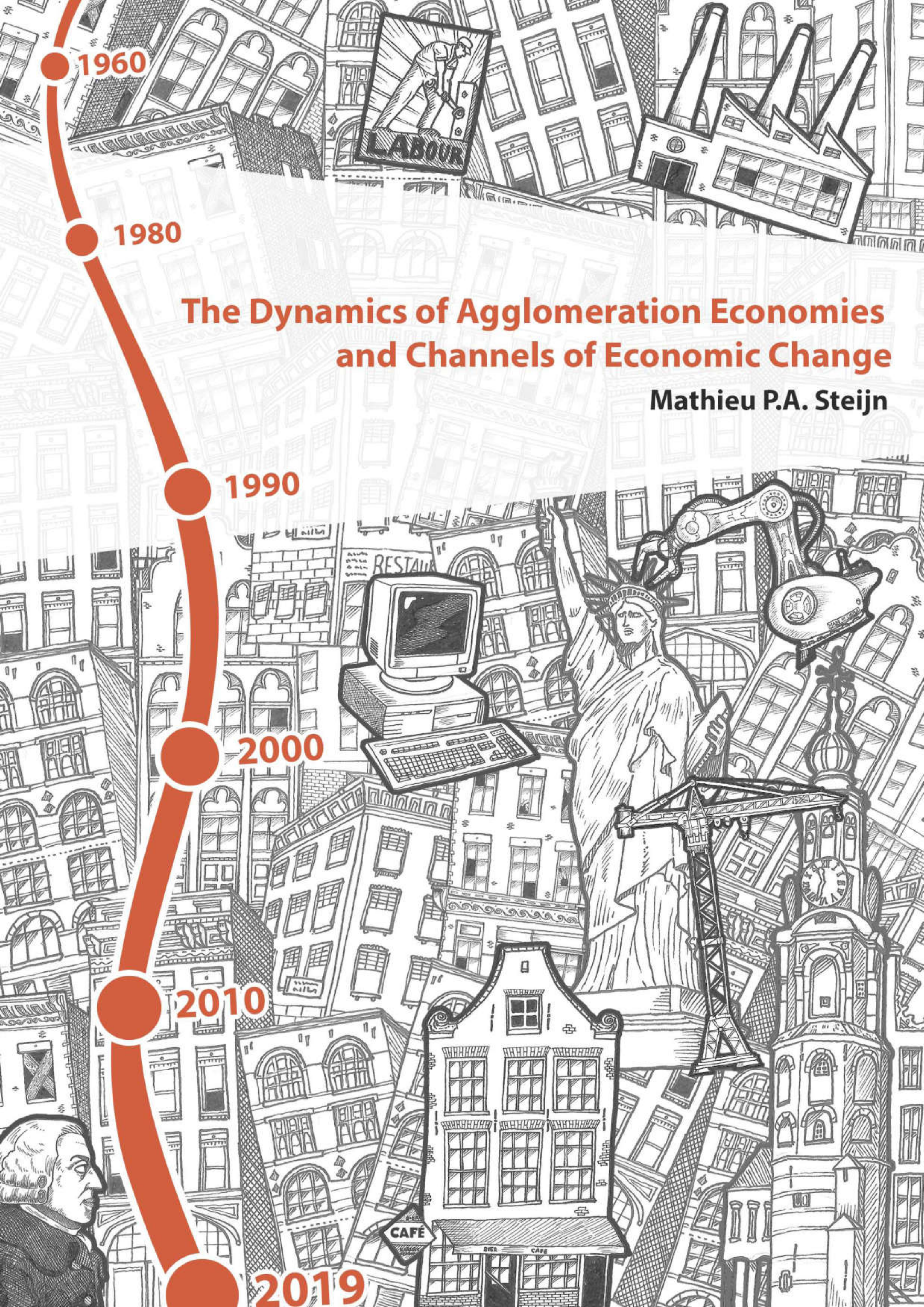
The Dynamics of Agglomeration Economies and Channels of Economic Change

Mathieu P.A. Steijn



The dynamics of agglomeration economies and channels of economic change

Mathieu P.A. Steijn

Doctoral dissertation

Steijn, M.P.A..
The dynamics of agglomeration economies and channels of economic change
ISBN: 978-90-393-7481-8
Layout by the author using LaTeX

Printed on 100% recycled paper by: Ruparo
Published by: Universiteit Utrecht
Cover designed by: Meyourim Benkemper en Thomas van Sonsbeek (Chaotic Bastards).

The dynamics of agglomeration economies and channels of economic change

De dynamiek van agglomeratievoordelen en mechanismen van economische verandering

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op:
vrijdag 3 juni 2022 des avonds te 6.15 uur

door

Mathieu Paul Adriaan Steijn

geboren op 2 februari 1989 te Amsterdam

Promotoren:	Prof.dr. R.A. Boschma
	Prof.dr. F.G. van Oort
	Prof.dr. H.R.A. Koster
Copromotor:	Dr. P.-A. Balland

The research reported in this thesis was partly funded by grant 438-13-406 from JPI Urban Europe.

Committee: Prof.dr. E.L. Glaeser
Prof.dr. C. Castaldi
Prof.dr.ir. E.S. van Leeuwen
Prof.dr. H.L.F. de Groot
Prof.dr. N.D. Lee

"*Parler de ce qu'on ignore finit par vous l'apprendre.*"[1]
(Camus, 1953, préface)

[1]English translation: Discussing what we do not know will end up teaching it to us. Nederlandse vertaling: Bespreken wat we niet weten zal het ons uiteindelijk leren.

Contents

Preface

This thesis embodies the journey I made as a Ph.D. student. It not only reflects the developments in research I made over the past seven years but also those in my personal life. When I started in early 2015, I came fresh out of the STREEM M.Sc. programme of the Vrije Universiteit Amsterdam (VU) and was hoping to further develop my research skills to the example of the authors of the works I admired during my studies, like Gilles Duranton and Ed Glaeser. Also, Amsterdam, my city of birth, had undergone great changes during my life time and was recovering after the 2008 crisis and the population loss of the 1960s-1980s. I was curious what forces made that some cities exerted a growing force of attraction on people while other cities did not and was not yet satisfied by the answers learned in class.

At Utrecht University, I found the opportunity to start a Ph.D. trajectory with Frank van Oort, Ron Boschma, and Pierre-Alexandre Balland on regional resilience, the ability of regions to withstand or overcome crises. With this position I got introduced not only to a slightly new topic but also a new field: Evolutionary Economic Geography (EEG), which was not covered in my spatial economics Master programme at the VU nor in my urbanism Master programme at l'institut d'urbanisme de Paris. Even though there was a shared topic of interest the difference in paradigms was larger than I expected and misunderstandings back and forth were numerous in the beginning and I had to learn a whole new vocabulary in some respects. Looking back, now I have overcome these epistemological barriers, I conclude that there are many places of common ground and respective advantages to each approach, to which I have dedicated section 5.3.2 in the conclusion.

The first project of the Ph.D. on changes in trade between European regions during the 2008 crisis was a bit of a false start as my first major finding was the rather disappointing conclusion that the datasets on trade were not sufficiently accurate to properly measure this effect. However, in the mean time I got picked up by Sergio Petralia, Pierre-Alexandre Balland, and David Rigby to help with their project on gathering the geographical locations of historical patents. This project was a great start to get familiar with EEG and in mastering R. In the end, that knowledge resulted in two scientific papers, one on technological diversification of cities during the great historical crises with Pierre-Alexandre Balland, Ron Boschma and David Rigby, see Chapter 4, and the other on methodological improvements to calculate relatedness, a fundamental concept of EEG, see the single-authored additional Chapter A. The development of R skills accumulated over time and resulted in, among others, skills in data cleaning/handling/analysing, (contributions to) R-packages, a personal website with 3D interactive maps, the ability to scrape data from websites and digitalized documents, and to automate emails for teaching purposes.

The collaboration on patent data also led to me joining the project of Sergio Petralia, Pierre-Alexandre Balland, and David Rigby on looking into the relation between patented inventions and agglomeration advantages over the large time span one could examine with geolocalised patent data. The idea to take the time and look into many possible angles to write something that could make impact was fitting for the exploratory phase I was in and my academic ambitions. The exploration of different angles meant that I got introduced to innovation studies, which is strongly linked to EEG, and that I witnessed how, over time, complexity studies, with its origins in physics, became more and more incorporated in the line of research at Utrecht University. This last part also led to César Hidalgo and Crístian Jara-Figueroa joining the project, which culminated in a publication in *Nature human behavior* that forms the basis of Chapter 3.

My interest in agglomeration advantages from the perspective of urban economics was not abated by the new perspectives I learned. Therefore, I set out on another project around 2016 with Frank van Oort and Hans Koster of the VU, who later also became one of my promoters.[2] Its goal was to map and understand the changes in agglomeration determinants as classified by the influential Alfred Marshall, see Marshall (1890). From the beginning, I felt that this paper, which became Chapter 2, was a great opportunity to extend the work of the authors I read during my studies and develop myself to possibly write something that would make an impact in the field.

As economic geographers know, most of the intellectual progress is not done by just reading the articles of others but by engaging in conversation with them. This is very much ingrained in the Utrecht approach to doing research through international collaboration. This had the downside that it was sometimes rather quiet at the Uithof in Utrecht but the upside that it allowed me to see and connect to many inspiring

[2]Funnily, at that time Hans sometimes joked that I was already too strongly influenced by the "Utrecht" way of thinking.

scholars. At seminars and conferences, I could not only listen to the authors whose work I had read but also talk to them in the breaks! During my Ph.D., I visited many research institutes, conferences, and Ph.D. schools, which brought me on short trips to places like Dublin, Toulouse, Bordeaux, San Francisco, Boston, Barcelona, New York, Cambridge (U.S.A.), and Groningen and longer stays in Medellín, Los Angeles, Buenos Aires, Rio de Janeiro, and Cambridge (U.S.A.). In the visits to these South American cities I could combine my research activities with spending free time on a continent that is dear to me. The U.S.A. visits were more based on the Utrecht network. Notably, with one or more supervisors, I spent over a month at UCLA with David Rigby after the AAG 2016 conference in San Francisco and several weeks at the Harvard Kennedy School (HKS), the home of former Utrecht researchers Frank Neffke, Matté Hartog, and Dario Diodato, after the AAG 2017 in Boston. These trips were not only characterised by hard work but also fun leisure time like night outs (I will likely never forget my first moment seeing famous professors dance), swimming, (foot)ball games, and hiking trips.

By building on this network I got to develop my own. At the AAG 2017 in Boston I got to meet Gilles Duranton, who was really helpful and provided me with large amounts of data for my quest on changes in Marshall's agglomeration determinants. At HKS I got to meet Bill Kerr of the Harvard Business School (HBS), who was one of the authors of Ellison et al. (2010), which was the main inspiration for that article. He invited me for a five-month Visiting Research Fellowship at his department in 2018, which I consider as one of the highlights of my Ph.D. Harvard is a breathtaking environment for a young scholar. There is a large concentration of inspiring scholars working in the direct vicinity, which also includes MIT and other renowned universities. Furthermore, most other researchers in the world that do groundbreaking work want to present their work in that place. This means that there are tons of seminars and lunches every week to listen to these scholars or occasions to meet them. One could go listen to former presidents and ministers or academic "superstars", like Thomas Piketty or Daron Acemoglu, live. At HBS and HKS I also connected to researchers in labour economics. Funnily, I met Anna Salomons, a researcher in labour economics of the Utrecht School of Economics, in Cambridge! Next to that, it was always easy to bike the few kilometres to the MIT media lab to hang out and learn on developments in complexity science with several of the co-authors of Chapter 3.

Inspired by these meetings, I set out to gather data from the many sources, including hundreds of pages of scanned historical documents, needed to calculate the importance of each of Marshall's agglomeration determinants between 1970 to 2014. This also included finding the best method in accommodating the many administrative changes in industry codes and county layout over time. The results proved interesting and inspired questions on what was underlying the trends in these agglomeration determinants. A topic that interested not only Bill Kerr but also Ed Glaeser, another author of Ellison et al. (2010) and one of the authors that inspired me during my studies. I was very grateful for the several precisely timed 15-minute meetings with him. Although these meetings were relatively short due to his busy agenda, I was impressed by the sharp, insightful and eloquent feedback I would receive every time.

More than I could have hoped for at the beginning of my Ph.D., I found that the article in progress was getting attention of many renowned scholars with useful feedback. In particular, I was honoured to present at HKS, MIT media lab, the NBER, and the UEA congresses in New York and Amsterdam in sessions with authors like William Strange, Stuart Rosenthal, and Gabriel Ahlfeldt.

On the other hand, this success also came at a cost. Travelling around the world and working hard in an inspiring environment also meant being far away from friends and family. I was lucky enough to always make new friends relatively easily, wherever I came, but I also came to understand that it takes time to develop good friendships. As a result, I started to appreciate more my life in Amsterdam, where I'm strongly rooted and life was less work-oriented.

Also, after 2018 my research time became less abundant. I originally started on a three year Ph.D. contract and managed to extend that to four years by taking up some teaching activities. However, by 2019, I had to dedicate almost 0.8FTE to teaching to finance my research activities, although I am very thankful to Fred Toppen and Bouke van Gorp for accommodating more research time in certain time periods. I very much enjoy teaching and I see it as critical in an academic career and in universities in general. In particular, I enjoyed the economics of cities course I coordinated and the dynamic atmosphere with the many (young) colleagues involved in teaching. At the same time, this also took a lot of time, especially the courses that were further away from my expertise like those in human geography, which was detrimental for my research progress and work-life balance.

Furthermore, the projects on technological diversification during crises and on Marshall's agglomeration determinants stalled around the same time. The former required the finding, understanding and programming of more rare econometric methods fit for the research set-up. While with the latter it was possible to apply the methodology by Ellison et al. (2010) to measure the changes in Marshall's agglomeration determinants over time but it was hard to find an explanation for these changes that survived the test of rigid econometric methods and the standards of the economic literature I was aiming for. Most of the literature on urban economics is based on transportation costs but the standard measures in this line proved irrelevant. Measures on the literature on manufacturing, covering the industries in the study, gave some insight but were hard to fit in the line of literature and sometimes yielded seemingly counter-intuitive results. Interestingly, answers came from labour economics and innovation studies, as I discuss in greater length in the introduction of this thesis. These studies describe channels of economic change that often come about during technological revolutions, which also proved to be the common denominator that tie together the chapters of my dissertation on agglomeration determinants.

To improve on these ideas, I got a few months without teaching and a visiting position at Northeastern University in 2020 with Gregory Wassall. Here I could reconnect with Bill Kerr, Ed Glaeser and also meet David Autor, a leading author in labour economics. They regularly publish in top journals, like the *American Economic Review*, and some of them saw a chance for my paper to make the cut, which I was willing to take. Then, the start of the corona pandemic meant an abrupt end to my stay in Boston but with their tips in my suitcase I headed home to finish the paper.

The corona pandemic, especially in the beginning, was a confusing time. With the restrictions on physical interaction, teaching became much more time-intensive and less rewarding and social activities were greatly limited. Despite all this, I found inspiration in the time I got for sport, hobbies like fixing up a 1980s racing bike, reading, and my research. In the summer of 2020, when lock-down measures were lifted, I took the 1980s racing bike on a bicycle trip from my *stamkroeg* de Spuyt to the Alpe d'Huez and back, while visiting friends and family on the way. When in the early winter there was a new lock down it was harder to remain optimistic. The thought of spending another few months at home was hard. Also for my students, as we went back to online teaching. Nevertheless, I finished Chapter 2 with my co-authors and we submitted it to the *American Economic Review*. I knew chances were not great but was hopeful that the hard work of incorporating virtually every advice of some of the most influential researchers would pay off. The paper made it to a first round of feedback, which is already a big sign of appreciation for which I feel honoured, but was then rejected. Shortly after it was rejected by another general interest journal. However, in the Journal of Urban Economics, which is, as the name suggests, more strongly dedicated to my field, I found more valuable feedback and an outlet that makes me feel that my work is very much appreciated.

Nevertheless, the earlier rejection set-backs that are common in academia but were new to me were tougher than I expected, likely also because they occurred during the isolated times of a corona lock-down. The one thing that makes the ups of an academic researcher higher is also the one that makes the downs lower. The possibility to pursue one's personal research interests and develop oneself to meet these research interests also ties ones personal happiness more firmly to the outcomes of one's professional endeavours than any other job I had before.

Furthermore, I came to realise that there were aspects to urban growth that I was not sufficiently paying attention to in my thesis. I had come to understand much better the questions I posed myself at the beginning of my Ph.D. on the local factors that contribute to growth and how their relevance changes over time but felt that I and most of the literature I was building on were overlooking how the benefits of that growth was distributed over people. At the start of my Ph.D. in early 2015, Amsterdam was a city that was recovering from the 2008 crisis and, longer ago, the population loss during the 1960s-1980s. With housing prices that were slowly increasing after having been relatively stagnant for almost 5 years. By the end of my Ph.D., housing prices had skyrocketed. Amsterdam was a city that was increasingly unaffordable for a large number of people and it was becoming clear that its growth had not benefited a considerable part of its population. The works I read of Thomas Piketty, see Piketty (2013) and Piketty (2019), Mariana Mazzucato, see Mazzucato (2019), labour economists, and the class materials of my master in urbanism and for the courses I taught in globalisation and human geography showed me that there were factors and topics that were not sufficiently taken into account in my line of research, such as distributional issues, governmental policies and the possibility of different societal groups to access the local sources of growth I was investigating. On the other hand, these other lines of study were missing many points made in Economic Geography. Also, it was certainly not necessary for the chapters to tackle all of these issues as limiting the scope by specialising is necessary to make progress, as also discussed in Chapter 3, but to understand the importance and potential societal impact the results can have it is necessary to understand the role of agglomeration mechanisms in society.

Therefore, I felt that my Ph.D. was not at its end without understanding these aspects and how research is used in policy-making to put my thesis in the right context and to find a line of future research in which I could possibly make contributions. This is why I dedicated the introduction and conclusion of this thesis to making the links between the scientific fields of the different epistemological communities on which I based my research and identifying other relevant perspectives on cities that could be incorporated to build a more complete image of economic development that can improve information for decision making. These building blocks formed the starting point for the post-doctoral trajectory I started in early 2022 on social mobility and collective facilities in the Amsterdam area at the VU, which also allows me to work closer to home.

All in all, I look back on a journey that took me to many places around the world, allowed me to connect to new persons and new ideas, and in the end allowed me to develop myself in a way I know is not accessible for many others on this planet and for which I'm very grateful. In the end, research is not only about getting to know the unknown but just as much about getting to know what still remains unknown. Therefore the challenge is perhaps equally large in finding the answers to questions as in finding the right questions. Like the topic of my thesis, this Ph.D. trajectory taught me that goals in research or even in life are *dynamic* and provided me with many useful experiences to continue this trajectory wherever it may take me.

Acknowledgments

This Ph.D. journey would not have been possible without the support and feedback of many individuals. First of all, I would like to thank my supervisors: Ron, Frank, Hans, and Pierre-Alexandre with whom it was pleasant to be able to speak either in my father's tongue, Dutch, or in my mother's tongue, French, and without whom this work would not have been possible. Next to the valuable feedback, I want to particularly thank Ron for the open discussions on other points of view and dedication to academic research; I want to thank Frank for his relentless enthusiasm for research; I want to thank Hans for his numerous feedback and availability, even when time zones were a hurdle; and I want to thank Pierre-Alexandre for encouraging me to think big and also for showing how to enjoy the perks of academic (travel) life. As a scholar in economic geography, I will proceed by detailing my acknowledgements by geographical location.

Amsterdam area

I want to thank my parents, Jack and Véronique, for being there to share in the ups and downs that come with a Ph.D. journey. My thanks also go out to the friends who I have known since my high school, *het Ignatius*. Even though since then our individual paths have taken us around the globe, I appreciate how often we make our paths cross again. For example, when many of you came over to Istanbul when I was on the way for a conference in 2016. In particular, I want to thank: Jurre; Thomas; Leonne; Jane; Joshua; Steven; Guy; Joyce; Leonie; Ruben; and Joanne. Special thanks to Jurre for his support during the ceremony as a paranimph and sharing in a part of this Ph.D. endeavour by exchanging our experiences at the same faculty at the VU.

My life as a student also brought me a lot of friends whose presence during the research process was very much appreciated. The *donderdorst* borrels, parties, and festivals were a welcome change to the Ph.D. struggles. In particular, I want to thank: Bertram;

Hannes; Stef; Mart; Puck; Gaby; Alex; Roy; Claudia; Tesse; Isis; Welmoed; Marcel; Laura; Luuk; Mara; Marleen; and Wilmer. Special thanks to my (former) roommates Stef and Hannes and also to Bertram for sharing in this Ph.D. experience at the VU.

A big source of support is also Café de Spuyt and the individuals that make up its microcosm. The comfort of a place one can always go to for distraction or discussion is invaluable and the effort so many of you have put in to help with the book production and party organisation after the Ph.D. defence has touched me. In particular, I want to thank: Kasper; Fabian; Steef; Emile; Esin; Nikkie; Debbie; Job; Thomas; Wouter; Susana; Trieu; Kana; Danny; Frans; Meyo & Thomas (who also designed the cover of my thesis); Mikko (who also designed the illustration in Figure 1.1); and Bregje. Extra thanks to Kasper for the long-term friendship and his support as paranimph during the ceremony and to Fabian for transforming the bar to accommodate the reception and party. My thoughts also go out to Steef, who would have very much loved to have witnessed the completion of my Ph.D. journey after being of great support in the run-up and first steps of the endeavour.

I also want to thank my other family members. My brother, Cedric, and Rianne but also *oma* Jeanne; Frank; Frank; Frank; Hanneke; Jera; Simon; Jelmer; Ilse; Eelco; Marlies; Isabel; Duco; Eveline; and Daniëlle. Also thanks to the many friends I made outside of the mentioned environments so far, in particular: Thomas; Sandra; Marco; Suus; Eva; Siënna; Sissy; Leonieke; Ollie and Piet of Café Brel; and Michel, Henk, Chaim, Willem, and the others of SBK.

At the Vrije Universiteit Amsterdam, I have felt at home ever since my high school visit to my later study program. I very much appreciate the guest researcher position I received at the end of my studies at the department of Spatial Economics and kept on during my entire Ph.D.. Those that were once my teachers became my colleagues and then my friends. In particular, I want to thank: Hans; Jos; Erik; Hadewijch; Eveline; Thomas; Maurice; Henri; Eduardo; Jesper; Jamie; Francis; Devi; Joris; Zhaoxin; Leon; Diego; Sacha; Dan; Jakob; Benjamin; Hedda; Elfie; Jenny and Maureen.

Utrecht
Although only some 40km from Amsterdam, Utrecht was a rather new place for me. A new place and new university was a great fit for a fresh start into this Ph.D. adventure. It did not take long to feel at home at the Human Geography & Spatial Planning department. I have much appreciation for the welcoming atmosphere and the activities, like the Ph.D. Sinterklaas and the Research Days, that were organised. In particular, I want to thank: Ron; Frank; Pierre-Alexandre; Marianne; Kyri; Sergio; Nicola; Teresa; Nynke; Marijke; Marielle; Summer; Karlijne; Karlijn; Carolina; Koen; Alje; Jeroen;

Gery; Oedzge; Martijn; Ton; Paulien; Hannah; Mathias; Sara; Emilinah; Ben; Peter; Leo; Véronique; Karin; Bianca; Kirsten; James; Egbert; Evert; Benjamin; Eduardo; Dongmiao; Ineke; Dario; Bo; Sofie; Ruud; Thomas; Stef; Leon; de Brammen; Patrick; Tessa; Tara; Anne; Anna; Sara; Eoin; Wiebe; Mirjam; Michiel; Stephan; Bouke; Fred; Geertje; Marleen; Veerle; and Ron.

Paris and the rest of France
Looking over the border my first thoughts go to Paris. Where part of my family lives and numerous friends that I have met during my Master programs at l'institut d'urbanisme de Paris. Although some 450km from Amsterdam, it feels like my second home. Remote working from my grandmother's place and catching up with friends in the area offered a nice change of scenery during the elaboration of my thesis. The occasional (road) trip to other parts of the country and Costa Rica during the holidays were also very welcome. Also a thank you to my other family members and friends in other parts of France, who were always willing to welcome me when I was in the area. In particular, I want to thank: *mamie* Marie-Paule; Diane; Tom; Pauline; François; Pierre; Philippe; Elise; Pauline; Juliette; Christine; Marouan; Stéphanie; Laurent; Liselotte; Philémon and Dimitri.

South America
During my Ph.D. journey I also got to know a few other remarkable places outside of Europe. Next to short trips for summer schools and conferences, I got to spend some longer time periods in South America, namely, in Medellín, Buenos Aires, and Rio de Janeiro. I want to thank Alejandro, Alejo, Thomas, Carlos, Alexandra, Alejandra, and Veronica for receiving me well at the EAFIT universidad in Medellín and Anabel for the reception at the CENIT in Buenos Aires. For taking me up in their group of friends and making me feel more than at home in Rio de Janeiro, I want to thank: Ana Clara; Bruna; Clara; Filipíu; Carol; Enrico; Zé; Diogo; Mari; and Isabelle.

United States of America
I also had the pleasure to spend longer time periods at renowned research institutes in the United States of America, namely, at UCLA, Harvard University, and Northeastern University. I want to thank David, Frank, Sin Gin, and Chris for receiving me well at UCLA. Special thanks for Chris for taking me in as his room mate during my visiting period. I want to thank Frank, Matté and Dario for receiving me well at the Harvard Kennedy School; Bill, Tommy, Jasmina, and Yosub for receiving me well at the Harvard Business School; César, Isabelle, and Crístian for receiving me well at the MIT Media lab; and Gregory for receiving me well at Northeastern University. Special thanks to Kathleen and Jackson for taking me in and making me feel part of the family during a large part of my visiting time in Cambridge.

Chapter acknowledgments

During the elaboration of each of the chapters I also benefited of many comments and suggestions by other scholars. I am grateful for this valuable feedback.

Chapter 2, "The Dynamics of Industry Agglomeration: Evidence from 44 Years of Coagglomeration Patterns" is based on a paper co-authored with Hans Koster and Frank van Oort. For this chapter, I would like to thank, in particular, Bill Kerr, Ed Glaeser, Frank Neffke, Dario Diodato, and the editor of the Journal of Urban Economics, Nathaniel Baum-Snow, but also David Autor, Pierre-Alexandre Balland, Ron Boschma, Jasmina Chauvin, Gilles Duranton, Koen Frenken, Henri de Groot, César Hidalgo, Jos van Ommeren, Anna Salomons, Michael Storper, Will Strange, two anonymous referees, and participants at seminars at NBER, Harvard University, MIT Media Lab, Utrecht University, Vrije Universiteit Amsterdam, UEA New York and Amsterdam, GEOINNO2018, and ERSA Groningen for useful comments on this paper. I also want to thank Gilles Duranton, Sergio Petralia, and Bill Kerr for providing data and Bill Kerr and Frank Neffke for receiving me as a visiting researcher at, respectively, the Harvard Business School, and the Harvard Kennedy School during the elaboration of this chapter.

Chapter 3, "Complex economic activities concentrate in large cities" is based on a paper co-authored with Pierre-Alexandre Balland, Sergio Petralia, Crístian Jara-Figueroa, David Rigby, and César Hidalgo. For this chapter, I would like to thank Ron Boschma and Koen Frenken, and participants at seminars at AQR Barcelona, RSA Dublin, Economic Geography PhD School Utrecht 2017,and GEOINNO2016 Toulouse for valuable comments. I would like to thank Önder Nomaler, Koen Franken, and Gaston Heimeriks for providing data and David Rigby and César Hidalgo for receiving me as a visiting researcher at, respectively, UCLA, and the MIT media lab during the elaboration of this chapter.

Chapter 4, "Technological diversification of U.S. cities during the great historical crises" is based on a paper co-authored with Pierre-Alexandre Balland, Ron Boschma, and David Rigby. I want to thank Sergio Petralia and participants at GEOINNO2016 Toulouse and DRUID Academy Conference 2016 Bordeaux for valuable comments. I also want to thank David Rigby for receiving me as a visiting researcher at UCLA during the elaboration of this chapter.

Additional Chapter A, "Improvement on the association strength: implementing a probabilistic measure based on combinations without repetition." is based on a single-authored methodological paper. I want to thank Ludo Waltman and three anonymous referees for valuable comments.

Chapter 1

Introduction

1.1 Motivation

The unequal distribution of human activities over space is the foremost *raison d'être* of urban economics and economic geography. A key focus in understanding (changes in) this spatial distribution is the role of economies of agglomeration, which are the productivity advantages activities experience when colocated. The trade-off that activities face in their location choice between these advantages and localised disadvantages, such as higher land prices, underpins our understanding of the concentration of activities in space since the development of the first cities.

Recently, most attention is given to the so-called great divergence, the growing disparity in welfare over space (Moretti, 2012). Around 1980, a clear structural break in the role of economies of agglomeration occurred. On the one hand, large innovative cities in high-income countries saw their population increase again after decades of decline. This triumph of the city, as labelled by Glaeser (2011), indicates that colocation has become more important, exactly in times of unprecedented reductions in transportation costs, innovations in communication technologies, and global integration (Gaspar and Glaeser, 1998; Storper and Venables, 2004; Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011; Moretti, 2012; PBL, 2015).

On the other hand, not all cities are triumphing, which suggests that this structural break is more complex than just an increase in the relevance of economies of agglomeration. The shining success of innovative cities is eclipsed by the fate of cities hosting less competitive activities. The same forces that allowed New York, Paris, and my birthplace Amsterdam to grow took a heavy toll on cities like Detroit, Amiens, and Heerlen. Around 1980, human capital levels, income levels, life expectancy, and living conditions stopped converging and started diverging between U.S. regions (Berry and

Glaeser, 2005; Moretti, 2012; Austin et al., 2018; Autor, 2019). Similar, albeit less strong, diverging trends have been denoted for French regions (Guilluy, 2014; Catin and Van Huffel, 2019), and Dutch regions (PBL, 2016; De Groot, 2019).[1]

The societal consequences of these locational trends put the understanding of economies of agglomeration in the limelight of attention. The "triumphing" regions are confronted with soaring housing costs, gentrification, segregation and growing local inequality (Florida, 2017; Milikowski, 2018), while the lagging regions are confronted with all kinds of social issues due to income and employment losses. Notably the rise in so-called deaths of despair, *i.e.* the premature mortality, in particular of young men, related to suicide, drug and alcohol abuse, liver diseases, and homicide (Case and Deaton, 2015; Austin et al., 2018; Autor et al., 2019; Pierce and Schott, 2020).

Furthermore, these issues have a clear spatial character and reinforce resentments between societal groups and regions, as exemplified by the regional rise of support for anti-establishment movements of the likes of Trump, Bolsonaro, Le Pen, and Wilders, the gilets jaunes movement in France, Brexit vote, and the rise of anti-gentrification mobilisations. As a consequence, academics working on understanding the spatial distribution of human activities are pressed to step up (Guilluy, 2014; Autor et al., 2016; The Economist, 2016; Florida, 2017; Austin et al., 2018; Le Figaro, 2018; Rodrik, 2018; Rodríguez-Pose, 2018; Storper, 2018; De Groot, 2019).

1.2 Goal and scope of this thesis

Economies of agglomeration are the centrepiece topic of urban economics and economic geography and have been described for centuries. Most cited is Marshall (1890) and his categorisation of agglomerational benefits in three categories: labour market pooling, input-output linkages, and knowledge spillovers. But the existence of some form of economies of agglomeration has already been denoted by Smith (1776) and has even been mentioned by Plato and Xenophon in Ancient Greece (Finley, 1973; Silvermintz, 2010).

[1]The U.S.A. do not have the largest divergence rates as Eastern European countries, notably Poland, have actually experienced stronger divergence rates (PBL, 2016). Storper (2018) notes that between NUTS-2 regions within the European Union inequality fell between 1980 and 1990, due to European integration, but went sharply up afterwards. However, within E.U. countries spatial inequality has generally increased (PBL, 2016; Catin and Van Huffel, 2019). Interestingly, PBL (2016) classifies France as a county where regional convergence occurs, because of the relatively slow growth of the capital region. Catin and Van Huffel (2019) confirms that Paris did not experience a large relative growth in the number of jobs but that this is mostly the case when one compares the growth to other large cities, notably Toulouse, Montpelier, and Nantes. A strong divergence does exist between larger cities and "la France périphérique", as coined and denoted by Guilluy (2014). This also explains why PBL (2016) classifies Belgium as a country of convergence whereas this likely is not the case when considering other large cities.

Despite this long track record and relevance in understanding locational patterns, economies of agglomeration are still largely considered to be a black box (Duranton and Puga, 2004a; Combes and Gobillon, 2015; Davis and Dingel, 2019). The lion's share of research focusses on (changes in) agglomeration patterns and the size of agglomeration effects but there are actually "*only rare attempts to distinguish the various channels behind agglomeration economies*" (Combes and Gobillon, 2015, p.249).

Furthermore, even rarer are the attempts to distinguish *changes* in the relevance of these agglomeration mechanisms, despite the extensive documentation on changes in agglomeration patterns (Ellison et al., 2010, p.1210, Moretti, 2012, p.124, and Storper, 2018, p.255). Most notably, the rising importance of knowledge spillovers has never been demonstrated convincingly even though it is suggested by a large literature to have led to the so-called triumph of the city. Also, mixed and sometimes contradictory suggestions have been made on changes in other agglomeration mechanisms, see McCann and Fingleton (1996); Glaeser and Kohlhase (2004); Duranton and Storper (2008); Glaeser (2011); Moretti (2012); Combes and Gobillon (2015) and Faggio et al. (2017).

Lastly, even less developed than our understanding of (changes in) the mechanisms of economies of agglomeration is our knowledge on the factors significantly interacting with the relevance of these mechanisms (Combes and Gobillon, 2015, p.336). In other words, what are the channels of economic change that make that certain mechanisms become more or less important. Therefore this thesis combines addressing the gaps in both the changes in agglomeration mechanisms and on the identification of possible causes.

A window of opportunity exists to make progress on these issues as the recent online publication of scanned old documents and recently published data combined with novel data collection methods allow for the development of consistent datasets over long time periods. In this thesis, industry statistics, concordance tables, and geolocalised data on activities by type, such as patents by technology class or the number of employees per industry, are gathered, cleaned and harmonised over time. In some cases, it allows me to go back as far as 1790. As a result, it is possible not only to unveil changes in the relevance of agglomeration mechanisms over time but also to test its association to possible channels of economic change.

Within the objective of assessing the extent and reasons for changes in the roles of economies of agglomeration, a number of more narrowly defined research questions are formulated bringing together insights of urban economics and evolutionary economic geography, but also from labour economics, international trade, innovation studies, and complexity theory. The remainder of this section is a general treatise on how these fields are combined and have led to the research questions of this thesis. Readers that are less familiar with one or more of these fields are advised to also read the extended theoretical background in Appendix 1A.

As said, economies of agglomeration are the *productivity gains* one experiences when *colocated*. This suggests that the roles are likely to change when the productivity gains of a certain agglomeration channel or the extent to which one needs to be colocated to obtain these gains changes, *i.e.* the possibility to transport these gains changes.

As economies of agglomeration increase the productivity of an agent (*e.g.* individual, establishment), it is in economics logically presented as a shifter of the production function: $g(A)f(x)$ (Rosenthal and Strange, 2004). Where $f(x)$ is the regular production function, with x being the land, capital, labour, and material inputs, to which the agent adds value in its productive activities, and $g(A)$ the enhancement due to agglomeration effects. Rosenthal and Strange (2004) capture the productivity gains due to agglomeration A of agent j, as follows:

$$A_j = \sum_{k \in K} q(x_j, x_k) a(d^G_{jk}, d^I_{jk}, d^T_{jk}) \tag{1.1}$$

Here, the economies of agglomeration are the sum of the advantages and disadvantages of the interactions with all other agents K. The importance of an interaction with agent k depends on a function of the sizes x of agents j and k: $q(x_j, x_k)$. and on three scopes or distances, as Rosenthal and Strange (2004) label them: geographical, industrial, and temporal. The first is just the physical distance between j and k. The second relates to distance in type of productive activities, the more dissimilar the less likely that colocation is useful argue Rosenthal and Strange (2004), although this will be nuanced further on in this thesis. The latter relates to distance in time. Previous colocation of productive activities may influence the present productivity of an agent. Notable examples are skills learned when in proximity to others and applied at a later time period, see for example Glaeser (1999) and De la Roca and Puga (2017), or the development of new productive activities by regions building on the routines it obtained in the past, see for example Vernon (1960) and Glaeser (2005). When agglomeration economies benefit current activities, *i.e.* $d^T_{jk} = 0$, these are labelled *static* economies of agglomeration, when there is an effect of past activities on more recent activities then these are labelled *dynamic* economies of agglomeration.

The industrial distance and the temporal distance are particularly considered in evolutionary economic geography. From this viewpoint, agents are heterogeneous collections of routines, or in other words, capabilities, obtained at a temporal distance and with varying industrial and geographical distances. In parallel with evolutionary theory, agents develop new routines and shed those that become redundant through survival of the fittest (Nelson and Winter, 1982; Boschma and Frenken, 2006). This thesis builds on ideas from this literature on these distances, in particular on the concept of relatedness to measure industrial distance, but also to acknowledge the existence of other forms of relevant distances such as social, organisational, and institutional, following Boschma (2005), that Rosenthal and Strange (2004) do not consider.

Although evolutionary economic geography and urban economics are in my eyes too much separate lines of research, a topic further discussed in the conclusion, this thesis demonstrates that ideas can be applied and combined when one understands the different perspectives. After all, both consider similar topics, such as the channels of economics of agglomeration, the topic of this thesis. In fact, by combining these insights it became possible to achieve the goal of shedding light on the dynamics of agglomeration economies and the channels of economic change.

A large share of previous research in urban economics and economic geography of the last decades focussed on using more refined data and identification strategies on estimating the productivity gains A over space and its geographical attenuation, or deducing these from agglomeration patterns (Puga, 2010; Combes and Gobillon, 2015). Thereby particular attention is paid to differences in workers, establishments, industries, and regions by building on works on heterogeneity from the 1960s by Vernon (1960), Chinitz (1961) and Jacobs (1969), which is relevant here as these may help to explain differences over time. From this literature, an image arises that so-called young production activities that compete on the basis of product and timing differentiation are often small-scale, innovative, and highly dependent on flexible external relations, involving face-to-face contact with agents at some industrial distance whereas mature production activities that compete on the basis of low prices and standardised production are often large-scale, routine intensive, with most interaction internal to the firm and external relations being characterised by long-term standard exchanges that can take place without face-to-face contact, *e.g.* telephone calls.

These differences in production types also lead to differences in agglomeration patterns. From the classic seminal works by Weber (1922); von Thünen (1826); Alonso (1960, 1964); Muth (1969), and Mills (1967), we know respectively that short distances between agents, high rents, and high densities indicate that locational benefits are large and hard to transport. This is mostly the case for the young production activities that are often in more expensive dense areas with a large variety of sectors and agents within a short geographical distance, whereas the mature production activities are often in less expensive low-dense areas with other agents of the same industrial sector at relatively larger geographical distances, see Vernon (1960); Chinitz (1961); Jacobs (1969); Breschi and Lissoni (2001); Duranton and Puga (2001); Arzaghi and Henderson (2008); Neffke et al. (2011b); Castaldi et al. (2015).

A large literature suggests that the younger type of production activities has become more important by showing that densities, local wages and local rents have increased while the distance between interacting agents has decreased or at least remained comparable, see Duranton and Storper (2008); McCann (2008); Rodríguez-Pose and Crescenzi (2008); Glaeser (2011); Moretti (2012). This line of research in particular counters ideas that geographical proximity no longer matters like "the death of distance" (Cairncross, 1997) and that "the world is flat" (Friedman, 2007) by showing that, even though unprecedented improvements in communication and transportation technologies have put the world within arm's reach, geographical proximity still matters and actually increasingly does so, notably for sharing ideas.

However, these works do not actually measure the changes in determinants of agglomeration and authors even suggest contradictory expectations on the direction of changes in the relevance of some mechanisms, see Glaeser and Kohlhase (2004); Duranton and Storper (2008); Glaeser (2011); Moretti (2012); Faggio et al. (2017). To address this gap, this thesis brings down the barrier of the lack of data and extends the application over a larger time period of a small but solid base on how to analyse agglomeration channels, see Combes and Gobillon (2015) for an overview.

When it comes to the reasons for these changes, the varying suggestions on changes in agglomeration channels show that our understanding is greatly underdeveloped. Therefore, the greatest challenge was identifying the potential channels of economic change for have impacted agglomeration mechanisms. Most of urban economics and economic geography considering the spatial concentration trend that started in the 1980s focus on transportation costs, which are fundamental to the field ever since the classic works by Smith (1776); von Thünen (1826); Marshall (1890) and Weber (1922).

Transportation costs consist of the pecuniary transportation costs like fares and the harder to estimate opportunity costs of the foregone value of time not being (fully) productive while travelling (Glaeser et al., 2001; Glaeser and Kohlhase, 2004; Small, 2012).[2] According to a large literature the opportunity costs of not having face-

[2]Note that people are often not fully unproductive while travelling as work and leisure, *e.g.* enjoying the ride, are often still possible (Small, 2012). In particular, in public transport, the possibility of being productive while travelling has increased with the democratisation of connected devices and improved connection (Adoue, 2016).

to-face contact, *i.e.* the value of time, have increased due to the possibilities for product and timing differentiation unleashed by the development of communication and transportation technologies, notably the internet McCann and Fingleton (1996); Leamer and Storper (2001); Glaeser and Kohlhase (2004); Duranton and Storper (2008); McCann (2008); Glaeser (2011); Moretti (2012).

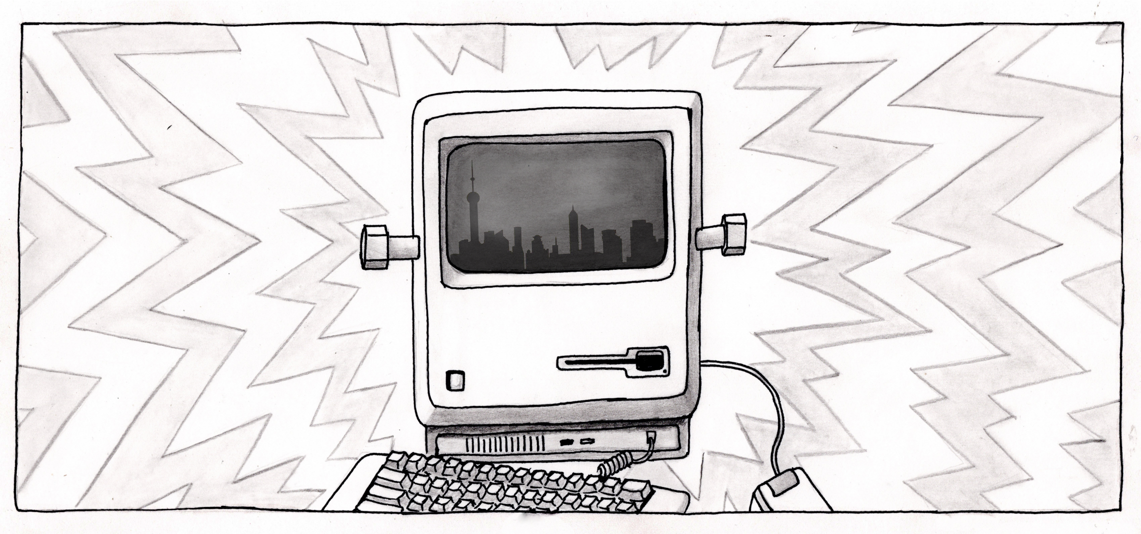
However, changes in communication and transportation technologies are insufficient to explain recent dynamics in agglomeration for two reasons: (1) the divergence in agglomeration trends starts around 1980, which is many years before the first notable breakthroughs in communication are made; and (2) for a brief period between 1950 and 1980 forces for dispersion actually were overtaking forces for agglomeration, see Anas et al. (1998); Berry and Glaeser (2005); Glaeser (2011), despite (or thanks to) an increase in transportation speed due to the democratisation of the combustion engine and telephone lines.

Therefore, I will argue that other trends, well documented in labour economics and innovation studies on the division and types of tasks executed by workers and the functioning of technological revolutions, are very insightful on how the value of time of different agents changed and with that the determinants of agglomeration and importance of geographical proximity.

Of particular interest is the computer revolution, as it is associated with a shift in profitability towards creative high-skilled jobs and niche customised products and away from routine middle-skilled jobs and mass production facilities in the 1980s more likely due to the general automation of routine tasks rather than just the automation of tasks related to communication and transportation (Storper and Scott, 1992; Pine, 1993; Goldin and Katz, 1998; Brynjolfsson and Hitt, 2000; Autor et al., 2003; Deming, 2017). This shift towards high-skilled product-and-timing-differentiated-forms of production is further encouraged by globalisation due to off-shoring and trade competition, as low-wage countries are cheaper alternatives for the production of standardised products and allow for the development of intricate global value chains, which requires the organisational capabilities of high-skilled workers (Storper and Scott, 1992; Pine, 1993; Bernard et al., 2006; Glaeser and Ponzetto, 2007; McCann, 2008; Holmes and Stevens, 2014; Autor et al., 2015; Pierce and Schott, 2016; Bloom et al., 2016). The structural break in the college wage premium, *i.e.* the wage difference between a worker with college education and a worker with high school education, that started increasing again in the 1980s after decades of decrease is a clear confirmation that the value of time of workers specialised in idea-intensive interactive tasks has increased (Freeman and Hollomon, 1975; Autor et al., 2003; Deming, 2017; Autor, 2019).

Thus, the changes in production methods and labour tasks attributed to the computer revolution coincide with the great divergence in agglomeration patterns and have similar characteristics in line with young production activities that most of urban economics and economic geography have observed but attribute mostly to the development of communication and transportation technologies instead of the larger movement of technological change of the computer revolution. The rising prominence of cities since the 1980s can not be separated from the rising importance of innovative knowledge-intensive activities (Glaeser, 2011).

Figure 1.1 – The computer set to revolutionise cities as visualised by Mikko Kuiper.



The structural break in agglomeration and wage trends following the computer revolution is a reminder that economic change is not a continuous incremental process but often punctuated by radical innovations that replace old technologies and production processes (Helpman and Trajtenberg, 1998). A force that Schumpeter (1942, pp.82-83) has dubbed creative destruction, the "*process of industrial mutation that continuously revolutionises the economic structure from within, incessantly destroying the old one, incessantly creating a new one*". A better understanding of how these channels of economic change interact with agglomeration mechanisms will improve our understanding of the spatial distribution of human activities in general and of the recent rise in spatial inequality in particular.

1.3 Research questions

Within this setting, four research questions are selected divided over three chapters. In each chapter, the focus is on a particular type of agglomeration economies, namely, the categorisation by Marshall (1890) of labour market pooling, input-output linkages and knowledge spillovers in Chapter 2 and the division of labour as often attributed to Smith (1776) in Chapter 3. In these two chapters, *static* sources of agglomeration benefits are considered. In Chapter 4, the benefits of past colocation, *dynamic* agglomeration benefits, are considered. More specifically, the development of new specialisations by regions based on past industrial activity.

All three chapters have in common that economic change plays a key role. Notably, industrial revolutions. Chapter 2 looks at the effect on Marshall's agglomeration forces of: routine-biased technological change in labour tasks; trade competition from low-income countries; and the transportation costs of goods since the computer revolution. Thereby, building on insights from labour economics on how human work changed due to technological progress and globalisation and therefore the reasons for local interaction between humans and transport economics on changing transportation costs and therefore the need for geographical proximity of different activities.

Chapter 3 evaluates the impact on the need for agglomeration resulting from the influence of the complexity of knowledge of activities on the division of labour within industries, jobs, scientific publications and patents. Data on this last activity is available since 1850, which allows for the examination of this relation over a long time period including two technological revolutions: the invention of electricity and that of the semiconductor (computer). This chapter builds on insights from innovation studies and complexity studies on the qualitative aspect of knowledge and how this affects team size and the need for face-to-face contact and thus geographical proximity.

Chapter 4 evaluates the difference in the possibility of specialised and diverse regions to diversify in new activities during the great historical crises of the long depression, great depression, and oil crisis, which coincide with moments of rapid technological change (Boschma, 1999). Here a particular focus is on regional resilience, industrial distance and temporal distance in combination with insights from innovation studies on technological change.

In each chapter, contributions are made along five dimensions: theory, data, methodology, code, and empirical results to address research gaps around the following four research questions[3]:

RQ1: To what extent has the importance of Marshall's determinants of agglomeration changed over time?

RQ2: Why has the importance of Marshall's determinants of agglomeration changed over time?

RQ3: To what extent do complex activities concentrate in large cities?

RQ4: To what extent do diverse cities differ from more specialised cities in terms of diversification behaviour during crises?

In the following, the research gaps and the respective contributions are further detailed.

1.3.1 Marshall's agglomeration determinants

RQ1: To what extent has the importance of Marshall's determinants of agglomeration changed over time?

In Chapter 2, with co-authors Hans Koster and Frank van Oort, we evaluate the changes in the importance of labour market pooling, input-output linkages, and knowledge spillovers over time. A large literature claims that the increasing importance of Marshall's knowledge spillovers is the reason for the trends in spatial concentration (Gaspar and Glaeser, 1998; Leamer and Storper, 2001; Storper and Venables, 2004; Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011; Moretti, 2012; Davis and Dingel, 2019). However, this has not been tested yet. Little is also known on how the relevance of the two other determinants of Marshall has changed: labour market pooling and input-output linkages, while the literature expresses sometimes contradictory, but empirically untested, expectations on this point. With respect to input-output linkages, Glaeser and Kohlhase (2004) suggest that its importance has decreased, as the transportation costs of goods have strongly decreased. On the other

[3]Due to the chosen focus many aspects lie outside the scope of this thesis, which are discussed in detail in the conclusion. Among others, this thesis pays less attention to natural advantages as a location reason and even less to sorting, the home-market effect, rent-seeking, and consumptive advantages, see Anas et al. (1998); Glaeser et al. (2001); Rosenthal and Strange (2004); Ellison et al. (2010) and Combes et al. (2012). This thesis also does not go into the debate if economies of agglomeration are actually externalities, see Breschi and Lissoni (2003). When it comes to the forces acting upon the role of economies of agglomeration, this thesis does not consider policy changes or changes in fiscal and social norms. These are seen as particularly relevant since the 1980s by some, see Harvey (2006); Raspe and van Oort (2007); Rodrik (2011); Raspe et al. (2012); Piketty (2013); Milikowski (2018); Piketty (2019); Milikowski (2020). Finally, this thesis focusses mainly on the U.S.A. due to data availability for a large variety of cities and an extensive time period.

hand, McCann and Fingleton (1996); Duranton and Storper (2008) and Moretti (2012) suggest that its relevance has grown as *total* transportation costs have increased, due to an increase in the frequency of deliveries, and required face-to-face contact for coordination and customisation. Labour market pooling, has become more important due to the rising skill levels of the population argues Moretti (2012). However, the findings by Faggio et al. (2017) are contradictory to most of these observations. They find that labour market pooling and input-output linkages matter more for low-skill/low-technology industries suggesting that when the skill/technology intensity of society increases these mechanisms should matter less. The finding by Faggio et al. (2017) that knowledge spillovers matter more for high-skill/high-technology industries would be in line with a rising importance of sharing ideas. Identifying the trends in agglomeration determinants addresses a major gap in the literature as mentioned by Ellison et al. (2010, p.1210), Moretti (2012, p.124), and Storper (2018, p.255).

We build on the approach of Ellison et al. (2010) to evaluate the importance of Marshall's agglomeration determinants over time. To do so a novel data set is developed using, among others, Optical Character Recognition (OCR) and data scraping tools to make use of the growing availability of scanned old documents and previously unpublished data sets. As such it is possible to build the coagglomeration index for 140 manufacturing industries, measures for labour market pooling and input-output linkages, following Ellison et al. (2010), and improve, both theoretically as empirically, on knowledge spillovers by building a measure based on the relatedness literature, see Hidalgo et al. (2018). A methodological improvement is made by developing a new measure to calculate relatedness and publish the relevant code in the EconGeo package for R maintained by Balland (2016).

We control for omitted variables, by adding dissimilarity indices following Faggio et al. (2017). The possibility of reverse causality is countered by using spatial instruments, see Ellison et al. (2010). Methodological improvements are made compared to previous studies in this line, by using industry-time fixed effects and by estimating bias-adjusted coefficients, following Oster (2019) further reducing omitted variable bias.

The empirical results show that cross-sectionally labour market pooling is the most important determinant of agglomeration followed by knowledge spillovers and input-output linkages. Technological relatedness, the new measure for knowledge spillovers, outperforms patent citations, used in previous studies. Over time there is a strong decline in the importance of labour market pooling and input-output linkages, whereas the importance of knowledge spillovers as an agglomeration determinant increases. This last finding is strong evidence of the increasing relevance of the sharing of ideas, which was brought forward but not yet tested by a large line of literature.

1.3.2 Trade, technology, and transportation costs

RQ2: Why has the importance of Marshall's determinants of agglomeration changed over time?

Chapter 2 also explores potential reasons for the trends observed under research question (1). The literature names many causes for the changes in agglomeration patterns, notably around the development of communication and transportation technologies, see McCann and Fingleton (1996); Leamer and Storper (2001); Glaeser and Kohlhase (2004); Duranton and Storper (2008); McCann (2008); Glaeser (2011); Moretti (2012). However, given that the structural break in spatial concentration trends occurred in the 1980s at the time of the computer revolution it is likely that the general automation of routine tasks and the associated shift towards young production activities in terms of the product life cycle is more relevant than just the automation of routine tasks in communication and transportation (Storper and Scott, 1992; Pine, 1993; Goldin and Katz, 1998; Brynjolfsson and Hitt, 2000; Autor et al., 2003; Deming, 2017). Following the results of Faggio et al. (2017), this would agree with the increasing relevance of knowledge spillovers and the decreasing relevance of labour market pooling and input-output linkages found under Research Question 1, see Section 1.3.1. However, the decreasing importance of input-output linkages is also in line with the prediction of Glaeser and Kohlhase (2004) based on the decreasing transportation costs of goods.

Therefore, the choice is made to focus on trade competition, technological progress, and the transportation costs of goods. Data are collected on respectively, the routine employment share following Autor and Dorn (2013), the share of imports from low-wage countries, following Bernard et al. (2006), and the share of expenditure on transportation services, following Glaeser and Kohlhase (2004) for each of the 140 SIC manufacturing industries and each of the eight time periods between 1970 and 2014.

Previous studies on the heterogeneity among industries in the preference for Marshall's agglomeration determinants by Faggio et al. (2017) and Diodato et al. (2018), respectively, divide the sample on the basis of a single industrial characteristic or do suggestions on the basis of literature to explain differences. Thereby running a large risk of omitted variable bias.

Therefore, this thesis improves methodologically on previous approaches by developing a two-step estimation procedure, which allows to identify the interaction strength of multiple variables. In this case: trade, technology, and transportation costs. In the first step, industry-time-specific coefficients for Marshall agglomeration determinants are estimated and then used as a dependent variable in the second step where the independent variables are measures for trade, technology, and transportation costs.

Once again, Oster-style bias-adjusted coefficients are estimated to counter omitted variable bias and possible reverse causality is addressed by constructing instrumental variables using trade data for other high-wage countries, following Autor et al. (2015), spatial instruments for the routine employment share, following Ellison et al. (2010), and the mean value of a ton to instrument for transportation costs (see Glaeser and Kohlhase, 2004), while controlling for its 1970 level.

The empirical results show that technological progress and trade competition are positively associated with knowledge spillovers and negatively with labour market pooling. This suggests that as routine labour tasks can be substituted by machines or offshored a common labour pool becomes less important while knowledge spillovers become more important to develop ideas and production structures to exploit these possibilities. There is some evidence that trade competition reduces the importance of input-output linkages. In robustness analyses, we show that this is more likely because locally supplied inputs can be substituted by alternatives originating from low-wage countries instead of a rise in skill/technology intensity as could be expected based on Faggio et al. (2017). At the same time and perhaps somewhat surprisingly we find that the decrease in the transportation costs of goods is not associated with the trend in input-output linkages.

1.3.3 Complex activities concentrate in large cities

RQ3: To what extent do complex activities concentrate in large cities?

In Chapter 3, with co-authors Pierre-Alexandre Balland, Cristian Jara-Figueroa, Sergio Petralia, David Rigby, and César Hidalgo, we take as starting point the first ever described mechanism of agglomeration economies, often attributed to Smith (1776): task specialisation through the division of labour. This division of labour is strongly related to the complexity of knowledge and technological progress. As this technological progress is seen as the key to economic growth (Schumpeter, 1942; Solow, 1956; Nelson and Winter, 1982; Romer, 1986), the extent to which it requires spatial concentration through the division of labour determines the extent to which cities are the engines of growth (Arrow, 1962).

Contrary to much previous research understanding the relation between innovation and spatial concentration requires measuring the qualitative aspect of knowledge instead of the quantitative aspect. Previous research on the link between innovation and agglomeration, started by Audretsch and Feldman (1996) and reviewed by Carlino and Kerr (2015), builds on count measures of inputs, intermediate outputs, and final outputs of innovation, like the number of patents. Other authors build on self-developed or borrowed, *e.g.* from Pavitt (1984) or Evangelista (2000), dichotomies of innovative and non-innovative activities/regions, see Glaeser and Ponzetto (2007); McCann (2008); Caragliu et al. (2016) and Faggio et al. (2017). These have strong limitations in the sense that count measures do not capture the quality of knowledge of each of the units, *e.g.* patents, counted and are time indifferent.[4] In this line, Balland and Rigby (2017) argue that complexity should be based on the qualitative aspect of knowledge.

Here, building on ideas from innovation studies and complexity theory the suggestion is made that more complex new knowledge requires more face-to-face-contact as it involves tacit unfamiliar knowledge (Breschi and Lissoni, 2003; Storper and Venables, 2004), and a finer division of labour as it requires individuals to strongly specialise in a limited number of tasks because it is impossible for a single person to possess all this knowledge (Leamer and Storper, 2001; Jones, 2009). As a result, with the advancement of knowledge over time the connection between complexity and geographical concentration has likely become more important.

To test these hypotheses, data are gathered on a wide range of activities on which several continuous non-geographical measures of complexity are applied. The main results are based on the size of teams involved in scientific papers, the average number of years of education per occupation or per industry, the average year of introduction of technology classes for patent categories. The results show that scientific papers with a larger team, industries and job occupations with more years of education, and patents with a more recent year of introduction concentrate more strongly in large cities.

The patent dataset allows for the analysis of the connection between complexity and spatial concentration since 1850. These results show that over time patents increasingly concentrate in large cities as technology advances, in particular during the industrial revolutions of 1870 based on electricity and 1970 based on the semiconductor (computer). Furthermore, around the time of the computer revolution, a divergence occurs in which less complex patents concentrate less strongly and more complex patents concentrate more strongly. This is in line with Leamer and Storper (2001) and is likely due to the dual effect of the improvement of communication and transportation technologies that on the one hand allows for the routinisation and dispersion of the less complex technologies but increases the need for physical connection for the more complex technologies.

[4]Note also that count indicators are also influenced by other economic factors, like changing patenting laws and strategies (Carlino and Kerr, 2015).

1.3.4 Technological diversification in times of crisis

RQ4: To what extent do diverse cities differ from more specialised cities in terms of diversification behaviour during crises?

Chapter 4, with co-authors Pierre-Alexandre Balland, David Rigby, and Ron Boschma, considers *dynamic* economies of agglomeration. The relevance of regions to develop new growth paths as old industries fade is demonstrated by the creative destruction associated with technological change. As Vernon (1960) and Glaeser (2005) discuss New York, respectively, Boston reinvented itself multiple times during its history by developing new industries when economic change led to the decline of old industries.

Many large cities like Paris and Amsterdam managed to remain the largest city in their country over centuries because they were able to reinvent themselves when technological paradigms changed. Understanding why agglomeration persists also requires understanding why certain cities are better able than others in overcoming periods of downturn by developing new growth paths. In contrast to the static model of Glaeser and Ponzetto (2007) where only good-producing and idea-producing cities exist, cities have a wide variety of industrial portfolios and are able to develop new industrial specialisations. The development of new specialisations by regions is the core topic of Evolutionary Economic Geography.

Regions do not start from scratch when diversifying: they tend to build on existing local capabilities, a process that has been labelled related diversification (Neffke et al., 2011a; Boschma, 2015; Rigby, 2015). This is not to say that unrelated diversification (i.e. the successful development of new activities unrelated to local activities) does not occur in regions, but the evidence shows it is a rare phenomenon (Hidalgo et al., 2007; Neffke et al., 2018; Pinheiro et al., 2021), as was already argued by Rosenthal and Strange (2004).

Where larger and more diverse cities first of all already have less chance of entering a crisis as "*their fortunes are not tied to the fortunes of a few industries.*" (Chinitz, 1961, p.281).[5] This chapter is concerned with the advantages of diverse cities when these have entered a crisis. Jacobs (1969) emphasises that diverse cities have more options for the recombination of ideas, which leads to "adding new work". The creative destruction of a technological revolution may exactly be the moment where radical innovations, "adding new work", is more advantageous than incremental innovations, "expanding old work". This advantage is complemented by the fact that in diverse regions, industries and vested interests are less likely to dominate the institutional and

[5] Frenken et al. (2007) find proof for this, but conclude that it holds in particular for unrelated variety instead of diversity per se.

policy network that can block new key developments (Grabher, 1993; Boschma, 2015; Neffke et al., 2018).

However, to this day these ideas have not been put to the test nor if unrelated (more radical) or related (more incremental) diversification is the norm during crises. Mostly because studies on dynamic economies of agglomeration relied primarily on case studies (Rosenthal and Strange, 2004, p.44) as it was empirically hard to quantify diversification and the industrial distance between former and new industrial activities before the development of the concept of relatedness, see Hidalgo et al. (2018). Kogler et al. (2013); Boschma et al. (2015); Balland et al. (2015) and Rigby (2015) already applied this concept to regions and technologies but have not investigated differences in diversification behaviour in periods of economic downturn and technological change. The focus lies on the three great historical crises of the U.S.A., the Long Depression (1873-1879), the Great Depression (1929-1934), and the Oil Crisis (1973-1975). These match moments of great technological change (Boschma, 1999). Notably the second and third industrial revolutions of respectively the 1870s and 1970s.[6]

For the analysis, data of the HISTPAT database of Petralia et al. (2016) was updated to be geographically consistent over time and to take into account code changes to match census records of the time. Also, a R package fastlogitME was developed to improve on current practices in R to calculate the marginal effects of logit models.

Results show that cities diversify less in new technologies during crises but when they do these are more likely related technologies. This confirms the demand-pull hypothesis of Schmookler (1966); Freeman et al. (1982) and Scherer (1982), which states that the development of new and less familiar technologies happens in periods of growth when there is more demand to fund these innovative activities. Diverse cities are shown to outperform more specialised cities in two ways. First, they have larger technological portfolios and therefore on average more relatedness density to possible new technologies, which increases the probability of entry. Second, when one controls for relatedness density diverse cities still outperform their more specialised counterparts. This suggests that agents in diverse cities are generally more open to new activities, which is in line with suggestions that vested interests against new developments are stronger in specialised cities, see (Grabher, 1993; Boschma, 2015; Neffke et al., 2018). A final analysis shows that there is no significant difference between more diverse cities, intermediately diverse cities, and specialised cities in the extent to which they switch to more unrelated diversification when entering a crisis.

[6]The Great Depression does not match an industrial revolution but economic change did occur around this time, see Boschma (1999).

1A Appendix: extended theoretical background

This thesis combines insights on economies of agglomeration, location choice and transportation costs from various subfields of urban economics and (evolutionary) economic geography with insights from labour economics, international trade, innovation studies, and complexity theory. As it requires a certain level of acquaintance with these fields to follow the line of reasoning set out in the goal and scope of this thesis in Section 1.2, this section presents a more elaborated and detailed discussion of the theoretical background

This section proceeds as follows: in Section 1A the theory behind the mechanisms of economies of agglomeration is discussed; in Section 1A a more detailed discussion of the heterogeneity among agents in the functioning of these mechanisms is discussed; Section 1A then discusses how these different agents then face different transportation costs; where Section 1A discusses how these different transportation costs are associated with different agglomeration patterns; Section 1A then discusses the evidence gathered by studies in urban economics and economic geography on changes in these agglomeration patterns and transportation costs and how these changes are mostly attributed to changes in communication and transportation technologies; Section 1A discusses how technological revolutions function and in particular how the computer revolution led to economic changes through routine-biased technological change and trade competition that alter the value of time of different agents. All together this provides the reader with the theoretical background from which the research questions in Section 1.3 are deduced.

Theoretical mechanisms of economies of agglomeration

Any treatment of economies of agglomeration invariably has as starting point the work by Marshall (1890). Here I will follow the associated categorisation of the sources of agglomeration benefits: labour market pooling, input-output linkages, and knowledge spillovers.[7]

Labour pooling

Over a century before Marshall (1890), Smith (1776) describes a first mechanism of agglomeration benefits, namely, the division of labour. When workers divide the labour tasks in production this creates efficiency gains for three reasons: first, the workers improve their skills at performing that skill, also known as learning by doing; second, a worker saves time from not having to switch between tasks; and third, it

[7]Note that Duranton and Puga (2004a) categorise the benefits by sharing, matching, and learning. While this categorisation may conceptually be more intuitive, to date it has not been possible to develop meaningful empirical metrics capturing sharing, matching, and learning.

allows for labour saving innovations such as the development of specific tools or the mechanisation of part of the tasks.[8] Rotemberg and Saloner (2000) and Duranton and Jayet (2011) show more recent evidence of the larger division of labour tasks in cities.

A second mechanism of agglomeration benefits is the sharing of risk, as already mentioned by Marshall (1890) and further worked out by Krugman (1991) as cities offer a constant market for skill, which means that firms that experience fluctuations in the demand for their products/services can scale up or down. If one firm faces a positive shock, workers with similar skills are readily available to join the production process. If, on the other hand, a firm faces a negative shock, workers can more easily find another job if similar firms are present. This alleviates negative wage effects a firm has to face during these shocks and thus the uncertainty.

A third advantage is related to matching, Duranton and Puga (2004a) detail that as more agents try to match the quality increase the probability of matching, quality of matches, of the match improves; the chances of matching, in a market with the chances of finding one that matches one needs increases. Most research on matching relates to job-seekers and job vacancies (Helsley and Strange, 1990) but Duranton and Puga (2004a) extend it to input-output linkages and knowledge spillovers. Matching also mitigates hold-up problems, when more matching parties are available the risk of hold-up problems decreases as alternatives for uncooperative partners are more readily available.[9]

However, there are also disadvantages related to labour market pooling, firms have less incentive to invest in the human capital of their workers when the risk of losing them to competitors is higher (Matouschek and Robert-Nicoud, 2005; Combes and Duranton, 2006). In this respect, labour market pooling acts as a dispersion force instead of an agglomeration force.

Input-output sharing
Marshall (1890) detailed how customer firms and supplier firms can save on transportation costs and more easily exchange information when colocated. The colocation of many customers-suppliers also leads to productivity benefits as there is sufficient demand for the development of niche products, that can more precisely fit specific

[8]Note that the division of labour itself has already been described in Ancient Greece by Plato and Xenophon, see Finley (1973) and Silvermintz (2010), but that Smith (1776) is generally more cited in the literature, see for example Duranton and Jayet (2011) and Hausmann et al. (2014).

[9]Hold-up problems occur when incomplete agreements between two parties allow one party to extort the other when that latter party made asset-specific investments in return for future profits depending on the former party. In this situation, the former party can renegotiate its share of the profits and has bargaining power if the latter party already has made investments, which it can't easily extract or can't easily replace the former part with a new partner

needs and because it becomes possible to share the costs of large indivisible goods and facilities like airports and broad-band internet (Duranton and Puga, 2004a).

Knowledge spillovers
The most famous quote of Marshall (1890, p.198) is: "*great are the advantages which people following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously.*"

Individuals working and interacting in close proximity allows less experienced workers to acquire the skills and knowledge of more experienced workers and become more productive. This also has a *dynamic* agglomeration component as knowledge accumulates in a person and the learned skills remain with that person even when one moves out of an area (Glaeser, 1999; De la Roca and Puga, 2017).

Heterogeneity

The work of Marshall (1890) was inspired by the large industrial complexes following the second technological revolution based on the invention of electricity. In the 1960s works by Vernon (1960), Chinitz (1961) and Jacobs (1969) pointed out a great heterogeneity between agents in the functioning of these economies of agglomeration and its industrial scope, *i.e.* the originating source.

Raymond Vernon is mostly associated with the product life cycle theory.[10] Vernon (1960) details how New York offers benefits to firms in the young phase of the product life cycle. This type of firms is engaged in "producing the unpredictable" (p.100). "*From one month to the next, a producer sometimes has no way of knowing what he may be expected to produce, what materials or processes may be involved, and what volume may be demanded.*" (p.101). These firms are small as they do not invest in specialised machinery for mass-production or workers specialised in only a few tasks as knowledge, production capital or inputs may change regularly. Therefore these firms build on *external* economies of agglomeration, *i.e.* unstable relations with other firms for requirements not in house.[11] Their business model is to compete by catering to small niche markets and being able to meet timely and specific requests for which New York offers benefits in speed and contact that compensate the high wages and high rents.

[10]He elaborated the product life cycle theory in international trade in Vernon (1966), but the basics of the theory focussed on urban economics were already strongly present in his summary, see Vernon (1960), of the New York Metropolitan Region study, which he conducted with among others Benjamin Chinitz.

[11]External to the firm should not be confused with agglomeration externalities, as these latter mean that the benefits are not directly paid for and hence are external to the market.

When a product and its production methods are standardised the product enters the mature phase of the product life cycle. Here long-term investments in knowledge, production capital and input production can be made. The economies of agglomeration are no longer *external* but *internal* to the firm.[12] The critical competitive question for these firms is no longer product design but the costs of labour, transport, and land. Speed and contact are no longer requirements, as Vernon (1960) argues that phone calls generally suffice for the contact required in this type of production, *e.g.* referencing product numbers in catalogues. Firms at the end of the product life cycle build on large mass production facilities to compete on the basis of low prices.

Chinitz (1961) describes this type of firms in Pittsburgh, which for example have each their own transportation fleet instead of sharing external transport providers. He argues this type of specialised city, *i.e.* few industrial sectors, with large firms inhibits the development of new firms because these new firms generally cannot make the large investments to obtain internal economies of scale but there are also no external economies of scale a firm can make use of, *e.g.* independent transportation services.

Jacobs (1969) adds that specialised cities have fewer options for the development of new industry specialisations, known as diversification, because new ideas originate from the recombination of existing knowledge. Notably by applying ideas from other sectors, which is more likely to happen in diverse cities. Furthermore, Jacobs (1969) argues that local economic growth comes from "adding new work", which are more radical developments leading to new products/production methods, in comparison to "expanding old work", which are incremental developments in more standardised production processes.

Because this latter type of process is more likely happening in specialised clusters, which are predominantly featured in the descriptions of Marshall (1890). Jacobs and Marshall are often contrasted in the literature, with, on one hand, Marshall-based localisation economies, *i.e.* economies of agglomeration within the same industrial sector, versus Jacobs-based urbanisation economies, *i.e.* economies of agglomeration between sectors, see Glaeser et al. (1992); van Oort (2002); Beaudry and Schiffauerova (2009); Caragliu et al. (2016).[13]

---

[12] Also known as economies of scale.

[13] Other authors, like Rosenthal and Strange (2004) and Faggio et al. (2017), would say that Marshall's descriptions of the mechanisms of agglomeration are not necessarily meant to only hold for specialised cities, despite the use of "*same skilled trade*" in his most famous quote. The mechanisms of labour market pooling, input-output linkages and knowledge spillovers can lead to agglomeration benefits within as well as between sectors, as shown by Faggio et al. (2017). After all, Marshall (1890) was only describing how they work in the industrial complexes of his time not how they would work in the future.

Relatedness
The dichotomy of localisation economies versus urbanisation economies basically forms the two options in the industrial scope of Rosenthal and Strange (2004). Even though, it is not hard to imagine that a continuous measure is more in place. Some sectors complement other sectors more than others do (Frenken et al., 2007). For example, for a car manufacturer, the ideas used in the semiconductor industry are likely more relevant than those used in the meat industry.

However, it is very hard to quantify such a qualitative aspect as how related two industries are in a certain dimension. To give an idea of the extent to which two activities use similar capabilities, the concept of relatedness is developed by Hidalgo et al. (2007) and has rapidly gained in popularity (Hidalgo et al., 2018). Hidalgo et al. (2007) build on the export of products by countries. They argue that if two products are more often produced together in a country than these likely use similar capabilities and are therefore more related. The probability of co-occurrence, therefore, gives an idea of the relatedness between products, which can be visualised as a network.[14] In the case of Hidalgo et al. (2007) called the product space which they use to predict the development of new specialisations by countries, based on the country's activities relatedness, known as relatedness density, to new activities. Since then, a large body of literature, see Hidalgo et al. (2018) for an overview, shows that for many geographical units (cities, regions, countries) and many activities (products, patents, occupations, research areas) the relatedness of an area to an (new) activity is positively associated with the growth of that activity. This suggests that there are different distances in the industrial scope than just diverse or specialised.

All in all, the literature on the industrial scope shows that this heterogeneity matters differently for the productivity gains of different production activities. Productive activities do not only differ in the industrial scope of economies of agglomeration but also the extent to which colocation is necessary, *i.e.* the geographical scope. Both the productivity gains and the geographical distance are strongly connected to transportation costs.

Transportation costs
Smith (1776) already wrote that transportation efficiency determines the extent of the market for each task and therefore the division of labour. A worker that specialises in a certain task needs to have sufficient demand in trading distance to be able to purchase other goods and services. As an example, Smith (1776) contrasts the remote dwellers of his time in the Scottish Highlands that are each their own farmer, brewer,

[14]Note that an improved formula for this probability is introduced in this thesis.

butcher and baker to city dwellers who specialise in only one of these tasks.

Marshall (1890) based the categorisation of labour market pooling, input-output linkages, and knowledge spillovers on the notion that firms can colocate to save on the transportation costs of people, goods, and ideas. However, each of these channels may require the transport of more than one of these. For example, sharing ideas may require exchanging prototypes (transporting goods) and possibly meeting in person (transporting people).

This last aspect is particularly important as face-to-face contact, is "*the most fundamental aspect of proximity*" (Storper and Venables, 2004, p.351) and therefore essential for understanding the need for colocation. Storper and Venables (2004) make the distinction between codifiable information and uncodifiable information, also known as tacit knowledge. Codifiable information can be easily and cheaply transmitted if sender and receiver understand that system, *e.g.* language and mathematical notations, and have the means of communicating it, *e.g.* letters, books, e-mail. In contrast, uncodifiable information can not fully be expressed in a symbol system, as often different dimensions of the problem at hand are only understood in relation to each other (Storper and Venables, 2004). This is what (Glaeser, 2011, p.24) calls the "*complex communication curse*", which can be resolved via face-to-face interaction as "*long hours spent one-on-one enable listeners to make sure that they get it right.*". Face-to-face contact is particularly useful for building trust and communicating complex information (Storper and Venables, 2004; Glaeser, 2011).[15]

However, determining to what extent geographical proximity is required to share information is challenging. Breschi and Lissoni (2001) and Boschma (2005) state that understanding codifiable information may require long study and shared experiences to understand jargon and background information and therefore may still require face-to-face contact. On the other hand, persons that have a great understanding of the same matter and each other's roles may be able to mainly communicate via codified information and video calls.

Where Rosenthal and Strange (2004) distinguish three forms of distances, Boschma (2005) labels these proximities, of which geographical and cognitive match to geographical and industrial respectively, while the temporal scope is not explicitly mentioned

[15] Storper and Venables (2004) summarise the advantages of face-to-face contact stating that it is an efficient communication technology, as it allows for instant interruption, feedback, and repair and builds on visual and body language cues; it allows for screening and socializing, which depend on identifying and assimilating tacit knowledge among group members; it generates psychological motivation; and it helps build trust as it aids in the detection of lying and meeting face-to-face requires a larger sacrifice of time to come to the same location compared to other communication tools, which signals commitment.

by Boschma (2005) but social, organisational and institutional is not mentioned by Rosenthal and Strange (2004). These last proximities relate, respectively, to sharing social ties, organisational practices, and values of conduct. When these are well developed geographical proximity is not necessary. However, geographical proximity can help overcome barriers in these dimensions.

These barriers hence give clues when geographical proximity is more relevant. In this line, Breschi and Lissoni (2001) suggest that this holds in the early stages of a project when the organisation and common language is still under development. This is in line with Vernon (1960) who discusses that the early stages of the product life cycle require speed and contact but the mature stages can do with phone calls and referencing catalogue numbers and the descriptions by Arzaghi and Henderson (2008) of the advertisement industry in which continuously networking in person is essential for new projects and new partners. Bridging cognitive/industrial/relatedness and social distances are likely more important here matching the advantages of face-to-face contact.

The transfer of ideas may therefore require the movement of people and this may be necessary not only for knowledge spillovers but also input-output linkages or labour market pooling. Mutual contact on the customisation and user guidelines of an input and on matching and dividing labour tasks may require the movement of people.

The transportation costs of people, goods, and ideas consist of two dimensions: pecuniary costs and opportunity costs, also known as money costs and time costs (Becker, 1965; Glaeser and Kohlhase, 2004). For example, it costs, at the time of writing, 7.50 euro to travel from Amsterdam Zuid to Utrecht Centraal by train but it also takes 23 minutes of someone's time, which could have been spent more productively.[16] People, goods and ideas may differ in their opportunity cost depending on context. Small firms with little inventory competing on timely niche products face larger opportunity costs for required inputs in knowledge, goods and labour than larger firms with standardised inputs and large inventories (Vernon, 1960; McCann and Fingleton, 1996). The opportunity cost of a person, good, or idea is hard to measure as theoretically the benchmark is the best use of that person, good, or idea at that time.[17]

[16] The loss in productivity while travelling has decreased through the democratisation of connected devices and improved connection in public transport (Adoue, 2016).

[17] Value of time also explains why Edward Glaeser only allows for meetings of 15 minutes for PhDs, while other professors are generally more generous with their time.

Agglomeration patterns

To get an idea of the relevance of total transportation costs, agglomeration patterns can be revealing. Agents will locate where transportation costs are the lowest, as already suggested by Weber (1922). von Thünen (1826) build a land-use model stating that agents that obtain the highest benefits of a location will outbid other agents for the land closest to that location. However, these other agents with lower benefits and lower transportation costs may be able to outbid the former agents for pieces of land further away. In the monocentric city model by Alonso (1960, 1964); Muth (1969) and Mills (1967) add to this land-use model that agents can substitute land for capital by building high-rise buildings and therefore allowing more agents on smaller plots of land, which increases the total willingness to pay for a plot of land. These sources show that the transportation costs of different agents, in the sense of pecuniary costs and missed opportunities for not being close to a location, can therefore be deduced from the distance to inputs, the location, density, and land values.

The empirical evidence on the importance of distance for economies of agglomeration dates back to Jaffe et al. (1993), who find that patents are more likely to cite patents in the same Metropolitan Statistical Area (MSA) or county than similar patents further away.[18] More recently, Inoue et al. (2019) show in Japanese patenting activity that geographical proximity matters more for inter-firm collaboration than intra-firm collaboration, in particular of small firms, confirming Vernon (1960) ideas on their needs for speed and contact. Rosenthal and Strange (2003) show that localisation effects are strong and attenuate more strongly for software industries than industries working on fabricated metals and machinery, of which the former is arguably more complex. Arzaghi and Henderson (2008) show that for advertisement agencies in New York the opportunities for networking attenuate extremely sharply with distance being equal to zero when located at more than 750 meters of another agency. This indicates that the opportunity costs of walking 750 meters for interaction are that high that the willingness to pay for proximity drives up the land prices around Madison Avenue. Advertisement agencies are relatively easy to stack upon each other, giving rise to the skyscrapers while Chinitz (1961) describes that the large complex firms in Pittsburgh described only have limited possibility to substitute land for capital.

In terms of geographical concentration, there is a large literature beginning with Audretsch and Feldman (1996) who show that the number of patents and R&D expenditures are more concentrated than production activities. Carlino and Kerr

[18]Note that Breschi and Lissoni (2003) who reviewed the paper argue that social ties and labour mobility are more likely to play a role and that geographical proximity per se is not sufficient, in line with Breschi and Lissoni (2001) and Boschma (2005).

(2015) review this literature, noting among others that venture capital is even more spatially concentrated than R&D expenditures and patent production. They argue that these innovative activities are reasonably more complex than other activities and therefore require more face-to-face contact. In line with the suggestions of Vernon (1960); Breschi and Lissoni (2001) and Arzaghi and Henderson (2008) that newer unfamiliar knowledge requires face-to-face contact, this kind of activities also locate close to inputs and strongly concentrate in space. However, how to exactly measure the innovativeness, newness or complexity remains unstated. This is further discussed in Chapter 3.

All in all, economies of agglomeration play different roles for mature activities compared to young activities both in the terms of the sources of productivity gains and the extent to which colocation is required. Localisation economies are associated with economic growth in low-density and less-knowledge/technology-intensive regions, see Caragliu et al. (2016), large establishments in the mature phase of the product life cycle building on internal economies of scale, see Duranton and Puga (2001) and Neffke et al. (2011b), which are more likely to make use of labour market pooling and input-output linkages, see Faggio et al. (2017). On the other hand, these studies show, respectively, that urbanisation economies are associated with dense knowledge/technology intensive areas and small establishments in the young phase of the product life cycle building on external economies of scale, which are more likely to make use of knowledge spillovers. This suggests that transportation costs, likely in the form of opportunity costs, are lower for mature production activities.

The heterogeneity in the role of economies of agglomeration between activities raises the main question of this thesis how these change over time and why.

Changes in agglomeration patterns and transportation costs

As a logical consequence of the focus on transportation costs and their relation to agglomeration patterns, a lot of attention in the literature is given to the role of communication and transportation technologies.

This relation over time is well detailed by Anas et al. (1998): Prior to 1840 cities were tied to ports as transportation was the cheapest using waterways. In the 19th century, trains started competing with ships. Nonetheless, within city transport remained expensive as it was done by slow and unreliable horse carriage and foot forcing all activities to cluster tightly around stations and harbours. This changed first around 1850 with the advent of streetcars and then with the development of telephone lines and the democratisation of motor vehicles. In particular, manufacturing firms left the transportation hubs to develop large scale assembly line production facilities in the cheaper outermost suburbs. Leaving the central cities to start "*their painful transition from manufacturing to service and office centers.*"(Anas et al., 1998, p.1430).[19]

Vernon (1960, p.109) and Leamer and Storper (2001, p.641) argue that technology has a double role in relation to agglomeration, at the one hand it allows for the routinisation of complex tasks, *i.e.* turning them into mature production activities, and thus allowing for its spatial dispersion, while on the other hand, it increases the complexity and time-dependence of productive activity, which requires more reliance on external economies and therefore more agglomeration, like young production activities. But where (Vernon, 1960, p.109) argues that guessing how the balance between the two plays out is "*predicting the thoroughly unpredictable.*" Leamer and Storper (2001) argue that cities overall will continue to grow with the increasing need for face-to-face contact, with each infrastructure improvement, even though shedding the jobs that as a consequence become routine and mobile enough. Thereby, pointing to the urban concentration trends in relation to transportation developments in earlier centuries, in a similar fashion as Anas et al. (1998).

Leamer and Storper (2001) in particular predict that the internet will lead to the opposite of what Cairncross (1997) calls "*the death of distance*". Thereby joining a large literature that face-to-face contact cannot be substituted by communication technology and that the development of communication and transportation technologies and the global integration of markets have actually increased the returns on physical encounters (Gaspar and Glaeser, 1998; Leamer and Storper, 2001; Storper and Venables, 2004; Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011; Moretti, 2012).[20] Like Leamer and Storper (2001) these sources unanimously emphasise the increase in complexity of the information is in one way or another. For example, McCann (2008, p.357) asserts that "*the time (opportunity) costs associated with not having continuous face-to-face contact have increased with the quantity, variety and complexity of the information produced.*". In this line, Glaeser and Kohlhase (2004) and Glaeser and Ponzetto (2007) discuss how the transportation costs of goods decreased over the last century, allowing for dispersion of certain activities, but that those of people increased, due to increasing wages and therefore higher opportunity costs of their time, leading to the concentration of face-to-face activities.

[19] Anas et al. (1998) note that in European cities less explicit trends in decentralisation occurred, likely due to the more important role of cultural amenities, as also described in the comparison between Paris and Detroit by Brueckner et al. (1999). Although better public transportation, less car accessibility and higher fuel taxes also play a role, see Glaeser et al. (2001), as well as differences in spatial planning, see Garreau (1991).

[20] Note also that New Economic Geography models predict a stronger concentration of activities with stronger local agglomeration economies and lower barriers to trade, both between regions as between countries (Rodríguez-Pose and Crescenzi, 2008; McCann, 2008).

The most cited piece of evidence is the increase in the spatial concentration of human activities (Storper and Venables, 2004; McCann, 2008; Rodríguez-Pose and Crescenzi, 2008; Glaeser, 2011; Moretti, 2012). Pointing out the trends around the world in agglomeration and urban concentration is convincing that economies of agglomeration have increased in importance but is somewhat of a circular argument, to explain why there is increased concentration one refers to the increasing importance of economies of agglomeration and the proof is again the increased concentration.

Much evidence is also directed to demonstrating that geographical proximity still matters, Gaspar and Glaeser (1998) show that business travel increases and that both denser cities use more communication technologies and that people closer together are more likely to contact each other by telephone, which indicates that interaction through technology is a complement instead of replacement to most local physical interaction. Leamer and Storper (2001); Duranton and Storper (2008) and McCann (2008) discuss evidence that there is no sign of a reduction but actually an increase in the importance of distance in international trade. Gaspar and Glaeser (1998); McCann (2008) and Glaeser (2011) discuss that the industries with access to the best communication technologies are also the most concentrated, such as those in Silicon Valley. While at the other hand less complex forms of production such as call centres, assembly, or back-offices can easily be coordinated from further away, as illustrated by anecdotes in McCann (2008) and Glaeser (2011).

However, the reasons why complexity and the need for proximity increases are very much underdeveloped, hampering the correct identification of the (changing) sources of agglomeration.[21,22] Furthermore, just pointing to the development of communication and transportation technologies is insufficient to explain agglomeration dynamics for two reasons: (1) the divergence in agglomeration trends starts around 1980, which is many years before the first notable breakthroughs in communication are made; and (2) for a brief period between 1950 and 1980 forces for dispersion actually were overtaking forces for agglomeration, see Anas et al. (1998); Berry and Glaeser (2005); Glaeser (2011), despite (or thanks to) an increase in transportation speed due to the democratisation of the combustion engine and telephone lines.

Therefore, before reviewing the evidence on the economies of agglomeration, I will go in more deeply into the specifics of the computer revolution and the changes it brought about in the nature of human activities. More specifically, the changes

[21] With an exception for a part of Leamer and Storper (2001) discussed further on.

[22] For example, during my master's program the book by Glaeser (2011) left me inspired but unsatisfied by claiming but insufficiently backing that (pp.37-38) "*as we acquire more efficient means of transmitting information, like e-mail or Skype, we spend more, not less, time transmitting information.*"

in the *opportunity costs* of routine versus creative abstract labour tasks, and the balance between the competitive advantages of mature and young production activities. Thereby building on insights from innovation studies and labour economics.

The computer revolution

The watershed in the trends of spatial concentration and spatial divergence occurred in the 1980s, which coincides with the democratisation of the computer. This discontinuity echoes in the innovation literature. Helpman and Trajtenberg (1998) argue that economists too often treat technological progress as being continuous and incremental, while it is often punctuated by a radical innovation that replaces old technologies and production processes. A force that Schumpeter (1942, p.82-83) has dubbed creative destruction, the "*process of industrial mutation that continuously revolutionises the economic structure from within, incessantly destroying the old one, incessantly creating a new one*". Based on the cycles of expansion, boom, recession, and depression by Kondratieff (1926), Schumpeter (1939) argues that radical innovations cluster in time and are then followed by a swarming of related innovations until the technology matures and recession hits. In analogy with the product life cycle theory, entire parts of the economy can switch to the phase of experimentation.

The most radical of these economic shifts occur during industrial revolutions, which are based on so-called *general purpose technologies* (GPTs) as dubbed by Bresnahan and Trajtenberg (1995). These are technologies of which the usefulness of the invention is not limited to its domain but spreads through many other fields giving rise to increasing returns to scale due to technical improvements and complementary innovations. As such, Bresnahan and Trajtenberg (1995); Helpman and Trajtenberg (1998) identify three industrial revolutions based on respectively steam, electricity, and the semiconductor.

The semiconductor technology originates in 1947 but it is the 1971 microprocessor that made the revolutionary use of this technology (Helpman and Trajtenberg, 1998). Followed swiftly by the first personal computer, the Altair 8800 in 1975, lending its name to this industrial revolution.

The nature of the technological revolution can also alter the balance between young production activities and mature production activities. Where the production methods of the electrical revolution based on mechanisation employed mainly routine middle-skilled machine operatives, these workers are virtually redundant in the production methods of the computer revolution, which instead rely on high-skilled problem-solving workers (Goldin and Katz, 1998; Autor et al., 2003; Goos et al., 2009; Autor and Dorn, 2013).[23]

More specifically, computerised machinery is highly suitable to perform routine tasks that follow explicitly programmable rules, thereby replacing the middle-skilled workers that are predominantly involved in these tasks (Autor et al., 2003, 2015).[24] At the same time, computers can not easily perform abstract tasks that require flexibility, creative thinking, and complex communication. However, it can complement workers with these tasks as these often require routine informational inputs (Autor et al., 2003). Tasks as record-keeping, calculation and searching information have been made infinitely easier due to the computer and related innovations, which mainly boosted the productivity of high-skilled workers.[25]

High-skilled workers are further benefited by the arrival of the computer, as the business value lies not in the trillionfold decline in the real price of computing power, see Nordhaus (2007), but in leveraging this capability by inventing and managing new processes and organisational structures, which raises the demand for creative, communicative and problem-solving skills (Brynjolfsson and Hitt, 2000; Autor et al., 2003).

