)	(0.66749)	(0.142649)	(-4.26086)
NON-BUSINESS R&D (LOG)				0.076805	0.017006	0.103494	0.030643	0.002383
				(0.910775)	(0.226165)	(1.466632)	(0.328202)	(0.033572)
NEW MEMBER STATES DUMMY					-0.9319***			
					(-7.80302)			
COUNTRY FIXED EFFECTS						yes		
Adjusted R^2	0.328	0.325	0.324	0.340	0.481	0.669	0.191	0.532

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

Table 7: Dependent variable: UNEMPLOYMENT GROWTH. *Weighted sector structure variables*

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7a) OLS 1998-2003	(7b) OLS 2003-2008
CONSTANT	-0.196	-0.559	-0.527	-0.963	5.521***	9.286***	2.374***	-3.932***
	(-0.509)	(-1.065)	(-0.964)	(-1.220)	(4.426)	(5.981)	(3.519)	(-4.817)
INITIAL UNEMPLOYMENT (LOG)	-0.56432***	-0.54767***	-0.54769***	-0.5706***	-0.594***	-0.63823***	-0.15828**	-0.52027***
	(-7.01025)	(-6.66788)	(-6.65227)	(-6.56728)	(-7.48252)	(-8.64182)	(-1.78311)	(-6.61637)
RELATED VARIETY (LOG)	-0.0699	-0.04174	-0.04399	-0.04199	0.062158	-0.06264	0.046767	-0.03707
	(-0.79275)	(-0.45178)	(-0.4721)	(-0.44631)	(0.711029)	(-0.87906)	(0.486508)	(-0.45426)
UNRELATED VARIETY (LOG)	-0.0563	-0.04323	-0.04186	-0.04315	-0.12687*	-0.01553	-0.26645***	0.137127**
	(-0.74882)	(-0.56693)	(-0.54582)	(-0.5596)	(-1.77205)	(-0.25728)	(-3.38242)	(2.035924)
SPECIALIZATION (LOG)	-0.12258	-0.09222	-0.09098	-0.08898	-0.01611	-0.08683	-0.242**	0.076075
	(-1.27657)	(-0.91725)	(-0.90142)	(-0.87813)	(-0.17287)	(-1.23388)	(-2.33771)	(0.869982)
POPULATION DENSITY (LOG)	0.179657**	0.185138***	0.184797***	0.183336**	0.289884***	0.377568***	0.25278***	0.014411
	(2.547836)	(2.618182)	(2.606518)	(2.573)	(4.316705)	(5.111754)	(3.472339)	(0.23159)
NUMBER OF FIRMS (LOG)	0.382166***	0.362804***	0.363668***	0.404011***	0.421093***	0.534856***	0.017653	0.416079***
	(3.453571)	(3.231543)	(3.229563)	(3.235728)	(3.694118)	(5.27413)	(0.138382)	(3.917367)
HUMAN CAPITAL (LOG)	-0.01698	-0.04633	-0.04667	-0.03051	0.063406	-0.48152***	0.104929	-0.10136
	(-0.25311)	(-0.63462)	(-0.63751)	(-0.39621)	(0.882981)	(-4.64146)	(1.333828)	(-1.50031)
WAGE (LOG)		0.082978	0.070698	0.124186	-1.14758***	-1.6112***	-0.45776***	0.5111***
		(1.018372)	(0.714743)	(1.040941)	(-5.04215)	(-5.05524)	(-3.75558)	(4.673515)
INVESTMENT (LOG)			0.018062	0.022588	0.007425	-0.00115	-0.10753	0.081966
			(0.220071)	(0.271623)	(0.097782)	(-0.01839)	(-1.26558)	(1.151428)
BUSINESS R&D (LOG)				-0.08573	0.093588	0.101197	-0.08535	-0.10917
				(-0.81266)	(0.932825)	(1.26041)	(-0.79187)	(-1.18063)
NON-BUSINESS R&D (LOG)				-0.0153	-0.00608	0.116783	0.074723	-0.0916
				(-0.1572)	(-0.06846)	(1.605185)	(0.751656)	(-1.08495)
NEW MEMBER STATES DUMMY					-1.17086***			
					(-6.36326)			
COUNTRY FIXED EFFECTS						yes		
Adjusted R^2	0.263	0.264	0.260	0.255	0.380	0.698	0.223	0.438

*Notes: standardized betas, with t-values in parentheses; significance is marked as ***0.01, **0.05, *0.10*

5. Conclusion and discussion

This study has used an EU-wide database to test the model of agglomeration externalities as proposed by Frenken et al. (2007), distinguishing related variety and unrelated variety besides specialization and urbanization. While several country-level tests have confirmed predictions based on this model, the empirical analysis in this research cannot find clear evidence in favor of the model. Related variety only has the expected significant positive effect on employment growth in the model with sector structure variables based on absolute numbers of firms (i.e. not weighted by the size of firms) and with country fixed effects (national-scale differences) accounted for, while no effect of specialization on employment growth can be found.

When labor productivity is used as dependent variable the expected positive impact of specialization is again only found in the the unweighted model, but its significance is ambiguous. And finally for unemployment growth as outcome variable no consistent relation can be found for unrelated variety, while unexpectedly specialization shows a stronger negative impact (i.e. reducing unemployment growth) in the unweighted models. However the relation between specialization and unemployment growth is only marginally significant and is not robust for changes in the period of study.

Besides testing a key model of agglomeration externalities, this study also made a number of other empirical contributions. Firstly this study demonstrates that measuring sector structure based on absolute numbers of firms or weighted by revenue has a significant impact on the model, sometimes even reversing the signs of sector structure variables. So far very few studies have used variables based on the number of firms per sector, while none of the studies that use related and unrelated variety have done so. The theoretical models proposed so far by the literature do not exclude either of the approaches, so further research is needed to assess the added value of measures based on absolute numbers of firms.

Secondly this study has begun to compare the relative strength of sector structure variables measured at different levels of industry sector detail. Specialization indicators were calculated at 2, 3 and 4 digit level, holding methodology constant. No significant changes were found, indicating that at least in the case of specialization the level of sector detail used to measure the indicator has almost no influence on the outcomes. Further research could test if the same is true for related and unrelated variety. Thirdly this study has followed the call by Gardiner et al. (2004) in controlling for differences between regions in the average work week, but again no significant impact on the models was detected. While the average workweek does differ markedly between European regions, these differences turned out to be insignificant compared to the much bigger differences in employment and labor productivity.