These new opportunities include the increased possibility of opening new markets due to the development of communication and transportation technologies and the global integration of markets (Leamer and Storper, 2001; Glaeser and Ponzetto, 2007; McCann, 2008). Manual tasks, often performed by low-skill workers, such as required in non-routine jobs are hard to automate but when these do not require close proximity they can be outsourced to lower-income countries (Autor et al., 2003, 2013, 2015).[26] It then further increases the demand for idea-intensive workers to develop the complex coordination of worldwide networks to increase product and timing differentiation (Leamer and Storper, 2001).[27]

At the firm level, the computer revolution is associated with a switch from mass production to mass customisation. Where single product assembly lines, à la Henry

[23]The switch is verbally put by Austin et al. (2018, p.9): "*Henry Ford's automated assembly lines depended on tens of thousands of less skilled workers, and hence his skills strongly complemented less skilled labor. Bill Gates' innovations primarily employed highly skilled software programmers.*"

[24]This not only includes the mentioned routine machine operatives but also clerical workers in administration and management.

[25]For example, much of the literature and data collection, visualisation, and analysis in this thesis would not have been possible without these innovations.

[26]Autor et al. (2013) give the example of cooks in restaurants versus cooks that make prepared meals for grocery stores, which can be done abroad.

[27]A notable example being the five-page report of Friedman (2007, pp.414-419) who retraces the origin of his Dell laptop is one of the many perplexing accounts of these possibilities.Friedman (2007, p.419) finds out that the total supply chain for his "*computer, including suppliers of suppliers, involved about four hundred companies in North America, Europe, and primarily Asia*".

Ford, were outcompeted by more flexible production processes able to switch and customise products and as such being able to cater to more fragmented niche markets and handle volatile demand (Storper and Scott, 1992; Pine, 1993). In accordance, Brynjolfsson and Hitt (2000) note that less vertically integrated, more knowledge-intensive, higher-skilled, and less standardised establishments were more effective in harnessing the potential of the computer.

The competitive advantage of developing new ideas is complemented by trade competition. It is notably the import competition by Japanese export industries such as machine tools, motorcycles and consumer electrical products in the 1980s that speared the switch to mass customisation (Storper and Scott, 1992; Pine, 1993). More recently, the attention has shifted to the rise of China. Between 2001 and 2007, U.S. manufacturing industries lost 18% of their workers but the value added kept increasing (Pierce and Schott, 2016). This is indicative of the major changes western manufacturing industries underwent due to import competition from low-wage countries. Bernard et al. (2006); Pierce and Schott (2016), and Bloom et al. (2016) document how U.S. and European manufacturing industries become more technology- and skill-intensive. They invest more in R&D and management quality, both through investments within surviving establishments and the reduction of employment and survival probability of more labour-intensive low-tech establishments. Relatedly, Holmes and Stevens (2014) document how larger establishments producing standardised products are outcompeted by Chinese imports, whereas smaller establishments producing customised goods for niche markets are less harmed. These traits bear close resemblance to those discussed earlier, therefore at the firm level, the effects of technological progress are likely to be similar to those of trade.[28] Further increasing the returns to skill (Autor, 2019).

All in all, the democratisation of the computer therefore not only coincides with a structural break in the divergence of prosperity over space but also with (1) a structural break in returns to skill and (2) a shift from mature towards young production processes.

The first makes it likely that the opportunity costs of face-to-face contact increased. The college wage premium, the difference in returns on a college degree versus a high school degree, can be seen as a reasonable measure of the opportunity cost of the lost productive time of college-educated workers (Glaeser et al., 2001; Glaeser and Kohlhase, 2004; Small, 2012). It was decreasing until the 1980s and has grown ever

[28]At the job level, there are differences in the extent to which jobs are threatened by automation or trade competition, see Autor et al. (2015).

since (Freeman and Hollomon, 1975; Autor, 2019).[29] Although the details of this rising wage inequality differ across skills and countries.[30] This also makes it likely that the automation of routine tasks, in general, is affecting these opportunity costs, complemented by trade competition, instead of just the automation of routine tasks in communication and transportation technologies. In this line, Autor et al. (2003) and Deming (2017) already noted an increase in the intensive and extensive margins of interactive tasks performed by workers and increasing returns to social skills. Michaels et al. (2019) show that the number of interactive tasks is even stronger in cities and has risen continuously over time.

The second makes it likely that the relative importance of sources of agglomeration and industrial scope has shifted to those more important for young production activities. In this line, Glaeser and Ponzetto (2007) build a model that explains that "*the death of distance*" has been beneficial to idea-producing cities but detrimental to good-producing cities. As proof, they compare the largest industries in New York, Chicago, San Francisco, Boston, Cleveland, and Detroit and show that the latter two, which are more strongly specialised in manufacturing, declined in population and real wages, while the former cities, which are specialised in idea-intensive sectors, grew in population and real wages. McCann (2008) build a similar model based on high value-added goods versus low value-added goods.[31] Relatedly, cities that had initially high levels of human capital or abstract skills experience more growth in the number of high-skilled jobs (Berry and Glaeser, 2005; Berger and Frey, 2016) and the college wage premium (Moretti, 2012). This clearly shows path dependence and the importance of the temporal scope. As the dynamic agglomeration effects of previous activities are strongly associated with the productivity of more recent activities.

But where these changes in agglomeration patterns are rather well documented the

[29]Ironically, Freeman and Hollomon (1975) published an alarming piece on the continuously decreasing difference between the wages of college-educated and of those with only high-school degrees. At this rate, they (p.29) argue "*a growing number of people may be destined to remain underemployed or - by implication - overeducated.*".

[30]Wage inequality has grown since 1980, mostly at the expense of middle-skilled routine workers and less for low-skilled in-person service workers (Goos et al., 2009; Acemoglu and Autor, 2011; Autor and Dorn, 2013; Autor et al., 2015). Note that U.S. workers that did not complete college actually earn on average less today than they would have in the 1970s, when corrected for inflation (Acemoglu and Autor, 2011; Moretti, 2012). In the Netherlands and France similar but less pronounced trends of wage inequality exist, although in France this is only pre-tax (Groot and De Groot, 2011; Verdugo, 2014; Bozio et al., 2016). In this sense, France is an odd duck compared to other advanced countries. Where a growing inequality in *gross* labour wages occurred but actually a decrease in *net* wage inequality, as changes in taxation more than compensate the underlying growing inequality in labour costs (Verdugo, 2014; Bozio et al., 2016).

[31]Interestingly low value-added goods also concentrate in this model due to the home market effect of the New Economic Geography models resulting in large cities that attract all activity and smaller cities disappearing, in a sort of extreme winner takes all scenario.

changes in the roles of the economies of agglomeration are not, which provides the motivation for the research questions presented in Section 1.3.

Chapter 2

The dynamics of industry agglomeration:

evidence from 44 years of coagglomeration

Abstract – Evidence abounds that agglomeration patterns have changed over time, but little is known on changes in the underlying determinants of agglomeration. We analyze 44 years of coagglomeration patterns of U.S. manufacturing industries. Our findings show that over time input-output linkages and labour market pooling become less important determinants of industry agglomeration, while knowledge spillovers have become more important. We show that trade and technology shocks are strongly associated with the decline in the importance of labour market pooling and the growing importance of knowledge spillovers. The downward trend in the importance of input-output linkages is associated with increased trade competition but not with a decrease in the transportation costs of goods.

This chapter is co-authored with Hans Koster and Frank van Oort. It is forthcoming in the Journal of Urban Economics.

2.1 Introduction

Economies of agglomeration are key in understanding the spatial distribution of economic activities (Ellison and Glaeser, 1999; Duranton and Overman, 2005). A well-established literature documents large changes in agglomeration patterns (see for example Glaeser, 2011; Moretti, 2012), but pays limited attention to the changing role of agglomeration determinants explaining these patterns (Ellison et al., 2010; Moretti, 2012; Combes and Gobillon, 2015; Storper, 2018).[1]

A natural starting point to study changes in agglomeration determinants is the classification by Marshall (1890) into labour market pooling, input-output linkages, and knowledge spillovers. Ellison et al. (2010) are the first to empirically distinguish between the importance of each of these agglomeration determinants by regressing pairwise coagglomeration intensity of U.S. manufacturing industries in 1987 on the extent to which they employ similar workers, sell or buy from each other, and use similar technologies. They find that input-output linkages matter most, followed by labour market pooling, and knowledge spillovers. Since Ellison et al. (2010), and prior work by Dumais et al. (2002), there have been a number of studies studying coagglomeration, but most studies neither allow for heterogeneity in agglomeration determinants between industries, nor measure changes over time (see *e.g.* Jacobs et al., 2013; Behrens, 2016; Hanlon and Miscio, 2017; Aleksandrova et al., 2020). Notable exceptions are Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020), as they show that there is strong heterogeneity in the intensity of agglomeration determinants between industries. Of particular interest here is the finding by Faggio et al. (2017) that technology and skill-intensive industries value knowledge spillovers more, while labour market pooling and input-output linkages are more relevant for low-skilled technology-extensive industries.[2] Industries have become more technology and skill-intensive, which therefore likely has changed agglomeration determinants. Hence, we think that analyzing heterogeneity between industries *over time* may be important in understanding the changing role of agglomeration determinants. Industry agglomeration changes through the growth, closure, opening and relocation of establishments. These establishment dynamics have

[1] Using historical data, but with different methodologies, studies by Kim (1995); Dumais et al. (2002); Klein and Crafts (2012, 2020), and Hanlon and Miscio (2017) suggest that agglomeration patterns, as well as its determinants, may change considerably over time.

[2] They relate this to the 'nursery city hypothesis' as introduced by Duranton and Puga (2001). The nursery city hypothesis implies that firms first learn about their ideal production process by making prototypes. They then benefit from being in diverse places. Once firms have found their ideal process, firms switch to mass production and relocate to specialised cities where production costs are lower. Faggio et al. (2017) hypothesise that industries in the early developmental phase of the industry's life cycle coagglomerate because of knowledge spillovers, when industries become more mature and standardise their production process they coagglomerate to take advantage of a common labour pool and input-output linkages.

likely been influenced by developments in trade competition, technological progress and transportation costs of goods, following Bloom et al. (2016); Brynjolfsson and Hitt (2000), and Glaeser and Kohlhase (2004), which are also associated with an increase in technology and skill intensity.

In this Chapter, we study the dynamic nature of agglomeration economies. We assess the changes of Marshall's agglomeration determinants over time and document how these changes are related to changes in trade competition, technological change, and transportation costs. Our analysis consists of three steps. First, we exploit panel data to explain manufacturing agglomeration by proxies for labour market pooling and input-output linkages, as well as an improved proxy for knowledge spillovers. Second, we identify changes in sources of agglomeration economies over time by estimating year-by-year regressions. Third, we explore industry-year heterogeneity and test to what extent these three channels of economic change can be associated with changes in agglomeration determinants.

We invest considerable effort to digitise hard-copy data in order to build a unique balanced panel dataset with consistent geographical units and industries covering relevant aspects of coagglomeration, occupations, input-output linkages and patented knowledge, for the years for which data are available: 1970, 1977, 1989 and then for every 5 years until 2014. Our main analyses focus on coagglomeration in Metropolitan Statistical Areas (MSA) of the U.S., as these approximately represent functional urban areas.[3]

Each of the three steps introduces innovative measures and produces novel insights. In the first step, we regress coagglomeration on proxies for Marshall's determinants of agglomeration. We build on Ellison et al. (2010) in defining coagglomeration and proxies for labour market pooling, and input-output linkages. For knowledge spillovers, we improve on Ellison et al.'s (2010) measure by focusing on the co-occurrence of technologies employed in patented inventions rather than patent citations between industries. We show that our so-called technological relatedness measure outperforms patent citations in explaining coagglomeration. Following Faggio et al. (2017), we control for simultaneous dependencies of industry pairs on non-manufacturing inputs that may be correlated with coagglomeration.

The preferred specification shows that labour market pooling is the most important determinant of agglomeration between 1970 to 2014. An increase of one standard deviation in the extent to which two industries can share workers is associated with an increase of 0.195 of a standard deviation in the extent to which these two industries

[3]We show that similar results hold at the county level, which cover the entire U.S.

are coagglomerated in the same MSA. The impacts of knowledge spillovers and input-output linkages are comparable in magnitude, as a standard deviation increase in the respective proxies leads to an increase in coagglomeration of respectively 0.104 and 0.090.

In the second step, we investigate the dynamics in agglomeration determinants. We estimate year-specific regressions of coagglomeration on Marshall's sources of agglomeration economies. We find that knowledge spillovers have become more important, as since 1970 the coefficient on knowledge spillovers has almost doubled. This is strong support for a large literature that suggests that the sharing of ideas is the reason that geographical proximity is still important despite the developments in transportation and communication technologies.[4] On the other hand, we find a clear downward trend in the importance of labour market pooling and input-output linkages, which decreased, respectively, by about 45% and 90%. Hence, the large changes in agglomeration patterns documented in the literature may have been caused by drastic changes in agglomeration determinants.

In a third step, we explore why the determinants of industry agglomeration have changed over time and are different between industries. We estimate industry-year-specific coefficients for each of the agglomeration determinants and project these on proxies for three major economic trends that considerably altered the composition of manufacturing industries. These are increased trade competition from low-wage countries, routine-biased technological change, as well as a large decrease in the transportation costs of goods (Glaeser and Kohlhase, 2004; Autor et al., 2013, 2015).[5] We then test whether trade, technology, and transportation shocks are associated with changes in the determinants of industry agglomeration. The results from the third step show that more intense trade competition is associated negatively with labour market pooling and positively with knowledge spillovers. Furthermore, the routine employment share of an industry is associated positively with labour market pooling, while negatively with knowledge spillovers. For example, a standard deviation increase in routine employment share and trade competition is associated with an increase of, respectively, about 100% and 40% of the median coefficient on labour market

[4] The evidence is in line with suggestions made in Chapter 3 and among others Gaspar and Glaeser (1998); Storper and Venables (2004); McCann (2008); Glaeser (2011); Moretti (2012); Michaels et al. (2019) who provide evidence that communication technologies complement face-to-face contact; and that knowledge-intensive interactive activities increasingly concentrate in space.

[5] As proxies for trade competition and technological progress, we closely follow Autor et al. (2013) and Autor et al. (2015). We measure trade through import competition from low-wage countries, and technological progress by the share of workers with routine task-intensive jobs. For transportation costs, we follow Glaeser and Kohlhase (2004) by calculating iceberg-like transportation costs for goods.

pooling. Hence, the effects are sizable. A likely interpretation is that increased trade competition and technological progress led to more knowledge-intensive and flexible industrial production processes, both through investments in surviving establishments and the closure of the least competitive establishments (see Brynjolfsson and Hitt, 2000; Bloom et al., 2016). The demand for standardised routine tasks, and therefore access to a 'common' labour pool, decreased, while knowledge spillovers related to new (production) technologies became more important. This interpretation is in line with the results on high-technology/high-education industries of Faggio et al. (2017).

Interestingly, we find that transportation costs of goods are not strongly associated with input-output linkages. However, we do find evidence that more intense trade competition negatively relates to input-output linkages. A standard deviation increase in trade competition is associated with a decrease of about 50% of the median coefficient on input-output linkages. An extended analysis provides support for the idea that local input-linkages are replaced by input-linkages originating in low-wage countries. In contrast to Faggio et al. (2017), we do not find evidence that the decline in input-output linkages can be explained by industries becoming more technology and skill-intensive.

Related literature. We contribute in several ways to the existing literature. First, most previous studies on coagglomeration are cross-sectional, while we use panel data (Ellison et al., 2010; Faggio et al., 2017, 2020). This enables us to improve on identification by including industry-by-year fixed effects, which matters for the results. Second, we improve on the proxy for knowledge spillovers by using a measure based on the co-occurrence of technologies mentioned on patents instead of patent citations. Third, compared to Diodato et al. (2018) who also analyse dynamics in coagglomeration patterns, we use more fine-grained data at the three-digit (SIC) industry level, as well as at the MSA level, and include a proxy for knowledge spillovers. Fourth, in terms of investigating industrial heterogeneity in coagglomeration, we improve on Faggio et al. (2017) and Diodato et al. (2018) by explicitly explaining industry heterogeneity in a multivariate setting. This is similar to Faggio et al. (2020) but we use more detailed data and also exploit temporal variation.

We think the Chapter relates to a broader literature on understanding changes in location patterns of (manufacturing) industries. Most notably, a large literature suggests that the increased demand for geographical proximity is due to an increasing importance of knowledge spillovers. Despite large improvements in communication and transportation technologies, knowledge spillovers still require face-to-face contact (Gaspar and Glaeser, 1998; Storper and Venables, 2004; Rodríguez-Pose and Crescenzi, 2008; McCann, 2008; Glaeser, 2011). The findings of the second step indeed confirm that localised knowledge spillovers have become more important in the last decades.

Contradictory predictions exist in the literature on labour market pooling. According to Moretti (2012), labour market pooling is expected to be on the rise due to the increase in skill levels of the workforce. By contrast, the results of Faggio et al. (2017) suggest that a shift towards more high-technology/high-education industries would lead to a decreasing importance of labour market pooling because labour is less standardised. Our results are in line with the latter study, as labour market pooling is becoming a less prominent determinant of industry agglomeration.

Regarding input-output linkages also contradictory predictions exist: on the one hand, Glaeser and Kohlhase (2004) suggest that because transportation costs of goods have been greatly reduced input-output linkages are likely less relevant today. On the other hand, McCann and Fingleton (1996), Duranton and Storper (2008) and McCann (2008) argue that input-output linkages may have become more relevant as more competitive knowledge intensive industries require more frequent deliveries, and more face-to-face interaction, leading to higher coordination costs. Our results suggest that input-output linkages have become less important, although we cannot attribute this to the reduction in transportation costs of goods.

The third step of is the analysis is related to a large literature that aims to understand why and how the spatial organisation of the economy changes. Recall that in this step, we project industry-year heterogeneity in agglomeration determinants on proxies for important economic trends related to trade, technology and transportation costs of goods. The reduction of trade barriers allowed for more intense trade competition from low-wage countries and meant that low-skilled work has been offshored. Further, the computer revolution brought about fundamental changes in manufacturing. Entire production processes and value chains were reinvented to fully exploit the possibilities of the computer (Brynjolfsson and Hitt, 2000). In the process, computerised machinery took over much of the performance of routine tasks but raised the productivity of workers performing abstract tasks, in particular those involving complex communication and coordination (Autor et al., 2003; Deming, 2017). As a result, both trade competition and technological progress led to the downsizing and closure of establishments that were more low-skill labour intensive, low-technology, and more likely to produce standardised products, while surviving establishments increased investments in R&D, workers' skills, and capital (Brynjolfsson and Hitt, 2000; Bernard et al., 2006; Holmes and Stevens, 2014; Bloom et al., 2016; Pierce and Schott, 2016). This change in the composition of establishments within industries altered colocation patterns and now particularly represents the location choices of knowledge and skill-intensive establishments. More specifically, we find that increased import competition and a decrease in the routine employment share are associated with stronger knowledge spillovers and weaker labour market pooling.

Glaeser and Kohlhase (2004) demonstrate that over time the transportation costs of goods have strongly decreased, which may also incentivise establishments to change locations. However, we find little evidence that the decline in the transportation costs of goods can explain the decreasing trend in input-output linkages.[6]

The rest of the Chapter is organised as follows. In Section 2.2, we introduce the econometric framework, followed by the discussion of various datasets used in the analyses in Section 2.3. We report and discuss the results in Section 2.4, while Section 2.5 concludes.

2.2 Empirical framework

This section outlines the econometric framework. We first focus on identifying the impact of Marshall's sources of agglomeration economies on coagglomeration patterns. Second, we aim to study changes over time in agglomeration determinants. Third, we project industry-year-level estimates of agglomeration determinants on proxies for trade, technology and transport costs of goods.

2.2.1 Step 1: Determinants of industry agglomeration

We aim to analyse the factors that impact coagglomeration of industries over time. Following Ellison and Glaeser (1997), coagglomeration $\mathcal{C}_{ijt}$ of industries i and j in year t is:

$$\mathcal{C}_{ijt} = \frac{\sum_{m=1}^{M}(s_{mit} - x_{mt})(s_{mjt} - x_{mt})}{1 - \sum_{m=1}^{M} x_{mt}^2}, \tag{2.1}$$

where s_{mit} is the share of industry's i employment in location m in year t. More specifically, $s_{mit} = E_{mit}/\big(\sum_{m=1}^{M} E_{mit}\big)$, where E_{mit} captures employment of industry i in location m. Further, x_{mt} is the size of location m in year t, measured by the employment share of the location in the total employment of the nation. Our main results are based on estimates from 1970 to 2014 at the Metropolitan Statistical Area (MSA) level, of which there are 363 in our data, but we will also show results at the county level, which covers the entire U.S.[7]

[6]This does not mean that transportation costs do not matter because wages of skilled workers have strongly increased in the last decades (*e.g.* due to technological progress Autor, 2019). This suggests that the transportation costs of *people* likely increased as their value of time has increased (Glaeser and Kohlhase, 2004; Koster and Koster, 2015).

[7]Note that equation (2.1) implies that if both industries are not present in an area, this leads to positive coagglomeration values. We think this makes sense, as the industries then do not locate where the other is *not* present. As a robustness check, in calculating the coagglomeration index, we also consider to remove areas where both industries are not present, which leads to very similar results. These results are available upon request.

Following the literature we will construct proxies for labour pooling, $\mathcal{LP}_{ijt}$, for input-output linkages, $\mathcal{IO}_{ijt}$, and an alternative proxy for knowledge spillovers, which we refer to as technological relatedness, $\mathcal{TR}_{ijt}$. Regarding labour market pooling, firms can be located near firms that employ workers with similar skills and expertise. There are three benefits to this common labour pool: first, employing or laying off workers in the face of fluctuating demand becomes easier (Krugman, 1991); second, the matches between available job positions and workers improve (Helsley and Strange, 1990); and, third, workers will invest more in acquiring industry-specific skills (Rotemberg and Saloner, 2000). On the other hand, labour market pooling also increases the risk of losing workers to competitors, so-called labour poaching, which gives firms less incentive to coagglomerate (Matouschek and Robert-Nicoud, 2005; Combes and Duranton, 2006). With respect to input-output linkages, industries can coagglomerate with upstream industries or downstream industries to save on the costs of transporting respectively inputs or outputs. Finally, colocation of firms that employ similar technologies can lead to useful ideas 'spilling over' from one firm to another.

We address potential correlation of unobservables with our variables of interest in multiple ways. First, following Faggio et al. (2017), we include dissimilarity indices, $\mathcal{DS}_{ijt,k}$, to capture shared dependence on various inputs, $k = 1, ..., 7$.[8] Second, we improve on the previous literature by controlling for the overall tendency of a three-digit industry to coagglomerate with other industries in a specific year. More specifically, we include industry $i\times$year and industry $j\times$year fixed effects. Our baseline specification is given by:

$$\mathcal{C}_{ijt} = \alpha\mathcal{LP}_{ijt} + \beta\mathcal{IO}_{ijt} + \gamma\mathcal{TR}_{ijt} + \sum_{k=1}^{7}\zeta_k\mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}, \tag{2.2}$$

where α, β, and γ are the main parameters of interest, ζ_k, $\forall k$ are additional parameters, λ_{it} and λ_{jt} are industry i-by-year and industry j-by-year fixed effects, and ϵ_{ijt} is an error term.[9]

Furthermore, we test for the presence of omitted variable bias by estimating bias-adjusted coefficients, following Oster (2019). She shows that the effects of the inclusion of observable control variables that lead to changes in R^2 and coefficient movements can be used to calculate the bias due to unobserved omitted variables. This relies

[8] Faggio et al. (2017) argue that these are the most obvious candidates for omitted variables, as other unobserved location characteristics need to have a very particular structure and imply that industry agglomeration is correlated to both coagglomeration and the strength of the linkages between sectors measured by labour market pooling, input-output sharing and knowledge spillovers.

[9] We also considered to exploit temporal variation in coagglomeration by including industry fixed effects. However, there is too little meaningful variation to obtain reasonably precise coefficients.

on two parameters: R^2_{max}, which is the R^2 from a hypothetical regressions of the dependent variables on all observables and unobservables; and δ, which depicts the relative degree of selection on observed and unobserved variables. Note that R^2_{max} is very unlikely to be 1 in most empirical applications due to measurement error in the dependent variable. Following Oster (2019), we set δ to 1 and R^2_{max} to 1.3 times the R^2 of the baseline regression with controls and industry-year fixed effects (see equation (2.2)).

Reverse causation may be another potential concern in equation (2.2). More specifically, firms in industries with strong Marshallian links may choose to locate together but, conversely, firms that are located close together may also forge Marshallian links.[10] To mitigate this issue we follow Ellison et al. (2010) by instrumenting the Marshallian agglomeration variables with proxies based on areas where one industry is present but the other is (virtually) not present, and vice versa. Even in industry pairs with high coagglomeration values, there will typically be some establishments that are not located near establishments in the other industry. By focusing on establishments in industry i that are not near establishments in industry j, their labour hiring decisions and knowledge spillovers are less likely to be driven by joint omitted factors or by the influence of proximity to the other industry, hence coagglomeration does not affect labour market pooling or knowledge spillovers[11]

2.2.2 Step 2: Changes in agglomeration determinants

To identify changes in the determinants of industry agglomeration over time, we estimate a similar specification as (2.2) but with year-specific coefficients for each of the agglomeration determinants:

$$\mathcal{C}_{ijt} = \alpha_t \mathcal{LP}_{ijt} + \beta_t \mathcal{IO}_{ijt} + \gamma_t \mathcal{TR}_{ijt} + \sum_{k=1}^{7} \zeta_{t,k} \mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}, \qquad (2.3)$$

where α_t, β_t, and γ_t are year-specific coefficients.

[10] Note that Faggio et al. (2017) argue that coagglomeration leading to productive links should be considered as agglomeration economies. For example, if two firms forge an input-output link *after* they have coagglomerated, then this is also a form of input-output sharing. Similar examples can be given for labour market pooling and knowledge spillovers. Hence, it is questionable whether reverse causation is really an issue that should be tackled. Fortunately, the OLS results are not fundamentally different from the IV results.

[11] Ellison et al. (2010) also mitigate reverse causality concerns by employing data from the United Kingdom to calculate Marshallian proxies in the U.K. to instrument for the corresponding U.S. variables. Apart from limited historical data availability for the U.K., we consider the 'spatial' instruments to be more convincing, as similar reverse causation issues cannot be ruled out in the U.K. data.

2.2.3 Step 3: Exploring industry-level and temporal heterogeneity

We expect that there is considerable heterogeneity in agglomeration forces across industries, as shown by Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020). Hence, changes in the economy driven by developments in *(i)* trade competition, *(ii)* technological progress, and *(iii)* transportation costs of goods likely influenced the importance of the determinants of industry agglomeration, which in turn altered coagglomeration patterns.

Trade competition is associated with the displacement of millions of workers within U.S. manufacturing industries in the last decades, even though value added kept growing (Pierce and Schott, 2016). So, industries have become more technology and skill intensive, focus more on niche products, and invest more in R&D. This happened because of *(i)* trade-induced technological change within surviving establishments, as well as *(ii)* a reduction in employment and survival probability of more labour intensive low-technology establishments. This change in the composition of establishments within industries likely altered colocation patterns and now particularly represents the location choices of knowledge and skill-intensive establishments (Bernard et al., 2006; Holmes and Stevens, 2014; Pierce and Schott, 2016; Bloom et al., 2016). Following the literature, we construct a measure – import penetration – that is based on the exposure of an industry to imports from low-wage countries in year t.

Technological change, particularly the computer revolution, likely fueled changes in agglomeration determinants (see Glaeser, 2011; Moretti, 2012) and Chapter 1. Computers excel at performing so-called routine tasks thereby replacing the often middle-skilled workers performing these tasks. On the other hand, high-skilled workers are complemented by technological progress (Brynjolfsson and Hitt, 2000; Autor et al., 2003).[12] In particular, the demand for labour performing interactive tasks increased (Autor et al., 2003; Deming, 2017). As these rely on face-to-face contact, the demand for geographical proximity increases (Storper and Venables, 2004; McCann, 2008; Glaeser, 2011). Furthermore, technological progress allowed establishments to vertically disintegrate, relocate and outsource parts of their activities, leading to fundamental changes in the distribution of remaining establishments and therefore in coagglomeration patterns (Brynjolfsson and Hitt, 2000; Duranton and Puga, 2005; Glaeser and Ponzetto, 2007). As a proxy for exposure to technology we therefore take the share of routine employment in a sector, following Autor et al. (2015).

Marshall (1890) already noted that geographical proximity implies reductions in

[12] Note that technological progress is routine-biased rather than skill-biased as low-skilled workers, who generally perform manual tasks, fared better than middle-skilled workers, a phenomenon known as job polarisation (Goos et al., 2009; Autor and Dorn, 2013).

transportation costs. Glaeser and Kohlhase (2004) show how the transportation costs of goods have decreased sharply in the last decades, which could be another reason for changes in agglomeration determinants. Following Glaeser and Kohlhase (2004), we proxy for transportation costs of goods by the share of expenditure in each sector spent on transportation.

We apply a two-step estimation approach, similar to Faggio et al. (2020), to explain the observed variation in agglomeration determinants by trade competition, technological progress and transportation costs. First, we estimate a specification, which is again very similar to the baseline regression (2.2) but with *industry*-year-specific coefficients for each of the agglomeration determinants:

$$\mathcal{C}_{ijt} = \alpha_{it}\mathcal{LP}_{ijt} + \beta_{it}\mathcal{IO}_{ijt} + \gamma_{it}\mathcal{TR}_{ijt} + \sum_{k=1}^{7}\zeta_{it,k}\mathcal{DS}_{ijt,k} + \lambda_{it} + \lambda_{jt} + \epsilon_{ijt}. \tag{2.4}$$

Equation (2.4) yields an estimated coefficient for each Marshallian force for each industry in each year.

Second, we regress $\hat{\alpha}_{it}$, $\hat{\beta}_{it}$, and $\hat{\gamma}_{it}$, of industry i in time period t on the three industry-level variables capturing trade, technology, and transportation costs $\mathcal{IC}_{it}$ while controlling for time fixed effects μ_t.

$$\{\hat{\alpha}_{it}, \hat{\beta}_{it}, \hat{\gamma}_{it}\} = \sum_{\ell=1}^{3}\eta_\ell\mathcal{IC}_{it,\ell} + \mu_t + \xi_{it}, \tag{2.5}$$

where η_ℓ, $\forall\ell$ are parameters to be estimated, μ_t are year fixed effects and ξ_{it} is an error term.

We mitigate concerns related to omitted variable bias by estimating Oster-style bias-adjusted coefficients by adding variables on average establishment size and the capital-labour ratio. Given that these could be considered as proxy controls, the results should be interpreted as lower bounds (Angrist and Pischke, 2008).

One may be concerned that reverse causality is also an issue here. Agglomeration determinants are potentially correlated with local economic conditions that affect the demand for imports from low-wage countries. To instrument for import penetration, we follow Autor et al. (2015) by calculating import penetration in other high-wage countries.[13] The predicted part of import penetration due to trade shocks in all of these countries are likely due to a rising comparative advantage of low-wage countries

[13] Autor et al. (2015) use data on Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain and Switzerland. We use data of the U.N. Comtrade database for the same countries but also include France and the Netherlands, which is to be explained later in more detail.

and/or a decrease in trade costs rather than import demand changes due to local conditions.

Furthermore, the investment in technology may be dependent on the size the local labour pool, which makes it easier to employ or replace workers when necessary. Also, when firms are coagglomerated because of knowledge spillovers, the availability of new tacit knowledge is likely to encourage firms not to standardise but to continuously reinvent production processes (see Duranton and Puga, 2001; Faggio et al., 2017). For the routine employment share we construct spatial instruments as before by focusing on areas where an industry does not coagglomerate to make use of, respectively, a common labour pool, buyer-supplier relations, and knowledge spillovers.[14] In these MSAs the share of routine employment is unlikely to be influenced by these agglomeration determinants.

Finally, the share of transportation expenditure may be influenced by the extent to which an industry coagglomerates with buyers or suppliers because stronger co-agglomeration leads to shorter distances and therefore less transport expenditures. Transportation expenditures as a share of total expenditure is dependent on the expenditure on transportation and other expenditures. The latter are unlikely to be influenced by agglomeration determinants. We capture other expenditures by the mean value of a ton. Although using value of a ton as an instrument for the share of transport expenditures addresses reverse causality, it may lead to omitted variable bias as a higher value of a ton is likely correlated with more knowledge/skill intensive products, which may not be fully captured by our trade and technology measures. Therefore, we also add the value of a ton in 1970 as a control variable.

To obtain standard errors, and to address the issue that $\{\hat{\alpha}_{it}, \hat{\beta}_{it}, \hat{\gamma}_{it}\}$ are estimated parameters, we bootstrap this two-step estimation procedure by randomly selecting industry ij-ji pairs.

2.3 Data and descriptives

This section discusses the construction of the data. We discuss our proxies for agglomeration sources, in particular with respect to our alternative proxy for knowledge spillovers: technological relatedness. We further gather data related to industry-level proxies for trade competition, technology exposure, and transportation costs of goods. We close this section by reporting descriptive statistics in Section 2.3.3.

[14] More specifically, we estimate the impact of agglomeration determinants *for each MSA m separately.* We then calculate the routine employment share of each industry in the 50 MSAs where, $\{\hat{\alpha}_{itm}$, $\hat{\beta}_{itm}$, and $\hat{\gamma}_{itm}\}$ are the smallest.

2.3.1 Determinants of industry agglomeration

Coagglomeration. We calculate industry-pair-specific coagglomeration measures using the County Business Patterns (CBP) gathered by the U.S. census bureau, which has made the data available online from 1986 onwards. Raw data from before 1987 was kindly provided by Duranton et al. (2014). We construct a balanced panel dataset of consistent counties and industries at the 3-digit SIC '87 classification, which is discussed in more detail in Appendix 2A.

Labour market pooling. We use the National Industrial-Occupation Employment Matrix (NIOEM) published by the Bureau of Labour Statistics (BLS). These data are digitally available online from 1989 onwards for manufacturing employment. Data for 1970 and 1978 are obtained from hard-copy reports by the BLS (1981) using optical character recognition (OCR). We then manually checked the digitised data to avoid errors. We built a composite job classification, in which job occupations are consistent over time (see Appendix 2A for further details). In line with the previous literature, we then calculate the correlation in the share of employees across occupations between each industry pair.

Input-output linkages. For the construction of input-output linkages, we employ the use tables of the U.S. Bureau of Economic Analysis (BEA) and their concordance tables. We define $Input_{i \leftarrow j}$ as the share of industry i's inputs that come from industry j, while $Output_{i \rightarrow j}$ is the share of output sold to industry j by industry i.[15] We then define input-output linkages between industry i and j as $\mathcal{IO}_{ij} = \max(Input_{i \leftarrow j}, Output_{i \rightarrow j})$.

In contrast to previous work, we do not combine $\mathcal{IO}_{ij}$ and $\mathcal{IO}_{ji}$ into a single measure as these are directional and therefore relevant for estimating industry-year-specific coefficients later on.

Knowledge spillovers. With respect to knowledge spillovers, we show results using the patent citation measure – which follows the existing literature – but prefer results based on a novel measure: technological relatedness. Patent citations are used as a proxy for knowledge spillovers by linking technology classes to industries and then calculating the share of citations to each industry by patents of each industry. However, not all technologies, and therefore industrial knowledge, mentioned in cited patents are actually used in the citing patent. For example, in our sample there are some patents that cite up to 1,500 other patents. Hence, it is likely that a large share of the technological knowledge of these cited industries is irrelevant for the described invention.

[15] Like Ellison et al. (2010), we calculate these shares relative to all suppliers and customers, some of whom may be outside manufacturing, see Appendix 2A for more details.

Furthermore, the share of patent citations of industry i to industry j only normalises for the size of industry i in terms of patents, and not for the size of industry j. This turns out to matter a lot. The correlation between the total number of patents of industry j and the share of patent citations received from industry i is 0.72. The industry with the most patents, *i.e.* SIC356 (General Industrial Machinery and Equipment), is one of the most important cited industry for about 65% of all the industries. As a result, the sheer size of industry SIC356 seems to suggest knowledge spillovers between industries that intuitively do not make much sense. We elabourate on this issue in Appendix 2A.

Instead, we prefer to link *technologies* in patent documents to industries and then focus on the *co-occurrence of industries* in each patent. Note that a single patent can mention more than one technology class. We do so by estimating a network-based probability measure, known as relatedness, that normalises for the size of both industry i and industry j (see Hidalgo et al., 2018).[16] As explained in the Additional Chapter A, we define the *technological relatedness*, $\mathcal{TR}_{ijt}$, between industry i and industry j in year t as:

$$\mathcal{TR}_{ijt} = \frac{\mathcal{O}_{ijt}}{\left(\frac{\mathcal{S}_{it}}{\sum_{n=1}^{N}\mathcal{S}_{nt}}\frac{\mathcal{S}_{jt}}{(\sum_{n=1}^{N}\mathcal{S}_{nt})-\mathcal{S}_{it}} + \frac{\mathcal{S}_{jt}}{\sum_{n=1}^{N}\mathcal{S}_{nt}}\frac{\mathcal{S}_{it}}{(\sum_{n=1}^{N}\mathcal{S}_{nt})-\mathcal{S}_{jt}}\right)\frac{\sum_{n=1}^{N}\mathcal{S}_{nt}}{2}}, \qquad i \neq j, \quad (2.6)$$

where $\mathcal{O}_{ijt}$ is the number of co-occurrences on patents between industry i and industry j in year t, *i.e.* a count of the number of times that industries i and j are associated with the same patent. $\mathcal{S}_{it}$ and $\mathcal{S}_{jt}$ are the number of co-occurrences involving respectively industry i and industry j in t, and $n = 1, ..., N$ also refer to industries. The rationale behind the measure is to divide the observed number of co-occurrences of two industries on patents by the expected co-occurrence if all occurrences of industries would have been assigned to patents randomly. Hence, $\mathcal{TR}_{ijt} = 1$ indicates that exactly the same amount of co-occurrences has been observed as could be expected from a random distribution.

To calculate $\mathcal{TR}_{ijt}$ we then need to link technologies listed on patents to industries. We employ the concordance table by Kerr (2008) between the technology classification by the United States Patent and Trademark Office (USPTO) and the SIC classification. This concordance is based on a limited time period between 1990 and 1993 that the Canadian Patent Office recorded both the technology classes involved in the invention

[16] An argument in favour of using patent citations is that a citation captures an actual spillover. However, Jaffe et al. (2000) surveyed inventors and found that only 18% of the citations can be considered as an actual knowledge spillover, defined as the direct contact between inventors or from attending a product demonstration.

on the patent and the industry class of the patenting firm (see Silverman, 2002; Kerr, 2008). Data on technology classes per patent is obtained from the USPTO (see Marco et al., 2015).[17]

Dissimilarity measures. Faggio et al. (2017) argue that two industries might be coagglomerated because both depend on an input that is unevenly distributed over space, such as transportation infrastructure (*e.g.* ports) or natural resources. As a result, the two industries end up coagglomerating not because of Marshallian linkages but because of these location endowments. We follow their approach in addressing this issue, by measuring how similar two industries are in their dependency on inputs from agriculture, mining, water, energy, transportation, finance, insurance, and real estate (FIRE), and other services. We refer to Appendix 2A for details.

Spatial instruments. We develop spatial instruments for labour market pooling and knowledge spillovers following Ellison et al. (2010) to address reverse causality concerns between coagglomeration and agglomeration determinants.[18] Using the CBP, we identify MSAs where industry i is present but industry j is (virtually) absent and vice versa. Then, we calculate the correlation in labour share per occupation of industry i in the former MSAs with that of industry j in the latter MSAs by employing the IPUMS census samples by Ruggles et al. (2018). We use the combined 1970 1% *metro fm1* and 1970 1% *metro fm2* samples for 1970, the 1980 5% sample for 1977, the 1990 5% state sample for 1989 and 1994, the 2000 5% sample for 1999 and 2004, the 2009 *ACS 3yr* sample for 2009, and the 2014 *ACS 5yr* sample for 2014. We use the location of inventors from Petralia et al. (2016) for 1965 to 1975 and from Hall et al. (2001) for more recent years to construct a spatial instrument for technological relatedness.

For each industry i, Ellison et al. (2010) select the 25 MSAs for which industry j is the least present but industry i is strongly present. We choose to increase the number of MSAs from 25 to 50 as the IPUMS sample size in 1970 is smaller than that of the

[17]We caution that no patent-based measure can capture the full extent of knowledge spillovers and that part of the knowledge spillovers operate via input-output linkages, as suggested by Duranton and Storper (2008), and via labour market pooling, as shown by Serafinelli (2019). We also looked at the concordance table between technologies and industries of Goldschlag et al. (2020) used by Diodato et al. (2018). This crosswalk is based on keyword analysis in both technology and industry classifications. We found that for 1994 the correlation between technological relatedness following the concordance of Kerr (2008), respectively, of Goldschlag et al. (2020) is only 0.36. This low correlation coefficient does not give us enough confidence to use the Goldschlag et al. (2020) concordance, as 1994 is the closest to the 1990-1993 reference years of the concordance by Kerr (2008), where Kerr (2008) should be the most accurate.

[18]Ellison et al. (2010) used the material input trailers to construct the spatial instruments for input linkages. These data are unfortunately unavailable to us. On the other hand, Ellison et al. (2010) did not have data to construct spatial instruments for knowledge spillovers.

1990 data used in Ellison et al. (2010). Recall that there are 363 MSAs in the data.

2.3.2 Trade, technology, and transportation costs

So far, we discussed the data to estimate the impact of agglomeration determinants on coagglomeration. In the third step of the analysis, we further aim to explore how these agglomeration determinants are related to proxies for trade, technology, and transportation costs.

Trade. The effects of trade on establishments, and therefore coagglomeration patterns, are thought to result from import competition from low-wage countries (Bernard et al., 2006; Bloom et al., 2016). Bernard et al. (2006) define these trade partners as countries that, across the entire time period, have a GDP per capita that is less than 5% of the GDP per capita in the U.S., but they also consider thresholds of 10% and 15%. We choose to use this last threshold as, due to our long sample period, China would fall out of the sample in 2014 if the other thresholds are used. Data on GDP per capita are obtained from the World Bank and the full list of low-wage countries is shown in Appendix 2A.

We deviate slightly from Autor et al. (2013, 2015) as our data are at the industry-level. We follow the value share approach of Bernard et al. (2006) and Bloom et al. (2016) by calculating the import penetration, $\mathcal{IMP}_{it}$, for industry i in year t as the share of imports M_{it} from low-wage countries in the total amount of imports in this industry, $\mathcal{IMP}_{it} = M_{it}^{\text{Low-wage countries}}/M_{it}^{\text{World}}$. Trade data and concordance tables are obtained from the U.N. Comtrade database.

This measure simplifies import penetration as it ignores nationally produced and consumed goods. Another popular measure is the share of low-wage imports in 'apparent consumption', which is defined as the total of imports plus domestic production minus exports (Bloom et al., 2016). We will show that similar results hold using this measure, by employing the NBER-CES manufacturing database by Bartelsman and Gray (1996). However, we do not prefer this measure as it has limitations and incomplete industrial coverage (see Appendix 2A).

Technology. We follow Autor and Dorn (2013) and Autor et al. (2013, 2015) in defining technological progress as the decreasing share of workers performing routine tasks. To this end, we use the IPUMS census samples by Ruggles et al. (2018).[19] The exact definition of routine task intensity is given in Appendix 2A.

Transport costs of goods. To obtain a proxy for the transport costs of goods, we employ the BEA's use tables to calculate the share spend on transportation sectors

[19]We use the same IPUMS census samples as for the spatial instruments.

(SIC41-47) of the total use value of an industry i in year t. Note that the total use (*i.e.* demand) value of an industry is equal to the total make (*i.e.* supply) value. We correct for the underestimation of transport expenditure due to stronger use of private trucks in earlier time periods by employing the Commodity Flow Surveys (CFS) (see Appendix 2A for details).

2.3.3 Descriptive statistics

Determinants of industry agglomeration. Before we report descriptives, please note that the data contain some outliers, which are due to measurement errors and extreme values. The latter are a result of *e.g.* strong dependencies in input-output linkages, or in the case of coagglomeration due to extremely small industries. For instance, industry SIC237 (fur goods) has about $5,000$ employees in 1970 but only 40 in 2014, which results in extreme coagglomeration values in the latter year.

In what follows, we cap outliers for all variables to limit the disproportionate impact of a few industries by setting values below the 1$^{\text{st}}$ percentile and above the 99$^{\text{th}}$ percentile to the respective 1$^{\text{st}}$ percentile and 99$^{\text{th}}$ percentile.

Table 2.1 reports the descriptive statistics of the main variables, while histograms of the variables, as well as the developments over time are reported in Appendix 2A. By construction the mean of the coagglomeration index is close to zero. The negative minimum value on labour market pooling reveals that some industries such as SIC241 (logging) and SIC372 (aircraft and parts) have a negative correlation in employment shares per occupation, while industries SIC233 (women's, misses', and juniors' outerwear) and SIC231 (men's and boys' suits, coats, and overcoats) virtually use the same type of workers and therefore have a correlation close to 1. The maximum value on technological relatedness indicates that technologies associated with SIC391 (jewelry, silverware, and plated ware) and SIC344 (secondary smelting and refining of nonferrous) are 36 times more likely to co-occur on a patent compared to a random assignment of technologies to patents. The maximum value on patent citations indicates that 6.6% of citations by patents associated, for example, with SIC201 (meat products) cite patents associated with SIC356 (general industrial machinery and equipment). The dissimilarity indices are measured as one half of the absolute difference in the share of inputs between industry i and industry j. For example, a maximum of 0.341 for the mining dissimilarity index indicates that there is an absolute difference of 68.2% in the share of inputs received from sectors related to mining between SIC201 (meat products) and SIC291 (petroleum refining).

Figure 2A2 in Appendix 2A shows that the mean of coagglomeration and each of the agglomeration determinants are relatively stable over time. Note that the mean of the

Table 2.1 – Descriptives

	Mean	St. Dev.	Min	Max
Coagglomeration	0.0001	0.007	−0.021	0.031
Labour market pooling	0.289	0.229	0.008	0.954
Input-output linkages	0.006	0.018	0.000	0.129
Technological relatedness	1.764	4.160	0.106	36.225
Patent citations	0.007	0.012	0.00001	0.066

Note: The number of observations is 155,680.

Table 2.2 – Descriptives for trade, technology and transportation costs

	Mean	St. Dev.	Min	Median	Max
Import penetration	0.181	0.231	0.00001	0.074	0.917
Routine employment share	0.212	0.134	0.066	0.165	0.673
Transportation costs	0.048	0.046	0.001	0.038	0.285

Notes: We report the independent variables here. The number of observations is 1120.

coagglomeration index per time period in Figure 2A2 is not indicative of changes in overall coagglomeration patterns as it is close to zero by construction. The correlation between coagglomeration values in 1970 and 2014 is only 0.51, which strongly suggests that coagglomeration patterns have changed considerably. Furthermore, the variance decreased by 60% since 1970. As Faggio et al. (2017) show that low-technology industries have more extreme coagglomeration values (*i.e.* a larger variance), this would be in line with industries becoming more technology-intensive over time. Note that the correlation between the values of 1970 and 2014 for labour market pooling, input-output linkages, and knowledge spillovers, are, respectively, 0.80, 0.62, and 0.98.

Trade, technology, and transportation costs. Table 2.2 gives the descriptive statistics of our proxies for trade, technology, and transportation costs. To limit the effect of outliers, we again cap outliers. As we have 140 industries and 8 time periods the number of observations is 1,120.

The mean of import penetration tells us that between 1970 and 2014 on average 18.1% of total imports in an industry are originating from low-wage countries. SIC376 (guided missiles and space vehicles and parts) experiences the lowest (capped) import penetration of 0.1%, while the maximum of 91.7% is in *i.a.* SIC302 (rubber and plastics footwear).

The mean routine employment share indicates that on average 21.2% of the employees have routine task-intensive jobs, which are susceptible to automation. The lowest shares are found in high-technology sectors like SIC357 (computer and office equipment), while high values are found in SIC231-239 (apparel and other textile products).

We show that on average 4.8% of all expenditure is spent on transportation and related services. The industries with the lowest expenditure share produce relatively expensive products (*e.g.* SIC372 (aircraft and parts)). SIC327 (concrete, gypsum, and plaster products) spends relatively the most on transportation.

Because of possible interdependence between trade, technology, and transportation costs, we checked the correlation between our measures and find this to be rather small. The Pearson correlation is 0.33 between trade and technology, -0.22 between trade and transportation costs, and -0.13 between technology and transportation costs.

Figure 2.1 plots the median value over time for each of the variables. As expected the median import penetration rises while the share of workers with routine tasks and the share of expenditure on transportation decreases over time.[20]

Our data provide several indications of structural changes in establishments within industries and their locations that drive changes in coagglomeration patterns. When comparing total employment per industry and MSA between 1970 and 2014 the correlation is only 0.58, which is in line with the low correlation between the coagglomeration measure. This suggests that indeed large changes occurred in the industry composition within and between cities.

Furthermore, in data used for robustness checks, we find that the mean average establishment size across industries in 1970 is about 186 workers while in 2014 this has been reduced to 77 workers, while R&D expenditure per employee increased from about $3,282$ to $13,659$ (in 1987 dollars). This is in line with the literature, as Brynjolfsson and Hitt (2000); Bernard et al. (2006); Holmes and Stevens (2014); Bloom et al. (2016) and Pierce and Schott (2016) show that larger establishments are more likely to engage in low-skilled standardised production processes and more likely to close down due to trade competition and technological change, while surviving establishments become more knowledge and skill-intensive.

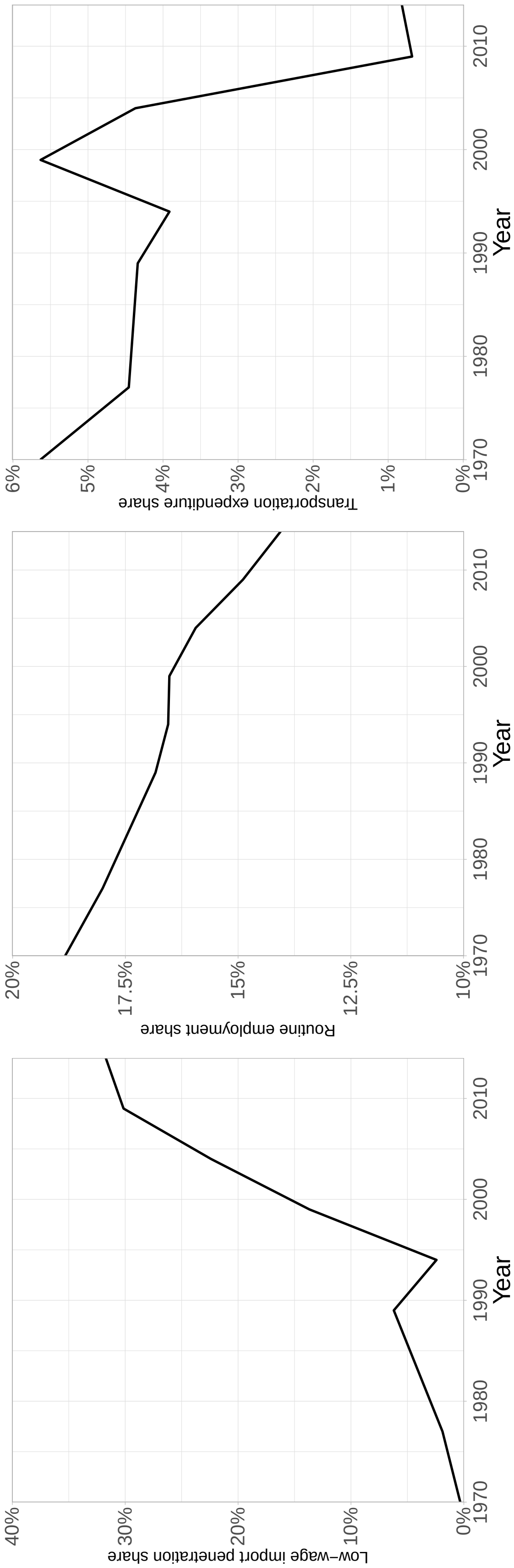
2.4 Results

2.4.1 Step 1: Determinants of industry agglomeration

The first step in the analysis is to replicate the results by Ellison et al. (2010), Faggio et al. (2017) and Diodato et al. (2018) for our data. All variables have been standardised to have a mean of zero and a standard deviation of one. As $coagg_{ij}$ and $coagg_{ji}$ are identical, we cluster standard errors at the industry pair $(ij - ji)$ by year level. We

[20] We double-checked values in 1999 for transportation costs as the values are somewhat higher than in the preceding years, but we did not observe any peculiarities. Note that we add year fixed effects to our regressions to mitigate average year-based measurement issues.

FIGURE 2.1 – MEDIAN VALUES OF ECONOMIC DEVELOPMENTS OVER TIME



(a) Trade

(b) Technology

(c) Transportation costs

TABLE 2.3 – BASELINE RESULTS
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*+ Dissimilarity measures*	*+ Industry× year f.e.*	*Patent citations*	*Tech. rel & pat. cit.*	*Separate input & output*	*Bias-adjusted*	*2SLS specification*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labour market pooling	0.114***	0.109***	0.195***	0.236***	0.195***	0.190***	0.148***	**0.294*****
	(0.008)	(0.008)	(0.013)	(0.013)	(0.013)	(0.013)	(0.026)	(**0.006**)
Input-output linkages	0.077***	0.076***	0.077***	0.090***	0.077***		0.073***	0.061***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)		(0.011)	(0.003)
Technological relatedness	0.161***	0.159***	0.104***		0.103***	0.099***	0.081***	**0.069*****
	(0.015)	(0.015)	(0.016)		(0.016)	(0.016)	(0.025)	(**0.003**)
Patent citations				0.051***	0.002			
				(0.014)	(0.013)			
Input linkages						0.048***		
						(0.006)		
Output linkages						0.063***		
						(0.007)		
Dissimilarity measures	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155680	155680	155680	155680	155680	155680	155680	155680
R^2	0.067	0.07	0.116	0.11	0.116	0.116		
R^2_{max}							0.151	
δ							1	
Kleibergen-Paap F-statistic								2632.19

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses. Instrumented variables are indicated in bold; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

present the results in Table 2.3.

In column (1), we estimate a naive specification where we only control for year fixed effects. We find that all proxies have a considerable positive effect on coagglomeration. For example, we find that a standard deviation increase in labour market pooling on average increases coagglomeration by 0.114 standard deviations. In contrast to the results based on 1987 of Ellison et al. (2010), we find a reverse order of importance for our pooled 1970-2014 data: knowledge spillovers, as proxied by technological relatedness, is the most important determinant of industry agglomeration, followed by labour market pooling and input-output linkages.[21]

Column (2) adds dissimilarity measures, in line with Faggio et al. (2017). The results do not materially change, which suggests that either omitted variable bias is not a main issue, or that having similar input requirements outside manufacturing is not a strong reason to coagglomerate.

In column (3), we present our preferred specification, in which industry $i\times$year and industry $j\times$year fixed effects are included to control for the overall tendency of an industry to coagglomerate. The estimated coefficient on labour market pooling is significantly larger compared to previous specifications. Here, a standard deviation increase in labour market pooling leads to an increase in coagglomeration of 0.195 of a standard deviation, which is almost twice as large as in the previous specification. The coefficient on input-output linkages is essentially unaffected. By contrast, the coefficient on technological relatedness is about two-thirds of the size compared to the previous specifications. Hence, this strongly suggests that it is important to control for unobservables at the industry-year level.

In column (4), we use patent citations instead of technological relatedness as a proxy for knowledge spillovers. The coefficient on patent citations is less than half of that on technological relatedness in column (3).[22] When including both proxies for knowledge spillovers in column (5), the coefficient on patent citations is close to zero, while the coefficient on technological relatedness remains virtually unchanged. Hence, the technological relatedness measure strongly outperforms patent citations in explaining coagglomeration.

Column (6) presents the results when including input linkages and output linkages

[21]We show results by year in Table 2B8 in Appendix 2B. In 1989, which is the closest to 1987, the year used in Ellison et al. (2010), we find that labour market pooling is the most important agglomeration force, followed by input-output linkages and knowledge spillovers.

[22]Note that Ellison et al. (2010) do not use industry fixed effects and only find significant positive effects for patent citations using univariate regressions in their 1987 sample.

separately. Output linkages appear to be a significantly stronger determinant of agglomeration than input linkages, while the coefficients on labour market pooling and technological relatedness are virtually unchanged.

In column (7), we present omitted variable bias-adjusted estimates, following Oster (2019). Reassuringly, the coefficients are not materially influenced when using this alternative estimation procedure. Unsurprisingly, the standard errors are higher as this methodology is less efficient than OLS.

The instrumental variable regression results are reported in column (8) of Table 2.3. The agglomeration determinants are instrumented by values taken from areas where the latter industry is not or hardly present. Recall that we do not have access to data to construct an instrument for input-output linkages so we only instrument for labour market pooling and knowledge spillovers. The first stage results are reported in Table 2.4. The coefficient on the spatial instrument of labour market pooling in column (1) shows that a standard deviation increase in the instrument is associated with 0.718 standard deviations increase in labour market pooling. Regarding the spatial instrument for technological relatedness in column (2), the effect is almost equal to one. By looking at the Kleibergen-Paap F-statistic in Table 2.3, we can confirm that the instruments are strong.

Going back to the second-stage results in column (8) of Table 2.3, we find a significantly higher coefficient for labour market pooling, in line with Ellison et al. (2010). By contrast, the coefficient on technological relatedness is somewhat lower.

Table 2.4 – First stage results

Dependent variable:	*Labour market pooling*	*Technological relatedness*
	(1)	(2)
Labour market pooling – spatial instrument	0.718***	0.010***
	(0.010)	(0.002)
Input-output linkages	0.064***	−0.001
	(0.005)	(0.001)
Technological relatedness – spatial instrument	0.129***	0.983***
	(0.008)	(0.003)
Dissimilarity measures	Yes	Yes
Industry i × year fixed effects	Yes	Yes
Industry j × year fixed effects	Yes	Yes
Observations	155,680	155,680
R^2	0.758	0.982

Notes: Standard errors are clustered at the industry ij-ji level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

2.4.2 Step 2: Changes in agglomeration determinants

So far, we have estimated the average of the coefficients between 1970 and 2014. However, we are particularly interested in how the determinants of agglomeration have changed over time. Therefore, we estimate year-specific coefficients for each agglomeration determinant. The estimated coefficients for each Marshallian determinant are plotted over time in Figure 2.2, while the full regression results are shown in Appendix 2B.

The graphs show a clear and more or less steady decline in labour market pooling and input-output linkages as determinants of agglomeration, whereas knowledge spillovers has been relatively stable until 1994 and then significantly increased. The positive trend in knowledge spillovers is strong evidence that firms aim to increase geographical proximity to share ideas, despite improvements in communication technologies see Chapter 3 or Rodríguez-Pose and Crescenzi (2008); McCann (2008); Glaeser (2011); Moretti (2012).

The decrease in labour market pooling is surprising as it is not in line with Moretti (2012) and Diodato et al. (2018).[23] By contrast, our results are in line with Faggio et al. (2017), who suggest that a shift towards more high-technology/high-education industries would lead to a decrease in labour market pooling and input-output linkages but an increase in knowledge spillovers because production is less standardised. However, Section 2.4.4 will show that the decrease in input-output linkages cannot be explained by increased technology and skill intensity.

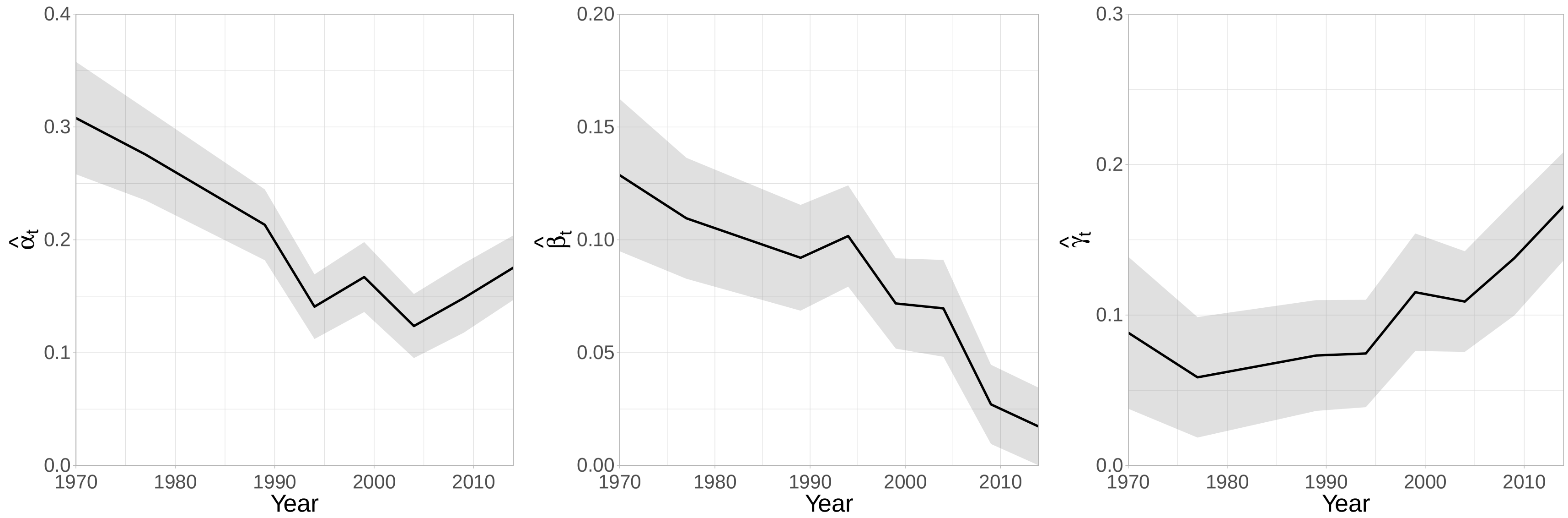
2.4.3 Robustness of Step 1 and 2

Constant definition of agglomeration determinants. Our results are robust regarding several different specifications. A concern may be that the trends in Marshallian proxies found are driven by changes in the measurement quality of the proxies over time.

To test whether this alternative explanation is important, in Appendix 2B we hold all variables, except coagglomeration, constant at their 1994 values and reproduce the main regressions. We choose 1994 as it is in the middle of our time period and the

[23]Note that suggestion by Moretti (2012) that labour market pooling has become more important is based on the "thickness" of the labour market, while coagglomeration analyses look at the importance of each force through the relative shares of industries within cities. Hence, coagglomeration in a small MSA counts the same as in a large MSA. Therefore, the results here cannot be interpreted as implying that larger labour markets do not matter more than smaller ones but only that labour market pooling became a less important determinant of coagglomeration. Furthermore, a decreasing importance of labour market pooling may not be contradicting the results of Diodato et al. (2018) as a close inspection of Figure 6 in their work reveals the possibility of a decrease in labour market pooling since 1970, even though the large standard errors prevent definite conclusions.

Figure 2.2 – Estimated coefficient per agglomeration determinant over time



(a) Labour market pooling (b) Input-output linkages (c) Technological relatedness

Note: The shaded areas indicate 95% confidence bands.

original data of 1994 is classified SIC'87, which reduces the risk of concordance errors.

These show similar results over time, which indicates that the results are not driven by changes in variable definitions or the exact measurement of independent variables.[24]

Industry-pair fixed effects. We also consider the inclusion of industry-pair fixed effects. Including these would imply that we solely rely on temporal variation in coagglomeration and agglomeration determinants to identify the effects of interest. However, industry-pair fixed effects also amplify measurement error in the agglomeration measures, as there is clearly more measurement error in the variables *within* industry pairs than *between* industry pairs. Hence, our estimates are expected to be biased towards zero.

We report results with industry-pair fixed effects in Appendix 2B. The inclusion of these fixed effects captures most of the variation, as suggested by the R^2 of over 0.7. Nonetheless, labour market pooling is positive and significant in both columns, whereas the effects of the other agglomeration forces are positive but statistically insignificant. The coefficients on the time trends have the expected sign but are not statistically significant. Hence, despite large standard errors, the coefficients seem to confirm the findings with industry-by-year fixed effects.

Coagglomeration at the county level. In Appendix 2B we show that similar results hold at a more refined geographical level. More specifically, we calculate the coagglomeration index at the county level instead of at the MSA level, leading to very comparable results.

Weighted regressions. The baseline results present the results for the average industry pair. However, industries vary greatly in size. In Appendix 2B, we reproduce the main results using weighted regressions where we weight observations by the log of employees, number of establishments, and value added. Because results would be entirely driven by a few very large industries, we take the log instead of weighting by the levels of industry size. We find that the weighted results are not statistically significantly different from the baseline regressions.

Two-way clustering of standard errors. In Appendix 2B, we reproduce the main results using two-way clustering by industry i and industry j instead of clustering at the industry pair (ij - ji) level. Unsurprisingly, these results show the same coefficients and standard errors that are considerably larger than in the main results. Still, all coefficients remain statistically significant.

[24] We obtain similar results when using 1970 or 2014 instead of 1994.

Table 2.5 – Industry-year-specific estimates

	Mean	Median	St.Dev.	Min	Max
$\hat{\alpha}$	0.157	0.111	0.281	-0.476	1.199
$\hat{\beta}$	0.066	0.034	0.129	-0.180	0.575
$\hat{\gamma}$	0.255	0.095	0.570	-0.946	2.624

Notes: We report the estimated dependent variables here. $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are the coefficients obtained in the first stage on, respectively, labor market pooling, input-output linkages, and technological relatedness. The number of observations is 1120.

Dealing with outliers. In the baseline results we have capped the values of the dependent and independent variables to the 1st percentile or the 99th percentile. This threshold is obviously somewhat arbitrary. In Appendix 2B, we show that the results are largely robust when dropping extreme observations and choosing different thresholds.

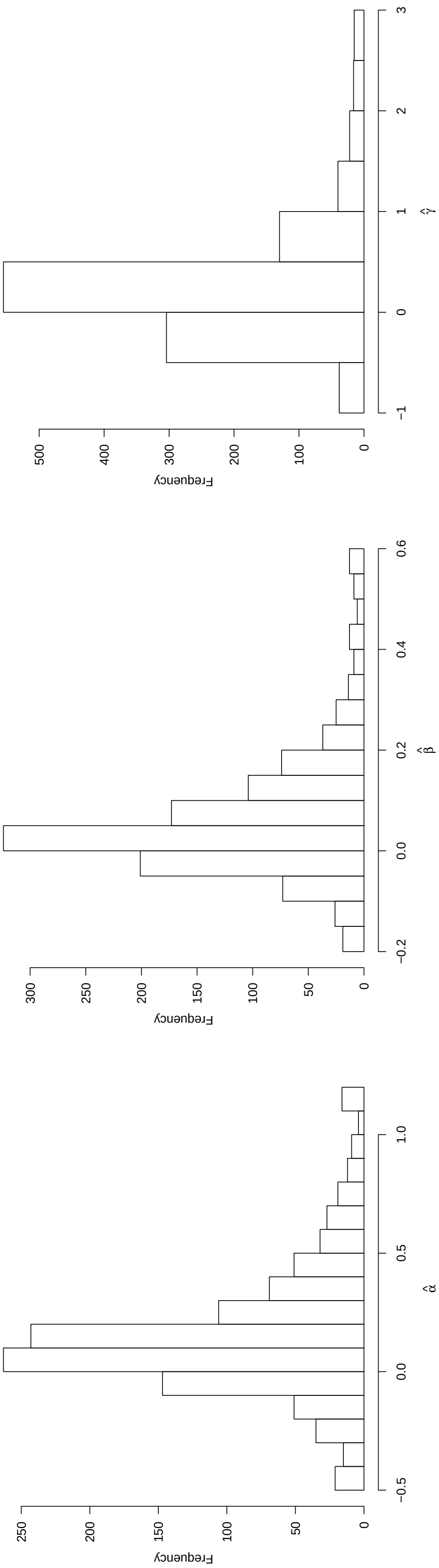
2.4.4 Step 3: Exploring industry-level and temporal heterogeneity

In this section, we explore industry-year-level heterogeneity and investigate whether our measures for trade, technology, and transportation costs are associated with the (trends in the) importance of agglomeration determinants. In the first step we obtain industry-year specific-coefficients for each of the determinants of agglomeration (see equation (2.4)). Table 2.5 reports the descriptive statistics of the industry-year-specific coefficients and Figure 2.3 shows histograms. As before, we limit the effect of outliers by capping these to the 1st percentile or the 99th percentile. As we have 140 industries and 8 time periods the number of observations is $1,120$.

Recall that the variables in the first step are standardised to have a mean value of 0 and a standard deviation of 1. As such, the mean value of the coefficient on labour market pooling $\hat{\alpha}$ indicates that an increase of one standard deviation in the extent to which labour can be pooled is associated with an increase in coagglomeration of 0.157 of a standard deviation. The means of the other coefficients input-output linkages $\hat{\beta}$, and technological relatedness $\hat{\gamma}$ also show that on average, industries have a positive appreciation for the respective agglomeration determinants. Still, 23.6%, 27.1%, and 30.4% of the values of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, respectively, are negative, although only very few of these estimates are statistically significant. We find some negative and significant labour market pooling effects for relatively skilled industries like SIC366 (communications equipment) where labour poaching may be a concern.

Each of the industry-year-specific coefficients is regressed on proxies for trade, technology, and transportation costs (see equation (2.5)). For each of the coefficients on Marshall's agglomeration determinants, Table 2.6 reports estimates from an OLS

FIGURE 2.3 – HISTOGRAMS OF THE ESTIMATED COEFFICIENTS OBTAINED IN THE FIRST STEP



(a) Labour market pooling

(b) Input-output linkages

(c) Technological relatedness

regression, bias-adjusted estimates, and estimates relying on instrumental variables. The independent variables are again standardised to have a mean of 0 and a standard deviation of 1, whereas the dependent variables are taken as they are as these are derived from regressions with standardised variables. Standard errors are obtained through bootstrapping.

Column (1) reports the OLS results for the coefficient on labour market pooling ($\hat{\alpha}$); an increase of a standard deviation in import penetration is associated with a decrease of 0.042 in the size of the coefficient on labour market pooling. As the median coefficient on labour market pooling is 0.111, a standard deviation increase in import penetration is associated with a decrease in labour market pooling of about 38% of the median. A standard deviation increase in routine employment share is associated with an increase of 0.105, which is almost equal to 100% of the median coefficient on labour market pooling. Hence, the effects are sizable and suggest that industries facing little trade competition and with highly routinised job tasks benefit more from a common labour market pool. The reduction in the number of routine task-intensive jobs, through technological progress, and the rise in import penetration complement each other and can explain the decreasing trend in labour market pooling.

Column (2) presents the results of the same specification but using the omitted variable bias-adjusted approach proposed by Oster (2019).[25] The results confirm those of the previous column. Column (3) presents the 2SLS results. For import penetration we use import penetration in other high-wage countries as an instrument, while we instrument for the routine employment share with the routine employment share calculated in areas where industries do not coagglomerate to make use of a common labour pool. The first-stage results show plausible signs and are reported in Appendix 2B and are also very similar to those in column (1).

Column (4) of Table 2.6 focuses on input-output linkages. It shows that import penetration is negatively and significantly associated with input-output linkages. A standard deviation increase in trade competition is associated with a considerable decrease of 56% of the median coefficient on input-output linkages. The coefficient on routine employment share is close to zero and highly statistically insignificant. Perhaps more surprisingly, this also holds for the coefficient on transportation costs of

[25]We compare a regression without controls to a regression with additional controls, namely the natural logarithm of the average establishment size and the capital-labour ratio, to observe movements in coefficient size and R^2, while $R^2_{\max}$ is estimated to be 0.172. Recall that these extra controls can be seen as proxy controls as these are partly capturing omitted variables but also partly the effect of trade and technological progress (see Angrist and Pischke, 2008). This procedure is therefore expected to provide a lower bound of the true effects of the variables of interest.

TABLE 2.6 – TRADE, TECHNOLOGY, AND TRANSPORTATION COSTS RESULTS

Dependent variable:	*Labour market pooling* $\hat{\alpha}$			*Input-output linkages* $\hat{\beta}$			*Technological relatedness* $\hat{\gamma}$		
	OLS	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Import penetration	-0.042^{***}	-0.057^{***}	$\mathbf{-0.051}^{***}$	-0.019^{***}	-0.008	$\mathbf{-0.017}^{*}$	0.022^{*}	0.049^{**}	**0.023**
	(0.008)	(0.009)	(**0.014**)	(0.007)	(0.009)	(**0.010**)	(0.013)	(0.021)	(**0.021**)
Routine emp. share	0.105^{***}	0.142^{***}	$\mathbf{0.128}^{***}$	-0.002	0.009	$\mathbf{-0.001}$	-0.049^{***}	-0.090	$\mathbf{-0.053}^{**}$
	(0.014)	(0.020)	(**0.016**)	(0.009)	(0.008)	(**0.011**)	(0.016)	(0.118)	(**0.021**)
Transportation costs				-0.001	-0.001	**0.013**			
				(0.008)	(0.009)	(**0.014**)			
Value of a ton in 1970 *(log)*						0.004			
						(0.004)			
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extra controls	No	Yes	No	No	Yes	No	No	Yes	No
Observations	1120	1120	1120	1120	1120	1120	1120	1120	1120
R^2	0.133			0.045			0.028		
$R^2_{\max}$		0.172			0.059			0.036	
δ		1			1			1	
Kleibergen-Paap F-stat.			112.73			16.15			118.24

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. Instrumented variables are indicated in bold. Extra controls consist of the natural logarithm of the average establishment size and the capital labour ratio. A dummy variable is added to indicate missing data in the CFS on the instrumental variable value of a ton for industries belonging to SIC27 but its coefficient is not reported in the table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

goods.[26] In column (5), we add extra controls and estimate bias-adjusted coefficients following Oster (2019). In this specification, none of the coefficients are statistically significant but they are also not significantly different from the OLS results in the previous column.

The 2SLS results are reported in column (6), we use the same instruments as before for trade and technology. For transportation costs we use the natural logarithm of the average value of a ton as an instrument. We control for the value of a ton in 1970, which mitigates the issue of omitted variable bias as the value of a ton also captures the complexity of a product. The coefficient on transportation cost is larger and positive when instrumented but not statistically significantly different from the OLS results.

Hence, the results unequivocally suggest that the 'pure' transportation costs of goods are not a relevant factor in coagglomerating with suppliers or customers, in contradiction to the expectations of Glaeser and Kohlhase (2004) and Diodato et al. (2018). However, total trade costs consist of much more than only these 'physical' transportation costs of goods. For instance, McCann and Fingleton (1996) and Duranton and Storper (2008) show that face-to-face contact and coordination are also important in sustaining input-output linkages. The transportation costs of persons has actually increased over time due to the increasing wages of (skilled) workers as Glaeser and Kohlhase (2004) suggest.

By contrast, we show that import penetration reduces the demand for input-output linkages. We find suggestive evidence in Appendix 2B that localised input linkages are replaced by input linkages with low-wage countries, as trade competition for input negatively affect input linkages while it is positively associated with output linkages (although coefficients are imprecise).

The decline in input-output linkages cannot be explained by industries becoming more technology and skill-intensive (see Faggio et al., 2017), as other industry characteristics closely related to industry skill and technology levels are unrelated to input-output linkages (see Appendix 2B).

Column (7) in Table 2.6 explores whether the importance of knowledge spillovers can be linked to changes in trade and technology. Import penetration is positively associated with the intensity of knowledge spillovers. According to column (7) the effect is about 23% of the median coefficient on technological relatedness (which is equal to 0.095). By

[26] One may argue that by controlling for the dissimilarity variable capturing transport inputs in the first stage, the coefficient of transport costs may be reduced. We have estimated regressions where we excluded the transport dissimilarity measure in the first stage leading to nearly identical results.

contrast, the routine employment share shows a negative association of about 52% of the median coefficient. This suggests that the increase in import penetration and ongoing computerisation that led to more innovative skill technology-intensive manufacturing establishments have raised the need to coagglomerate in proximity of establishments that use similar technologies. This is likely due to the increased relevance of new ideas in the production process and the need to meet face-to-face to exchange ideas (see Holmes and Stevens, 2014; Storper and Venables, 2004). Column (8), which are bias-adjusted results, display stronger but not significantly different coefficients. The 2SLS results in column (9) show very similar results as the bias-adjusted estimates.

All in all, the results presented in this section indicate a complementary impact of increasing import competition and decreasing routinisation of labour tasks on labour market pooling and knowledge spillovers. Input-output linkages seem unaffected by the decrease in the pure transportation costs of goods but the increase in import competition does seem to have had a negative effect.

2.4.5 Robustness of Step 3

In Appendix 2B we further investigate the robustness of the results by showing: *(i)* the effects of transportation costs of goods on labour market pooling and knowledge spillovers; *(ii)* the effects of the control variables (*i.e.* average establishment size and the capital-labour ratio), used in the bias-adjusted estimation procedure; and *(iii)* the effects of increased R&D expenditure and skill intensity, following Bloom et al. (2016) and Pierce and Schott (2016), which are closely associated to trade competition and technological progress. The results show that the main results are robust to these specifications. We further find evidence that the decreasing importance of labour market pooling and increasing importance of knowledge spillovers is likely related to the rise of the high technology/high education firms, as R&D expenditures are strongly associated with labour market pooling and knowledge spillovers. By contrast, none of the additional industry characteristics capturing technology and skill levels are statistically significantly associated with input-output linkages.

Further, in the main results in Table 2.6, we attach equal weight to each industry-by-year observation. By contrast, Faggio et al. (2020) weight each observation by the inverse of the standard deviation of the coefficient obtained in the first step. The results reported in Appendix 2B are largely similar.

In Appendix 2B, we employ an alternative measure of import penetration, namely the imports from low-wage countries divided over the so-called apparent consumption in the U.S. The apparent consumption is equal to domestic production minus exports plus imports. The results are not significantly different from our main results in Table

2.6. This demonstrates that there is no issue in using the value share measure, for which more data and instruments are available.

Appendix 2B finds similar effects if we estimate everything in one step, instead of our proposed two-stage approach (recall equations (2.4) and (2.5)); the signs on the coefficients all point in the same direction with only minor differences in significance levels and effect sizes.

Finally, in Appendix 2B, we further explore the results on input-output linkages when: *(i)* using the value of a ton as a proxy for transportation costs instead as an instrument; *(ii)* using the import penetration within sectors from which inputs are obtained instead of import penetration within the own sector; and *(iii)* calculating separate coefficients for input linkages and output linkages using these as separate dependent variables in the second step. The results are similar to the main results and provide further evidence that the decrease in transportation costs of goods is not a relevant factor explaining input-output linkages. By contrast, both import competition within producing industries as well as within supplying industries influence input linkages. This further suggests that the substitution of inputs from low-wage countries for local inputs is likely behind the decline in input-output linkages, rather than the effect of trade competition on increasing the intensity of technology and skill of industries, (following Faggio et al., 2017).

2.5 Conclusion

In the last 50 years, the economy underwent large and fundamental changes due to more intense trade competition, technological progress, and reductions in transportation costs of goods. Evidence abounds that this has resulted in large changes in agglomeration *patterns*. In this Chapter, we assess changes in agglomeration *determinants* over time and explore whether industry-year-level heterogeneity can be explained by changes in trade competition, technological progress and reductions in transport costs.

Using an alternative proxy for knowledge spillovers, we find that between 1970 and 2014 knowledge spillovers have become more important. This is strong evidence that geographical proximity is becoming more relevant for exchanging ideas, despite strong improvements in communication technologies. On the other hand, we find that labour market pooling and input-output linkages have become less important agglomeration determinants.

Furthermore, we show that trade competition and technological progress are strongly related to labour market pooling and knowledge spillovers. These results suggest that the computer revolution and trade competition, which led to less standardised, less vertically integrated and more knowledge-intensive establishments, altered the composition of industries and therefore the relevance of labour market pooling and knowledge spillovers in explaining agglomeration. Maybe surprisingly, we do not find that transportation costs of goods are associated with input-output linkages. On the other hand, we find a negative effect of increasing trade competition. We find suggestive evidence that this is likely due to import substitution of local inputs in input-output linkages.

Our study opens up avenues for further research. First, future studies could look more closely into the heterogeneity in agglomeration benefits of establishments *within* industries, related to *e.g.* skill and capital-intensity of establishments. Second, we note that the current framework overlooks the effects of agglomeration size, because the coagglomeration index is a relative measure.[27] Third, the two-step methodology introduced to explain changes in the determinants of agglomeration could be expanded to include various other industrial or regional characteristics. Fourth, a more obvious step forward would be to include measures of knowledge spillovers in the services industry. Finally, we note that Duranton and Puga (2004a) distinguish between sharing, matching, and learning, rather than using Marshall's categorization. While Duranton and Puga's categorisation may conceptually be more intuitive, to date it has not been possible to develop meaningful empirical metrics. Future research could aim to find meaningful proxies for sharing, matching, and learning.

All in all, this Chapter demonstrates the importance of explicitly considering the determinants of agglomeration and underlying economic trends in understanding (changes in) agglomeration patterns.

[27] The coagglomeration index considers the relative joint presence of two industries in a city. Hence, coagglomeration in a small city is just as important as coagglomeration in a large city.

2A Appendix A: Data

Coagglomeration

We obtain data on the number of employees per manufacturing industry and county from the CBP, which is available online from 1986 onwards. Raw data for 1970, 1971, 1977, and 1978 was kindly provided by Duranton et al. (2014). We base ourselves on their procedure to estimate censored values and accommodate changes in county boundaries. This procedure also involves taking the county level mean employment per industry of two subsequent years, *i.e.* employment of an industry in a county for 1970 is based on the average employment for 1970 and 1971.

Data of 1970 and 1971 uses the SIC '67 classification, and data of 1977 and 1978 uses the SIC '72 classification. We use the Census of Manufactures of 1972 and 1987 to concord employment values to the 3-digit SIC '87 classification. For the industrial concordance between NAICS to SIC '87, we employ the economic censuses of 1997, 2002, 2007, and 2012 in which a classification change took place. Hence, data are given according to both classifications, *e.g.* the economic census 1997 contains data on the number of establishments, number of employees, value of shipments, and payroll according to both the NAICS1997 classification as well as the SIC'87 classification. Our baseline results are based on estimates at the Metropolitan Statistical Area (MSA) level but we will also show results at the county level.

Agglomeration determinants

labour market pooling

We employ the NIOEM of the BLS to calculate a proxy for the extent to which two industries can share workers. Data of 1970 and 1978 were published in the hard-copy BLS (1981) report, along with projected values for 1990. Data was retrieved using OCR.[28] For the other time periods, we try to stay close to the years of the CBP. This is exactly possible in 1989, 2004, 2009, and 2014. There is no data for 1994, therefore we combine data from 1992 and 1995, and data for many occupations is missing in 1999, therefore we use data from 1998 instead.

In 44 years, the layout of the NIOEM as well as the occupations recorded underwent numerous changes. There is a risk that measured changes in the importance of labour market pooling over time are due to these definition changes rather than actual changes in the potential to share workers. Therefore, we employ a concordance of the BLS

[28] This is not without risk as numbers can be misread. However in this case, OCR errors could easily be manually identified and corrected as the values are stated in percentages, which should add up to 100%. The tabulizer package for R by Leeper (2018) proved to be particularly useful in our data collection endeavours.

between the OES codes used for job occupations in 1989 to 1998 and the SOC codes used in 2004 to 2014. When multiple occupation categories of one classification match a single occupation category of another classification, we group these occupations into the respective single category. We do the same when one or more years publish data at a more aggregate occupation category. As such, we create a composite classification consisting of 260 occupations that are (more or less) consistent over time.

Using the Dictionary of Occupation Titles of 1965 and O-NET job descriptions, the job occupations in 1970 and 1978 are matched to this job classification. When a single occupation applies to multiple job occupations in BLS (1981) we use the industry-specific projected 1990 values in comparison to the respective values of 1989 in our composite classification to estimate the industry-specific share of each occupation that belongs to the respective occupations.

Input-output linkages

To estimate the share of inputs/outputs bought from/sold to a certain industry we employ the detailed use tables and accompanying concordance tables between the IO codes and the industry classification in each year. These are published by the BEA every 5 years, in years ending on 2 and 7. When multiple SIC codes match a single IO code we divide the associated values according to the relative employment size of the involved SIC industries as reported in the CBP at that time period. For data after the switch to NAICS in 1997, we concord IO codes to NAICS codes and then to SIC codes using the share of value of shipments obtained from the economic censuses at every classification update. The data in the SIC '67 and SIC '72 classification are taken as they are as not all values in IO codes have detailed matching SIC codes.

By employing only the use tables, we follow Ellison et al. (2010), who simplify matters by disregarding the fact that industries can produce other commodities than their own. Diodato et al. (2018) do take this fact into account by employing both the make (supply) and use (demand) tables to construct input-output linkages. We also tried their procedure and found the same trends in agglomeration forces over time but noted that coefficients on input-output linkages are noticeably smaller when using the latter approach.

Knowledge spillovers

We build on patent data to develop a proxy for the extent to which two industries can learn from each other. The historical USPTO data (see Marco et al., 2015) gives among other the technology classes of each patent between 1836 and 2014. We pool patents granted between 5 years before and 2 years after each year in our data. For example, for calculating our measures in 1970, we use patents granted in 1965 to

Table 2A1 – Source years of the data

CBP	Labour market pooling	Input and Output	Patents
1970 and 1971	1970	1972	1965-1972
1977 and 1978	1978	1977	1972-1979
1989 and 1990	1989	1987	1984-1991
1994 and 1995	1992 and 1995	1992	1989-1996
1999 and 2000	1998	1997	1994-2001
2004 and 2005	2004	2002	1999-2006
2009 and 2010	2009	2007	2004-2011
2014 and 2015	2014	2012	2009-2014

1972. Then we use the probability that the respective technologies were manufactured by an industry according to the concordance table by Kerr (2008) to associate SIC '87 industries to each technology. As such, we obtain to what extent industry pairs co-occur in patents and to which industries patents belong when cited or citing. Note that each patent can make reference to multiple technology classes.

Data on patent citations before 1975 was kindly provided by Petralia et al. (2016), whereas more recent data are obtained from the NBER patent database (Hall et al., 2001). Following the previous literature, we construct the patent citation measure by calculating the share of patents associated with industry i citing patents associated with industry j in the total number of patent citations by industry i, for each i and each j.

Table 2A1 shows the respective years of the data sources per time period.

We decided not to use patent citations because technological relatedness outperforms this measure both empirically (see Section 2.4), and conceptually. Conceptually, both proxies for knowledge spillovers build on linking technological classes on patents to industries, where patents can make reference to multiple technology classes. When using patent citations as a proxy for knowledge spillovers, we consider the share of citations of industry i's patents towards patents of industry j.

However, not al of the technology classes used in the cited patents may be used in the citing patent. In our sample, some patents cite up to $1,500$ other patents. As a result, this superfluous information is strongly reducing the quality of the patent citations proxy as a large share of technologies cited is likely irrelevant. On the other hand, technological relatedness builds on the co-occurrence of technological classes listed on the patents and therefore only uses the technological classes that are actually used in the invention described in the patent.

Furthermore, the patent citations measure is likely distorted as they are only normalised by the number of patent citations of the industry itself and not by the sizes of other industries. It is a directional measure in the sense that $PC_{ij} \neq PC_{ji}$ where the number of citations of industry i to industry j is divided over the total number of citations by patents of industry i. Therefore, the number of patent citations is only normalised by the size of industry i and not by the size of industry j.

This has strong consequences because industries vary greatly in the number of patents filed. For example, industry SIC 356 General Industrial Machinery And Equipment is associated with about $1,273,495$ (out of $21,278,149$) patents used to estimate knowledge spillovers in 1994, while industry SIC 201 Meat Products is only associated with about 6,194 patents. As a result, a large share of the citations of SIC 201 are in the direction of industry SIC 356, likely not because these are as useful but because these are overrepresented in the patent data. Actually, the correlation between the number of patents of industry j and the share of patent citations by industry i it receives is substantial with 0.72. As a result, SIC 356 is in the top 3 of the most important cited industry for almost 65% of the industries.

More specifically, we give a ranking of the industries that are in the top 3 of most important knowledge spillover partner of other industries in Table 2A2, where the columns show the industry, the frequency in which it appears in the top 3 of other industries, and an overview of these corresponding industries. We note that a large share cites industries with other two digit codes and that many do not make much sense intuitively. It is highly unlikely that earlier mentioned example SIC 201 (Meat Products) can learn the most from SIC 283 (Drugs) and SIC 355 (Special Industry Machinery), next to SIC 356. These industries also happen to be the industries with the most patents and the most received citations.

On the other hand, technological relatedness normalises the size effect of both industries. The measure captures how often industries co-occur on patents compared to if industries were randomly distributed over patents. As industries with more patents are more likely to be present on a patent if distributed randomly it corrects for the size effect of both. As a result, technological relatedness is an undirected measure and $\mathcal{TR}_{ij} = \mathcal{TR}_{ji}$. The correlation between the number of patents of industry j or i and how technologically related it is to industry i, respectively, j is essentially zero (*i.e.* -0.1).

When looking at the frequency and industries that appear in the top 3 of most important knowledge partners according to technological relatedness in Table 2A3 industries appear much less often and are more often within the same 2-digit industry. For example, SIC 271 (Newspaper) is the most important partner for all other industries in SIC27 (Printing and Publishing): SIC 272-SIC279, which are likely to use similar technologies. When the 2-digit codes do not match the combinations intuitively make

Table 2A2 – Industries in the top 3 of most cited by other industries

Industry	Frequency	Corresponding industries
SIC356 General Industrial Machinery And Equipment	90	SIC201-SIC223; SIC226-SIC229; SIC239-SIC267; SIC274; SIC279; SIC291-SIC301; SIC305; SIC308; SIC317; SIC321-SIC355; SIC358; SIC359; SIC363; SIC371-SIC374; SIC376; SIC379; SIC391; SIC393; and SIC395-SIC399
SIC308 Misc. Plastics Products	76	SIC202-SIC206; SIC208; SIC209; SIC223-SIC225; SIC227-SIC238; SIC241-SIC244; SIC249-SIC252; SIC254-SIC267; SIC273; SIC275-SIC278; SIC282; SIC284; SIC285; SIC289; SIC295; SIC301-SIC306; SIC311-SIC329; SIC334; SIC341; SIC344; SIC345; SIC347; SIC348; SIC353; SIC355; SIC375; SIC391; and SIC394
SIC357 Computer And Office Equipment	37	SIC239; SIC271-SIC279; SIC285; SIC301; SIC306; SIC342; SIC349; SIC356; SIC358-SIC362; SIC364-SIC369; SIC375; SIC381-SIC387; and SIC393-SIC399
SIC355 Special Industry Machinery, Except Metalworking	36	SIC201; SIC227; SIC243; SIC245; SIC249; SIC253; SIC261; SIC323; SIC325-SIC334; SIC336; SIC339; SIC342-SIC346; SIC351-SIC354; SIC356; SIC363; SIC371-SIC373; SIC376; SIC379; and SIC391
SIC382 Laboratory Apparatus And Instruments	30	SIC221-SIC226; SIC228-SIC239; SIC251; SIC259; SIC262; SIC267; SIC274; SIC315; SIC349; SIC384; SIC385; SIC387; SIC393; SIC395; and SIC396
SIC366 Communications Equipment	23	SIC273; SIC275-SIC279; SIC321; SIC322; SIC333; SIC335; SIC336; SIC357-SIC362; SIC364; SIC365; SIC367; SIC369; SIC381; SIC386; and SIC399
SIC385 Ophthalmic Goods	17	SIC221; SIC222; SIC224-SIC226; SIC231-SIC238; SIC315; SIC382; SIC384; and SIC387
SIC283 Drugs	14	SIC201-SIC209; SIC281; SIC282; SIC284; SIC286; and SIC287
SIC353 Construction, Mining, And Materials Handling	14	SIC331; SIC332; SIC339; SIC343; SIC351; SIC352; SIC354-SIC356; SIC363; SIC371-SIC373; and SIC376
SIC289 Misc. Chemical Products	13	SIC207; SIC281-SIC284; SIC286; SIC287; SIC291-SIC299; SIC308; SIC324; and SIC348

Table 2A3 – Industries in the top 3 of the most technologically related to other industries

Industry	Frequency	Corresponding industries
SIC271 Newspaper	14	SIC261; SIC262; SIC272-SIC279; SIC306; SIC375; SIC394; and SIC399
SIC272 Periodical	13	SIC261; SIC262; SIC271; SIC273-SIC279; SIC306; SIC394; and SIC399
SIC237 Fur Goods	10	SIC225; SIC231-SIC236; SIC238; SIC315; and SIC391
SIC324 Cement, Hydraulic	10	SIC285; SIC295; SIC321-SIC323; and SIC325-SIC329
SIC202 Dairy Products	9	SIC201; SIC203-SIC209; and SIC283
SIC234 Women's, Misses', Children's, And Infants' Undergarments	9	SIC225; SIC231-SIC233; SIC235-SIC238; and SIC315
SIC205 Bakery Products	8	SIC201-SIC204; and SIC206-SIC209
SIC379 Misc. Transportation Equipment	8	SIC239; SIC245; SIC346; SIC352; SIC371; and SIC373-SIC375
SIC371 Motor Vehicles And Motor Vehicle Equipment	7	SIC239; SIC245; SIC346; SIC351; SIC372; SIC376; and SIC379
SIC204 Grain Mill Products	6	SIC202; SIC203; SIC205; SIC206; SIC209; and SIC283

more sense. For example, SIC 261 (Pulp Mills) and SIC 262 (Paper Mills) are likely to use similar technologies as producers in SIC 271 (Newspaper). Note that the correlation between patent citations and technological relatedness is actually only 0.146.

Dissimilarity measures

We control for coagglomeration due to co-dependence on a similar input by developing dissimilarity measures, following Faggio et al. (2017). We employ the BEA's use tables mentioned under input-output linkages. We discern the following groups: Agriculture related inputs (SIC: 7-9); Mining related inputs (SIC: 10-15); Water related inputs (SIC: 494); Energy related inputs (SIC: 491 and 492); Transport related inputs (SIC 40-48); Finance, Insurance and Real Estate (FIRE) Services (SIC: 60-67); Other services (SIC: 70-89). The dissimilarity in dependency on inputs from each of these groups between industry i and industry j is measured as one half of the absolute value of the difference in these shares.

Trade, technology, and transportation costs

Trade

To calculate import penetration we employ trade data from the U.N. comtrade database. As data are available on a yearly basis we use the same years as for the CBP, namely 1970, 1977, 1989, 1994, 1999, 2004, 2009, and 2014. U.N. comtrade also has concordance tables between the respective product codes and the SIC '87 classification used here. We correct imports for re-exports where necessary before dividing imports from low-wage countries over the total amount of imports, according to the value share approach (see Bloom et al., 2016).

Table 2A4 reports the names of countries that are defined as low-wage, *i.e.* they have a GDP per capita of less than 15% of that of the U.S. over the entire time period. We use data of the World Bank to obtain the GDP per capita of countries for each time period in our data.

An alternative measure for import penetration than the one in the main analysis is to divide imports from low-wage countries over the so-called apparent consumption in the U.S., which is equal to domestic production minus exports plus imports (see Bloom et al., 2016). Export data are available as well in the U.N. comtrade database, while the total value of shipments is obtained from the NBER-CES manufacturing industry database (see Bartelsman and Gray, 1996). This measure may be closer to the concept of import penetration as it gives the share of low-wage imports in the total value of products on the national market, but it has the drawback that data on the value of shipments for SIC industries 241, and 271 to 277 is missing after 1997, as these are no longer seen as manufacturing in the NAICS classification adopted in that year. Furthermore, the two databases derive their data from different sources, leading to discrepancies in the values. A final advantage of using the value share approach instead of the share in apparent consumption is that it allows to derive instruments

Table 2A4 – List of low-wage countries

Afghanistan	Albania	Algeria
Angola	Armenia	Azerbaijan
Bangladesh	Belize	Benin
Bhutan	Bolivia	Bosnia Herzegovina
Botswana	Burkina Faso	Burundi
Cabo Verde	Cambodia	Cameroon
Central African Republic	Chad	China
Colombia	Comoros	Congo
Côte d'Ivoire	Democratic Republic of the Congo	Djibouti
Dominica	Dominican Republic	Egypt
El Salvador	Eritrea	Ethiopia
Fiji	Gambia	Georgia
Ghana	Guatemala	Guinea
Guinea-Bissau	Guyana	Haiti
Honduras	India	Indonesia
Jordan	Kenya	Kiribati
Kosovo	Kyrgyzstan	Laos
Lesotho	Liberia	Madagascar
Malawi	Mali	Marshall Islands
Mauritania	Micronesia	Moldova
Mongolia	Montenegro	Morocco
Mozambique	Myanmar	Nepal
Nicaragua	Niger	Nigeria
North Macedonia	Pakistan	Papua New Guinea
Paraguay	Peru	Philippines
Rwanda	Samoa	São Tomé and Principe
Senegal	Serbia	Sierra Leone
Solomon Islands	Somalia	South Sudan
Sri Lanka	St Vincent and Grenadines	Sudan
Syria	Tajikistan	Thailand
Timor-Leste	Togo	Tonga
Tunisia	Turkmenistan	Tuvalu
Uganda	Ukraine	United Republic of Tanzania
Uzbekistan	Vanuatu	Vietnam
West Bank and Gaza	Yemen	Zambia
Zimbabwe		

for each industry for the entire time period. Nonetheless, we also present results in the robustness analysis using the import share in apparent consumption. These confirm our results, as is the case in Bloom et al. (2016).

Technology

To proxy for technological progress we calculate the routine employment share per industry and time period, following Autor et al. (2013).[29] We use the same IPUMS census samples of (Ruggles et al., 2018) that were also used for the spatial instruments. Autor et al. (2013) employ data from the Dictionary of Occupational Titles from 1977

[29]We note that this measure underestimates the decrease in routine task intensity, because the task content *within* occupations also has seen a decrease in routine task intensity, as discussed by Autor et al. (2003) and Autor and Dorn (2013).

to estimate the routine (T_k^R), manual (T_k^M), and abstract (T_k^A) task inputs on a scale from zero to ten per occupation.

Subsequently they derive the routine task intensity ($\mathcal{RTI}_k$) per occupation k through the following formula:

$$\mathcal{RTI}_k = \log(T_k^R) - \log(T_k^M) - \log(T_k^A). \tag{2A.1}$$

Then they apply a binary approach according to which all occupations in the 1980 IPUMS census sample with a routine task intensity value above the employment weighted 66$^{\text{th}}$ percentile are defined as routine task intensive jobs. Like Autor et al. (2013), we set zero values of task inputs to the score of the 5$^{\text{th}}$ percentile.

We follow this approach and then calculate the routine employment share $\mathcal{RSH}_{it}$ for each industry i for each year t, according to the following formula:

$$\mathcal{RSH}_{it} = \frac{\sum_{k=1}^{K} L_{ikt} \times 1(\mathcal{RTI}_k > \mathcal{RTI}^{P66})}{\sum_{k=1}^{K} L_{ikt}}, \tag{2A.2}$$

where L_{ikt} is the employment in occupation k in industry i at time t and $1(\mathcal{RTI}_k > \mathcal{RTI}^{P66})$ is an indicator function, which takes the value one when the occupation is defined as routine task intensive.

Transportation costs

We employ the use tables of the BEA to calculate the share of expenditure on transportation sectors (SIC 41-47) in the total use value of an industry.[30] This underestimates the transportation costs because the use of company-owned trucks driven by company personnel is not included in transportation cost expenditure. This underestimation is larger in the earlier periods than in the later periods as industries increasingly outsource transportation services to dedicated firms. For example the Commodity Flow Survey (CFS) of 1977 indicates that 35.8% of the ton-miles of U.S. manufacturing was performed by private trucks against 9.3% in the CFS of 2012.

To correct for this underestimation, we first derive expenditure per ton-mile on trucking by dividing the expenses on SIC 42 Trucking and warehousing in the use table over the ton-miles executed by motor carriers in the CFS. Then we multiply this by the number of ton-miles performed by private trucks. This estimated amount spend on private trucking is deduced from the total use value outside transportation and added to the expenditure on transportation. Due to coverage limitations and availability

[30]Note that by definition the total use value is equal to the total make value and therefore this definition approaches the often used iceberg transportation cost definition used in economic models.

issues, we only use the CFS of 1977, 1993, 2002, 2007, and 2012.

In a robustness analysis, we also show results using another sectoral transportation cost measure proposed by Glaeser and Kohlhase (2004), namely the value of a ton. As a higher value of a ton suggests that transportation costs are a small fraction of the total price. We use the data from the aforementioned CFS to calculate the value of a ton. Because data from 2002 onwards is only available at an aggregate industry level we calculate the measure at the 2-digit level. Furthermore, SIC 27 is not considered in the CFS so we set its value at the average of all manufacturing and include a dummy variable to indicate that the value is missing. Next to these data limitations this measure also has the disadvantage that changes in the value of a ton over time are not indicative of changes in transportation costs. Nonetheless, we show results using this measure in Appendix 2B.

Controls
To mitigate omitted variable bias we add extra control variables to the main specification and follow Oster (2019) in using the changes in coefficients of the variables of interest, together with changes in the R^2, to estimate the effect of other omitted variables that cannot be observed. We add the natural logarithm of the average establishment size and the capital-labour ratio per year and industry. The former is calculated by dividing the number of employees over the number of establishments in the CBP, while the latter is calculated by dividing the capital value over the total payroll using the NBER-CES manufacturing database by Bartelsman and Gray (1996). Note that in this last data set values are missing for SIC 241, and 271 to 277 after 1997, as mentioned earlier.

Additional descriptives

Table 2A5 presents the descriptive statistics of the dissimilarity indices, which indicate how dissimilar an industry pair is in their dependence on a certain input. The minimum value of 0, indicates that both industries have the same dependence on an input. Whereas the maximum of 0.336, after capping, for the agricultural dissimilarity index indicates that one industry obtains 67.2% more of its inputs from the agricultural sector (SIC 7) than the other. This is for example the case for the industry pair SIC 201 (Meat Products) and SIC 283 (Drugs), where the former receives 69.9% of its inputs from the agricultural sector and the latter only 0.2%.

Figure 2A1 presents histograms of the main variables of interest. The more skewed a distribution is, the larger the risk is that outliers drive results. In this respect, the distribution of coagglomeration values presents little risk as it resembles more or less a normal distribution and is only slightly skewed to the right. labour market pooling is

Table 2A5 – Descriptive statistics cont'd

Statistic	Mean	St. Dev.	Min	Max
Agricultural Dissimilarity Index	0.032	0.072	0.000	0.336
Mining Dissimilarity Index	0.020	0.050	0.000	0.341
Water Dissimilarity Index	0.017	0.020	0.0001	0.107
Energy Dissimilarity Index	0.011	0.012	0.0001	0.064
Transport Dissimilarity Index	0.030	0.030	0.0003	0.146
FIRE Dissimilarity Index	0.012	0.018	0.00004	0.096
Other Services Dissimilarity Index	0.0003	0.001	0.000	0.004

Note: The number of observations is 155680.

more strongly right skewed, but particularly input-output linkages and technological relatedness are right skewed. Note that the variables would be more skewed when outliers were not capped.

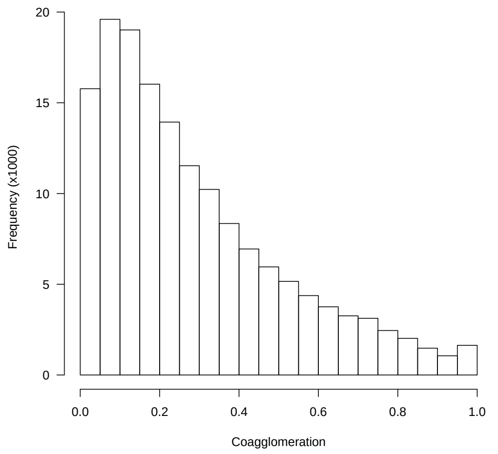
Figure 2A2 depicts the mean and variance of variables of interest over time. The mean is relatively stable over time for each of the variables. We emphasise that this does not mean that there are no changes in the extent to which industries coagglomerate, employ similar workers, buy or sell to each other, and can learn from each other. Large changes of multiple standard deviations can be found for all variables between industries. This is obscured in Figure 2A2 because the average value over all industries remains very much unchanged.

Also, coagglomeration, input-output linkages, and technological relatedness are measures that are normalised by each year. This means that their average values over time are not influenced by changes in region size, industry size, total number of inputs/outputs, total number of co-occurrences of technologies on patents but only by the relative distribution within these indicators across industries.[31] We emphasise that the main results of this Chapter are not driven by the fluctuations in the independent variables, as holding them equal at their 1994 values leads to similar results as in the main analysis, see Appendix 2B.

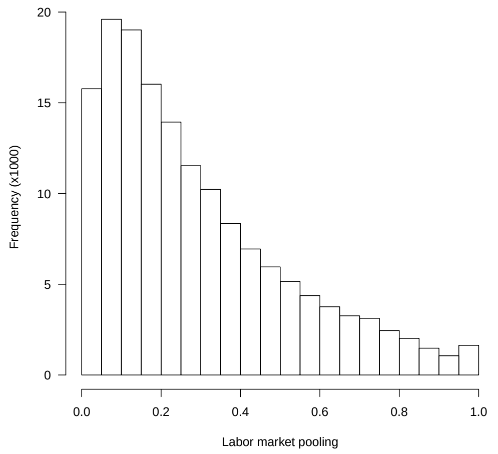
One can also observe in Figure 2A2 that the variance of each variable is also relatively stable over time, except in the case of coagglomeration. Here a steady decline over time occurs to about 40% of its original 1970 value. This observation is particularly interesting as Faggio et al. (2017) notes that low-technology industries have larger, more extreme, coagglomeration values. A decrease in the variance over time would then be in line with coagglomeration patterns becoming less low-technology oriented.

[31]On the other hand, labour market pooling is not normalised as it is a correlation measure and therefore its values are technically possible to move freely between 0 and 1 over time. Interestingly, the correlation of employment shares across job categories between industries is relatively stable.

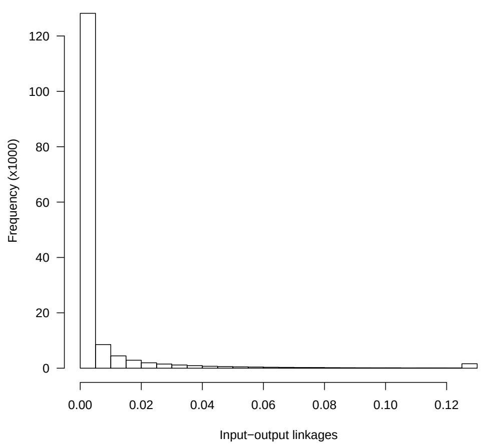
Figure 2A1 – Histograms of the variables of interest



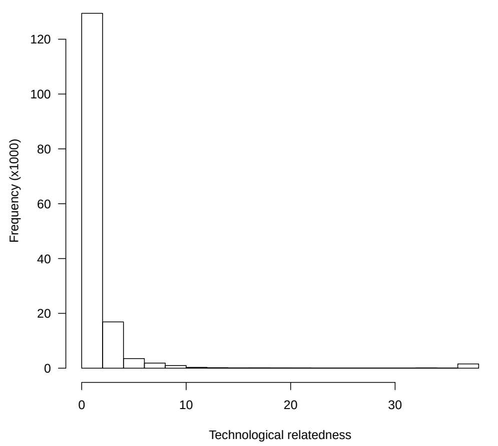
(a) Coagglomeration



(b) labour market pooling

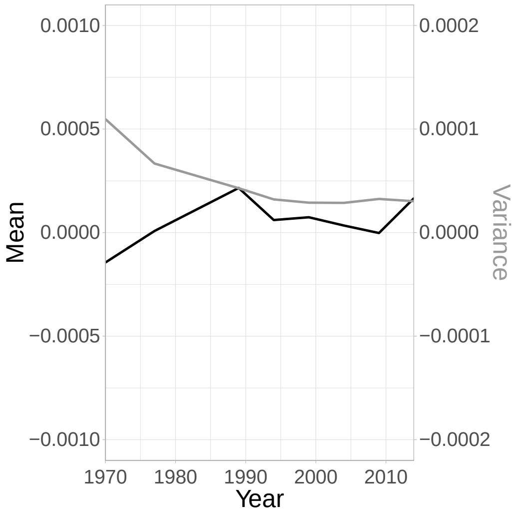


(c) Input-output linkages

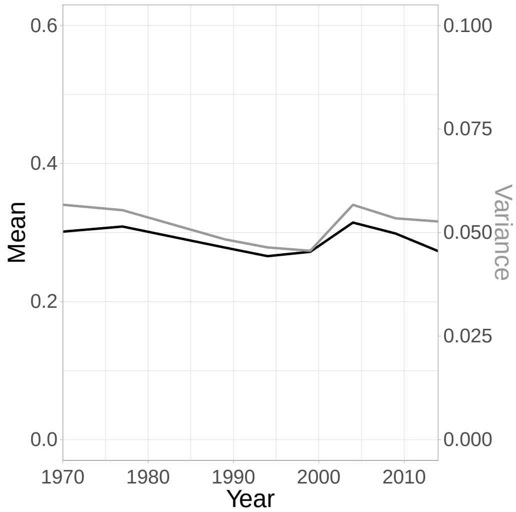


(d) Technological relatedness

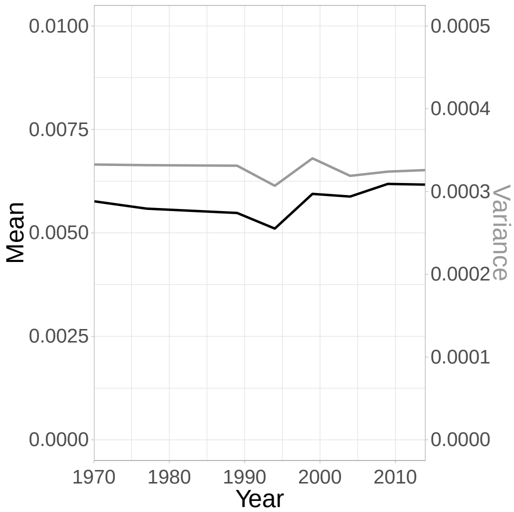
Figure 2A2 – Mean and variance of the variables of interest over time



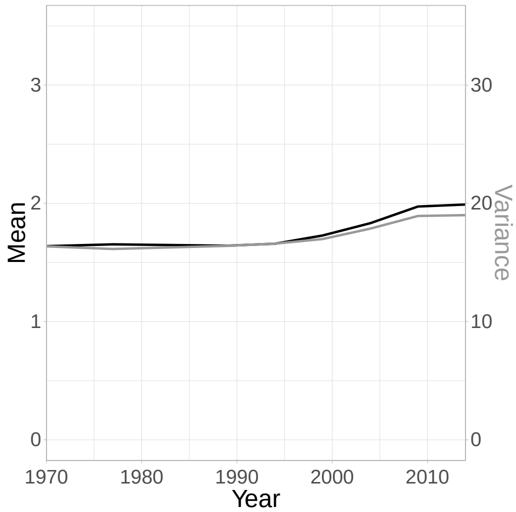
(a) Coagglomeration



(b) labour market pooling



(c) Input-output linkages



(d) Technological relatedness

Note: The black lines represent the mean value, while the gray lines represent the variance.

Table 2B1 – Main results w/ 1994 values
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*+ Dissimilarity measures*	*+ Industry× year fixed effects*	*Time trends*
	(1)	(2)	(3)	(4)
Labour market pooling	0.115***	0.109***	0.163***	0.216***
	(0.009)	(0.009)	(0.013)	(0.020)
Input-output linkages	0.096***	0.096***	0.098***	0.129***
	(0.010)	(0.010)	(0.010)	(0.015)
Technological relatedness	0.153***	0.150***	0.106***	0.071***
	(0.015)	(0.015)	(0.016)	(0.022)
Labour market pooling×				−0.022***
(year-1970)/10				(0.005)
Input-output linkages×				−0.013***
(year-1970)/10				(0.003)
Technological relatedness×				0.014***
(year-1970)/10				(0.005)
Dissimilarity measures	No	Yes	Yes	Yes
Industry i × year fixed effects	No	No	Yes	Yes
Industry j × year fixed effects	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	155,680	155,680	155,680	155,680
R^2	0.070	0.074	0.122	0.123

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

2B Appendix B. Sensitivity analyses

Step 1: Determinants of industry agglomeration

Constant definition of agglomeration determinants

We check that the changes in the importance of agglomeration determinants over time are not driven by changes in the Marshallian proxies or in the quality of measurement of these. To this end, we hold all variables but coagglomeration constant at their 1994 values and show the main findings in Table 2B1.

Variables are standardised to have zero mean and a standard deviation of one. As such the coefficient on labour market pooling in the naive specification shown in column (1), indicates that a standard deviation increase in the extent to which two industries can share workers leads on average to an increase of 0.115 of a standard deviation in coagglomeration. Like in the main results (in Table 2.3), the addition of industry×year fixed effects increases the coefficient on labour market pooling but decreases that on technological relatedness.

When we compare the coefficients of our preferred specification, column (3), we note that the coefficient on labour market pooling is significantly smaller when using 1994 values for the proxies (0.163 vs. 0.195). On the other hand, the coefficients on input-output linkages and technological relatedness are slightly larger but these differences are not statistically significant.

Even though the specific values of a variable in a year do seem to influence the results, the differences are not that large. Most importantly, column (4) shows that the time trends are strongly significant and have the same sign as in the main results. This confirms that changes in the importance of agglomeration over time are not driven by changes in (the quality of) our Marshallian proxies.

Industry-pair fixed effects

Table 2B2 presents the results when introducing industry-pair fixed effects. Therefore the coefficients are estimated based on within variation. These fixed effects capture most of the variation, as suggested by the R^2 exceeding 0.7. The coefficients are greatly reduced compared to all other results. According to column (1), a standard deviation increase in labour market pooling is only associated with 0.037 of a standard deviation increase in coagglomeration. The coefficients on input-output linkages and technological relatedness are also positive but even smaller and highly insignificant. Hence, the effects are considerably lower than in the preferred specification.

When adding the interaction with the time trends in column (2), the coefficients on the Marshallian proxies are higher and in the case of labour market pooling and input-output linkages significantly different from zero. Due to the interaction variables these coefficients represent the value of the base year 1970. In that year the coefficients on labour market pooling and input-output linkages were much larger compared to later years (see Table 2B8). The interaction terms do have the expected sign but are only significant for labour market pooling and input-output linkages.

Hence, we think the results with industry-pair fixed effects amplify measurement error, which leads to a strong bias towards zero. Even so, although not all the coefficients presented here are strongly statistically significant, these results point towards similar effects.

Coagglomeration at the county level

Table 2B3 reproduces the main results when calculating the coagglomeration index at the county level instead of the MSA level. When comparing the results of our preferred specification, we notice that a standard deviation increase in the labour market pooling measure leads to an increase of 0.223 of a standard deviation in coagglomeration at the county level versus 0.195 of a standard deviation at the MSA level. This difference is statistically significant and suggests that labour market pooling

Table 2B2 – Industry pair results
(Dependent variable: coagglomeration of industries i and j)

	Full model	*+ Time trends*
	(1)	(2)
Labour market pooling	0.037***	0.058***
	(0.010)	(0.014)
Input-output linkages	0.001	0.043***
	(0.008)	(0.013)
Technological relatedness	0.007	0.005
	(0.035)	(0.051)
Labour market pooling×		−0.009***
(year-1970)/10		(0.004)
Input-output linkages×		−0.017***
(year-1970)/10		(0.003)
Technological relatedness×		0.007
(year-1970)/10		(0.006)
Dissimilarity measures	Yes	Yes
Year fixed effects	Yes	Yes
Industrypair fixed effects	Yes	Yes
Observations	155,680	155,680
R^2	0.663	0.664

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

is a stronger determinant at finer geographical levels. The coefficients on input-output linkages and technological relatedness are not significantly different with respect to geography. The interactions with the time trends given in column (4) are also not significantly different from those at the MSA level in column (3) of Table 2B7. Hence, our results are robust to the choice of geographical area.

Weighted regressions

The baseline results in Table 2.3 present the results for the average industry pair. However, industries vary greatly in size. In Table 2B4, we take the log of respectively the number of employees, number of establishments, and value added to weight observations.

Columns (1), (3), (5) show our preferred specifications by, respectively, the number of establishments, employment, and value added. These results are not statistically significantly different the baseline specification in column (3) of Table 2.3.

In columns (2), (4), (6) we add time trends. These results are also not statistically significantly different from the corresponding unweighted results in column (3) of Table 2B7. However, note that the trend on technological relatedness is not statistically significant here because the standard errors are somewhat larger.

Table 2B3 – Main results at the county level
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*+ Dissimilarity measures*	*+ Industry× year fixed effects*	*Time trends*
	(1)	(2)	(3)	(4)
Labour market pooling	0.136***	0.132***	0.223***	0.347***
	(0.008)	(0.008)	(0.013)	(0.021)
Input-output linkages	0.072***	0.072***	0.073***	0.135***
	(0.008)	(0.008)	(0.008)	(0.015)
Technological relatedness	0.164***	0.163***	0.100***	0.070***
	(0.014)	(0.014)	(0.015)	(0.023)
Labour market pooling×				−0.051***
(year-1970)/10				(0.006)
Input-output linkages×				−0.026***
(year-1970)/10				(0.004)
Technological relatedness×				0.013**
(year-1970)/10				(0.006)
Dissimilarity measures	No	Yes	Yes	Yes
Industry i × year fixed effects	No	No	Yes	Yes
Industry j × year fixed effects	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	155,680	155,680	155,680	155,680
R^2	0.076	0.080	0.153	0.158

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

TABLE 2B4 – MAIN RESULTS WEIGHTED BY INDUSTRY SIZE

(Dependent variable: coagglomeration of industries i and j)

Weighted by:	*log(number of establishments)*		*log(employment)*		*log(value added)*	
	Base specification	*+ Time trends*	*Base specification*	*+ Time trends*	*Base specification*	*+ Time trends*
	(1)	(2)	(3)	(4)	(5)	(6)
Labour market pooling	0.193***	0.300***	0.195***	0.297***	0.190***	0.291***
	(0.011)	(0.021)	(0.011)	(0.021)	(0.011)	(0.021)
Input-output linkages	0.080***	0.137***	0.080***	0.133***	0.079***	0.133***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
Technological relatedness	0.090***	0.053***	0.095***	0.062***	0.098***	0.060***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Labour market pooling× (year-1970)/10		−0.044***		−0.042***		−0.042***
		(0.011)		(0.011)		(0.011)
Input-output linkages× (year-1970)/10		−0.023***		−0.023***		−0.023***
		(0.008)		(0.008)		(0.008)
Technological relatedness× (year-1970)/10		0.015		0.013		0.015
		(0.012)		(0.012)		(0.012)
Dissimilarity measures	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155,680	155,680	155,680	155,680	152,900	152,900
R^2	0.112	0.116	0.112	0.115	0.111	0.115

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

TABLE 2B5 – MAIN RESULTS W/ TWO-WAY CLUSTERING
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*+ Dissimilarity measures*	*+ Industry× year fixed effects*	*Time trends*
	(1)	(2)	(3)	(4)
Labour market pooling	0.114***	0.109***	0.195***	0.293***
	(0.018)	(0.018)	(0.027)	(0.040)
Input-output linkages	0.077***	0.076***	0.077***	0.135***
	(0.015)	(0.014)	(0.017)	(0.029)
Technological relatedness	0.161***	0.159***	0.104***	0.061*
	(0.031)	(0.031)	(0.031)	(0.034)
Labour market pooling× (year-1970)/10				−0.040***
				(0.009)
Input-output linkages× (year-1970)/10				−0.024***
				(0.007)
Technological relatedness× (year-1970)/10				0.017*
				(0.009)
Dissimilarity measures	No	Yes	Yes	Yes
Industry i × year fixed effects	No	No	Yes	Yes
Industry j × year fixed effects	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	155,680	155,680	155,680	155,680
R^2	0.067	0.070	0.116	0.119

Notes: Standard errors are two-way clustered at industry i and industry j, and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

Two-way clustering of standard errors

In the main results we cluster the standard errors at the industry pair (ij - ji) level. This means that each cluster contains 16 observations and coefficients are very precisely estimated, *i.e.* the standard errors are rather small. Other choices in clustering likely leads to larger standard errors. Therefore, we reproduce the main results using two-way clustering by industry i and industry j in Table 2B5.

Two-way clustering does not affect the estimated coefficients but imply larger standard errors. Two-way clustering implies that standard errors are roughly twice as large. Hence, the p-values decrease but virtually all coefficients remain statistically significant at the 99% confidence level. The exceptions are the coefficient on technological relatedness and its interaction with the time trend, which is now statistically significant at the 90% level.

Outliers – levels

In order to limit the influence of extreme values, which are present in all of the main variables, we cap values to the respective 1st percentile and 99th percentile. We compare the results of this approach to different alternative thresholds and the possibility of

TABLE 2B6 – OUTLIER TREATMENT – BASELINE

(Dependent variable: coagglomeration of industries i and j)

	Cap < 1% |> 99%	*Drop* < 1% |> 99%	*Cap* < 0.1% |> 99.9%	*Drop* < 0.1% |> 99.9%	*Cap* < 2.5% |> 97.5%	*Drop* < 2.5% |> 97.5%
	(1)	(2)	(3)	(4)	(5)	(6)
Labour market pooling	0.195***	0.165***	0.188***	0.191***	0.172***	0.122***
	(0.013)	(0.012)	(0.015)	(0.013)	(0.012)	(0.011)
Input-output linkages	0.077***	0.061***	0.067***	0.066***	0.073***	0.056***
	(0.009)	(0.007)	(0.011)	(0.010)	(0.008)	(0.006)
Technological relatedness	0.104***	0.043***	0.169***	0.113***	0.124***	0.086***
	(0.016)	(0.010)	(0.025)	(0.019)	(0.012)	(0.009)
Dissimilarity measures	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155,680	146,274	155,680	154,769	155,680	132,743
R^2	0.116	0.092	0.127	0.105	0.121	0.111

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

dropping outliers instead of capping these in Table 2B6.

Column (1) shows the main results from the preferred specification. Column (2) shows the results when *dropping* observations with at least one value below the respective 1st percentile or above the 99th percentile in coagglomeration, labour market pooling, input-output linkages or technological relatedness. As a result, the number of observations decreases from $155,680$ to $146,274$. Without these observations, the coefficients on all agglomeration determinants are somewhat smaller. In the case of technological relatedness the coefficient is even less than half the size of column (1). We prefer capping outliers over dropping as the information of the industry pairs that have high values for the variable of interest are paramount for understanding agglomeration patterns.

Columns (3) and (4) show the results when capping at the 0.1th/99.9th percentile. In this case only 911 observations are capped or dropped, respectively. Only the values on technological relatedness are significantly larger than when using the 1st and 99th threshold, see column (1) and (2), whereas labour market pooling and input-output linkages show more similar results. The differences between capping and dropping are generally small.

Columns (5) and (6) show the results when capping at the more strict 2.5th/97.5th percentile. In this case $22,937$ observations are dropped or capped. As this is almost 15% of all observations, excluding further modifications due to capped values on the control variables, we see this as a rather stringent way of dealing with outliers. The estimated coefficients in column (5) are not statistically different from those obtained in the main results, column (1). On the other hand, dropping once again leads to significantly smaller coefficients on all agglomeration determinants.

We also check the robustness of the trends in agglomeration determinants to these different outlier treatments in Section 2B.

Step 2: Changes in agglomeration determinants

Year-by-year results

Table 2B8 presents the by-year regression results that underlie Figure 2.2. Variables are standardised to have zero mean and a standard deviation of one. As such the coefficient on labour market pooling in 1970, see column (1), indicates that a standard deviation increase in the extent to which two industries industry can share workers leads on average to an increase of 0.311 of a standard deviation in coagglomeration.

As became clear from Figure 2.2, the coefficients on labour market pooling and input-

Table 2B7 – Time trend results
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*+ Dissimilarity measures*	*+ Industry× year fixed effects*	*Bias-adjusted*	*2SLS specification*
	(1)	(2)	(3)	(4)	(5)
Labour market pooling	0.130***	0.134***	0.293***	0.274***	**0.415*****
	(0.014)	(0.014)	(0.022)	(0.027)	(**0.010**)
Input-output linkages	0.137***	0.136***	0.135***	0.134***	0.117***
	(0.015)	(0.014)	(0.016)	(0.020)	(0.006)
Technological rel.	0.151***	0.148***	0.061***	0.059**	**0.014****
	(0.021)	(0.021)	(0.023)	(0.027)	(**0.007**)
Labour market pooling×	−0.006*	−0.010***	−0.040***	−0.051***	**−0.048*****
(year-1970)/10	(0.003)	(0.004)	(0.006)	(0.006)	(**0.004**)
Input-output linkages×	−0.025***	−0.024***	−0.024***	−0.025***	−0.023***
(year-1970)/10	(0.003)	(0.003)	(0.004)	(0.005)	(0.002)
Technological rel.×	0.004	0.004	0.017***	0.011*	**0.022*****
(year-1970)/10	(0.005)	(0.005)	(0.006)	(0.006)	(**0.002**)
Dissimilarity measures	No	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry i × year F.E.	No	No	Yes	Yes	Yes
Industry j × year F.E.	No	No	Yes	Yes	Yes
Observations	155680	155680	155680	155680	155680
R^2	0.069	0.072	0.119		0.114
R^2_{max}				0.154	
δ				1	
Kleibergen-Paap F-stat.					1093.21

Notes: Standard errors are clustered at the industry *ij*-*ji* level and in parentheses. Instrumented variables are indicated in bold; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

output linkages become smaller over time, whereas the coefficient on technological relatedness increases.

Interactions with year trends

Table 2B7 presents the results in which we test for linear time trends by interacting the determinants of agglomeration with the year of observation, which allows us to easily evaluate if these trends are statistically significant.

Variables are standardised to have zero mean and a standard deviation of one. The value of the interaction terms is by definition zero in 1970. The interaction between labour market pooling and the time trend in column (1) indicates that in ten years the coefficient on labour market pooling is 0.006 smaller.

Like in the baseline results reported in Table 2.3, the inclusion of dissimilarity measures has no significant effect on the coefficients (see column (2)). The addition of industry×year fixed effects in column (3) amplifies the coefficient on labour market

TABLE 2B8 – CROSS-SECTIONAL RESULTS BY YEAR
(Dependent variable: coagglomeration of industries i and j)

	1970	*1977*	*1989*	*1994*	*1999*	*2004*	*2009*	*2014*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Labour market pooling	0.308***	0.276***	0.213***	0.141***	0.167***	0.124***	0.148***	0.175***
	(0.025)	(0.021)	(0.016)	(0.015)	(0.016)	(0.014)	(0.016)	(0.015)
Input-output linkages	0.129***	0.110***	0.092***	0.102***	0.072***	0.070***	0.027***	0.017**
	(0.017)	(0.014)	(0.012)	(0.011)	(0.010)	(0.011)	(0.009)	(0.009)
Technological relatedness	0.088***	0.059***	0.073***	0.074***	0.115***	0.109***	0.138***	0.172***
	(0.026)	(0.020)	(0.019)	(0.018)	(0.020)	(0.017)	(0.019)	(0.018)
Industry i fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry j fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dissimilarity measures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,460	19,460	19,460	19,460	19,460	19,460	19,460	19,460
R^2	0.086	0.116	0.129	0.125	0.139	0.155	0.164	0.201

Notes: Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

pooling but decreases the one on technological relatedness. Recall that the specification in column (3) is our preferred specification as it controls for the year-specific tendency of each industry to coagglomerate.

Column (4) presents the Oster-style bias-adjusted coefficients. To verify whether other unobserved omitted variables are important, we use changes in coefficients of the variables of interest, as well as changes in the R^2, once the dissimilarity measures are included. The results are not significantly different from our preferred specification.

Column (5) presents the results of the instrumental variable regression. Here we use Marshallian proxies calculated in areas where industry i is present but industry j is (virtually) absent to instrument our measures for labour market pooling and technological relatedness. Proxies obtained from these areas preclude reverse causality in which coagglomeration leads to the formation of Marshallian links. The coefficient on labour market pooling is stronger, whereas the one on technological relatedness is smaller. The trends remain strongly significant and of a similar magnitude.

Outliers – trends

In Table 2B9, we repeat a similar exercise as in Section 2B but now include the trends in agglomeration determinants to the regressions. Column (1) again shows the baseline specification from the main results. As in Table 2B6, column (2), where outliers are dropped, we show much smaller coefficients, this also holds for the trends, of which the trend in technological relatedness even becomes statistically insignificant.

Columns (3) and (4) report the results when capping at the smaller 0.1^{st}/ 99.9^{th} percentile. Both columns are not statistically significantly different from column (1). Nonetheless, the coefficients on the base and trend of technological relatedness in column (3) stand out. Where the trend is positive, but insignificant, and the base is much higher than in other specifications. This result is mainly driven by the year 1970. The year-by-year results reported in Table 2B8 show that the coefficient on technological relatedness in 1970 is larger than in 1977 to 1994. Small changes in outlier treatments can therefore have strong effects on the results. When dropping 1970, however, the trend on technological relatedness is strong and significantly positive, as can be seen in Table 2B10. Note as well that the standard errors on the coefficients are higher, in particular in the case of technological relatedness. This is likely because outliers still exert a disproportionally large influence on the variation.

Columns (5) and (6) show the results when capping at the more restrictive $2.5^{\text{st}}/97.5^{\text{th}}$ percentile level. As in Table 2B6, coefficients are notably smaller when compared to the first column, particularly when dropping outliers. The notable exception is technological relatedness. Once again, this is due to the values of 1970. When removing

TABLE 2B9 – DEALING WITH OUTLIERS – TRENDS IN AGGLOMERATION DETERMINANTS
(Dependent variable: coagglomeration of industries i and j)

	Cap < 1% |> 99%	*Drop* < 1% |> 99%	*Cap* < 0.1% |> 99.9%	*Drop* < 0.1% |> 99.9%	*Cap* < 2.5% |> 97.5%	*Drop* < 2.5% |> 97.5%
	(1)	(2)	(3)	(4)	(5)	(6)
Labour market pooling	0.293***	0.241***	0.293***	0.304***	0.244***	0.182***
	(0.022)	(0.021)	(0.026)	(0.024)	(0.021)	(0.022)
Input-output linkages	0.135***	0.093***	0.114***	0.113***	0.125***	0.080***
	(0.016)	(0.013)	(0.019)	(0.016)	(0.014)	(0.012)
Technological relatedness	0.061***	0.038**	0.151***	0.053	0.116***	0.105***
	(0.023)	(0.017)	(0.046)	(0.035)	(0.020)	(0.018)
Labour market pooling×	−0.040***	−0.030***	−0.043***	−0.046***	−0.029***	−0.024***
(year-1970)/10	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)	(0.006)
Input-output linkages×	−0.024***	−0.013***	−0.020***	−0.019***	−0.021***	−0.010***
(year-1970)/10	(0.004)	(0.003)	(0.005)	(0.005)	(0.004)	(0.003)
Technological relatedness×	0.017***	0.002	0.008	0.023**	0.004	−0.007
(year-1970)/10	(0.006)	(0.004)	(0.013)	(0.010)	(0.005)	(0.005)
Dissimilarity measures	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	155,680	146,274	155,680	154,769	155,680	132,743
R^2	0.119	0.093	0.130	0.109	0.123	0.112

Notes: Standard errors are clustered at the industry ij-ji level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

1970 from the data, the base coefficients on technological relatedness decrease and those on the trends increase. Hence, the presumed absence of a trend in technological relatedness is entirely due to 1970.

Table 2B10 shows the results of the same specifications as in Table 2B9 but when 1970 is left out of the data. Unsurprisingly, the total number of observations decreases from 155,680 to 136,220 when outliers are not dropped.

The results are particularly interesting with respect to the trend in knowledge spillovers, which is positive and significant in all instances where outliers are capped and also when these are dropped using the $0.1^{st}/99.9^{th}$ percentile thresholds. As said, we find it more reasonable to cap than to drop outliers as those with more extreme coagglomeration patterns or Marshallian links should not to be disregarded when trying to understand coagglomeration.

All in all, the different outlier treatments presented here and in Section 2B reveal that there is a non-negligible effect of outliers on the regressions and that different ways of handling these can lead to somewhat different results. Nonetheless, the interpretation of both levels and trends in agglomeration determinants does not materially change.

Step 3: Exploring industry-level heterogeneity

Including more detailed industry characteristics

In the main results, we have shown the effect of the transportation costs of goods on input-output linkages but not on labour market pooling and knowledge spillovers. Also, we add extra control variables, on the average establishment size and the capital-labour ratio, in the bias-adjusted procedure but did not add these control variables to the OLS estimation.

The average establishment size is likely to influence the importance of external economies of agglomeration as large establishments can internalise its benefits reducing the need to coagglomerate (see Vernon, 1960; Chinitz, 1961). For similar reasons Faggio et al. (2020) use the average establishment size of incumbent firms as an explanatory variable. However, our data do not allow us to distinguish between new and incumbent firms. The capital-labour ratio is likely to influence the demand for certain agglomeration externalities according to the product life cycle (see Duranton and Puga, 2001; Neffke et al., 2011b).

On the other hand, the establishment size and capital-labour ratio are also clearly influenced by trade competition and technological progress, (see Brynjolfsson and Hitt, 2000; Holmes and Stevens, 2014; Bloom et al., 2016; Pierce and Schott, 2016), thereby capturing part of the effect of our proxies for trade and technology. As such, the

TABLE 2B10 – OUTLIER TREATMENT – TRENDS IN AGGLOMERATION DETERMINANTS W/O 1970

(Dependent variable: coagglomeration of industries i and j)

	Cap < 1% |> 99%	*Drop* < 1% |> 99%	*Cap* < 0.1% |> 99.9%	*Drop* < 0.1% |> 99.9%	*Cap* < 2.5% |> 97.5%	*Drop* < 2.5% |> 97.5%
	(1)	(2)	(3)	(4)	(5)	(6)
Labour market pooling	0.314***	0.278***	0.302***	0.338***	0.268***	0.211***
	(0.024)	(0.024)	(0.029)	(0.028)	(0.022)	(0.024)
Input-output linkages	0.158***	0.112***	0.128***	0.123***	0.149***	0.096***
	(0.017)	(0.015)	(0.021)	(0.019)	(0.016)	(0.015)
Technological relatedness	0.031	0.041**	0.097*	−0.016	0.100***	0.101***
	(0.026)	(0.019)	(0.055)	(0.041)	(0.021)	(0.020)
Labour market pooling×	−0.040***	−0.038***	−0.039***	−0.049***	−0.033***	−0.031***
(year-1970)/10	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)	(0.007)
Input-output linkages×	−0.029***	−0.016***	−0.021***	−0.019***	−0.026***	−0.012***
(year-1970)/10	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)
Technological relatedness×	0.028***	0.001	0.029*	0.045***	0.011**	−0.006
(year-1970)/10	(0.007)	(0.005)	(0.016)	(0.013)	(0.005)	(0.005)
Dissimilarity measures	Yes	Yes	Yes	Yes	Yes	Yes
Industry i × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry j × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	136,220	128,038	136,220	135,438	136,220	116,114
R^2	0.135	0.104	0.149	0.124	0.139	0.118

Notes: Standard errors are clustered at the industry ij-ji level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

variables added can be seen as proxy controls, which allows us to better control for omitted variable bias but also absorbs part of the overall effects of trade and technology (see Angrist and Pischke, 2008). Therefore the estimated coefficients should be seen as a lower bound of the true effects. The addition of these variables also yields insights on the channels through which trade and technology influence industrial organisations and therefore coagglomeration patterns.

We also explore the effects of increased R&D expenditure and increasing skill intensity, which are also closely associated with trade competition and technological progress. Industrial R&D expenditure data are obtained from the National Science Foundation and divided over the number of employees according to the CBP.[32] Following Pierce and Schott (2016), the skill intensity is measured as the ratio of non-production workers to production workers, which is obtained from the NBER-CES manufacturing database.This measure is related to the share of highly educated workers used by Faggio et al. (2020). Results are reported in Table 2B11.

Column (1) shows us that the addition of transportation costs of goods does not change the coefficients on trade and technology much, in comparison to the main results in Table 2.6. The coefficient on transportation costs is positive and significant. However, an industry like SIC 372 (aircraft and parts), which has the lowest transportation expenditure share, differs in more dimensions than just transportation costs from SIC 327 (concrete, gypsum, and plaster products), which has the highest expenditure share. Hence, transportation costs of goods are likely negatively correlated to a higher average product value, which is in turn correlated with trade competition and technological progress. The coefficient on the capital-labour ratio is positive and strongly statistically significant, which is at a first glance the opposite of what one would expect given the increasing trend in this variable and the decreasing trend in labour market pooling.

However, the coefficient is in line with the 'nursery city hypothesis' (see Duranton and Puga, 2001), because an industry that highly invests in machinery is likely to standardise its production process and therefore move out of the experimental phase (associated with knowledge spillovers) to the mass producing phase, where labour market pooling is more important.

In column (3), we also add the R&D expenditure per employee and the skill intensity to the specification, which are the most clearly associated with trade competition and

[32]The data are at a more aggregate industry level for most SIC industries, therefore we sum employment at this industrial level and assign the same R&D expenditures per employee for the respective SIC industries involved. We deflate the dollar values to 1987 prices using industry-specific deflators for investments developed in the NBER-CES manufacturing database by Bartelsman and Gray (1996).

TABLE 2B11 – INCLUDING MORE INDUSTRY CHARACTERISTICS RESULTS

Dependent variable:	*Labour market pooling* $\hat{\alpha}$			*Input-output linkages* $\hat{\beta}$			*Technological relatedness* $\hat{\gamma}$		
	Base	*+ Extra controls*	*+ R&D & skill int.*	*Base*	*+ Extra controls*	*+ R&D & skill int.*	*Base*	*+ Extra controls*	*+ R&D & skill int.*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Import penetration	−0.043***	−0.037***	−0.038***	−0.019***	−0.015**	−0.015**	0.024*	0.021	0.023*
	(0.008)	(0.008)	(0.008)	(0.007)	(0.006)	(0.006)	(0.013)	(0.013)	(0.013)
Routine employment share	0.108***	0.112***	0.093***	−0.002	0.001	0.003	−0.057***	−0.058***	−0.030*
	(0.014)	(0.014)	(0.014)	(0.009)	(0.008)	(0.009)	(0.016)	(0.016)	(0.016)
Transportation costs	0.016**	0.006	0.004	−0.001	−0.0004	−0.0001	−0.043**	−0.036**	−0.032*
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.018)	(0.017)	(0.017)
Average establishment size		−0.016	−0.006		0.012	0.012		0.013	0.001
(log)		(0.010)	(0.010)		(0.010)	(0.009)		(0.027)	(0.027)
Capital labour ratio		0.057***	0.063***		0.013	0.013		−0.031	−0.037*
		(0.008)	(0.008)		(0.012)	(0.012)		(0.021)	(0.021)
R&D per employee			−0.063***			0.002			0.075***
(log)			(0.011)			(0.009)			(0.029)
Skill intensity			0.023*			0.003			−0.009
			(0.012)			(0.010)			(0.028)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
R^2	0.135	0.164	0.187	0.045	0.067	0.068	0.032	0.034	0.044

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. Two dummy variables are included in the specification to indicate missing values in the census of manufactures and the CFS, respectively, but not reported in this table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

technological progress. Again, this does not materially influence the coefficients related to import penetration and routine employment share. We find that R&D expenditure is strongly and negatively associated with labour market pooling, which is in line with the nursery city hypothesis. The positive coefficient on skill-intensity cannot be reconciled with this hypothesis, but we note that it is negative and significant when using an univariate regression, in line with expectations. Faggio et al. (2020) find a statistically insignificant coefficient on the share of highly-educated workers both in an univariate setting as in a multivariate setting. The coefficient on average establishment size is not statistically significant in columns (2) and (3). This is in line with Faggio et al. (2020), who also do not find a statistically significant effect of average establishment size.

Columns (4), (5), and (6) show the results when the coefficient on input-output linkages is used as the dependent variable. The addition of the other variables in columns (5) and (6) does not lead to stronger and more significant coefficients. At the same time, the coefficient on trade competition hardly changes. This is in line with the results of Faggio et al. (2020), who also do not find statistically significant results for average establishment size and the share of highly educated workers in explaining input-output linkages. We further analyze the impacts on input-output linkages in Appendix 2B.

Columns (7), (8), and (9) show the results when the coefficient on technological relatedness is used as the dependent variable. Transportation costs of goods have a considerable impact on the importance of knowledge spillovers. The effect is negative, which is in line with the observation that more sophisticated and high-value products exhibit relatively low transportation costs, but are knowledge intensive. Relatedly, Behrens et al. (2018) show that industries with low transport costs are also the most geographically concentrated, which makes sense given that knowledge spillovers attenuate sharply over distance and therefore drive concentration. The addition of the other variables in columns (8) and (9) does reduce the coefficient of transportation costs, but does not fully explain the effect. This may be because the other proxies are also imperfect. The coefficients on the capital-labour ratio and R&D expenditure are particularly strong and significant and in line with expectations. The coefficients on average establishment size are insignificant just as in Faggio et al. (2020) and the coefficient on our measure for skill intensity is insignificant here, while Faggio et al. (2020) find a positive statistically significant coefficient on the share of higher educated workers.

In Table 2.6 we have also shown results of the effect of transportation on input-output linkages with bias-adjusted coefficients following Oster (2019) and using instruments. We have also ran these estimation strategies for labour market pooling and knowledge spillovers. We found that these results did not lead to very different insights compared to the OLS results presented here in columns (1), (4), and (7). Therefore, we chose not to include them here and focus on other industry characteristics.

Weighted regressions

In the main results, we have shown unweighted regressions. Faggio et al. (2020) weight their results by the inverse of the standard error. In Table 2B12, we reproduce the results of Table 2.6 when using this approach.

The interpretation of the coefficients is the same as in the main results. The coefficient on import penetration in column (1) indicates that an increase of a standard deviation in import penetration is associated with a decrease of 0.026 in the size of the coefficient on labour market pooling. This is significantly smaller than the 0.042 found in Table 2.6. However, the interpretation that import penetration strongly negatively influences the need to coagglomerate to make use of labour market pooling remains valid as both coefficients are strongly statistically significant. The same holds for the coefficient on routine employment share, which is significantly smaller here but shows the same implications.

Column (2) and column (3) shows respectively the bias-adjusted estimates, following Oster (2019), and the instrumented estimates, where import penetration is instrumented by values from other high-income countries and routine employment share by values from areas where each industry does not coagglomerate to make use of labour market pooling. These are again significantly smaller than the corresponding estimates in Table 2.6.

The results on input-output linkages are given in Columns (4), (5), and (6). The coefficients are not statistically significantly different from corresponding estimates in Table 2.6. However, they are more precisely estimated. As a result, the negative association between the relevance of input-output linkages and import penetration is strongly statistically significant in this set-up. This increases our confidence that import penetration plays a strong role in the decreasing importance of input-output linkages over time. The control variable log value of a ton in 1970 is also statistically significant in this specification.

The results on knowledge spillovers are given in columns (7), (8), and (9). Like with labour market pooling, these are smaller than the corresponding estimates in Table 2.6. For input-output sharing we find near-identical results, while for knowledge spillovers the effect of routine employment share is somewhat smaller and ceases to be strongly statistically significant.

TABLE 2B12 – TRADE, TECHNOLOGY, AND TRANSPORTATION COSTS – WEIGHTED RESULTS

Dependent variable:	*Labour market pooling* $\hat{\alpha}$			*Input-output linkages* $\hat{\beta}$			*Technological relatedness* $\hat{\gamma}$		
	OLS	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*	*OLS*	*Bias-adj.*	*2SLS*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Import penetration	−0.026***	−0.028***	**−0.037*****	−0.015***	−0.008	**−0.018*****	0.019***	0.025***	**0.010**
	(0.004)	(0.005)	(**0.006**)	(0.003)	(0.008)	(**0.005**)	(0.005)	(0.009)	(**0.007**)
Routine employment share	0.071***	0.086***	**0.094*****	0.001	0.009	**0.003**	−0.012	−0.015*	**−0.013**
	(0.010)	(0.012)	(**0.011**)	(0.006)	(0.006)	(**0.007**)	(0.008)	(0.009)	(**0.010**)
Transportation costs				−0.003	0.0001	**0.007**			
				(0.004)	(0.005)	(**0.010**)			
Value of a ton in 1970 *(log)*						0.006**			
						(0.003)			
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extra controls	No	Yes	No	No	Yes	No	No	Yes	No
Observations	1120	1120	1120	1120	1120	1120	1120	1120	1120
R^2	0.092			0.061			0.009		
R^2_{max}		0.119			0.08			0.012	
δ		1			1			1	
Kleibergen-Paap F-statistic			77.79			22.35			100.23

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. Instrumented variables are indicated in bold. Extra controls consist of the natural logarithm of the average establishment size and the capital labour ratio. A dummy variable is added to indicate missing data in the CFS on the instrumental variable value of a ton for industries belonging to SIC27 but its coefficient is not reported in the table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

Import penetration in apparent consumption
Our main results are based on the value share approach, the share of imports from low-wage countries in the total of imports. However, this does not actually measure the import penetration in the economy as the nationally produced and consumed goods are not taken into account.Therefore, we show results here using the import penetration in apparent consumption (see Bloom et al., 2016).

We obtain import and export data from the U.N. Comtrade database as before, while the total value of shipments, *i.e.* domestic production, is obtained from the NBER-CES manufacturing industry database (see Bartelsman and Gray, 1996). In Appendix 2A, we discussed several drawbacks of this measure in its application; namely missing data values on the value of shipments for SIC industries 241, and 271 to 277 after 1997, the mismatch between the two different data sources, and the impossibility to obtain data on the value of shipments for other high-wage countries over the same time period and in a similarly detailed industry classification to build an instrument.

We present the results from the OLS specification employing this measure in Table 2B13. The results are hardly different. The impact of import penetration on input-output linkages in column (2) is now slightly less strong and not statistically significant at conventional levels, whereas it is stronger and more strongly significant in the case of technological relatedness. However, all the differences in the coefficients are not statistically significantly different from the coefficients in the main analysis. The other coefficients are virtually unchanged.[33]

Hence, these results support the use of the value share measure in the main analyses and are in line with Bloom et al. (2016), who also obtain similar results when using import penetration in apparent consumption next to the value share approach.

Trade, technology, and transportation costs as interaction variables
In Table 2B14, we present the results of introducing trade, technology, and transportation costs as interaction variables. Column (1) presents the results of a naive specification, with only the dissimilarity measures and year fixed effects added as controls. Routine employment share and transportation costs are strongly and positively associated with coagglomeration. This is in line with the finding of Faggio et al. (2017)

[33]We also used the import penetration in apparent consumption by all countries instead of just the low-wage countries. This does not materially influence the results.

TABLE 2B13 – ALTERNATIVE TRADE MEASURE RESULTS

	Labour market pooling	Input-output linkages	Technological relatedness
Dependent variable:	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
	OLS	*OLS*	*OLS*
	(1)	(2)	(3)
Import penetration	-0.033^{***}	-0.007	0.058^{***}
(in apparent consumption)	(0.010)	(0.007)	(0.018)
Routine employment share	0.103^{***}	-0.006	-0.066^{***}
	(0.014)	(0.009)	(0.019)
Transportation costs		-0.002	
		(0.008)	
Year fixed effects	Yes	Yes	Yes
Observations	1,120	1,120	1,120
R^2	0.131	0.038	0.034

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. A dummy variable is included in the specification to indicate missing values in the census of manufactures but not reported in this table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

that low-tech industries have larger coagglomeration values.[34]

In column (2), we present the results of a specification that includes industry $i\times$year and industry $j\times$ fixed effects. As a result, the coefficients on import penetration, routine employment share, and transportation costs are somewhat lower. We compare the effect size with respect to those found in the main analysis, by dividing the coefficient on the interaction by the main effect of the agglomeration determinant. The effect of the interaction between labour market pooling and import penetration is a reduction of about 25% ($-0.053/0.194$), while that with the routine employment share is an increase of about 71% ($0.138/0.194$). These percentages are similar to the direction and size of the coefficients in the main OLS results in Table 2.6, when these would be compared to the average coefficient on labour market pooling from the first step. In the case of technological relatedness, the size effect is much larger here in the case of import penetration with about 96% ($0.045/0.047$) and for routine employment share, which is about -43% ($-0.020/0.047$). This is also due to the larger average coefficient on technological relatedness in the first step. The effect of import penetration on input-output linkages is also larger here about -43% ($-0.037/0.086$).

Interestingly, the coefficient on the interaction between input-output linkages and trade,

[34]Faggio et al. (2017) find this counterintuitive as they expected high-technology industries to have larger coagglomeration values according to the nursery city hypothesis (Duranton and Puga, 2001). However in diverse (usually large) cities, high-tech industries are relatively more present but no industry is dominant and coagglomeration values are therefore smaller. By contrast, low-tech industries are more likely to cluster together in small specialised towns, where co-agglomeration values will be high.

Table 2B14 – One-step estimator with interactions
(Dependent variable: coagglomeration of industries i and j)

	Naive specification	*Including industry fixed effects*
	(1)	(2)
Labour market pooling	0.123***	0.194***
	(0.009)	(0.012)
Input-output linkages	0.084***	0.086***
	(0.008)	(0.008)
Technological relatedness	0.096***	0.047***
	(0.015)	(0.016)
Import penetration	−0.003	
	(0.004)	
Routine employment share	0.075***	
	(0.011)	
Transportation costs	0.010**	
	(0.004)	
Labour market pooling×	−0.014***	−0.053***
import penetration	(0.005)	(0.006)
Labour market pooling×	0.113***	0.138***
routine employment share	(0.013)	(0.015)
Input-output linkages×	−0.039***	−0.037***
import penetration	(0.006)	(0.006)
Input-output linkages×	0.0005	−0.006
routine employment share	(0.009)	(0.009)
Input-output linkages×	−0.007	−0.008*
transportation costs	(0.005)	(0.005)
Technological relatedness×	0.028***	0.045***
import penetration	(0.007)	(0.007)
Technological relatedness×	−0.013	−0.020*
routine employment share	(0.010)	(0.011)
Dissimilarity measures	Yes	Yes
Year fixed effects	Yes	Yes
Industry i fixed effects.	No	Yes
Industry j fixed effects.	No	Yes
Industry i × year fixed effects.	No	Yes
Industry j × year fixed effects.	No	Yes
Observations	153,456	153,456
R^2	0.078	0.124

Notes: SIC 237 (Fur goods) is left out of the regression as it only has 40 employees in 2014 in total. Standard errors are clustered at the industry *ij-ji* level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

and the one on technological relatedness and trade is more precisely estimated, whereas the interaction between routine employment share and technological relatedness is less statistically significant.[35]

Trade, transportation costs, and importance of input-output linkages

In this subsection, we further explore sensitivity with respect to the impacts of trade competition and transportation on the importance of input-output linkages.

First of all, we can use the value of a ton measure as an alternative proxy for transportation cost. When the value of a ton is higher transporting it becomes a smaller fraction of the total selling price and therefore cheaper, as suggested by Glaeser and Kohlhase (2004). This proxy has the advantage of not being directly impacted by input-output linkages but it has the drawback that changes in this value over time may not reflect changes in transportation costs.[36] Moreover, some industries are not covered in the CFS.

Secondly, one may argue that it is no so much import penetration within the sector itself that matters for input-output linkages but *import penetration within the sectors from which inputs are obtained*. This is measured as the sum of the share of import penetration in each industry times the share of inputs received from these industries.

Thirdly, the distinction between trade competition in outputs and inputs suggests it may be useful to discern between input linkages and output linkages. Therefore, we, thirdly, reproduce the results for input and output linkages separately.

Column (1) in Table 2B15 replicates the baseline result. Recall that only trade competition has an economically and statistically significant impact on input-output linkages. In column (2), the iceberg transportation cost measure is replaced by the natural logarithm of the average value of a ton. The coefficient on this proxy is highly insignificant and close to zero. This seems to confirm that transportation costs of goods are unlikely to be an important determinant of the intensity of input-output linkages.

In column (3) import penetration is replaced by *input* import penetration. The coefficient on this estimate is not significantly different from its counterparts in the previous columns but the standard error on this measure is much higher. This suggests that import penetration *within supplying sectors* and import penetration *in the sector*

[35]Note that we dropped SIC 237 (fur goods) from the sample as it only has 40 employees in 2014, which results in extreme coagglomeration levels. The interaction variables are more susceptible to outliers than the other variables in this specification compared to those in the two-step approach, which are capped. Therefore, we left this industry out as including it would have a strong influence on some of the coefficients on the interaction variables, even though it is such a small industry.

[36]Notably, the price-adjusted average value of a ton actually decreases over a time. According to the transportation cost interpretation this would suggest that products become *more* expensive to transport.

TABLE 2B15 – INPUT-OUTPUT LINKAGES RESULTS

Dependent variable:	*Input-output linkages* $\hat{\beta}$			*Input linkages* $\hat{\beta}_I$			*Output linkages* $\hat{\beta}_O$		
	Base	*Value of a Ton.*	*Input-penetration*	*Base*	*Value of a Ton.*	*Input-penetration*	*Base*	*Value of a Ton.*	*Input-penetration*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Import penetration	−0.019***	−0.018***		−0.019**	−0.019***		−0.018	−0.015	
	(0.007)	(0.006)		(0.007)	(0.007)		(0.053)	(0.054)	
Input import penetration			−0.014			−0.019			0.018
			(0.013)			(0.016)			(0.075)
Routine employment share	−0.002	0.0001	−0.004	0.007	0.010	0.007	−0.004	−0.003	−0.017
	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.035)	(0.037)	(0.030)
Transportation costs	−0.001		−0.002	−0.008		−0.009	0.020		0.020
	(0.008)		(0.008)	(0.008)		(0.008)	(0.040)		(0.040)
Value of a ton *(log)*		−0.006			−0.002			−0.018	
		(0.006)			(0.009)			(0.036)	
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
R^2	0.045	0.049	0.038	0.046	0.046	0.042	0.017	0.017	0.016

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. Standard errors are bootstrapped. A dummy variable is added to indicate missing data in the CFS on the value of a ton for industries belonging to SIC27 but its coefficient is not reported in the table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

itself have similar effects on the importance of input-output linkages. This provides additional support for the claim that the impact of import penetration on input-output linkages runs through trade competition, rather than through the level of technological sophistication as suggested by Faggio et al. (2017). This is likely because localised input linkages are replaced by input linkages in low-wage countries.

Columns (4), (5), and (6) show the results when distinguishing between input and output linkages in the first step and using the coefficient on input linkages as the dependent variable in the second step. The coefficients and standard errors are very similar to when using input-output linkages combined. As could be expected the coefficient on input import penetration, in column (6), is a bit larger. This provides additional, albeit suggestive, evidence that localised input linkages (rather than output linkages) are replaced by input linkages in low-wage countries.

Columns (7), (8), and (9) show the results when using the coefficient on output linkages as the dependent variable in the second step. Interestingly, none of the coefficients in these specifications are statistically significant. The standard errors are much larger and the explained level of variance is smaller, as indicated by the low R^2. This is likely because output linkages contain more extreme values, which indicates that industries are more likely to sell strongly to, rather than buy from, a few manufacturing industries.

First-stage results

In columns (3), (6), and (9) of Table 2.6, we show the second stage results of a 2SLS approach in which we use instruments for import penetration of low-wage countries, routine employment share, and transportation costs. These instruments are the import penetration by low-wage countries in other high-wage countries; spatial instruments for routine employment share based on the routine employment share in areas where industry i does not coagglomerate with industry j because of, respectively, labour market pooling, input-output linkages, or knowledge spillovers; and the natural logarithm of the value of a ton. Table 2B16 reports the corresponding regression results.

The coefficient on import penetration (in other high-wage countries) in column (1) shows that a standard deviation increase in the instrument is associated with 0.780 standard deviations increase in import penetration. The impact of the spatial instrument for routine employment share based on areas where industry i and industry j do not coagglomerate to make use of a common labour pool in column (2) the the effect is slightly higher: the routine employment share leads to 0.823 standard deviations increase in routine task intensity. This confirm that our instruments are very strong,

TABLE 2B16 – FIRST STAGE RESULTS (STEP 3)

Dependent variable:	*Import penetration*	*Routine task intensity*	*Import penetration*	*Routine task intensity*	*Transportation costs*	*Import penetration*	*Routine task intensity*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Import penetration	0.780***	0.202***	0.789***	0.214***	0.147***	0.785***	0.199***
(in other high-wage countries)	(0.019)	(0.018)	(0.019)	(0.018)	(0.030)	(0.019)	(0.018)
Routine employment share	0.025	0.823***					
(spatial instrument LMP-based)	(0.016)	(0.016)					
Routine employment share			0.013	0.825***	−0.116***		
(spatial instrument input-output-based)			(0.017)	(0.016)	(0.026)		
Routine employment share						0.014	0.823***
(spatial instrument tech. rel.-based)						(0.016)	(0.016)
Value of a ton *(log)*			0.007	−0.016	−0.455***		
			(0.015)	(0.015)	(0.024)		
Value of a ton in 1970 *(log)*			−0.025*	−0.033**	0.012		
			(0.015)	(0.014)	(0.022)		
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,120	1,120	1,120	1,120	1,120	1,120	1,120
R^2	0.774	0.787	0.775	0.799	0.469	0.774	0.788

Notes: independent variables are standardized to have a mean value of 0 and a standard deviation of 1. A dummy variable is added to indicate missing data in the CFS on the instrumental variable value of a ton for industries belonging to SIC27 but its coefficient is not reported in the table; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

as already suggested by the large Kleibergen-Paap F-statistic in column (3) of Table 2.6.

Columns (3), (4) and (5) in Table 2B16 repeat this analysis but consider as an additional instrument for transportation costs the value of a ton. We *control* for the value of a ton in 1970 to mitigate the issue that more advanced industries have a higher value of a ton. Column (5) shows that the coefficient of value of a ton is negative, as expected, because a larger value of a ton means that the cost of transporting a ton becomes less important compared to all other costs of producing and selling the product plus added value. Therefore, a higher value of a ton naturally leads to a lower percentage of expenditure on transportation costs. The coefficient implies that transportation costs increase by $(\log 2 - \log 1) \times -0.455 = 0.315$ standard deviations if the value of a ton doubles. This instrument is less strong than those for import penetration and routine employment share. Therefore the corresponding Kleibergen-Paap F-statistic is somewhat lower but is still well above 10.

Going back to Table 2B16, in columns (6) and (7) we show the regression results of the first-stage for column (9) in Table 2.6, which consider technological relatedness. Here only two instruments are used for import penetration and routine employment. The former is instrumented by the import penetration of low-wage countries in other high-wage countries, as before, while the latter is based on the routine employment share of MSAs where industry i does not coagglomerate to profit of knowledge spillovers. Like in columns (1) and (2) both of these instruments are very strong resulting in a large Kleibergen-Paap F-statistic in column (9) in Table 2.6.

Chapter 3

Complex activities concentrate in large cities

Abstract – Knowledge-intensive activities are known to concentrate more strongly in space than other activities. Most evidence on this topic is based on counts of innovative inputs or outputs but these fail to appreciate the qualitative aspect of knowledge. In this chapter, knowledge complexity is proposed to explain spatial concentration based on ideas from innovation studies and complexity theory and on the division of labour. We show that empirical proxies of complexity are universally strongly related to spatial concentration across industries, occupations, technologies and scientific fields. Furthermore, the results on technologies show that since 1850 the association between complexity and spatial concentration has increased over time, in particular during industrial revolutions.

This chapter is a personal extended version of Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., and Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, 4(3):248–254, which is the only work that should be considered for research and citing purposes.

3.1 Introduction

Smith (1776) and Marshall (1890) already noted that the productivity of activities is higher in larger cities. This positive association between productivity and city size drastically increased since the 1980s, in particular, due to the rising concentration of knowledge-intensive activities (Berry and Glaeser, 2005; Glaeser, 2011; Moretti, 2012; Austin et al., 2018). As knowledge production is key in long-term economic growth (Schumpeter, 1942; Solow, 1956; Nelson and Winter, 1982; Romer, 1986) understanding why knowledge-intensive activities concentrate in space is key in understanding the growth in spatial inequality.

Currently, most evidence on the relation between activities and city size is based on the quantity of certain activities taking place, like the number of patents filed, both in urban economics/economic geography, see Audretsch and Feldman (1996) and Carlino and Kerr (2015) and in the urban scaling, see Bettencourt et al. (2007). This last line of literature shows that complex knowledge-intensive activities scale more strongly than other activities, *i.e.* on average a city of twice the population of another city has more than twice the amount of a certain knowledge-intensive activity, such as the number of patents. However, count measures as used in both lines of literature fall short in providing a theoretical foundation for the need for geographical proximity and appreciating the difference in innovativeness within the activity that is counted, both cross-sectionally and over time (Schumpeter, 1942; Kleinknecht, 1981; Dosi, 1984). For example, a patent on a type of semiconductor is likely more complex than one on a type of saw and the first invention of a semiconductor is likely more complex than the versions patented decades later. These qualitative aspects of activities are fundamental to theoretically understand and empirically measure their connection to economies of agglomeration (Pavitt, 1984; Balland and Rigby, 2017).

This chapter takes the mechanism of the division of labour as often attributed to Smith (1776) as theoretical starting point and builds on insights on the qualitative aspects of production activities from innovation studies and complexity theory to develop measures to establish to what extent activities are complex and why this requires geographical proximity. Then the extent to which complex activities concentrate in large cities and how this has changed over time is evaluated by measuring the scaling coefficient of these activities. Hereby the focus is on jobs, industries, invention, and scientific research.

The division of labour has been portrayed by Smith (1776) as the engine of prosperity. By allowing workers to specialise and exchange the fruits of their tasks higher levels of productivity can be achieved compared to self-sufficiency. With the development of communication and transportation technologies, the ability to exchange the fruits of tasks has increased and individuals can more narrowly specialise. As a result, the division of labour has become finer in societies (Leamer and Storper, 2001; Hausmann et al., 2014). Building on complexity theory, Hausmann et al. (2014) suggest that the economic complexity of nations is given by how intricate the networks are within the division of labour of countries. However, the measure they develop is not suitable to evaluate the link between the complexity of activities and urban scaling as part of their complexity measure is based on how geographically rare, *i.e.* how spatially concentrated, these are. Therefore, this measure is unsuitable to explain why complex activities concentrate in space and why this is increasingly the case.

Two components of the division of labour are important in understanding why activities concentrate: when they require proximity and when they require to be split up over more than a person. With regards to the first component, communication and transportation technologies have allowed for some activities to take place at a distance in recent decades, however, it increased the importance of geographical proximity overall. In particular, knowledge-intensive jobs have become more interactive and more concentrated in cities see Autor et al. (2003); Deming (2017) and Michaels et al. (2019). This is because cities offer better possibilities to easily meet a large variety of individuals face-to-face, which is known to be essential to build trust and communicate unfamiliar complex information of a tacit nature (Storper and Venables, 2004; Glaeser, 2011). Breschi and Lissoni (2001) discuss that this is particularly the case when new unfamiliar knowledge is involved. With regards to the second component, Jones (2009) demonstrates that with the progress of knowledge, new inventors compensate for the educational burden by specializing in a smaller set of technologies and then work in larger teams to recombine knowledge to be able to innovate. Hence, complexity influences the spatial concentration of activities through both components: on the one hand, it increases the number of members required in a team and, on the other hand, it increases the need to communicate unfamiliar tacit knowledge. As the relevance of matching with the right potential team members and communicating complex knowledge increases the importance of face-to-face contact, and therefore geographical proximity, increases, as already suggested by Vernon (1960); Duranton and Puga (2004a)

Both these elements come together in the model of Fleming and Sorenson (2001) based on complexity theory. They demonstrate that matching and recombining bits of knowledge in patent increases in complexity along two dimensions (1) the number of knowledge components, N, and (2) how related these components are to each other, K. van der Wouden (2020) shows that indeed more complex innovations are also associated with larger teams as these bits of knowledge are likely spread over multiple persons, in line with Jones (2009).

All in all, the discussion of these sources brings up several non-geographical measures for complexity that can be applied to different categories of jobs, industries, innovation and scientific research:

- **The number of members in a team** is indicative of the division of labour following Jones (2009) the complexity of the activity it is engaged in. This measure is available per Metropolitan Statistical Area in the U.S.A. for patents and research articles.

- **The average year of introduction of the subclasses on a patent** is indicative of the novelty of knowledge, which following Breschi and Lissoni (2001), is likely to require face-to-face contact for the communication of unfamiliar knowledge and therefore following Jones (2009) requires larger teams.

- **The NK measure** by Fleming and Sorenson (2001) itself, which can be applied to patent data.

- **Years of education** The years of education can be used to measure the educational burden as proposed by Jones (2009), which likely indicates their need to match and recombine knowledge in large networks. This is in line with the growth of interactive tasks by highly-skilled workers, particularly in cities found by Autor et al. (2003); Deming (2017) and Michaels et al. (2019).

Unlike the other data sources, the patent data goes back to 1850. This allows us to evaluate the urban scaling of complex activities over time. Notably during two industrial revolutions, *i.e.* the application of electricity around 1870 and that of the semiconductor around 1970. This last revolution is particularly interesting as Leamer and Storper (2001) suggest that the development of communication and transportation technologies around that time have allowed for the spatial dispersion of some non-complex activities but to an increase in the concentration of complex activities. The possibility to measure the complexity of activities allows testing this claim.

In the results, first replications of previous findings of the literature on scaling, see Bettencourt et al. (2007), and urban concentration and innovation, see Carlino and Kerr (2015). It is shown that patents, scientific articles, jobs and GDP scale superlinearly with city size. In a second step, we find that there is a strong correlation between the complexity of activities and urban scaling across invention, scientific research, jobs, and industries. For invention, this holds for all three measures. In a third step, we extend the analysis to the urban scaling of complex and less complex patents over time. The results show that over time the urban scaling of complex activities has increased. In particular during the two industrial revolutions in our sample. On the other hand, less complex activities scale more strongly over time but disperse after the 1970s. Plausibly following the possibility to routinise the coordination of tasks with the development of communication technologies following the invention of the semiconductor, in accordance with Leamer and Storper (2001).

This chapter proceeds as follows: in section 3.2, theory on the link between knowledge intense activities and urban concentration is discussed. Then the division of labour and how this can be used to measure the qualitative aspect of knowledge, *i.e.* complexity, following insights from innovation studies and complexity theory. In section 3.3, data and methodology on the activities and the measurement of scaling is introduced. Section 3.4 presents the results and Section 3.5 concludes.

3.2 Theory

Two parallel lines of literature, urban economics/economic geography and urban scaling, show similar findings that more knowledge-intensive activities concentrate in larger cities. Urban scaling is based on scaling laws, which in combination with the associated vocabulary are notably used by academics with a background in evolutionary biology and physics, who apply these mathematical laws to social phenomena, see *e.g.* Bettencourt et al. (2007).

Scaling laws, also known as power laws, are used to describe the functional relationship between two quantities according to the $y \approx x^{\beta}$. In urban scaling, y is often the measure of the activity, *e.g.* the number of patents filed in a city, and x the population of a city. β then gives scaling coefficient. When $\beta = 1$ an increase in city size is associated with a proportional increase in the size of the activity. When $\beta > 1$ the activity is said to scale superlinearly, which means that the size of the activity increases more strongly with city size. For example, Bettencourt et al. (2007) find $\beta \approx 1.2$ for patent production in cities in the U.S.A.. When $\beta < 1$ the activity is said to scale sublinearly. For example, Bettencourt et al. (2007) find in the same study that $\beta \approx 0.8$ for the amount of infrastructure.

In different wording, similar relationships have been denoted in urban economics/economic geography. Here the functional relationship between two quantities is known as elasticities and a large line of literature starting with Ciccone and Hall (1996) and reviewed by Rosenthal and Strange (2004) has empirically demonstrated that productivity per person increases when city size and or density increases, as already suggested by Smith (1776) and Marshall (1890). The use of univariate power law formulas in different vocabulary for these phenomena is also not uncommon in urban economics. For example, Combes et al. (2010) use the same formula and produce similar graphs like Bettencourt et al. (2007) to denote the relation between productivity per worker and density in French urban areas.

More specifically, on the relation between innovation and agglomeration, a line of research started with Audretsch and Feldman (1996) and recently reviewed Carlino and Kerr (2015) exists that show that the number of patents, R&D expenditures, and venture capital are more geographically concentrated than other production activities or population. Generally, an image is drawn that more complex knowledge-intensive activities concentrate more strongly.

However, Carlino and Kerr (2015, p.6) review that three types of indicators are used: "*1) by the inputs used in the innovation process, such as R&D expenditures or venture capital (VC) investment;(2) by intermediate outputs of the innovation effort, such as the number of patents; or (3) by some final measure of innovative work, such the count of new product announcements.*". In these count measures, also used in the urban scaling literature, each patent or euro spent in R&D expenditure is valued as equally innovative even though a large literature asserts that there is much heterogeneity in innovativeness between these, both cross-sectionally as over time (Schumpeter, 1942; Kleinknecht, 1981; Dosi, 1984).[1]

Other authors attempted to measure the qualitative aspect of knowledge by developing dichotomies of non-innovative and innovative activities but do not come to clear continuous measures. For example, Rosenthal and Strange (2003) compare the geographical attenuation of spillovers in software industries compared to industries in fabricated metals and machinery. Similar dichotomous self-made industrial classifications are used in Neffke et al. (2011b); Caragliu et al. (2016) and Faggio et al. (2017). Others divide cities in similar dichotomies like the idea, respectively, high value-added goods-producing cities versus good, respectively, low value-added goods-producing cities in the models of McCann (2008) and Glaeser and Ponzetto (2007). As Pavitt (1984) claims measuring the complexity of knowledge remains difficult even though this is needed to both theoretically understand and empirically measure why certain activities concentrate more strongly in large cities. Here complexity is proposed to tackle both the theoretical and empirical dimensions.

[1] Also note that a measure like the number of patents is also subject to changes in patenting laws instead of innovativeness. As an example, Carlino and Kerr (2015, p.9) state that "*founded in 1975, Microsoft had just five patents by 1990 and over a billion dollars in revenue; by 2009, the company held 10,000 patents.*".

3.2.1 Complexity, proximity and the division of labour

A first starting point is the first mechanism of agglomeration benefits ever described, namely, the division of labour, for which Smith (1776) is often accredited.[2] When workers divide the labour tasks in production this creates efficiency gains for three reasons: first, the workers improve their skills at performing that task, also known as learning by doing; second, a worker saves time from not having to switch between tasks; and third, it allows for labour-saving innovations such as the development of specific tools or the mechanisation of part of the tasks.

Using a similar line of reasoning, Hausmann et al. (2014) posit that modern societies do not distinguish themselves from traditional hunter-gatherer societies because of more productive knowledge per society member but because of a larger diversity of productive knowledge across its members, as a finer division of labour allows members to specialise and access the bits of knowledge, expressed in products and services, they don't possess through networks.

The size of the network, or the fineness of the division of labour, is dependent on transport efficiency, as already noted by Smith (1776). A worker that specialises in a certain task needs to have sufficient demand for that task so the person can specialise and purchase the other goods and services she wants to consume but doesn't produce. The extent of the market is then determined by transportation costs as one requires to exchange the fruits of their tasks produced for those one consumes. The further customers/suppliers are located the higher the price for selling/buying the products/services. Cities are therefore the natural place where a larger division of labour takes place. As an example, Smith (1776) contrasts the remote dwellers of his time in the Scottish Highlands that are each their own farmer, brewer, butcher and baker to city dwellers who through a larger division of labour make a profession out of a single of these tasks leading to higher productivity in cities.

However, developments in communication and transportation technologies also allow individuals living outside cities to specialise and receive goods and services from all over the world, like the current inhabitants of the Scottish Highlands. Geographical proximity is hence not required in all cases, but the death of distance as Cairncross (1997) predicted has also not materialised. In fact, the contrary has happened as cities have become more important than ever Gaspar and Glaeser (1998); Leamer and Storper (2001); Glaeser (2011); Moretti (2012).

[2] Note that the division of labour and its relation to the city has already been described in Ancient Greece by Xenophon and to a lesser extent Plato, see respectively Finley (1973) and Silvermintz (2010). Furthermore, Smith (1776) is considered to have re-used a lot of material earlier available in French, see Peaucelle (2006).

Leamer and Storper (2001, p.643) explain the two effects that developments in communication and transportation technologies have, namely: "*(1) the constant transformation of complex and unfamiliar coordination tasks into routine activities that can be successfully accomplished at remote but cheaper locations (e.g., commodification), and thus an ongoing tendency toward deagglomeration or dispersion of production; and (2) bursts of innovations that create new activities requiring high levels of complex and unfamiliar coordination, which, in turn, generate bursts of agglomeration*".

Leamer and Storper (2001) argue that with a finer division of labour comes a greater need for coordination. A part of this coordination of tasks can be done through the exchange of routinised codified knowledge and can therefore be performed at a distance but a large part may require the communication of much tacit knowledge and therefore geographical proximity. That this latter part outweighs the previous part is demonstrated by the rise of knowledge-intensive activities in cities, and the growth in time spent on interactive tasks by highly-skilled workers, particularly in cities, found by Autor et al. (2003); Deming (2017) and Michaels et al. (2019). As such, one sees both a finer division of labour and an increasing advantage of cities to divide labour. Insights on these two aspects are combined here to develop measures of the quality of knowledge, expressed in complexity, to explain geographical concentration.

To understand why cities and therefore geographical proximity are increasingly in demand one needs to understand *the most fundamental aspect of proximity: face-to-face contact*" (Storper and Venables, 2004, p.351). Storper and Venables (2004) make the distinction between codifiable information and uncodifiable information, also known as tacit knowledge. Codifiable information can easily and cheaply transmitted if sender and receiver understand that system, *e.g.* language and mathematical notations, and have the means of communicating it, *e.g.* letters, books, e-mail. In contrast, uncodifiable information can not fully be expressed in a symbol system, as often different dimensions of the problem at hand are only understood in relation to each other (Storper and Venables, 2004). Breschi and Lissoni (2001) and Boschma (2005) add that even when tacit knowledge can be codified and shared openly not everyone may understand the language of a much closer and restricted community, as understanding the jargon and background information of certain professionals, requires long study and common experiences. These persons lack cognitive proximity in the words of Boschma (2005). Boschma (2005) details three other relevant dimensions of proximity for the sharing of information: organisational proximity, which relates to the coordination and hierarchy of agents exchanging information, social proximity, which relates to friendship, kinship and trust; and institutional proximity, which relates to sharing the same norms and values of conduct. The distances in these proximities are summarised in what (Glaeser, 2011, p.24) calls the "*complex communication curse*", which can be resolved via face-to-face interaction as "*long hours spent one-on-one enable listeners to make sure that they get it right.*".

Storper and Venables (2004) summarise the advantages of face-to-face contact stating that it is an efficient communication technology, as it allows for instant interruption, feedback, and repair and builds on visual and body language cues (Storper and Venables, 2004); it allows for screening and socializing, which depend on identifying and assimilating tacit knowledge among group members; it generates psychological motivation; and it helps build trust as it aids in the detection of lying and meeting face-to-face requires a larger sacrifice of time to come to the same location compared to for example sending an email, which signals commitment.

Geographical proximity therefore may seem paramount when reading these works but this essence needs to be relaxed as detailed by Breschi and Lissoni (2001) and Boschma (2005). Members that have developed the knowledge to be part of an epistemic community may only require occasional geographical proximity, *i.e.* occasional meetings, but can mainly communicate via codified information and video calls when other dimensions of proximity are sufficiently developed.

Despite these nuances, evidence shows that a useful indication of knowledge that likely requires face-to-face contact is its novelty. Breschi and Lissoni (2001) state that co-location is generally helpful in the early stages of a project when the organisation and common language is still under development. Thereby, echoing thoughts of Vernon (1960) who stated that the development of new products, *i.e.* the young phase of the product life cycle, require speed and contact but when the product and its production methods are standardised, *i.e.* the mature phase of the product life cycle, communication can do with phone calls and referencing catalogue numbers. Similarly, Arzaghi and Henderson (2008) conclude after inquiries among managers of advertisement agencies that networking in person is essential because of the short-term projects consisting of different compositions of collaborators and goals in the business.

The novelty of information for a person can arguably be measured by cognitive proximity or industrial distance, in the words of Rosenthal and Strange (2004). When agents have different knowledge bases a larger share of unfamiliar knowledge needs to be communicated. This is plausibly empirically visible in the higher densities and shorter distances between agents in diverse places, where cognitive proximity between agents is likely to be large, compared to specialised places, see (Vernon, 1960; Chinitz, 1961; Jacobs, 1969; Caragliu et al., 2016).

Jacobs (1969) already argued that these diverse cities have more options for matching and recombining ideas of different sectors, which leads to "adding new work", *i.e.* technological breakthroughs that lead to new products and services. These activities can be considered as complex as innovation studies also suggest that finding and recombining ideas is key in the development of new knowledge (Kauffman, 1993; Fleming and Sorenson, 2001; van der Wouden and Rigby, 2019). To empirically measure the complexity of a new piece of knowledge, in this case, a patent, Fleming and Sorenson (2001) build on complexity theory, which is the analysis of large networks with many interacting components (Frenken, 2006). They combine both the idea of recombination as well as of cognitive proximity. According to them, the complexity of a patent increases along two dimensions, N and K, in which N is the number of subclasses combined and K is a measure of the ease of recombination in their words, similar to cognitive distance. The more classes are combined, and the less these have been combined in inventions on previous patents, the more complex the patent is. They show that the higher a patent scores on this NK-measure the more impactful it is, as measured using patent citations.

This likely leads to an increase in the demand for geographical proximity as other evidence shows that these knowledge components are increasingly spread over a larger number of humans, who need to match and interact to collaboratively produce new knowledge (Wuchty et al., 2007; Jones, 2009). As an explanation, Jones (2009) argues that the educational burden to reach the frontier of knowledge has grown so large with the advances of knowledge over time that not a single person can hope to have all the knowledge required to make great advancements in multiple fields, like a *homo universalis*, such as Leonardo da Vinci, could. He describes how inventors decrease the educational burden by specializing in a narrow field and collaborating with others to recombine knowledge into new advancements.

Previous studies have also built on complexity theory to not evaluate the complexity of an activity but of a geographical area. Hausmann et al. (2014, p.18) define complexity as "*a measure of how intricate this network of interactions is and hence of how much productive knowledge a society mobilises.*" The complexity scores they calculate correlate strongly with the level of economic development of countries. In geography, Balland and Rigby (2017) apply this method to cities and find a similar pattern.

However, this measure of complexity is not useful to evaluate the relation between geographical concentration and complex knowledge, as it is partly defined by how geographically rare an activity is. The method of Hidalgo and Hausmann (2009) and Hausmann et al. (2014) consists of estimating the complexity of a country's economy based on: (1) the diversity of products produced in a country and (2) the ubiquity of these products, *i.e.* how many other countries are specialised in this product. They apply a so-called method of reflection, which attributes higher complexity values to countries with products that are not ubiquitous but mostly produced by more diverse countries. As the ubiquity measure is based on geographical concentration, the literature so far has not produced a geographically unbiased measure for complexity but has produced valuable insights to develop one.

3.2.2 Complexity measures

Based on the previous discussion the following complexity measures are withheld to evaluate the association between urban scaling and the complexity of invention, scientific research, jobs, and industries.

- **The number of members in a team** is indicative of the division of labour following Jones (2009) the complexity of the activity it is engaged in. This measure is available per Metropolitan Statistical Area in the U.S.A. for patents and research articles.

- **The average year of introduction of the subclasses on a patent** is indicative of the novelty of knowledge, which following Breschi and Lissoni (2001), is likely to require face-to-face contact for the communication of unfamiliar knowledge and following Jones (2009) requires larger teams.

- **The NK measure** by Fleming and Sorenson (2001) demonstrates the innovativeness, a proxy for complexity, of patents.

- **Years of education** The years of education can be used to measure the educational burden as proposed by Jones (2009), which likely indicates their need to match and recombine knowledge in large networks. This is in line with the growth of interactive tasks by highly-skilled workers, particularly in cities found by Autor et al. (2003); Deming (2017) and Michaels et al. (2019).

3.2.3 Complexity and urban scaling over time

The more interesting question in light of the increase in spatial inequality, see for example Moretti (2012), is how the relation between urban scaling and knowledge complexity has changed over time. Therefore, the complexity of patents since 1850 is calculated. As a result, a time period of over 150 years is considered including two industrial revolutions, *i.e.* the application of electricity in 1870 and that of the semiconductor in 1970.