Taken together the results of this study pose a challenge to the theoretical model of agglomeration externalities, while they also call for a reassessment of the empirical techniques used. Even though this study cannot corroborate existing theories on sector structure, this should be interpreted as a call for further research to find out why this Europe-wide test found such different effects than the country-level studies carried out previously. More specifically the findings of this research could be signs of a number of very different issues. One possibility is that it is purely due to the limitations of currently used measurement techniques. This study found clear signs that the measures of sector structure commonly used are significantly sensitive to the modifiable areal unit problem. While this study has tried to correct for this problem by adding the number of firms as a control variable, a better solution would be to develop a continuous space model that is free from any interference caused by subdividing space into unequal units (Burger et al. 2010).

Another explanation for the paucity of confirmatory findings may be that regional sector structures do not in fact have a significant influence on the overall economic landscape, but their effects are limited to some specific industrial sectors. Bishop and Gripaios (2010) demonstrate that in the UK related and unrelated variety are indeed of influence to some but not all sectors they study. Further research should take up the empirical challenge of finding out whether at the EU-scale the effects of sector structure are also different between sectors, and the theoretical challenge of explaining why the same model would not apply equally to firms in different sectors.

Related to the latter explanation, it is also possible that sector structure does not have the same effects in regions facing different economic circumstances. Besides differences in industrial composition, regions also differ in among others their institutional structure. Other things equal regions with an institutional structure that facilitates inter-firm co-operation are expected to benefit more from the effects of sector structure proposed in models of agglomeration externalities. Differences in academic and private research systems and differences in the willingness and ability of workers to start up new firms may also interfere with the workings of agglomeration externalities, while some of these are likely to be determined at the national rather than regional level. The strong influence of the country-fixed effects dummies on this analysis would be compatible with this interpretation of the research findings.

A final possibility is that the weakness of the measured effects of sector structure is due to the temporal scale (10 years) used in this research. While the time periods used are broadly consistent with earlier research, the discussion of the theoretical model in this study suggested that there are in fact reasons to suspect that longer time periods may give quite different results. Knowledge externalities are associated with the ability of regional economies to avoid long-term declined caused by the inability to anticipate structural change and to deal with asymmetric shocks. Both effects are theoretically expected to become visible only in the long run, which may suggest that using a temporal scale of 10 years is too short to fully capture them.

Further research should disentangle these possible explanations of the lack of confirming evidence and thereby strengthen our empirical and theoretical understanding of the role of sector structure in driving agglomeration externalities. Besides the scientific value of the proposed refinements of the model, it is also important for policymakers to know to what extent and in what specific circumstances the model of regional sector structure and agglomeration externalities can be used as a tool for stimulating regional economic performance in the long run.

Besides the challenges already mentioned, two more directions for further research can be mentioned. Firstly as discussed in the theoretical section of this study sector structure is only expected to have an indirect effect on economic performance, while its direct effects are linked to process and product innovation. If a suitable indicator for these forms of innovation can be found, the effects of sector structure could be tested more directly and more accurately. And secondly an open question is whether to base calculations of sector structure indicators on input or output data. Most previous studies used employment data, which shows whether the main input (labor) of firms is related or not. This study is one of the few which uses sector structure variables that are closer to indicators of relatedness of output, since firms and revenue are distinguished by the type of product they produce rather than the inputs they use. Theoretical work could elucidate whether relatedness of input or of firm output is most likely to capture the effect of sector structure.

Literature

- Baldwin & Wyplosz (2006), *The Economics of European Integration* 2nd edition, Berkshire: McGraw Hill
- Beaudry & Schiffaurova (2009), Who's right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy*, vol. 38, pp. 318-337.
- Bishop & Gripaios (2010), Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies*, vol. 44, pp. 443-454
- Boschma (1997), New industries and windows of locational opportunity. A long-term analysis of Belgium, *Erdkunde*, vol. 51, pp. 12-22
- Boschma (2005), Proximity and Innovation: A Critical Assessment, *Regional Studies*, vol. 39, pp. 61-74
- Boschma (2007), Path creation, path dependence and regional development. In: J. Simmie and J. Carpenter (eds.), *Path Dependence and the Evolution of City Regional Development*, Chapter 3, Working Paper Series No. 197, Oxford Brookes University, Oxford, pp. 40-55
- Boschma, Eriksson & Lindgren (2009), How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography*, vol. 9, pp. 169-190
- Boschma & Frenken (2006), Why is economic geography not an evolutionary science? Towards an Evolutionary Economic Geography. *Journal of Economic Geography*, vol.6, pp. 273-302
- Boschma & Frenken (2009) Technological relatedness and regional branching, in: Bathelt, Feldman & Kogler (eds.), *Dynamic Geographies of Knowledge Creation and Innovation*, Routledge, Taylor and Francis
- Boschma, & Iammarino (2009), Related Variety, Trade Linkages and Regional Growth, *Economic Geography*, vol. 85, pp. 289-311
- Boschma & ter Wal (2007), Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy. *Industry and Innovation*, vol. 14, pp. 177-199
- Boschma and Wenting (2007), The spatial evolution of the British automobile industry: does location matter? *Industry and Corporate Change*, vol. 16, pp. 213-238
- Burger, van Oort & van der Knaap (2010), A treatise on the geographical scale of agglomeration externalities and the modifiable areal unit problem. *Scienze Regionali*, vol. 9, pp. 19-40
- Burger, van Oort & Raspe (2011), Agglomeration and new establishment: a mixed hierarchical and cross-classified model. In: Kourtit, Nijkamp & Stough (eds), *Drivers of Entrepreneurship, Innovation and Regional Dynamics*, Berlin: Springer
- Cutrini (2006), The Balassa Index meets the Dissimilarity Theil Index: a decomposition methodology for Location Studies, *Working papers 274*, Universita' Politecnica delle Marche (I), Dipartimento di Scienze Economiche e Sociali
- Duranton & Puga (2000), Diversity and specialisation in cities: why, where and when does it matter? *Urban Studies*, vol. 37, pp. 533-555
- Essletzbichler (2007), Diversity, Stability and Regional Growth in the United States, 1975-2002. In Frenken (ed.), *Applied Evolutionary Economics and Economic Geography* , Cheltenham, UK: Edward Edgar