This last revolution is particularly interesting as Leamer and Storper (2001) suggest that the development of communication and transportation technologies around that time have allowed for the dispersion of some non-complex activities but to an increase in the concentration of complex activities. In contrast, the model of McCann (2008) based on the New Economic Geography (NEG) suggests that all activities concentrate in space because of home market size, agglomeration economies and the decrease in transportation costs. The possibility to measure the complexity of activities over time allows us to test these claims.

3.3 Data and Methodology

3.3.1 Patent data

Patents are registered by the United States Patent and Trademark Office (USPTO). Information on patents is available from NBER patent data (Hall et al., 2001). For the cross-sectional analysis on the relation between complexity and spatial concentration data is used from 2000 to 2010 per 2-digit NBER sub-categories, as defined by Hall et al. (2001).

Figure 3.1 shows the 2-digit NBER sub-categories used in the Chapter and their respective number of patents granted between 2000 and 2010 for the 353 cities for which data is available for all activities. In the main results the sectors Agriculture, Husbandry, Food; Agriculture, Food, Textiles; and Earth Working & Wells are left out as these are often geographically concentrated in less-populated areas for natural reasons. These results are nonetheless shown in the robustness analysis.

For the analysis of the relation between complexity and spatial concentration over time the NBER patent data set 1974-2009 by Hall et al. (2001) is combined with data from HISTPAT, which originates from the efforts of Petralia et al. (2016) to obtain geographical locations for all patents over the period $1836 - 1974$ from Google scans of historical U.S. patents. Figure 3.2 shows the 2-digit SOC occupations used in the Chapter and their respective number of employees in 2015 for the 353 cities for which data is available for all activities.

3.3.2 Industry data

Here we use 2015 GDP data from the Bureau of Economic Analysis to quantify the economic output of MSAs in 18 industries as defined by the North American Industry Classification System (two-digit NAICS). In the main analysis, we consider only industries for which there are more than 200 cities with any recorded activity and we remove categories that are based on natural advantages: ‘agriculture, forestry, fishing, and hunting’ and ‘utilities’ but do show these results in the appendix. Figure 3.3 shows the 2-digit NAICS industries used in the Chapter and their respective GDP in 2015 for the 353 cities for which data is available for all activities.

FIGURE 3.1 – NUMBER OF PATENTS FOR EACH TECHNOLOGICAL SUB-CATEGORY BETWEEN 2000 AND 2010.

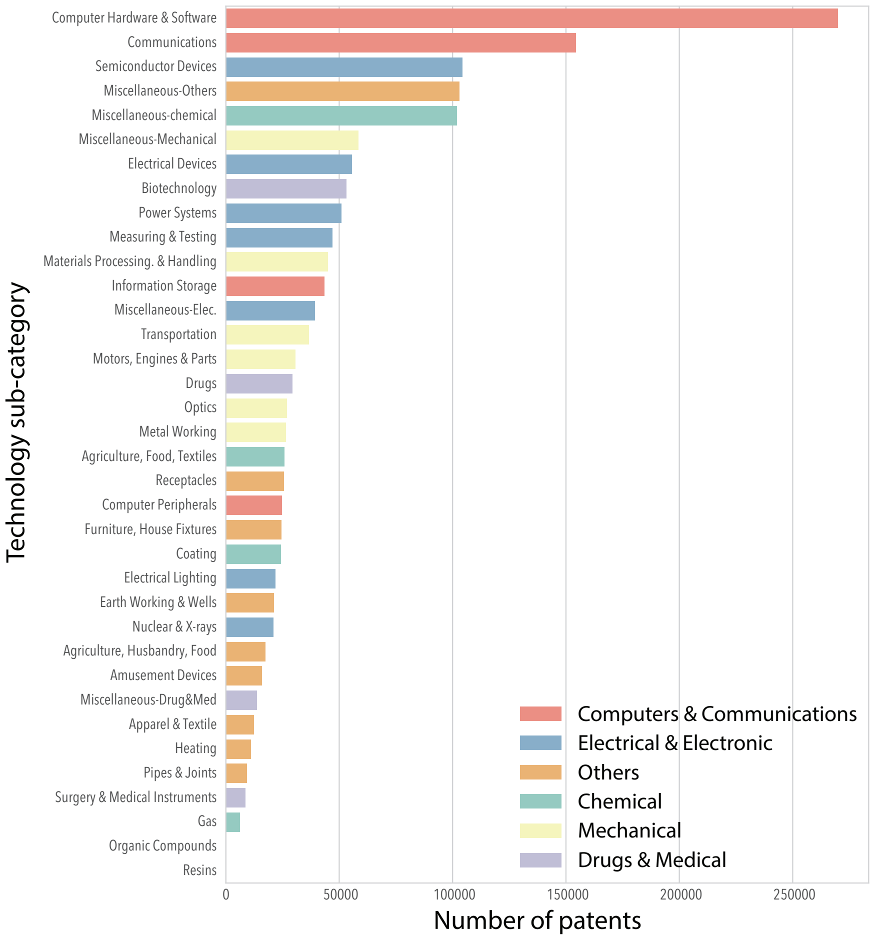


FIGURE 3.2 – NUMBER OF EMPLOYEES IN EACH OCCUPATION CATEGORY IN 2015.

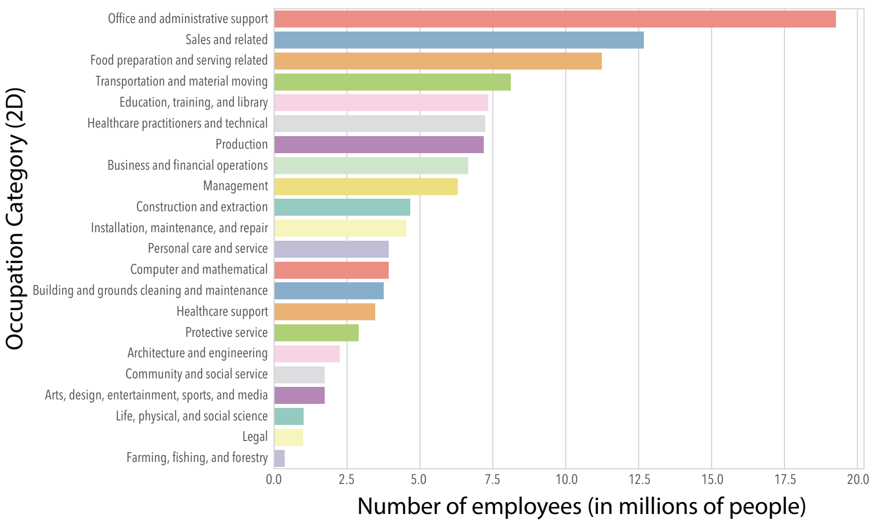
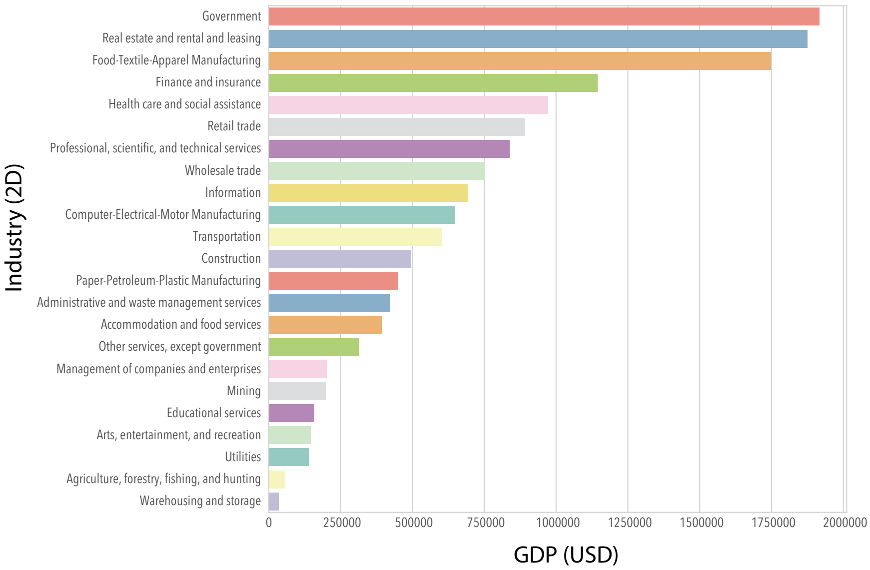


FIGURE 3.3 – GDP OF EACH INDUSTRY IN 2015.



3.3.3 Scientific research data

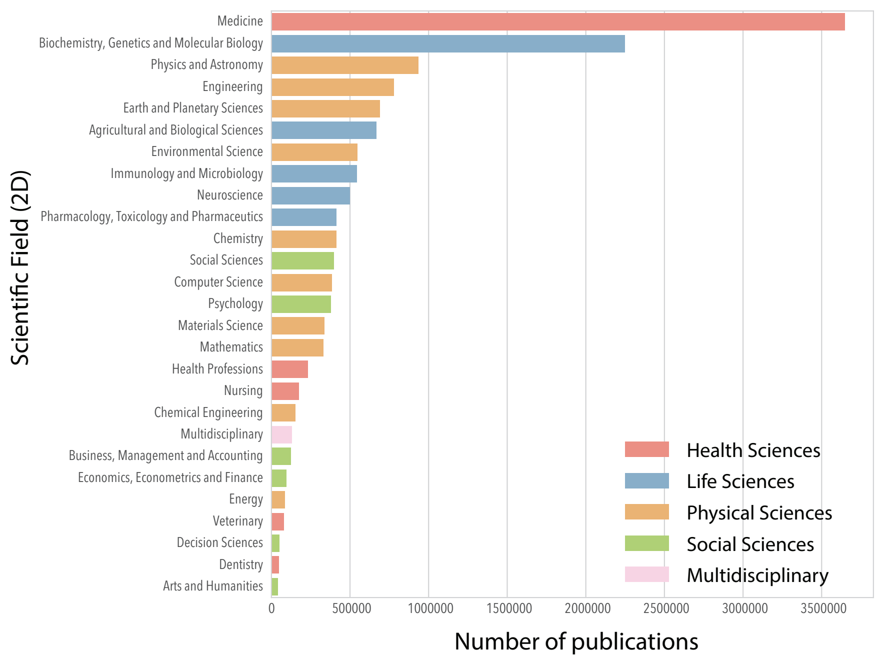
For scientific papers, we use publication data from Elsevier's Scopus database covering the time period 1996–2008. Publications are disaggregated into 23 scientific disciplines as defined by the Scopus classification (two-digit major thematic categories). These data have kindly been provided by Nomaler et al. (2014). We analyse a total of 4,4000,000 scientific publications. The data include documents that have at least one author who has (at least) one affiliation to a US scientific organisation. In the main analysis, we consider only scientific fields for which there are more than 200 cities with any recorded activity and like before we remove categories that are based on natural advantages: 'agricultural and biological sciences', 'environmental science', 'Earth and planetary sciences' and 'veterinary' but do show these results in the appendix.

Figure 3.4 shows the 23 main scientific fields used in the Chapter and their respective number of papers for the 353 cities for which data is available for all activities.

3.3.4 Employment data

Here, we use 2015 employment statistics from the Bureau of Labor Statistics disaggregated into 22 occupations according to the Standard Occupational Classification system (two-digit SOC). For occupations, we use 2015 employment statistics from the Bureau of Labor Statistics. In the main analysis, we consider only occupations for which there are more than 200 cities with any recorded activity and we remove one category that is based on natural advantages: 'farming, fishing, and forestry' but do show these results in the appendix.

FIGURE 3.4 – NUMBER OF PAPERS IN EACH SCIENTIFIC FIELD BETWEEN 1998 AND 2008.



3.3.5 *Empirics*

In this approach, use is made of scaling coefficients. As said, scaling coefficients are obtained by evaluating $y \approx x^{\beta}$. Where, y is measured by the number of patents, scientific papers, jobs, and GDP per subcategory and x is the population of a city. β then gives scaling coefficient, which is expected to be larger for more complex subcategories. The complexity of each category within activities is given by the complexity measures mentioned in Section 3.2.2.

The analysis proceeds in three steps. First, the scaling laws of employment, GDP, patents, and scientific articles are compared to the results of Bettencourt et al. (2007) to set the baseline results for the overall scaling of these activities. In the second step, the scaling coefficient of each subcategory within activities is measured and related to the complexity of that category. In the third step, the relation between complexity and urban scaling is evaluated over time since 1850 for patent activities.

3.4 Results

3.4.1 *Baseline results*

In a first step, the analyses of Bettencourt et al. (2007) are applied to our data. We compare the spatial distribution of population, shown in Figure 3.5 to that of employment, GDP, patents, and scientific articles, shown in Figures 3.6-3.13.

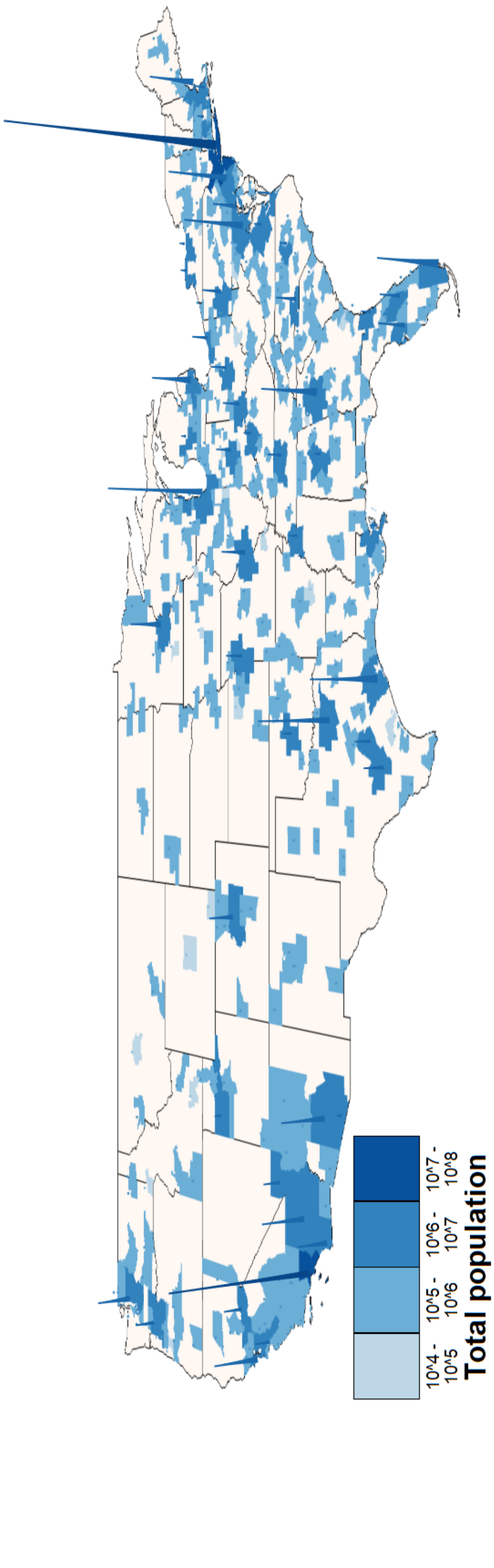


Figure 3.5 – Spatial concentration of Population

Figure 3.6 – Spatial concentration of employment

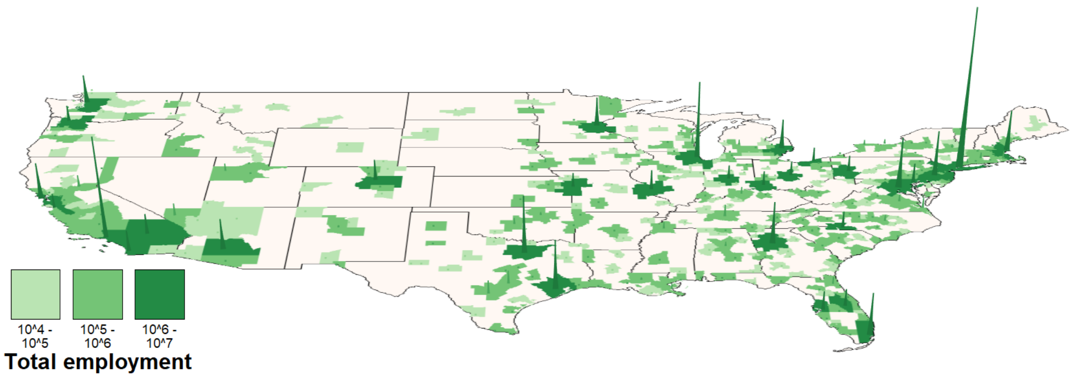


Figure 3.7 – Scaling of employment

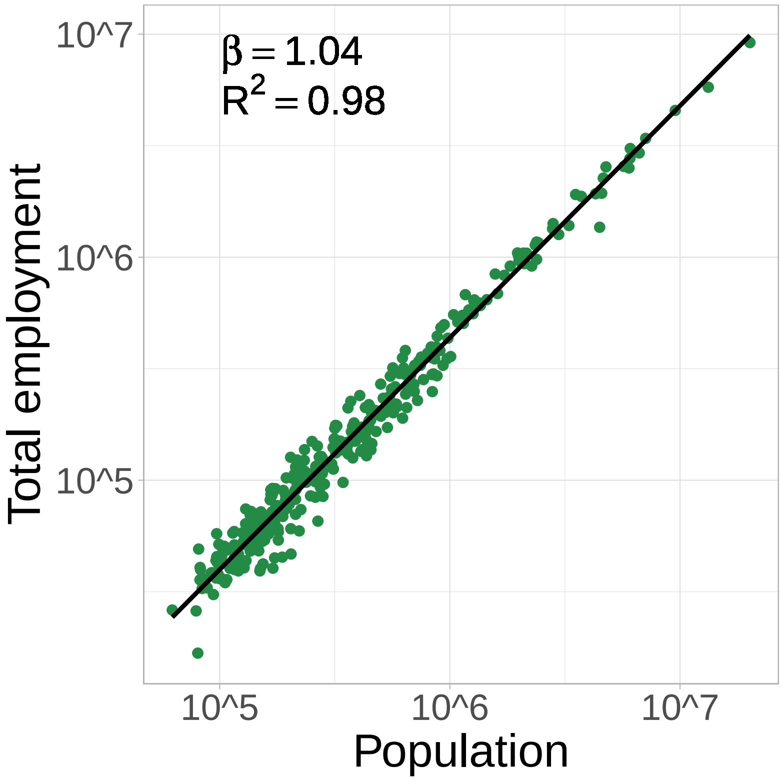


FIGURE 3.8 – SPATIAL CONCENTRATION OF GDP

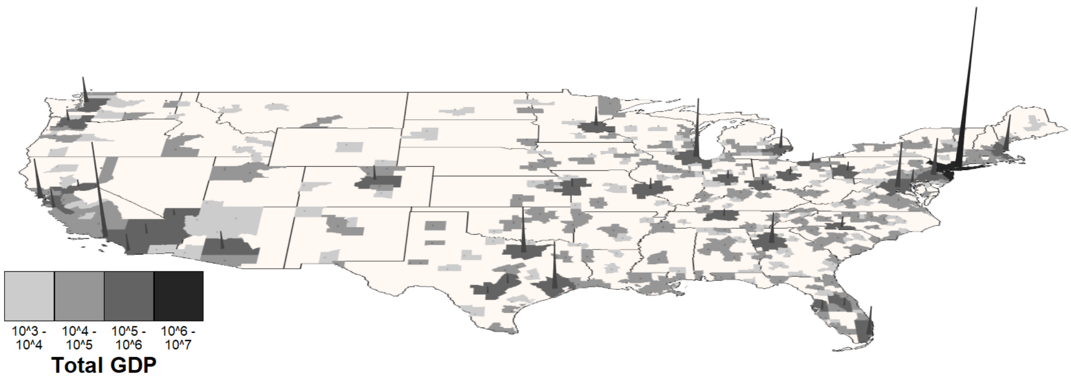


FIGURE 3.9 – SCALING OF GDP

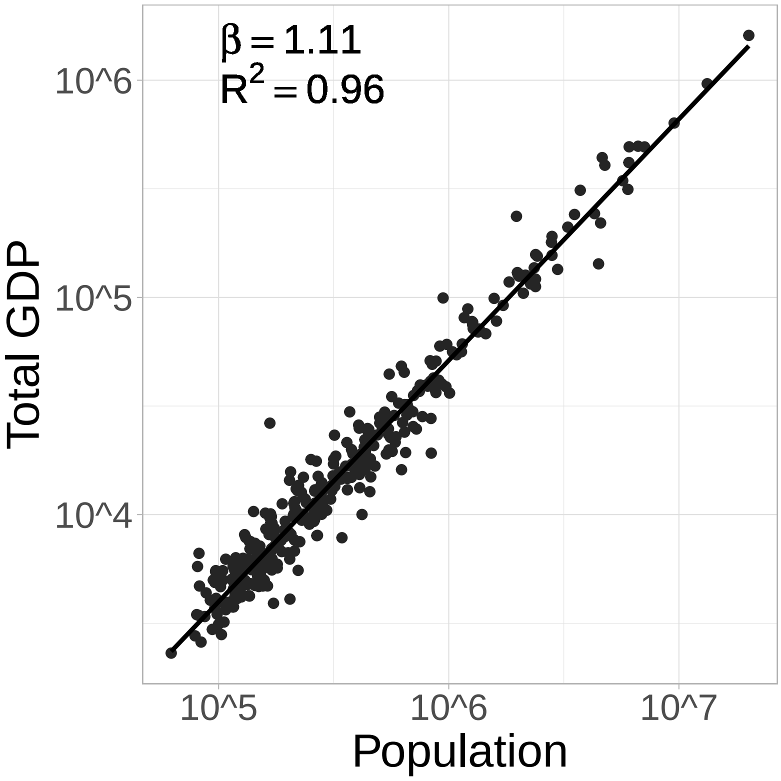


FIGURE 3.10 – SPATIAL CONCENTRATION OF PATENTS

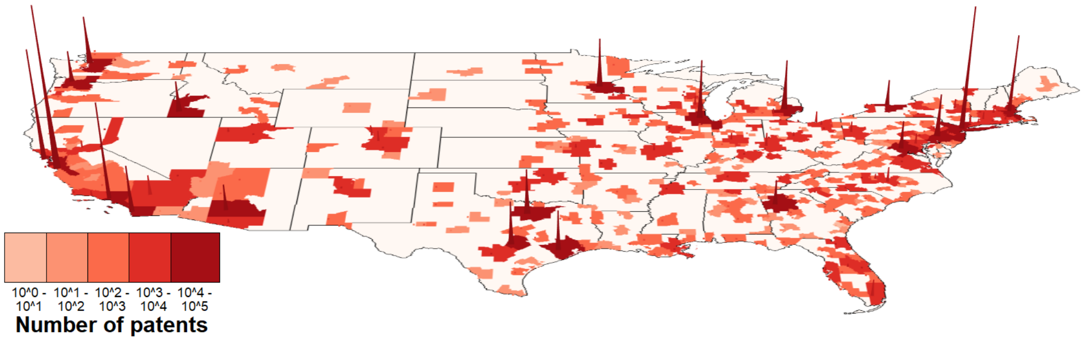


FIGURE 3.11 – SCALING OF PATENTS

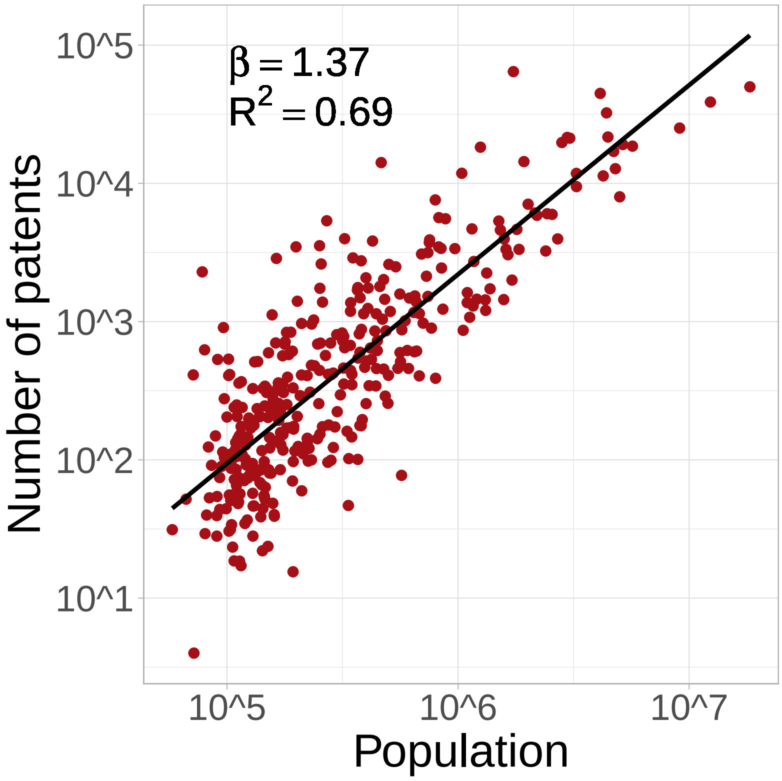


FIGURE 3.12 – SPATIAL CONCENTRATION OF PUBLICATIONS

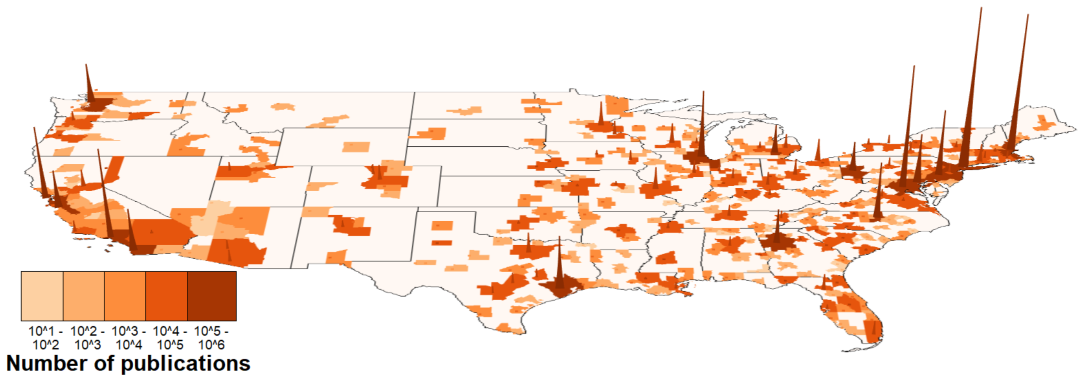


FIGURE 3.13 – SCALING OF PUBLICATIONS

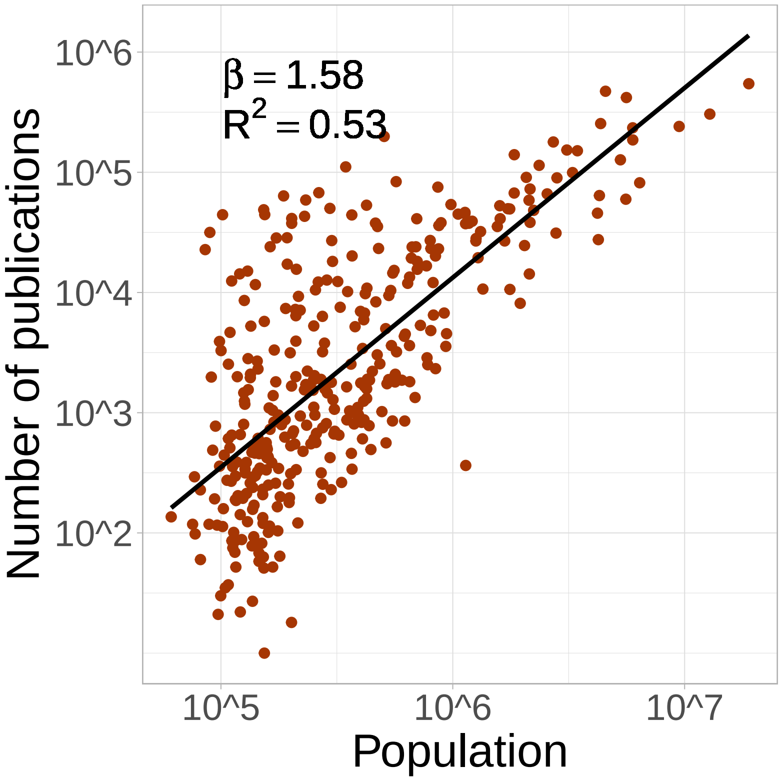


Figure 3.5 shows the spatial distribution of population. The two largest cities are clearly visible on the opposite sides of the country. The largest spike corresponds to the New York MSA, which also includes Newark and Jersey City, totalling a population of 20,118,063 in 2015. Los Angeles MSA, which also includes Long Beach and Anaheim, had 13,268,828 inhabitants in 2015. The top 10 in population size is completed by the MSAs of Chicago, Dallas, Houston, Washington, Philadelphia, Miami, Atlanta, and Boston. Note that the less populated regional divisions Mountain and West North Central, as defined by the U.S. census bureau, also have relatively few places that meet the threshold of being considered a MSA.[3]

Figures 3.6-3.13 show the spatial distribution in employment, GDP, patents, and scientific articles. Figure 3.6 shows the spatial distribution of employment, which looks rather similar to that of population. The top 10 MSAs also consists of the same 10 cities but Washington and Boston score a few positions higher while Atlanta and Miami rank lower. Superlinear scaling is not so visible in the map but Figure 3.7 shows that the scaling coefficient of employment is 1.04, which suggests that a 1% increase in the total population of a city is associated with a 1.04% increase in total employment. As an example, New York City has 0.4558 of a job per inhabitant while the smallest MSA in terms of jobs, Hinesville in Georgia, only offers 0.2089 of a job per inhabitant. Clearly demonstrating that the size of cities is associated with increasing returns making it more profitable (or necessary?) to be employed in larger cities.

Figure 3.8 shows the spatial distribution of GDP. An interesting feature that stands out is that New York is the only city in the highest category in GDP. The difference in city size and resulting GDP is more clearly visible compared to earlier maps. In the top 10, the San Francisco metro area makes an entry while Miami has dropped out. Los Angeles and Chicago are much smaller in terms of GDP compared to the population difference with New York and Chicago seems larger than Detroit in terms of GDP than their population difference. This superlinear scaling is confirmed by the scaling coefficient (β) of 1.11 in Figure 3.9. As an example, New York City has a GDP per capita of \$ 79,945.- while the smallest MSA Lewiston, Idaho has a GDP per capita of \$ 37,005. This is in line with observations by Smith (1776) and Marshall (1890) and more systematically analysed by Ciccone and Hall (1996) that productivity of individuals is higher in larger cities.

Figure 3.10 shows the spatial distribution of patent production. As patents are only used for technological innovation it only represents a subset of innovative activities

[3]The U.S. census bureau classifies the U.S.A. into 9 divisions. Division 4 West North Central consists of Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota. Division 8 Mountain consists of Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming.

in a few industrial sectors. Inventions in services or organisation structures are for example not patented. As a result, the highest peak in this map is not New York City but the San José metro area in California, commonly known as Silicon Valley and famed worldwide for technological innovation. For similar reasons, San Francisco, Boston, Seattle, and San Diego enter the top 10 while the Southeastern cities only show modest peaks. Boise, Idaho, headquarters of HP, also come up high even though only being in 80th place according to population.

Figure 3.10 shows the spatial distribution of patent production. As patents are only used for technological innovation it only represents a subset of innovative activities in a few economics sectors. Inventions in services or organisation structures are for example not patented. As a result, the highest peak in this map not New York City but the San José metro area in California, commonly known as Silicon Valley and famed world-wide for technological innovation. For similar reasons, San Francisco, Boston, Seattle, and San Diego enter the top 10 while the Southeastern cities only show modest peaks. Boise, Idaho, headquarters of HP, also come up high even though only being in 80th place according to population.

Nonetheless, the relation between city size and patent production is clear. Despite the clear differences in the hierarchy of patent production compared to population size, which are articulated in the smaller R^2 and the larger spread of observations in Figure 3.11. The scaling coefficient is 1.37, which is indicative of the suggestion that more complex activities like patent production concentrate more strongly in large cities than more general activities as gathered under employment or GDP. The scaling coefficient of 1.37 on our 2000 to 2009 data is also statistically significantly higher than the 1.2 found by Bettencourt et al. (2007) using data from 1980 to 2001.[4] This increase in scaling over time is in line with the suggestion that knowledge-intensive activities have come to concentrate more strongly in large cities (Glaeser, 2011; Moretti, 2012).

Figure 3.10 shows the spatial distribution of scientific publications. Like patents, scientific publications are a measure limited to a small subset of knowledge-intensive activities. In this case, academic research. As a result, larger discrepancies can be seen when compared to the sizes of the spikes of population in Figure 3.5 than in the case of employment and GDP. Although the New York area rates are on the top, there are notable differences in the relative spatial distribution. The second place in scientific publications is the Boston-Cambridge area, home to many renowned universities like Harvard, MIT, and Boston University. The Durham-Chapel Hill metro area in North Carolina is number eight in the top ten even though only being number

[4] The standard deviation on the scaling coefficient is 0.0489 shown in Figure 3.11.

97 in terms of population size. The area is part of the so-called research triangle of North Carolina comprising Duke University, and the University of North Carolina at Chapel Hill. Outside of the top ten, a remarkable observation is the size of the Ann Arbor metro area, home to Michigan State University, compared to the much more populated neighbouring Detroit area. Like with the production of patents these differences between the hierarchy in population and the hierarchy in publications lead to a larger variance and therefore a smaller R^2 in Figure 3.13.

Nonetheless, the relation between city size and the production of scientific articles is clear. The scaling coefficient (β) of 1.58 shows the strongest relation between city size and activity of all the four activities here.

3.4.2 Robustness of the first step

In Appendix 3A, we show that similar superlinear scaling relationships hold when scaling coefficients are calculated separately for cities with less, respectively, more than one million inhabitants.

3.4.3 Concentration of complex economic activities in large cities

The fact that arguably more knowledge-intensive activities like patent production and scientific article production have stronger scaling coefficients than total employment and total GDP is a first indication that more complex activities concentrate more strongly in space.

To establish this connection we exploit the difference within jobs, industries, patents and research articles in terms of complexity to evaluate their connection to urban scaling. For each 2-digit SOC job category, 2-digit NAICS industry, NBER technological subcategory, and 2-digit Scopus major research a scaling coefficient is calculated.

Figures 3.14 and 3.15 give the example of urban scaling of employment in, respectively, the production job occupations, and computer and mathematical job occupations. The number of workers in production occupations like assemblers and fabricators scale sublinearly with city size, as $\beta < 1$. This indicates that a larger city actually has a smaller proportion of workers in production occupations than a smaller city. On the other hand, workers in computer and mathematical occupations, like computer and information research scientists, scale superlinearly with a scaling coefficient β of 1.35, which is far larger than the β of employment in general of 1.04 found in Figure 3.7.

Figure 3.14 – Scaling of jobs in production

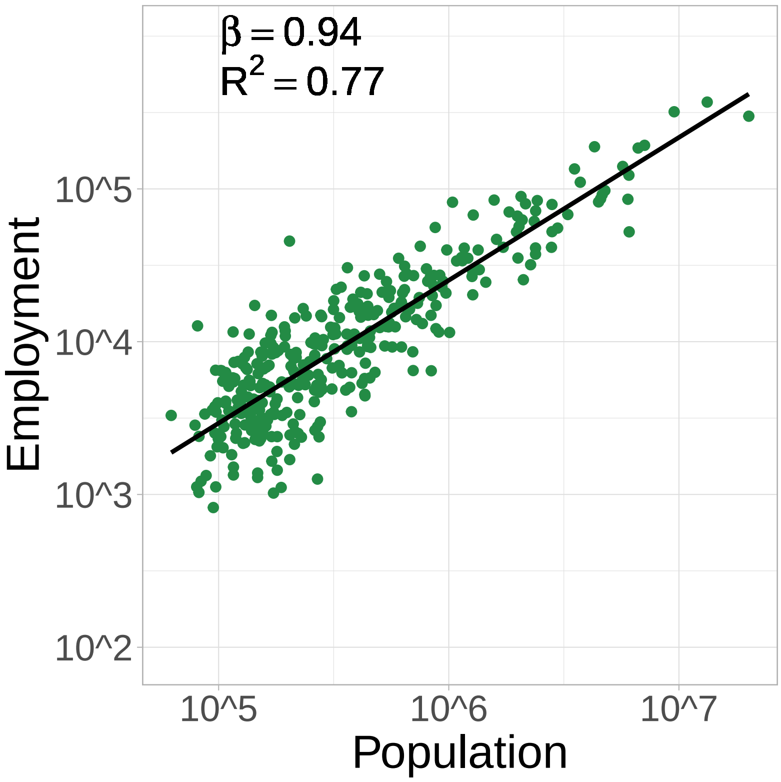


Figure 3.15 – Scaling of jobs in Computer and mathematical

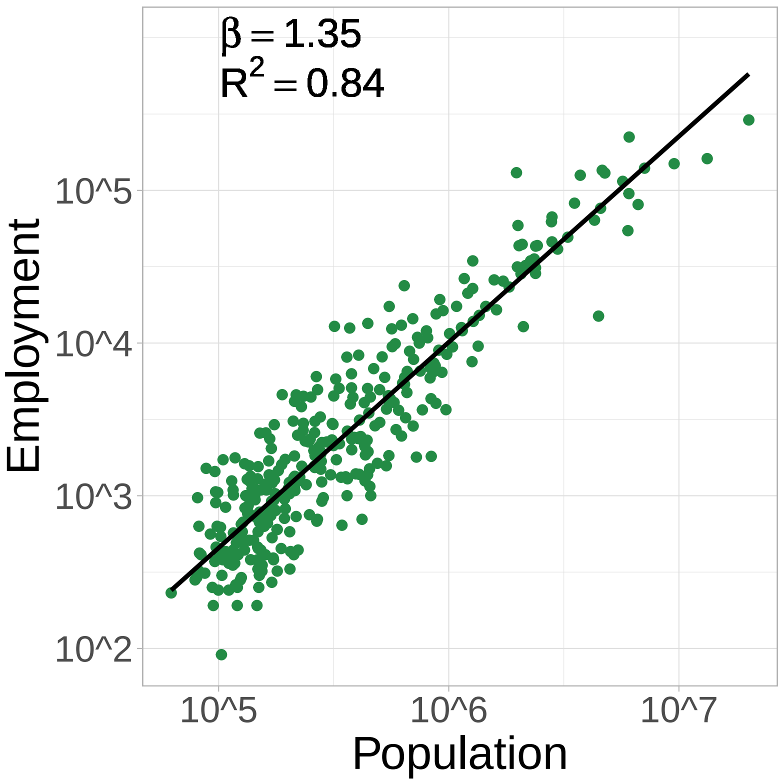
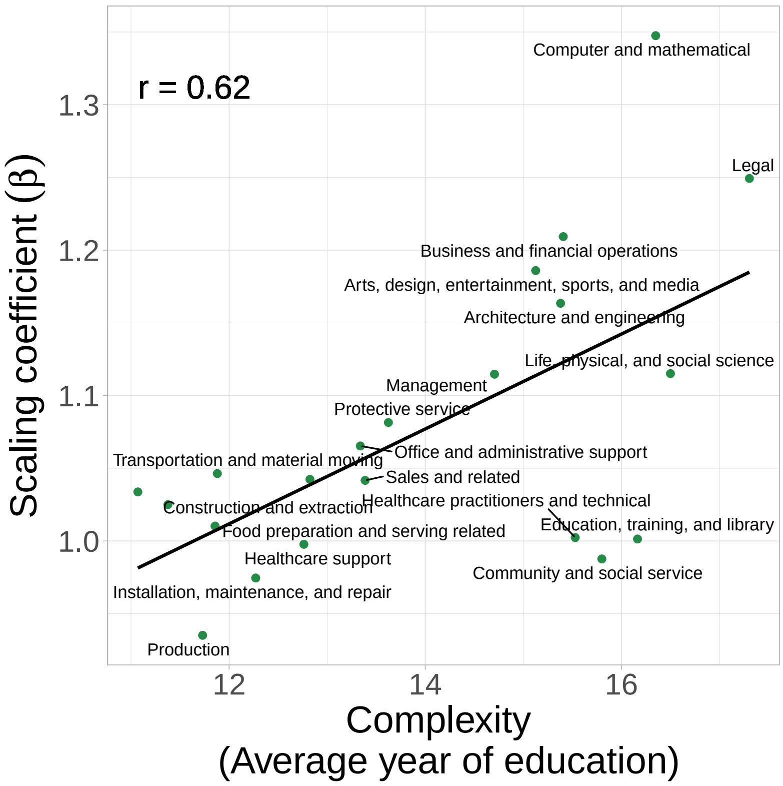


FIGURE 3.16 – SCALING AND COMPLEXITY OF JOB CATEGORIES



This result is in line with the findings of Audretsch and Feldman (1996); Rosenthal and Strange (2003), and Carlino and Kerr (2015) that general production activities are less concentrated than more information technology-intensive sectors. To explain this tendency we introduce the concept of complexity and novel ways to measure these.

The complexity indicators are average number of years of education for jobs and industries, average year of subclass introduction for patents, and average number of authors for publications. In the robustness analysis, the complexity of patents is also measured by the NK measure of Fleming and Sorenson (2001) and the average number of inventors.

Figures 3.16-3.19 show the scaling coefficient of each category within each activity on the y-axis, the measure of complexity on the x-axis, and the Pearson correlation coefficient r between these two. All figures show a strong correlation between the extent to which an activity scales and the extent to which it can be considered complex. Individual graphs of each category within each activity can be consulted in Appendix 3A.

Figure 3.16 shows a correlation between urban scaling and the average number of years of education of job categories. With the earlier examples of the occupations in

Figure 3.17 – Scaling and complexity of industries

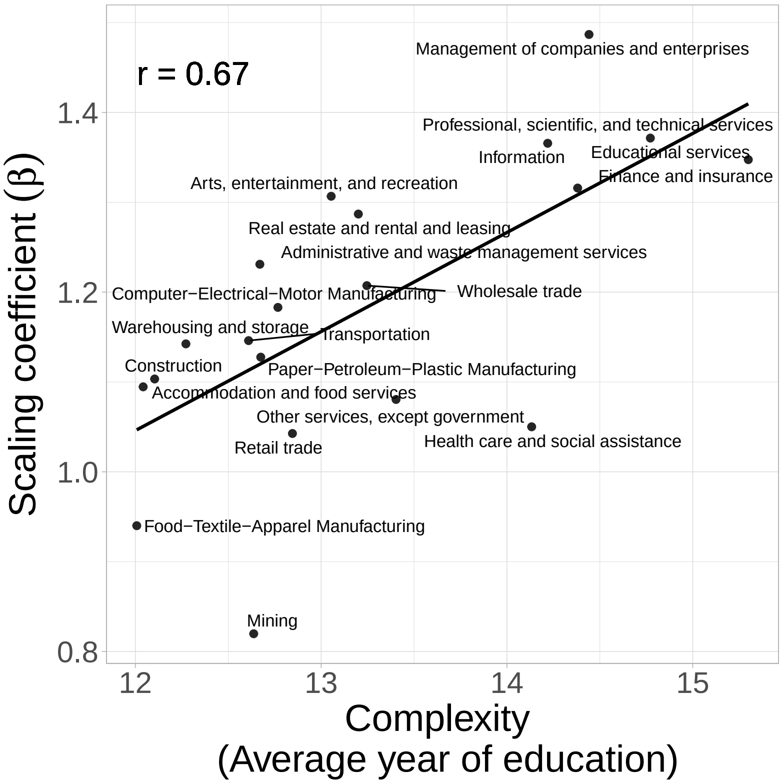


Figure 3.18 – Scaling and complexity of technological classes

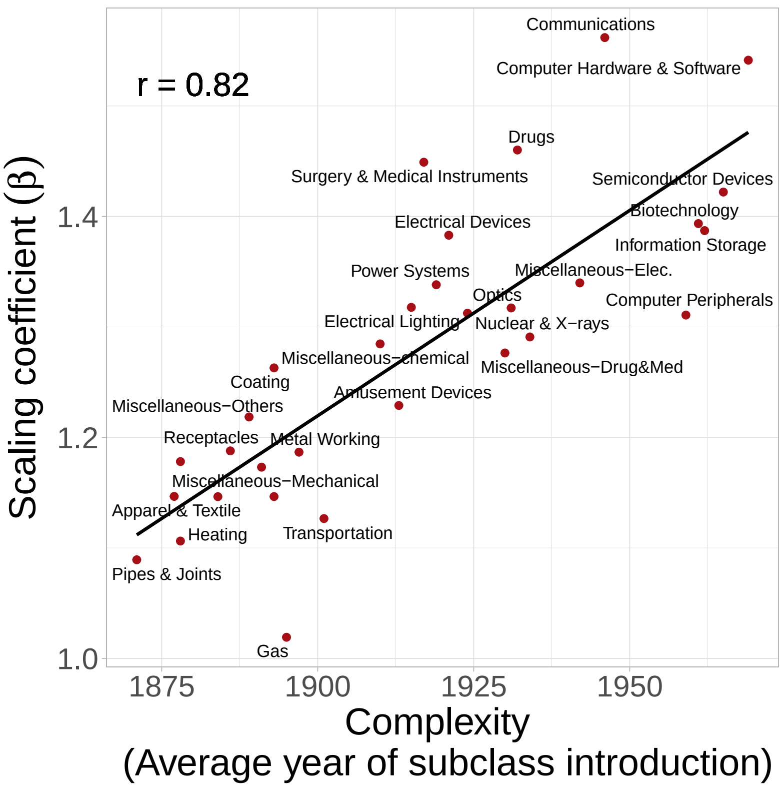
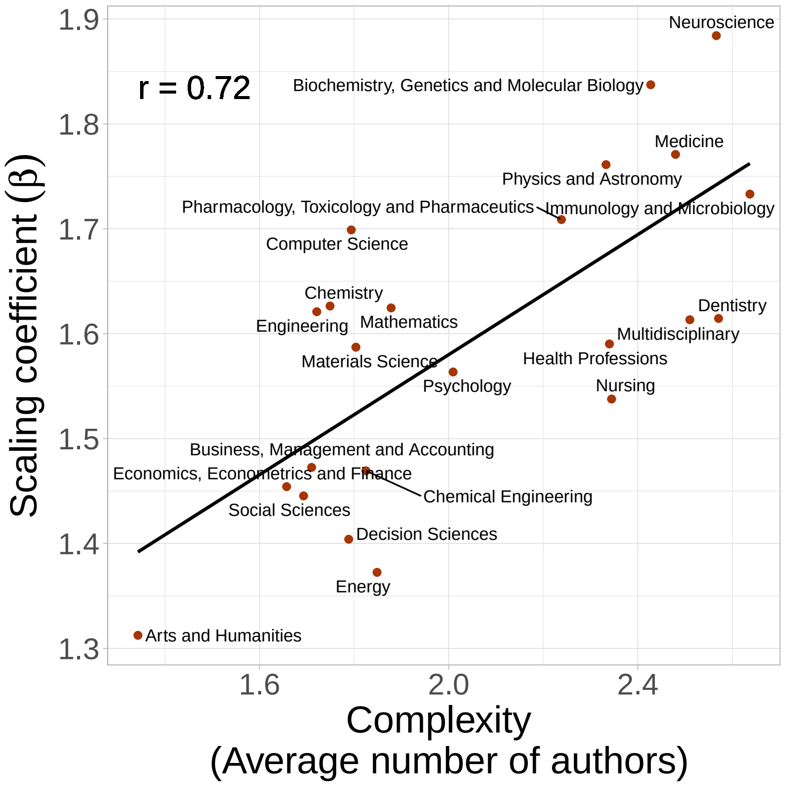


FIGURE 3.19 – SCALING AND COMPLEXITY OF RESEARCH FIELDS



production, and computer and mathematical, considered in Figures 3.14 and 3.15, at the two extremes of the y-axis. The correlation between scaling and complexity is 0.62. As suggested by Jones (2009) more education likely indicates a larger educational burden and therefore the need to divide labour. This division of labour likely requires face-to-face contact to profit from agglomeration benefits, as education is the embodiment of tacit knowledge, see Hausmann et al. (2014), which cannot be communicated easily through telecommunications (Gaspar and Glaeser, 1998; Storper and Venables, 2004; Glaeser, 2011).

Figure 3.17 shows a similar pattern for industries. Industries that employ workers that have had more years of education tend to concentrate more strongly in space. Food-textile-apparel manufacturing, which employs more workers in production, scales less strongly than information, which includes software publishing and employs more computer and mathematical occupations. Mining is a bit off compared to the others but this is likely due to the specific often less-populated areas where this activity takes place.[5]

[5]Note that this also holds for the sectors Utilities, and Agriculture, forestry, fishing, and hunting, which are not shown in this figure but can be consulted in the appendix.

Figure 3.18 shows that urban scaling of technological classes is strongly associated with the average year of subclass introduction. This indicates new and unfamiliar knowledge is involved, which as suggested by a large literature, *e.g.* Breschi and Lissoni (2001); Leamer and Storper (2001); Storper and Venables (2004); McCann (2008) and Glaeser (2011), requires face-to-face contact to communicate and make use of agglomeration benefits and therefore scales strongly. Following Jones (2009), it also is likely that a stronger division of labour is needed when more recent knowledge is recombined, as is shown in the appendix.

Figure 3.19 shows that urban scaling of academic fields is strongly associated with the average team size involved in an article. Confirming again that a larger division of labour in knowledge-intensive activities leads to a larger need for the benefits of proximity to a large number of other individuals.

3.4.4 Robustness of the second step

In Appendix 3A, we show that previous results are robust to a variety of other measures and approaches. Regression results show that the relation between spatial concentration and knowledge complexity are statistically significant even when using activity and city fixed effects. Results are also robust to using alternative measures for knowledge complexity, like the NK measure by Fleming and Sorenson (2001). This is also the case when alternative measures are used for spatial concentration, proxied in the main analysis by urban scaling. Results using Brazilian data show that the pattern of complex activities concentrating in large cities is not limited to the U.S.A.

3.4.5 Density results

Density is often seen as more important for productivity than the size of cities in the line of literature started by Ciccone and Hall (1996) and reviewed by Rosenthal and Strange (2004). In Appendix 3A, we show results based on the scaling of density instead of population size. Results show that activities scale even more strongly based on density than population size with some categories within activities attaining scaling coefficients over 3.5, which is more than double for most of the maximum scaling coefficients obtained in the main results. Nonetheless, note that this is not sufficient proof that density matters more than population size, as both measures are strongly correlated but on different scales.

3.4.6 Concentration of complex economic activities in large cities over time

In this section, patent data since 1850 is considered to evaluate if complex activities have increasingly concentrated in cities over time. The trend in the rising importance of cities over time and the strong association between complexity and urban scaling may suggest that this relation has increased.

Figures 3.20 and 3.21 depicts the evolution in the scaling of patents based on, respectively, their level of complexity and per NBER category. Figure 3.20 shows that the 25% most complex patents, those that recombine new knowledge, have become increasingly concentrated in large cities over time, in particular during industrial revolutions. In 1870, at the time of the electrical revolution, the scaling coefficient (β) was about 1.15 and grows to about 1.55 around 1930 then it plateaus until the computer revolution of 1970. After which it increases again. This would be in line with the increasing need for proximity in times of rapid economic change, as explained by Vernon (1960); Chinitz (1961); Jacobs (1969) and Duranton and Puga (2004a), which characterizes the initial phase of technological revolutions following Schumpeter (1942). When new technology matures the urban scaling plateaus.

The least complex patents (light yellow line) have always been less geographically concentrated than the most complex patents but follow a more or less similar trend up to 1970. There the two diverge with the urban scaling of the most complex increasing and the least complex patents decreasing. This is in line with the suggestion of Leamer and Storper (2001) that the rise of communication and transportation technologies has a two-fold effect: (1) an increase in the routinisation of low-complex knowledge allowing it to mature faster and therefore be performed at a distance, exemplified here by the trend of the least complex patents; (2) an increase in the complexity of new innovations, which requires proximity to develop, exemplified here by the trend of the most complex patents. This observation contradicts the results of the NEG-based model of McCann (2008) that suggested that both low complex and high complex goods concentrate in large cities due to larger home markets and reduced transportation costs.

The relation between urban scaling and technological cycles is more strongly exhibited in the evolution of individual NBER technology categories shown in Figure 3.21. In 1870, patents in mechanical, chemical, and others, which in this time period refers mostly to textile-related inventions were among the most important. The technologies in mechanical and others plateau and then decrease in scaling as these technologies mature. With a strong decrease after 1960, which suggests that a large part of the drop

Figure 3.20 – Evolution of the urban scaling of technologies by complexity

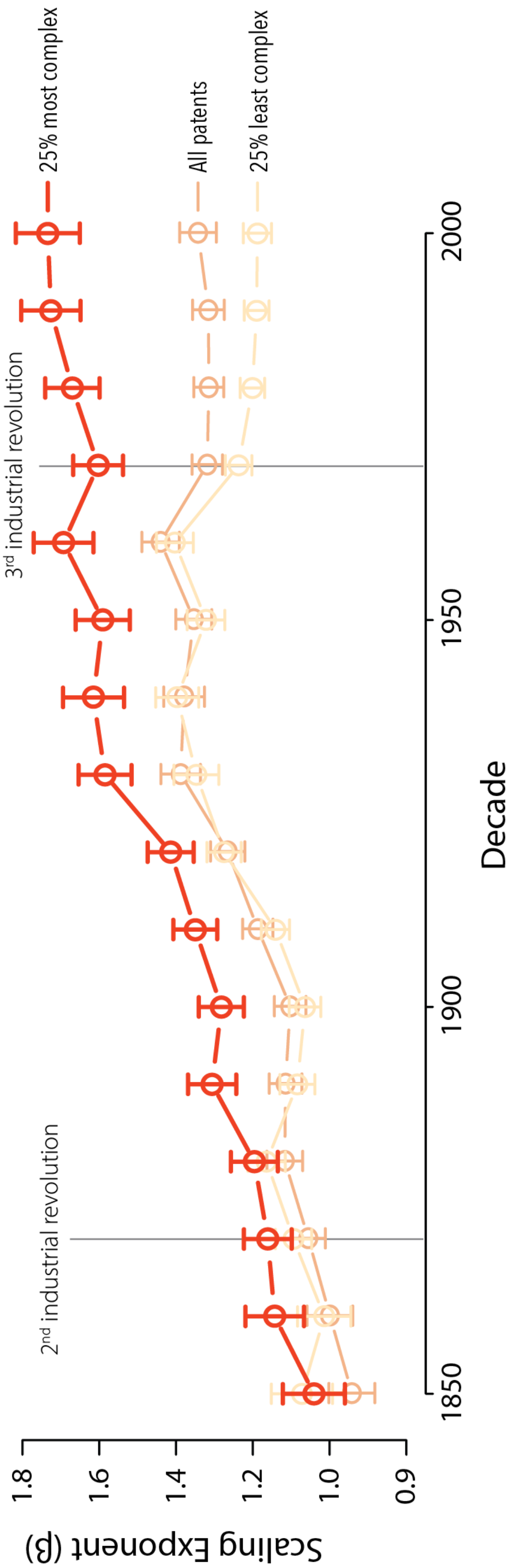
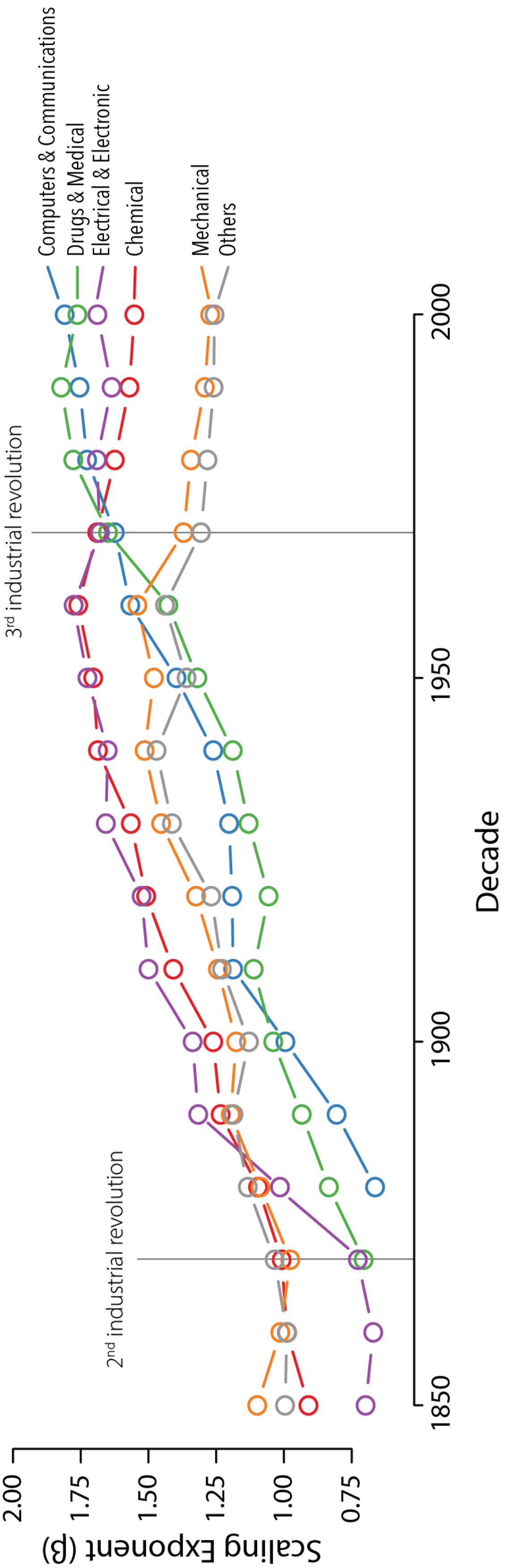


FIGURE 3.21 – EVOLUTION OF THE URBAN SCALING OF TECHNOLOGIES BY NBER CATEGORY



in urban scaling of the 25% least complex patents is due to the decrease in scaling of these technologies. Chemical shows a similar trend but scores much higher in scaling and complexity due to different new subcategories coming to being, which requires more concentration.

A sharp increase in scaling of patents in electrical & electronic can be denoted just after the 1870 electrical revolution. When rapid developments in these technologies take place the urban scaling also increases. This suggests that most of the increase in scaling in the most complex patents in Figure 3.20 around 1870 is due to this category. By 1930 the scaling level plateaus and starts decreasing after 1960.

Just after the 1970 computer revolution Computers & Communications become the dominant technology where rapid development takes place. Thereby driving up the upwards trend of the 25% most complex patents after 1970 in Panel A with Drugs & Medical.

All in all, this section gives strong suggestive evidence that there is a strong increase in the relation between urban scaling and the complexity of activity over time. As van der Wouden (2020) shows that complexity has gone up since 1870, whether measured in team size or according to the NK measure, this gives insight on why cities have become more important over time and particularly since 1980. As cities increase their relevance for a finer division of labour, they also increase their potential as engines of growth.

The fact that mostly the 25% most complex patents have come to scale more strongly suggests that a smaller and smaller number of cities manages to host a larger and larger proportion of the most innovative ideas its benefits, which sheds light on the growth in spatial inequality.

Note that the results of this section cannot be due to cities growing faster than rural areas, since scaling is a relative proportional measure. This indicates that even though cities have grown faster than rural areas the rate at which they came to harbour economic complex activities has increased even faster.

3.4.7 Robustness of the third step

In Appendix 3A, we show that the results hold when respectively using the number of claims of each patent instead of the number of patents; when reducing the sample of cities to the 353 cities used in the second step; and when using different quantiles of the most, respectively, least complex patents.

3.5 Conclusion

More knowledge-intensive activities are known to concentrate in larger cities. This tendency has been thought to have increased over time and likely plays a major role in the growing inequality of activities over space. Despite this relevance, the understanding of the spatial concentration of knowledge-intensive activities is limited in both the literature of urban economics/economic geography and in urban scaling to aggregate count measures, like the number of patents. These fail to appreciate the diversity in knowledge intensity within these activities, *i.e.* not every patented invention is equally innovative, and to explain why these activities tend to concentrate more.

Here, use is made of ideas in innovation studies and complexity theory on when activities require geographical proximity between humans to develop theoretically sound empirical measures of knowledge intensity within activities and over time. The theories behind the measures are related to the division of labour, as often attributed to Smith (1776), and the transmission of tacit knowledge. We show that each of these measures is strongly related to a stronger spatial concentration in large urban areas. When considering patent data from 1850 onwards, we show that the spatial concentration of complex activities has been increasing over time, in particular during industrial revolutions. All in all, this demonstrates that there is an almost universal tendency for complex activities to concentrate in large cities and that this tendency is accelerating.

3A Appendix: additional results and robustness checks

Additional results - Urban Scaling of all economic activities

In this section we show the scaling laws followed by all economic activities used in the main analysis: 2-digit NBER sub-categories (technologies), 2-digit Scopus AJSC (research areas), 2-digit NAICS (industries), and 2-digit SOC (occupations).

Technological classes

The different scatter plots show the relation between the population of an MSA and the number of patents produced in different 2-digit NBER sub-categories from 2000 to 2009.

FIGURE 3A1 – SCALING RELATIONSHIP FOR EACH PATENT SUB-CATEGORY.

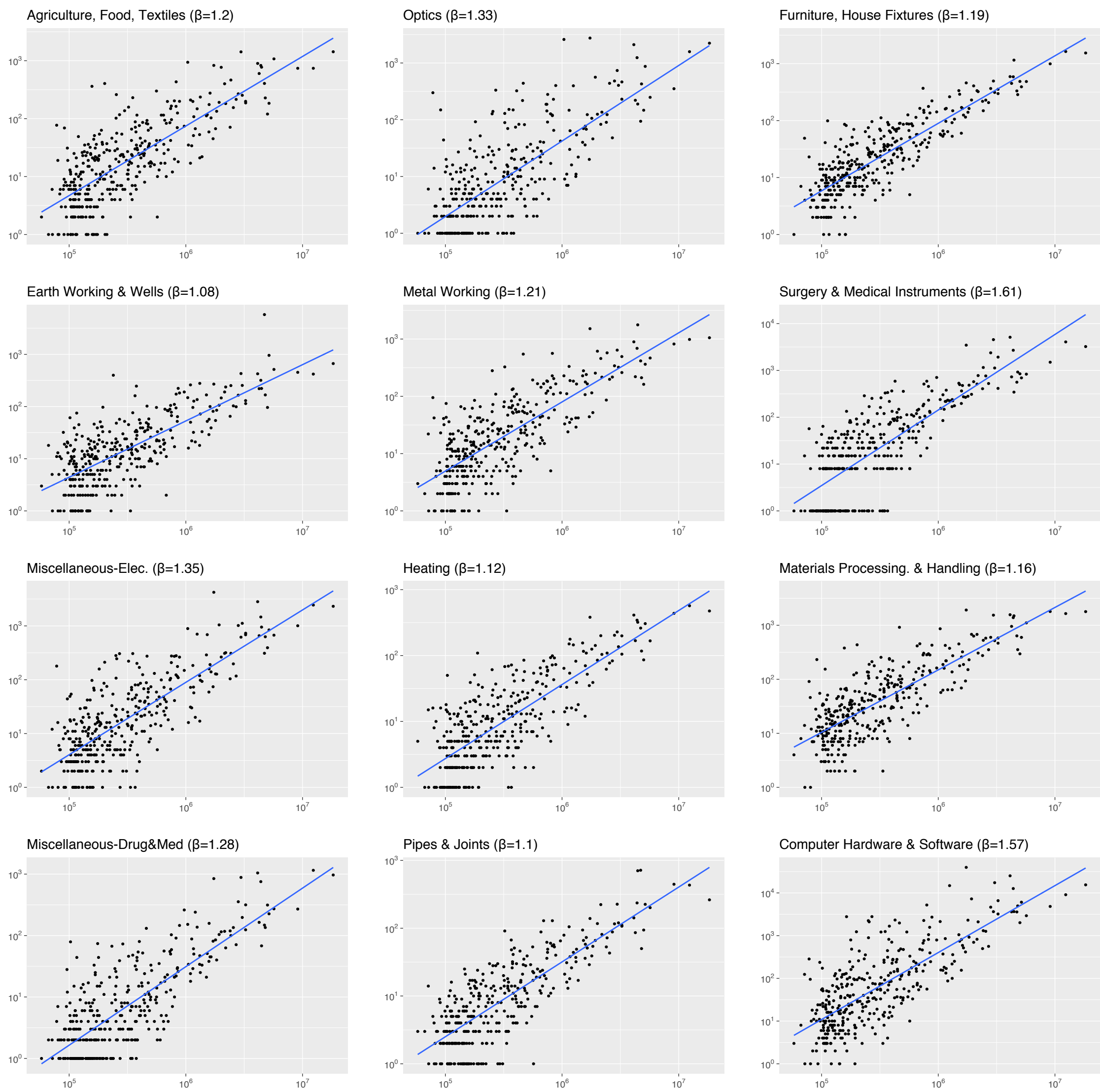


Figure 3A2 – Scaling relationship for each patent sub-category.

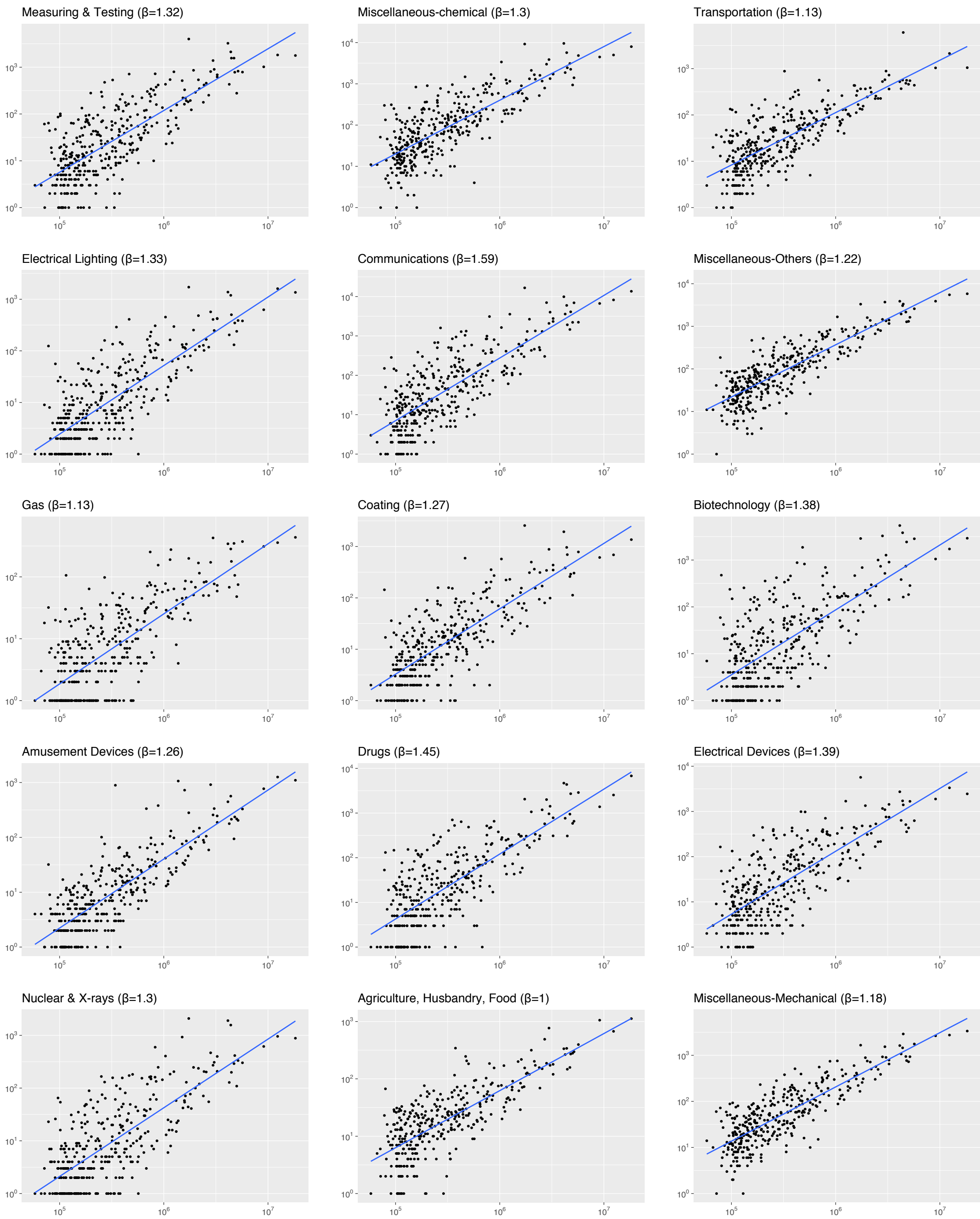
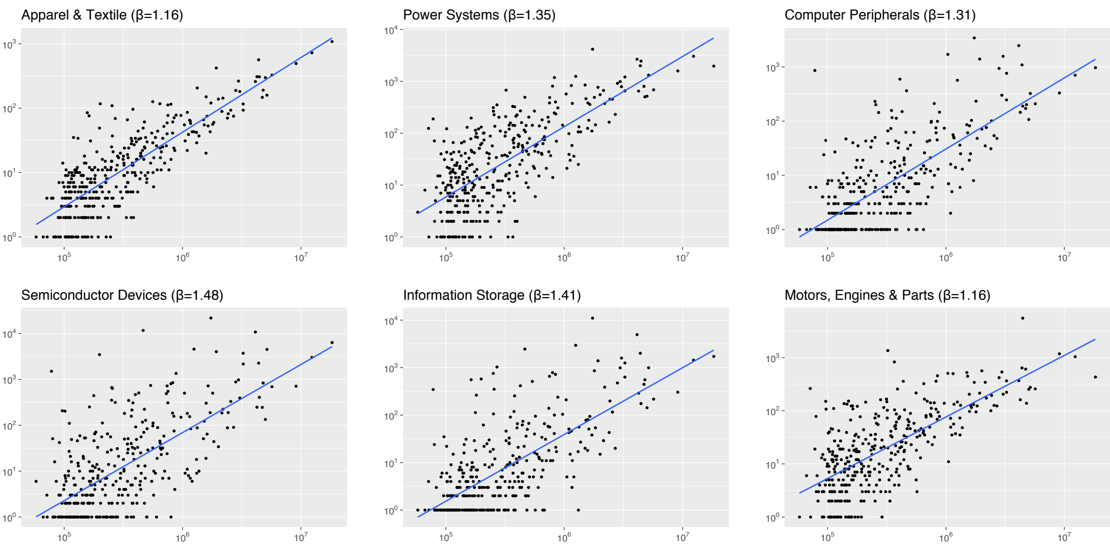


FIGURE 3A3 – SCALING RELATIONSHIP FOR EACH PATENT SUB-CATEGORY.



Scientific fields

The different scatter plots show the relation between the population of an MSA and the number of research papers produced in different 2-digit Scopus AJSC research areas from 1996 to 2008.

FIGURE 3A4 – SCALING RELATIONSHIP FOR EACH SCIENTIFIC FIELD.

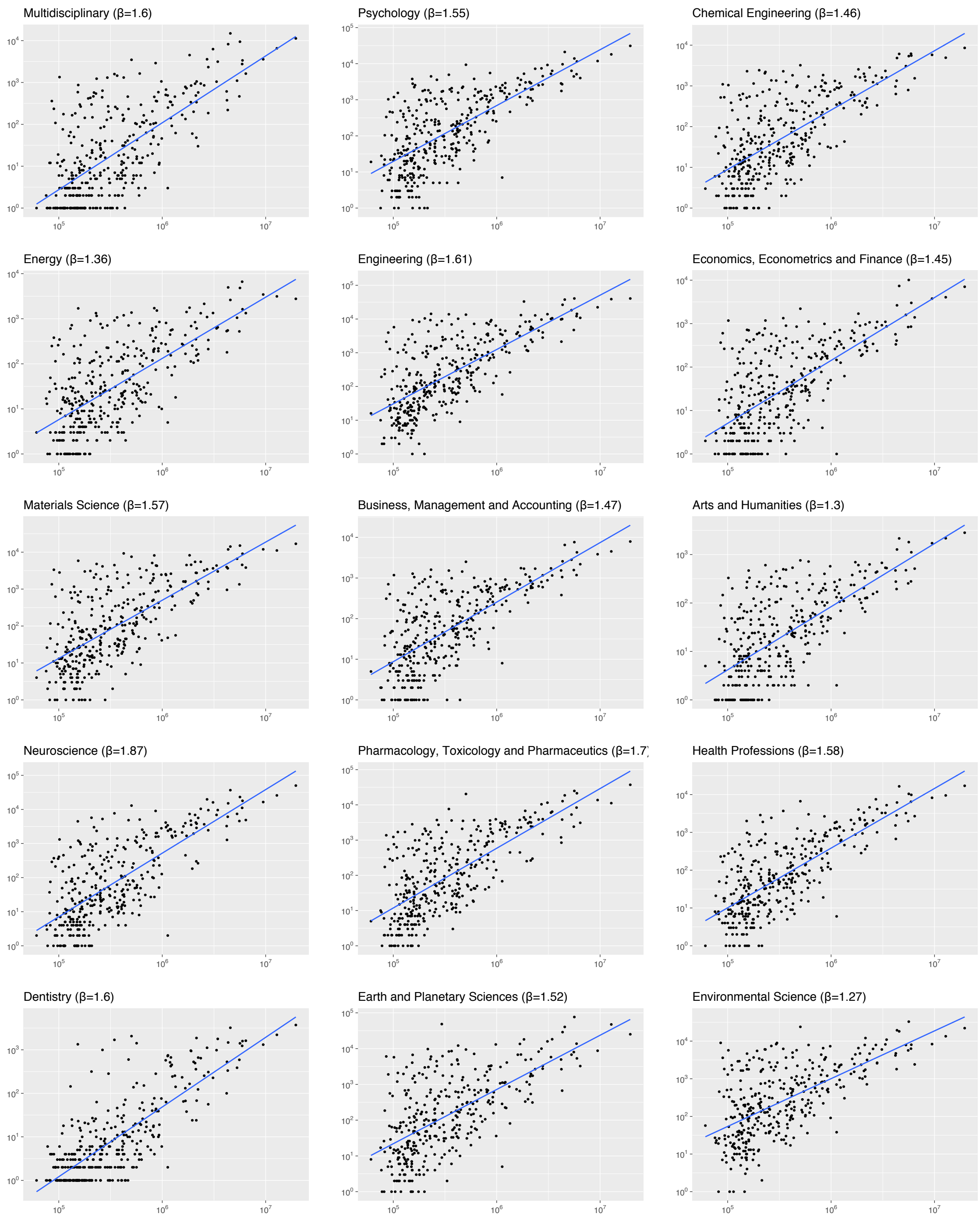
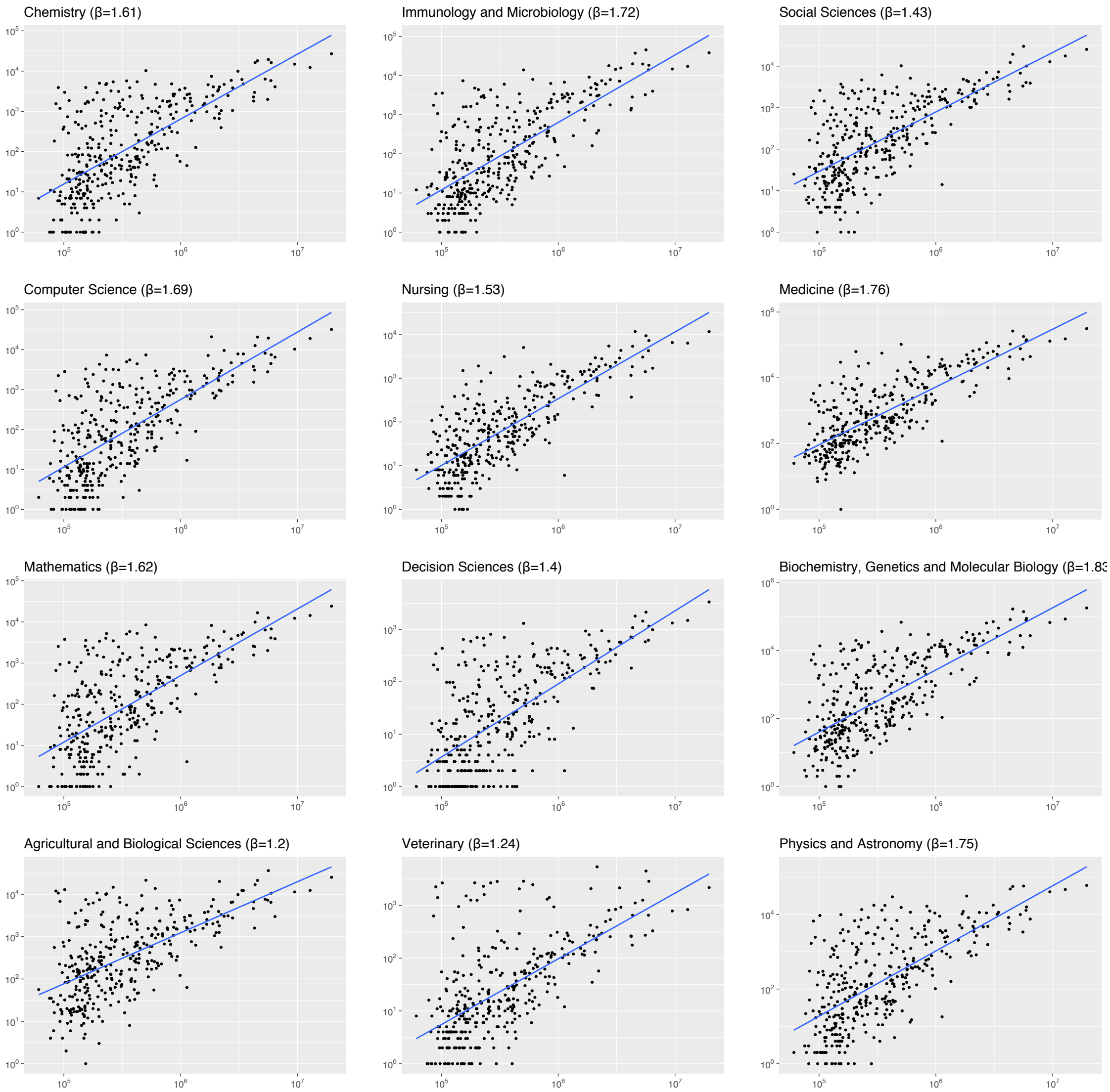


FIGURE 3A5 – SCALING RELATIONSHIP FOR EACH SCIENTIFIC FIELD.



Industries

The different scatter plots show the relation between the population of an MSA and GDP produced in different 2-digit NAICS sectors in 2015.

FIGURE 3A6 – SCALING RELATIONSHIP FOR EACH INDUSTRY.

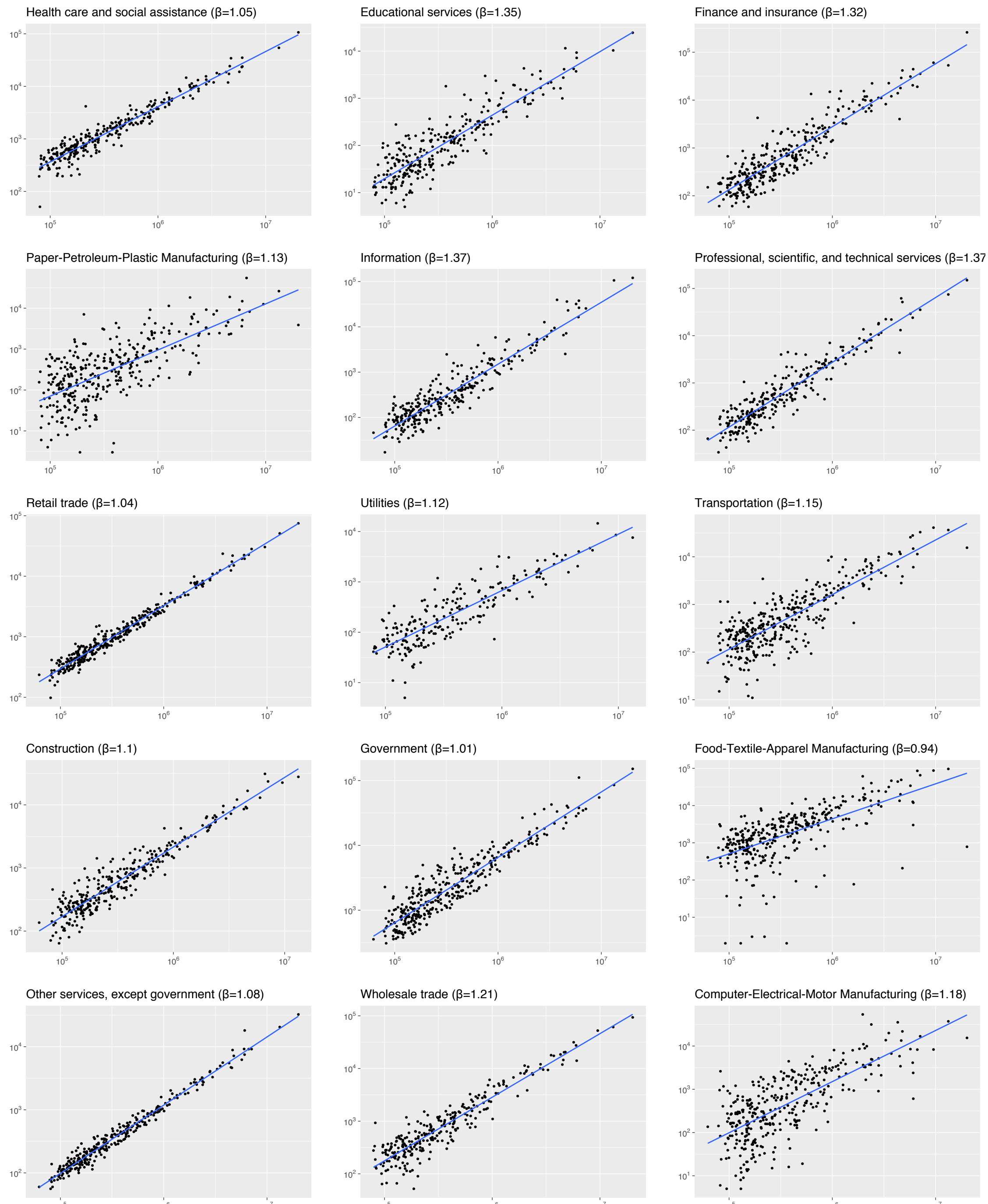
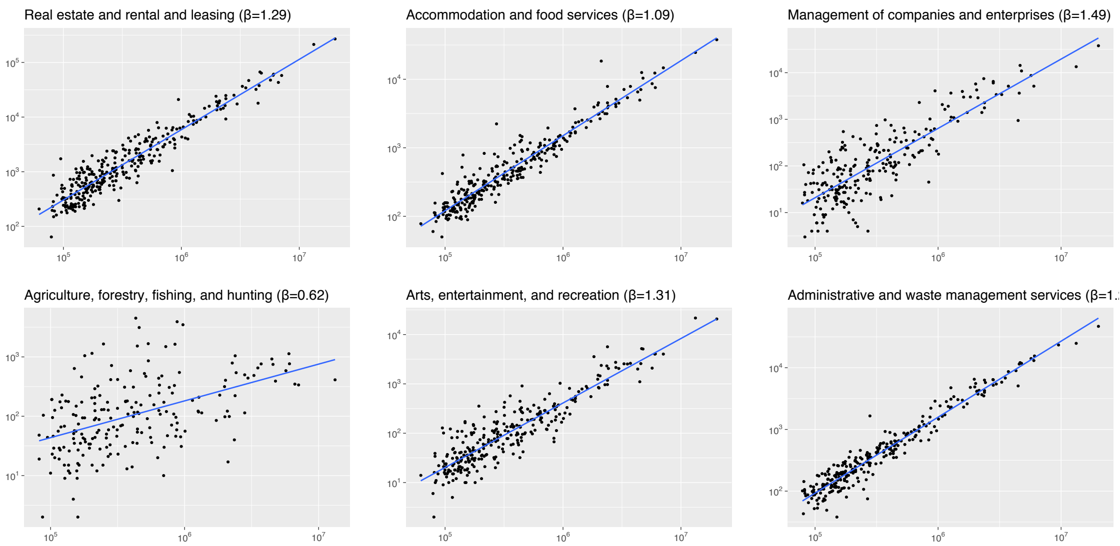


FIGURE 3A7 – SCALING RELATIONSHIP FOR EACH INDUSTRY.



Occupations

The different scatter plots show the relation between the population of an MSA and the number of employees in different 2-digit SOC in 2015.

Figure 3A8 – Scaling relationship for each occupational category.

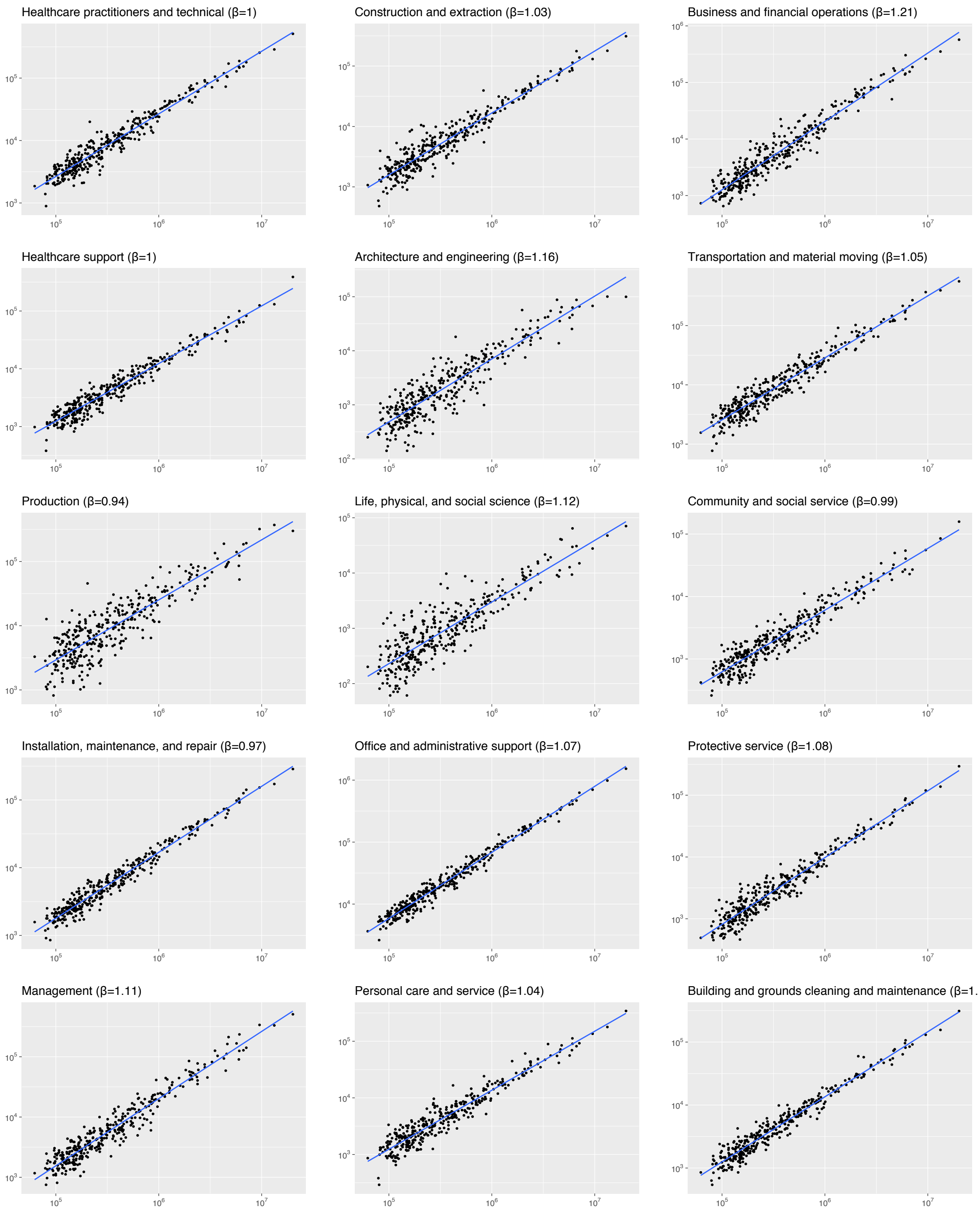


Figure 3A9 – Scaling relationship for each occupational category.

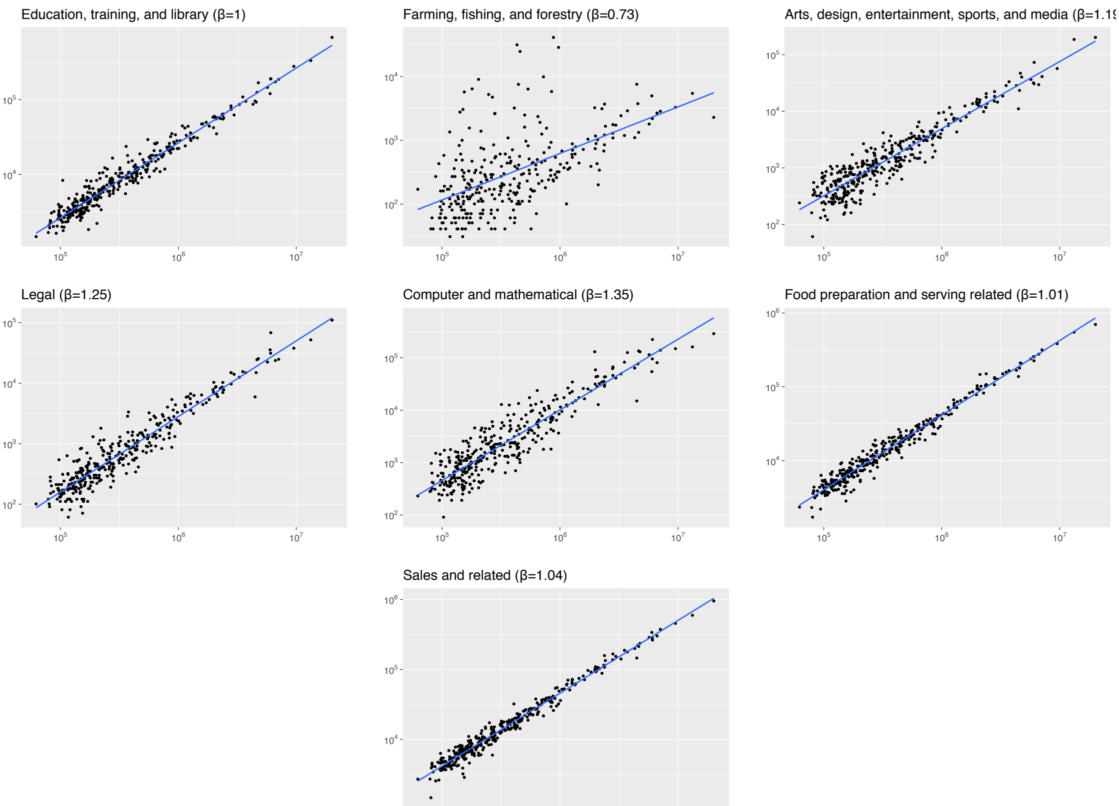
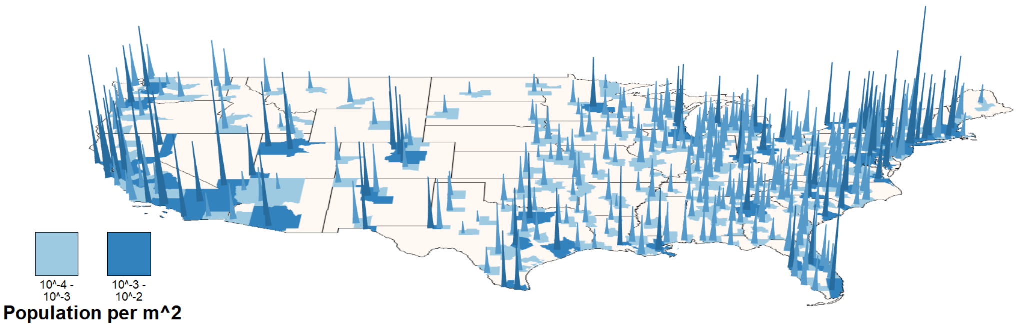


Figure 3A10 – Spatial concentration of Population in terms of density



Additional results - Density scaling

In this section results are shown when scaling is not measured on the basis of the population size but the population density of cities, as put forward in Ciccone and Hall (1996) and many studies reviewed by Rosenthal and Strange (2004). Density is measured by the natural logarithm of the inhabitants per square meter of built area. The built area per MSA is obtained using the National Land Cover Database.

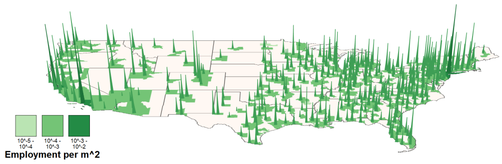
Figure 3A10 shows the population density across the United States. Compared to population per city shown in Figure 3.5 it is clear that the differences in population density are less large than in population size. Nonetheless, New York and Los Angeles stand out like before. Chicago is in terms of inhabitants per square meter of built area no longer in the top 10 while on the other hand the lesser known Salisbury Metro Area in Maryland makes its appearance.

Figures 3A11-3A14 replicate the baseline results presented in Figures 3.6-3.13 when using urban density instead of population size as independent variable when explaining the size of activities taking place in a city. The left-hand panel shows maps of the density of each activity. Like in the maps on population, the maps on density show much less variation between cities than when taking the totals, like in the main results. This is particularly the case for employment and GDP, which are more generally spread than publications and patents, in which a few cities are strongly specialised.

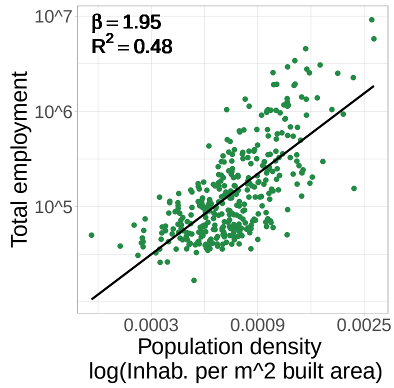
The right-hand panel shows the scaling relationship between density and production in an activity. Two major differences with respect to the baseline results in Figures 3.6-3.13 stand out: first, the scaling coefficient β is much higher, meaning that an increase in density leads to a larger increase in production than an increase in population. An 1% increase in density is even associated with a 3.22% increase in patent production;

second, the R^2 is much smaller, meaning that there is a larger variation in the extent to which density is associated to more production compared to population size.

Figure 3A11 – Urban density scaling of employment

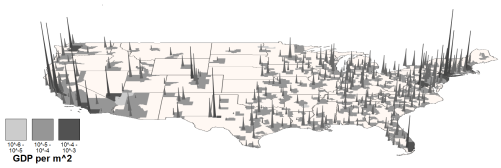


(a) Spatial concentration of employment

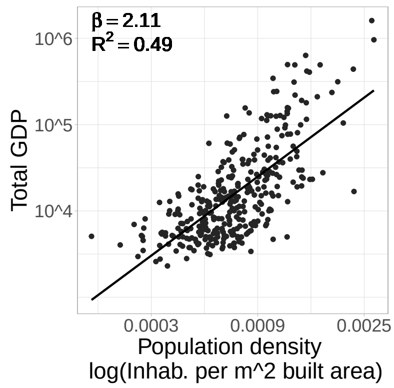


(b) Density scaling of employment

Figure 3A12 – Urban density scaling of GDP

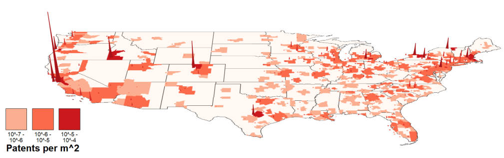


(a) Spatial concentration of GDP

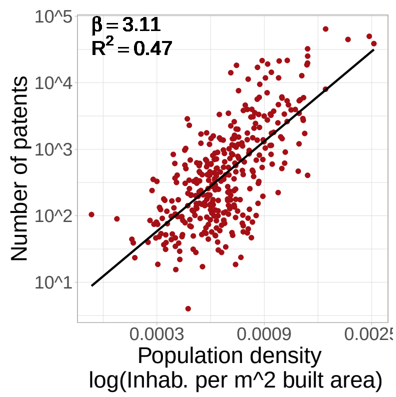


(b) Density scaling of GDP

Figure 3A13 – Urban density scaling of patents

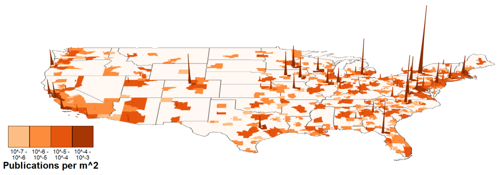


(a) Spatial concentration of patents

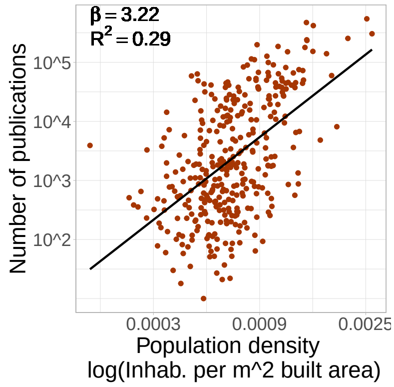


(b) Density scaling of patents

Figure 3A14 – Urban density scaling of publications

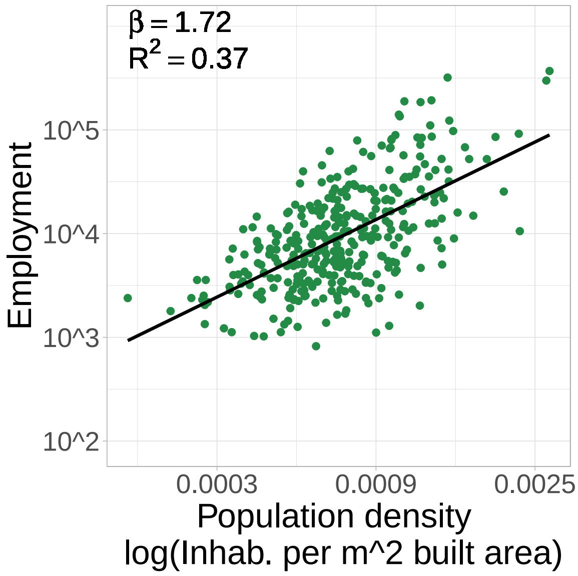


(a) Spatial concentration of Publications

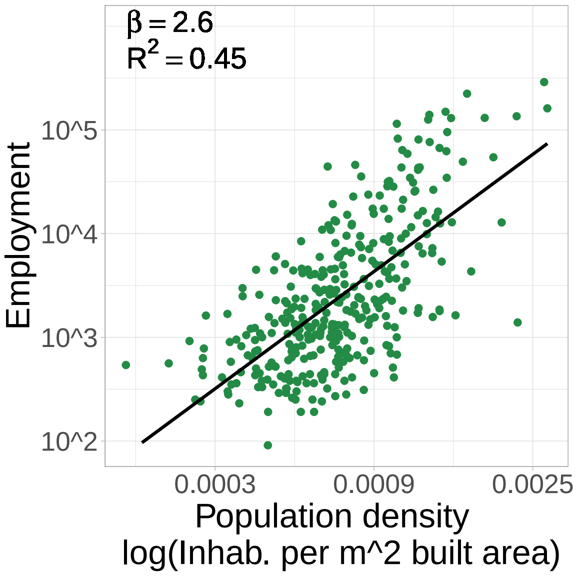


(b) Density scaling of publications

Figure 3A15 – Urban scaling of two job categories



(a) Production

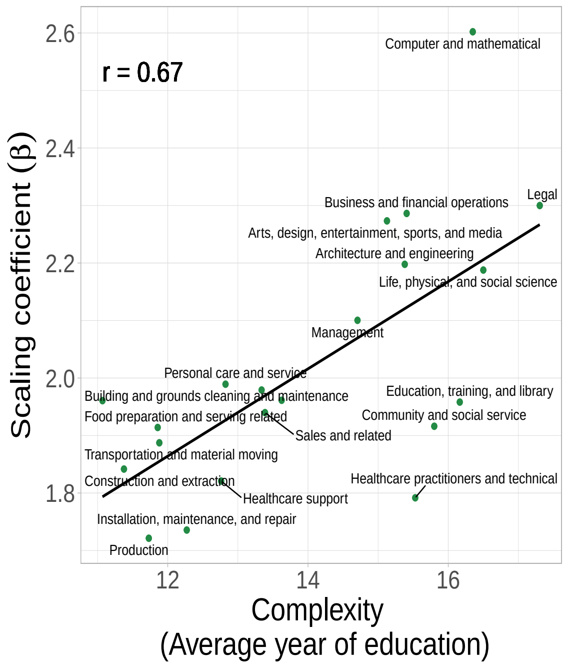


(b) Computer and mathematical

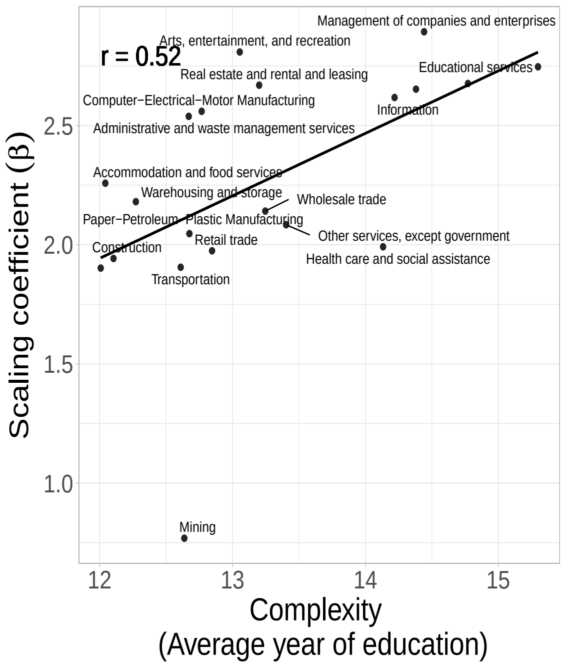
In Figure 3A15 a replication of Figures 3.14 and 3.15 when using population density instead of population size is used for the example job categories of production workers and computer and mathematical professions. Like before, the beta coefficients and variance are larger when based on population density. Nonetheless, the more complex job occupation scales much more strongly than the less complex one.

Figure 3A16 replicates Figures 3.16-3.19 when using population density instead of population size. Two things stand out: first, all scaling coefficients are higher than before as could be expected by the results in Figures 3A15 and Figures 3A11- 3A14; second, the regression lines are much steeper, which suggests that more complex activities scale even more strongly with larger densities than with larger population sizes. Nonetheless, this is only suggestive evidence that density matters more for complex activities than population size and further research with a more rigorous statistical or experimental design is needed to establish this with certainty.

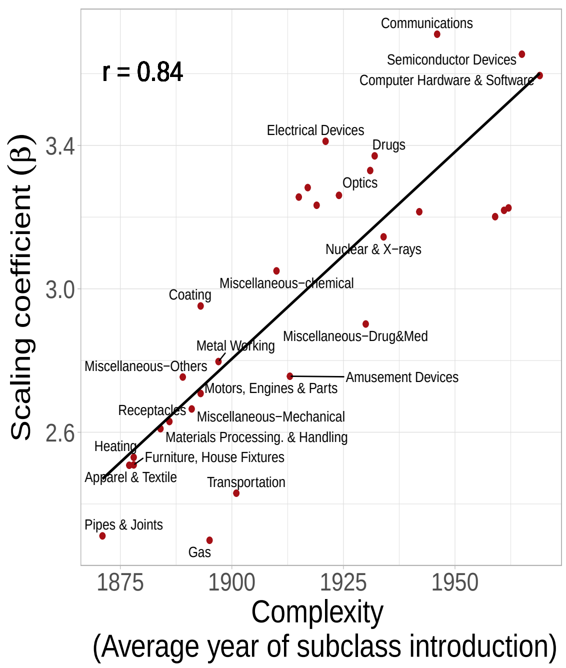
Figure 3A16 – Urban density scaling and complexity



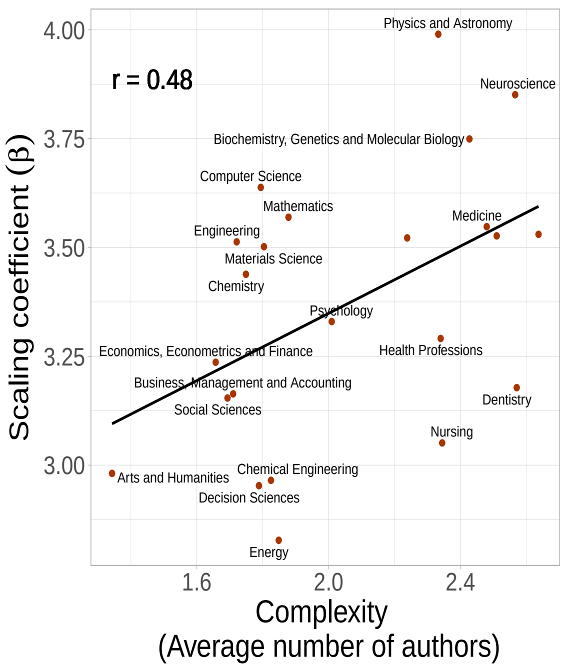
(a) Density scaling and complexity of job categories



(b) Density scaling and complexity of industries



(c) Density scaling and complexity of technological classes

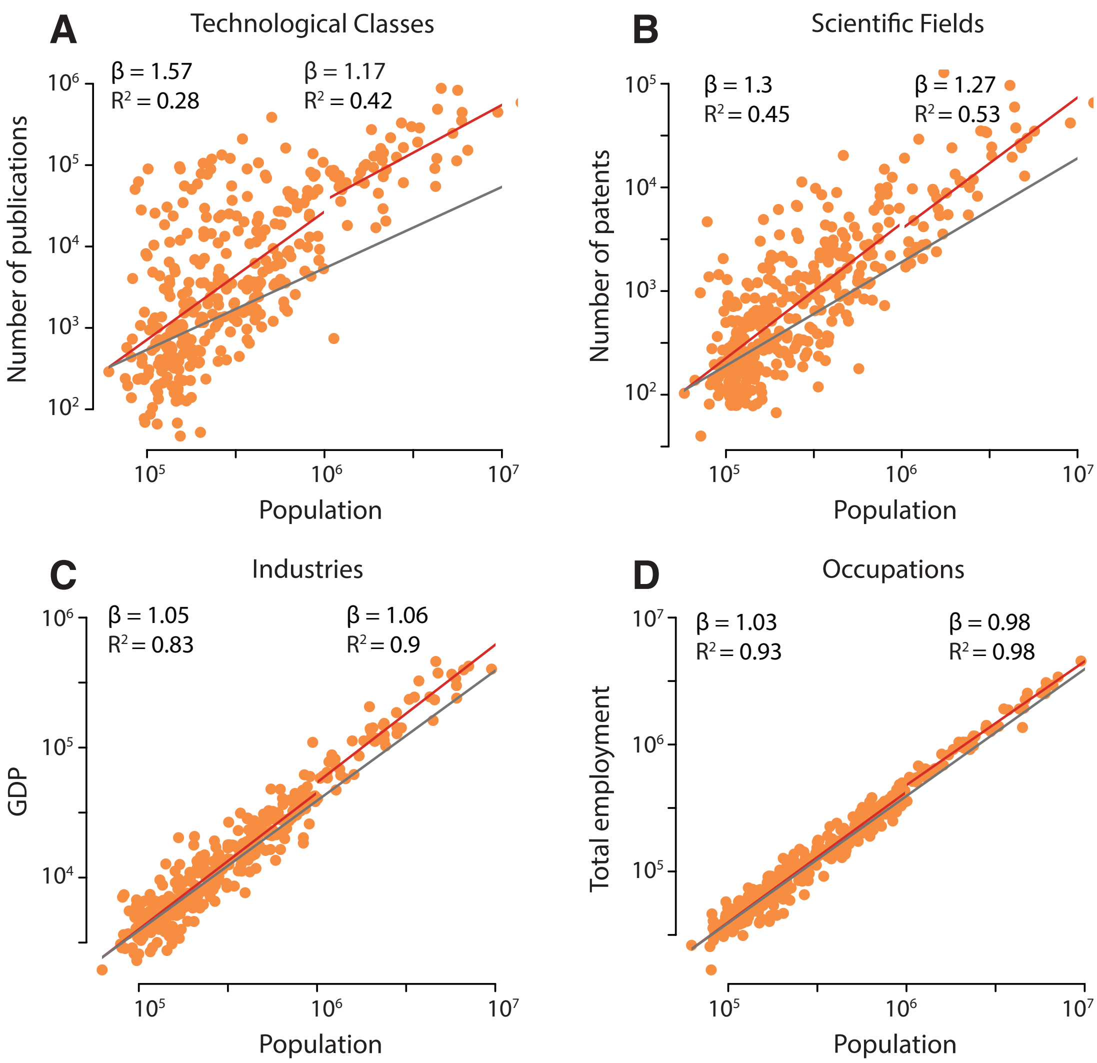


(d) Density scaling and complexity of research fields

Overall scaling for different population thresholds

Figure 3A17 shows that the beta coefficients estimated from a pool of large (>1M) or medium-sized cities (<1M) are of a similar magnitude for scientific fields, industries, and occupations. In the case of patents, we find superlinear scaling for both samples, but interestingly we find higher superlinear scaling for medium sized cities (1.57) than for large cities (1.17).

Figure 3A17 – Scaling for scientific fields, technology classes, industries, and occupations for cities above and below 1 million people.



Robustness checks - Knowledge Complexity and Urban Scaling

Multivariate regression model for the relation between knowledge complexity and concentration

The main hypothesis of this Chapter is that knowledge complexity drives urban concentration. Here we formalise this hypothesis using a regression model that we can fit using our data. Let Y_{ic} be the output in city c of economic activity i, meaning number of patents for technological classes, number of papers for scientific fields, GDP for industries, and number of employees for occupation categories. The scaling is captured by the scaling exponent β_i from the following model:

$$\log(Y_{ic}) = \mu_i + \beta_i \log(Pop_c) + \epsilon_{ic}, \tag{3A.1}$$

where Pop_c is the population of city c. In the main analysis we found a positive correlation between $\hat{\beta}_i$ and K_i, where K_i is the knowledge complexity of economic activity i. This relation can be formalised be $\beta_i = \alpha_0 + \alpha_1 K_i$, where the correlation presented in the main analysis is proportional to α_1. To estimate the value of α_1 we will combine both equations into the following model:

$$\log(Y_{ic}) = \mu_i + \alpha_0 \log(Pop_c) + \alpha_1 K_i \log(Pop_c) + \epsilon_{ic}. \tag{3A.2}$$

The coefficient of the interaction term between the knowledge complexity of the economic activity and the population of the city captures the hypothesis that more complex activities are more concentrated. We also add city-level fixed effects and estimate the following model:

$$\log(Y_{ic}) = \mu_i + \eta_c + \alpha_1 K_i \log(Pop_c) + \epsilon_{ic}. \tag{3A.3}$$

Table 3A1 shows that in all four cases, the interaction term between knowledge complexity and city population is positive and significant, even after adding city fixed effects. This means that complex industries located in large cities tend to generate more economic output than the same industry located in a smaller city, and also more than a less complex industry located in the same city. Table 3A2 presents results of the same model over all economic activities, including the natural resources that were ignored in the main analysis. Finally, Table 3A3 shows estimations for the same model at a finer level of aggregation, when available. All the results are consistent.

TABLE 3A1 – RELATION BETWEEN CONCENTRATION AND KNOWLEDGE COMPLEXITY

	Dependent variable:											
	Economic Output											
	Number of patents			Number of papers			GDP per industry			Number of employees		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Area (log)	−0.295***	−0.295***		0.010	0.010		−0.088***	−0.088***		0.014**	0.015**	
	(0.015)	(0.015)		(0.028)	(0.028)		(0.015)	(0.014)		(0.007)	(0.007)	
Population (log)	1.412***	1.412***		1.580***	1.580***		1.234***	1.234***		1.071***	1.071***	
	(0.012)	(0.012)		(0.022)	(0.022)		(0.011)	(0.011)		(0.005)	(0.005)	
Pop.; Knowledge int.		0.109***	0.109***		0.104***	0.104***		0.096***	0.100***		0.062***	0.061***
		(0.010)	(0.007)		(0.018)	(0.009)		(0.009)	(0.009)		(0.004)	(0.004)
Constant	−8.716***	−7.602***	1.429***	−17.371***	−19.000***	−0.168	−7.265***	−5.760***	7.132***	−5.251***	−5.540***	7.658***
	(0.293)	(0.310)	(0.162)	(0.529)	(0.599)	(0.222)	(0.272)	(0.307)	(0.232)	(0.129)	(0.129)	(0.079)
Activity f.e.	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
MSA f.e.	no	no	yes	no	no	yes	no	no	yes	no	no	yes
Observations	11,296	11,296	11,296	8,050	8,050	8,050	5,457	5,457	5,457	7,382	7,382	7,382
R^2	0.644	0.647	0.834	0.562	0.564	0.906	0.803	0.807	0.845	0.928	0.930	0.951
Adjusted R^2	0.643	0.646	0.828	0.561	0.563	0.901	0.803	0.806	0.834	0.928	0.930	0.949
F Statistic (α_1)		116***	238***		32.9***	146***		105.9***	130.0***		202***	274***
F Statistic	616***	608***	142***	430***	415***	199***	1,168***	1,136***	75.22***	4,338***	4,271***	367***

Notes: Knowledge complexity measures are standardised; natural resources are excluded; also $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

TABLE 3A2 – RELATION BETWEEN CONCENTRATION AND KNOWLEDGE COMPLEXITY WHEN INCLUDING NATURAL RESOURCES.

	Dependent variable:											
	Economic Output											
	Number of patents			Number of papers			GDP per industry			Number of employees		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Area (log)	−0.275***	−0.275***		0.048*	0.048*		−0.085***	−0.085***		0.032***	0.032***	
	(0.015)	(0.014)		(0.026)	(0.026)		(0.014)	(0.014)		(0.008)	(0.008)	
Population (log)	1.375***	1.375***		1.522***	1.522***		1.229***	1.229***		1.050***	1.050***	
	(0.012)	(0.011)		(0.020)	(0.020)		(0.011)	(0.011)		(0.006)	(0.006)	
Pop.; Knowledge int.		0.121***	0.121***		0.069***	0.069***		0.094***	0.097***		0.082***	0.081***
		(0.010)	(0.007)		(0.017)	(0.008)		(0.009)	(0.009)		(0.005)	(0.004)
Constant	−9.587***	−8.983***	0.080	−17.483***	−18.412***	0.482**	−8.473***	−8.682***	4.259***	−5.369***	−5.806***	7.518***
	(0.278)	(0.281)	(0.147)	(0.491)	(0.541)	(0.205)	(0.269)	(0.267)	(0.185)	(0.145)	(0.145)	(0.095)
Activity f.e.	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
MSA f.e.	no	no	yes	no	no	yes	no	no	yes	no	no	yes
Observations	12,002	12,002	12,002	9,450	9,450	9,450	5,670	5,670	5,670	7,684	7,684	7,684
R^2	0.640	0.645	0.820	0.554	0.555	0.895	0.803	0.807	0.843	0.916	0.919	0.935
Adjusted R^2	0.639	0.644	0.815	0.553	0.554	0.891	0.802	0.806	0.832	0.916	0.919	0.932
F Statistic (α_1)		160***	308***		16.81***	68.80***		104.8***	127.5***		288.68***	338.36***
F Statistic	609***	604***	141***	418.2***	405.0***	206.2***	1,152***	1,123***	76.89***	3,642***	3,633***	282.8***

Notes: Knowledge complexity measures are standardised; natural resources are not excluded; also $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

TABLE 3A3 – RELATION BETWEEN CONCENTRATION AND KNOWLEDGE COMPLEXITY FOR THE LOWER LEVEL OF AGGREGATION.

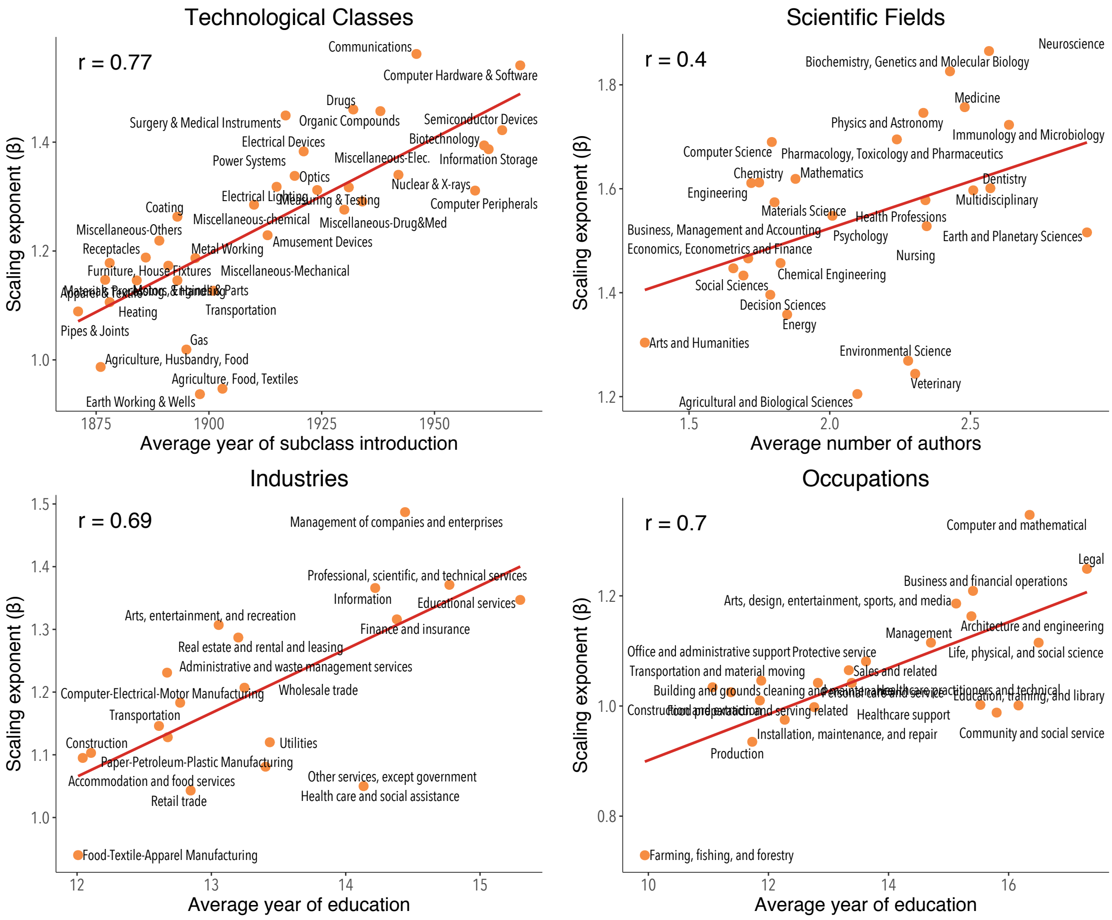
	Dependent variable:								
	Economic Output								
	Number of patents			Number of papers			Number of employees		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Area (log)	−0.191***	−0.191***		0.030***	0.030***		0.019***	0.020***	
	(0.006)	(0.006)		(0.008)	(0.008)		(0.006)	(0.006)	
Population (log)	1.111***	1.111***		1.334***	1.334***		1.110***	1.109***	
	(0.005)	(0.005)		(0.006)	(0.006)		(0.005)	(0.005)	
Pop.; Knowledge int.		0.152***	0.152***		0.120***	0.120***		0.073***	0.073***
		(0.004)	(0.003)		(0.005)	(0.003)		(0.004)	(0.004)
Constant	−8.384***	−6.850***	1.544***	−14.695***	−16.192***	−0.220***	−6.872***	−7.394***	6.273***
	(0.130)	(0.135)	(0.085)	(0.166)	(0.177)	(0.082)	(0.115)	(0.117)	(0.080)
Activity f.e.	yes	yes	yes	yes	yes	yes	yes	yes	yes
MSA f.e.	no	no	yes	no	no	yes	no	no	yes
Observations	47,302	47,302	47,302	84,350	84,350	84,350	14,618	14,618	14,618
R^2	0.601	0.612	0.746	0.512	0.515	0.841	0.913	0.915	0.932
Adjusted R^2	0.600	0.611	0.744	0.510	0.514	0.840	0.913	0.915	0.930
F Statistic (α_1)		1337.51***	2028.364***		582.084***	1766.394***		317.934***	389.576***
F Statistic	527.084***	547.869***	283.350***	364.139***	367.541***	749.996***	2,828.928***	2,843.720***	477.842***

Notes: Knowledge complexity measures are standardised; natural resources are not excluded; also $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Results including natural resources

Figure 3A18 shows the results when the natural resource sectors left out in the main analysis are included. The correlation for industries becomes smaller, though remain significant (see Table 3A2).

FIGURE 3A18 – SCALING RESULTS INCLUDING THE NATURAL RESOURCE SECTORS.



Results for all cities available in the data

Figure 3A19 – Scaling results with all the cities available for each economic activity.

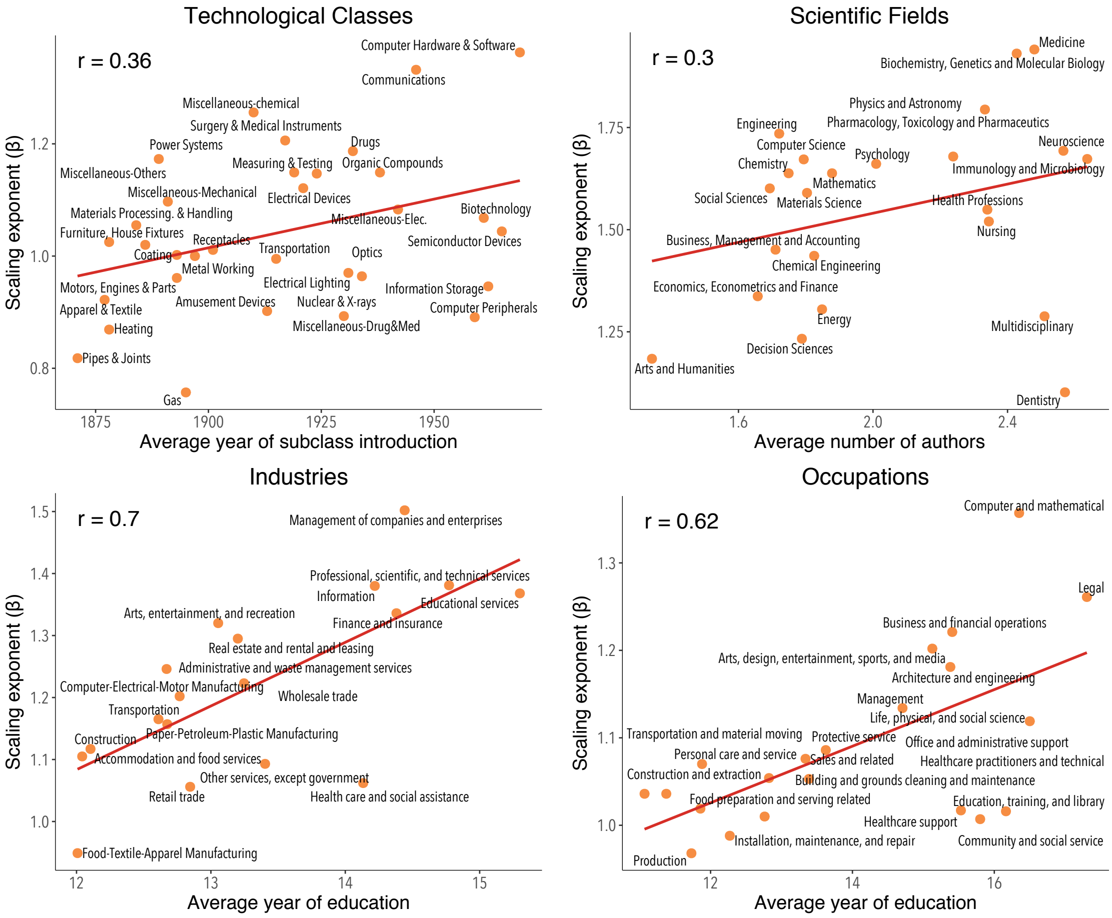


Figure 3A19 shows the results when data on all the cities available for each economic activity are used. This means 940 cities for technological classes, 923 for scientific fields, 353 for industries (which coincides with the main analysis), and 388 for occupations.

Knowledge complexity and concentration at different levels of aggregation

Figure 3A20 – Relationship between knowledge complexity and concentration for technological classes and scientific fields at a lower level of aggregation.

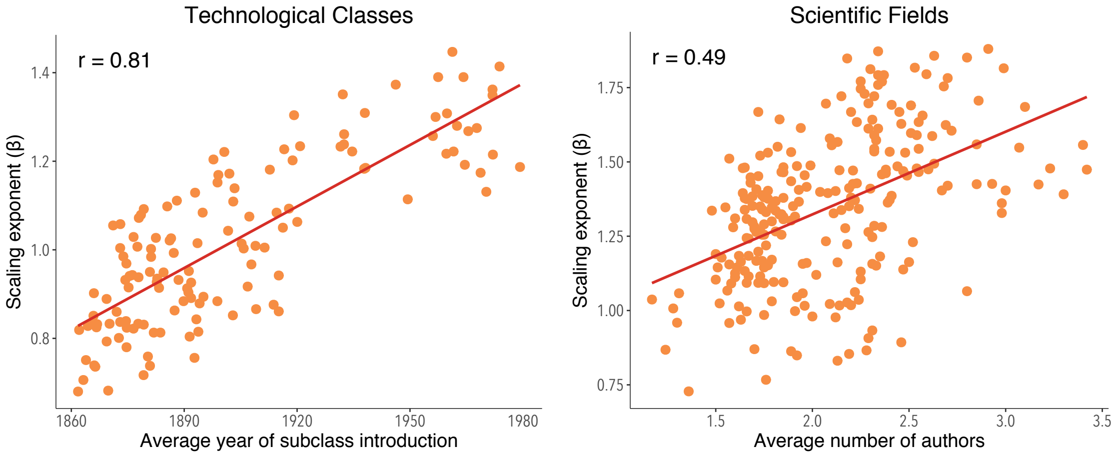


Figure 3A20 shows the scaling results for technological classes and scientific fields at a more detailed level: Technological classes are defined following the 3-digit United States Patent Classification (USPC) and scientific fields are defined as following the 4-digit All Science Journal Classification Codes (ASJC). The relation between complexity and scaling remains similar.

Alternative measures of knowledge complexity

In this section, we present the relation between scaling and other commonly used measures of knowledge complexity for occupations, industries, technologies, and scientific publications.

For technologies, a first alternative measure is the NK measure as proposed by Fleming and Sorenson (2001) and we computed it using the function implemented by Balland and Rigby (2017) in the EconGeo R package. NK measures are computed at the patent level and are based on how many sub-classes are listed on a patent and how often these sub-classes have been recombined in the past. We then average per technology category. As a second measure of knowledge complexity we use the date at which a given technology has been officially established by the USPTO. Finally, in line with the idea of division of knowledge we use the average number of inventors as a measure of complexity. The left-hand shows the relation between each alternative measure and the average year of subclass introduction used in the main analysis, while the right hand side reproduces the main results, *i.e.* the relation between scaling and knowledge complexity, using these alternative measures. As can be seen, the main results presented in the Chapter are robust to the use of these alternative measures of knowledge complexity.

Table 3A4 shows the spatial concentration of economic output for patenting activity at the 3-digit level using specification in Eq. 3A.2 for alternative measures of knowledge complexity. To make both measures comparable, we use the 3-digit patent classes because at the 2-digit level both measures of knowledge complexity (year of subclass introduction and NK-complexity) are highly correlated so the specification suffers from multicollinearity. Both measures confirm that complex activities concentrate in space.

For scientific publications an alternative measure of knowledge complexity is the average age of references cited in a given scientific publication. The data has been kindly provided by Patience et al. (2017).Figure 3A22 shows that these results are highly similar to the ones in the main analysis.

For industries an alternative measure of knowledge complexity is the share of Science, Technology, Engineering and Mathematicians (STEM) workers, as computed by Rothwell (2013). Figure 3A23 shows that results are highly similar to those in the main analysis.

For occupations we use two alternative measures of knowledge complexity: the originality index provided by the O*NET classification (https://www.onetonline.org/) and the average wage of workers (data from the Bureau of Labor Statistics). Figure 3A24 shows that results are highly similar.

Figure 3A21 – Three alternative measures for knowledge complexity of technologies (NK, date established, and average number of inventors).

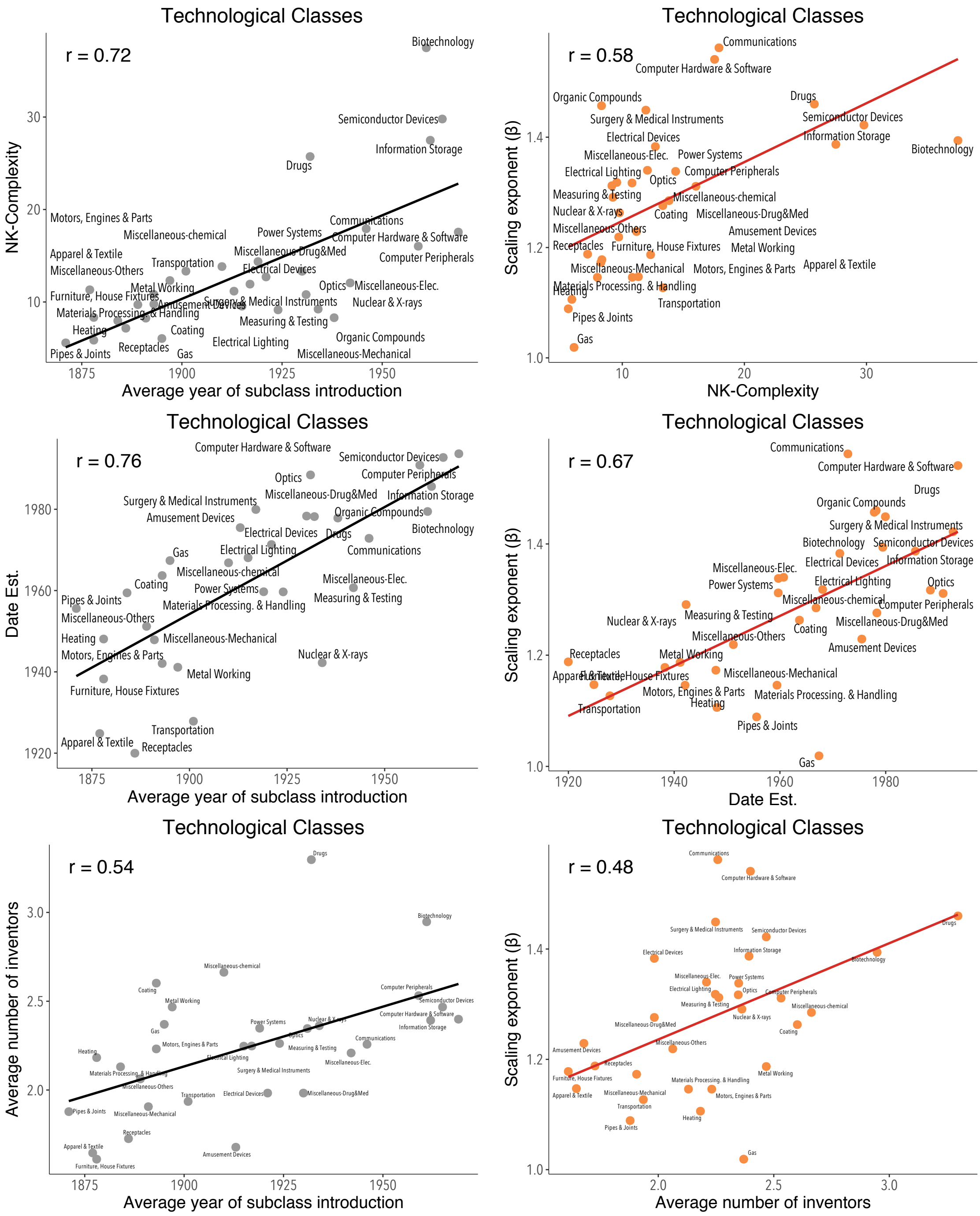


Table 3A4 – Relation between concentration of patenting activity and knowledge complexity.

	Dependent variable:				
	Economic Output Number of patents				
	(1)	(2)	(3)	(4)	(5)
Area (log)	−0.186***	−0.186***	−0.186***	−0.186***	
	(0.006)	(0.006)	(0.006)	(0.006)	
Population (log)	1.097***	1.097***	1.097***	1.097***	
	(0.005)	(0.005)	(0.005)	(0.005)	
Pop.; Year of subclass introduction		0.149***		0.128***	0.128***
		(0.004)		(0.005)	(0.004)
Pop.; NK-Complexity			0.112***	0.035***	0.035***
			(0.004)	(0.005)	(0.004)
Constant	−8.310***	−6.802***	−7.899***	−6.888***	1.467***
	(0.127)	(0.132)	(0.127)	(0.132)	(0.084)
Activity f.e.	yes	yes	yes	yes	yes
MSA f.e.	no	no	no	no	yes
Observations	47,302	47,302	47,302	47,302	47,302
R^2	0.599	0.610	0.605	0.611	0.747
Adjusted R^2	0.598	0.609	0.604	0.609	0.745
F Statistic	522.088***	542.913***	532.034***	539.816***	284.192***

*Notes: Knowledge complexity measures are standardised; also *p<0.1; **p<0.05; ***p<0.01*

Figure 3A22 – Alternative measure for knowledge complexity of scientific fields.

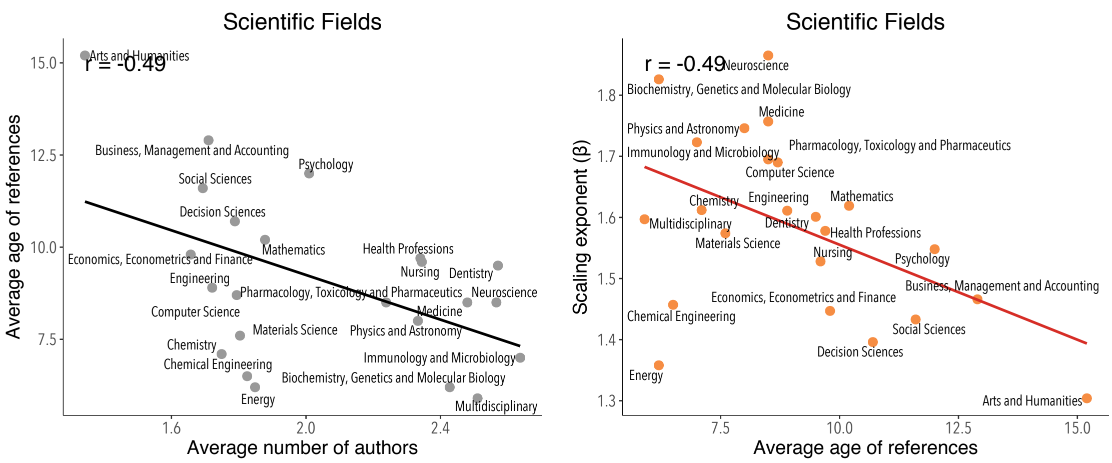


FIGURE 3A23 – ALTERNATIVE MEASURE FOR KNOWLEDGE COMPLEXITY OF INDUSTRIES.

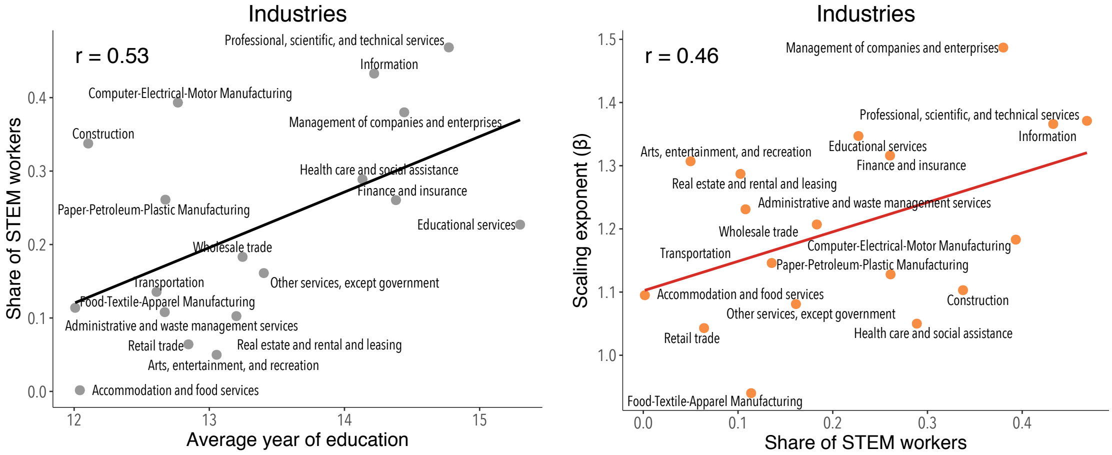
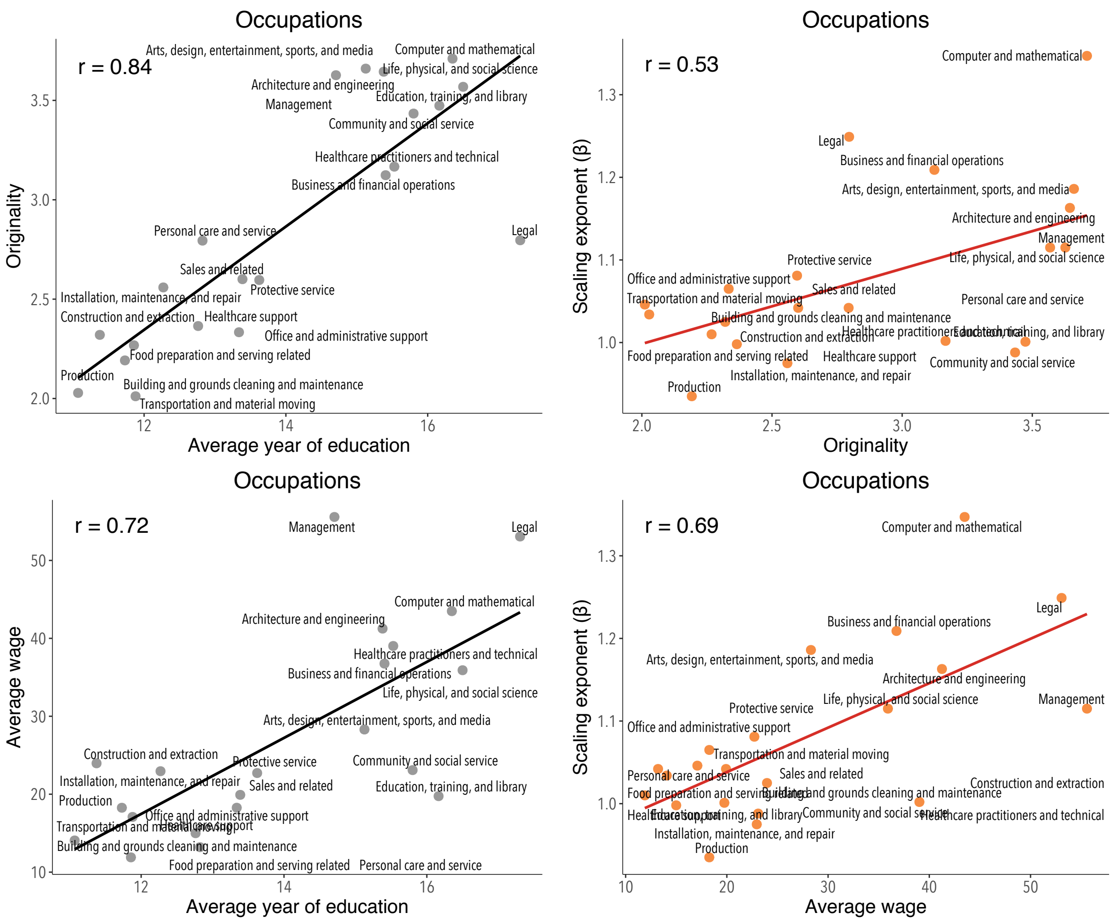


FIGURE 3A24 – TWO ALTERNATIVE MEASURES FOR KNOWLEDGE COMPLEXITY OF OCCUPATIONS.



Alternative measures of concentration

This subsection tests the robustness of our results to changes in the way we define concentration across geographical units. The following figures show, on their left panels, the relationship between our baseline concentration measure (the scaling exponent) and the alternative one (Hoover Gini coefficient, as computed using the EconGeo package Balland and Rigby (2017)). The right panels show the relationship between the alternative measure of concentration and our baseline knowledge complexity measures across economic activities. In all cases results do not vary significantly.

FIGURE 3A25 – CORRELATION BETWEEN HOOVER GINI AND THE SCALING EXPONENT, AND BETWEEN HOOVER GINI AND KNOWLEDGE COMPLEXITY FOR PATENTS.

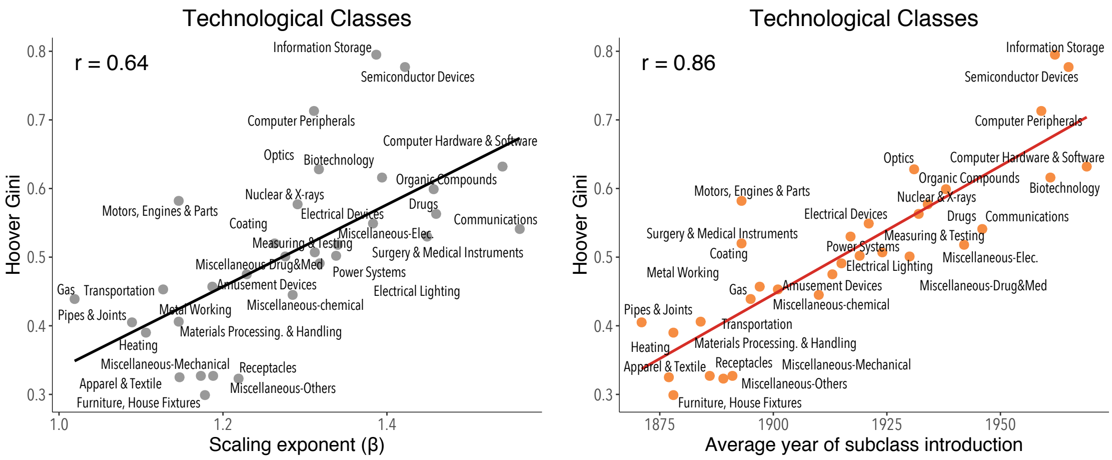


FIGURE 3A26 – CORRELATION BETWEEN HOOVER GINI AND THE SCALING EXPONENT, AND BETWEEN HOOVER GINI AND KNOWLEDGE COMPLEXITY FOR PAPERS.

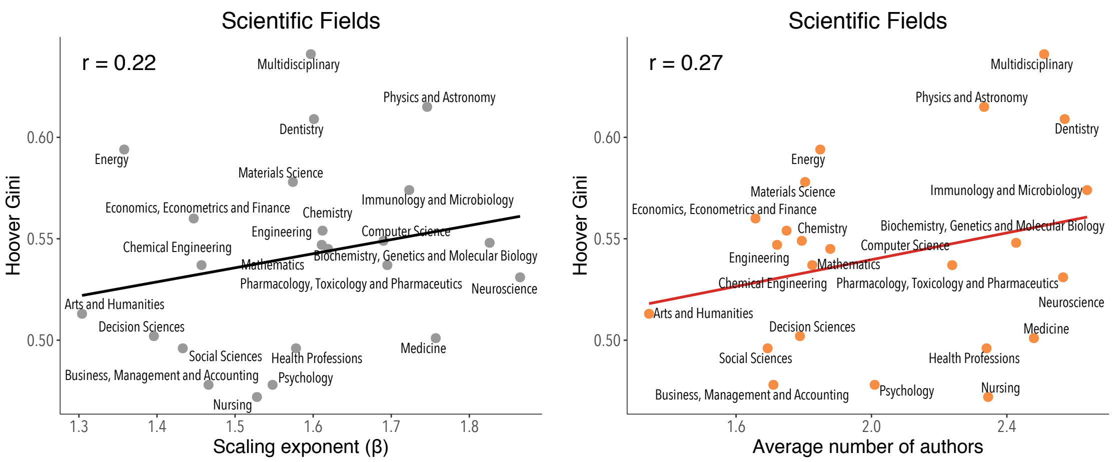


Figure 3A27 – Correlation between Hoover Gini and the scaling exponent, and between Hoover Gini and knowledge complexity for industries.

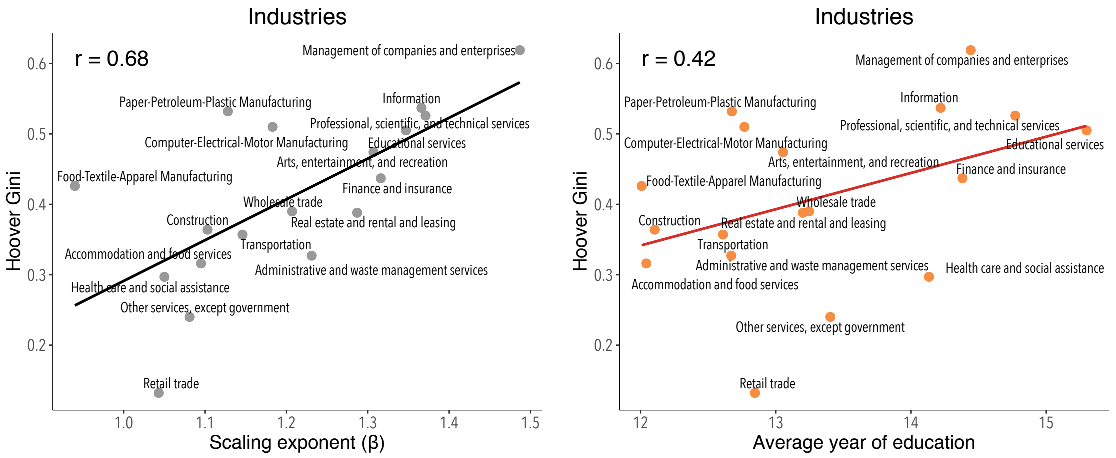
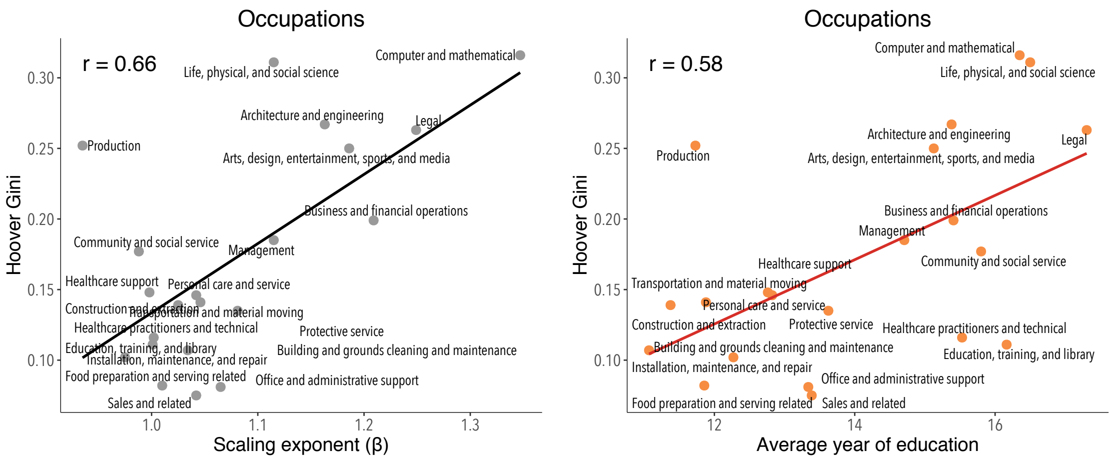


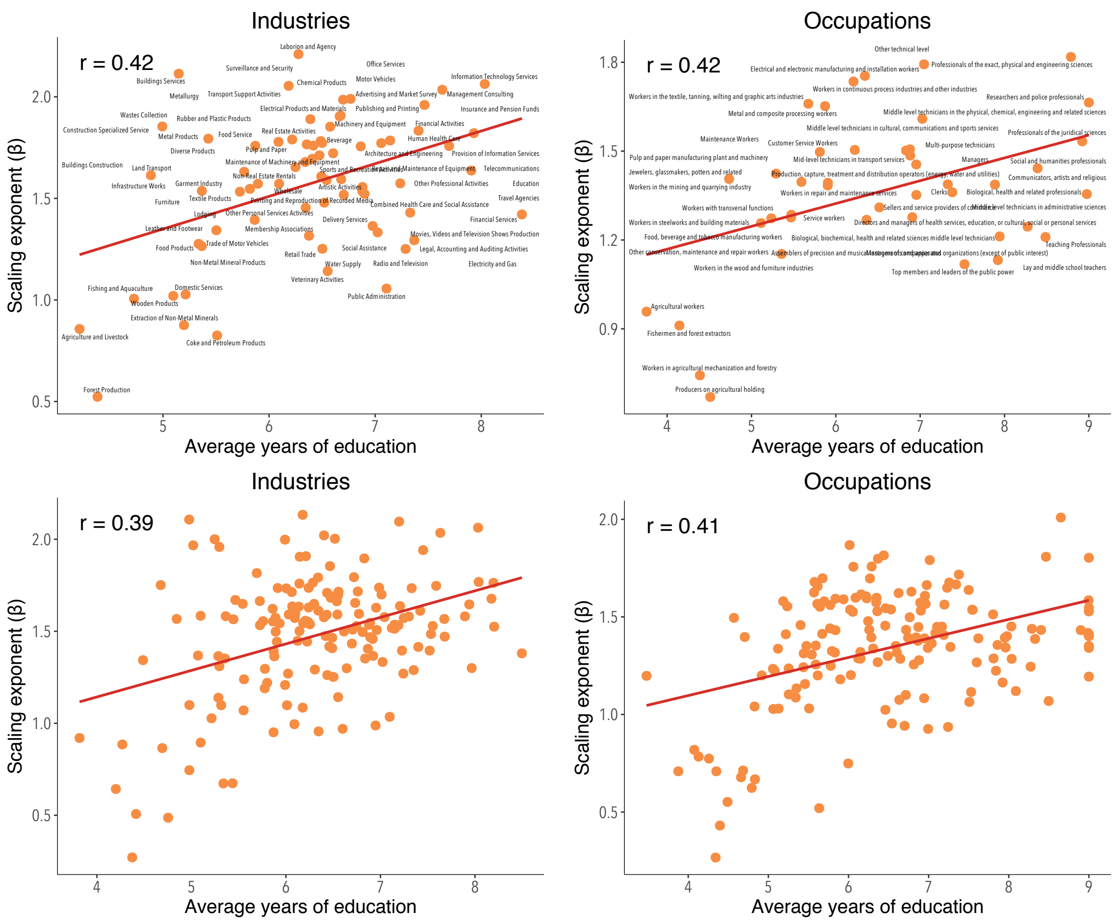
Figure 3A28 – Correlation between Hoover Gini and the scaling exponent, and between Hoover Gini and knowledge complexity for occupations.



Knowledge complexity and concentration for Brazilian industries and occupations

This subsection explores the relation between knowledge complexity and concentration for industries and occupations in Brazil, using the RAIS dataset. We use the mesoregion level of aggregation. For more information visit `http://legacy.dataviva.info/en/`Dataviva.

Figure 3A29 – Relation between knowledge complexity (average years of education) and concentration (scaling exponent) for Brazilian mesoregions, at different levels of aggregation.

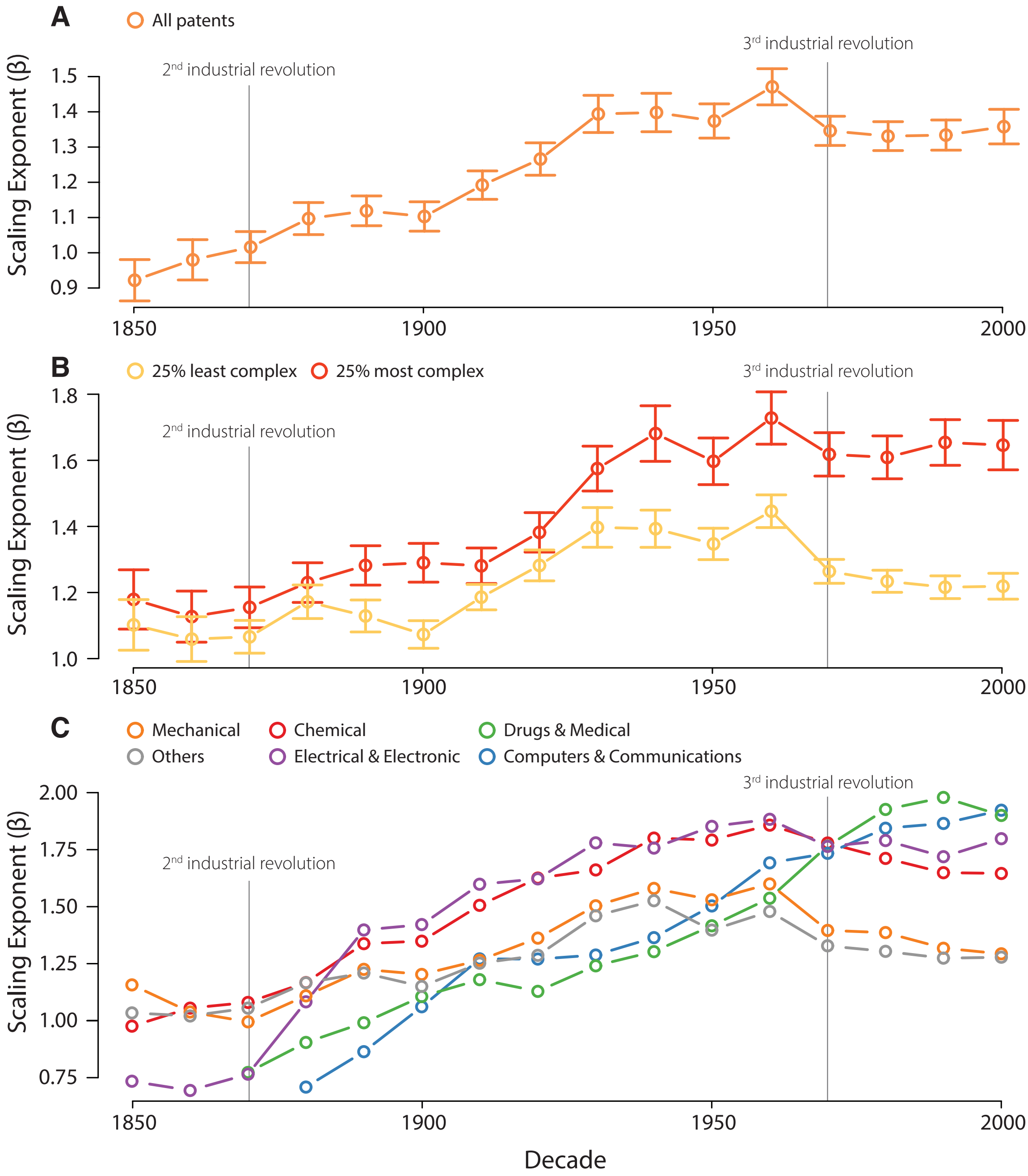


Robustness checks - Historical scaling of patenting activity

Computing scaling exponent using number of claims instead of number of patents

This subsection tests the robustness of our results related to historical trends in the evolution of the urban concentration of technologies. In the main analysis, we estimated scaling exponents based on the number of patents in a given city, here we estimate scaling exponents based on the number of patent claim (a patent can make multiple claims).

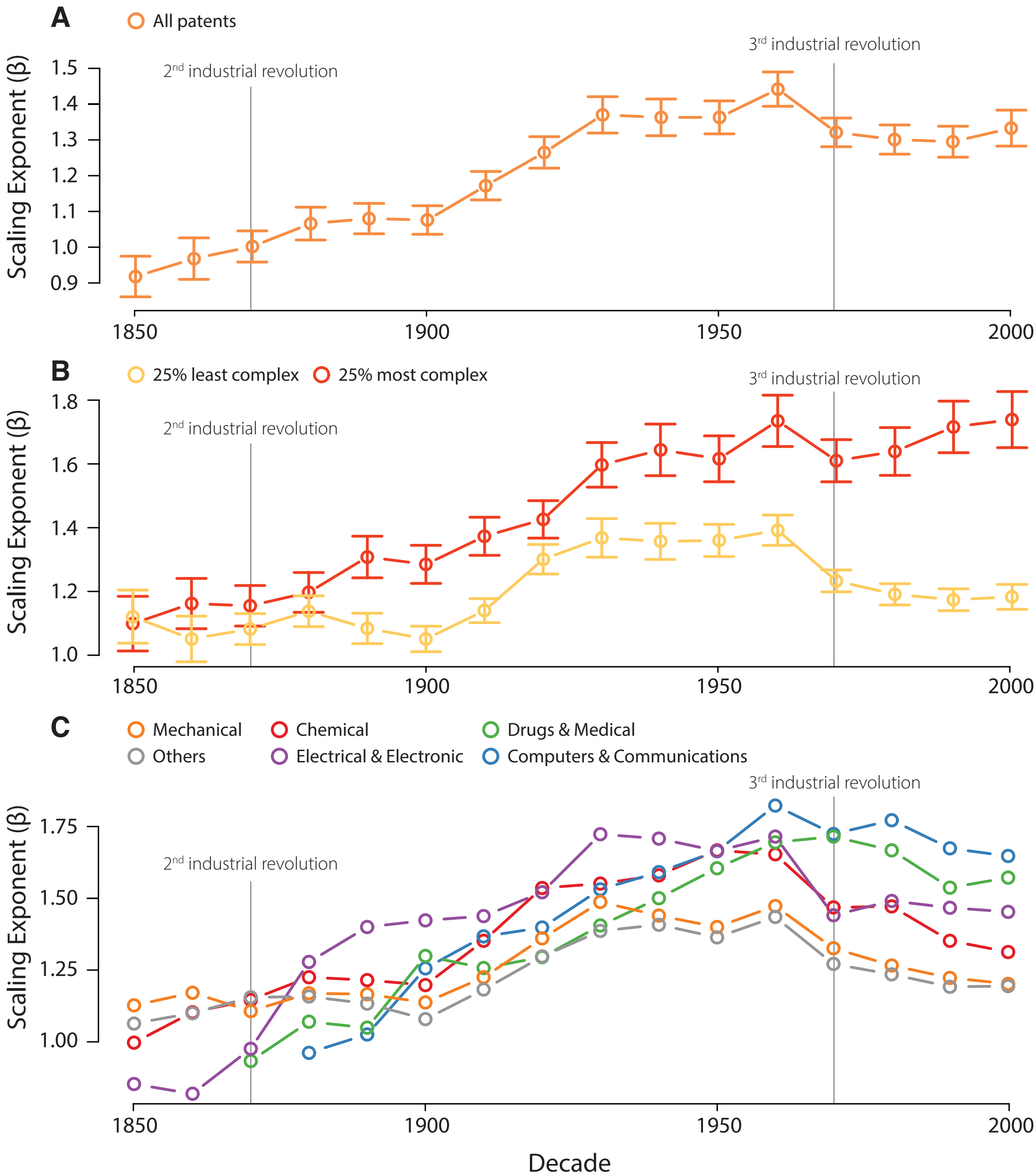
Figure 3A30 – Historical scaling using number of claims to calculate the scaling coefficient.



Alternative pool of cities

This subsection tests the robustness of our results related to historical trends in the evolution of the urban concentration of technologies by changing the cities considered in the analysis. The following figure replicates the results of Figures 3.20 and 3.21 presented in the main analysis, but uses the 353 cities for which we have data on occupations, industries, and scientific fields. Results do not vary significantly.

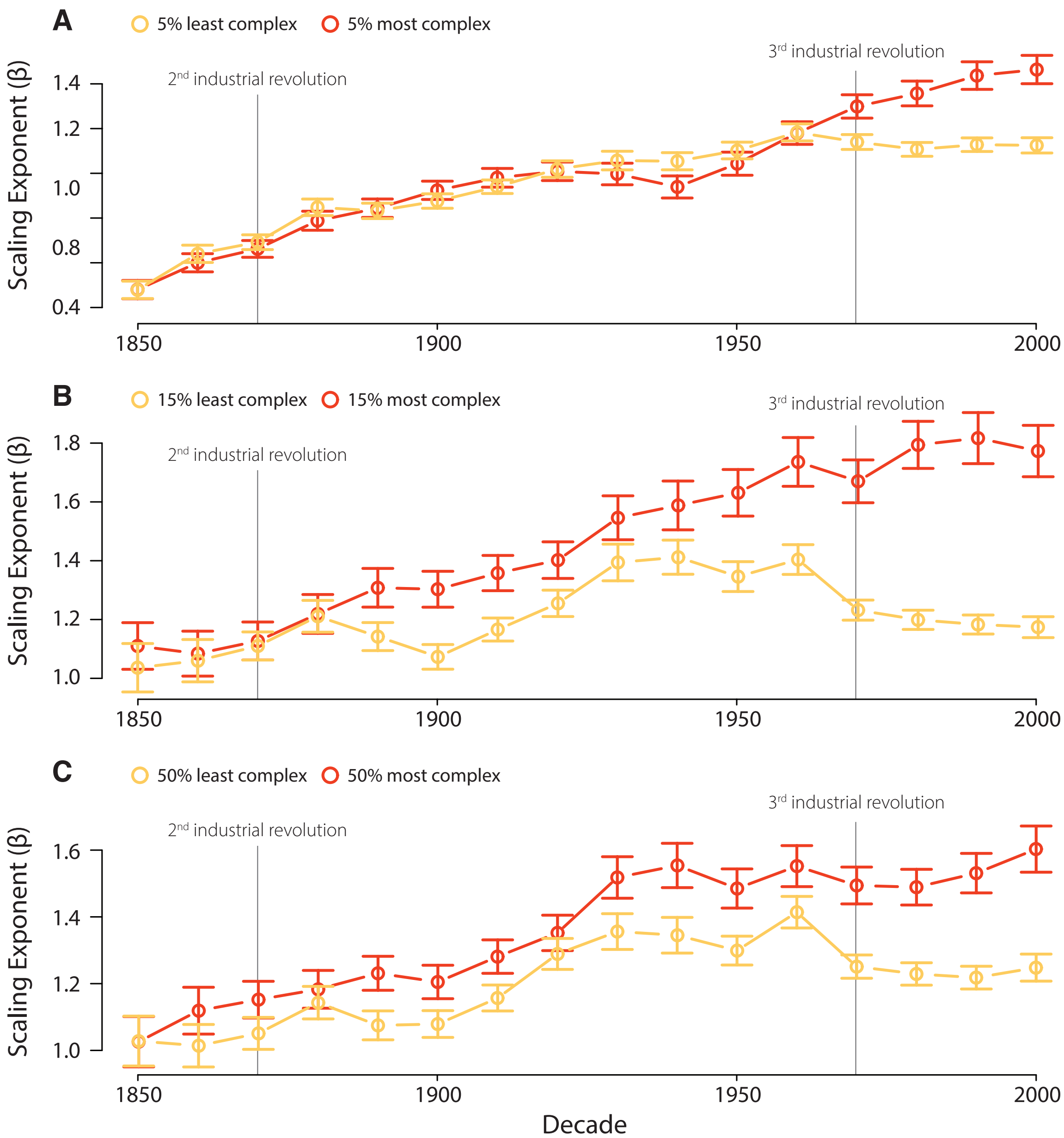
Figure 3A31 – Historical scaling using the same 353 cities as in the main analysis.



Different knowledge complexity thresholds

This subsection tests the robustness of our results related to historical trends in the evolution of the urban concentration of technologies by changing the knowledge complexity thresholds. In this Chapter, we define the most (least) complex patents as the top (bottom) 25% most (least) complex ones. Here we present results using the 5, 15, and 50% threshold instead. Results do not vary significantly.

Figure 3A32 – Changes in scaling for the most complex and least complex patents using different thresholds.



Chapter 4

Technological diversification of U.S. cities during the great historical crises

Abstract – Regional resilience is high on the scientific and policy agenda. An essential feature of resilience is diversifying into new activities but little is known about whether major economic crises accelerate or decelerate regional diversification. this Chapter offers systematic evidence on the effects of three of the largest crises in U.S. history, the Long Depression (1873-1879), the Great Depression (1929-1934), and the Oil Crisis (1973-1975), on the development of new technological capabilities within U.S. metropolitan areas. We find that crises reduce the pace of diversification in cities but they also narrow the scope of diversification to more closely related activities, in particular during the Great Depression. We also find that more diverse cities diversify more strongly during times of crisis.

This chapter is co-authored with Pierre-Alexandre Balland, Ron Boschma, and David Rigby. It is currently in the third round of revision at Journal of Economic Geography.

4.1 Introduction

The shock of the current covid-19 crisis, like the financial crisis of 2007-2008, has global economic consequences but is characterised by strong intra-country disparities in vulnerability (Martin, 2012; Odendahl and Springford, 2020). As a result, questions on how to prevent regions from entering crises and how to alleviate the impacts of crises on regions have once more returned to prominence on the research agenda. However, despite the wide interest, the literature on regional resilience is still largely considered as work in progress (Boschma, 2015).

A crucial component of regional resilience is the ability of regions to diversify into new activities (Pike et al., 2010; Boschma, 2015; Xiao et al., 2018; Rocchetta et al., 2022). When regions are hit by a shock, it may be crucial to develop new growth industries to speed up the recovery process in regions during times of crisis. Several case studies (Grabher, 1993; Glaeser, 2005) indeed suggest that diversifying into new activities may alleviate crises. However, little is actually known on how much diversification occurs in crises relative to periods of regular economic activity. Theories inspired by Schumpeter have expressed divergent views on this issue (Filippetti and Archibugi, 2011): some scholars claim major crises trigger technological breakthroughs (Schumpeter, 1939; Kleinknecht, 1987), while others suggest that dramatic drops in demand prevent the introduction of new (major) technologies during unsettled times (Schmookler, 1966; Scherer, 1982). Which of these theories prevails at the regional level remains unclear.

Previously, empirical evidence on these questions relied primarily on case studies. Work of Hidalgo et al. (2007); Kogler et al. (2013); Boschma et al. (2015); Balland et al. (2015); Rigby (2015), among others, made it possible to quantify such a qualitative phenomenon as the relatedness between technologies, opening up the way for more systematic analyses. Advances in data availability complement this development. The HISTPAT U.S. patent data set (Petralia et al., 2016) reaching back to 1836, allows us to examine some of the deepest crises the United States has experienced. We focus on patterns of technological diversification within Metropolitan Statistical Areas (MSAs) during three of the most devastating economic shocks in U.S. history: the Long Depression, the Great Depression and the Oil Crisis.[1]. These historical crises coincide with moments of great technological change, notably two industrial revolutions (Boschma, 1999) as the neo-Schumpeterian long wave theory demonstrated (Freeman et al., 1982; Kleinknecht, 1990). Adapting to new major technologies through diversification is likely essential to prevent lock-in and assure long-term

[1]The current covid-19 crisis and the financial crisis (2007-2008) are too recent to be included in the analysis as we insist that for successful diversification, new technologies should persist in a region for a certain time period, as further explained in Section 4.3.

regional prosperity (Marshall, 1987; Glaeser, 2005).

The analysis yields the following insights. First, we find that U.S. cities diversify less during major crises. Second, in periods of crisis, cities diversify more in closely related activities than during periods of prosperity, especially during the Great Depression.

Additionally, we find that diversification during the Oil Crisis is very distinct from this pattern in the sense that diversification is less impacted, in particular in upcoming technologies, such as in computers & communication. Furthermore, we find that more diverse cities have a higher probability of diversifying during crises than do specialised cities.

The structure of the Chapter is as follows. In Section 4.2, we discuss recent theorizing on regional resilience and diversification, and how that is related to periods of crisis and technological change. Based on these theoretical considerations, we derive two hypotheses on diversification in times of crisis. In Section 4.3, we explain the data and the methodology used. In Section 4.4, we present the main empirical findings. Section 4.5 of the Chapter will conclude and discuss the findings in light of a future research agenda.

4.2 Resilience of Regions and Diversification in Times of Crisis

In recent years, studies have investigated the ability of regions to bounce back after a crisis (Martin, 2012; Balland et al., 2015; Dijkstra et al., 2015; Diodato and Weterings, 2015; Cuadrado-Roura et al., 2016; Crescenzi et al., 2016; Sedita et al., 2016; Bristow and Healy, 2018; Fratesi and Perucca, 2018; Rocchetta and Mina, 2019). The regional resilience literature is fundamentally interested in the capacity of regions to recover from a shock, and what processes drive that recovery. Many resilience studies follow an equilibrium approach, *i.e.* looking at the ability of regions to return to a pre-existing equilibrium state after a shock or to move into a new equilibrium state (Fingleton et al., 2012). These studies tend to overlook the fact that a substantial part of the recovery process may depend on the ability of regions to develop new growing activities that offset processes of decline (Boschma, 2015; Balland et al., 2019; Rocchetta et al., 2022). As such, tackling the question of regional resilience requires an understanding of how regions diversify into new activities.

A large empirical literature on diversification suggests that regions do not start from scratch when diversifying: they tend to build on existing local capabilities, a process that has been labelled related diversification (Neffke et al., 2011a; Boschma et al., 2015; Rigby, 2015; Hidalgo et al., 2018). This is not to say that unrelated diversification (i.e. the successful development of new activities unrelated to local activities) does not occur in regions, but the evidence shows it is a rare phenomenon(Hidalgo et al., 2007; Neffke et al., 2018; Pinheiro et al., 2021). However, diversification during crises has not been considered yet.

Inspired by scholars who advocate an evolutionary approach to regional resilience (e.g. Christopherson et al., 2010; Pike et al., 2010; Simmie and Martin, 2010; Martin and Sunley, 2015; Webber et al., 2018; Cainelli et al., 2018), Boschma (2015) proposed connecting the literature on regional diversification to regional resilience. He links resilience to the ability of regions to diversify and create new growth paths, to offset stagnation and decline during shocks. This implies that, besides looking at the vulnerability of regions to a shock (conventionally measured as a decline in output levels) and the ability to recover from a shock (conventionally measured as a return to previous output levels, or to new equilibrium output levels), there is a need to examine to what extent shocks impact the ability of regions to diversify (Xiao et al., 2018; Rocchetta et al., 2022).

Diversification is considered to be crucial for regions to overcome a crisis, Rigby et al. (2022) has shown that technological diversification in regions is in general associated with economic growth. This is even more relevant in major crisis periods that are often associated with moments of great technological change (Boschma, 1999) that have pervasive economic consequences in the long run (Freeman et al., 1982; Marshall, 1987; Kleinknecht, 1990).

When regions are confronted with such major crisis periods and shifts in technological paradigms, the need for diversification is even more urgent. This is echoed in certain case studies. Glaeser (2005), for example, describes how Boston reinvents itself by developing new leading industries when others fade. In contrast, Detroit does not manage to develop new industries when their dominance in car-producing technology fades (Hill et al., 2012).

The differences in economic prosperity between Boston, which reinvents itself after the computer revolution in the 1970s, and Detroit, which did not, shows the relevance of understanding (technological) diversification during crises. But although such case studies give some suggestions on the diversification that takes place during crises and the extent to which new specialisations are related to previous activities, more systematic evidence is missing to show how generalizable and precise these patterns are, which is the goal of this Chapter.

This topic has not received a lot of attention in the regional resilience literature. However, a related debate has been taking place in the long wave literature for many years. Innovation theories, inspired by Schumpeter, that developed in the 1980s (Dosi et al., 1988) viewed radical innovations as clustering in waves rather than occurring randomly over time. Schumpeter referred to this as the 'swarming of innovations' which he believed happened during the downswing period of the long wave. In his work on basic innovation, Mensch (1975) developed the depression trigger hypothesis to explain the tendency for radical innovations to bunch during periods of crisis. This hypothesis was challenged by other scholars (Clark et al., 1981; Duijn, 1983) who argued that most innovations take place just after the crisis, during the upswing of a long wave. Kleinknecht (1981) reconciled both views, stating that "*the argument that depression is acting as a trigger for major innovations .... does not exclude the existence of a swarm of related innovations which accompany the diffusion of newly introduced products*" (p.295).

Kleinknecht (1981, 1987) supported the depression trigger hypothesis, claiming that in periods of crisis, demand drops dramatically and returns on further improvements of mature products and technologies are low, and therefore the relative risk of introducing radical innovations for firms decreases. This incentive becomes even stronger when productive resources are set free during the downswing of the economy, leading to declining wages and lower capital costs, which makes it more attractive to invest (Krugman, 1993; Glaeser, 2005). Moreover, many innovative breakthroughs are technologically related to each other, showing interdependencies and complementarities (Rosenberg, 1982; Carlsson and Stankiewicz, 1991) which makes them cluster in time (Rosenberg and Frischtak, 1983; Boschma, 1999). And once radical innovations are introduced, they will attract new investments that will lead to a large stream of additional innovations, known as the 'bandwagon effect' (Clark et al., 1981).

Diametrically opposing this depression trigger hypothesis is the 'demand-pull' hypothesis suggesting that dramatic drops in demand during crises prevent the introduction of new (major) technologies (Schmookler, 1966; Freeman et al., 1982; Scherer, 1982). Freeman et al. (1982) argued that R&D activity is reduced considerably in long wave depressions. Instead, the rise in demand during the upswing provides more favourable conditions for firms to introduce breakthroughs and major innovations (Geroski and Walters, 1995). Schmookler (1966) claimed that upswings in inventive activity followed upswings in demand (Coombs et al., 1987). Moreover, depression phases are characterised by a mismatch between major technologies and institutions (Perez, 1983; Dosi, 1984): the successful introduction and diffusion of major breakthroughs in the economic system requires a set of new institutions that take a long time to develop (Freeman and Perez, 1988). The demand-pull claims suggest that new major technologies are more likely to enter the economy in the growth phase of the long wave.

The agents introducing new technologies operate strongly in regional settings. Studies have shown that entrepreneurs and firms are very much influenced by local capabilities when diversifying and innovating (Klepper (2007); Neffke et al. (2018); Lo Turco and Maggioni (2016) and Hazir et al. (2019)). Agents faced by a drop in demand can opt to innovate in other technologies and possibly more specifically into technologies that are new to the region or postpone diversification until demand rises again. Reformulating the neo-Schumpeterian ideas into the framework of the regional diversification literature, we could expect regions to introduce and develop new activities during downswings as much as during upswings.

Therefore, we develop a set of competing hypotheses on how regions adopt technologies new to them. Hypothesis 1a builds on the depression trigger hypothesis, stressing that diversification is more likely to occur during periods of crisis. As stated above, economic agents might be more willing to take risks and to try out something new when current products and technologies show decreasing returns. Institutional agents (like regional governments) may see major enduring crises as windows of opportunity and are therefore more prone to promote new ways of getting out of the crisis. By contrast, Hypothesis 1b builds on the demand-pull hypothesis and states that diversification is even more unlikely to take place in regions during periods of crisis. Inventions have to wait until upswings in demand arise. We, therefore, formulate the following two competing hypotheses:

Hypothesis 1 (a). *cities diversify more during crises than during non-crisis periods*

Hypothesis 1 (b). *cities diversify less during crises than during non-crisis periods*

Furthermore,the contrasting Schumpeterian views on adopting new major technologies also yield different expectations on the level of unrelated diversification during crises. On the one hand, the depression trigger hypothesis suggests that unrelated diversification is more likely as returns on related diversification have decreased. On the other hand, the demand-pull hypothesis suggests that related diversification is more likely as unrelated diversification would just add to the high uncertainty that is already inherent to a crisis period. . Furthermore, unrelated diversification is generally more costly because underlying capabilities need to be transformed completely (Neffke et al., 2018). Therefore, we formulate the following hypotheses:

Hypothesis 2 (a). *cities diversify more in less related technologies during crises than non-crisis periods*

Hypothesis 2 (b). *cities diversify more in related technologies during crises than non-crisis periods*

4.3 Data and Methodology

The hypotheses outlined above are tested with a unique dataset of U.S. patents covering the period $1836 - 2002$. This long time span allows us to test the hypotheses across three major crises in U.S. history: the Long Depression, the Great Depression and the Oil Crisis. Although we are aware of the limitations of patent records (Griliches, 1981), patent records hold a wealth of information regarding the process of invention and the nature of additions to the expanding stock of knowledge. The patent data originates from the efforts of Petralia et al. (2016) to obtain geographical locations for all patents over the period $1836 - 1974$ from Google scans of historical U.S. patents. Information on patents since 1974 is available from NBER patent data (Hall et al., 2001).

Diversification in a MSA is captured by the development of a comparative advantage in a new technology within that MSA.[1] MSAs are defined by the U.S. Census Bureau

[1] We note that if a region diversifies in activities where patenting is uncommon this will not be captured by our methodology.

as core areas with at least 50,000 inhabitants and adjacent countries that are socially and economically integrated, as measured through commuting patterns. As these commuting patterns do not necessarily hold for the same areas 100 years ago, we also run robustness checks using county-level data. We use the Metropolitan Core-Based Statistical Areas delimitations of 2013 as the definition of MSAs. As agents interact within this area it is also indicative of the technological knowledge that is well known to agents and the area in which resources set free during crises find new purposes, see Boschma (2015) and Rigby (2015). We remain agnostic on which agents within the area are involved in the development of new technological specialisations, this topic is studied elsewhere, see Neffke et al. (2018). We do introduce measures below that capture the linkages between MSAs through inventors, as innovation does not necessarily take place in isolation and global pipelines matter (Bathelt et al., 2004; van der Wouden, 2018).

Technologies are represented by the 438 different primary classes of the United States Patent and Trademark Office (USPTO) patent classification system.[2] When diversification occurs, the data allows us to calculate the relatedness of the new technology to the technologies present in the MSA in the previous time period.

We restrict our sample to MSAs within the contiguous U.S.A. We also impose a minimum of 10 patents per year for a time period of a MSA to be taken into account and a minimum of 0.5 patents[3] per year in a certain primary technology class. As a result, data is drawn from a sample of 274 MSAs and $2,171$ MSA-time periods. Below, we introduce our definitions and measurements of crises, diversification, relatedness and diversity.

4.3.1 Crises

Like Balland et al. (2015), we build on trends in patenting per region to indicate when regions are in crisis, as patent counts are highly correlated with economic performance (Glaeser and Gyourko, 2007; Rothwell et al., 2013). To ascertain this link with economic performance, we focus on the great historical crises of the United States, identified independently of the patent data, while using patent counts to indicate the breadth and depth of these crises per MSA.

Each nationwide crisis is regarded as a shock at the regional level. A metropolitan area can then either enter into a crisis or not. At the regional level, the emergence and the

[2] Primary technology classes are comparable over time as the USPTO reclassifies all patents when new class definitions are introduced.

[3] Patents that are assigned to inventors in multiple MSAs, only count as 1 divided by the number of MSAs on that patent for each of the MSAs.

duration of crises are identified from patent records using an adapted version of the business cycle algorithm of Harding and Pagan (2002), after Balland et al. (2015). We follow the definition of technological crises by Balland et al. (2015, p.6) as sustained periods of negative growth in patent activity: "*more formally, a time series recording yearly patenting activity can be defined as a continuum of local maxima (peaks) and minima (troughs) that divide the series into periods of technological growth from trough to peak and technological crisis from peak to trough.*".

The algorithm to detect business cycles "*identifies potential turning points as the local minima (trough) and maxima (peak) in the series. Let P_t be a patent count yearly series. A trough is identified as $(p_{(t-j)}, \ldots, p_{(t-1)}) > p_t^{trough} < (p_{(t+j)}, \ldots, p_{(t+1)})$ while a peak follows the condition that $(p_{(t-j)}, \ldots, p_{(t-1)}) < p_t^{peak} > (p_{(t+j)}, \ldots, p_{(t+1)})$.*" (Balland et al., 2015, p.172). To prevent "noise" due to years of random growth or decline, two extra conditions are imposed: "*The phases (technological growth or technological crisis) should be at least 2 years long, while complete cycles (period between 2 peaks or between 2 troughs) should be at least 5 years long.*"

As a result of this procedure, time periods are defined separately for each MSA and therefore do not necessarily match or have the same duration. For each MSA, all periods of crisis and growth are identified between 1836 and 2002. Crises that do not overlap with one of the three great U.S. crises or have a decrease in patenting activity of less than 35 percent during the crisis are ignored.[4]

The decision to ignore regional downturns in patenting that do not occur during a nationwide shock decreases the risk of including local crises that are unrelated to major economic downturns or are spurious decreases in patent counts. Regional periods of growth are kept regardless of when they occur. We give further detail on the dynamics of regional patenting during the great historical crises in Appendix 4A.