- Essletzbichler & Rigby (2010), Generalized Darwinism and Evolutionary Economic Geography. In Boschma & Martin (eds.), *The Handbook of Evolutionary Economic Geography,* Cheltenham, UK: Edward Elgar
- European Commission (2010), *Europe 2020: a strategy for smart, sustainable and inclusive growth*, available online: http://ec.europa.eu/europe2020/
- Feldman & Audretsch (1999), Innovation in cities: Science-based diversity, specialization and localized competition, *European Economic Review*, vol. 32, pp. 409-429
- Frenken & Boschma (2007), A theoretical framework for economic geography: industrial dynamics and urban growth as a branching process, *Journal of Economic Geography*, vol. 7, pp. 635-649
- Frenken, van Oort, Verburg & Boschma (2004), *Variety and regional economic growth in the Netherlands*. The Hague: Ministry of Economic Affairs
- Frenken, van Oort & Verburg (2007), Related variety, unrelated variety and regional economic growth, *Regional Studie*, vol. 41, pp. 685-697
- Gardiner, Martin & Tyler (2004), 'HEAD', *Regional Studies*, vol. 38, pp. 1045-1067
- Glaeser, Kallal, Scheinkman & Shleifer (1992), Growth in cities, *Journal of Political Economy*, vol. 100, pp. 101-149
- De Groot, Poot, & Smit (2009), Agglomeration, Innovation and Regional Development: Theoretical Perspectives and Meta-Analysis, in Capello & Nijkamp (eds.), *Handbook of regional growth and development theories*, Cheltenham: Edward Elgar, pp. 256-281
- Henderson, Kuncoro & Turner (1995), Industrial development in cities, *Journal of Political Economy*, vol. 103, pp. 1067-85
- Jacobs (1969), *The economy of cities*, New York: Vintage
- Jaffe, Trajtenberg & Henderson (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, vol. 108, pp. 577-598
- Klepper (2007), Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry, *Management Science*, vol. 53, pp. 616-631
- Markusen (1996), Sticky places in slippery space: a typology of industrial districts. Economic Geography, vol. 72, pp. 293-313
- Marshall (1890), *Principles of Economics*. London: MacMillan
- Martin & Sunley (2003), Deconstructing clusters. Chaotic concept or policy panacea? *Journal of Economic Geography*, vol. 3, pp. 5-35
- Martin & Sunley (2006), Path dependence and regional economic evolution. *Journal of Economic Geography*, vol. 6, pp. 395–437
- Martin & Sunley (2010), Path Dependence and the Evolution of the Economic Landscape. In *The Handbook of Evolutionary Economic Geography*, Boschma and Martin (eds.) Cheltenham, UK: Edward Elgar
- Neffke, Henning & Boschma (2011), How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, *Economic Geography*, vol. 87, pp. 237-265
- Nooteboom (2000), *Learning and Innovation in Organizations and Economies*. Oxford: Oxford University Press.
- Van Oort (2004), *Urban growth and innovation: spatially bounded externalities in the Netherlands*, Aldershot: Ashgate

- Nordic Council of Ministers (2010), *Nordic Statistical Yearbook 2010*, Copenhagen: Nordic Council of Ministers
- Ponds, van Oort & Frenken (2010), Innovation, spillovers, and university-industry collaboration: an extended knowledge production function approach, *Journal of Economic Geography*, vol. 10, pp. 231-255
- Saxenian (1996), Regional Advantage: Culture and Competition in Silicon Valley and Route 128, Cambridge, US: Harvard University Press
- Setterfield (1997), Should economists dispense with the notion of equilibrium?, *Journal of Keynesian Economics*, vol. 20, pp. 47–76
- Schürmann & Talaat (2000), *Towards a European Peripherality Index, Report for General Directorate XVI Regional Policy of the European Commission*, Dortmund: Institute of Spatial Planning.
- Storper & Walker (1989), *The Capitalist Imperative: Territory, Technology and Industrial Growth,* Oxford: Blackwell,
- Weterings, van Oort, Raspe & Verburg (2007), *Clusters en Economische Groei*, Den Haag: RPB