4.3.2 Diversification

We use the notion of Revealed Comparative Advantage (RCA) (see Hidalgo et al., 2007) to identify in which technologies each MSA is specialised across the time periods examined. In equation 4.1, x represents the number of patents, c denotes the city-region (MSA), i is the primary technology class, and t indicates the time period. RCA values are bounded on the left by zero. A RCA value of 1 indicates that a MSA has the same share of patenting activity in a particular technology class as the national average. RCA values of 1 or greater indicate regional specialisation in a technology. A technology enters the technological portfolio of a MSA when a MSA develops a specialisation in a technology class that it did not have in the previous time period.

[4]In the robustness check, we show that using different thresholds lead to similar results.

An entry is considered a diversification of the region. To account for spurious entries of technologies we add the condition that an entering technology has to remain present in the portfolio of a MSA (with $RCA => 1$) for at least two time periods.

$$RCA_{cit} = \frac{\frac{x_{cit}}{\sum_{i=1}^{I} x_{cit}}}{\frac{\sum_{c=1}^{C} x_{cit}}{\sum_{c=1}^{C}\sum_{i=1}^{I} x_{cit}}}, \tag{4.1}$$

4.3.3 Relatedness

Technologies that are not in the technological portfolio of a MSA in time period $t-1$ (those for which the RCA value is below one) enter or do not enter in time period t. An important predictor of the entry of a technology within a MSA is how closely related it is to technologies that are already present in the region (Boschma et al., 2015; Balland et al., 2019). This notion of relatedness is essential for hypothesis 2, where we focus on less related diversification. The co-occurrence of technology classes on patents is used to measure the relatedness between *technologies*. Technology classes are more related to one another as they co-occur with a frequency that is greater than that which would be predicted based on the overall counts of classes in the population of patents of a given time period. The formula for relatedness, outlined Additional Chapter A, is reported in equation 4.2. Where C_{ijt} is the number of co-occurrences between technology i and technology j in time period t. S_{it} and S_{jt} is the number of co-occurrences involving respectively technology i and technology j in time period t, N is the total number of technologies, and m is the total number of co-occurrences.

$$TR_{ijt} = \frac{C_{ijt}}{\left(\frac{S_{it}}{\sum_{n=1}^{N} S_n}\frac{S_{jt}}{(\sum_{n=1}^{N} S_n)-S_{it}} + \frac{S_{jt}}{\sum_{n=1}^{N} S_n}\frac{S_{it}}{(\sum_{n=1}^{N} S_n)-S_{jt}}\right)m}, \tag{4.2}$$

Building on relatedness, relatedness density (see Hidalgo et al., 2007) gives the relatedness of a *region* to a *technology* that is not yet present in its technological portfolio. Relatedness density is equal to the sum of relatedness values of the technologies in the region to the possibly entering technology divided over the sum of relatedness values of all technologies to this technology, as can be seen in equation 4.3.

$$Rel.density_{cit} = \frac{\sum_{j\in c, j\neq i} TR_{ijt}}{\sum_{j\neq i} TR_{ijt}}, \tag{4.3}$$

4.3.4 Control variables

Presence of technology in neighbouring MSA

Other factors that are correlated with our variables of interest may influence the development of specialisation in a new technology by a MSA. Having MSAs nearby that have an RCA in a technology can be expected to positively influence the entry of that technology to the technological portfolio of a city because knowledge flows tend to be geographically conditioned (Rigby, 2015; Boschma, 2017). Therefore, we develop a spatial weight matrix using the inverse distance for the presence of technology in neighbouring MSAs.

Population

We also include the average population of MSAs in the time periods based on census data.

Diversity

Some scholars argue that it is the diversity of capabilities in a city that is more important than the size of a city. The regional resilience literature argues that variety is crucial for resilience because it can accommodate sector-specific shocks (Essletzbichler, 2007, 2015; Diodato and Weterings, 2015; Rocchetta et al., 2022). This is in line with numerous case studies on specialised regions that showed structural problems of adjustment (Boschma and Lambooy, 1999; Pike et al., 2010). Specialised regions may have a low capacity to diversify in new activities, because they are cognitively, socially and politically locked-in (Grabher, 1993; Hassink, 2005).

To control for this, we follow Duranton and Puga (2000) who propose a simple diversity index, known as the Relative Diversity Index (RDI). The intuition is that if the relative distribution of patenting activity over technology classes in a MSA resembles the national distribution, then the city is relatively diverse. On the other hand, when the patents of a MSA cluster strongly above the national average in a few classes then it is seen as specialised.

Following Duranton and Puga (2000) the formula is given in equation 4.4, like before x stands for the number of patents, c indicates the MSA, i the respective technology, and t the respective time period. A value close to zero denotes a specialised city, whereas the larger the value the more diverse a city is.

$$RDI_{ct} = \frac{1}{\sum_{i=1}^{I} \left| \frac{x_{cit}}{x_{ct}} - \frac{x_{it}}{x_t} \right|}, \tag{4.4}$$

Degree centrality
Agents in regions may also have strong connections external to their area that are not fully captured by adding a variable on the presence of a technology in neighbouring cities. For example, multinational corporations are known to be more capable of wielding knowledge from distant areas (Iammarino and McCann, 2013). To somewhat control for the extent to which diversification within and outside of crises may be influenced by these so-called "pipelines" (Bathelt et al., 2004; van der Wouden, 2018), we use the degree centrality of MSAs in the collaboration network of inventors of patents.

Fixed effects
Due to the historical nature of the HISTPAT data, there is much less information available on other confounding variables as in related approaches, such as inventor characteristics. However, this shortfall can largely be mitigated by the inclusion of fixed effects at the level of the time period, technology, and MSA. Table 4.1, gives the descriptive statistics of our variables.

4.3.5 Empirics

Entry models are a common tool in the literature that yield insight on the role of relatedness in diversification (e.g. Boschma et al., 2015; Balland et al., 2019). Despite the popularity, some underestimation of risks exists concerning two particular traits of the econometric specification that prove to be important here and are likely to hold in related applications. Namely, the extreme right skewness of its main variables of interest: entry and relatedness density. This means that often-used linear probability models do not lead to correct estimations and that the coefficient is strongly influenced by outliers. Therefore, we choose to use a logit model and substitute the continuous relatedness density variable for an ordered categorical variable by creating dummy variables for each quantile of relatedness density values. We give further explanation on these reasons in Appendix 4.5.

Equation 4.5 gives our preferred regression formula for Hypotheses 1 and 2 and is in line with previous work like Boschma et al. (2015). If a technology i enters the technological portfolio of city c in time period t, the value of the dependent variable is 1. If it was not in the portfolio of city c and it did not enter its value is 0. The dependent variable is regressed on the relatedness density (RlD) of the technology class to the portfolio of each city in the previous time period, on a dummy variable which indicates if a city is experiencing a crisis (Cris) or not, on the interaction between these first two terms (RlD $\times$ Cris), city characteristics (City) at time t, which consist of the relative diversity index, population and degree centrality, and on the presence

TABLE 4.1 – DESCRIPTIVE STATISTICS

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Entry	724,752	0.031	0.174	0	0	0	1
Crisis	724,752	0.141	0.348	0	0	0	1
Relatedness density	724,752	0.089	0.126	0	0.01	0.1	1
Population	724,752	416,093.600	961,537.100	20,402	100,534	398,751	17,019,060
Present $\times$ **W**	724,752	0.00003	0.00003	0.00000	0.00001	0.00004	0.0003
Diversity	724,752	1.086	0.303	0.255	0.886	1.309	1.833
Degree centrality	724,752	60.002	250.444	0.000	4.750	38.500	8,460.667

(Pr) of technology i in the technological portfolio of neighbouring MSAs multiplied by a spatial weight matrix $\mathbf{W}$, and fixed effects (F.E.) that consist of a city-fixed effect, a technology-fixed effect, and a time-fixed effect.

$$Entry_{cit} = \sum_{k=1}^{5} \alpha_k RlD_{cit-1,k} + \beta Cris_{ct} + \sum_{k=1}^{5} \gamma_k RlD_{cit-1,k} \times Cris_{ct} + \delta City_{ct} + \eta Pr_{it} \times \mathbf{W} + F.E. + \epsilon_{cit}, \tag{4.5}$$

To facilitate interpretation we standardise the relative diversity index, population, degree centrality, and Present$\times\mathbf{W}$ to have 0 mean and a standard deviation of 1 and use sum-to-zero contrasts for the fixed effects. As such, converting the intercept to probabilities gives the probability of entry at the average of all cities, technologies, and time periods instead of the reference category for each dummy variable.

Based on equation 4.5, we can calculate the marginal effect of crisis per relatedness density group to see if Hypothesis 1 and 2 can be accepted or rejected.[5]

We note that this empirical strategy aims at *describing* how regions diversify during crises and does not allow us to ascertain that crises *cause* these changes.

4.4 Results

4.4.1 Diversification in times of crisis

Table 4.2 gives the marginal effects based on the results for specification 4.5. The full regression results can be consulted in Table 4A2 in the Appendix. The first and second columns give the marginal effects for a specification without fixed effects, and in the case of column (1) also without interaction terms. These results are highly similar to the preferred specification given in column (3). Note that the parentheses give the 95% confidence interval.[6]

The marginal effect of relatedness density values falling between the 20% and 40% lowest values in column (3) indicates that technologies within this category are 0.27% more likely to enter the technological portfolio of a city than those of the reference category with the lowest 20% values *ceteris paribus.* Although small, such an increase

[5] We experienced that most packages in R currently available to calculate marginal effects are rather slow and computational intensive when having many observations and fixed effects as is the case here. Therefore we developed and published a R-package called fastlogitME, which uses less CPU and is compatible with speedglm.

[6] We prefer to give the 5^th and 95^th percentile instead than a to probabilities converted standard error as the response scale (probabilities) is linear whereas the scale of the underlying link function is non-linear. When using a standard error derived from bootstrapping or the delta-method it may result in a confidence interval that exceeds the range of 0% to 100% entry probability, which is technically impossible and the reason in the first place to use logit models instead of linear models.

is not negligible. After all, the development of a new specialisation by a region is a rare event. The probability of entry is namely only 0.63% for technologies that fall in the reference category of the lowest 0%-20% relatedness density values. Therefore, an increase of 0.27% is substantial.[7]

The probability of entry increases with relatedness density as can be seen by the increase in coefficient size up to 0.0293 with each step in relatedness density. This indicates that a region is more likely to develop a specialisation in a technology that is more strongly related to its technological portfolio, as could be expected based on the previous literature (see among others Boschma et al., 2015 and Hidalgo et al., 2018).

The marginal effect of technological diversity, as measured by the Relative Diversity Index, is positive and significant. The effect of entry is substantial and of a similar size of increasing relatedness density from the 0%-20% quantile to the 20%-40% quantile.[8] This is the first systematic evidence corroborating earlier suggestions based on case studies (Grabher, 1993; Boschma and Lambooy, 1999; Hassink, 2005; Pike et al., 2010; Boschma, 2015; Neffke et al., 2018), which claim that there is more to diverse cities that makes them open and interested in developing new activities. In such diverse settings, there is a lower probability that established industries and vested interests that dominate the institutional and policy network can block new key developments. This comes on top of the advantage that diverse cities have by having a larger technological portfolio and therefore increased proximity to possibly entering technologies, which is captured by the relatedness density variable and further discussed by Balland et al. (2015) and Boschma (2015). In Appendix 4.5, we further explore the differences in diversification patterns during crises of more diverse regions versus more specialised regions.

Contrary to expectation, the marginal effect of population size is insignificant and virtually zero and the marginal effect of degree centrality is even significantly negative. However, when the diversity variable is omitted from the regression the population variable is positive and statistically significant and when also the population variable and fixed effects are dropped this also holds for the degree centrality variable. This suggests that the industrial composition, proxied by diversity, is more important for the development of new specialisations than just agglomeration size, proxied by population, or having a central position in inventor networks proxied for degree centrality.

[7] More specifically, an entry probability of 0.63% holds for the lowest relatedness density category outside of crisis for the mean values of population size, degree centrality, and the presence $\times$ spatial weight matrix when.

[8] The reference category for which the marginal effects are calculated are non-crisis periods but results are similar when this is switched to crisis periods.

TABLE 4.2 – MARGINAL EFFECTS (HYPOTHESIS 1 AND 2)

(Dependent variable: entry of technology class i in the technological portfolio of city c at time t)

	Naive specification	*+ crisis× relatedness density*	*+ fixed effects*
	(1)	(2)	(3)
Relatedness density (20%-40%)	0.0033*** (0.0024, 0.0043)	0.0034*** (0.0024, 0.0044)	0.0027*** (0.0019, 0.0035)
Relatedness density (40%-60%)	0.0102*** (0.0088, 0.0116)	0.0105*** (0.0091, 0.0120)	0.0074*** (0.0063, 0.0085)
Relatedness density (60%-80%)	0.0222*** (0.0201, 0.0244)	0.0223*** (0.0201, 0.0245)	0.0150*** (0.0134, 0.0167)
Relatedness density (80%-100%)	0.0428*** (0.0394, 0.0463)	0.0414*** (0.0381, 0.0450)	0.0293*** (0.0266, 0.0322)
Crisis	−0.0023*** (−0.0026, −0.0021)	−0.0053*** (−0.0063, −0.0040)	−0.0032*** (−0.0042, −0.0019)
Diversity	0.0034*** (0.0033, 0.0035)	0.0035*** (0.0034, 0.0036)	0.0023*** (0.0021, 0.0025)
Population	0.0001 (−0.0001, 0.0003)	0.0001 (−0.0001, 0.0002)	0.00004 (−0.0001, 0.0002)
Present×**W**	0.0032*** (0.0032, 0.0033)	0.0033*** (0.0033, 0.0034)	0.0020*** (0.0018, 0.0021)
Degree centrality	−0.0012*** (−0.0014, −0.0010)	−0.0012*** (−0.0014, −0.0010)	−0.0005*** (−0.0007, −0.0004)
Time F.E.	No	No	Yes
Technology F.E.	No	No	Yes
MSA F.E.	No	No	Yes
Observations	724752	724752	724752

Notes: The relatedness density groups and crisis are dummy variables with as reference category, respectively, the 20% lowest relatedness density values and non-crisis time periods; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

The positive marginal effect of $Present \times \mathbf{W}$ indicates that the presence of the technology in nearby cities increases the likelihood that said technology enters the technological portfolio of a region, which is to be expected and in line with (Boschma et al., 2017).

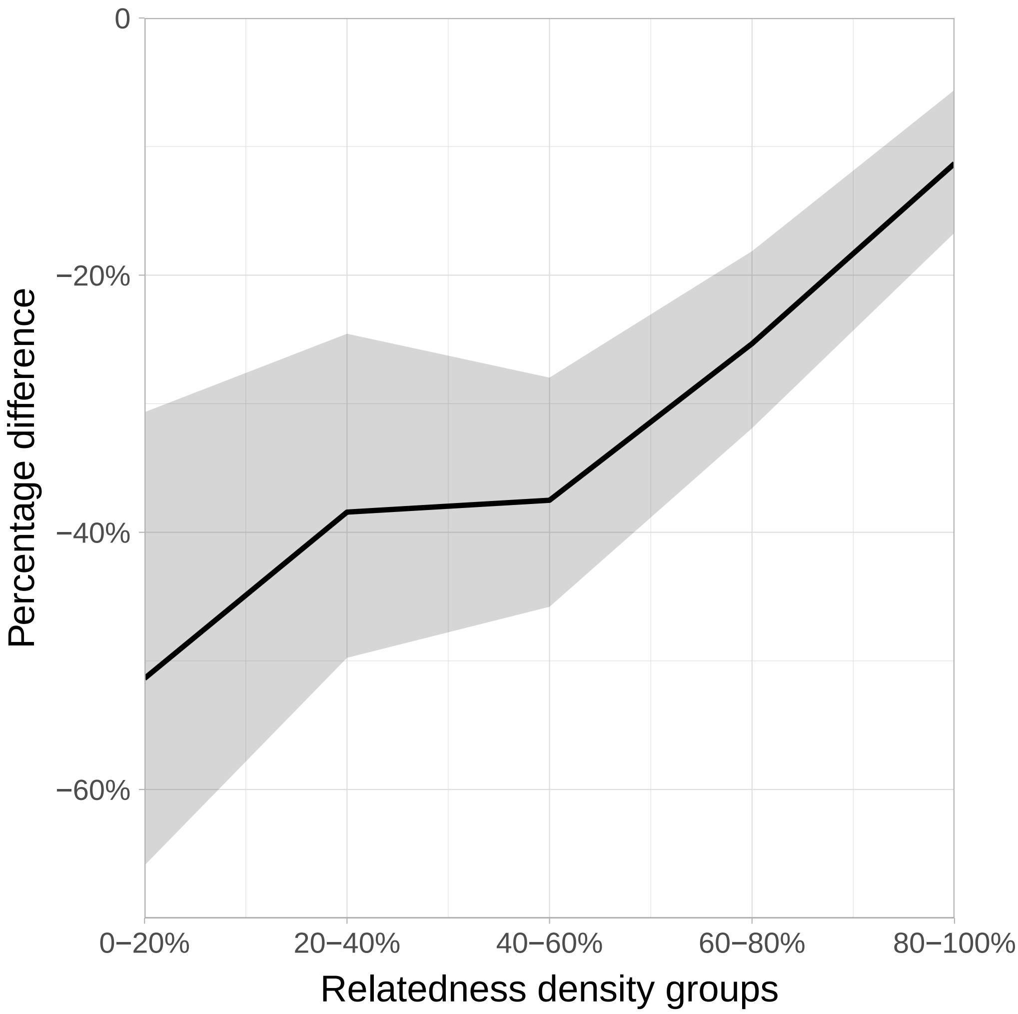
The variable of interest here is crisis, for which the marginal effect is negative and significant, which indicates that on average the probability of entry decreases by 0.32% in crisis for the reference category, according to the full specification in column (3). Once again this effect is sizeable as for this reference category the probability of entry is only 0.63%, which suggests a reduction of on average more than 50%. This suggests that even though diversification is rare it becomes even much rarer during crises.

Cities also diversify less when entering a crisis in the other relatedness density values than those of the reference category of the 20% lowest relatedness density values, as shown in Appendix 4.5. This rejects the depression trigger Hypothesis 1a and confirms the demand-pull Hypothesis 1b suggesting that when a crisis hits agents in cities see their resources to develop new activities diminish and cities end up diversifying less during crises.

For Hypothesis 2 on the changes in relatedness density of entering technologies during crises, we turn to Figure 4.1 below, which gives the relative size of the marginal effect of entering a crisis compared to the average probability of entry outside of a crisis per relatedness density group.

As said the probability of entry of a technology decreases by $0.32/0.63 \approx 51\%$ when entering a crisis for the lowest relatedness density values, whereas, for technologies with relatedness density values in the highest quintile, the entry probability is only about 11% smaller during crises, see Appendix 4.5 for more details. Hence Figure 4.1 confirms Hypothesis 2b: cities diversify more in related technologies during crises. Apparently, in times of high uncertainty, diversification is more likely in technology classes that are more closely related to the knowledge core of the region. This likely reflects the uncertainty of economic agents in terms of future technological development during the highly turbulent phases of major crises.

Figure 4.1 – Percentage difference in probability of entry between crisis and no crisis across quintile groups



Interestingly it seems that when national economies undergo radical technological change, such as the electrical revolution in Figure 4A12, local economies that enter a crisis switch to more conservative, *i.e.* related, diversification. This may give insight into why industrial revolutions shift prosperity from certain cities to other cities, as technological proximity to new technologies is apparently particularly important when a crisis hits. This is in line with the role of local capabilities during the computer revolution and its shift of prosperity between cities documented by Glaeser and Ponzetto (2007) and Berger and Frey (2016). The extent to which this observation holds is an interesting avenue for future research.

4.4.2 Robustness and extensions

The results presented should be interpreted as being descriptive, in the sense that they show to what extent diversification patterns change during crises, but the results do not give certainty that crises *cause* these changes. Nevertheless, to check and expand on the results, we briefly discuss results derived from other specifications in this section, while the full results can be consulted in Appendix 4.5.

We show that similar results are found when the threshold of crisis depth is increased or lowered, in Appendix 4.5; when entries are not defined by the RCA passing the threshold of one but when larger steps are necessary, in Appendix 4.5; when also crisis periods outside of the great historical crises are taken into account, in Appendix 4.5; when looking at the county level instead of the MSA level, in Appendix 4.5; when entering technologies are compared to the previous technological portfolio of a region, instead of the idiosyncratically varying boom-bust-cycle based time periods, in Appendix 4.5; and when crisis and non-crisis periods of cities within the great historical crises are considered within equal time periods while controlling for observables and certain unobservables through fixed effects, in Appendix 4.5. This latter approach approximates a difference-in-difference approach.

We also consider some extensions to the main results. The main results show that diverse cities have a strong advantage in diversifying into new technologies. In Appendix 4.5, we show that diverse cities, nevertheless, do not have a stronger tendency to focus on unrelated diversification when entering a crisis compared to more specialised cities.

We also reproduce the main results in Figure 4.1 for each of the three great historical crises, in Appendix 4.5. These results show that the Long Depression and the Great Depression show a very similar pattern as Figure 4.1, although the Great Depression is significantly deeper, but that the Oil Crisis is very distinct, in the sense that the probability of entry is not much lower during a crisis and unrelated technologies even have a larger probability of entry.

A first possible reason is that the Oil Crisis is not a financial crisis like the other two but rather based on fossil energy-intensive inputs, which may mean there is sufficient incentive and funding to promote diversification.

A second reason is related to the computer revolution and import competition based on that technology, notably from Japan (Storper and Scott, 1992; Helpman and Trajtenberg, 1998; Brynjolfsson and Hitt, 2000), which encourages adopting these technologies regardless of how related they are to the technological portfolio.

We find further evidence for this second suggestion in Appendix 4.5. In that section, we relate the diversification in technologies during crises to those that are upcoming or become outdated according to technological change that coincides with these crises, as shown in Figure 4A12.

We distinguish between upcoming technologies and outdated technologies. For example, technologies under electrical & electronics are considered upcoming during the electric revolution that coincides with the Long Depression but become outdated when the computer revolution comes about that coincides with the Oil Crisis. We also distinguish between cities that are specialised in upcoming technologies or outdated technologies.

We then find that upcoming technologies have a larger probability of entry, during or outside of crises, in most time periods. Whereas outdated technologies have less or equal probability of entry compared to other technologies. There is not much difference in the probability of entry for cities specialised in either upcoming or outdated technologies. Also in this case the level of diversity is the most important city-level variable that predicts diversification.

When reproducing the main results on the difference in diversification patterns when entering a crisis, see Figure 4.1. The results show once again that there is not much difference between cities specialised in outdated or upcoming technologies. Outdated technologies do see a larger drop in the probability of entry, particularly when less related. On the other hand, upcoming technologies show a much smaller decrease in the probability of entry, which is even not statistically significantly different from non-crisis periods when these are the most related or the most unrelated. Further analysis shows that this is mostly due to strong diversification in upcoming technologies during the Oil Crisis. This is in line with the earlier suggestion on the importance of diversifying into computer & communication technologies, even when less related, in this time period due to the computer revolution and rising import competition.

4.5 Conclusion

In this Chapter, we provide systematic evidence on the diversification patterns of regions in times of major crisis. Diversification is considered to be a crucial part of regional resilience, as developing new capabilities may allow regions to come out of crises. For a long time, questions like the ones asked here relied on case studies, which although insightful were difficult to generalise. Combining developments in data availability and in methods to quantify relatedness, we were able to examine the technological diversification of MSAs in the U.S. during the Long Depression, the Great Depression, and the Oil Crisis.

We found that crises have a strong dampening effect on diversification, and that especially diversification in less related technologies is reduced compared to more prosperous times, which is in line with the demand-pull hypothesis (Schmookler, 1966; Freeman et al., 1982; Scherer, 1982).

Additionally, we also show that the Great Depression was particularly deep and that the Oil Crisis has a very distinct pattern. Considerably less diversification was lost during the Oil Crisis, in particular in upcoming technologies, like those in computers & communication. This may be due to the computer revolution and strong import competition based on that technology during that time period (Storper and Scott, 1992; Helpman and Trajtenberg, 1998; Brynjolfsson and Hitt, 2000).

Furthermore, we also show that more diverse cities manage to diversify more than their more specialised counterparts during crises, which is in line with suggestions that there are less vested interests in the policy and institutional context that block new developments (Boschma, 2015; Neffke et al., 2018). This comes on top of the advantage that diverse cities have because of increased technological proximity to more technologies due to the larger technological portfolio.

These results give a detailed description of the diversification of regions during major crises. However, the study remains largely descriptive, causal mechanisms can be suggested from theory but are not tested directly. Future research could develop on the features of diverse regions and upcoming technologies that allow them to be involved more strongly in diversification than their relatedness density would suggest.

Another interesting avenue for future research is the observation that during great crises there are radical technological changes, like industrial revolutions, at the national level but at the local level cities diversify less and more into closely related technologies. This suggests that as nations move into new technological paradigms it does so through cities that have related activities whereas cities with unrelated activities are stuck, thereby leading to great changes in the distribution of welfare, as, for example, after the computer revolution, see (Moretti, 2012). Although the Oil Crisis shows that unrelated diversification can also occur in crises, the extent to which this observation holds and the conditions in which it does not are fruitful avenues for future research.

Furthermore, this Chapter describes how regions diversify during times of crisis but not how this impacts the depth and duration of crises. Do regions that diversify more strongly or more into less related activities experience less damage from crises, and under which circumstances? Related diversification is suggested to be more sustainable in the long run in a city because it can build on local capabilities (Balland et al., 2019). Rigby et al. (2022) indeed showed that growth in GRP and employment have been higher in European cities that diversified into related and complex technologies but this has not been extended to the context of regional resilience.

The framework also allows to retrace previous case studies in the data and compare them with a large sample of other cities. This would for example allow us to examine if the story of “Reinventing Boston” (Glaeser, 2005) is a story of unrelated diversification against the odds or a more common case of related diversification, and whether major crises like the ones we examined had a major impact on the diversification pattern in the Boston region. The described diversity of economic activities in Boston and the associated diversification through economic downturns is in line with the results here that more diverse regions outperform more specialised regions in diversification during crises. In this sense, the results also shed light on how large cities like New York remain among the top largest cities of the country through economic cycles.

This Chapter is limited by its focus on technological diversification based on patent data. Consequently, it picks up only that part of new knowledge that is embodied in patents. To get a more comprehensive picture of the resilience of cities, it is important to account for other forms of knowledge that may provide opportunities for cities to diversify. This would include other forms of new activities like new products, industries or new jobs in which cities can diversify, which are not captured by patent data, like in most tertiary activities (Xiao et al., 2018).

Finally, a possible improvement for future research would be to include the role of institutions in regional resilience research (Boschma, 2015). Recent research has shown that regional institutions like bridging social capital matter for the ability of regions to diversify (Cortinovis et al., 2017). This might be especially relevant in times of crisis when high demands are put on institutional agents to renew their economies, adapt their institutions, and enable the development of new growth paths (Freeman and Perez, 1988; Amable, 2000; Hall and Soskice, 2001). This requires more understanding of the effect of regional institutions on regional resilience, and whether institutional agents like policymakers can make the difference during major crises (Bristow and Healy, 2014; Dawley, 2014; Evenhuis, 2017; Sotarauta et al., 2017).

For policymakers, there are several implications. If crises hamper (unrelated) diversification then these activities may need stimulation, in particular, to reabsorb resources set free because of the crisis. To alleviate crises and overcome lock-in a diverse industry base seems strongly relevant. However, improving diversity and therefore regional resilience through diversification is currently not part of diversification policies, such as the Horizon 2020 and Smart Specialisation strategy of the European Union. The impact of crises on diversification and therefore possible avenues for recovery should not be neglected.

4A Appendix: additional results and robustness checks

Background information on crises

This section presents more information on (changes in) patent dynamics during the great historical crises. Figure 4A1 depicts the number of MSAs entering a period of growth in green, respectively a period of crisis in red per year, during the time periods associated with each of the great historical crises. The impact of the crises on patenting activity is clearly visible in the number of MSAs that start a period of crisis in red compared to those that start a period of growth in green. This suggests that patenting is a suitable proxy for regional economic activity.

One can also note a small time lag between the actual start of the great historical crises and MSAs entering a period of downturn in patenting for the first two major crises whereas the effect of the Oil crisis is immediately noticeable.[9] Because of the time lag in the reaction of patenting activity, we retain the regional crises that start in years when more MSAs enter a crisis period tha MSAs enter a period of growth. For the Long depression, this is 1876 to 1878, for the Great depression 1932 to 1938, and for the Oil crisis 1972 to 1976. All other crisis periods are dropped from the sample. Regional periods of growth are kept regardless of when they occur.

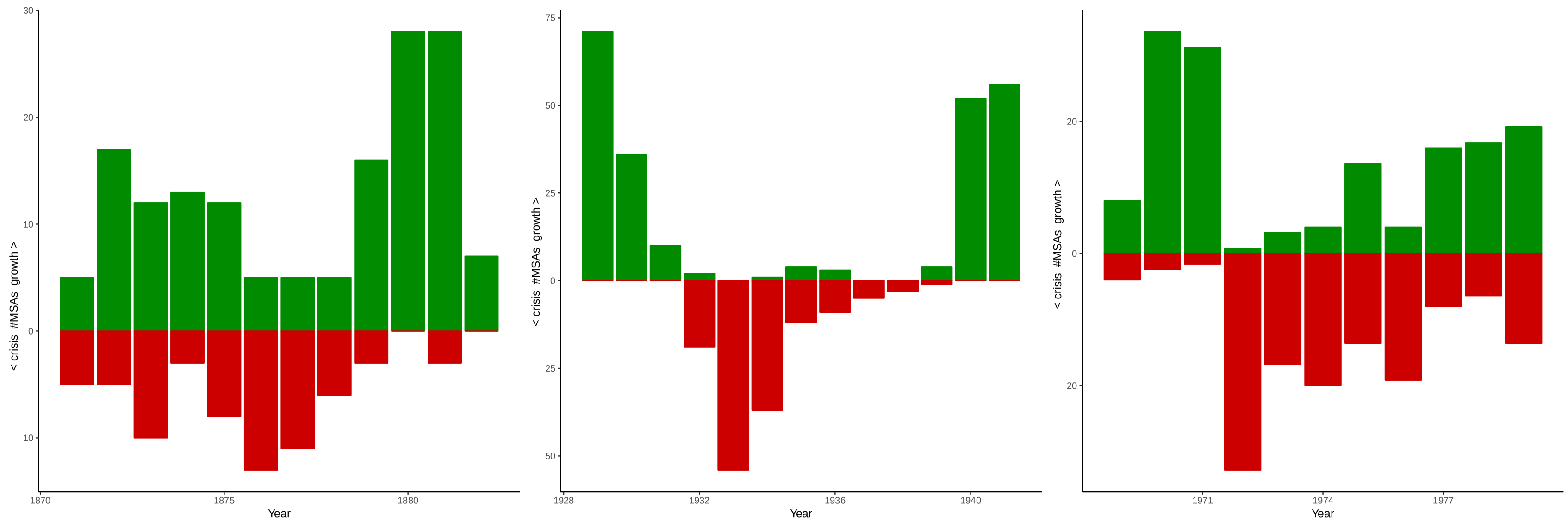
Table 4A1 shows the strong impact of the great historical crises on the patent production at the regional level. Affected MSAs, in the second column, indicates the number of MSAs that enter a crisis that meets the aforementioned requirements and respective time period. Unaffected MSAs are MSAs that were in a growth phase before the start of the crisis and remain so over the course of the crisis. #MSAs gives the total number of MSAs that meets the requirement of producing on average ten patents per year in that time period. This is not equal to the sum of unaffected MSAs and affected MSAs as MSAs could already be in crisis upon entering the respective time periods or could enter a crisis in which the requirement of losing more than 35% of patenting activity is not met. The last two columns respectively give the average duration of the crises, and the average percentage of patent activity lost at the trough compared to the peak for the affected MSAs. In these respects, the Great Depression stands out as the heaviest crisis.

[9] Note that the years indicate the end year of the previous cycle period and the start year of the next cycle period. E.g. a period of crisis starting in 1972 indicates that the peak was in 1972 and the first year of downturn is 1973.

Table 4A1 – The Regional Impact of the Great Historical Crises

Crisis	Aff. MSAs	Unaff. MSAs	#MSAs	Avg. Crisis Length	Avg. Crisis Depth
Long depression	30	19	101	$\sim$ 3.7 years	$\sim$ -53.6%
Great depression	139	2	205	$\sim$ 6.4 years	$\sim$ -74.3%
Oil crisis	128	44	252	$\sim$ 6.7 years	$\sim$ -59.4%

Figure 4A1 – Number of MSAs starting a period of growth (green) versus a period of crisis (red)



(a) Long Depression (b) Great Depression (c) Oil Crisis

Background information empirics

In this section, we further provide information on two underestimated risks in the empirics of the widely used entry models. First, the dependent variable entry is strongly right-skewed, *i.e.* there are very few incidences of successful entries (values of 1) compared to technologies that do not enter (values of 0). Second, the independent variable relatedness density is strongly right skewed, i.e. values range from 0 to 1 but are more strongly concentrated to the left of the mean, as can be seen in Figure 4A2.

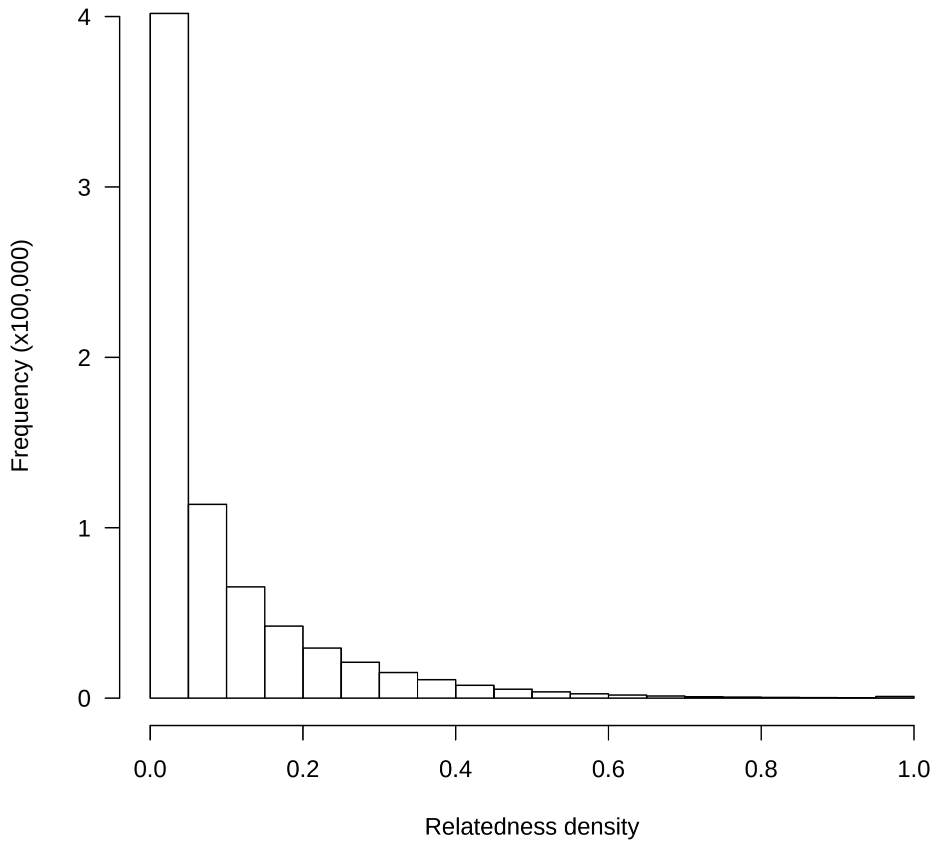
The first has already been noticed by Boschma et al. (2015), referring to work by King and Zeng (2001). They argue that the coefficient estimates of nonlinear models might not be consistent when there are too many zeros in the dependent variable. They, therefore, use an OLS to estimate the entry model. The use of such a Linear Probability Model (LPM) has certain risks that can be considered to be outweighed by the benefits of easier interpretation (see Hellevik, 2009). However, when the probability of "success" of the dependent variable is on the extreme ends of the distribution, as is the case here, the slope of a logit or probit is not well approximated by the slope of a linear regression and the flaw of the LPM in predicting probabilities outside the possible range of 0 to 1 generally becomes apparent. Von Hippel (2015) suggests that probabilities of success should be in the range of 20% to 80% for logit and linear models to be used interchangeably.

Therefore, a logistic regression seems more appropriate.[10] As said, this is not without risk as King and Zeng (2001) warn for inconsistent estimates when probabilities are extremely low. King and Zeng (2001) also provide guidelines when this risk is more likely to exist. They show in a simulation that the inconsistency tends to zero as the sample size tends to infinity and/or the percentage of ones tend to 50%. In our data, there are $724,752$ observations and an average probability of entry of 3.1%. Following guidelines and simulation results of King and Zeng (2001), the risk can be assumed to be negligible. We, therefore, argue that a logit model is the appropriate method to estimate the entry model.

Then there is the second issue related to the main variable of interest: relatedness density. As it is strongly right-skewed there is a sizeable risk that outliers exert a strong influence on the estimated coefficient on relatedness density.

[10]Note that Boschma et al. (2015) do run a logit model as a robustness check, which confirms their results.

FIGURE 4A2 – HISTOGRAM OF RELATEDNESS DENSITY



Therefore, we substitute the continuous relatedness density variable for an ordered categorical variable by creating dummy variables for each quantile of relatedness density values. *I.e.* we rank the relatedness density values and create five dummy variables so that each designates a fifth of these values from the 20% lowest values to the 20% highest values. The categorisation of relatedness density also has the added benefit that it controls for the possibility of this variable having a non-linear relation in log odds with the dependent entry variable.[11]

[11]Note that we did explore several monotonic transformations to reduce the risk of outliers but found these to be unfruitful. The most common option, a log transformation, cannot be applied because there are zeros among the values. Other common transformation methods, which are able to deal with zeros, *e.g.* box-cox, taking the square, or an inverse hyperbolic sine, also fail to give a distribution in which a strong influence of outliers can be ruled out. This is likely due to that an impressive 12.9% of observations involve zero relatedness density.

TABLE 4A2 – REGRESSION RESULTS (HYPOTHESIS 1 AND 2)

(Dependent variable: entry of technology class i in the technological portfolio of city c at time t)

	Naive specification	*+ crisis× relatedness density*	*+ fixed effects*
	(1)	(2)	(3)
Relatedness density (20%-40%)	0.340***	0.335***	0.357***
	(0.042)	(0.044)	(0.045)
Relatedness density (40%-60%)	0.814***	0.811***	0.779***
	(0.039)	(0.040)	(0.042)
Relatedness density (60%-80%)	1.328***	1.307***	1.229***
	(0.037)	(0.038)	(0.041)
Relatedness density (80%-100%)	1.866***	1.811***	1.758***
	(0.036)	(0.037)	(0.041)
Crisis	−0.334***	−0.988***	−0.724***
	(0.022)	(0.178)	(0.182)
Diversity	0.412***	0.414***	0.366***
	(0.006)	(0.006)	(0.013)
Population	0.012	0.009	0.007
	(0.010)	(0.010)	(0.014)
Present×**W**	0.396***	0.396***	0.312***
	(0.005)	(0.005)	(0.009)
Degree centrality	−0.145***	−0.142***	−0.081***
	(0.011)	(0.011)	(0.012)
Relatedness density (20%-40%) × crisis		0.321	0.236
		(0.205)	(0.206)
Relatedness density (40%-60%) × crisis		0.332*	0.249
		(0.191)	(0.194)
Relatedness density (60%-80%) × crisis		0.548***	0.426**
		(0.183)	(0.187)
Relatedness density (80%-100%) × crisis		0.821***	0.600***
		(0.180)	(0.184)
Constant	−4.790***	−4.757***	−5.057***
	(0.033)	(0.034)	(0.444)
Time Fixed Effects	No	No	Yes
Technology Fixed Effects	No	No	Yes
MSA Fixed Effects	No	No	Yes
Observations	724752	724752	724752
Log Likelihood	-87613	-87568.1	-81544.2
Akaike Inf. Crit.	175246	175164.2	164552

Notes: The relatedness density groups and crisis are dummy variables with as reference category, respectively, the 20% lowest relatedness density values and non-crisis time periods; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

Additional results

Supplementary material to the main results

The marginal effects estimated in the main results of Table 4.2 are based on the regression results presented below in Table 4A2.

As the marginal effects in the main results of Table 4.2 only holds for the reference category of the 20% lowest relatedness density values and some of the interaction terms between crisis and relatedness density values in Table 4A2 are positive and significant, we show the marginal effects for other relatedness density values in comparison to the average probability of entry outside of crises in Figure 4A3.

Here the sample average probability of entry per relatedness density group outside of crises is given in blue. For the first group of relatedness density values, the probability of entry is, as said in the main text, 0.63%.[12] The red line gives the marginal effect of crisis and its 95% confidence intervals vis-à-vis the blue baseline. As the marginal effect of crisis is minus 0.32% for the first relatedness density group the probability of entry is about 0.31% during crises.

Clearly, the red line is significantly lower than the blue line across all relatedness density groups, which indicates that the marginal effect of crisis is statistically significant for all relatedness density values. This confirms that the probability of a MSA entering a new technological specialisation is lower during a crisis regardless of relatedness density.

[12] This average probability is equal to the intercept in column (3) of Table 4A2 converted to probabilities as population and Present$\times\mathbf{W}$ have been scaled to have a mean of zero and we use sum-to-zero contrasts for the fixed effects. Note that there is obviously also a margin of error to this estimate but this is not shown in the figure as we are interested in the marginal effect of crisis with respect to the average probability of entry outside of crises.

Figure 4A3 – Probability of entry according to crisis status

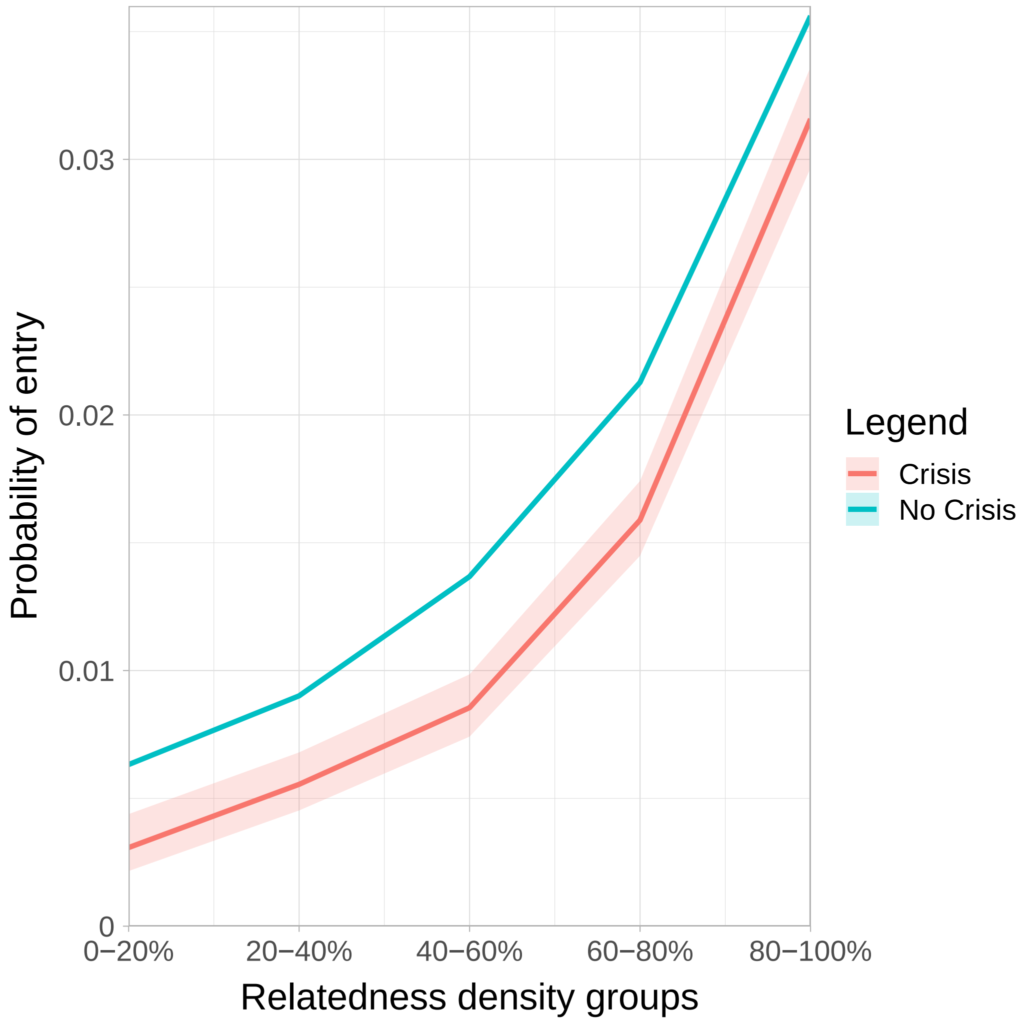


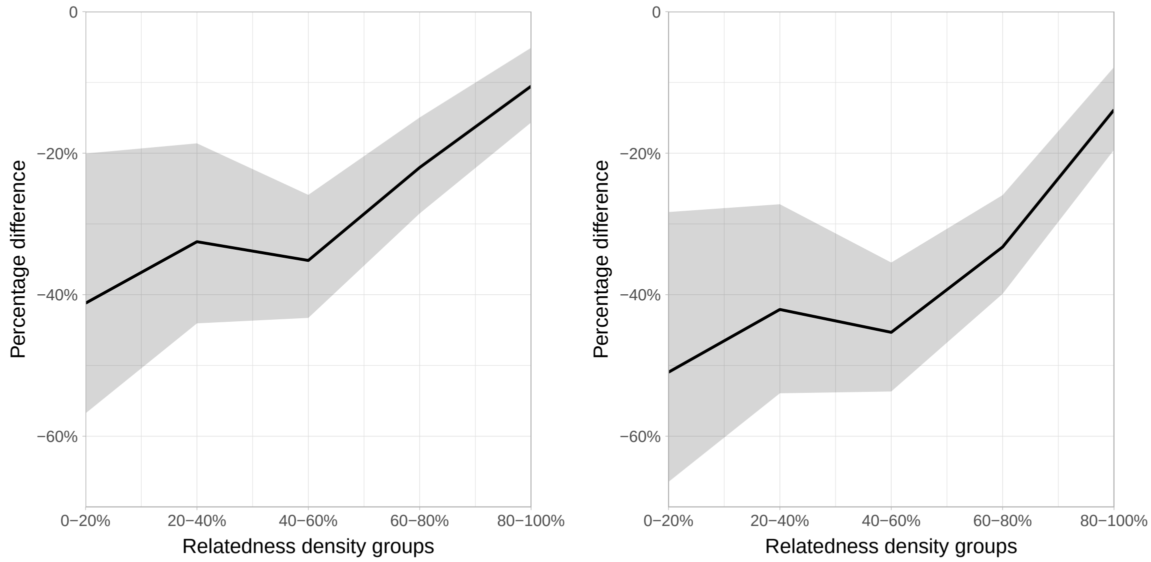
Figure 4A3 also gives insight on the foundation of Figure 4.1 in the main text, which is based on the difference expressed in percentages between the blue line and the red line and its confidence intervals.

Crisis depth variations

A first interesting check is to see how the depth requirement of the crisis impacts the results. In the main analysis, regions had to lose at least 35% of patenting activity during one of the major crises to be taken into account. In Figure 4A4a and Figure 4A4b, we reproduce Figure 4.1 putting the depth requirement at, respectively, at least 25% and at least 45%.

In the former case, the loss during crises in entry probability is reduced across all relatedness density groups, which suggests that diversification patterns during smaller crises are similar to that outside of crises, which is to be expected. Although we have to note that these results are not statistically significantly different from those with a 35% depth requirement. The crises with a 45% depth requirement lead to a similar reduction in entry probability compared to the main results.

FIGURE 4A4 – DIFFERENCE IN PROBABILITY OF ENTRY DURING CRISES (≥25% AND ≥45% CRISES)



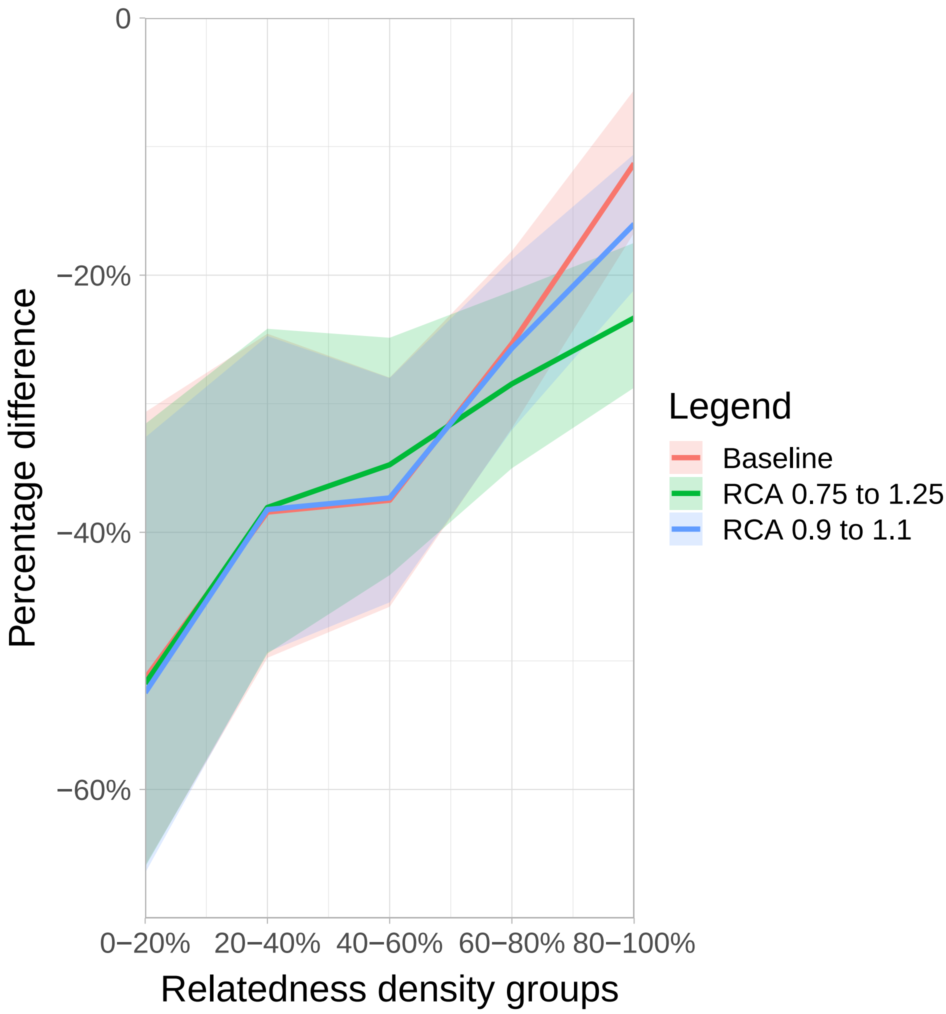
(a) ≥25% crises (b) ≥45% crises

Entry with variations in RCA threshold

Another interesting check is to see if the entry variable is influenced notoriously differently during crisis compared to non-crises due to the choice in the threshold of RCA, which we use to define entry. In the main results, an entry is seen as an increase in the RCA from below 1 to above 1. This means that a change from 0.99 to 1.01 is seen as an entry even though it is a negligible change in the technological portfolio of a region. Although this issue plays a role both in crisis and in non-crisis periods and therefore may not directly impact the difference in diversification patterns between these types of time periods, we explore how the results would be when defining entry only when larger changes in RCA are observed, respectively from 0.9 to 1.1 and from 0.75 to 1.25. The results are shown in Figure 4A5.

These results are highly similar to the main results depicted in Figure 4.1, reproduced by the red line. We do notice that there are more small movements in RCA during crisis periods for the most related technologies because when these are left out the drop in entry probability becomes larger when entering a crisis, as indicated by the difference between the red and the green line for the 80-100% most related technologies.

Figure 4A5 – Difference in probability of entry during crises (Entry RCA thresholds)

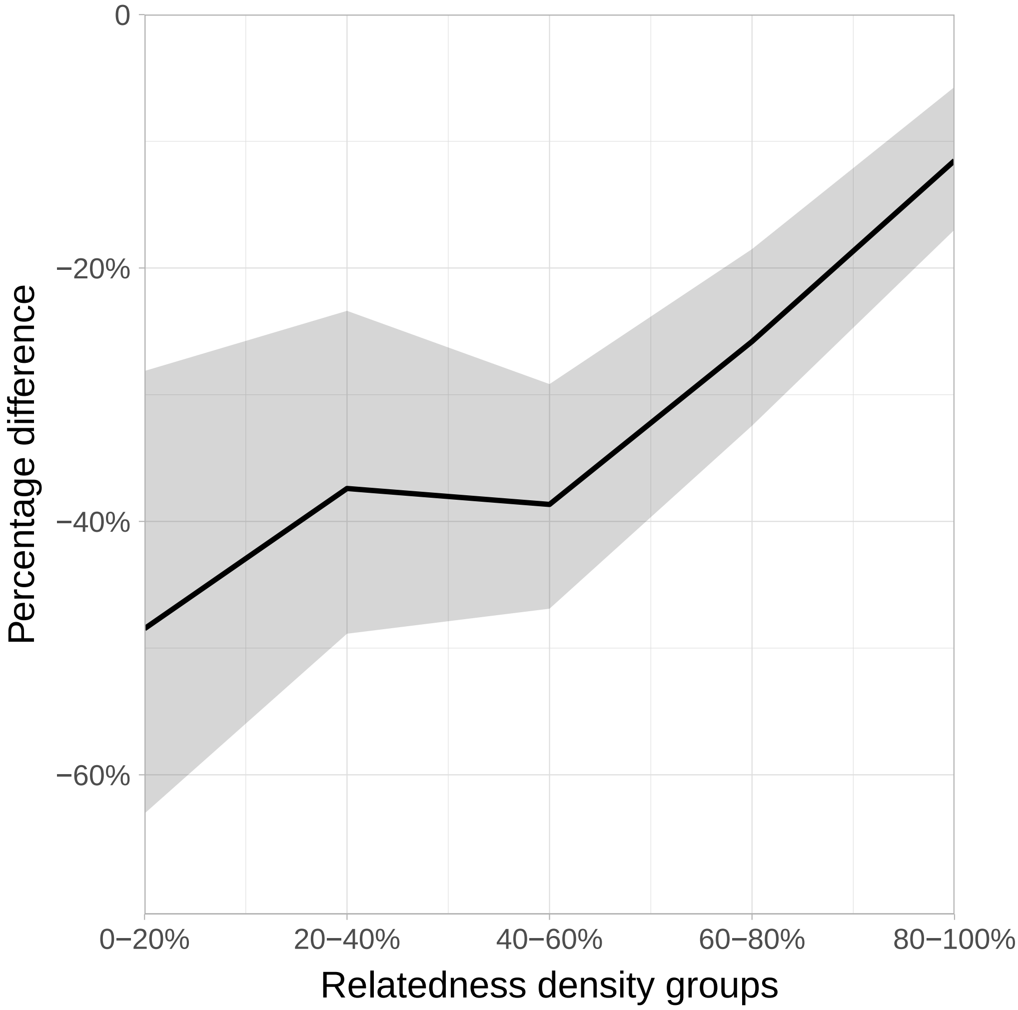


Diversification during local crises

Another data choice in this Chapter was to only retain downturns in patenting activity as periods of crisis when they occurred during one of the three great historical crises to ascertain that these local downturns were not due to changes in patent activity unrelated to actual economic downturns. The downside of this method is that differences between these periods and those where most of the non-crisis periods are not fully captured by our time fixed effects and may therefore influence our results. Therefore, we reproduce the main results in Figure 4A6, in which we take all downturns in patenting activity of at least 35% into account.

The results confirm those in the main analysis. However, more interesting is that the confidence intervals are much smaller. This indicates that these crises behave in a very similar way as those in larger economic crises and that therefore statistical efficiency increases by adding these other crisis periods.

FIGURE 4A6 – DIFFERENCE IN PROBABILITY OF ENTRY DURING CRISES (ALL CRISES)



Diversification during crises at the county level

Another data choice in this Chapter was to look at the MSA level. However, MSAs are based on counties that are socially and economically integrated according to recent Census Bureau definitions. These levels of integration likely much less existed in history when infrastructure was less well developed and agglomerations not as large. Agents in an area were, therefore, more likely to interact and redistribute resources during crises within a smaller area. The finest spatial level at which data on patents and control variables is available is the county level. We reproduce the main results using county-level data in Figure 4A7.

Although the results are not statistically significantly different from the main results and show a similar pattern the underlying data is much more problematic than that for the main results, as the much larger confidence intervals already suggest. It is much more difficult at the county level to meet thresholds for the number of patents in total and per class. As a result, the number of observations is much lower here than in the main results $202,112$ versus $724,752$. Furthermore, not every county and time period, and not every entering technology, knows both crises and non-crises periods, which means that the marginal effect of crisis is based on comparing different instead of the same counties, time periods, and classes, which is not the case in the main results.[13] Although it is reassuring about the main results to see that there is a similar pattern we advise not to highly value these results.

Relatedness density values based on uniform time periods

In the main results, we define time periods based on the boom-bust cycle algorithm of Harding and Pagan (2002). As a result, some time periods last for the minimum of two years, while the longest time period is 30 years with the median being 5 years. As we calculate relatedness density values based on the previous time period this also means that the technologies in the portfolios of cities and the relatedness between technologies is calculated at different distances. To make sure that these definitions do not influence results we calculate relatedness density values using the 0-5 years before the start of a time period for all time periods and reproduce the main results in Figure 4A8.

This confirms the main results but also shows that using this definition even leads to a stronger reduction of diversification in the least related technologies during crises and smaller confidence intervals compared to the main results. We also experimented with calculating relatedness density values in even earlier time periods but found that these results were not reliable.[14]

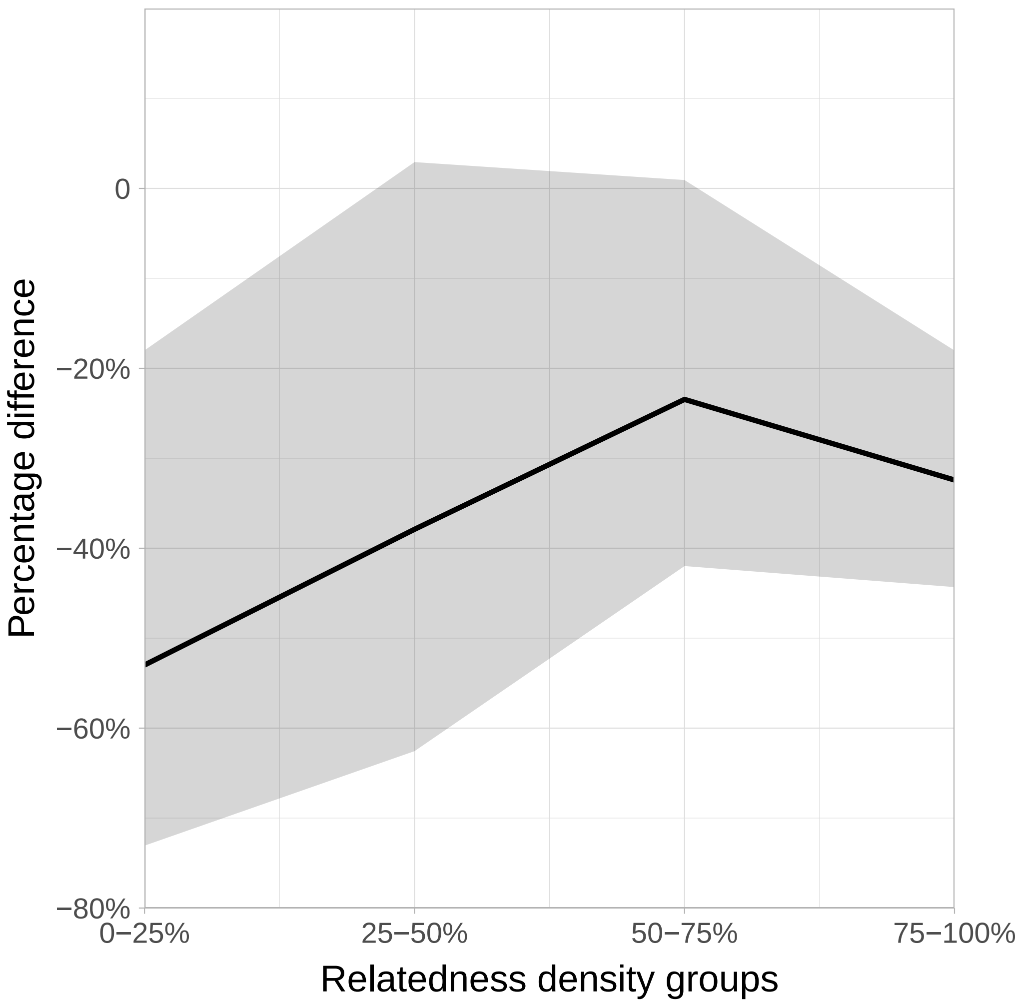
Quasi differences-in-differences approach

In the previous section, we used the same time period to calculate relatedness density values for all MSA-time periods instead of the time periods based on the boom-bust

[13] As a result, one cannot use fixed effects in this regression. We tried to alleviate this effect by discerning four relatedness density groups instead of five as in the main results. This increases the chance that counties, time periods and entering technologies are present both in crisis and non-crisis for each relatedness density group.

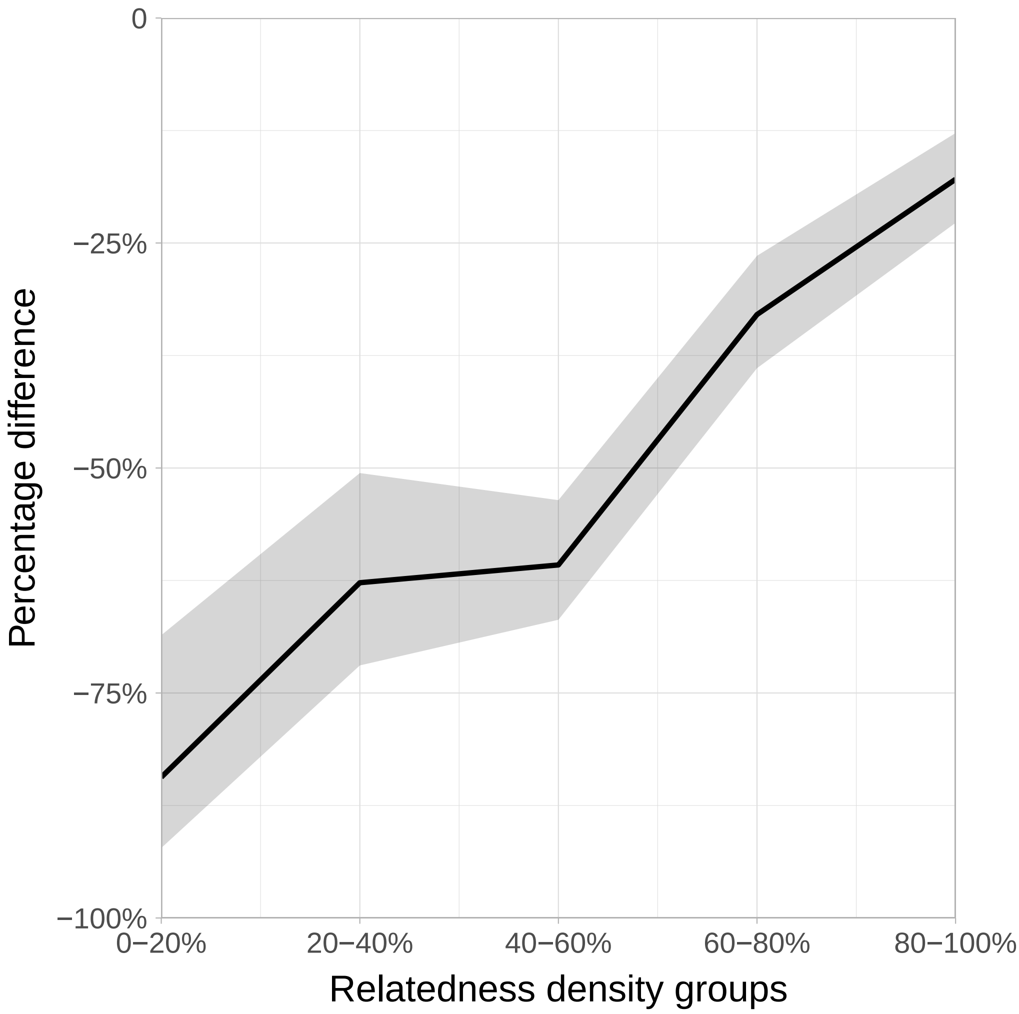
[14] Over time technological portfolios of regions change. This means that entering technologies are related to technologies present in the region 5 years ago but not so much to those 20 or 30 years ago. The longer ago relatedness density values are calculated the lower relatedness density values to entering technologies are and the less predictive this variable is on entry, *i.e.* the lines in Figure 4A3 become flatter and flatter. We found that this effect is purely driven by the technological portfolio of regions in earlier time periods because large regions, which have fewer large changes in their technological portfolio over time, experience this effect to a much lesser extent. Also, when calculating relatedness based on patents from longer ago but keeping technological portfolios as in the main analysis leads to the same results.

Figure 4A7 – Difference in probability of entry during crises (county level)



cycle algorithm of Harding and Pagan (2002) in the main results. In this section, we take the approach one step further by setting all time periods equal and only using observations from the great historical crises. This means that not only the entering technologies are compared to the technological portfolio of a region at the same number of years earlier, like in the previous section, but also the time length in which these technologies can enter. Furthermore, by dropping observations outside of the great historical crises. Diversification in times of growth and crisis are compared at the exact same time period. Through this approach, we compare the diversification patterns of regions that enter a crisis to those that do not while controlling for observables, *i.e.* diversity, population, degree centrality, and relatedness density, plus for unobservable at the MSA, technology, or time level. Through this approach, we come closer to a difference-in-difference approach, as first experimented by Card and Krueger (1994).

Figure 4A8 – Difference in probability of entry during crises (Adj. relatedness density)



Although we have much less data to observe the extent to which non-crisis and crisis counties are similar except for the mentioned observable characteristics and unobservable characteristics that are fixed at the regional, technological, or time level.

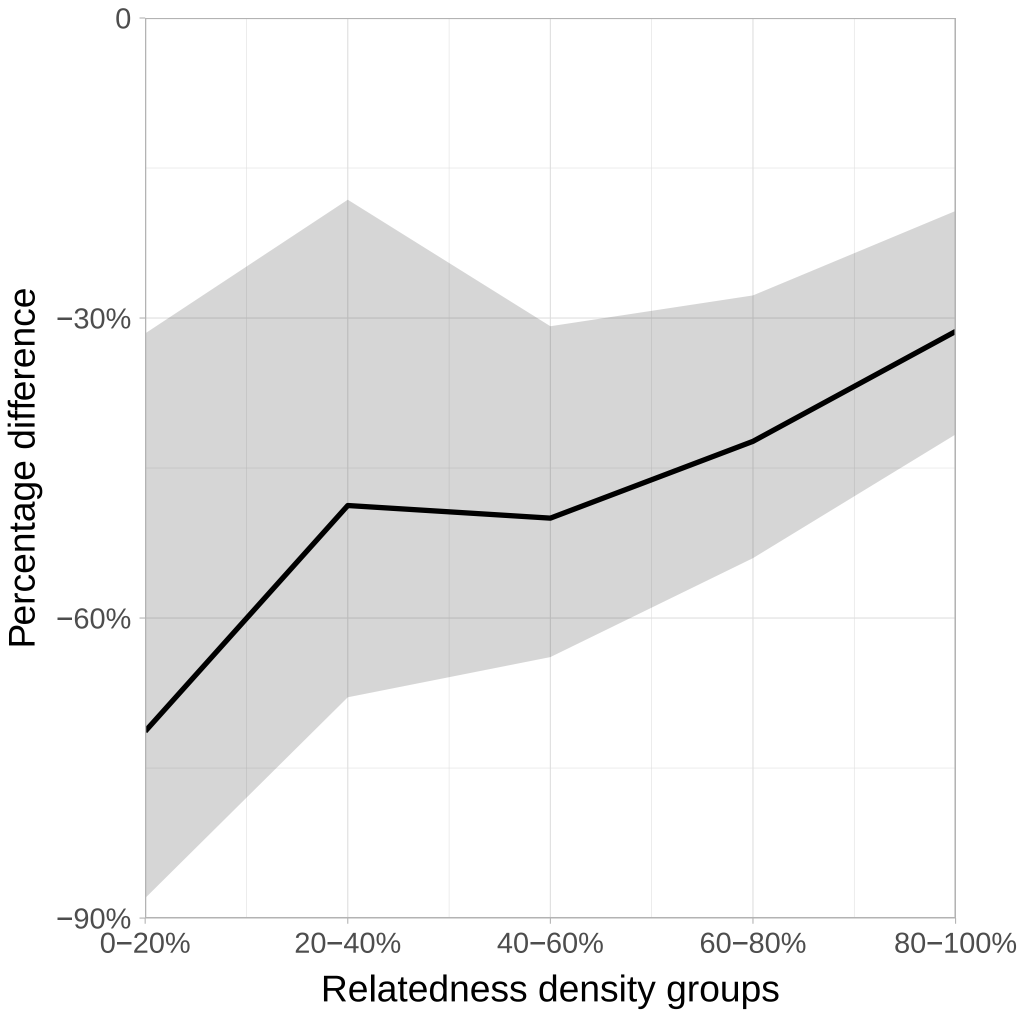
For the Long Depression, we look at the crisis status and calculate entering technologies between 1876 and 1879 comparison to technological portfolios of these areas and relatedness density between 1873 to 1876. To be considered a successful entry, technologies still have to be present in the area between 1879 and 1882. For the Great Depression, these years are, respectively, 1932 to 1938, 1926 to 1938 and 1938 to 1944. For the Oil Crisis, these years are, respectively, 1972 to 1976, 1968 to 1972, and 1976 to 1980. The time span is chosen to mirror the length of each of the crises.[15]

Figure 4A9 shows that when one of the great historical crises hit regions that enter a crisis diversify less, in particular in unrelated technologies, than regions that are unaffected in the same time periods. This pattern is not statistically significantly different and similar to the main results. Nevertheless, the confidence intervals are much larger, in particular for the unrelated technologies. This is likely because there are notably fewer observations, in particular of non-crisis periods, as only few regions are not affected by the great historical crises, and of unrelated technologies, as these know relatively less entry than more related technologies.

That this approach confirms the main results suggests more strongly that there may be a causal link between entering a crisis and diversification patterns. Although it remains impossible to fully isolate the effect of crises on diversification patterns in this historical setting, like a natural experiment would.

[15]Note that in differences-in-differences approaches it is also custom to take the period before the event, in this case, crises into account. In this way, crisis counties are not only compared to non-crisis counties in the same time period but also to themselves before the crisis. As these latter observations are very strongly present in the main results, where most non-crisis periods occur outside the great historical crises, we decided not to add these here to ascertain that the main results also hold for the comparison within the same time period.

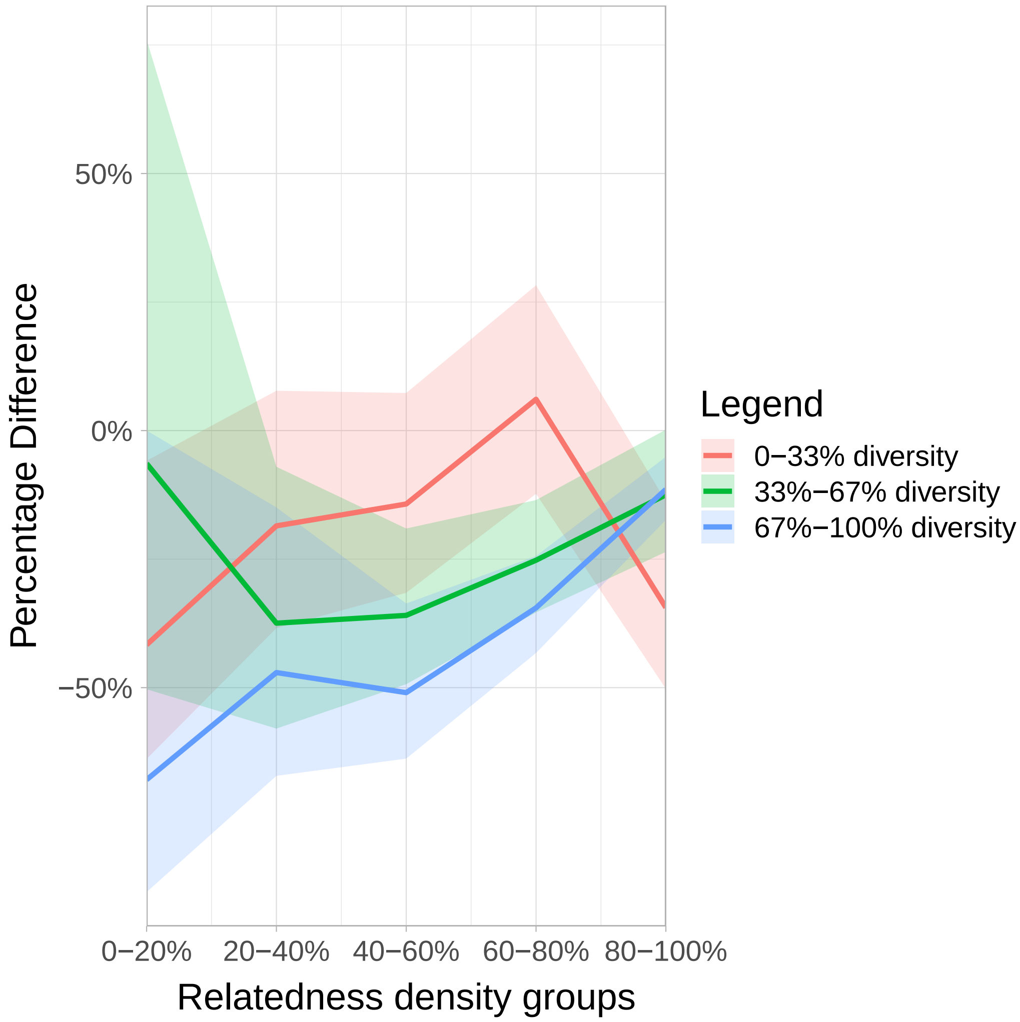
Figure 4A9 – Difference in probability of entry during crises (quasi diff-in-diff)



Diversity and diversification

In the main results of Table 4.2, we conclude that diverse regions have an advantage compared to their specialised counterparts to develop new activities outside and during crises. This brings up the question on how diverse regions change the focus of their diversification when entering a crisis compared to more specialised regions, as done in Figure 4.1. In other words: do diverse cities switch more strongly to less related technologies during crises than specialised cities? To this aim, we reproduce figure 4.1 but by estimating the effect for different groups according to RDI. The resulting Figure 4A10 is shown below.

Figure 4A10 – Percentage difference in probability of entry between crisis and no crisis across RDI groups



When entering a crisis, the 67%-100% percentile of most diverse regions, lose over 67.9% of the diversification in the least related technologies. While the most specialised regions, in the 0%-33% least diverse percentile, only lose 41.7%, and the intermediate group only lose about 6% during crises than outside of crises.

A possible explanation is that diverse regions are more likely to have unrelated variety between industrial sectors, see Frenken et al. (2007), meaning that some sectors are not affected by regional crises and that developing technologies related to these unaffected sectors is a secure and reasonable use of resources. On the other hand, more specialised regions when hit by a crisis are less likely to have unaffected industries to continue to develop and as such there's more incentive to focus on locally less common, and therefore, less related technologies. The fact that averagely diverse cities

(33%-67% diversity values) focus more strongly on unrelated technologies than their most specialised counterparts (0-33% diversity values) may be due to the latter being in a state of technological lock-in, see Grabher (1993) and Boschma (2015), in which the knowledge of actors and views in an area are focused on such a way on current core activities that it inhibits the development of new sectors and technologies.

However, the confidence intervals are so large that these differences are not statistically significant.[16] All in all, one cannot claim that there are significant differences in diversification patterns in relation to the diversity of a region.

Diversification per crisis

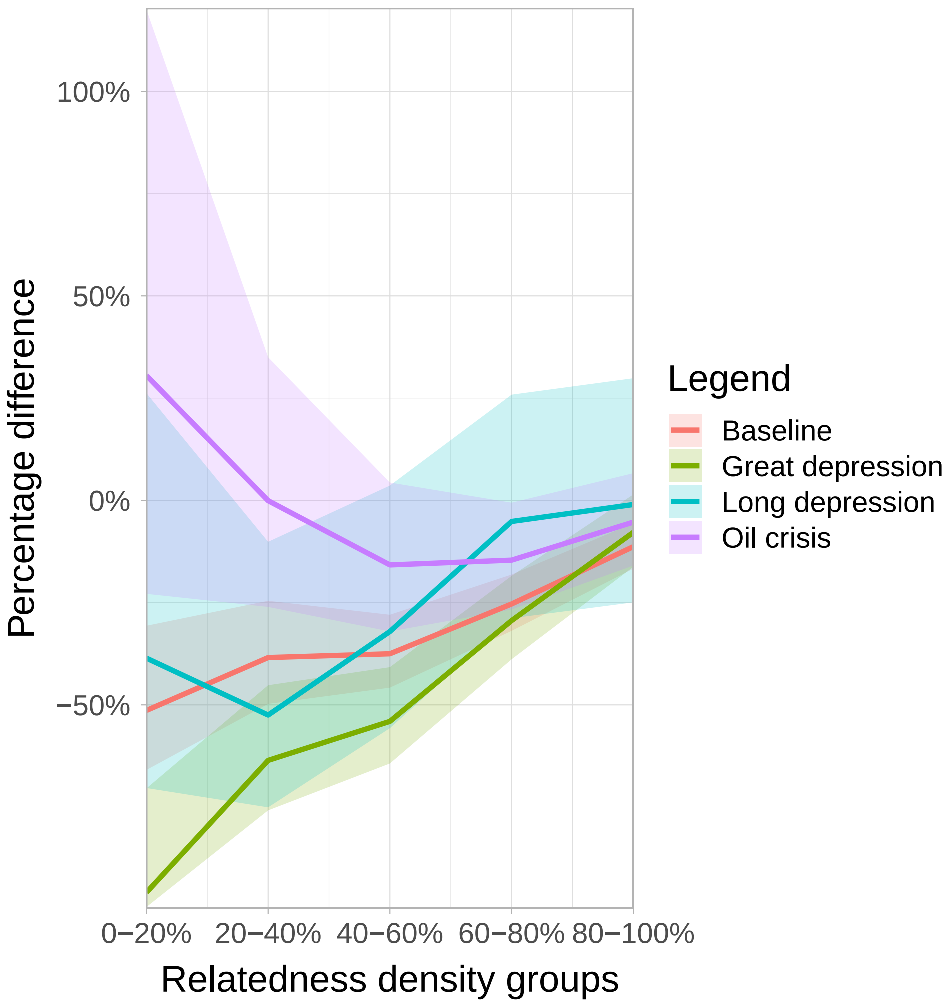
In the main results, all three great historical crises are lumped together in a single analysis. In Figure 4A11 we reproduce the main results per crisis for individual scrutiny.

The Long Depression show similar results as the main results, given by the baseline curve. In comparison, the Great Depression knows a significantly larger drop in the probability of entry of unrelated technologies. Most interesting is that the Oil Crisis shows a very distinct pattern where diversification during crises is not statistically significantly different from periods of growth. Diversification in unrelated technologies even seems to increase during this crisis although this is not statistically significant.

There may be several possible explanations for this difference although future research is necessary to find the exact reasons. First of all, the nature of the oil crisis, rising costs of energy inputs, was very distinct from the other two, which were financial crises. This may suggest that during the oil crisis there were more means to back investments and incentives to divert to new technologies that were less reliant on oil. Also, the oil crisis co-occurs with the rise of automation, following the computer revolution, and import competition, most notably from Japan, based on customisation using this technology, which motivated firms in regions to switch to these technologies (Storper and Scott, 1992; Helpman and Trajtenberg, 1998; Brynjolfsson and Hitt, 2000).

[16]Note that there are no formal testing methods available as each marginal effect has a different baseline (depicted in blue in Figure 4A3). Therefore, we have to rely solely on the confidence intervals.

Figure 4A11 – Difference in probability of entry during crises per crisis

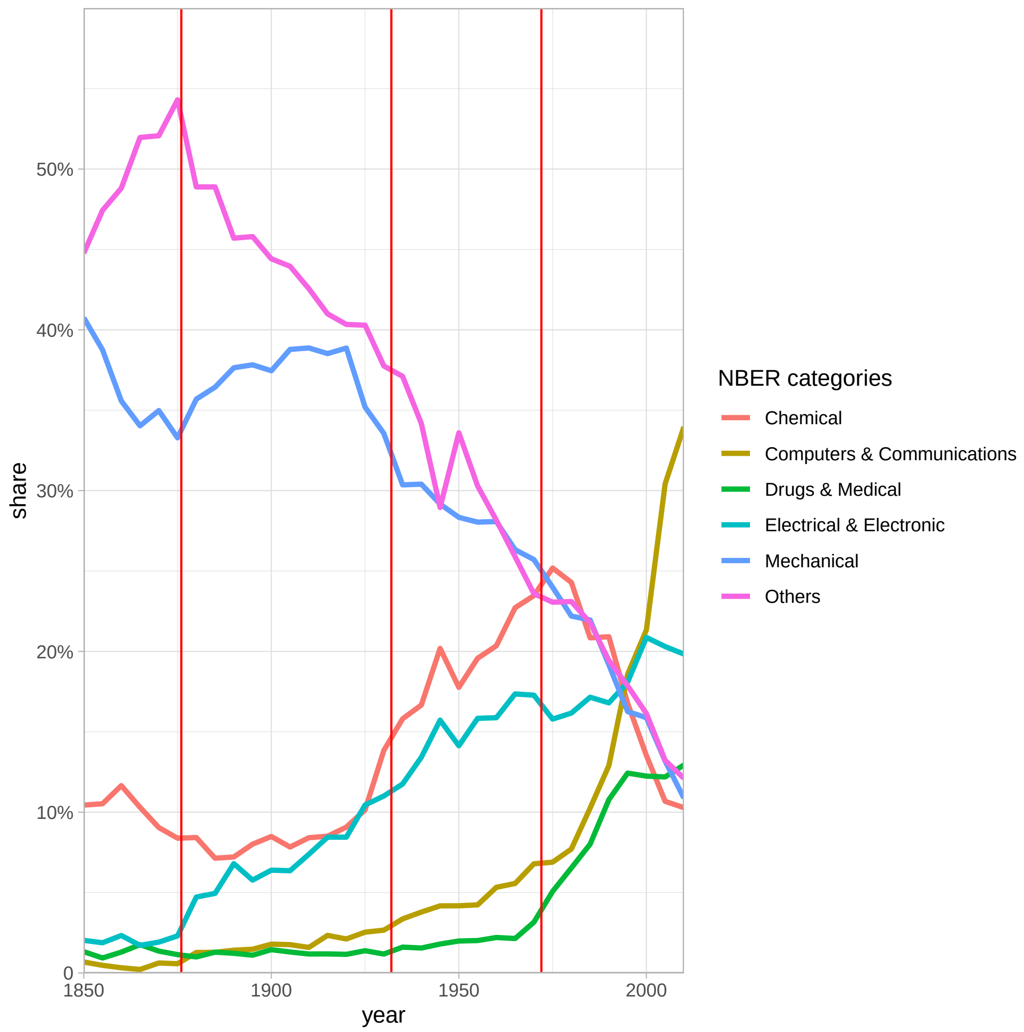


Crises, diversification and technological change

The great historical crises all occur during periods of great technological change. Notably, the Long Depression and the electrical revolution, and the Oil Crisis and the computer revolution. The radical nature of technological change likely implies that some technologies and cities show different diversification patterns than others as technologies are coming up while others become outdated. This section explores these differences.

Our patent data captures the coincidence of technological change and the great historical crises, as is illustrated in Figure 4A12, where the share of patents per NBER technological category is given over time with the red vertical lines indicating the start of each of the respective crises. Here one can clearly see that at the time of the Long Depression, which coincides with the electrical revolution, the upcoming technologies belong to the categories mechanical and electrical & electronic while those in others, textiles among others, become outdated. During the Great Depression, the chemical and electrical & electronic technologies are upcoming while those in others and mechanical are outdated. During the Oil Crisis, which coincides with the computer revolution, the previously upcoming technologies in chemical and electrical & electronic become outdated while computers & communication and drugs & medical come up.

FIGURE 4A12 – SHARE OF PATENTS PER NBER CATEGORY OVER TIME



Upcoming technologies may give rise to different diversification patterns. After all, when in crisis it may be worth the risk, *i.e.* having less related capabilities, to invest in promising technologies. On the other hand, outdated technologies may be a less interesting option no matter how related.

Furthermore, cities that are specialised in outdated technologies may have a different diversification pattern than other cities. There is a strong risk of technological lock-in (Boschma and Lambooy, 1999; Hassink, 2005; Pike et al., 2010; Boschma, 2015). However, it is unknown if this shows in diversification patterns. In the main text, we showed that diverse regions have an advantage in diversifying compared to specialised regions, within and outside crises, here the question is to what extent this holds for cities specialised in outdated technologies.

To this end, we classify technologies per time period into upcoming, outdated or neither based on the observations from Figure 4A12, discussed above. We classify MSAs as upcoming, outdated or neither if they have a LQ in one of these categories.[17] Because of this classification technology fixed effects and MSA fixed effects are dropped from the analysis.

Table 4A3 shows the marginal effects of the baseline regression, as shown in equation 4.5, with the addition of dummy variables indicating if technologies, respectively, MSAs, belong to, respectively are specialised in, upcoming or outdated technologies. As well, as an interaction of these variables with the crisis dummy. Per column, we use data around each crisis and period of technological change: 1870 to 1919 for the Long Depression; 1920 to 1969 for the Great Depression; 1970 to 2000 for the Oil Crisis. The final column groups together all three of these datasets.

The relatedness density variables behave in a similar fashion as before, as expected. Although note that relatedness density seems to matter more during the time period since the Great Depression. The crisis variable is (close to) insignificant in the case of the Great Depression and the Oil Crisis for the reference category, which in this case is not only the non-crisis periods of the 20% lowest relatedness density values but also technologies and cities that are neither upcoming nor outdated. Earlier, we saw in Section 4.5 that the entry of unrelated technologies in the Oil Crisis are not significantly impacted by the crisis. Whereas for the Great Depression this only seems to hold for technologies and cities that are neither classified as upcoming nor outdated, *i.e.* the reference category.

[17] When a MSA has a LQ in two of these categories we choose the category with the largest LQ. It is technically impossible to have a LQ in all three.

Table 4A3 – Regression results - technological change - marginal effects

(Dependent variable: entry of technology class i in the technological portfolio of city c at time t)

	Long Depression	*Great Depression*	*Oil crisis*	*All crises*
	(1)	(2)	(3)	(4)
Relatedness density	0.0078***	0.0062***	0.0038***	0.0042***
(20%-40%)	(0.0054, 0.0106)	(0.0042, 0.0085)	(0.0020, 0.0062)	(0.0033, 0.0052)
Relatedness density	0.0155***	0.0180***	0.0104***	0.0110***
(40%-60%)	(0.0122, 0.0193)	(0.0149, 0.0214)	(0.0072, 0.0143)	(0.0096, 0.0125)
Relatedness density	0.0272***	0.0403***	0.0226***	0.0237***
(60%-80%)	(0.0224, 0.0326)	(0.0353, 0.0459)	(0.0170, 0.0296)	(0.0214, 0.0261)
Relatedness density	0.0454***	0.0827***	0.0464***	0.0473***
(80%-100%)	(0.0383, 0.0535)	(0.0739, 0.0922)	(0.0361, 0.0592)	(0.0433, 0.0515)
Crisis	−0.0034***	−0.0016	−0.0008*	−0.0015***
	(−0.0049, −0.0009)	(−0.0039, 0.0015)	(−0.0018, 0.0005)	(−0.0023, −0.0006)
Upcoming tech.	0.0004**	0.00001	0.0016***	0.0003***
	(−0.0001, 0.0010)	(−0.0006, 0.0007)	(0.0010, 0.0024)	(0.0001, 0.0006)
Outdated tech.	0.0005**	−0.0013***	−0.0004***	−0.0004***
	(−0.0001, 0.0011)	(−0.0018, −0.0008)	(−0.0007, −0.0001)	(−0.0007, −0.0002)
Upcoming MSAs	−0.0011***	0.0011***	0.0003*	−0.000002
	(−0.0015, −0.0006)	(0.0005, 0.0017)	(−0.0001, 0.0007)	(−0.0002, 0.0002)
Outdated MSAs	−0.0002	0.0020***	−0.0001	0.0006***
	(−0.0007, 0.0003)	(0.0010, 0.0031)	(−0.0006, 0.0005)	(0.0003, 0.0009)
Diversity	0.0035***	0.0045***	0.0023***	0.0024***
	(0.0033, 0.0037)	(0.0042, 0.0048)	(0.0021, 0.0025)	(0.0023, 0.0025)
Population	−0.0028***	0.0006***	0.00004	0.0001
	(−0.0035, −0.0022)	(0.0002, 0.0009)	(−0.0001, 0.0002)	(−0.0001, 0.0002)
Present×**W**	0.0036***	0.0054***	0.0019***	0.0028***
	(0.0035, 0.0038)	(0.0052, 0.0056)	(0.0018, 0.0019)	(0.0028, 0.0029)
Degree centrality	0.0010***	−0.0022***	−0.0008***	−0.0009***
	(0.0003, 0.0017)	(−0.0025, −0.0018)	(−0.0010, −0.0006)	(−0.0010, −0.0007)
Time F.E.	Yes	Yes	Yes	Yes
Technology F.E.	No	No	No	No
MSA F.E.	No	No	No	No
Observations	200348	338464	172912	711724

Notes: The relatedness density groups, crisis, and technology related groups are dummy variables with as reference category, respectively, the 20% lowest relatedness density values, non-crisis time periods, and technologies/MSAs that are neither upcoming or outdated; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

Interestingly, in the time period around the Oil Crisis upcoming technologies have a larger chance of entry, 0.16% to be exact, see column (3) in Table 4A3. This may seem small but the probability of entry for the reference category in this time period is only 0.14%, which suggest a doubling of the probability of entry.[18] This may be because of the trade competition in that time period, notably from Japan, that already adopted computerised machinery (Storper and Scott, 1992), which increased pressure to diversify to upcoming technologies in computers & communications.

[18]Note that this relationship is not mechanical because of the definition of upcoming technologies as entry is based on the fact that a region obtains a *relative* specialisation in this technology and not an *absolute* specialisation.

Outdated technologies, on the other hand, have a smaller probability of entry, which is to be expected. The only exception being the Long Depression, which may be due to less competition and less circulation of new ideas in relation with the more limited connectivity between cities in that time period. In this line, Perlman et al. (2015) shows that patenting activity in the 19th century is strongly related to railroad access, which was not fully developed at the time of the Long Depression.

Cities specialised in upcoming technologies show mixed results in terms of diversification. Whereas those in outdated technologies show either virtually non-existent or positive marginal effects. All in all, the coefficient on diversity remains strongly positive and much larger across all specifications, which suggests that advantages at the city level are derived from this aspect rather than a specialisation in a certain category of technologies.

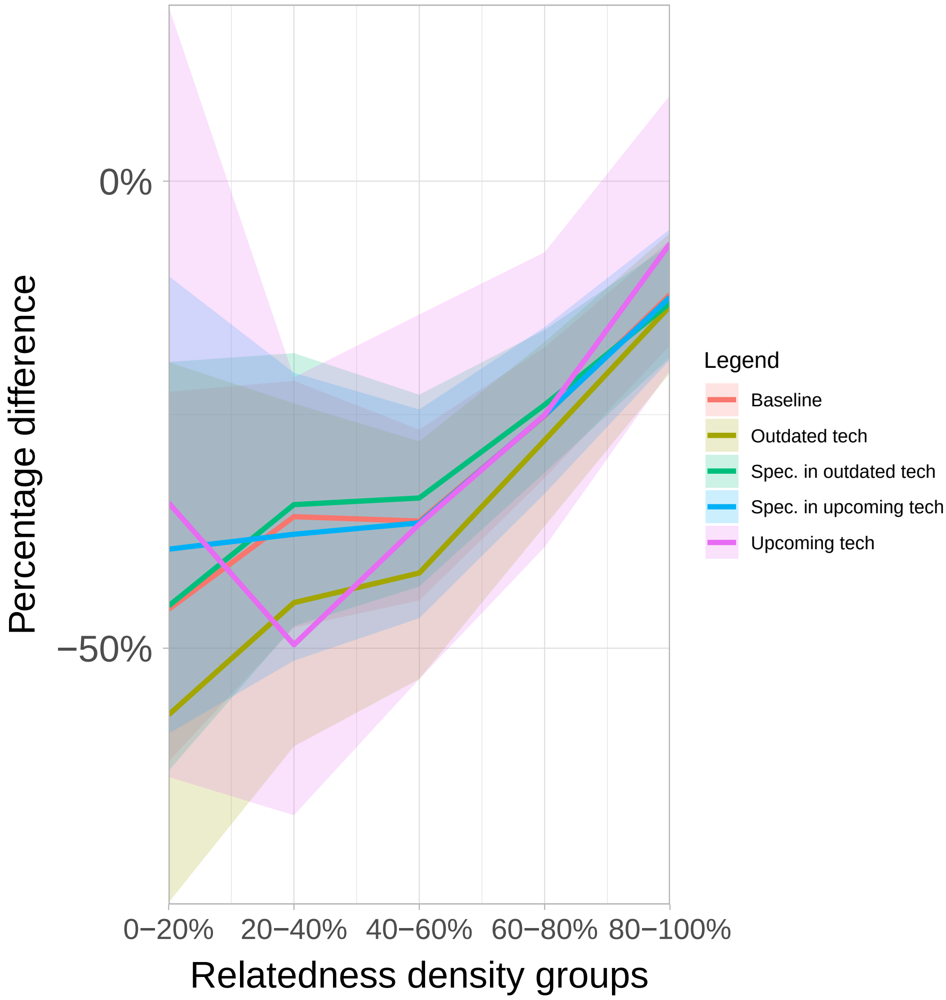
Table 4A3 gives differences in the probability of entry but not how diversification changes in a relative sense when entering a crisis. To show this latter aspect, we reproduce Figure 4.1 per type of technology and MSA in Figure 4A13.

The change in diversification patterns when entering a crisis is similar to the baseline results in red for most of these types. MSAs that are specialised in outdated technologies show virtually the same pattern as MSA that are specialised in upcoming technologies and the baseline results. Outdated technologies show a stronger relative decrease in entry probability when entering a crisis, in particular, when unrelated to the technological portfolio of a region. This suggests that diversifying into these technologies is even less attractive when entering a crisis, plausibly to avoid technological lock-in. However, confidence intervals are too large to state that this pattern is significantly different from the baseline results.

Most interesting is the line of upcoming technologies. The effect of a crisis is not statistically different from a period of non-crisis for both the most related as well as the most unrelated technologies. This suggest that these technologies remain (close to) equally as attractive to diversify into even when in crisis. This is likely because these technologies contain the General Purpose Technologies of that industrial paradigm (Helpman and Trajtenberg, 1998).

When we reproduced Figure 4A13 per crisis, we found that it was during the oil crisis that unrelated upcoming technologies, in particular, showed less decrease in entering probability. This further lends credibility to the statement in the previous section that computer based import competition motivated regions to diversify in the upcoming computer & communication technologies of that time period, even when these were not so related to the technologies that were previously patented in the area.

Figure 4A13 – Difference in probability of entry during crises (technological change)



Chapter 5

Conclusion

5.1 Overview

Economies of agglomeration are at the core of our understanding of the spatial distribution of human activities. In recent times, the topic is in the limelight of attention because of the so-called great divergence between a number of growing regions, which are confronted by rising housing prices and social exclusion, and a relatively large number of lagging regions, which are confronted by reductions in employment and liveability. Despite these societal challenges, economies of agglomeration are still considered a black box (Duranton and Puga, 2004a; Combes and Gobillon, 2015; Davis and Dingel, 2019). In particular: the mechanisms of agglomeration, how their roles changed over time, and the channels of economic change impacting these roles (Ellison et al., 2010; Moretti, 2012; Combes and Gobillon, 2015; Storper, 2018), which is the focus of this thesis.

Economies of agglomeration have been extensively described in one way or another since ancient history, see for an overview Finley (1973) and Silvermintz (2010). Notable are the categorisation of Marshall (1890) into labour market pooling, input-output linkages, and knowledge spillovers, which is the focus in the first chapter, and the division of labour, often attributed to Smith (1776), which is the focus in the second chapter. Also, there have been notions on the dynamic aspect of economies of agglomeration, from early on. Notably, how regions can develop new specialisations based on their industrial past, see Vernon (1960); Chinitz (1961), and Jacobs (1969), which is the focus of the third chapter.

Less work exists on measuring these agglomeration forces. A notable step forward in measuring the relative importance of each of Marshall's agglomeration determinants is the work by Ellison et al. (2010). The division of labour can be analysed straightforwardly in certain well-documented cases, like patents and research articles, see for example Jones (2009). Analysing the development of new regional specialisations has received an impressive boost with the development of the concept of relatedness, see Hidalgo et al. (2007). By building on these foundations it is relatively straightforward to extend the measurement over larger time periods when limitations in data availability have been overcome. Thereby addressing the gaps in our understanding of the changes in these roles.

The greatest challenge of this thesis lies in understanding the reasons for these changes as theory and methods on the subject are less well developed. To address this challenge, this thesis combines insights on the heterogeneity of the roles of economies of agglomeration between activities with those between time periods: notably during technological revolutions. Specific attention is paid to technological change, trade competition, transportation costs, economic complexity, regional resilience, and the diversity of sectors in the local environment. The four research questions chosen within this broad domain are as follows:

RQ1: To what extent has the importance of Marshall's determinants of agglomeration changed over time?

RQ2: Why has the importance of Marshall's determinants of agglomeration changed over time?

RQ3: To what extent do complex activities concentrate in large cities?

RQ4: To what extent do diverse cities differ from more specialised cities in terms of diversification behaviour during crises?

5.1.1 Contributions

In this thesis, contributions are made along five dimensions: theory, data, methodology, code, and empirics. In terms of theory, connections are made between various related concepts in urban economics/economic geography and evolutionary economic geography as well as by connecting these to related ideas in labour economics, innovation studies, and complexity theory.

Data building is a quintessential part of this thesis as evaluating changes in the economies of agglomeration and its impacting factors requires data over long time periods, which was previously unavailable. To overcome barriers in data limitation use is made of novel techniques, such as OCR, and the rising online availability of scanned documents and datasets from the pre-internet era. Most notable is the development of a unique balanced panel dataset with consistent geographical units and industries on employment numbers, occupations, input-output linkages, patented knowledge, trade competition, transportation costs in Chapter 2 and an update to the HISTPAT database (Petralia et al., 2016) to make it geographically consistent for the analysis of historical patent data in Chapters 2 and 3.

In terms of methodology, improvements are made in comparison to previous approaches to mitigate omitted variable bias, by adding more fixed effects and the approach of Oster (2019) in Chapter 2; and reverse causality by developing new instruments in Chapter 2. Also, new questions like the interaction of economic factors and the determinants of agglomeration require new approaches, such as the two-step approach introduced in Chapter 2. In Chapter 4, improvements are made in taking into account the functional form of the regression. Furthermore, an improved formula to calculate relatedness is introduced in this thesis.

In terms of code, the new relatedness measure has been released as a function in the EconGeo package for R and also a R package called fastlogitME was developed, which allows for a less computationally intensive method to calculate the marginal effects of logit models in R.

Empirically, the results show that there is considerable dynamism in the economies of agglomeration; whether in the changing relevance of mechanisms of agglomeration over the past 44 years in Chapter 2; the growing importance of agglomeration since 1850 in Chapter 3; or the temporal effect of agglomeration across the cycle of boom and bust in Chapter 4. These dynamics are strongly associated with channels of economic change, more specifically: routine biased technological change and trade competition in Chapter 2; the rising complexity of knowledge-intesive activities in Chapter 3; and the technological change of the great historical crises in Chapter 4. In the following, the empirical results of each chapter are shortly summarised.

Agglomeration determinants change over time: knowledge spillovers become more important

The first part of Chapter 2 addresses the research gap mentioned by Ellison et al. (2010, p.1210), Moretti (2012, p.124) and Storper (2018, p.255) on the changes over time in the relevance of agglomeration forces as categorised by Marshall (1890): labour market pooling, input-output linkages, and knowledge spillovers. By gathering new data from historical sources and using technological relatedness instead of patent citations to proxy for knowledge spillovers improvements are made with respect to earlier studies using similar approaches, such as Ellison et al. (2010); Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020).

The results show that over time the importance of labour market pooling and input-output linkages as a determinant of agglomeration has decreased whereas the relevance of knowledge spillovers has increased.

This last point is one of the strongest proofs to date that the importance of sharing ideas has increased as is claimed by a wide literature, see for example Gaspar and Glaeser (1998); Leamer and Storper (2001); Storper and Venables (2004); Rodríguez-Pose and Crescenzi (2008); McCann (2008); Glaeser (2011); Moretti (2012). Previous research on this point was based on changes in the spatial concentration of idea-intensive activities. However, such concentration could also come about for other reasons, such as consumer preferences following the increases in the value of time, due to increasing returns to skill, see Glaeser et al. (2001). The results here estimate changes in the *reasons* for agglomeration, instead of changes in agglomeration, and show that knowledge spillovers have considerably gained in importance.

Trade and technological progress are strongly associated with changes in agglomeration determinants

The second part of Chapter 2 focusses on explaining the documented changes in the relevance of Marshall's agglomeration determinants. Thereby addressing the gap noted by (Combes and Gobillon, 2015, p.336) that there has been no attempt yet to interact the relevance of the determinants of agglomeration with economic factors explaining these.

Where economic geography and urban economics generally consider transportation and communication technologies to understand changes in agglomeration (McCann, 2008; Duranton and Storper, 2008; Glaeser, 2011; Moretti, 2012) and would focus on decreasing pecuniary transportation costs, see Glaeser and Kohlhase (2004), here changes in agglomeration are explained by factors derived from labour economics and innovation studies that have more likely impacted the opportunity cost aspect of transportation costs of goods, people, and ideas since the computer revolution: trade competition and routine-biased technological change, *i.e.* technology taking over routine tasks of humans.

To empirically test for the association of trade, technology, and transportation costs with the relevance of the determinants of agglomeration a novel two-step estimation procedure is used, thereby improving on previous analyses of heterogeneity by Faggio et al. (2017) and Diodato et al. (2018).

Empirically, the findings show that trade competition and technological progress are negatively associated with labour market pooling and positively with knowledge spillovers. Trade competition is somewhat negatively associated with input-output linkages, whereas the transportation costs of goods do not seem to matter. Further exploration shows that it is more likely that input-output linkages decrease in importance due to import substitution following trade competition rather than an increase in skill/knowledge intensity as suggested by Faggio et al. (2017). This is because robustness checks using more specific indicators of skill/knowledge intensity like wages and R&D expenditures, just like technological progress, do not impact input-output linkages, whereas more detailed measures of trade competition, like input import competition, do show strong relations.

Complex activities concentrate in large cities

In Chapter 3, the focus is on the division of labour as mechanism of agglomeration and the expanding frontier of knowledge as channel of economic change. To understand the link between the division of labour and spatial concentration a qualitative understanding of knowledge is needed rather than quantitative count measures previously used in the literature, see Carlino and Kerr (2015) for an overview. Qualitative measures that are nonetheless measurable instead of more arbitrary static self-developed classifications as in Pavitt (1984) and Faggio et al. (2017).

Here several continuous non-geographical measures of the complexity of knowledge are developed by building on ideas from innovation studies and complexity theory. These measures are based on the insight that more complex new knowledge requires more face-to-face-contact as it involves tacit unfamiliar knowledge (Breschi and Lissoni, 2003; Storper and Venables, 2004), and a finer division of labour as it requires individuals to strongly specialise in a limited number of tasks because it is impossible for a single person to possess all this knowledge (Leamer and Storper, 2001; Jones, 2009). The measures are the size of teams involved in scientific papers, the average number of years of education per occupation or per industry, the average year of introduction of technology classes for patent categories.

The results show that scientific papers with a larger team, industries and job occupations with more years of education, and patents with a more recent year of introduction concentrate more strongly in large cities. This suggests that complex activities concentrate in large cities.

The time dimension of the patent data show that over time patents increasingly concentrate in large cities as technology advances, in particular during the industrial revolutions of 1870 based on electricity and 1970 based on the semiconductor (computer). Furthermore, around the time of the computer revolution a divergence occurs in which less complex patents concentrate less strongly and more complex patents concentrate more strongly. This is in line with Leamer and Storper (2001) and is likely due to the dual effect of the improvement of communication and transportation technologies that on the one hand allows for the routinisation and dispersion of the less complex technologies but increases the need for physical connection for the more complex technologies. This suggests that over time the importance of proximity increases to facilitate the division of labour as the knowledge frontier is being pushed further making larger cities increasingly important as the engines of growth.

Diverse cities have diversification advantages in times of crisis
In Chapter 4, the focus is on the industrial distance and temporal distance of agglomeration, more specifically the development of new specialisations by regions, in relation to the boom and bust cycle, as channel of economic change. The development of new growth paths is seen as an important part of regional resilience (Boschma, 2015). However, not much is known on how regions diversify during crises. Furthermore, it is known that diverse cities are more resilient in preventing crises but not if more diverse cities have an advantage compared to more specialised cities in developing new growth paths (Chinitz, 1961; Balland et al., 2015; Boschma, 2015). Certain case studies have suggested that specialised cities may have a low capacity to diversify in new activities, because they are cognitively, socially and politically locked-in (Grabher, 1993; Boschma and Lambooy, 1999; Hassink, 2005; Pike et al., 2010). However, there is no systematic evidence on the difference between these types of regions. This question is particularly relevant as it helps to understand why certain cities top the national hierarchy in city size over centuries despite changes in technological paradigms over time.

Using geolocalised historical patent data, the analysis focusses on regional technological diversification during the great historical crises: Long Depression (1873-1879), the Great Depression (1929-1934), and the Oil Crisis (1973-1975), which coincide with moments of rapid technological change (Boschma, 1999).

Results show that cities diversify less in new technologies during crises but when they do these are more likely to be related to previous technologies in which the city already had a comparative advantage. This confirms a regional version of the demand-pull hypothesis of Schmookler (1966); Freeman et al. (1982) and Scherer (1982).

Furthermore, diverse cities are shown to outperform more specialised cities in two ways. First, they have larger technological portfolios and therefore on average more relatedness density to possible new technologies, which increases the probability of entry. Second, when one controls for relatedness density diverse cities still outperform their more specialised counterparts. This suggests that agents in diverse cities are generally more open to new activities, which is in line with suggestions that vested interests against new developments are stronger in specialised cities (Grabher, 1993;

Boschma, 2015; Neffke et al., 2018). A final analysis shows that there is no significant difference between more diverse cities, intermediately diverse cities, and specialised cities in the extent to which they switch to more unrelated diversification when entering a crisis.

5.2 Policy implications

This thesis has not directly considered policies. Nevertheless, current ideas and debates on policy can benefit from the insights in this thesis. It is, therefore, best to talk about policy *implications* and their context. To come to actual full-fledged policy *advice* new lines of research are necessary, which is discussed in Section 5.3.

5.2.1 Industrial policies

First of all, Chapter 2 confirms the suggestions of a large literature that the local sharing of knowledge has become more important as an agglomeration advantage. On the other hand, a strong common labour market pool has become less important and input-output linkages have even become almost irrelevant as an agglomeration motive. This helps policymakers understand why firms cluster in proximity in cities. Cities hosting clusters based on input-output linkages may need to consider the stimulation of industries with related technologies to enhance knowledge spillovers as its rising appreciation suggests it is important for productivity. Also when attempting to attract new firms through industrial policies it is important to know which local assets to promote, in this case, those related to knowledge spillovers seem to become more attractive for firms. Next to Chapter 2, it may help for policies targeted at certain industries to also consider Faggio et al. (2017); Diodato et al. (2018) and Faggio et al. (2020), who consider different dimensions of industry heterogeneity.

However, policymakers should be aware of the larger debates on industrial policies as these are contested tools as also discussed further on in Section 5.3. There are 220 cluster initiatives in the Netherlands attempting to mimic the success of a high-skilled cluster like Silicon Valley (Zeemeijer, 2016). It does not require expert insight to know that not all of these initiatives can be successful.

The literature shows that industrial policies are contested tools because first of all, “successful” expenditure is highly dependent on the quality of institutions (Acemoglu and Robinson, 2012; Rodríguez-Pose et al., 2018). Second, it is very hard for policymakers to guess which industries will be the next success, in the words of Moretti (2012, p.201) “*they need to be a little like venture capitalists*”. This makes local industrial policies some sort of betting with public money in the hope of attracting and developing the next successful industries, although the smart specialisation framework gives some guidance, see Balland et al. (2019). On top of this risk, regions tend to overbid to outcompete other regions in attracting firms. Where it is at least questionable if social benefits outweigh the local costs let alone if it is societally justifiable to redistribute welfare towards the owners of these companies (Moretti, 2012; Kuijpers and Thomas, 2016). Furthermore, the firms receiving these subsidies are often large multinational companies (Greenstone et al., 2010; Moretti, 2012; Kuijpers and Thomas, 2016). This increases their market power, which following basic microeconomic theory leads to market failures as it prevents supply from reaching the societal optimum resulting in welfare loss.

In general, from a societal welfare perspective, the attraction of a firm is a zero-sum game in aggregate national (or global) welfare as the production lost in some place appears in another place assuming there are no market failures or system failures.[1] This is a strong assumption but most politicians will likely mostly consider if an investment makes their own citizens better off rather than these types of market/system failures.

More research is needed to fully understand the different welfare effects for different agents and regions from industrial policies to come to balanced policy advice. Ideas for this type of research are discussed in Section 5.3.2. The implications of the results of this thesis on certain forms of market failures are discussed next.

5.2.2 Market failures

Market failures are generally regarded as an acceptable reason for government intervention. One important case of market failure is externalities and therefore related to the topic of this thesis: agglomeration externalities.[2] As externalities are not priced by the market, the agent at the origin of the benefits of these externalities is not compensated for the productivity gains these entails for other agents. Therefore a Pigouvian subsidy would seem in place to correct for the market failure (Baumol, 1972).

Promoting knowledge spillovers

Some forms of policy are warranted to motivate the sharing of ideas as not only the rising importance of knowledge spillovers demonstrated in this thesis suggests but also the rising college wage premium, Acemoglu and Autor (2011), the long-term effects

[1] An example of an externality, in this case, may be that knowledge spillovers generated by the firm may benefit a larger number of other firms in the area to which it is attracted thereby generating more economic growth. An example of market failure may be that due to imperfect information or non-rational behaviour, in the economist sense, a firm has chosen a suboptimal location, which is common as discussed in behavioural economics (Pred, 1967). An example of system failure is when the attracted firm can fulfil a key missing link in the regional innovation system and thereby help the entire region prosper (Boschma, 2008). In these cases, it may not exactly be a zero-sum game.

[2] Although the extent to which agglomeration externalities are actually externalities is somewhat contested by Breschi and Lissoni (2003) and Fitjar and Rodríguez-Pose (2017).

on individual productivity, Glaeser (1999) and De la Roca and Puga (2017), and the social returns of experienced workers (Moretti, 2004b,a).

In general, the policies should aim at compensating individuals that have valuable knowledge for sharing information that renders other individuals more productive.[3] Because if these former are not compensated in some way they will share suboptimal amounts of knowledge, even though it is in society's benefit to increase the amount of knowledge shared.

One way knowledge spillovers are often promoted is through industrial policies in attracting and subsidising businesses, see Greenstone et al. (2010); Moretti (2012); Kuijpers and Thomas (2016). As discussed previously these tools are highly contested and may increase one type of market failure, *i.e.* market power of large firms, while mitigating another market failure, *i.e.* knowledge externalities. The extent to which the market failure of knowledge externalities is corrected remains also to be seen as firms have incentives to not disclose knowledge. In particular large firms are known to internalise externalities and block access by outsiders (Chinitz, 1961) but are also suggested to be more effective in generating benefits in Feldman (2003). Future research will need to improve on measuring the welfare effects of different channels as discussed in Section 5.3.2.

Nevertheless, if one wants to correct for the externality without amplifying another market failure or transferring public money to wealthy firm owners a possible sensible solution to study would be to spend money on hiring some of the bright minds that else end up in the private sector and put them to work in the public sector in a position dedicated to more public knowledge dissemination: teaching and academic research. This in combination with lowering barriers to access these sources.

Promoting agglomeration advantages

Another market failure that needs to be addressed is related to the geographical reach of knowledge spillovers. These advantages are known to attenuate sharply with distance, likely because of the high value of time of the agents that need to meet face-to-face to transact knowledge (Arzaghi and Henderson, 2008; Rosenthal and Strange, 2008; Ellison et al., 2010). From the monocentric city model (Alonso, 1960, 1964; Muth, 1969; Mills, 1967), we know that when the benefits and transportation costs of local knowledge spillovers increase, as demonstrated in this thesis, then densities should increase.

Therefore, these results bring new prominence to the debate on local zoning policies

[3]For example older experienced workers that share knowledge with younger inexperienced workers in the framework of Glaeser (1999)'s model.

because in the real world cities do not adapt instantly to new realities as in models. This is not only because buildings are durable goods and can not easily be modified but because local agents, in particular those well-endowed, contest new development: known as NIMBYs in Glaeser (2011); Moretti (2012) or as the new urban Luddites in Florida (2017).[4]

The strongest commonality in Glaeser (2011); Moretti (2012) and Florida (2017) is their call upon policymakers to act against zoning regulations. They argue that these zoning regulations mostly cater to the growing influence of NIMBYism and that the benefits to society nor those that would be able to live there are properly taken into account. Hence, giving rise to externalities and therefore a market failure.

Opposing new housing also majorly benefits these incumbents not only by mitigating a possible disamenity but also by keeping the housing stock low and driving the prices up. Florida (2017, p.26) argues that "*urban rentiers have more to gain from increasing the scarcity of usable land than from maximizing its productive and economically beneficial uses.*" Piketty (2013) and Florida (2017) even warn for the possible societally destabilizing effects now that a larger and larger share of the wealth is absorbed by the rentier behaviour of landlords.

The societal losses due to inefficient land use itself are calculated with much detail in various works. For example, Hsieh and Moretti (2015) and Gyourko and Molloy (2015) produce some telling estimations on the U.S. suggesting, respectively, that it costs the total economy about 9% of total GDP annually and that real house prices are about 55% above their real construction costs. Relatedly, Cheshire and Hilber (2008) and Koster et al. (2012) show large societal costs due to regulatory policies in the United Kingdom, respectively, the Netherlands.

The evidence in this thesis on the rising importance of geographical proximity for knowledge spillovers adds to this support for less restrictive regulatory policies and to take more into account the benefits of those under-represented in the decision processes, *i.e.* those that are willing to live in the housing stock that is not built yet. Nonetheless, this should be particularly targeted at reducing NIMBYism and is not as straightforward as it may seem.

There are a few considerations: A first consideration is the positive externalities associated with keeping historical amenities and public spaces, which have societal value and therefore should be protected. Another consideration is that the back-

[4]NIMBY refers to "Not In My BackYard". The Luddites were a movement of textile workers in the 19th century that destroyed weaving machinery that threatened their work. It later became a term for those who generally oppose new technologies.

of-the-envelop calculations in Hsieh and Moretti (2015) are unlikely to hold in real life. It is not that easy to double the size of the city and at the same time double its liveliness. Jacobs (1969) already suggested that skyscrapers lead to different interactions compared to lower density areas to which Florida (2017, p.28) agrees by stating that "*The world's most innovative and creative places are not the high-rise canyons of Asian cities but the walkable, mixed-use neighbourhoods in San Francisco, New York, and London.*" More consideration of the functioning of agglomeration economies at the microscale is likely relevant before enacting urban development programs. In a similar line, Rodríguez-Pose (2018) argues that just increasing city size is not sufficient to increase productivity, as numerous examples show that some smaller innovative cities, like San Francisco, outcompete larger cities with older industries, like Los Angeles. Economic vitality thus also depends on the dynamics of a city and the sectors it hosts. Another consideration is brought up by Storper (2018), namely, that just lifting housing restrictions will likely lead to more high-skilled workers leaving lagging regions and migrating to more prosperous regions, thereby disadvantaging these lagging regions even more nor helping the less skilled in prosperous regions. It may also lead to larger real income advantages in prosperous regions as a decrease in housing costs will lead to larger real income increases in large cities thereby increasing spatial inequality in real income.

A final consideration, when deciding on urban development programmes, is that just facilitating access to places is not sufficient for knowledge spillovers to occur as other dimensions of proximity, *i.e.* cognitive, social, institutional, and organisational, also need to be developed (Boschma, 2005). Just increasing density or providing better infrastructure may not lead to everyone making equal use of these assets. Although, geographical proximity may allow for the development of other dimensions of proximity (Boschma, 2005) when different actors meet and interact. From this perspective, it would be interesting to combine insights from urban studies on how different groups have access and interact in spaces to be able to address their ability to develop these proximities. In particular, to consider the welfare distributional effects of such policies. How to improve research to come this kind of policy advice is discussed in Section 5.3.2.

5.2.3 Reducing income inequality

The geographical reach of knowledge spillovers discussed in this thesis also has implications for policies now that these are increasingly focussed on reducing inequalities (European Commission, 2017). A particular point of focus for policy should be the accessibility of the sources of knowledge for lower-income groups as Piketty (2013) describes that the spread of productive knowledge to these income groups is an equalizing force on income inequality. The fact that this thesis suggests that knowledge spillovers increasingly requires geographical proximity and are ever more concentrated in a smaller number of expensive cities, see Chapters 1 and 2, implies that lower-income groups have seen a reduction in the access to productive knowledge thereby hindering this equalizing mechanism. To reduce inequalities policies should therefore increase the geographical proximity of these income groups to the sources of knowledge, as well as the other dimensions of proximity of Boschma (2005). This could be done, for example, through social housing programs and by reducing barriers to education and other forms of productive knowledge embedded in firms and workers.

5.2.4 Reducing spatial inequality

Spatial inequality was one of the motivating factors to write this thesis and of growing concern among researchers and policymakers (Rodríguez-Pose, 2018; Storper, 2018). Chapter 3 has a particularly gloomy forecast with respect to spatial inequality: If one extrapolates to the future the tendency that making innovational progress requires the combination of an ever-larger number of narrowly-skilled individuals for which large cities offer advantages then at some point all complex activities will end up in a few large cities leaving all other places to whither. For these other places, it will become ever more difficult to attract complex activities as the number of assets required to develop these activities is ever increasing.

Naturally, the extrapolation of this tendency may not prove true. Technological progress is not incremental but may have drastically different repercussions based on the nature of the technologies that are developed (Helpman and Trajtenberg, 1998). For example, how the invention of electricity in the 1870s mostly complemented routine-skilled workers and that of the semiconductor in the 1970s abstract-skilled workers (Goldin and Katz, 1998). This means that technological change may reduce the need for agglomeration. One way this can happen is that technologies may decrease the number of persons required in the development of complex activities, for example by reducing the educational burden on new scholars, which means that it becomes easier for other regions to develop such an activity. The other way is by reducing the need for geographical proximity in these activities. For example, if the so-called metaverse currently being developed can substitute for real-life meetings then it would no longer make sense for agents to pay the high rents for space that large cities require.[5]

However, most signs point towards an increasing need for agglomeration in the future. Over time, Chapter 3 and Leamer and Storper (2001) note that the effect of new

[5] This also suggests that policies aiming at reducing the educational burden and/or need for geographical proximity may reduce spatial inequality.

innovations on increasing complexity and therefore promoting agglomeration is larger than the effect of new innovations on the possibility of relocating the most routine and most mobile jobs and therefore promoting deagglomeration. This means that the results of this thesis bring even more prominence to the debate on spatial inequality and what policymakers could do about it.

Policies around the topic underwent great changes over time. In the 1980s, the Netherlands, like many Western countries, made a switch from policies focussed on equity, *i.e.* reducing inequalities, to policies focussed on efficiency, *i.e.* increasing growth and the competitive advantages of the country (Raspe and van Oort, 2007).

These policies went from supporting lagging regions, by investing in declining industries and placing public sector facilities there, to supporting winning regions, by investing in the industries of tomorrow that locate predominantly in prosperous growing regions and by turning over public services to the private sector (Raspe and van Oort, 2007; Milikowski, 2020).

Relatedly, policies focussed on efficiency in terms of human capital target the stimulation of workers with high human capital, for example, the 30% tax discount for high-skilled expat workers in the Netherlands. As Chapter 3 shows these become more and more concentrated in larger cities. This suggests that the most routine and most mobile jobs that the larger cities shed through technological change, following Leamer and Storper (2001), may not offer the lagging cities that attract them a sufficiently decent livelihood if subsidies/tax breaks benefit the jobs that concentrate in large cities.

Recently, the thought that equity should be taken more strongly into account next to efficiency is gaining momentum among policymakers and advisors, where concerns about inequalities are growing (Guilluy, 2014; Autor et al., 2016; The Economist, 2016; Florida, 2017; European Commission, 2017; Le Figaro, 2018; Rodrik, 2018; Rodríguez-Pose, 2018; Storper, 2018; De Groot, 2019; Oevering and Raspe, 2020; Meijers and van Rietbergen, 2021). As even though investing in the knowledge-intensive industries and workers in a region like Amsterdam may bring about more growth and competitive advantages to the country than investing in smaller lagging regions it may not justify the increasing gap it helps bring about between the prosperous regions and lagging regions nor between high-skilled workers and middle/low-skilled workers.

However, although policies with a stronger focus on equity are becoming more popular it is not so clear yet which regional policies to enact. Within the literature, there is a divide between so-called people-based strategies and place-based strategies. People-based strategies focus on government intervention in improving the mobility of people and firms, which should lead to productive assets moving to the location where these are the most productive and lead to the largest welfare for all. On the other hand, place-based strategies focus on government intervention in improving the productivity of people and firms based on the context of the regions they are in, which should lead to each region developing their own competitive advantage and lead to the largest welfare for all (Barca et al., 2012).

Each approach sees flaws in the other approach. On the one hand, advocates of people-based approaches see place-based approaches as "*bribing*" (Moretti, 2012, p.208) production factors, like labour, to remain stuck in unproductive places. On the other hand, advocates of place-based approaches see people-based approaches, or as they sometimes call them spatially-blind approaches, as unrealistic because enacting policies that reduce market frictions will inevitably have different repercussions among people and regions due to the differences in the local contexts these operate in, thereby implying that people-based approaches are needed (Barca et al., 2012). The academic debate between the two approaches and their foundational origins is further discussed in Section 5.3.2.

For policymakers, the answer lies likely in between the two approaches. On the one hand, only people-based strategies are not sufficient because people and firms, in particular those with lower income/productivity, find increasing difficulty in moving to productive areas, as even previously ardent advocates of people-based policies agree, see Austin et al. (2018). Furthermore, regions have specific local institutional and sectoral context that means that specific programmes rather than reducing market frictions may in certain cases stimulate growth. On the other hand, only place-based strategies are not sufficient because there are too many examples of firms receiving unreasonable amounts of public money to remain/locate in unproductive places, see Moretti (2012, p.208-209), and one cannot hope that regions can sustain the same number of citizens through periods of economic downturn when there is more productive use for them elsewhere. Therefore, future policy should consist of a balance between the two approaches.

Both advocates of people-based and place-based policies agree that past place-based policies have not been successful (Martin et al., 2016; Rodríguez-Pose, 2018). But where people-based advocates see that fact as evidence that these policies do not work place-based advocates argue that the policies have only been poorly executed. policymakers should thus proceed cautiously when enacting these.

A first piece of advice from the literature is that this requires tailor-made approaches unique to each region, thereby focussing on local underused potential. As each approach is unique it is hard to give concrete advice but a must is to develop activities related to capabilities present in the region as suggested in the smart specialisation literature, see Balland et al. (2019). Particular effort should also be devoted to improving local institutions by reducing the political or economical influence of strong local actors that mostly pursue personal benefits, as described in Rodríguez-Pose et al. (2018). Another way forward for lagging regions may be to build on 'borrowed size', *i.e.* network connections with more performing regions for the urban functions it lacks, see Burger et al. (2015). Finally, the effects of policy on inequality within regions should also be taken into account. In the view and examples of Rodríguez-Pose et al. (2018) the expenditures and benefits of the projects aimed at helping lagging regions are perceived by most to end up in the pockets of the elites, which actually increases within region inequality and fuels discontent.

Enacting people-based strategies and their pitfalls have been discussed earlier in the context of promoting knowledge spillovers and agglomeration in Section 5.2.2.

The exact balance between place-based policies and people-based policies, and the balance between policies based on equity versus efficiency should depend on what societies see as a "just" distribution of welfare. It is up to us academics to provide the insights and tools to allow for these goals to be achieved. How to improve academic research on the gap between people-based and place-based approaches is discussed in Section 5.3.2; on the role of institutions and regulations in Section 5.3.2; and on welfare effects for different agents and different locations in Section 5.3.2.

5.2.5 Diversification

Chapter 4 has some more straightforward policy implications. The long-term benefits of colocation of activities is hard to take into account for agents deciding on their location. Nonetheless, this positive externality may lead to the development of new activities in the future, which can possibly set off losses in other sectors (Vernon, 1960; Rosenthal and Strange, 2004; Glaeser, 2005; Boschma, 2015).

Chapter 4 shows that having a larger diversity in sectors leads to greater diversification opportunities also in times of crisis. This comes on top of the advantage that diverse cities have in being less vulnerable to crises, as already noted by Chinitz (1961) and Frenken et al. (2007). Because the entire fortune of the city is not tied to that of a few industries. The impact of diversity on diversification may be due to the actual diversity of sectors or due to the fact that it is correlated to having less vested interests in a single industry that block new developments as suggested by Grabher (1993); Boschma (2015) and Neffke et al. (2018). This makes it hard to define the exact goal of a policy.

Nonetheless, it would make sense for single industry-oriented regions to encourage the development of other industries to better withstand economic shocks in that particular industry. In particular, it would help for the new industries to use similar workers but not have input-output linkages to the main industry. Therefore, it can more easily absorb unemployed workers when the main industry is hit without itself facing a fall in demand. Furthermore, an extra advantage is that the relatedness in type of workers is also known to increase the probability of successful diversification (Farinha et al., 2019).

Focussing on diversification to improve regional resilience is currently not part of the Horizon 2020 and smart specialisation strategy of the European Union. The programmes focus on the development of new activities by regions that can give these regions and the EU as a whole a competitive advantage but do not evaluate to what extent this leads to regional resilience through a diverse portfolio. By adding this component a contribution to future resilience can be achieved.

5.3 Limitations and future research

"*All theory depends on assumptions which are not quite true.*" (Solow, 1956, p.65). This also holds for the chapters in this thesis and the respective fields they belong to. In this section, attention is paid to possibilities for future research with regards to these assumptions divided over two subsections.

The first subsection considers assumptions revolving around the measurements of the interaction between phenomena of interest in this thesis and how these impact each other. These consist of the methods, in particular in establishing causality; choice of proxies; and variables of interest. Most of these topics already received some attention in the chapters and are familiar to authors in the field.

In contrast, the second subsection considers assumptions that result from paradigms and are generally less clearly mentioned in the chapters or on the mind of authors. Through studying, researching, meeting, and writing within an epistemic community the implicit assumptions that are shared within the community are too little questioned and may even fall out of sight for some members. Therefore, in the second subsection, a step back is taken and these paradigms are considered. A first and relatively straightforward point of interest is the difference between the two most influential fields in this thesis: urban economics and evolutionary economic geography. Second and third are insights from outside the field on the impact of institutions on spatial inequality, respectively, methods to measure the distribution of welfare between and within regions that are relevant for the field if the understanding of spatial inequality is to be improved in future research. This will also help to improve the policy advice mentioned in Section 5.2.

5.3.1 Measurements

Causality

One strong limitation of this thesis is that it cannot fully claim causality in the answers to the research questions. This is inherent to the focus on the bigger picture of trends over long time periods with data aggregated at the local level instead of the individual level.

Nevertheless, this thesis made progress on this point compared to earlier approaches on similar questions. Notably by controlling more strongly for omitted variable bias by using more fixed effects and the approach by Oster (2019) and by explicitly testing for explanations, including instrumental variable approaches, in explaining the link between changes in agglomeration economies and channels of economic change in Chapter 2. In Chapter 3, most progress is made by the use of a plethora of data sources covering many different activities, time periods and areas of the world. In Chapter 4, more attention to the regression model, functional form, causality, and robustness checks on endogeneity issues in the diversification literature were introduced.

In future research, small-scale approaches complementary to analyses of the bigger picture, as used here, can help establish a stronger sense of causality. For example, the analysis of developments in comparable regions that receive a different exogenous shock of the investigated effect in a certain time period. The challenge is finding suitable ones.

In the case of technology shocks, see Chapter 2 and Chapter 3, a possibility may be to exploit the "randomness" of break-through General Purpose Technologies (GPTs) as the semiconductor. As knowledge is "*geographically sticky*", such technology shocks are likely to affect local industries in the area of invention earlier than those further away, even if these industries are not involved in the invention of the GPT itself.

Causality in relation to diversification, as in Chapter 4, is particularly tricky as relatedness is not conceptually a cause to diversify. This is discussed in more detail in Section 5.3.2. However, unpacking the different dimensions of relatedness, such as in Farinha et al. (2019) and a more detailed study of the agents involved and processes of diversification, as in Neffke et al. (2018), may yield descriptive insights on the causes of (successful) diversification.

Proxies for mechanisms of agglomeration

This thesis is an attempt to better understand the mechanisms of agglomeration, a relatively underexplored part of the field (Duranton and Puga, 2004a; Combes and Gobillon, 2015; Davis and Dingel, 2019). Like most research in this line, this thesis builds on detecting the positive association between a measure for colocation and a proxy for interaction. This suggests that through some agglomeration mechanism interaction takes place and can therefore be the reason for colocation.[6] This section presents a number of considerations about this interpretation.

A first consideration is (the quality of) the proxies as a representation of their respective agglomeration mechanisms. Chapter 2 demonstrated that when technological relatedness is used instead of patent citations to proxy for knowledge spillovers the importance of knowledge spillovers greatly increases. Where it was the least important mechanism in Ellison et al. (2010), it is the second most important one in Chapter 2. This tells that the relevance of agglomeration mechanisms is not only dependent on their actual strength but to a large extent also on how one tries to measure these. Furthermore, the proxy for knowledge spillovers is based on patented inventions. These differ in quality and may not necessarily lead to innovation, see Carlino and Kerr (2015), but more importantly do not capture other relevant forms of knowledge like organisational/work practices or unpatented innovations like new products or software. Therefore, patents may not capture the relevant ideas for each sector, in particular for services. It is hard to make good suggestions on how to improve our proxies but borrowing ideas from other fields, as was done here by building on relatedness, may help to improve the search. Nonetheless, one will have to keep in mind that proxies are imperfect.

A second consideration is how agglomeration mechanisms are conceptualised. The fact that the importance of labour market pooling is smaller when measuring knowledge spillovers using technological relatedness instead of patent citations tells that a portion of the proxy captures the effect of the other. This is also because labour market pooling, input-output linkages, and knowledge spillovers may require the movement of people, goods, and ideas and therefore are hard to distinguish both conceptually and empirically. For these reasons, many authors currently prefer the categorisation by Duranton and Puga (2004a) in matching, sharing, and learning but the literature has not been able yet to empirically distinguish these (Combes and Gobillon, 2015).

[6] The measures of colocation and interaction are: coagglomeration and the extent to which industries employ similar workers and ideas, and buy and sell to each other in Chapter 2; geographical concentration and activity in patents, scientific articles, industry, and jobs in Chapter 3; and geographical colocation and the development of new industries in Chapter 4.

A third consideration is that when a positive association between proxy and colocation exists it does not necessarily mean that interaction actually takes place. For example, two industries that are strongly coagglomerated and employ similar technologies do not necessarily actually need to exchange ideas. In this sense, the proxies are better thought of measuring the extent to which interaction is possible.

A fourth consideration is the extent to which the positive association between proxy and colocation is reflective of the actual relevance of that determinant in location choice. By only regressing proxies of production advantages on geographical measures, the false impression is easily made that these drive location choice. However, in this thesis and certain of the works the chapters build on other factors are left out, even though these can help explain location choice. First of all, Chapter 2 does consider natural advantages through the dissimilarity indices but its role is not taken into account in Chapters 2 and 3. Also, agglomeration resulting from the home market effect and rent-seeking, see Rosenthal and Strange (2004), instead of agglomeration benefits is also not considered. Sorting is also not considered. Industries may colocate more often with those that use similar technologies over time, as posited in Chapter 2, or occupations with more education may be more often in cities, as posited in Chapter 3, because of preferences, like neighbourhood type, and not necessarily because of changes in agglomeration benefits. Related to sorting is a large literature following Glaeser et al. (2001) and Florida (2002) that consumption advantages are increasingly relevant. Taking a more nuanced stance, Storper and Scott (2009) argue that there is a strong interplay between production advantages and consumption advantages. Therefore, an interesting research avenue would be to incorporate both types of advantages to look at this interplay. Behavioural economists would also argue that location choice is not the result of a rational analysis of all locations and their advantages and disadvantages as entrepreneurs do not have perfect information and do not make rational decisions (Pred, 1967). Some other factors less often considered by economists are mentioned in Section 5.3.2.

A fifth consideration, not considered in this thesis, is the discussion to what extent agglomeration channels are actually externalities in the sense if these are (fully) priced by the market or not. In this thesis, like in Vernon (1960), agglomeration determinants are seen as (dis)advantages external to the agent. However, a large literature frames these advantages as externalities, *i.e.* advantages unpriced by the market. However, Breschi and Lissoni (2003) argue that many knowledge spillovers are paid for by for example hiring workers of neighbouring firms. Fitjar and Rodríguez-Pose (2017) interview Norwegian firms on innovation and find that most interactions are the result of purpose-built searches instead of randomly being at the same location, as would be the case if knowledge spillovers are externalities. The debate on externalities is relevant as it has strong policy implications, see the previous Section 5.2. Providers of positive, respectively, negative externalities are not fully compensated, respectively, priced for these impacts with as a result an undersupply, respectively, oversupply of these externalities, making policy intervention fruitful (Baumol, 1972).

A sixth consideration, related to the debate on externalities, is to what extent geographical colocation is sufficient to profit from the mechanisms. The positive association between geographical colocation and proxies for agglomeration economies may give the false idea that geographical colocation is sufficient to profit from these advantages. However, two agents may not be able to understand each other because of different knowledge bases. In that case, the industrial distance or cognitive proximity is too large in the words of, respectively, Rosenthal and Strange (2004) and Boschma (2005). Boschma (2005) details three other relevant dimensions of proximity for the sharing of information: organisational proximity, which relates to the coordination and hierarchy of agents exchanging information, social proximity, which relates to friendship, kinship and trust; and institutional proximity, which relates to sharing the same norms and values of conduct.

A seventh consideration is related to the spatial scale used. The chapters build on counties and aggregates of counties like MSAs. While many agglomeration mechanisms occur at a finer spatial scale, see for example Arzaghi and Henderson (2008). As already suggested by Duranton and Puga (2004a) and Carlino and Kerr (2015), it would be fruitful to exploit the growing availability of micro-data to look at the micro-scale instead of the meso-scale of agglomeration mechanisms.

5.3.2 Paradigms

Urban economics and evolutionary economic geography

A limitation for future research to some extent overcome in this thesis is the barrier between urban economics and evolutionary economic geography (EEG). During my Ph.D. thesis, I had the pleasure to work with researchers from both sides and I was surprised to sense great distance and even some aversion between the two communities.[7]

From my point of view, the images stem from unfamiliarity with each others work and

[7] As these experiences are unrecorded, I would roughly generalise the arguments overheard in saying that, one side, urban economics/economic geography, is accused of upholding unrealistic assumptions in too abstract mathematical models while the other side, EEG, is generally accused of vague theories supported by poor econometrics. This contrast is similar to the notions of Duranton and Rodríguez-Pose (2005); Duranton and Storper (2006) and the statement on the contrast between Proper Economic Geographers (PEG) and New Economic Geography (NEG) by Garretsen and Martin (2010) that "*PEG followers [are] typically dismissing NEG as irrelevant, and NEG theorists [are] viewing PEG as non-scientific*".

jargon and are in need of a revision. This unfamiliarity is nicely summarised by Martin (1999, p.83) as a "*dialogue between deaf* and has also been noted in previous analyses on a larger debate between so-called geographical economists and economic geographers, from the perspective of geographical economists, see Mulder et al. (2001); Brakman and Garretsen (2003); Overman (2004); that of economic geographers: Martin (1999); Boschma and Frenken (2006); Boschma (2015) and Boschma and Frenken (2018); but also and likely more productively in collaborative works, see: Duranton and Rodríguez-Pose (2005); Duranton and Storper (2006) and Garretsen and Martin (2010). In this thesis, two subfields of geographical economics and economic geography are strongly present, respectively: urban economics and evolutionary economic geography (EEG).

It is crucial to attempt to bring down the barrier between the two fields. Because insights from different sides lead to different understandings of agglomeration economies and therefore lead to different policy advice, in particular on the most pressing issue within the field: the rise in spatial inequality. If researchers on the topic are unable to properly weigh the arguments of each side we cannot expect policymakers to do so. Here, I will separately consider each side on the nuances needed in their view of the other and the useful insights that each can apply from the other to improve future research.

Evolutionary economic geography
It is relatively easy to consider the view of EEG on other fields, as numerous works in this field actively discuss the differences between their accomplishments and of other subfields, see for example Boschma and Frenken (2006); Boschma (2015) and Boschma and Frenken (2018).

According to Boschma and Frenken (2018), the main reason for its success in this field is that EEG is able to incorporate relational issues between firms into its analysis, as it assumes the heterogeneity of its agents. However, it is an outdated idea that urban economics does not consider heterogeneity. This may have held at the time where Nelson and Winter (1982) wrote the founding work on evolutionary economics mostly aimed at models as used by Solow (1956) but developments have been made ever since and as early as Romer (1986), as discussed by Mulder et al. (2001).[8].

Duranton and Puga (2004b, p.48) even acknowledge in their chapter on the foundations of agglomeration mechanisms that the "*heterogeneity (of workers and firms) is at the root of most if not all the mechanisms explored in this chapter*". Empirical examples also exist like Duranton and Puga (2001); Glaeser and Ponzetto (2007); McCann

[8]Also as Elisabeth Perlman once said at a conference economics is an "*imperialist*" field able to take over ideas from other fields and absorb them in its own theory.

(2008); Koster et al. (2016) and Faggio et al. (2017) where heterogeneity is taken into account in, respectively, firms, regions, households, and industries, albeit it sometimes rather crudely.[9]

Furthermore, ideas like incomplete information, see Duranton and Puga (2004b), path dependency beyond the conceptualisation as in New Economic Geography (NEG), see Vernon (1960); Duranton and Puga (2004b); Glaeser (2005) and Glaeser and Ponzetto (2007), and imperfect competition, following Dixit and Stiglitz (1977), are also present in the literature, which contrasts with the ideas of Boschma and Frenken (2006, 2018) on mainstream urban economics/economic geography. So a more nuanced and updated view of urban economics is in place.

Insights from urban economics can even be beneficial for EEG both in terms of methodology and theory. In terms of methodology, the criticism on the quality of estimation strategies overheard is to some extent justified. There is much reason for endogeneity concerns. For example, in Chapter 4, patent data are used to define crisis periods, entry to an industrial portfolio, relatedness, and the diversity of cities. Even after the many robustness analyses undertaken in the chapter, it cannot be ruled out that some mechanical relation may be influencing part of the results. This chapter is not an isolated case. In Hidalgo et al. (2007) the entry of industries to an export portfolio of a regio is regressed on a relatedness measure based on the co-occurrence of industries in export portfolios of regions. In Boschma et al. (2015) the entry of technologies to a portfolio of regions is regressed on a relatedness measure based on the co-occurrence of technologies on patents, which are also located in regions. As a result, there is a risk that mechanical relations in the data explain (part of) the association between relatedness and diversification. Although some less risk-prone approaches also exist as in Neffke et al. (2011a).

Furthermore, often strong conclusions hinting at causality are suggested in these studies. Where Chapter 4 is rather careful and refrains to posit that the results indicate causality this is not always the case. For example, Boschma (2017) and Farinha et al. (2019) already suggest in their title that relatedness drives diversification. While the estimation strategies do not allow for such a strong conclusion. Other related approaches as in Klepper (2007) also draw strong conclusions from not so

[9]As discussed further on, a deficit is that these studies model heterogeneity in very simplistic categorisations and take this heterogeneity as given and generally don't allow for agents to move between categories, which is not a problem in the used cross-sectional setting but limits the understanding on developments in agents characteristics.

strong identification strategies.[10]

The use of more rigid identification strategies as is more common in economics could benefit EEG to establish causal relationships. For example, looking for forms of (quasi-)natural experiments or diff-in-diff strategies has so far been an unventured avenue. In Chapter 4, spatial econometrics, attention to functional form, and a quasi difference-in-difference approach commonly used in urban economics was applied to reduce endogeneity concerns of entry models compared to earlier approaches in EEG.

Furthermore, the theory behind diversification could be augmented using ideas on demand and supply from economics. The greatest issue is that relatedness is also not conceptually a cause for diversification. Firms do not move into another activity because it is related to their current activity but because they sees an opportunity to make a profit, *i.e.* answer a demand in a cost-efficient manner. For example, the Amsterdam chain ijscuypje sells ice cream in the summer and stamppot (Dutch stew) in the winter not because these products are related but because there is a demand. When stating that relatedness is the driver behind diversification, as in Boschma (2017) and Farinha et al. (2019), the implicit assumption is made that demand and cost-efficient production for a product/service is given.[11]

In the current approach, only a part of the supply side, *i.e.* relevant capabilities, is taken into account. Other supply factors are not included such as the prices of inputs of land, products, capital, and labour; market structure, *e.g.* oligopoly; and product differentiation. By not including these supply factors and demand factors it is not surprising that the results in Chapter 4 show that when an average region has 100% of the related capabilities to a new technology the chance that the region develops a competitive advantage in this technology is still only 3.5% and only about 2.5% larger than the chance of entry when the region has none of the related capabilities. The development of a new specialisation is a rare event, therefore it requires more parameters than just relatedness density. Ideally one would want that when a range

[10] In Klepper (2007) the claim is made that clusters thrive through spin-off processes and that Marshallian externalities are irrelevant. He shows this by regressing the survival rate of firms in the automobile industry on among others the fact that they are located in Detroit, which proxies for Marshallian externalities. This dummy variable is then insignificant while the dummy variable for spin-off is strongly and negatively associated with the hazard of not surviving. A first problem is that not being in Detroit not naturally means that there are no Marshallian externalities in the area of the firms. Second, the survival rate is a poor measure of productivity gains due to local circumstances, which is what agglomeration economies entail. It is well known that agglomeration on the one hand leads to productivity gains but also to stronger local competition leading to more closures, known as firm selection, see Puga (2010); Combes et al. (2012). Therefore a lower survival rate does not necessarily indicate that no local productivity gains from Marshallian externalities exist.

[11] This assumption is never explicitly written down. The only time I experienced that the assumption was made explicit is when I asked Ricardo Hausmann about it and he replied that it is safe to assume that there is demand as the products/services considered in the studies are tradables.

of parameters is fully met, *i.e.* 100%, the likelihood of diversification is close to 100%.

Therefore it would be more fruitful to see relatedness as one of the many cost parameters within the diversification process. If an entrepreneur and workers have related capabilities to a new activity then the costs of switching to that activity are likely lower, *i.e.* (re)education costs and adaptations to capital are lower. Relatedness can serve as a proxy for these transition costs. However, then the firm also needs to be cost-efficient with respect to competitors and have enough demand. The firm must be able to produce the product at a sufficiently low cost and/or in a niche that other firms do not serve. Measures of the production cost parameters are therefore also useful as this factor cannot be assumed to be uniform over space nor is demand. Future research on diversification could attempt to include these factors, which are relatively well studied in economics. To complete the picture ideas on-demand ideas from managerial economics can be borrowed or from other evolutionary approaches that do consider demand-side dynamics, such as sales growth in Colombelli et al. (2014), or the approaches in Saviotti (1996).

It has to be noted that there are already attempts to bring supply and demand into EEG but these are still deficient. One approach is by relating supply to relatedness and demand to the complexity of a new activity, see Balland et al. (2019). Relatedness gives how easily a region can transition to supplying the product and complexity indicates that, by definition, few regions can produce it hence there is little competition and thus likely much demand to fulfil. However, complexity, which following the method of Hidalgo and Hausmann (2009) is defined as the fact that few diverse regions produce a product/technology, does not necessarily mean that there is actually a high demand for that product/technology just like that relatedness does not fully capture supply. Furthermore, complexity also captures part of the supply side as with rising complexity the transition costs to produce this product are likely higher. Therefore, it would be useful for future research to try and disentangle the demand and supply aspects within complexity and to consider inputs from more traditional economic measures to demand and supply.

A final development following urban economics that would be beneficial for EEG is to do more policy evaluation. Research to measure the effectiveness of different policies is omnipresent in urban economics, see the numerous examples in Section 5.3.2, but virtually absent in EEG.[12] Even though this type of study would improve policy advice.

Urban economics

[12] A notable exception is Uhlbach et al. (2017).

The fact that urban economics does not actively consider the evolutionary economic geography view does not mean that it does not have much to learn from its approaches.[13]

Where the importance of heterogeneity in agents is acknowledged by scholars in the field and taken into account in models. This is often limited to simplified dichotomies, as also noticed in a different context by Mulder et al. (2001). For example, goods-producing cities versus idea-producing cities in Glaeser and Ponzetto (2007), or high-value-added activities versus low-value-added activities in McCann (2008), or high-technology versus low-technology industries in Faggio et al. (2017). Also, the industrial distance in Rosenthal and Strange (2004) is often operationalised by the dichotomy localisation economies à la Marshall versus urbanisation economies à la Jacobs.[14]

Heterogeneity is obviously much more complex than this. The use of different forms of relatedness in measuring industrial distance as operationalised in EEG is a major stride forward in understanding the particularities of the mechanisms associated with heterogeneity and in particular in measuring the industrial distance of Rosenthal and Strange (2004). In Chapter 2 technological relatedness proved to be more effective conceptually and empirically in measuring the relationship between the technological knowledge of industries than patent citations previously used in the literature.[15]

More importantly than refining views on heterogeneity is the need to incorporate the possibility of agents moving from one category to another. The mentioned studies that consider heterogeneity, see it as a fixed state. In which the agents be it firms, industries, or regions do not move from one state to another. When it is considered, such as in Duranton and Puga (2001) where firms move from the young exploration phase of the product life cycle to the mature phase, this happens for reasons exogenous to the model, where exactly these reasons are essential to understand how agents evolve in their capabilities and needs.

Therefore, it is difficult from this framework to understand the modalities agents undertake to diversify into new activities. This is relevant as economic change "*continuously revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one*" as put by Schumpeter (1942, pp.82-83). Studying the divergence in prosperity between idea producing cities and goods producing cities, as conceptualised in Glaeser and Ponzetto (2007), is incomplete without studying the

[13]For example, during my Bachelor program and Master program, I never even heard about it. Although the difference between geographical economics and economic geography as discussed by Brakman and Garretsen (2003) was in one of the textbooks used in class.

[14]Note that the heterogeneity in households in Koster et al. (2016) is an exception as households are considered at the individual level.

[15]Another notable example is the difference in related variety and unrelated variety compared to a more general measure of diversity à la Jacobs in general as discussed in Frenken et al. (2007).

possibilities of cities to move from the production of goods to the production of ideas. Studies on this topic within urban economics are limited to case studies, see Vernon (1960) and Glaeser (2005).

It is also the understanding of the radical innovations that alter modes of production and therefore its spatial distribution that led to the breakthrough in Chapter 2 that computer-based routine-biased technological change was a channel of economic change worth testing where pecuniary transportation, traditionally suggested in the literature, failed to provide an explanation. Clearly demonstrating the usefulness of ideas often-employed in EEG to research in urban economics.

Another relevant aspect on which the field can learn from EEG is how knowledge diffuses over space. Work on the relevance of geographical proximity for transmitting knowledge in works like Gaspar and Glaeser (1998); Glaeser (2011) and Carlino and Kerr (2015) is relatively limited to the usefulness of face-to-face contact but not as developed on the exact circumstances of when this is needed as in for example Leamer and Storper (2001); Breschi and Lissoni (2001); Storper and Venables (2004); Boschma (2005) and Duranton and Storper (2008).[16] Boschma (2005) also identifies other forms of distance, namely institutional, social, and organisational that relate to productivity gains of agglomeration in addition to the temporal and industrial distance mentioned by Rosenthal and Strange (2004), which are not picked up by the urban economics literature following in the steps of Rosenthal and Strange (2004).

In general, it seems that EEG has incorporated to a larger extent useful insights from innovation studies and applying them to understand the transmission of knowledge over space. Understanding the changes in the spatial distribution of activities can hardly be complete without a broader understanding of the transmission of knowledge over space; economic change; and diversification of agents.

Part of this may be due to the narrow focus with more room for identification strategies and less for theory in economics that I experienced when writing Chapter 2, the most urban economics-like piece. Where the original 30-page theory section I wrote got smaller and smaller but more focussed over time and in the end even ended up in the appendix. In contrast, the part on robustness checks and empirical strategies only

[16] An exception within mainstream urban economics/economic geography is Arzaghi and Henderson (2008), who detail the circumstances when workers in advertisement agencies require face-to-face contact. Furthermore, in most of the literature on the difference between geographical economists and economic geographers, the second author in Storper and Venables (2004) and the first authors in Leamer and Storper (2001) and Duranton and Storper (2008) are considered geographical economists. It is telling that when these authors collaborate with an economic geographer like Michael Storper more detailed theory comes to the forefront and less on how to capture these in mathematical models allowing to more fundamentally understand face-to-face contact.

grew over time. I also noticed this difference in the advice given by scholars and my supervisors from both backgrounds throughout the different pieces I worked on during my Ph.D. Where those more strongly rooted in economics would encourage to narrow the theory down to a small part and focus on the testing, those more strongly rooted in geography would advise to make broader connections to related ideas and subjects and focus less on empirics.

As a result, and also noted by Duranton and Rodríguez-Pose (2005) economic geographers have more broad ideas being more eclectic in the theories to incorporate by writing papers that combine theory and empirics, and as Boschma and Frenken (2006) note, more pluriform in methods being both quantitative and qualitative and inductive or deductive even though the empirics are less well developed compared to those of geographical economists who on the other hand remain in narrow less contested topics and with certain exceptions favour quantitative deductive approaches.

In my opinion, this practice may end up hurting economics by challenging authors less to think out of the box and incorporate valuable ideas from outside the field as was done in this thesis. In terms of approaches, the few case studies such as Glaeser (2005) and Arzaghi and Henderson (2008) are often cited and therefore show that there is value to these approaches. On the other hand, the pitfalls of more qualitative approaches like the overuse of jargon, as also noted by Duranton and Rodríguez-Pose (2005), should be avoided.

Conclusion
All in all, both sides could do with a more nuanced view of each other's work. It is in the academic spirit to be open to other ideas regardless of where these come from. Furthermore, there are useful aspects on either side that can help improve future research on the spatial distribution of activities and help spatial sciences as a whole confront the current societal challenges. Because the greatest victim of this "*dialogue between deaf*" is there where the societal relevance of spatial science is currently the strongest: the issue of spatial inequality.

With respect to this issue, there are two schools of thought offering policy solutions, respectively, summarised under people-based policy interventions, which are generally associated with removing barriers so that labour and capital can flow to the place where these are the most efficient, and place-based policy interventions, which are generally associated with fostering economic development in a certain region with explicit consideration for the local context, for example through local industrial policies (Barca et al., 2012). The former is generally associated with urban economics with some authors such as Koster (2013); Austin et al. (2018) and Duranton and Venables (2018) defining a few conditions for exceptions to the rule.[17] The latter contains virtually all policy interventions brought forward by evolutionary economic geographers.

Although a bit generalizing it can be suggested that each approach shows its foundational origins, the first is focussed on equilibria, arguing that barriers to market clearance, such as building limits or local social benefits keep productive factors, like labour, from moving to other places to achieve their greatest productivity and therefore reducing regional income disparities. The other refuting the existence of equilibria and building on the concept of local routines and the possibility of diversification argues that there is a way for each place to develop new competitive advantages and flourish.

The shortcomings in each approach discussed above come up in this debate. Put in a generalizing way, if one assumes in models that regions move from one activity to another or lack the understanding of these capabilities then it makes sense to spend government expenditures on allowing people and capital from less performing regions to move to more performing regions. If one assumes that demand and cost-efficiency of activities is given and that regions can diversify into new activities based on former activities then it makes sense to invest in lagging regions. The most fitting solution likely lies in between the two views but requires academics on both sides to step over the boundaries of their comfort zone and think more critically on the implicit assumptions in their world views that put a limit to understanding reality.[18]

Researching to what extent conflicting assumptions hold can provide a way out of this stalemate. For example, if economists insist that people from lagging regions do not move to more innovative regions due to housing shortages, see Austin et al. (2018), and geographers that they do not move due to other reasons than housing, see Rodríguez-Pose (2018), then it makes sense to turn this issue into a research topic rather than the subject of assumptions. Other examples of conflicting assumptions that require empirical scrutiny are suggested by Storper (2018). To achieve this level of integration the line of collaborations between geographical economists and economic geographers, see Duranton and Rodríguez-Pose (2005); Duranton and Storper (2006) and Garretsen and Martin (2010), on the differences and commonalities of these fields, merits an updated follow-up.[19]

[17] Telling is the quote by Austin et al. (2018, p.2) that "*Traditionally, economists have been sceptical towards these [place-based] policies because of a conviction that relief is best targeted towards poor people not poor places*".

[18] Storper (2018) provides an overview of certain other points of discussion between the two fields in understanding spatial inequality.

[19] These analyses are a bit outdated as Duranton and Rodríguez-Pose (2005) note that in economic geography quantitative approaches are virtually absent and Garretsen and Martin (2010, p.130) state that Evolutionary Economic Geography "*is still very much in its infancy.*

To end on a positive note, there seems to be an integration of concepts between the two sides, as Hidalgo et al. (2018) and Hidalgo (2021) show that authors and concepts are compatible, as also shown in this thesis. With the rising availability of data and push for interdisciplinary thinking, future research, may bring more insights on the extent to which theories and assumptions of different fields hold and allow for the integration of the strengths of both sides. As such, paving the way for a world with scholars without clear schools of thought.

Institutions

Next to learning from different subfields within spatial sciences to understand the spatial distribution of activities, useful ideas can also be found in other fields. For example, in this thesis ideas are borrowed on the disruptive and non-incremental nature of innovation during industrial revolutions from innovation studies; on trade competition and routine-biased technological progress from labour economics; and on economic complexity from complexity science.

However, other factors are nonetheless outside the scope of this thesis, which are worth considering in future research on the spatial distribution of human activities and economies of agglomeration. These are related to institutions, most notably, regulations; and social and fiscal norms.[20] This is important because when only looking at channels of economic change like technological progress the false suggestion can be made that societies are at the mercy of factors that are hard to control or that agglomeration patterns are the natural result of a value-free and neutral economical system. This is certainly not the case.

Where this thesis focusses on the simultaneity of the great divergence in welfare over space and the computer revolution around 1980, Harvey (2006); Rodrik (2011) and Piketty (2013, 2019) would point out that around that same time drastic changes in social and fiscal norms were set in motion allowing for greater economic inequality and less state intervention. This also translated to policies supporting the most promising workers, industries and regions instead of the weaker ones both at the national level (Raspe and van Oort, 2007; Raspe et al., 2012; Milikowski, 2020) as the local level (Florida, 2017; Milikowski, 2018).

These societal choices surely also impacted locational patterns. As discussed in the introduction, the location choice of different agents is dependent on their benefits of being close to a location and how hard these benefits are to transport. In the transportation costs the value of time of an agent, often proxied by their wage, is

[20]Note that these are another type of institutions than those based on the quality of government, which are often considered in geography, see Acemoglu and Robinson (2012) and Rodríguez-Pose et al. (2018).

taking up an important part. Through regulations and public spending, the locational benefits and wages of different agents are influenced and therefore the locational patterns.

A notable recent example is how in times of little financial regulation the banking sector occupied the most expensive piece of land in the Netherlands: the Zuidas in Amsterdam with its rapid connection to the airport and inner city. After the 2008 financial crisis, the ING and ABN-AMRO banks in exchange for governmental aid faced strong regulations. Instead of setting up risky ventures in foreign countries activities were limited to the national level. As a result, the locational advantages and reductions in travel time of the Zuidas no longer outweighed the land prices and the banks decided to move their headquarters (Rooijers and van Rein, 2020; Stil, 2020).

The effects of regulations, fiscal benefits, and public expenditure on location choices remains an underdeveloped branch within the studies in urban economics/economic geography considered in this thesis. Even though it is considered in other branches of geography and from a non-geographical perspective in public economics. Future research should incorporate these effects and compare them to those of technological change and trade competition in a horse race-like setting.

Welfare distribution effects

The topic of this thesis was the dynamics of agglomeration economies in relation to channels of economic change. The motivation was the societal challenge around the growing divergence in prosperity between regions and the associated growing discontent, which is likely the greatest issue the field is interested in. To move from the changes in the microfoundations of agglomeration to these welfare questions future research should consider welfare distribution effects of agglomeration. This requires a more fundamental shift in focus within the field.

The prime topic of urban economics and economic geography is the spatial distribution of human activities across space. The quintessential question within this topic is why certain places grow/decline more strongly than others. Therefore, the many studies in the field try to identify the local building blocks of growth, like in this thesis on the economies of agglomeration, or directly test this question on aggregate growth measured by GDP, number of jobs, or wages following the line of Glaeser et al. (1992)'s "Growth in cities.", see Beaudry and Schiffauerova (2009) for an overview, or by testing alternative explanations like human capital, see Glaeser et al. (2001) and Florida (2002), or the development of regional competitive advantages, see Balland et al. (2019); Rigby et al. (2022).

However, this focus on aggregate growth is rarely questioned even though one can wonder if it is societally justifiable that there is such a strong focus on mere growth? It seems that here the field still largely upholds the optimist axiom of the 1950s that "*growth is a rising tide that lifts all boats.*" (Piketty, 2014, p.11). Where the field made remarkable progress since Solow (1956) by targeting his assumption on the exogeneity of the factors that entail the Solow residual leading to productivity growth, be it via building on the evolutionary approaches of Nelson and Winter (1982) or the new neo-classical approach of Romer (1986), it did not target the assumption on "*balanced growth*", which suggests that every societal group profits to the same degree of this productivity growth.

Many reasons are emerging that this assumption does not hold. Most notably, the structural break in the 1970s-1980s that reversed trends in income inequality and wage inequality from converging to diverging in many countries (Piketty, 2013; Alvaredo et al., 2018). In the U.S. low-skilled workers even earn less than in 1970 when corrected for inflation, fewer citizens can expect to be better off than their parents, and life expectancies are decreasing (Acemoglu and Autor, 2011; Naidu et al., 2019a). Related to this is the rise in non-employment and the associated deaths of despair, see Case and Deaton (2015); Eberstadt (2016); Austin et al. (2018); Autor et al. (2019); Pierce and Schott (2020), and the growing tensions between the establishment and the societal groups that feel left behind, see Guilluy (2014); Autor et al. (2016); Florida (2017); Austin et al. (2018); Rodrik (2018); Storper (2018); Le Figaro (2018); Rodríguez-Pose (2018); De Groot (2019).[21,22] The field is well aware that these issues are in relation with the rise in spatial inequality (Glaeser, 2011; Moretti, 2012; Austin et al., 2018; Storper, 2018; Rodríguez-Pose, 2018).

It is not that the growth in inequality between societal groups within regions has been ignored in the literature used in this thesis.[23] Or that there are not a few attempts

[21] Statistics and words fall short in describing these situations, for those who speak Dutch I can recommend the documentary on Amiens in France by Wilfred de Bruijn entitled "*Op zoek naar Frankrijk: De Franse proteststem.*"

[22] If one takes into account the current concerns about climate change then one can also add the question if future societal groups benefit equally from growth as current societal groups.

[23] For example, Glaeser and Ponzetto (2007, p.3) state that the "*heterogeneity of ability determines decreasing returns to the size of the innovative sector, and it also predicts that the economy will become more unequal if it becomes more innovative.*" and reach the conclusion further on that "*As communication costs decline and the size of the innovative sector increases, within-city inequality increases.*" or McCann (2008, p.366) who states that "*In terms of economic geography, the only aspect in which individuals are becoming more empowered by the current processes of globalisation is in terms of the increased ability of highly skilled individuals to move between locations in order to reap the rewards of their human capital. A rapidly widening income gap between high- and low-skilled individuals has already emerged within advanced economies (Scheve and Slaughter, 2007)*"

that look into inequality.[24]

However, by continuing to mainly focus on economic growth in research the inequality and social discontent associated with these issues will unlikely be resolved. Now that many authors on growth and welfare acknowledge that the fruits of progress are not equally divided over societal groups it is also time for spatial scientists to focus on other objectives.

From this perspective, it is very clear that following the advice of some of the rather bold statements in the past may have numerous side effects in terms of welfare distribution and may have actually increased inequality. For example, Moretti (2012, p.13) states that "*the best way for a city or state to generate jobs for less skilled workers is to attract high-tech companies that hire highly skilled ones.*" and Glaeser et al. (2001, p.29) state that "*Traditional cities will only succeed when they provide amenities that are attractive to high human capital residents. In principle, it may be beneficial for the poorer residents of a community for that community to attract wealthier residents. After all, it cannot help the poor to live in isolated communities filled with poverty.*", which is similar to the influential statements on attracting the so-called creative class put forward by Florida (2002).

However, there are numerous side effects to these policy suggestions that in the end may mean that the people one aims to help could be worse off. For example, the statement by Moretti (2012, p.13) is based on a regional multiplier model that shows that one high-skilled job is associated with five low-skilled jobs but one manufacturing job only with two low-skilled jobs. Moretti (2012) claims that high-skill jobs through consumption lead to more job creation for the low-skilled. This may make one jump to the conclusion that investing in high-skilled jobs will automatically improve the fate of the less fortunate. However, the world is more complex than this and there are multiple channels that impact welfare. For example, if high-skilled workers are stimulated then their innovations may displace middle-skilled manufacturing workers through automation, lead to an increase in the market power of the firms that employ these workers or lead to increases in local rents, which do not benefit those that are not high-skilled workers or owners of innovative firms and housing stock. On the other hand, the less skilled may profit if the influx of high-skilled leads to a safer environment, more local prosperity due to a stronger local competitive advantage, cheaper products, or through increased social mobility for them or their children due

[24] Research on this actual topic has been rather limited. Some notable exceptions are: the local effects on social mobility, see Chetty and Hendren (2018); the rising inequality in cities, see Baum-Snow and Pavan (2013); and the local trickle-down effects of technology investments, see Lee and Rodríguez-Pose (2016).

to local learning effects.[25]

These side effects are important as the resulting policy suggestions from Glaeser et al. (2001); Florida (2002); Moretti (2012) and many others from the literature focussed on growth, innovation, and human capital have likely helped foster the rise in inequality, as they provide support for among others: the gentrification policies described in Florida (2017) and Milikowski (2018) that disproportionally benefit the rich/high-skilled and lead to the social exclusion of other groups; the changing social and fiscal norms that allow a small share of individuals to claim a larger and larger share of the world's income (Piketty, 2013); and the rat race of competitiveness between countries, respectively, regions to attract anchor firms, and capital, to the profit of the few that control these assets, see Rodrik (2011), respectively, Moretti (2012). Taking it a step further, to some extent one could even argue that the literature fits a so-called neo-liberal agenda in which reducing market frictions is put forward as bringing welfare to all, see Harvey (2006). Even though inequality tends to increase when markets become competitive (Harvey, 2006; Piketty, 2013).[26]

All together these policies have likely increased the market power of high-skilled workers, which was already increasing due to technological change and globalisation (Brynjolfsson and Hitt, 2000; Autor et al., 2003, 2015). As Chapter 3 shows that high-skilled workers are disproportionally located in larger cities it may be that policy suggestions from the literature have exacerbated the very problem it aims at solving, namely, leading spatial inequality to increase. However, we cannot be sure if we do not fully map all side effects of policy suggestions.

Therefore, our mission as spatial scientists should not end when having identified the factors that contribute to local growth but when we have identified to what extent changes in these factors can be governed and to which welfare changes for different

[25]Note that Glaeser et al. (2001, p.29) do mention an important premise to their statement in a footnote namely "*However, it is not obvious that if cities increasingly work to attract the highly educated the poor on net will be better off. The outcome would depend on whether positive spillovers arising from the presence of highly educated neighbors outweigh rent increases for the poor.*". These kinds of statements are too few in number and visibility in the literature and have not gotten the attention they deserve.

[26]This narrative is most clear in the so-called people-based approach to spatial inequality, where calls are made to reduce market frictions to migration like housing shortages and local social benefits, see Austin et al. (2018). But it is also present in the place-based approach where the development of regional competitive advantages is key, see Balland et al. (2019). Like in the narrative described by Harvey (2006), this emphasises that regions are unsuccessful because of a lack of competitiveness, *i.e.* unattractive for investment in innovation, rather than choices made in how the economic system is organised. Although it is not hard to claim that the language that an author like Harvey (2006) uses is not politically neutral, the point that the narrative of many pieces in our field is not neutral as well remains valid and authors should consider the implications of the larger frameworks on how research results are used in the real world.

societal groups these lead.[27] To establish this three steps need to be taken: defining and identifying different social groups in the data; identifying and estimating the channels through which the welfare of these societal groups are impacted following changes in local factors of growth; estimating the extent to which public spending can alter this distribution.

To make these extra steps there are luckily numerous ideas and concepts within and outside the field that may help to rise to this complicated challenge of understanding inequalities. An ideal end product would be cost-benefit analyses, that evaluate pros and cons, even those unpriced by the market, per societal group in comparison to public spending in policies. This is already to some extent common in Dutch infrastructural cost-benefit analyses, which include the travel time benefits for different societal groups at different locations.[28]

Identifying societal groups
For the first step, inspiration can be drawn from the many different attributes that have been studied in labour economics, public economics, and urban studies; such as education, labour tasks, age, gender, ethnicity, location, spoken language etc.

To move from aggregate welfare measures to group-specific welfare measures one can for example look to labour economics, which has a long history of calculating the impact of economic changes, like changes in trade policy or technological change, on the wages of different types of workers. A different kind of approach from complexity economics based on simulations is to use agent-based modelling, as suggested by Arthur (2021).

Identifying and measuring welfare channels
The greatest challenge is identifying and measuring the possible side effects of policy suggestions or the working of local growth mechanisms. Where most academics take up research questions by attempting to isolate the effect of X on Y, the reality is that the world is a more complex system of interconnected phenomena. As such, influencing X to influence Y may also lead to all sorts of side effects.

A large part of these channels may be unpriced and therefore involves estimating the importance of positive and negative externalities. These are hard to estimate as by

[27] Following the line of reasoning in Piketty (2014, p.264) the key issue is not to eradicate inequalities but to justify them. For example, encouraging a public policy on promoting a certain factor of growth that leads to an increase in overall welfare and only a minor increase in inequality may be totally acceptable for those elected by society to enact such a policy. However, without studying these dynamics of inequalities there is no information on the magnitudes of public expenditure required and changes in the distribution of welfare to act upon.

[28] The example of Dutch cost-benefit analyses should not be too closely followed as the costs are often very poorly calculated. As notoriously shown by the example of the Noord-Zuid metro line in Amsterdam, which was projected at 681 million euros but in the end cost 3.1 billion euro.

definition these are not priced by the market. *Dynamic* externalities are particularly hard to grasp and quantify as they only lead to benefits in the far future such as diversification opportunities when main industries fall, as discussed in Chapter 4, Vernon (1960); Glaeser (2005) and Boschma (2015). Although for dynamic externalities on worker productivity progress has been made in De la Roca and Puga (2017). For *static* externalities, hedonic pricing approaches offer possibilities to estimate the gains captured in housing or office rents, as in the example of the shopping mall given by Feldman (2003) or more recent elaborate approaches like Koster et al. (2012) and Dericks and Koster (2016). Other approaches consist of measuring externalities are based on gains by neighbouring incumbent firms after policy enactment, see Greenstone et al. (2010); extrapolating from wage differences, see Hsieh and Moretti (2015), or estimating the difference between marginal costs and marginal benefits, see Gyourko and Molloy (2015). These examples show that the identification strategies commonly used in urban economics are particularly useful in estimating these welfare gains/losses. Although the equation of value and price/wage/revenue in these studies is questionable, as higher wages or higher revenues do not necessarily mean that more value is created in the economy (Mazzucato, 2019). Furthermore, the future studies should consider multiple externalities simultaneously and their effects on different societal groups instead of aggregate or average productivity growth.[29]

This also means understanding to what extent each group has access to these externalities, an important issue that is underresearched in the field. As discussed in Section 5.3.2, innovation studies and evolutionary approaches have already made considerable advances in how ideas move within networks and are more productively used by some than others, see for example Kemeny et al. (2016), but these differences are not yet strongly linked to an inequalities perspective, *i.e.* to which societal groups the persons in the networks belong.

For inspiration, urban studies contain numerous attempts and descriptions on how and by whom space is used and can therefore interact, see for example Florida (2017) and Milikowski (2018) for an overview. From this literature the irrefutable image arises that knowledge is not such a *public* good as often assumed in the literature on growth and agglomeration externalities, see Romer (1986); Rosenthal and Strange (2004); Raspe and van Oort (2007). Despite the epistemic barriers, there is great advantage in combining insights from the two sides as urban studies does not consider the welfare gains that economic geography does consider. Or as Florida (2017, p.259) puts it we need to see the entire picture, which "*neither urban economics, with its traditional*

[29] In transportation economics some examples exist of assessing multiple externalities simultaneously, see Arnott (2007) and Hörcher et al. (2020).

preoccupation with concentrated advantage, nor urban sociology, with its preoccupation with concentrated disadvantage" can see.

Another step forward is to consider other channels of welfare than just externalities, often limited to agglomeration externalities in the field and this thesis. Insights from public economics suggest that welfare changes can occur through other market failures. Notably, market power likely impacts welfare outcomes of propositions on externalities by spatial scientists. For example, public expenditure to stimulate local knowledge spillovers, a sort of Pigouvian subsidy on a positive externality, is often reaped by large multinationals that hold market power such as the Amazon HQ2 bid, Danone in the Netherlands, see Kuijpers and Thomas (2016) or the many examples by Moretti (2012, pp.208-209). This means that attempts at correcting one market failure, *i.e.* a positive externality, leads to the amplification of another type of market failure, *i.e.* market power. Let alone that it also redistributes welfare towards the owners of these companies, which may not sit well with those that enact such a policy.

Another approach on mapping multiple welfare channels may also be found in complexity economics that actively tries to move away from a simple system approach in which the focus is on the effect of X on Y to a complex systems approach in which many factors are interconnected. These types of models have the advantage that they can incorporate a great variety of agents and types of interactions but the disadvantage is that they are less empirically tractable compared to models used in economics. Arthur (2021) gives some introduction and examples of complexity economics, although the image given here of mainstream economics can be a bit more nuanced, as also suggested in a similar context by Naidu et al. (2019b).

Public intervention

The overview of welfare channels can be converted into cost-benefit analyses that should also give insight into the extent to which public intervention is warranted. Peculiarly, even the reviews on agglomeration externalities, Rosenthal and Strange (2004); Puga (2010) and Combes and Gobillon (2015), hardly mention how to correct for these market failures let alone the conditions in which public action is justified. To increase the societal impact of the knowledge on agglomeration and geography the extent to which policies can influence should be centrepiece, see also the inaugural lecture of Hans Koster.

Most inspiration on calculating overall welfare gains of the different channels may likely come from public economics.[30] Concepts like social welfare functions, Pareto efficiency and Rawlsian utility are omnipresent in the microeconomics courses taught to bachelor students but when these students become researchers specialised in urban economics/economic geography these concepts end up on the bookshelf. These could improve the measurement of welfare changes and help spatial scientists to take the necessary extra step.

Conclusion
Much of the advice for future research asks spatial scientists to be more aware of ideas in related fields. The islands that groups of academics form allow for the specialisation and collaboration necessary to make progress in a well-defined area. However, the interactions between authors and conference participants within a field also resemble echo chambers where shared assumptions are questioned too little and one own's view of the world is confirmed.[31] Breaking out of this comfort zone requires questioning ourselves and bringing down the epistemic barriers to other fields by learning their jargon and worldview. This does not require becoming an expert in the other fields but knowing enough to understand each other and collaborate. This is by no means easy but is essential for more societal impactful research.

All in all, a more humble and self-reflecting approach to research will benefit our understanding of the world. After all, as notably discussed in Chapter 3, the world is a complex system in which phenomena are interconnected beyond our comprehension as it is impossible to hope that a single person can have all the knowledge to fully understand the entire problem at hand nor all the real-world consequences of the conclusions of our research.

[30] Although one can argue that the comments on the awareness of the distribution of welfare gains also applies to the models used in microeconomics, where often a focus is on total aggregate welfare rather than that of individual societal groups. For example, the distributional concerns for not taxing a monopolist in Buchanan (1969) or the premise that the distribution of property rights does not matter as total social welfare outcomes will be the same in Coase's theorem although these distributional concerns matter a lot for the welfare of the individual agents.

[31] This line of thought is somewhat related to the two opposing views by Smith (1776), discussed by West (1964), on the division of labour, which on the one hand allows individuals to become very adept at certain tasks but on the other hand reduces their capacity to experiment solutions for new problems. Although in the context of Smith (1776) this relates to persons whose job has been reduced to continuously executing a small number of tasks.

Bibliography

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D. and Robinson, J. (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. Crown Business, New York.

Adoue, F. (2016). Mobilité connectée et continuité spatio-temporelle des activités. *Reseaux*, 200(6):87–115.

Ahlgren, P., Jarneving, B., and Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, 54(6):550–560.

Aleksandrova, E., Behrens, K., and Kuznetsova, M. (2020). Manufacturing (co)agglomeration in a transition country: Evidence from Russia. *Journal of Regional Science*, 60(1):88–128.

Alonso, W. (1960). A theory of the urban land market. *Papers in Regional Science*, 6(1):149–157.

Alonso, W. (1964). *Location and Land Use. Toward a General Theory of Land Rent.* Harvard University Press, Cambridge, MA.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). World inequality report 2018. Technical report, World inequality lab.

Amable, B. (2000). Institutional complementarity and diversity of social systems of innovation and production. *Review of International Political Economy*, 7(4):645–687.

Anas, A., Arnott, R., and Small, K. A. (1998). Urban Spatial Structure. *Journal of Economic Literature*, 36(3):1426–1464.

Angrist, J. D. and Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

Arnott, R. (2007). Congestion tolling with agglomeration externalities. *Journal of Urban Economics*, 62(2):187–203.

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies*, 29(3):155–173.

Arthur, W. B. (2021). Foundations of complexity economics. *Nature Reviews Physics*, 3(2):136–145.

Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue. *Review of Economic Studies*, 75(4):1011–1038.

Audretsch, D. B. and Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *American Economic Review*, 86(3):630–640.

Austin, B., Glaeser, E. L., and Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, 2018(1):151–255.

Autor, D., Dorn, D., and Hanson, G. (2019). When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men. *American Economic Review: Insights*, 1(2):161–178.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2016). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. Technical report, National Bureau of Economic Research, Cambridge, MA.

Autor, D. H. (2019). Work of the Past, Work of the Future. *AEA Papers and Proceedings*, 109:1–32.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–1597.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The Geography of Trade and Technology Shocks in the United States. *American Economic Review*, 103(3):220–225.

Autor, D. H., Dorn, D., and Hanson, G. H. (2015). Untangling Trade and Technology: Evidence from Local Labour Markets. *Economic Journal*, 125(584):621–646.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Balland, P.-A. (2016). EconGeo: Computing Key Indicators of the Spatial Distribution of Economic Activities.

Balland, P.-A., Boschma, R., Crespo, J., and Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9):1252–1268.

Balland, P.-A. and Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, 93(1):1–23.

Balland, P.-A., Rigby, D. L., and Boschma, R. (2015). The technological resilience of US cities. *Cambridge Journal of Regions, Economy and Society*, 8(2):167–184.

Barca, F., McCann, P., and Rodríguez-Pose, A. (2012). The Case for Regional Development Intervention: Place-based versus Place-neutral Approaches. *Journal of Regional Science*, 52(1):134–152.

Bartelsman, E. J. and Gray, W. (1996). The NBER Manufacturing Productivity Database. *NBER working paper series, 205*.

Bathelt, H., Malmberg, A., and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1):31–56.

Baum-Snow, N. and Pavan, R. (2013). Inequality and City Size. *The Review of Economics and Statistics*, 95(5):1535–1548.

Baumol, W. J. (1972). On Taxation and the Control of Externalities. *The American Economic Review*, 62(3):307–322.

Beaudry, C. and Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, 38(2):318–337.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 75(299):493–517.

Behrens, K. (2016). Agglomeration and clusters: Tools and insights from coagglomeration patterns. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(4):1293–1339.

Behrens, K., Mark Brown, W., and Bougna, T. (2018). The world is not yet flat: Transport costs matter! *Review of Economics and Statistics*, 100(4):712–724.

Berger, T. and Frey, C. B. (2016). Did the Computer Revolution shift the fortunes of U.S. cities? Technology shocks and the geography of new jobs. *Regional Science and Urban Economics*, 57:38–45.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2006). Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of U.S. manufacturing plants. *Journal of International Economics*, 68(1):219–237.

Berry, C. R. and Glaeser, E. L. (2005). The divergence of human capital levels across cities. *Papers in Regional Science*, 84(3):407–444.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17):7301–7306.

Bloom, N., Draca, M., and Van Reenen, J. (2016). Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity. *The Review of Economic Studies*, 83(1):87–117.

BLS (1981). The National Industry-Occupation Employment Matrix, 1970, 1978, and Projected 1990, Vols. I and II, Bulletin 2086. Technical report.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1):61–74.

Boschma, R. (2015). Towards an Evolutionary Perspective on Regional Resilience. *Regional Studies*, 49(5):733–751.

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

Boschma, R., Balland, P.-A. A., and Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24(1):223–250.

Boschma, R. and Lambooy, J. (1999). The prospects of an adjustment policy based on collective learning in old industrial regions. *GeoJournal*, 49(4):391–399.

Boschma, R., Martín, V., and Minondo, A. (2017). Neighbour regions as the source of new industries. *Papers in Regional Science*, 96(2):227–245.

Boschma, R. A. (1999). The rise of clusters of innovative industries in Belgium during the industrial epoch. *Research Policy*, 28(8):853–871.

Boschma, R. A. (2008). Regional innovation policy. In *Micro-foundations for innovation policy*, pages 315–342. Amsterdam University Press, Amsterdam.

Boschma, R. A. and Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6(3):273–302.

Boschma, R. A. and Frenken, K. (2018). Evolutionary Economic Geography. In Clarke, G., Feldman, M. P., Gertler, M., and Wojcik, D., editors, *Economic Geography*, pages 213–229. Oxford University Press, Oxford.

Bozio, A., Breda, T., and Guillot, M. (2016). Taxes and Technological Determinants of Wage Inequalities: France 1976-2010.

Brakman, S. and Garretsen, H. (2003). Rethinking the "New' Geographical Economics. *Regional Studies*, 37(6-7):637–648.

Breschi, S. and Lissoni, F. (2001). Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Industrial and Corporate Change*, 10(4):975–1005.

Breschi, S. and Lissoni, F. (2003). Mobility and Social Networks: Localised Knowledge Spillovers Revisited.

Bresnahan, T. F. and Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics*, 65(1):83–108.

Bristow, G. and Healy, A. (2014). Regional Resilience: An Agency Perspective. *Regional Studies*, 48(5):923–935.

Bristow, G. and Healy, A. (2018). Innovation and regional economic resilience: an exploratory analysis. *The Annals of Regional Science*, 60(2):265–284.

Brueckner, J. K., Thisse, J. F., and Zenou, Y. (1999). Why is Central Paris Rich and Downtown Detroit Poor? An Amenity-based Theory. *European Economic Review*, 43(1):91–107.