Appendix A: Maps

Figure A1: Employment Growth 2003-2007

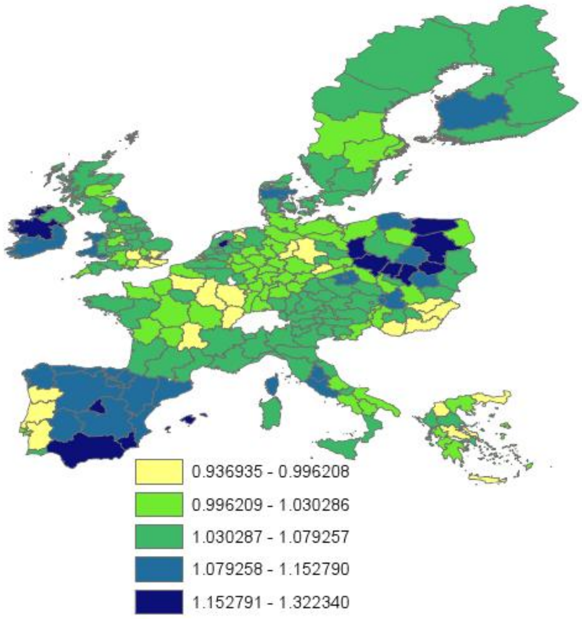


Figure A2: Labor Productivity Growth 2003-2007

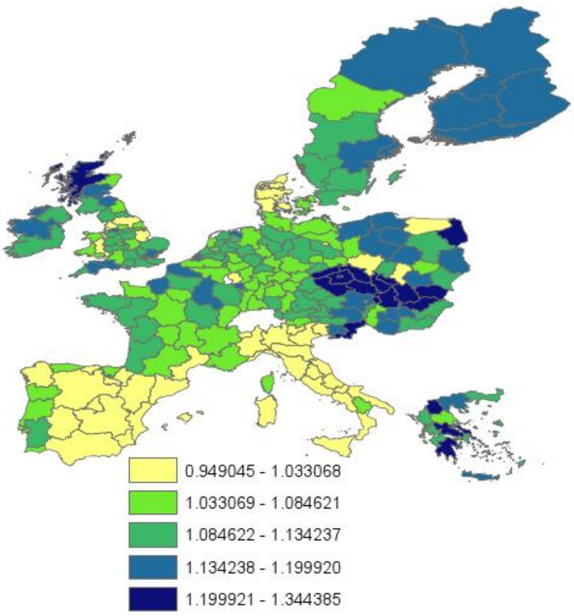


Figure A3: Unemployment Growth 2003-2007

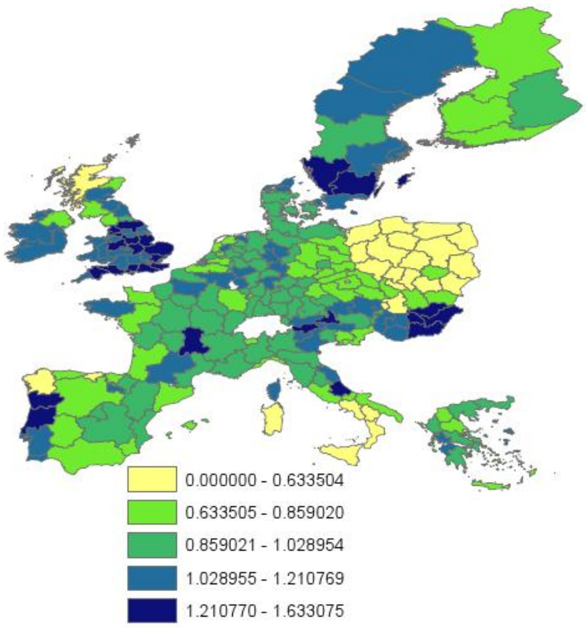


Figure A4: Specialization, unweighted, 2digit

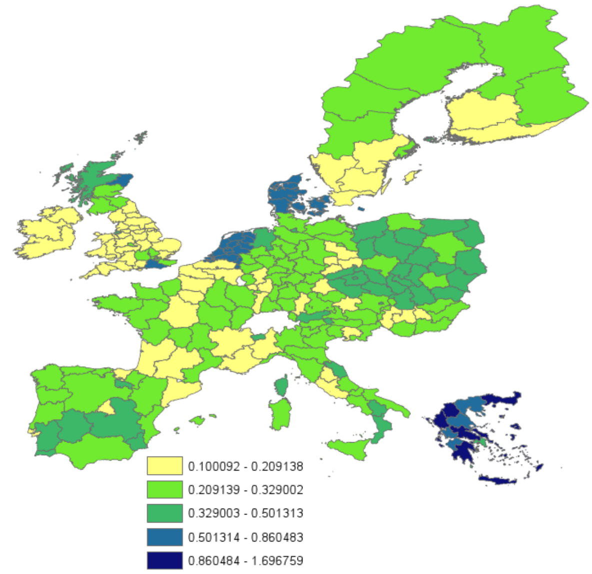


Figure A5: Specialization, unweighted, 4digit

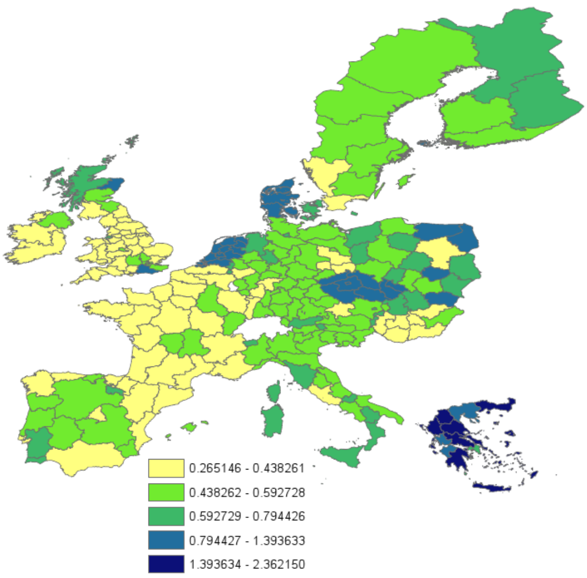


Figure A6: Specialization, weighted, 2 digit

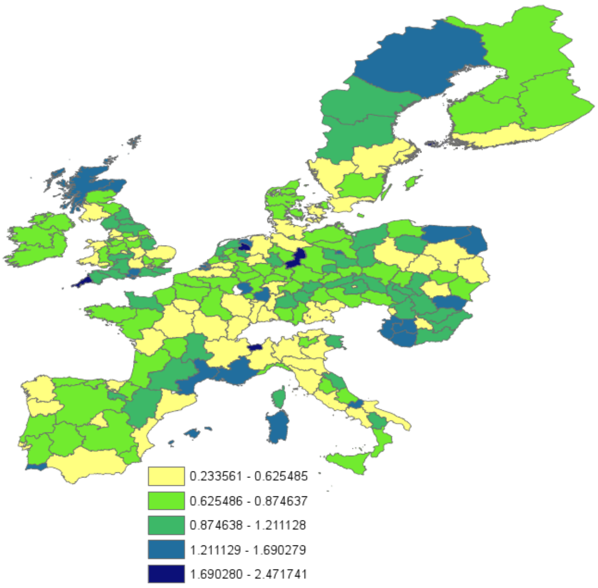


Figure A7: Specialization, weighted, 4 digit

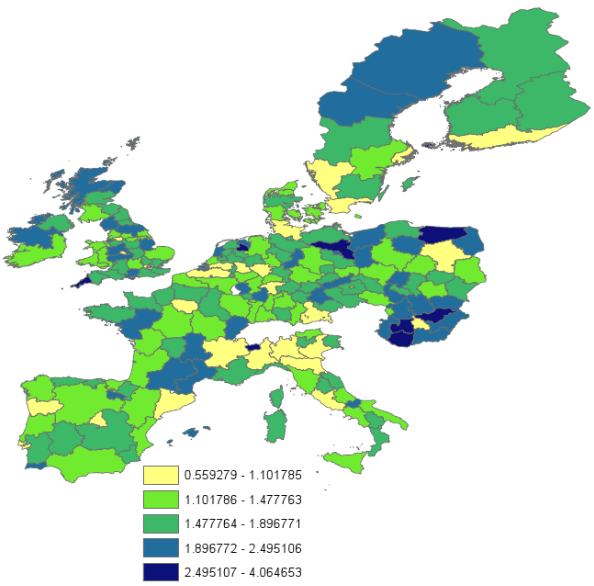


Figure A8: Number of Firms

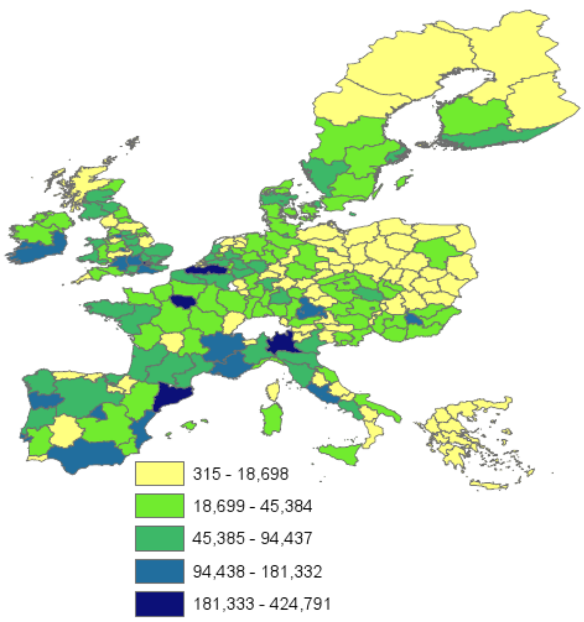


Figure A9: Number of Firms, weighted by revenue

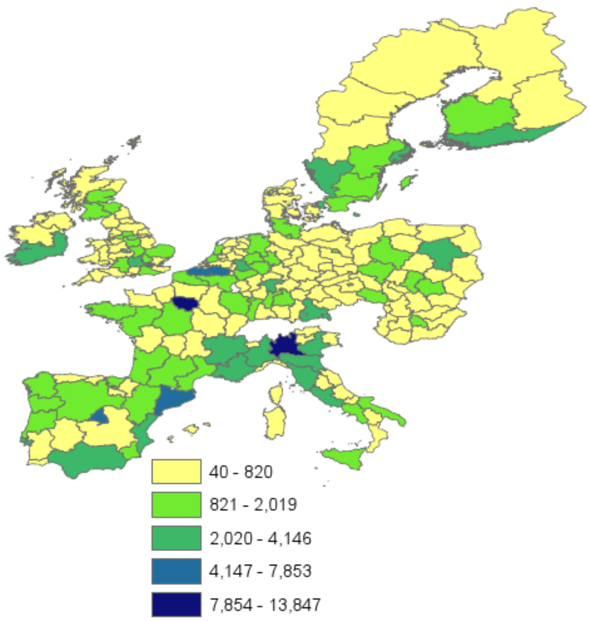


Figure A10: Population Growth 1998-2008

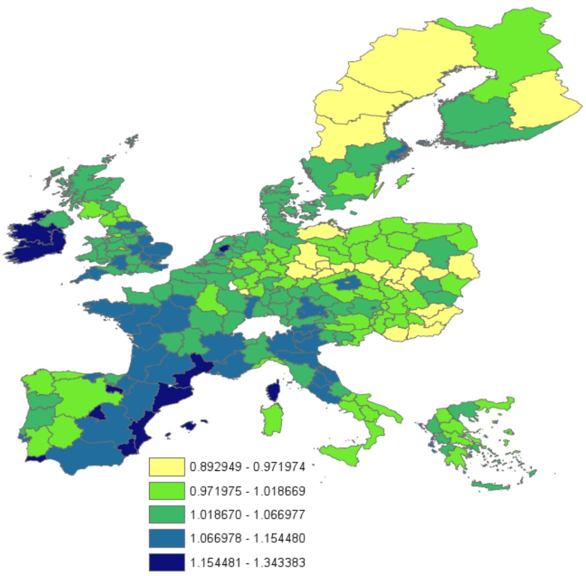


Figure A11: Population Density 1998

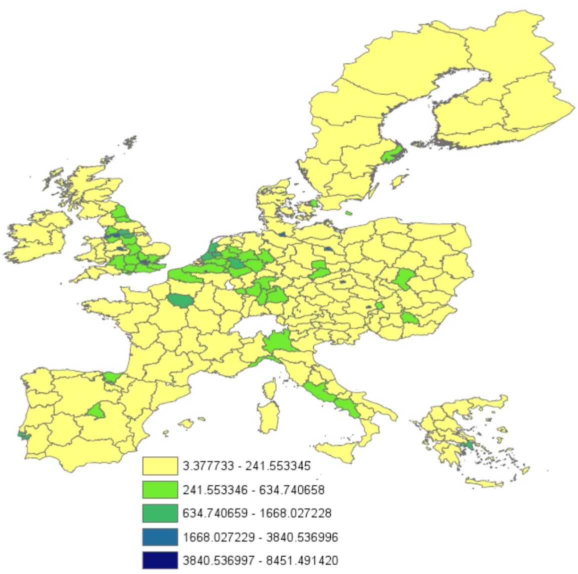


Figure A12: Wage Level 1998

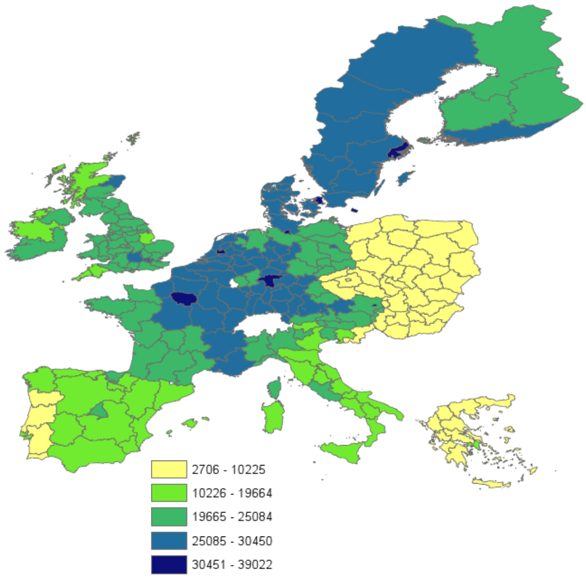


Figure A13: Wage Level 2003

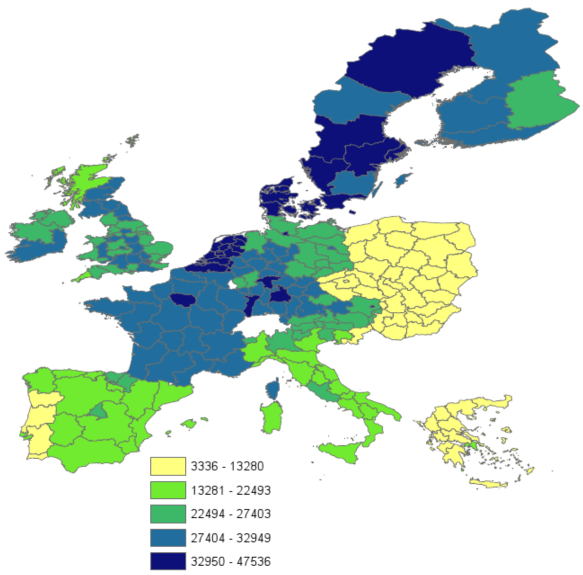


Figure A14: Human Capital, 2007

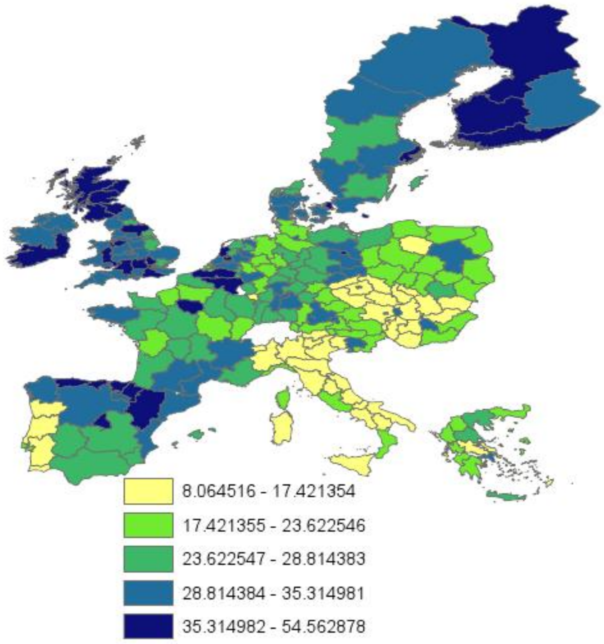


Figure A15: Investment per Worker 1998

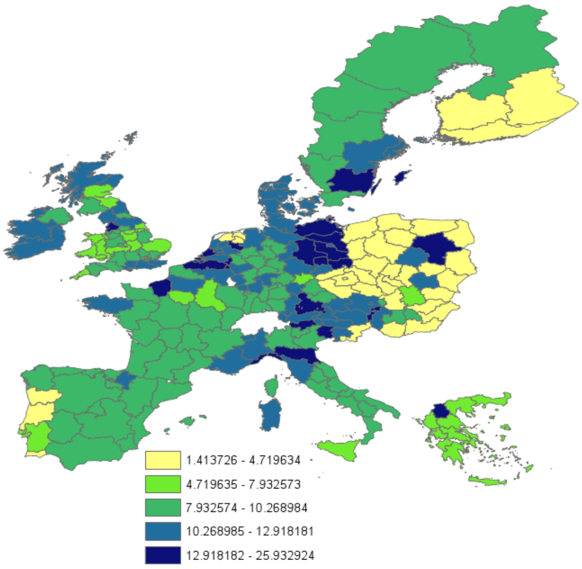


Figure A16: Investment per Worker 2003

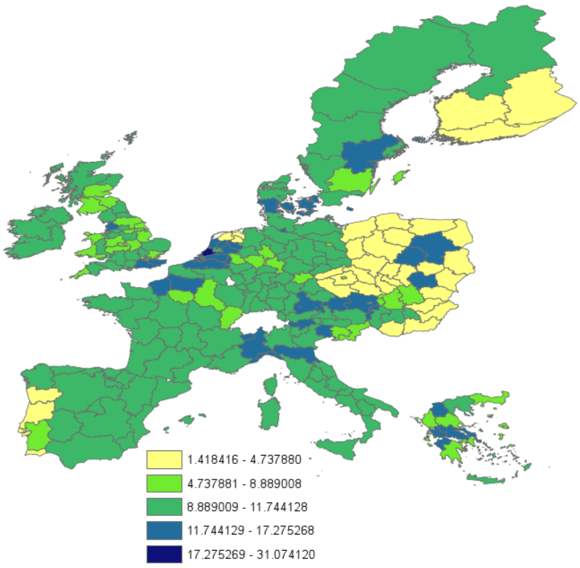


Figure A17: Business R&D

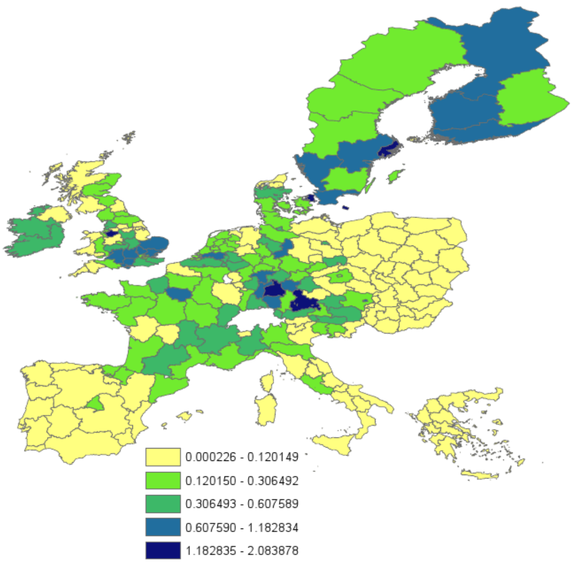


Figure A18: Non-Business R&D

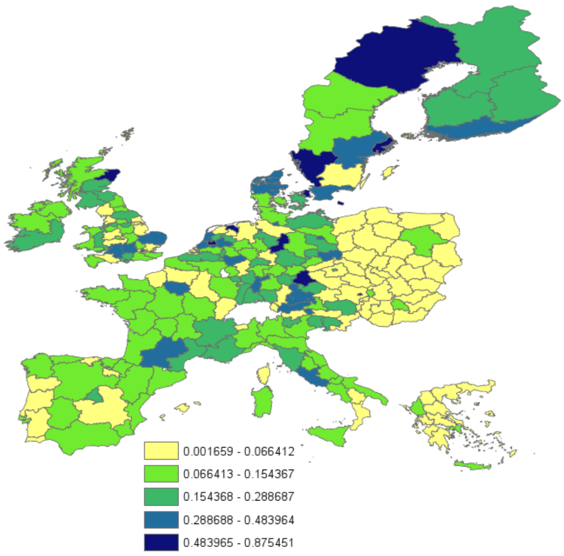


Figure A19: Capital/Labor Ratio 1998

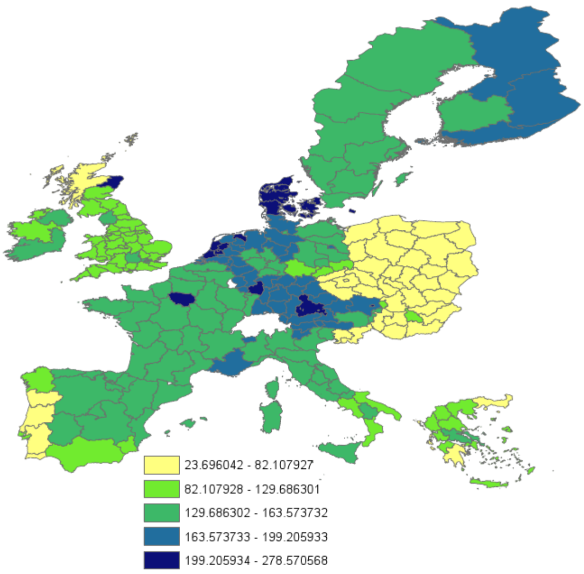


Figure A20: Capital/Labor Ratio 2003

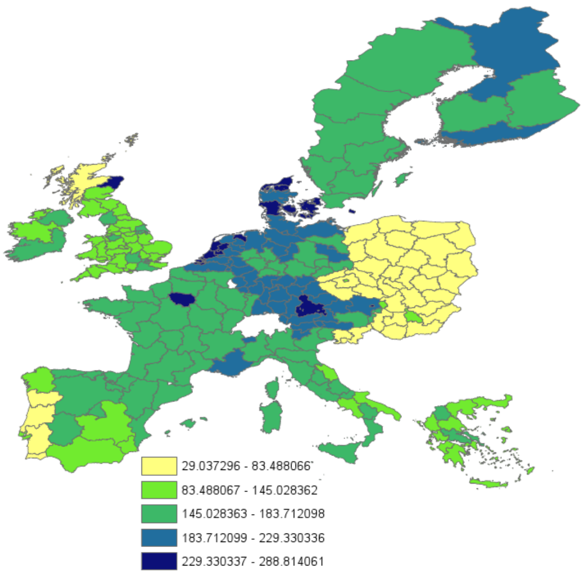
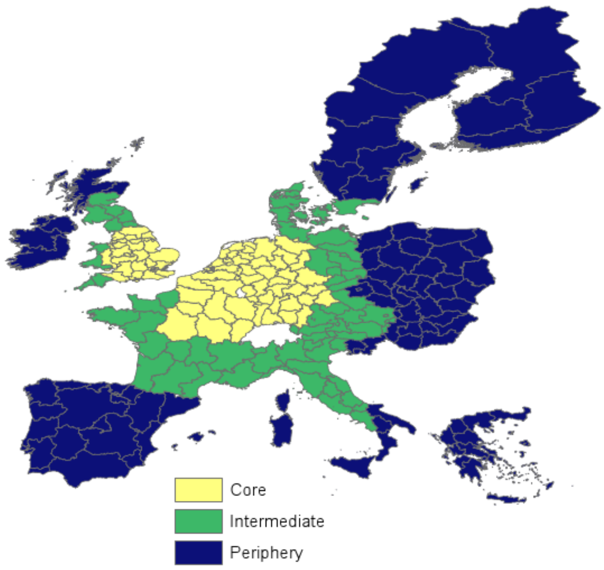


Figure A21: Accessibility dummy



Appendix B: Tables

Table B1: Descriptives for dependent variables

	Minimum	*Maximum*	*Mean*	*Standard Deviation*
Employment growth (1998 - 2008)	0.8722	1.4646	1.1067	0.1047
Employment growth (1998 - 2003)	0.7701	1.2596	1.0430	0.0779
Employment growth (2003 – 2008)	0.9173	1.3827	1.0610	0.0539
Employment growth, corrected for hours worked (1998 – 2008)	0.8770	1.4276	1.0805	0.1000
Labor productivity growth (1998 – 2008)	0.9214	1.7708	1.1638	0.1515
Labor productivity growth (1998 – 2003)	0.7821	1.5510	1.0823	0.1001
Labor productivity growth (2003 – 2008)	0.9171	1.3778	1.0731	0.0652
Labor productivity growth, corrected for hours worked (1998 – 2008)	0.9151	1.7466	1.1918	0.1530
Unemployment growth (1998 – 2008)	0.3235	2.4520	0.8915	0.3240
Unemployment growth (1998 – 2003)	0.3468	2.1395	0.9850	0.3035
Unemployment growth (2003 – 2008)	0.2500	2.6364	0.9511	0.3335

Table B2: Descriptives for main independent variables

	Minimum	*Maximum*	*Mean*	*Standard Deviation*
Related variety (unweighted)	0.5878	2.6705	1.9040	0.4302
Unrelated variety (unweighted)	3.4485	5.1562	4.6918	0.3121
Specialization (unweighted, 2digit)	0.1001	1.6968	0.3357	0.2435
Specialization (unweighted, 3digit)	0.2006	1.9908	0.4917	0.3017
Specialization (unweighted, 4 digit)	0.2652	2.3622	0.6100	0.3410
Related variety (weighted)	0.1357	2.3591	1.2767	0.4306
Unrelated variety (weighted)	2.2725	4.9548	4.0716	0.5193
Specialization (weighted, 2digit)	0.2336	2.4717	0.8379	0.3465
Specialization (weighted, 3digit)	0.4379	3.8249	1.2804	0.4638
Specialization (weighted, 4digit)	0.5593	4.0647	1.6232	0.5156

Table B3: Descriptives for other independent variables

	Minimum	*Maximum*	*Mean*	*Standard Deviation*
Number of firms	315	424,791	42,040.5086	51,996.3509
Number of firms (weighted by revenue)	40	13,847	1,160.9762	1,740.1958
Population growth (1998 – 2008)	0.8930	1.3434	1.0383	0.0631
Population growth (1998 – 2003)	0.9435	1.1991	1.0149	0.0308
Population growth (2003 – 2008)	0.9465	1.1531	1.0233	0.0354
Population density (1998, inhabitants per km^2)	3.3777	8,451.4914	347.7714	842.8098
Population density (2003, inhabitants per km^2)	3.3162	9,013.7067	354.5692	875.6132
Human capital (% higher educated)	8.0645	54.5629	26.37427	8.2079
Investment (1000s euro per worker, 1998)	1.4137	25.9329	8.9864	3.5192
Investment (1000s euro per worker, 2003)	1.4184	31.0741	9.3913	3.5938
Business R&D (1000s euro per capita)	0.0002	2.0839	0.2438	0.3165
Non-business R&D (1000s euro per capita)	0.0017	0.8755	0.1478	0.1520
Wage (euro, 1998)	2,706	39,022	19,745.1624	8,793.1404
Wage (euro, 2003)	3,336	47,536	23,518.3419	9,846.0958
Capital/labor ratio (1000s euro, 1998)	23.6960	278.5706	134.2626	50.6538
Capital/labor ratio(1000s euro, 2003)	29.0373	288.8141	152.3189	55.4131

Table B4: Correlation matrix

	Related Variety	*Unrelated Variety*	*Specialization 2digit*	*Specialization 3digit*	*Specialization 4digit*	*Related Variety (weighted)*	*Unrelated Variety (weighted)*	*Specialization (weighted, 2digit)*	*Specialization (weighted, 3digit)*	*Specialization (weighted, 4digit)*	*Employment (1998)*	*Labor Productivity (1998)*	*Unemployment (1998)*	*Number of firms (weighted)*	*Number of firms*	*Population growth (1998 – 2008)*	*Population density (1998)*	*Human capital*	*C/L ratio*	*Wage*	*Investment (1998)*	*Business R&D*	*Non-business R&D*	*Accessibility – Intermediate zone*	*Accessibility – Periphery*	*New Member States*
Related Variety	1																									
Unrelated Variety	,390	1																								
Specialization 2dig.	-.041	-,679	1																							
Specialization 3dig.	-.065	-,608	,965	1																						
Specialization 4dig.	-.081	-,570	,940	,992	1																					
Related Variety (weighted)	,480	,353	-,197	-,228	-,246	1																				
Unrel. Variety (weighted)	-.011	,353	-,336	-,325	-,314	,577	1																			
Specialization (weighted, 2dig)	-,175	-.107	,306	,325	,353	-,595	-,525	1																		
Specialization (weighted, 3dig)	-,136	-.051	,278	,285	,316	-,544	-,476	,936	1																	
Specialization (weighted, 4dig)	-.067	.003	,253	,259	,294	-,439	-,415	,864	,961	1																
Employment (1998)	,141	,224	-,401	-,441	-,478	,460	,374	-,599	-,643	-,682	1															
Labor prod. (1998)	-,336	-,245	-,246	-,250	-,250	-,304	-,152	-,194	-,244	-,290	.052	1														
Unemployment (1998)	,360	,374	-,307	-,341	-,360	,482	,371	-,432	-,452	-,445	,803	-,127	1													
# Firms (weighted)	.049	.031	-,279	-,316	-,361	,473	,410	-,685	-,739	-,790	,866	,210	,633	1												

# Firms	-,122	.098	-,530	-,590	-,645	,249	,297	-,541	-,589	-,661	,767	,324	,475	,856	1											
Pop. growth (1998 – 2008)	-,117	-,218	-.096	-,148	-,162	-.065	-.039	-,134	-,128	-,167	.019	,340	-.108	,222	,275	1										
Population density (1998)	-,263	-,203	-,139	-,186	-,216	-.014	.017	-,242	-,296	-,365	,537	,182	,279	,459	,500	.051	1									
Human capital	-,289	-.083	-,276	-,309	-,319	-,207	-.028	-,155	-,195	-,209	,187	,474	.099	,223	,333	,209	,249	1								
Capital / Labor ratio	-,194	-,284	-,162	-,175	-,192	-,298	-,242	-,123	-,170	-,208	-.018	,888	-,166	,138	,273	,282	,133	,385	1							
Wage	-,287	-,173	-,356	-,369	-,378	-,289	-,167	-,191	-,243	-,282	,114	,971	-.080	,205	,420	,283	,221	,516	,879	1						
Investment (1998)	-,168	-,118	-,255	-,276	-,275	-.105	-.105	-,200	-,239	-,253	.066	,673	-.070	,183	,224	,189	,142	,334	,680	,677	1					
Business R&D	-,202	.016	-,490	-,494	-,517	-.102	-.022	-,297	-,347	-,394	,371	,666	.037	,393	,639	,217	,304	,488	,620	,743	,499	1				
Non-business R&D	-,231	-,131	-,241	-,241	-,257	-,149	-.063	-,276	-,336	-,400	,278	,666	.069	,401	,457	,210	,282	,557	,623	,687	,521	,615	1			
Accessibility – Intermediate	-.016	-.004	-,138	-,167	-,158	.045	.058	-.051	-.050	-.039	.026	,238	-.023	.061	.087	.053	-.050	-,137	,255	,242	,319	,166	,175	1		
Accessibility – Periphery	,277	,170	,279	,318	,324	,164	,123	.102	.107	,126	-,207	-,627	.066	-.104	-,324	-.071	-,427	-,172	-,559	-,668	-,480	-,565	-,378	-,494	1	
New Member States	,300	,267	,136	,138	,130	,296	,130	,160	,209	,240	.052	-,911	,141	-,128	-,151	-,338	-.051	-,374	-,792	-,858	-,604	-,458	-,564	-,213	,490	1

note: significance is indicated by color: 0.10, 0.05, 0.01; -0.10, -0.05, -0.01