Brynjolfsson, E. and Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4):23–48.

Buchanan, J. M. (1969). External Diseconomies, Corrective Taxes, and Market Structure. *The American Economic Review*, 59(1):174–177.

Burger, M. J., Meijers, E. J., Hoogerbrugge, M. M., and Tresserra, J. M. (2015). Borrowed Size, Agglomeration Shadows and Cultural Amenities in North-West Europe. *European Planning Studies*, 23(6):1090–1109.

Cainelli, G., Ganau, R., and Modica, M. (2018). Industrial relatedness and regional resilience in the European Union. *Papers in Regional Science*.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Is Changing Our Lives*. Harvard Business School Press, Boston, MA.

Camus, A. (1953). La maison du peuple - Préface. In Guilloux, L., editor, *La maison du peuple*, pages i–iii. Grasset et Fasquelle, Paris.

Caragliu, A., De Dominicis, L., and De Groot, H. L. F. (2016). Both Marshall and Jacobs were right! *Economic Geography*, 92(1):87–111.

Card, D. and Krueger, A. B. (1994). A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4):773–793.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and Innovation. In *Handbook of Regional and Urban Economics*, volume 5, pages 349–404. Elsevier B.V.

Carlsson, B. and Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of Evolutionary Economics*, 1(2):93–118.

Case, A. and Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, 112(49):15078–15083.

Castaldi, C., Frenken, K., and Los, B. (2015). Related Variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting. *Regional Studies*, 49(5):767–781.

Catin, M. and Van Huffel, C. (2019). Disparités de croissance d'emploi des aires urbaines françaises et processus de métropolisation. *Revue d'Économie Régionale & Urbaine*, (4):749–779.

Cheshire, P. C. and Hilber, C. A. L. (2008). Office Space Supply Restrictions in Britain: The Political Economy of Market Revenge. *Economic Journal*, 118(529):F185–F221.

Chetty, R. and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chinitz, B. (1961). Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review*, 51(2):279–289.

Christopherson, S., Michie, J., and Tyler, P. (2010). Regional resilience: theoretical and empirical perspectives. *Cambridge Journal of Regions, Economy and Society*, 3(1):3–10.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *The American Economic Review*, 86(1):54–70.

Clark, J., Freeman, C., and Soete, L. (1981). Long waves, inventions, and innovations. *Futures*, 13(4):308–322.

Colombelli, A., Krafft, J., and Quatraro, F. (2014). High-growth firms and technological knowledge: do gazelles follow exploration or exploitation strategies? *Industrial and Corporate Change*, 23(1):261–291.

Combes, P.-P. and Duranton, G. (2006). Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1):1–28.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating Agglomeration Economies with History, Geology, and Worker Effects. In Glaeser, E. L., editor, *Agglomeration Economics*. The University of Chicago Press.

Combes, P.-P. P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics Vol.5*, volume 5, pages 247–348. Elsevier, Amsterdam.

Coombs, R., Saviotti, P., and Walsh, V. (1987). *Economics and technological change*. Rowman & Littlefield, Totowa.

Cortinovis, N., Xiao, J., Boschma, R., and van Oort, F. G. (2017). Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography*, 17(6):1179–1208.

Crescenzi, R., Luca, D., and Milio, S. (2016). The geography of the economic crisis in Europe: national macroeconomic conditions, regional structural factors and short-term economic performance. *Cambridge Journal of Regions, Economy and Society*, 9(1):13–32.

Cuadrado-Roura, J. J. R., Martin, R., and Rodríguez-Pose, A. (2016). The economic crisis in Europe: urban and regional consequences. *Cambridge Journal of Regions, Economy and Society*, 9(1):3–11.

Davis, D. R. and Dingel, J. I. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–170.

Dawley, S. (2014). Creating New Paths? Offshore Wind, Policy Activism, and Peripheral Region Development. *Economic Geography*, 90(1):91–112.

De Groot, H. L. F. (2019). Groei & krimp: uitdagingen voor governance en solidariteit. In *Land in samenhang - Krimp en regionale kansengelijkheid*, pages 30–49. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, Den Haag.

De la Roca, J. and Puga, D. (2017). Learning by Working in Big Cities. *Review of Economic Studies*, 84(1):106–142.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Dericks, G. and Koster, H. R. A. (2016). Density Restrictions and the Urban Spatial Structure: Evidence from Blitz Bombings. *Mimeo, Oxford University.*

Dijkstra, L., Garcilazo, E., and McCann, P. (2015). The effects of the global financial crisis on European regions and cities. *Journal of Economic Geography*, 15(5):935–949.

Diodato, D., Neffke, F., O'Clery, N., and O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, 106:1–26.

Diodato, D. and Weterings, A. B. (2015). The resilience of regional labour markets to economic shocks: Exploring the role of interactions among firms and workers. *Journal of Economic Geography*, 15(4):723–742.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308.

Dosi, G. (1984). *Technical change and industrial transformation: the theory and an application to the semiconductor industry.* The Macmillan Press LTD, London and Basingstoke.

Dosi, G., Freeman, C., Nelson, R., Silverberg, G., and Soete, L. (1988). *Technical change and economic theory.* Pinter, London.

Duijn, J. V. (1983). *The long wave in economic life.* George Allen & Unwin, London.

Dumais, G., Ellison, G., and Glaeser, E. L. (2002). Geographic Concentration as a Dynamic Process. *Review of Economics and Statistics.*

Duranton, G. and Jayet, H. (2011). Is the division of labour limited by the extent of the market? Evidence from French cities. *Journal of Urban Economics*, 69(1):56–71.

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and Trade: Evidence from the US. *The Review of Economic Studies*, 81(2):681–724.

Duranton, G. and Overman, H. G. (2005). Testing for Localization Using Micro-Geographic Data. *The Review of Economic Studies*, 72(4):1077–1106.

Duranton, G. and Puga, D. (2000). Diversity and Specialisation in Cities: Why, Where and When Does it Matter? *Urban Studies*, 37(3):533–555.

Duranton, G. and Puga, D. (2001). Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products. *American Economic Review*, 91(5):1454–1477.

Duranton, G. and Puga, D. (2004a). Micro-foundations of urban agglomeration economies. *Economic Policy*, 4(04):2063–2117.

Duranton, G. and Puga, D. (2004b). Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics Vol.4.*

Duranton, G. and Puga, D. (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics*, 57(2):343–370.

Duranton, G. and Rodríguez-Pose, A. (2005). When economists and geographers collide, or the tale of the lions and the butterflies. *Environment and Planning A*, 37(10):1695–1705.

Duranton, G. and Storper, M. (2006). Agglomeration and growth: a dialogue between economists and geographers. *Journal of Economic Geography*, 6(1):1–7.

Duranton, G. and Storper, M. (2008). Rising trade costs? Agglomeration and trade with endogenous transaction costs. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(1):292–319.

Duranton, G. and Venables, A. J. (2018). Place-Based Policies: principles and developing country applications. *NBER working paper series*, (24562).

Eberstadt, N. (2016). *Men without work: America's invisible crisis.* Templeton Foundation Press, West Conshohocken.

Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5):889.

Ellison, G. and Glaeser, E. L. (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review: Papers and Proceedings*, 89(2):311–327.

Ellison, G., Glaeser, E. L., and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(6):1195–1213.

Essletzbichler, J. (2007). Diversity, stability and regional growth in the United States, 1975–2002. In Frenken, K., editor, *Applied evolutionary economics and economic geography*, chapter 10. Edward Elgar, Cheltenham.

Essletzbichler, J. (2015). Relatedness, Industrial Branching and Technological Cohesion in US Metropolitan Areas. *Regional Studies*, 49(5):752–766.

European Commission (2017). Addressing inequalities. *European semester thematic factsheet.*

Evangelista, R. (2000). Sectoral Patterns Of Technological Change In Services. *Economics of Innovation and New Technology*, 9(3):183–222.

Evenhuis, E. (2017). Institutional change in cities and regions: a path dependency approach. *Cambridge Journal of Regions, Economy and Society*, 10(3):509–526.

Faggio, G., Silva, O., and Strange, W. C. (2017). Heterogeneous Agglomeration. *Review of Economics and Statistics*, 99(1):80–94.

Faggio, G., Silva, O., and Strange, W. C. (2020). Tales of the city: what do agglomeration cases tell us about agglomeration in general? *Journal of Economic Geography*, 20(5):1117–1143.

Farinha, T., Balland, P.-A., Morrison, A., and Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation*, 26(9):988–1022.

Feldman, M. (2003). The locational dynamics of the us biotech industry: knowledge externalities and the anchor hypothesis. *Industry and Innovation*, 10(3):311–329.

Filippetti, A. and Archibugi, D. (2011). Innovation in times of crisis: National systems of innovation, structure, and demand. *Research Policy*, 40(2):179–192.

Fingleton, B., Garretsen, H., and Martin, R. (2012). Recessionary shocks and regional employment: Evidence on the resilience of u.k. regions. *Journal of Regional Science*, 52(1):109–133.

Finley, M. I. (1973). *The ancient economy.* University of California Press, Berkeley.

Fitjar, R. D. and Rodríguez-Pose, A. (2017). Nothing is in the Air. *Growth and Change*, 48(1):22–39.

Fleming, L. and Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30(7):1019–1039.

Florida, R. (2002). *The rise of the creative class.* Basic Books, New York.

Florida, R. (2017). *'The New Urban Crisis.* Basic Books, New York.

Fratesi, U. and Perucca, G. (2018). Territorial capital and the resilience of European regions. *The Annals of Regional Science*, 60(2):241–264.

Freeman, C., Clark, J., and Soete, L. (1982). *Unemployment and technical innovation: a study of long waves and economic development.* Pinter, London.

Freeman, C. and Perez, C. (1988). *Structural crises of adjustment: business cycles.* Pinter, London.

Freeman, R. and Hollomon, J. H. (1975). The Declining Value of College Going. *Change: The Magazine of Higher Learning*, 7(7):24–62.

Frenken, K. (2006). Technological innovation and complexity theory. *Economics of Innovation and New Technology*, 15(2):137–155.

Frenken, K., van Oort, F. G., and Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5):685–697.

Friedman, T. L. (2007). *The World Is Flat: A Brief History of the Twenty-First Century*. Picador, New York, 3rd editio edition.

Garreau, J. (1991). *Edge city: Life on the new frontier*. Doubleday, New York.

Garretsen, H. and Martin, R. (2010). Rethinking (New) Economic Geography Models: Taking Geography and History More Seriously. *Spatial Economic Analysis*, 5(2):127–160.

Gaspar, J. and Glaeser, E. L. (1998). Information Technology and the Future of Cities. *Journal of Urban Economics*, 43(1):136–156.

Geroski, P. A. and Walters, C. F. (1995). Innovative Activity over the Business Cycle. *The Economic Journal*, 105(431):916.

Glaeser, E. L. (1999). Learning in Cities. *Journal of urban Economics*, 46(2):254–277.

Glaeser, E. L. (2005). Reinventing Boston: 1630–2003. *Journal of Economic Geography*, 5(2):119–153.

Glaeser, E. L. (2011). *Triumph of the City: How our Greatest Invention makes US Richer, Smarter, Greener, Healthier and Happier*. Penguin Press, New York.

Glaeser, E. L. and Gyourko, J. (2007). Arbitrage in Housing Markets. *NBER working paper series, 13704*.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126.

Glaeser, E. L. and Kohlhase, J. E. (2004). Cities, Regions and the Decline of Transport Costs. *Papers in Regional Science*, 83(1):197–228.

Glaeser, E. L., Kolko, J., and Saiz, A. (2001). Consumer City. *Journal of Economic Geography*, 1(1):27–50.

Glaeser, E. L. and Ponzetto, G. A. (2007). Did the Death of Distance Hurt Detroit and Help New York? Technical report, National Bureau of Economic Research, Cambridge, MA.

Goldin, C. and Katz, L. F. (1998). The Origins of Technology-Skill Complementarity. *The Quarterly Journal of Economics*, 113(3):693–732.

Goldschlag, N., Lybbert, T. J., and Zolas, N. J. (2020). Tracking the technological composition of industries with algorithmic patent concordances. *Economics of Innovation and New Technology*, 29(6):582–602.

Goos, M., Manning, A., and Salomons, A. (2009). Job polarization in Europe. *American Economic Review*, 99(2):58–63.

Grabher, G. (1993). The weakness of strong ties: the lock-in of regional development in the Ruhr area. In Grabher, G., editor, *The Embedded Firm*, pages 255–277. Routledge, London.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3):536–598.

Griliches, Z. (1981). Market value, R&D, and patents. *Economics Letters*, 7(2):183–187.

Groot, S. P. T. and De Groot, H. L. F. (2011). Wage inequality in the Netherlands: Evidence, trends and explanations. *CBP discussion paper*, (186):1–30.

Guilluy, C. (2014). *La France périphérique. Comment on a sacrifié les classes populaires.* Flammarion, Paris.

Gyourko, J. and Molloy, R. (2015). Regulation and Housing Supply. In Duranton, G., Vernon Henderson, J., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 1289–1337. Elsevier.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *NBER working paper series, 8498.*

Hall, P. A. and Soskice, D. (2001). *Varieties of Capitalism. The Institutional Foundations of Comparative Advantage.* Oxford University Press, New York.

Hanlon, W. W. and Miscio, A. (2017). Agglomeration: A long-run panel data approach. *Journal of Urban Economics*, 99:1–14.

Harding, D. and Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics*, 49(2):365–381.

Harvey, D. (2006). Neo-liberalism as creative destruction. *Geografiska Annaler, Series B: Human Geography*, 88(2):145–158.

Hassink, R. (2005). How to unlock regional economies from path dependency? From learning region to learning cluster. *European Planning Studies*, 13(4):521–535.

Hausmann, R., Hidalgo, C. A., Coscia, M., Simoes, A., and Yildirim, M. A. (2014). *The Atlas of Economic Complexity.* MIT Press, Cambridge, MA.

Hazir, C. S., Bellone, F., and Gaglio, C. (2019). Local product space and firm-level churning in exported products. *Industrial and Corporate Change*, 28(6):1473–1496.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, 43(1):59–74.

Helpman, E. and Trajtenberg, M. (1998). Diffusion of General Purpose Technologies. In Helpman, E., editor, *General purpose technologies and economic growth.*, pages 85–120. MIT Press, Cambridge, MA.

Helsley, R. and Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics*, 20(2):189–212.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3(2):92–113.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E. L., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., and Zhu, S. (2018). The Principle of Relatedness. In *Unifying Themes in Complex Systems IX*, pages 451–457. Springer, Cham.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of the Sciences of the United States of America*, 106(26):10570–10575.

Hidalgo, C. A., Kilinger, B., Barabási, A.-L., and Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science*, 317(July):482–487.

Hill, E., Clair, T., Wial, H., Wolman, H., Atkins, P., Blumenthal, P., Ficenec, S., and Friedhoff, A. (2012). Economic shocks and regional economic resilience. In Pindus, N., Weir, M., Wial, H., and Wolman, H., editors, *Building Resilient Regions: Urban and Regional Policy and Its Effects*, pages 193–274. Brookings Institution Press, Washington, DC.

Hoekman, J., Frenken, K., and Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5):662–673.

Holmes, T. J. and Stevens, J. J. (2014). An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade. *Journal of Political Economy*, 122(2):369–421.

Hörcher, D., De Borger, B., Seifu, W., and Graham, D. J. (2020). Public transport provision under agglomeration economies. *Regional Science and Urban Economics*, 81:103503.

Hsieh, C. T. and Moretti, E. (2015). Why Do Cities Matter? Local Growth and Aggregate Growth. *NBER Working Papers 21154*.

Iammarino, S. and McCann, P. (2013). *Multinationals and economic geography: Location, technology and innovation*. Edward Elgar Publishing, Cheltenham.

Inoue, H., Nakajima, K., and Yukiko, U. S. (2019). Localization of collaborations in knowledge creation. *The Annals of Regional Science*, 62(1):119–140.

Jacobs, J. (1969). *The Economy of Cities*. Random House, New York.

Jacobs, W., Koster, H. R. A., and Van Oort, F. G. (2013). Co-agglomeration of Knowledge-intensive Business Services and Multinational Enterprises. *Journal of Economic ...*, 14(2):443–475.

Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S. (2000). Knowledge Spillovers And Patent Citations: Evidence From A Survey Of Inventors. *American Economic Review*, 90(2):215–218.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577–598.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *Review of Economic Studies*, 76(1):283–317.

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

Kemeny, T., Feldman, M., Ethridge, F., and Zoller, T. (2016). The economic value of local social networks. *Journal of Economic Geography*, 16(5):1101–1122.

Kerr, W. R. (2008). Ethnic Scientific Communities and International Technology Diffusion. *Review of Economics and Statistics*, 90(3):518–537.

Kim, S. (1995). Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860-1987. *The Quarterly Journal of Economics*, 110(4):881–908.

King, G. and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(02):137–163.

Klein, A. and Crafts, N. (2012). Making sense of the manufacturing belt: determinants of U.S. industrial location, 1880–1920. *Journal of Economic Geography*, 12(4):775–807.

Klein, A. and Crafts, N. (2020). Agglomeration externalities and productivity growth: US cities, 1880–1930. *The Economic History Review*, 73(1):209–232.

Kleinknecht, A. (1981). Observations on the Schumpeterian swarming of innovations. *Futures*, 13(4):293–307.

Kleinknecht, A. (1987). *Innovation patterns in crisis and prosperity: Schumpeter's long cycle reconsidered.* Macmillan, Basingstoke, Hampshire.

Kleinknecht, A. (1990). Are there Schumpeterian waves of innovations? *Cambridge Journal of Economics*, 14(1):81–92.

Klepper, S. (2007). Disagreements, Spinoffs, and the Evolution of Detroit as the Capital of the U.S. Automobile Industry. *http://dx.doi.org/10.1287/mnsc.1060.0683*, 53(4):616–631.

Kogler, D. F., Rigby, D. L., and Tucker, I. (2013). Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, 21(9):1374–1391.

Kondratieff, N. D. (1926). Die langen Wellen der Konjunktur. *Archiv für Sozialwissenschaft und Sozialpolitik*, 56(3):573–609.

Koster, H. R. A. (2013). Economische effecten van het krachtwijkenbeleid. In *Soort zoekt soort: clustering en sociaal-economische scheidslijnen in Nederland.*, pages 128–135. Platform 31, Den Haag.

Koster, H. R. A., Van Ommeren, J. N., and Rietveld, P. (2012). Bombs, Boundaries and Buildings: a Regression-discontinuity Approach to Measure Costs of Housing Supply Restrictions. *Regional Science and Urban Economics*, 42(4):631–641.

Koster, H. R. A., van Ommeren, J. N., and Rietveld, P. (2016). Historic amenities, income and sorting of households. *Journal of Economic Geography*, 16(1):203–236.

Koster, P. and Koster, H. R. A. (2015). Commuters' Preferences for Fast and Reliable Travel: A Semi-Parametric Approach. *Transportation Research Part B: Methodological*, 81:289–301.

Krugman, P. (1991). *Geography and Trade.* MIT Press, Cambridge, MA.

Krugman, P. (1993). *Lessons of Massachusetts for EMU, "Adjustment and growth in the European Monetary Union", CEPR.* Cambridge University Press, Cambridge.

Kuijpers, K. and Thomas, C. (2016). De juiste yoghurt. *De Groene Amsterdammer.*

Latapy, M., Magnien, C., and Vecchio, N. D. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48.

Le Figaro (2018). Christophe Guilluy: Les "gilets jaunes" attestent la révolte de la France périphérique.

Leamer, E. E. and Storper, M. (2001). The economic geography of the internet age. *Journal of International Business Studies*, 32(4):641–665.

Lee, N. and Rodríguez-Pose, A. (2016). Is There Trickle-Down from Tech? Poverty, Employment, and the High-Technology Multiplier in U.S. Cities. *Annals of the American Association of Geography*, 106(5):1114–1134.

Leeper, T. J. (2018). *tabulizer: Bindings for Tabula PDF Table Extractor Library.*

Leydesdorff, L. and Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science and Technology*, 57(12):1616–1628.

Lo Turco, A. and Maggioni, D. (2016). On firms' product space evolution: the role of firm and local product relatedness. *Journal of Economic Geography*, 16(5):975–1006.

Marco, A. C., Carley, M., Jackson, S., and Myers, A. F. (2015). The USPTO Historical Patent Data Files Two Centuries of Innovation. *SSRN Electronic Journal.*

Marshall, A. (1890). *Principles of Economics.* MacMillan and Co., London.

Marshall, M. (1987). *Long waves of regional development.* MacMillan, Basingstoke, Hampshire.

Martin, R. (1999). The new 'geographical turn' in economics: some critical reflections. *Cambridge Journal of Economics*, 23(1):65–91.

Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks. *Journal of Economic Geography*, 12(1):1–32.

Martin, R., Pike, A., Tyler, P., and Gardiner, B. (2016). Spatially Rebalancing the UK Economy: Towards a New Policy Model? *Regional Studies*, 50(2):342–357.

Martin, R. and Sunley, P. (2015). On the notion of regional economic resilience: conceptualization and explanation. *Journal of Economic Geography*, 15(1):1–42.

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science (New York, N.Y.)*, 296(5569):910–3.

Matouschek, N. and Robert-Nicoud, F. (2005). The role of human capital investments in the location decision of firms. *Regional Science and Urban Economics*, 35(5):570–583.

Mazzucato, M. (2019). *The Value of Everything: Making and taking in the global economy*. Penguin Books Ltd, London.

McCann, P. (2008). Globalization and Economic Geography: The World is Curved, not Flat. *Cambridge Journal of Regions, Economy and Society*, 1(3):351–370.

McCann, P. and Fingleton, B. (1996). The regional agglomeration impact of just-in-time input linkages: Evidence from the Scottish Electronics Industry. *Scottish Journal of Political Economy*, 43(5):493.

Meijers, E. J. and van Rietbergen, T. (2021). Vergeet de dure wensen van de Randstad, investeer snel weer in de buitengewesten. *De Volkskrant*.

Mensch, G. O. (1975). *Das technologische Patt*. Umschau, Frankfurt.

Michaels, G., Redding, S. J., and Rauch, F. (2019). Task specialization in U.S. Cities from 1880 to 2000. *Journal of the European Economic Association*, 17(3):754–798.

Milikowski, F. (2018). *Van wie is de stad: de strijd om Amsterdam*. Atlas Contact, Amsterdam.

Milikowski, F. (2020). *Een klein land met verre uithoeken: Ongelijke kansen in veranderend Nederland*. Atlas Contact, Amsterdam.

Mills, E. S. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review*, 57(2):197–210.

Moretti, E. (2004a). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2):175–212.

Moretti, E. (2004b). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, 94(3):656–690.

Moretti, E. (2012). *The new geography of jobs.* Houghton Mifflin Harcourt Publishing Company, New York.

Mulder, P., De Groot, H. L. F., and Hofkes, M. W. (2001). Economic growth and technological change: A comparison of insights from a neo-classical and an evolutionary perspective. *Technological Forecasting and Social Change*, 68(2):151–171.

Muth, R. F. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* University of Chicago Press, Chicago.

Naidu, S., Rodrik, D., and Zucman, G. (2019a). Economics After Neoliberalism. *Boston Review.*

Naidu, S., Rodrik, D., and Zucman, G. (2019b). Forum Final Response - Economics After Neoliberalism: Economists should be enablers of democratic priorities, not oracles channeling a supply-and-demand deity. *Boston Review.*

Neffke, F., Hartog, M., Boschma, R., and Henning, M. (2018). Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification. *Economic Geography*, 94(1):23–48.

Neffke, F., Henning, M., and Boschma, R. (2011a). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87(3):237–265.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J. J., and Olander, L.-O. O. (2011b). The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. *Regional Studies*, 45(1):49–65.

Nelson, R. and Winter, S. (1982). *An evolutionary theory of economic change.* Harvard University Press, Cambridge, MA.

Nomaler, Ö., Frenken, K., and Heimeriks, G. (2014). On Scaling of Scientific Knowledge Production in U.S. Metropolitan Areas. *PLoS ONE*, 9(10):e110805.

Nordhaus, W. D. (2007). Two centuries of productivity growth in computing. *Journal of Economic History*, 67(1):128–159.

Odendahl, C. and Springford, J. (2020). Three ways COVID-19 will cause economic divergence in Europe. Technical report, Centre for European Reform, London.

Oevering, F. and Raspe, O. (2020). Regionaal-economische verschillen fors toegenomen. *Economische Statistische Berichten*, 105(4790):456–459.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Overman, H. G. (2004). Can we learn anything from economic geography proper? *Journal of Economic Geography*, 4(5):501–516.

Patience, G. S., Patience, Christian, A., Blais, B., and Bertrand, F. (2017). Citation analysis of scientific categories. *Heliyon*, 3(5):e00300.

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13:343–73.

PBL (2015). *De stad: magneet, roltrap en spons*. Planbureau voor de Leefomgeving, Den Haag.

PBL (2016). *De verdeelde triomf: verkenning van stedelijk-economische ongelijkheid en opties voor beleid: Ruimtelijke Verkenningen 2016*. Planbureau voor de Leefomgeving, Den Haag.

Peaucelle, J.-L. (2006). Adam Smith's use of multiple references for his pin making example. *The European Journal of the History of Economic Thought*, 13(4):489–512.

Peres-Neto, P. R. (2004). Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, 140(2):352–360.

Perez, C. (1983). Structural change and assimilation of new technologies in the economic and social systems. *Futures*, 15(5):357–375.

Perlman, E. R., Margo, R., Nicholas, T., Atack, J., Frydman, C., Paserman, D., Olivetti, C., Meurer, M., Naidu, S., Darby, T. M., Sims, P., Berntsen, M. C., Gross, H., and Corliss, R. (2015). Dense Enough To Be Brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America Market Access.

Petralia, S. G., Balland, P.-A. A., and Rigby, D. L. (2016). Unveiling the geography of historical patents in the United States from 1836 to 1975. *Scientific data*, 3:1–14.

Pierce, J. R. and Schott, P. K. (2016). The Surprisingly Swift Decline of US Manufacturing Employment. *American Economic Review*, 106(7):1632–1662.

Pierce, J. R. and Schott, P. K. (2020). Trade Liberalization and Mortality: Evidence from US Counties. *American Economic Review: Insights*, 2(1):47–63.

Pike, A., Dawley, S., and Tomaney, J. (2010). Resilience, adaptation and adaptability. *Cambridge Journal of Regions, Economy and Society*, 3(1):59–70.

Piketty, T. (2013). *Le Capital au XXIe siècle*. Éditions du Seuil, Paris.

Piketty, T. (2014). *Capital in the Twenty-First Century*. The Belknap Press of Harvard University Press, Cambridge, MA.

Piketty, T. (2019). *Capital et idéologie*. Éditions du Seuil, Paris.

Pine, B. (1993). *Mass customization: the new frontier in business competition*. Harvard Business School Press, Boston, MA.

Pinheiro, F. L., Hartmann, D., Boschma, R., and Hidalgo, C. A. (2021). The time and frequency of unrelated diversification. *Research Policy*, Advance On.

Pred, A. (1967). *Behaviour and Location: Foundations for a Geographic and Dynamic Location Theory: Part I*. Lund Studies in Geography, Lund.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219.

Raspe, O. and van Oort, F. G. (2007). *Ruimtelijk-economisch beleid in de kenniseconomie*. NAi Uitgevers/Ruimtelijk Planbureau, Rotterdam/Den Haag.

Raspe, O., Weterings, A. B., Geurden-Slis, M., and van Gessel, G. (2012). *De ratio van ruimtelijk-economisch topsectorenbeleid*. Planbureau voor de Leefomgeving, Den Haag.

Rigby, D. L. (2015). Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes. *Regional Studies*, 49(11):1922–1937.

Rigby, D. L., Roesler, C., Kogler, D., Boschma, R. A., and Balland, P.-A. (2022). Do EU regions benefit from smart specialization principles? *Regional Studies*, Advance On.

Rocchetta, S. and Mina, A. (2019). Technological coherence and the adaptive resilience of regional economies. *Regional Studies*, 53(10):1421–1434.

Rocchetta, S., Mina, A., Lee, C., and Kogler, D. F. (2022). Technological knowledge spaces and the resilience of European regions. *Journal of Economic Geography*, 22(1):27–51.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, 11(1):189–209.

Rodríguez-Pose, A. and Crescenzi, R. (2008). Mountains in a Flat World: Why Proximity Still Matters for the Location of Economic Activity. *Cambridge Journal of Regions, Economy and Society*, 1(3):371–388.

Rodríguez-Pose, A., Crescenzi, R., and Cataldo, M. D. (2018). Institutions and the Thirst for 'Prestige' Transport Infrastructure. In Glückler, J., Suddaby, R., and Lenz, R., editors, *Knowledge and Institutions*, pages 227–246. Springer, Cham.

Rodrik, D. (2011). *The Globalization Paradox: Democracy and the Future of the World Economy.* W.W. Norton & Company, Inc., New York.

Rodrik, D. (2018). Populism and the economics of globalization. *Journal of International Business Policy*, 1(1-2):12–33.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5):1002–1037.

Rooijers, E. and van Rein, E. (2020). De nieuwe Zuidas: van intimiderende kantoorkathedraal naar een bos op het dak. *Het Financieele Dagblad.*

Rosenberg, N. (1982). *Inside the black box: technology and economics.* Cambridge University Press, Cambridge.

Rosenberg, N. and Frischtak, C. R. (1983). Long Waves and Economic Growth: A Critical Appraisal. *The American Economic Review*, 73:146–151.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, 85(2):377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Chapter 49 Evidence on the nature and sources of agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 49, pages 2119–2171. Elsevier, Amsterdam.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 65(2):373–389.

Rotemberg, J. J. and Saloner, G. (2000). Competition and human capital accumulation: a theory of interregional specialization and trade. *Regional Science and Urban Economics*, 30(4):373–404.

Rothwell, J. (2013). *The hidden STEM economy.* Brookings, Washington D.C.

Rothwell, J., Lobo, J., Strumsky, D., and Muro, M. (2013). Patenting prosperity: invention and economic performance in the United States and its metropolitan areas. Technical report, Brookings Institution, Washington.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2018). *IPUMS USA: Version 8.0 [dataset]*. IPUMS, Minneapolis, MN.

Saviotti, P. (1996). *Technological Evolution, Variety and the Economy*. Edward Elgar Publishing, Cheltenham.

Scherer, F. M. (1982). Demand-Pull and Technological Invention: Schmookler Revisted. *Journal of Industrial Economics*, 30(3):225.

Scheve, K. F. and Slaughter, M. F. (2007). A New Deal for Globalization. *Foreign Affairs*.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press, Cambridge.

Schumpeter, J. A. (1939). *Business cycles*. McGraw-Hill, New York.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Routledge, London.

Schutze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97–123.

Sedita, S. R., De Noni, I., and Pilotti, L. (2016). Out of the crisis: an empirical investigation of place-specific determinants of economic resilience. *European Planning Studies*, 25(2):155–180.

Serafinelli, M. (2019). "Good" firms, worker flows, and local productivity. *Journal of Labor Economics*, 37(3):747–792.

Silverman, B. (2002). *Technological Resources and the Logic of Corporate Diversification*. Routledge, London.

Silvermintz, D. (2010). Plato's supposed defense of the division of labor: A reexamination of the role of job specialization in the republic. *History of Political Economy*, 42(4):747–772.

Simmie, J. and Martin, R. (2010). The economic resilience of regions: towards an evolutionary approach. *Cambridge Journal of Regions, Economy and Society*, 3(1):27–43.

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, 1(1-2):2–14.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford University Press, Oxford.

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1):65–94.

Sotarauta, M., Beer, A., and Gibney, J. (2017). Making sense of leadership in urban and regional development. *Regional Studies*, 51(2):187–193.

Stil, H. (2020). Zuidas moet het gaan stellen zonder ABN: hoe erg is dat? *Het Parool*.

Storper, M. (2018). Separate Worlds? Explaining the current wave of regional economic polarization. *Journal of Economic Geography*, 18(2):247–270.

Storper, M. and Scott, A. J. (1992). *Pathways to Industrialization and Regional Development*. Routledge, New York & London.

Storper, M. and Scott, A. J. (2009). Rethinking human capital, creativity and urban growth. *Journal of Economic Geography*, 9(2):147–167.

Storper, M. and Venables, A. J. (2004). Buzz: Face-to-face Contact and the Urban Economy. *Journal of Economic Geography*, 4(4):351–370.

The Economist (2016). Rage against the Dying of the Light. Regional Inequality is becoming too dangerous to ignore.

Uhlbach, W., Balland, P.-A., and Scherngell, T. (2017). R&D Policy and Technological Trajectories of Regions: Evidence from the EU Framework Programmes. *Papers in Evolutionary Economic Geography*, 17(22):1–21.

van der Wouden, F. (2018). *Exploring the Changing Structures of Inventor Collaboration in U.S. Cities between 1836 and 1975*. PhD thesis, UCLA.

van der Wouden, F. (2020). A history of collaboration in US invention: changing patterns of co-invention, complexity and geography. *Industrial and Corporate Change*, 29(3):599–619.

van der Wouden, F. and Rigby, D. L. (2019). Co-inventor networks and knowledge production in specialized and diversified cities. *Papers in Regional Science*, 98(4):1833–1853.

van Eck, N. J. and Waltman, L. (2009). How to Normalize Cooccurrence Data? An Analysis of SomeWell-Known Similarity Measures. *Journal of the American Society for Information Science*, 60(8):1635–1651.

van Oort, F. G. (2002). *Agglomeration, Economic Growth and Innovation. Spatial Analyses of Knowledge Externalities in the Netherlands.* PhD thesis, Erasmus University Rotterdam.

Verdugo, G. (2014). The great compression of the French wage structure, 1969-2008. *Labour Economics*, 28:131–144.

Vernon, R. (1960). *Metropolis 1985.* Harvard University Press, Cambridge, MA.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics*, 80(2):190–207.

Von Hippel, P. (2015). Linear vs. Logistic Probability Models: Which is Better, and When? [Blog post].

von Thünen, J. H. (1826). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Perthes, Hamburg.

Waltman, L., van Eck, N. J., and Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4):629–635.

Webber, D. J., Healy, A., and Bristow, G. (2018). Regional Growth Paths and Resilience: A European Analysis. *Economic Geography*, pages 1–21.

Weber, A. (1922). *Über den Standort der Industrie.* Auflage, Tübingen.

West, E. G. (1964). Adam Smith's Two Views on the Division of Labour. *Economica*, 31(121):23.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Xiao, J., Boschma, R., and Andersson, M. (2018). Resilience in the European Union: the effect of the 2008 crisis on the ability of regions in Europe to develop new industrial specializations. *Industrial and Corporate Change*, 27(1):15–47.

Zeemeijer, I. (2016). Nederland barst van de innovatievalleys, clusters en delta's. *Het Financieele Dagblad*.

Additional Chapter A

Improvement on the association strength:

implementing a probabilistic measure based on combinations without repetition.

Abstract – The use of co-occurrence data is common in various domains. Co-occurrence data often needs to be normalised to correct for the size-effect. To this end, van Eck and Waltman (2009) recommend a probabilistic measure known as the association strength. However, this formula, based on combinations with repetition, implicitly assumes that observations from the same entity can co-occur even though in the intended usage of the measure these self-co-occurrences are non-existent. A more accurate measure based on combinations without repetition is introduced here and compared to the original formula in mathematical derivations, simulations, and patent data, which shows that the original formula overestimates the relation between a pair and that some pairs are more overestimated than others. The new measure is available in the EconGeo package for R maintained by Balland (2016).

This additional chapter has been published as Steijn, M. P. A. (2021). Improvement on the association strength: Implementing a probabilistic measure based on combinations without repetition. *Quantitative Science Studies*, 2(2):778–794.

A.1 Introduction

The use of co-occurrence data is popular in numerous scientific domains like scientometrics (see for example Leydesdorff and Vaughan, 2006; van Eck and Waltman, 2009), computational linguistics (see for example Schutze, 1998), community ecology (see for example Peres-Neto, 2004), development economics (see for example Hidalgo et al., 2007), molecular biology (see for example Maslov and Sneppen, 2002) and evolutionary economic geography (see for example Boschma et al., 2015). Its use is widespread and in close relation with the popularity of network analysis across disciplines.

Co-occurrence data is used to infer the relation, referred to as relatedness here following Hidalgo et al. (2007), between entities, which can be species of fish, authors or technological classes, by observing how each of these co-occur with others in places, like streams, articles or patents. However, the total number of co-occurrences between a pair of entities cannot be used straightforwardly to reflect the relatedness between them because entities with more observations are more likely to co-occur than entities with fewer observations. To correct for this size-effect a normalisation measure is applied to the data.[1] van Eck and Waltman (2009) review the most popular normalisation measures and make a convincing case for the use of a probability-based measure known as the association strength. This measure is based on dividing the *observed* number of co-occurrences over the *expected* numbers of co-occurrences when assuming observations are randomly distributed over co-occurrences.[2]

In this Chapter, it is shown that the probability formula of the association strength, as proposed by van Eck and Waltman (2009), is not optimised to calculate the expected number of co-occurrences. The formula of van Eck and Waltman (2009) is proportional to probability calculations based on combinations with repetition, which means that when estimating the probability that two entities co-occur an observation drawn in the first draw is assumed to be available for drawing again when drawing the second observation. However, in the use of co-occurrence data the co-occurrence of observations from the same entity is disregarded.[3] Authors, for example, do not co-author papers with themselves (see Leydesdorff and Vaughan, 2006). Therefore, van

[1] Note that it depends on the goal of the research if it is necessary to correct for the size-effect or that absolute counts are more relevant. In the research cited here and in van Eck and Waltman (2009) normalisation is assumed to be necessary. The exact definitions of occurrences, co-occurrences and the size-effect are given in Section A.3

[2] As such, a value of one indicates that exactly the same amount of co-occurrences are observed as expected. While a value above one or below one indicates respectively a stronger relation or a weaker relation between the two entities.

[3] This holds for the work referred to in this Chapter and those by van Eck and Waltman (2009).

Eck and Waltman (2009) suggest to set these self co-occurrences to missing values.[4] This makes the possibility of drawing the same observation or any other observation from the same entity impossible in the second draw once an observation from this entity has been drawn in the first draw.

Therefore, an improved formula for the association strength is introduced using a probability measure based on combinations without repetition but with a noticeable change. In combinations with repetition one can not draw an observation in the second draw if it has been drawn in the first draw. In this setting none of the observations belonging to the same entity as the first observation can be drawn in the second draw. Furthermore, two refinements are proposed in this Chapter regarding the inputs to the formula, which in the current definition do not properly take into account how the number of observed co-occurrences are calculated.

The improved formula is compared to the original formula in a theoretical setting, a number of simulations, and a real world application using patent data. It is shown that: firstly, the original formula overestimates the relatedness between a pair of entities, when this pair has one co-occurrence. This indicates that the original formula can wrongfully identify two entities as related whereas in fact they are not; and, secondly, the original formula overestimates the relatedness between some pairs more than other. This indicates that the overestimation is not proportional and that the differences between the relatedness values for each pair are also distorted.

In the theoretical analysis, the improved formula is subtracted from the original formula, to obtain a formula for the difference. By considering the domain of each variable, it is shown that the original formula underestimates the number of expected occurrences in all cases and therefore overestimates the relationship between two entities when there is at least one observed co-occurrence. Continuing the theoretical exploration, the first order partial derivatives of the difference with respect to each variable is taken, which shows that the overestimation is not equal across all possible types of co-occurrence matrices.

Just taking the partial derivatives is not sufficient to show the size of the difference for each case, as the values of the variables are interconnected in ways that do not allow for analytical solving. Therefore, simulations are ran in which four different exemplary cases are taken to the extreme to demonstrate the effect on the difference. The simulations show that the overestimation by the original formula can be close to 0% but also close to 100% of the relatedness value given by the improved formula

[4] In this Chapter, the suggestion is made to set them to zero, see Section A.2, which is also often used (Ahlgren et al., 2003).

depending on the specificities of the co-occurrence matrix.

To measure to what extent these theoretical simulations are representative of real world applications of research on co-occurrence data, a number of patent samples, containing data on the technology classes per document, is treated to compare the results of both formulas. In these samples the overestimation of relatedness values for individual pairs varies between close to 0% to up to 3.234% of the value given by the improved formula and therefore does not attain the most extreme values obtained in the simulation. Nonetheless, it clearly confirms that some pairs are more overestimated than others. The results also show that some pairs are misidentified as being related by the original formula but that this is only the case for a rather small share of the pairs up to about 0.29% of the number of pairs identified by the original formula.

All in all, it is advisable to use the improved formula when working with co-occurrence data, where self co-occurrences are non-existent or irrelevant. The reformulation of the probability measure does not in any way alter the conclusion by van Eck and Waltman (2009) that probability based measures outperform so-called set-theoretic measures in normalising co-occurrence data. The improved measure, including the recommended method of implementation, is available in the EconGeo package for R maintained by Balland (2016).

this Chapter is organised as follows: Section A.2 gives a short overview of the use of co-occurrence data and the association strength; Section A.3 discusses the refinements; Sections A.4 to A.6 explore the overestimation by the original formula respectively in a theoretical setting, simulations, and in a real world example using patent data; and Section A.7 concludes.

A.2 Normalising co-occurrence data

Co-occurrence data is generally derived from a binary occurrence matrix O of some order $m \times n$. The rows of O correspond to the places in which the observations occur and the columns to the entities to which they belong.[5] There is a large variety of what these places and entities can be.[6] The example in Matrix 1 shows three patents that contain a reference to, respectively, only class c; class c & class d; and all classes a to d.

[5]This type of matrix in which two sets of vertices, here places and entities, are connected by the co-occurrences in such a way that each link is between one entity and one place is also known as a bipartite matrix in graph theory (see Latapy et al., 2008)

[6]There are for example occurrence matrices of: scientific publications by authors (e.g. Leydesdorff and Vaughan, 2006) or by research institutions (e.g. Hoekman et al., 2010); countries by industries (e.g. Hidalgo et al., 2007); streams by fish species (e.g. Peres-Neto, 2004); and patent documents by technology classes (e.g. Boschma et al., 2015).

Matrix 1

$$\begin{array}{cc} & \begin{array}{cccc} \mathit{Class}\ \mathrm{a} & \mathit{Class}\ \mathrm{b} & \mathit{Class}\ \mathrm{c} & \mathit{Class}\ \mathrm{d} \end{array} \\ \begin{array}{c} \mathit{Patent}\ 1 \\ \mathit{Patent}\ 2 \\ \mathit{Patent}\ 3 \end{array} & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

By multiplying the transpose of O by O itself the co-occurrence matrix C is obtained[7]. In which both the rows and the columns represent the entities and the matrix gives how often they co-occur with the other.

In the case of our example, this would yield the co-occurrence matrix C given in Matrix 2. Where class a co-occurs once with b, c, and d; class b co-occurs once with a, c and d; class c co-occurs once with a and b, and twice with d; and class d co-occurs once with a and b, and twice with c.

The diagonal is set to zero as the reference to a certain class does not entail a co-occurrence between that class and itself in the line of research for which the formula is intended. Ahlgren et al. (2003); Leydesdorff and Vaughan (2006) and van Eck and Waltman (2009) suggest setting the diagonal to missing values. This leads to the same results. However, it is advisable to use zeros because missing values often results in errors when using statistical software.[8] Setting the diagonal to zero has important implications down the line.

Matrix 2

$$\begin{array}{cc} & \begin{array}{cccc} \mathit{Class}\ \mathrm{a} & \mathit{Class}\ \mathrm{b} & \mathit{Class}\ \mathrm{c} & \mathit{Class}\ \mathrm{d} \end{array} \\ \begin{array}{c} \mathit{Class}\ \mathrm{a} \\ \mathit{Class}\ \mathrm{b} \\ \mathit{Class}\ \mathrm{c} \\ \mathit{Class}\ \mathrm{d} \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 2 & 0 \end{pmatrix} \end{array}$$

In many applications of co-occurrence data, such as the concept of relatedness, the raw numbers of co-occurrences between entities cannot straightforwardly be interpreted as giving the strength of the relation between each pair of entities. There is a so-called size-effect, as some classes co-occur more often with others for the simple reason that

[7] If the rows of O indicate the entities and the columns indicate the places where they co-occur then it is the other way around and O should be multiplied by its transpose.

[8] Ahlgren et al. (2003) also mentions the option of setting the diagonal equal to the number of times an entity occurs at least twice in a place. This option is unsuitable for probabilistic similarity measures, like the association strength, because the number of times an entity occurs at least twice does not entail a co-occurrence between i and j therefore when estimating the probability of a co-occurrence between i and j one cannot draw the observations on the diagonal even though these are added to the total, and therefore the pool of observations from which one can draw. This becomes more clear when discussing the formula in Section A.3.

these classes have more occurrences in the first place. Like in our example, where d has more co-occurrences with c than with a or b but c also has more occurrences in total and therefore is more likely to co-occur with any class.

To correct the absolute number of co-occurrences for the size-effect a normalisation procedure is applied to the data (van Eck and Waltman, 2009).[9] Correcting co-occurrence data for the size-effect to derive relationships between entities is done through direct similarity measures.[10] van Eck and Waltman (2009) wrote an extensive review on the most popular direct similarity measures, being: the cosine, the Jaccard index, the inclusion index and the association strength. Of these the last is a probabilistic measure, while the others are set-theoretic measures. The authors show that set-theoretic measures do not properly correct for the size effect and argue in favour of the association strength.

The usability of their formula exceeds the domain of scientometrics. Hidalgo et al. (2007) developed an influential network analysis tool to derive the relatedness between entities on the basis of co-occurrences. Although they use a different probabilistic direct similarity measure than the ones covered by van Eck and Waltman (2009), other authors (*e.g.* Balland et al., 2015) building on the framework of Hidalgo et al. (2007) do opt for the association strength, as defined by van Eck and Waltman (2009).[11]

Albeit influential, refinements to the work of van Eck and Waltman (2009) are in place. The probabilistic formula should be based on a specific case of combinations *without* repetition instead of *with* repetition. Furthermore, the definitions of the inputs for the formula are imprecise. These points will be treated in the following section. It should be noted that the refinements to the measure do not undermine in anyway the statement of van Eck and Waltman (2009) that probabilistic measures outperform set-theoretic measures in normalising co-occurrence data to control for the size-effect.

[9] In some cases, more normalisation measures are deemed necessary. For example, Neffke et al. (2011a) who look at the co-occurrence of products in the production process of the same plant also correct for the profitability of the respective products.

[10] Another option to derive similarities or relationships between entities is by comparing co-occurrence profiles of the entities, which are known as indirect similarity measures (see van Eck and Waltman, 2009).

[11] Hidalgo et al. (2007) look into the co-occurrence of specialisations in exporting industries in a country. Their formula consists of taking the smallest value of the conditional probability of effectively exporting product j knowing that a country effectively exports i and the conditional probability of effectively exporting product i knowing that a country effectively exports j. This does not correctly correct for the size effect because each conditional probability corrects for the size of only one of the two, the former of i the latter of j, by picking the smallest of the two conditional probabilities the other size effect still remains. Furthermore, the probability if a country meets the condition of effectively exporting product j or i is neglected by taking the conditional probabilities. These reasons make it understandable that other authors following the line of Hidalgo et al. (2007) have opted for the association strength of van Eck and Waltman (2009)

A.3 Refinement to the association strength

The objective of the association strength is to estimate the number of expected co-occurrences for each pair assuming that these are randomly distributed and compare this to the number of observed co-occurrences to give an indication of the relation between a pair of entities when corrected for the size-effect. The challenge therefore is to correctly estimate the number of expected co-occurrences per combination.

As an intuitive example Matrix 3 gives a co-occurrence matrix C in which three classes (a, b, and c) exist and co-occur exactly once with each other:[12]

Matrix 3

$$\begin{pmatrix} & \mathit{Class}\,\text{a} & \mathit{Class}\,\text{b} & \mathit{Class}\,\text{c} \\ \mathit{Class}\,\text{a} & 0 & 1 & 1 \\ \mathit{Class}\,\text{b} & 1 & 0 & 1 \\ \mathit{Class}\,\text{c} & 1 & 1 & 0 \end{pmatrix}$$

As each class has two observations and two possible other classes to co-occur with the expected number of co-occurrences is logically $\frac{2}{2} = 1$ for each combination (a & b, a & c, and b & c).

In this case, the matrix of expected co-occurrences is exactly the same as the matrix of observed co-occurrences given in Matrix 3. Therefore, we observe as many co-occurrences as expected and $\frac{Observed}{Expected}$ should be equal to one for each combination.

For the association strength, van Eck and Waltman (2009) use a simplified formula in the main text but describe formula A.1 on p.1636:[13,14]

$$S_{Original}(C_{ij}, S_i, S_j, T, m) = \frac{C_{ij}}{(\frac{S_i}{T}\frac{S_j}{T} + \frac{S_j}{T}\frac{S_i}{T})m}, i \neq j, \tag{A.1}$$

In which S_i and S_j are the number of occurrences of entity i respectively j involved in co-occurrences where $i \neq j$. To calculate S_i one can use the row sum of row i

[12]This matrix C would result from our example O in Matrix 1 if one would remove class d and its observations.

[13]I argue that it is more advantageous to use the full formula, which entails exactly dividing the number of observed co-occurrences over the number of expected co-occurrences as it gives a clear threshold of one when $Observed = Expected$. As such, values below one indicate that less co-occurrences are observed than could be expected given a random distribution, whereas values above indicate the opposite. This threshold holds in all cases, even when matrices with different numbers of occurrences are compared. In contrast, the simplified formula would have a different value indicating that the number of observed co-occurrences equals expected depending on the matrices, even though it is proportional to the more detailed formula by a factor of $2m$.

[14]This formula is also presented in rewritten form in equation 1 in Waltman et al. (2010).

of the matrix C when the diagonal is set to zero.[15] This slightly diverges from the explanation of van Eck and Waltman (2009).[16] T is the total number of occurrences and equal to $\sum_{i=1}^{n} S_i$ with n being the total number of entities, and m is the total number of co-occurrences and therefore equal to $\frac{\sum_{i=1}^{n} S_i}{2}$, which is half of T as each co-occurrence involves 2 occurrences. This definition also diverges from van Eck and Waltman (2009).[17] C_{ij} is the number of *observed* co-occurrences between i and j.

In essence, the denominator gives that the chance of encountering a co-occurrence between an observation of class i and an observation of class j is equal to the probability of first drawing one of the observations of class i out of the total number of occurrences times the chance of drawing an observation belonging to class j out of the total number of occurrences plus the probability of first drawing j and then i times the total number of co-occurrences.

Calculating this formula for our example C in Matrix 3 would yield Relatedness Matrix R given in Matrix 4 below:

Matrix 4

$$\begin{pmatrix} & \mathit{Class}\text{ a} & \mathit{Class}\text{ b} & \mathit{Class}\text{ c} \\ \mathit{Class}\text{ a} & 0 & 1.5 & 1.5 \\ \mathit{Class}\text{ b} & 1.5 & 0 & 1.5 \\ \mathit{Class}\text{ c} & 1.5 & 1.5 & 0 \end{pmatrix}$$

It is clear that the formula does not provide the intuitive answer of 1 but actually overestimates the relationship by returning that each pair co-occurs more often than

[15]Taking the column sum of column i gives the same value as the row sum of row i.

[16]van Eck and Waltman (2009, p. 1636) state that for S_i both the number of occurrences of entity i can be used or the number of co-occurrences in which i is involved. However, it is important to emphasize that single occurrences, as in Patent 1 of the example O in Matrix 1, should be ignored as these do not lead to co-occurrences. This also holds for self co-occurrences of i with i as both of these cannot be part of C_{ij} where $i \neq j$. Setting the diagonal to zero resolves both these issues. This is also the reason that setting the diagonal equal to the number of times an entity occurs at least twice in a place, as suggested by Ahlgren et al. (2003), is unsuitable for this probabilistic measure.

[17]van Eck and Waltman (2009, p. 1648) state that m should be equal to "the number of documents". However, this only holds when the number of documents is equal to the number of co-occurrences. In the example O in Matrix 1 patent 1 is one document but only refers to one class so it does not involve any co-occurrences and is therefore not equal to one co-occurrence. Patent 3, on the other hand, is also a single document but refers to all classes a to d and therefore leads to 6 unique co-occurrences (a&b, a&c, a&d, b&c, b&d, c&d). All together the example consists of three documents and seven unique co-occurrences. As a result, in this case using the number of documents for m would underestimate the expected number of co-occurrences as the probability of encountering a co-occurrence is multiplied by a too small number of co-occurrences than are actually possible. This explanation is the same as in Waltman et al. (2010). From this follows that the size-effect is the result of the fact that some entities are involved in more co-occurrences than others, which means more observations and therefore an increased likelihood to co-occur with any other entity. This means that raw probabilities of co-occurrence cannot be compared straight away and a normalisation measure is needed, such as the one introduced in this Chapter.

could be expected given a random distribution.

The flaw cannot lie in the numerator, which is equal to the number of observed co-occurrences. Therefore the problem lies in the denominator. The formula to calculate the expected number of co-occurrences includes the possibility that when an occurrence of a certain entity is drawn the same occurrence or another occurrence of the same entity (if present) can be drawn in the next draw to complete the co-occurrence. This is known as combinations with repetition. However, as self co-occurrences are non-existent one knows that one cannot redraw the same occurrence, but also none of the other occurrences of that class.

In the case of our example, the denominator of formula A.1 yields an expected number of $\frac{2}{3}$ co-occurrences. This is because the formula observes 2 occurrences for each class and 3 possible partners to co-occur with even though there are only 2 possible partners. Class a can co-occur with class b and class c but not with itself.[18]

In the case of co-occurrence data in which none of the observations belonging to the previously drawn entity can be drawn in the second draw the correct probabilistic measure would be formula A.2:

$$S_{Improved}(C_{ij}, S_i, S_j, T, m) = \frac{C_{ij}}{(\frac{S_i}{T}\frac{S_j}{T-S_i} + \frac{S_j}{T}\frac{S_i}{T-S_j})m}, i \neq j, \tag{A.2}$$

Here, the denominator gives that the chance of encountering a co-occurrence between an observation of class i and an observation of class j is equal to the probability of first drawing one of the observations of class i times the chance of drawing an observation belonging to class j knowing that none of the observations of class i can be drawn plus the chance of first drawing one of the observations of class j times the chance of drawing an observation belonging to class i knowing that any other observations of class j cannot be drawn.

The implications of using formula A.1 instead of formula A.2 are that the relatedness between a pair is overestimated when at least one co-occurrence is observed and that the overestimation is larger for certain pairs than others. These implications are demonstrated and further explored in the following parts. First in a theoretic setting, then by running simulations and concluding with the analysis of a real world example using patent data.

[18]To be exact the denominator of formula A.1 would be equal to $(\frac{2}{6}\frac{2}{6} + \frac{2}{6}\frac{2}{6})3$ for each pair outside of the diagonal in the matrix of this example.

A.4 Theoretical exploration of the overestimation.

An obvious first notion from observing formula A.1 and formula A.2 is that there is no difference in outcome when the number of observed co-occurrences is zero, as the numerator C_{ij} will then be zero.

Furthermore, it can be assumed that formula A.1 overestimates the relation between two entities when there is at least one co-occurrence. The assumption in the probabilistic measure of formula A.1 is that the same observation and other observations from the same entity can be drawn again while this is not possible. This enlarges the total pool from which observations can be drawn and therefore decreases the likelihood that a certain co-occurrence can be drawn. This leads to the denominator, which contains the expected number of co-occurrences, in formula A.1 being smaller than the one in formula A.2 in all cases. As was the case for the example Matrix 3, where the denominator indicated a co-occurrence probability of $\frac{2}{3}$ for each pair where actually only two options instead of three existed and therefore $\frac{2}{2}$ should have been the answer.

Due to the smaller expected probability, formula A.1 divides the number of observed co-occurrences over a too small number of expected co-occurrences and therefore the relatedness between these two entities is overestimated, when at least one co-occurrence is observed.

That the denominator of formula A.1 underestimates the expected number of co-occurrences can also be proven analytically. The original probabilistic measure of van Eck and Waltman (2009) in the denominator of formula A.1 is rewritten and given in formula A.3, while the improved probabilistic measure used in the denominator of formula A.2 is rewritten and given in formula A.4:

$$E(C_{ij})_{Original}(S_i, S_j, T) = \frac{S_i S_j}{T}, i \neq j, \tag{A.3}$$

$$E(C_{ij})_{Improved}(S_i, S_j, T) = \frac{S_i S_j (2T - S_i - S_j)}{2(T - S_i)(T - S_j)}, i \neq j, \tag{A.4}$$

Let $D_{probability}$ be equal to $E(C_{ij})_{Improved} - E(C_{ij})_{Original}$. It can be shown that this difference $D_{probability}$ is equal to formula A.5.

$$D_{probability}(S_i, S_j, T) = \frac{S_i S_j (S_i T + S_j T - 2 S_i S_j)}{2T(T - S_i)(T - S_j)}, i \neq j, \tag{A.5}$$

For $E(C_{ij})_{Improved}$ to be larger than $E(C_{ij})_{Original}$ formula A.5 gives that $S_i T + S_j T$

must be larger than $2S_iS_j$. As $S_i \geq 1$, $S_j \geq 1$, and $T = S_i + S_j + S_k + ... + S_n$ it is clear that $T > S_i$ and $T > S_j$ and therefore $S_iT + S_jT > 2S_iS_j$ must hold.[19]

This means that $D_{probability}$ is positive in all circumstances, which indicates that the improved formula predicts in all cases that more co-occurrences can be expected between i and j. This makes sense as the improved formula excludes the possibility of drawing a combination of i and i making it more likely to draw a combination between i and j.

Because the number of observed co-occurrences, C_{ij}, is divided over the number of expected co-occurrences, the original formula A.1 leads to larger results than the improved formula A.2 in all possible cases, when $C_{ij} > 0$. This can also be shown mathematically: Let $D_{Formula}$ be equal to $S_{Original}(C_{ij}, S_i, S_j, T) - S_{Improved}(C_{ij}, S_i, S_j, T)$.[20] It can be shown that the difference $D_{Formula}$ is equal to formula A.8 after rewriting formula A.1 to formula A.6 and formula A.2 to formula A.7.

$$S_{Original}(C_{ij}, S_i, S_j, T) = \frac{TC_{ij}}{S_iS_j}, i \neq j, \tag{A.6}$$

$$S_{Improved}(C_{ij}, S_i, S_j, T) = \frac{2(T - S_i)(T - S_j)C_{ij}}{S_iS_j(2T - S_i - S_j)}, i \neq j, \tag{A.7}$$

$$D_{Formula}(C_{ij}, S_i, S_j, T) = \frac{(S_iT + S_jT - 2S_iS_j)C_{ij}}{S_iS_j(2T - S_i - S_j)}, i \neq j, \tag{A.8}$$

Three important notions can be derived from formula A.8. First, it is confirmed that when there are no observed co-occurrences, i.e. $C_{ij} = 0$, the difference is zero. Second, if and only if $C_{ij} > 0$ then $S_i \geq S_j \geq 1$ and $T \geq S_i + S_j$ and therefore $(S_iT + S_jT > 2S_iS_j)$. This indicates that formula A.1 yields larger outcomes than formula A.2 in all possible cases, with at least one observed co-occurrence. Effectively overestimating the relation between entity i and j. Third, for different values of S_i, S_j, C_{ij} and T the difference between formula A.1 and formula A.2 will also vary. This means that the difference between the formulas is not proportional for each pair but the relatedness between certain pairs is more strongly overestimated than for other pairs.

To explore the difference due to different values of S_i, S_j, C_{ij} and T the partial

[19] If entities can partially occur in a place then the values for S_i and S_j can be below one but in any case not below or equal to zero and therefore same statements hold.

[20] Note that the order of the original formula and the improved formula has been altered compared to the previous calculation of the difference of the respective probabilistic measures.

derivatives are taken of $D_{Formula}$ with respect to each. Because T is a function of S_i, S_j, and all other co-occurrences, $\sum_{k\neq i,j}^{n} S_k$. T is replaced by $S_i + S_j + L$ in formula A.8 in which $L = \sum_{k\neq i,j}^{n} S_k$ and its range is equal to or larger than zero.

The partial derivatives $\frac{\delta D_{Formula}}{\delta C_{ij}}$, $\frac{\delta D_{Formula}}{\delta S_i}$, and $\frac{\delta D_{Formula}}{\delta L}$ are respectively given in formulas A.9, A.10, and A.11.[21]

$$\frac{\delta D_{Formula}}{\delta C_{ij}} = \frac{(S_i^2 + S_j^2 + S_iL + S_jL)}{S_iS_j(S_i + S_j + 2L)}, i \neq j, \tag{A.9}$$

$$\frac{\delta D_{Formula}}{\delta S_i} = \frac{C_{ij}(S_i^2S_j + S_i^2L + 2S_iL^2 - 2S_iS_jL - S_j^3 - 3L - 2S_iL - 2S_jL^2)}{S_i^2S_j(S_i + S_j + 2L)^2}, i \neq j, \tag{A.10}$$

$$\frac{\delta D_{Formula}}{\delta L} = \frac{-C_{ij}(S_i - S_j)^2}{S_iS_j(S_i + S_j + 2L)^2}, i \neq j, \tag{A.11}$$

Given the domain of each formula, formula A.9 is always positive, and, when at least one co-occurrence exists, formula A.10 can be positive or negative depending on the respective inputs and formula A.11 is always negative.

This last statement suggests that a relationship between two entities will be more overestimated by formula A.1 when there is a smaller amount of other possibilities to co-occur with.

Despite being informative, partial derivatives give an incomplete picture of the discrepancy between the two formulas as these give the direction of a function with respect to an infinitesimal increase in one of the variables while keeping the others equal, even though in reality it is impossible to keep the other variables equal as the inputs are all related to each other. Necessarily C_{ij} consists of S_i and S_j, and if not all S_i co-occur with S_j then L must at least have enough occurrences to co-occur with the remaining i and js. In other words, the following logical conditions hold: $C_{ij} \leq min\{S_i, S_j\}$; and $L \geq |S_i - S_j|$. In the next section theoretical simulations are run in which these conditions can be met.

[21] The partial derivatives $\frac{\delta D_{Formula}}{\delta S_i}$ and $\frac{\delta D_{Formula}}{\delta S_j}$ are very similar in the sense that one can interchange the S_i and S_j to obtain the same formula, therefore $\frac{\delta D_{Formula}}{\delta S_j}$ is not shown.

A.5 Simulational exploration of the overestimation

For the theoretical simulations a simple co-occurrence matrix C depicted in Matrix 5 is used. Albeit it simple, this matrix allows for some exploration of the numerical difference between formula A.1 or formula A.2 for different values of S_i, S_j, C_{ij}, and L. In four different simulations, hypothetical and rather extreme situations are simulated to get insight on the effects of increasing the values of each of the variables S_i, S_j, C_{ij}, and L, while meeting the conditions $C_{ij} \leq min\{S_i, S_j\}$; and $L \geq |S_i - S_j|$.

Matrix 5

$$\begin{array}{c c}
Classes & \begin{array}{cccc} \text{a} & \text{b} & \text{c} & \text{d} \end{array} \\
\begin{array}{c} \text{a} \\ \text{b} \\ \text{c} \\ \text{d} \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}
\end{array}$$

In the first simulation, Matrix 5 is taken and the number of co-occurrences between c & d is increased by 1 in each step k, ceteris paribus. Matrix 6 gives this simulation:

Matrix 6

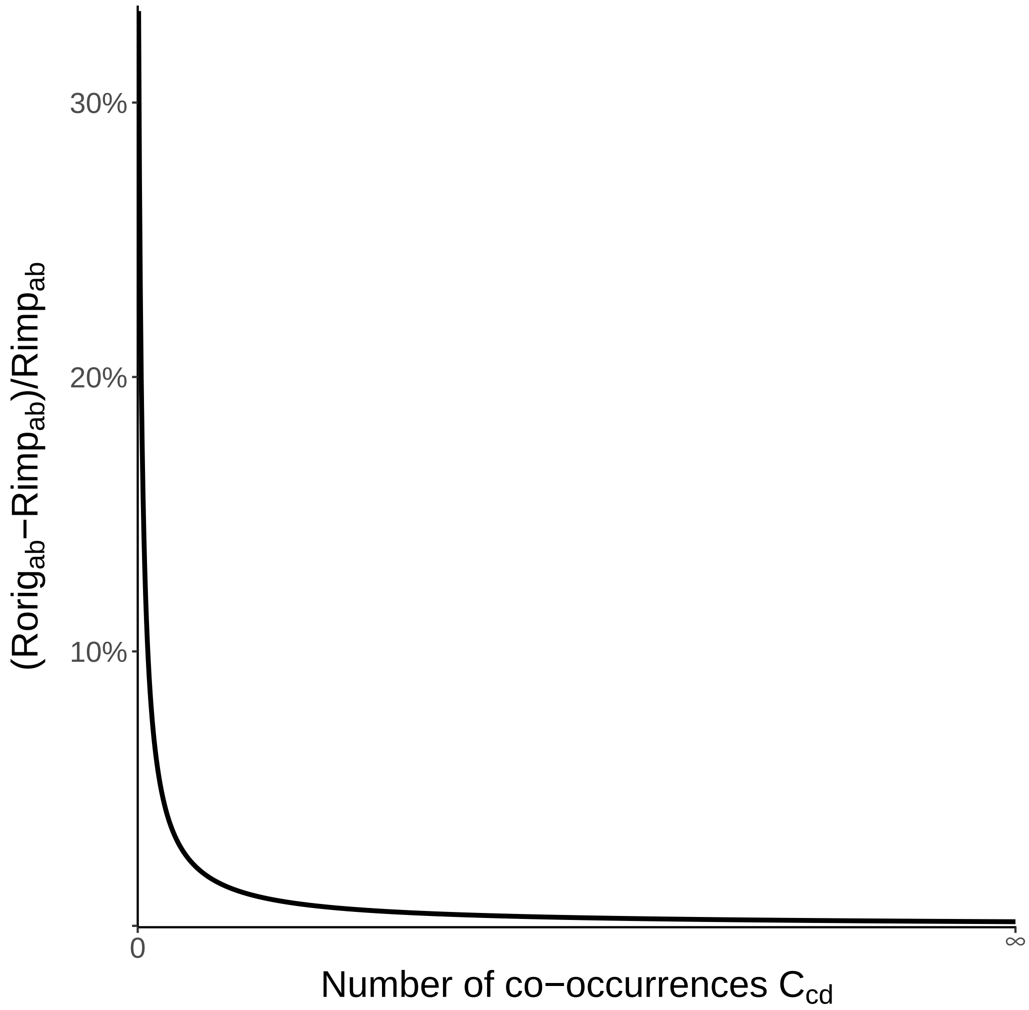
$$\begin{array}{c c}
Classes & \begin{array}{cccc} \text{a} & \text{b} & \text{c} & \text{d} \end{array} \\
\begin{array}{c} \text{a} \\ \text{b} \\ \text{c} \\ \text{d} \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & \mathbf{1+k} \\ 1 & 1 & 1+k & 0 \end{pmatrix}
\end{array}$$

In each step k the resulting relatedness matrix using formula A.1 is subtracted from the resulting relatedness matrix using formula A.2 and divided over the value of formula A.2 to express the difference in percentages. The relatedness values for the pairs a & b, and c & d are then plotted for each step. Each of these two changing relationships represent a different scenario:

- ***a* & *b*.** The changing difference in relatedness for the pair a & b simulates a steady increase in L, keeping $C_{ij} = 1$ and $S_i = S_j = 3$. This result is depicted in Figure A.1.

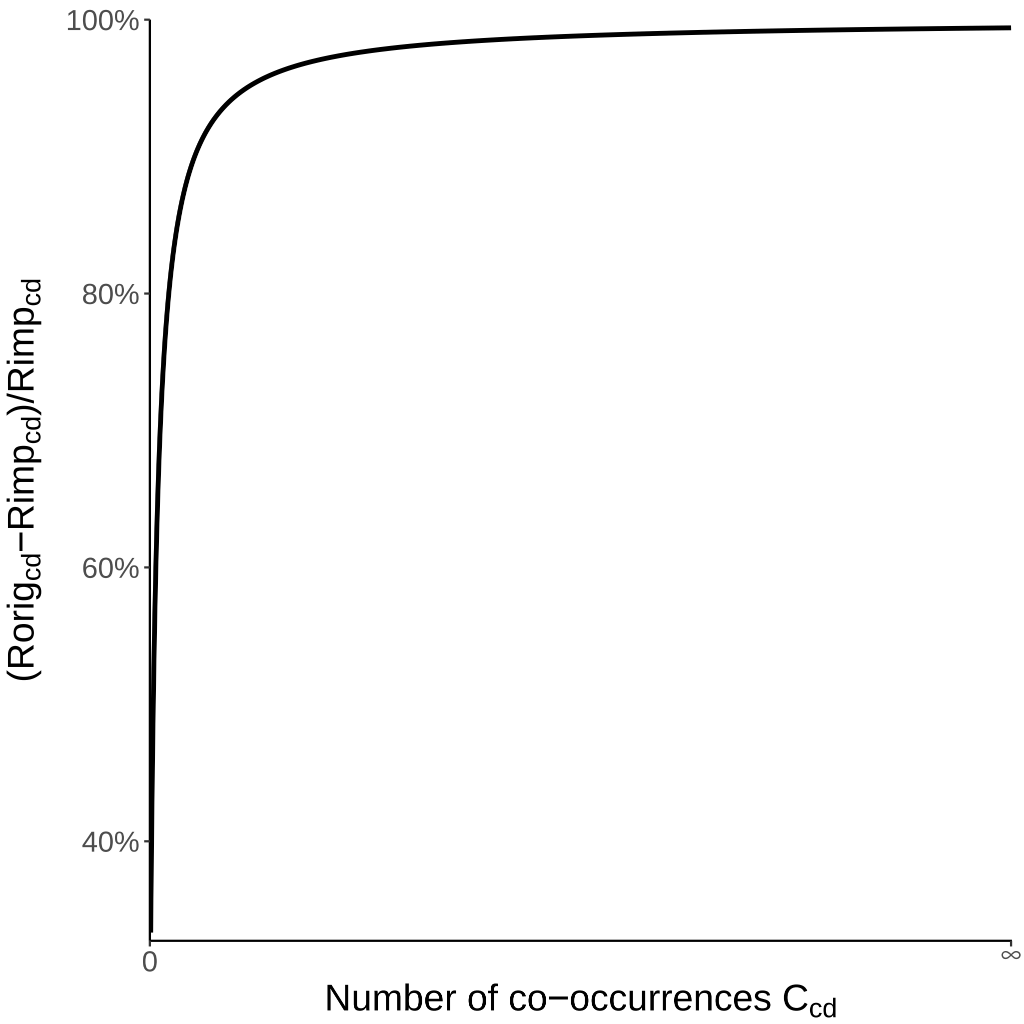
- ***c* & *d*.** The changing difference in relatedness between classes c & d simulates a steady increase in C_{ij} but also in S_i and S_j, keeping $L = 6$. To increase C_{ij} beyond the maximum value of S_i and S_j, S_i and S_j also have to increase. From the partial derivatives can be derived that an increasing C_{ij} would increase the difference whereas an increase in S_i and S_j can both increase or decrease the difference. The result of the simulation is depicted in Figure A.2.

FIGURE A.1 – THE DIFFERENCE IN RELATEDNESS BETWEEN THE ORIGINAL FORMULA AND THE IMPROVED FORMULA FOR CLASS a & b WHEN L INCREASES.



The absolute difference between the calculated relatedness of formula A.1 and formula A.2 for the pair a & b is equal to 1/3 across the entire simulation. However, as the number of other co-occurrences L increases, potential co-occurrence candidates increase as well and therefore the expected number of co-occurrences for a & b decreases. As a result, relatedness values are higher as L increases and the relative difference decreases, as can be seen in Figure A.1.

FIGURE A.2 – THE DIFFERENCE IN RELATEDNESS BETWEEN THE ORIGINAL FORMULA AND THE IMPROVED FORMULA FOR CLASS c & d WHEN C_{cd}, S_c AND S_d INCREASE.



For pair c & d, L remains equal to 6 but C_{cd}, S_c and S_d increase. Figure A.2 depicts how the difference in the estimated relatedness increases asymptotically converging from 33.3% to the value of 100%. As the $\frac{Observed}{Expected}$ should be close to one when two entities are close to having 100% of the occurrences in the sample but the values of the original formula A.1 converges to two the difference is close to 100% of the correct value.

To simulate an increase in C_{ij} while keeping S_i, S_j, and L equal, ceteris paribus, another simulation is needed: matrix 1 is altered by replacing the number of co-occurrences between entities a & b and c & d by a large amount of co-occurrences x.

Then in each step k of the simulation a co-occurrence is subtracted from this amount

x and added to the co-occurrences between entities a & d and b & c. See matrix 6. This keeps S_i, S_j and L equal but increases C_{ij} for the relatedness between a & d. Note that the result is insensitive to the exact value of x as the resulting change in the denominator and numerator cancel each other out.

Matrix 7

$$\left(\begin{array}{ccccc} Classes & \text{a} & \text{b} & \text{c} & \text{d} \\ \text{a} & 0 & x-k & 1 & 1+k \\ \text{b} & x-k & 0 & 1+k & 1 \\ \text{c} & 1 & 1+k & 0 & x-k \\ \text{d} & \mathbf{1+k} & 1 & x-k & 0 \end{array}\right)$$

The result is a stable overestimation of 33.3% for all values of k. When a & d co-occur more often but the total number of co-occurrences in the sample stays the same the relatedness between a & d naturally increases. Nonetheless, the increase in relatedness is proportional for the two formulas and therefore the difference remains 33.3%.

Lastly, an increase in S_i and S_j while keeping C_{ij} equal is simulated. The simulation is very similar to the first simulation except that next to increasing the co-occurrences between c & d also those between b & c is increased in each step k, see matrix 4. As a result, S_b and S_c increases while C_{bd} is kept at one. L increases necessarily as well in the form of S_c to match the added co-occurrences of S_b and S_d.

Matrix 8

$$\left(\begin{array}{ccccc} Classes & \text{a} & \text{b} & \text{c} & \text{d} \\ \text{a} & 0 & 1 & 1 & 1 \\ \text{b} & 1 & 0 & 1+k & \mathbf{1} \\ \text{c} & 1 & 1+k & 0 & 1+k \\ \text{d} & 1 & 1 & 1+k & 0 \end{array}\right)$$

Once again the percentual difference between calculating the level of relatedness for the pair b & d using formula A.1 and formula A.2 is stable at 33.3% for all values k. This time the relatedness between b & d decreases as k increases because their total number of occurrences S_b and S_d increase but their number of co-occurrences remains 1.

The simulations in this section show that the difference can range between close to 100% and close to 0. In real world applications of co-occurrence data the bias introduced by using formula A.1 instead of formula A.2 will be somewhere in between the extreme scenarios simulated here. In which each respective value in the relatedness matrix will be closer to a specific scenario than others.

A.6 Real world data-based exploration

The theoretical and simulational explorations demonstrate that formula A.1 overestimates the relatedness between entities compared to formula A.2 in a way that disproportionally affects certain pairs more than other pairs. However, the question remains how close these examples are to real world applications.

Therefore, the outcomes of formula A.1 and formula A.2 are compared using USPTO technology class data, see Hall et al. (2001) and USPTO, from utility patents in periods of 5 years from 1855 to 2014.[22]

In the occurrence matrix O of each time period the rows indicate patent numbers and the columns technology classes, like the example in Matrix 1. By multiplying the transpose of O by O itself a technology classes by technology classes co-occurrence matrix C is obtained. As before, the diagonal of C is set to zero and S_i can then be calculated as the column sum of column i or the row sum of row i.[23] Next formula A.1 and formula A.2 are calculated using the C of each time period and the results are compared in Tables A.1 and A.2.

Tables A.1 and A.2 give a number of statistics for each time period mentioned in the respective header. The first row gives the number of different technology classes (n) referred to on the patents. This number is equal to the number of columns/rows in C. The second line gives the number of pairs that have a value higher than 1 according to formula A.1 by van Eck and Waltman (2009), these relatedness pairs have more or just as much observed co-occurrences as expected and are therefore seen as related in research within this domain (see for example Balland et al., 2015). The third line gives the same statistic but employs the improved formula A.2. On line four the difference between the number of related pairs according to each formula is given.[24]. Difference (%) expresses this difference as a percentage of the number of related pairs according to the improved formula A.2.

Focussing on these first five statistics it can be seen that in 1855 to 1859 patents made references to 327 different technology classes and that according to formula A.1 5154 pairs of technology classes can be seen as related, while formula A.2 identifies 5150 related pairs. As a result, formula A.1 identifies 4 pairs or $\frac{4}{5150} \times 100 = 0.07\%$ more as related than formula A.2.

In later time periods the differences increase both in absolute terms as in relative

[22]A period of 5 years is also used by Boschma et al. (2015).

[23]Note that the relatedness function in the EconGeo package for R (see Balland, 2016) sets the diagonal of the input co-occurrence matrix to zero automatically.

[24]Note that there are no pairs identified as related by formula A.2 that are identified as unrelated by formula A.1, as formula A.1 > formula A.2, when $C_{ij} > 0$. See also Section A.4.

terms with a maximum in relative terms of 0.29% in 1885-1889 and a maximum in absolute terms with 62 pairs wrongly seen as related in 1955-1959.

Next to the overestimation another problem of using formula A.1 instead of formula A.2 is that the relatedness between some pairs is more overestimated than between other pairs. The last four statistics explore this disproportionality. The largest difference in value gives the largest difference in the relatedness value of a single pair between formula A.1 and formula A.2, while its percentage counterpart gives the largest overestimation relative to the value given by formula A.2. In relative terms the highest over estimation is 3.23% and occurs in 2000-2004, this percentage is way below some of the extreme scenarios simulated in Section A.5. The largest absolute difference is 0.837 in 1860-1864.

The last two statistics are similar but give the smallest difference, when $C_{ij} > 0$.[25] When at least one co-occurrence exists between a pair its relation is overestimated as already shown mathematically in Section A.4. The values are close to zero both in absolute terms as in relative terms and therefore in strong contrast to the highest values, showing that some pairs get more overestimated than others.

The results also show that there is not necessarily a direct connection between the number of technology classes and the number of related pairs or the overestimation. In 2000-2004, there is the second highest number of different technology classes, while the number of related pairs is lower than in 1950-1954 when fewer technology classes were in use.

When comparing these specific time periods, 2000-2004 turns out to have a much more concentrated co-occurrence matrix C than the one in 1950-1954. In 2000-2004 each row or column i contains a few pairs with a lot of observations while others have relatively few observations. This contrasts with the more even spread of observations across C in 1950-1954. The average Gini coefficient per row of C in 2000-2004 is 0.936 versus 0.909 in 1950-1954.

Very much like the simulation based on matrix 7, where S_i and S_j was increased while keeping C_{ij} equal, the pairs with little co-occurrences are less overestimated when there are more occurrences of the same technology class with other classes, as is more the case in 2000-2004. The pairs with relatively high numbers of co-occurrences have a larger share of the sample in 2000-2004 compared to 1950-1954, like in matrix 6, where C_{ij} is increased while S_i and S_j are kept equal, these pairs are more overestimated in 2000-2004. The pairs with relatively many co-occurrences are likely to pass the

[25] When $C_{ij} = 0$ both formulas return 0 and the difference is therefore also zero and obviously the smallest.

threshold of 1 using either formula, the stronger overestimation for these pairs in 2000-2004 does not lead to much change with respect to passing this threshold. This is not the case for the pairs with relatively fewer co-occurrences, which are less overestimated in 2000-2004 than in 1950-1954. Therefore in 2000-2004, these are less likely to pass the threshold irrespective of whether formula A.1 or formula A.2 is used. While in 1950-1954 these pairs are more likely to pass the threshold using formula A.1 but not when using formula A.2. As a result, 2000-2004 has larger overestimations of individual relatedness values but less pairs that are wrongly identified as related.

The comparison shows that using formula A.1 instead of formula A.2 in research can lead to non-negligible differences and that some pairs and matrices are affected disproportionally. Note that with an incorrect specification of S_i, S_j and m formula A.1 becomes even more inaccurate, see Section A.4. It is unlikely that papers employing formula A.1 instead of formula A.2 would have reached fundamentally different conclusions but a risk is more present in some cases than others. It is recommended to use formula A.2 in future research.

Table A.1 – Patent comparison results - Part 1

	1855-9	1860-4	1865-9	1870-4	1875-9	1880-4	1885-9	1890-4
Number of technology classes	327	335	343	356	361	372	379	385
Number of related pairs (Original formula)	5154	4902	7910	8954	10100	12396	13438	13484
Number of related pairs (Improved formula)	5150	4898	7892	8934	10080	12370	13398	13464
Difference	4	4	18	20	20	26	40	20
Difference (%)	0.07	0.08	0.22	0.22	0.19	0.21	0.29	0.14
Largest difference in value	0.827	0.837	0.788	0.786	0.822	0.662	0.593	0.63
Largest difference (%) in value	2.643	2.177	2.009	1.961	2.258	2.333	2.425	2.36
Smallest difference in value	0.00599	0.00505	0.00234	0.00169	0.0011	0.00107	0.00084	0.00082
Smallest difference (%) in value	0.0294	0.0268	0.01	0.0075	0.0085	0.0037	0.004	0.0032
	1895-9	1900-4	1905-9	1910-4	1915-9	1920-4	1925-9	1930-4
Number of technology classes	385	387	390	394	403	404	405	415
Number of related pairs (Original formula)	14196	15866	16372	16742	17784	18036	19560	21432
Number of related pairs (Improved formula)	14160	15842	16338	16694	17754	17990	19528	21396
Difference	36	24	34	48	30	46	32	36
Difference (%)	0.25	0.15	0.20	0.28	0.16	0.25	0.16	0.16
Largest difference in value	0.625	0.515	0.586	0.666	0.645	0.753	0.711	0.677
Largest difference (%) in value	2.568	2.341	2.303	2.441	2.536	2.173	1.933	1.872
Smallest difference in value	0.00063	0.00056	0.00042	0.00051	0.00038	0.00039	0.00023	0.00018
Smallest difference (%) in value	0.0026	0.0054	0.0055	0.0071	0.0052	0.0023	0.0036	0.0071

Notes: A pair is seen as related when the respective formula returns a value of 1 or higher for a certain pair. The statistics expressed in percentages are taken with respect to the value returned by the improved formula A.2.

TABLE A.2 – PATENT COMPARISON RESULTS - PART 2

	1935-9	1940-4	1945-9	1950-4	1955-9	1960-4	1965-9	1970-4
Number of technology classes	414	417	413	423	427	430	432	434
Number of related pairs (Original formula)	22852	23430	23336	25104	24422	25326	25932	25590
Number of related pairs (Improved formula)	22814	23388	23280	25060	24360	25280	25902	25544
Difference	38	42	56	44	62	46	30	46
Difference (%)	0.16	0.17	0.24	0.17	0.25	0.18	0.11	0.18
Largest difference in value	0.557	0.56	0.525	0.492	0.557	0.529	0.579	0.661
Largest difference (%) in value	1.641	1.76	1.772	1.726	1.51	1.561	1.602	1.892
Smallest difference in value	0.00015	0.00015	0.00022	0.00014	0.00014	0.00011	0.00009	0.00008
Smallest difference (%) in value	0.003	0.0034	0.0063	0.0019	0.0029	0.0018	0.0015	0.0006
	1975-9	1980-4	1985-9	1990-4	1995-9	2000-4	2005-9	2010-4
Number of technology classes	436	435	435	435	431	437	436	438
Number of related pairs (Original formula)	25350	25012	24712	23982	24120	24422	24356	26382
Number of related pairs (Improved formula)	25324	24980	24676	23928	24084	24388	24310	26348
Difference	26	32	36	54	36	34	46	34
Difference (%)	0.10	0.12	0.14	0.22	0.14	0.13	0.18	0.12
Largest difference in value	0.684	0.694	0.69	0.524	0.501	0.581	0.592	0.64
Largest difference (%) in value	2.29	2.52	2.192	2.293	2.404	3.234	3.176	2.834
Smallest difference in value	0.00008	0.00008	0.00006	0.00005	0.00005	0.00004	0.00003	0.00002
Smallest difference (%) in value	0.0018	0.0033	0.004	0.0028	0.0033	0.0036	0.0013	0.0012

Notes: A pair is seen as related when the respective formula returns a value of 1 or higher for a certain pair. The statistics expressed in percentages are taken with respect to the value returned by the improved formula A.2.

A.7 Conclusion

Co-occurrence data is commonly used in various domains. Researchers generally apply normalisation measures to correct for the size-effect. To this end, van Eck and Waltman (2009) make a convincing case to use a probability-based measure known as the association strength. In which the number of observed co-occurences is divided over the number of expected co-occurrences, assuming that observations are randomly distributed over co-occurences.

However, the probability formula to calculate the expected number of co-occurrences is not suited for the co-occurrence analysis it is recommended for, which is when self co-occurrences are non-existent or irrelevant.[26] The formula assumes combinations with repetition meaning that an observation from an entity can be drawn again after been picked in the first draw even this occurrence nor any other occurrence belonging to the same entity can be drawn in this line of work.

this Chapter introduces a formula that is based on, but not equal to, combinations *without* repetition in which the probability of drawing entity i and j together is calculated as the probability of drawing i first and then j, knowing that none of the observations pertaining to i can be drawn plus the probability of drawing j and then i, knowing that none of the observations pertaining to j can be drawn. This formula gives the correct results, as was demonstrated in an intuitive example.

Furthermore, it is shown that the original formula overestimates the relatedness between a pair of entities compared to the improved formula introduced here, when there is at least one observed co-occurrence, and that the overestimation is not proportional across pairs. Simulations show that the over estimation of the relatedness can range between virtually 0% and almost 100% of the correct value given by the improved formula. In a real world example, a number of patent samples showed that the overestimation of individual values was between virtually 0% and 3.234%, while the difference in the number of pairs that can be seen as related can be 0.29% more than the number of pairs identified as related by the improved formula.

All in all, this Chapter shows that the formula presented here is better equipped for the analysis of co-occurrence data. The formula, following all recommendations for inputs and treatment, is available in the EconGeo package for R maintained by Balland (2016).

[26] An interesting avenue for future research may be to more clearly determine in which situations self co-occurrences can be disregarded or not.

Nederlandse samenvatting

Agglomeratievoordelen staan centraal in het begrijpen van de ruimtelijke verdeling van menselijke activiteiten. Deze voordelen voor individuen en organisaties komen voort uit de nabijheid van anderen. De afgelopen jaren is de interesse in dit onderwerp toegenomen omdat het verschil in inkomen tussen regio's groeit sinds de jaren '80. Aan de ene kant zijn er enkele innovatieve stedelijke regio's die sterk groeien maar ook geconfronteerd worden met sterk stijgende huizenprijzen en buitensluiting. Terwijl er aan de andere kant vele ruralere en meer afgelegen regio's zijn die werkgelegenheid en leefbaarheid zien afnemen. Ondanks deze maatschappelijke kwesties is de kennis over agglomeratievoordelen nog sterk onderontwikkeld, met name: punt 1, via welke mechanismes deze lokale voordelen worden overgebracht; punt 2, hoe ze veranderd zijn over de tijd; en, punt 3, welke factoren van economische verandering ze beïnvloeden (Glaeser, 2011; Moretti, 2012; Combes and Gobillon, 2015).

Met betrekking tot het eerste punt zijn er vele theorieën die als uitgangspunt kunnen dienen zoals: de onderverdeling van agglomeratievoordelen van Marshall (1890) in het uitwisselen van werknemers, producten/diensten en ideeën, die in onderzoeksvragen #1 en #2 behandeld wordt; en arbeidsdeling, het opsplitsen van arbeidstaken over meerdere personen, vaak toegekend aan Smith (1776), dat in onderzoeksvraag #3 behandeld wordt. Daarnaast wordt er in het laatste hoofdstuk ook onderscheid gemaakt of agglomeratievoordelen direct tot meer productiviteit leiden of pas over langere tijd. De colocatie van activiteiten kan namelijk ook tot nieuwe activiteiten leiden die economische groei kunnen creëren lang nadat de relevantie van oorspronkelijke activiteiten is afgenomen, zoals de haven van Amsterdam de kiem heeft gelegd voor de huidige financiële sector. Dit is het onderwerp van onderzoeksvraag #4.

Empirisch is er ook een uitgebreide basis om deze conceptualisaties van agglomeratievoordelen te meten. Voor onderzoeksvragen #1 en #2 kan er gebouwd worden op Ellison et al. (2010), die hebben gemeten in hoeverre paren van industrieën samen voorkomen in dezelfde stad en dat vergeleken met de mate waarin deze industrieën dezelfde werknemers en ideeën gebruiken en producten/diensten van elkaar afnemen. Voor onderzoeksvraag #3 kan arbeidsdeling inzichtelijk worden gemaakt door te kijken naar het aantal academici, uitvinders en werknemers per, respectievelijk, wetenschappelijk artikel, patent en baancategorie/industrie en in relatie tot de mate van ruimtelijke concentratie van deze activiteiten. Voor onderzoeksvraag #4 kan er omtrent de evolutie van nieuwe activiteiten uit voorgaande activiteiten gebouwd worden op een uitgebreide literatuur met betrekking tot *relatedness* (gerelateerdheid) dat de mogelijkheid biedt om te meten in hoeverre twee activiteiten vergelijkbare vaardigheden en kennis vereisen.

Door nieuwe datatechnieken toe te passen zoals *Optical Character Recognition (OCR)* wordt er in dit proefschrift gedetailleerdere data over een grotere tijdspanne verzameld waardoor het niet alleen mogelijk is om de agglomeratiemechanismes accurater te meten dan voorgaande literatuur, zie punt 1, maar vooral om inzicht te geven in de verandering van de relevantie van deze agglomeratiemechanismes over de tijd, zie punt 2.

De grootste uitdaging van dit proefschrift is echter met betrekking tot punt 3, het verklaren van de verandering in agglomeratiemechanismes, omdat de theoretische literatuur en empirie hierover niet eenduidig zijn. Om dit op te lossen pas ik in dit proefschrift ideeën toe uit aanpalende vakgebieden: arbeidseconomie, complexiteitstheorie en innovatiewetenschappen.

Een belangrijk inzicht hieruit is dat innovatie niet incrementeel is maar af en toe onderbroken wordt door radicale uitvindingen die de economie ingrijpend veranderen (Helpman and Trajtenberg, 1998). De derde industriële revolutie, bijgenaamd de computerrevolutie, verhoogde namelijk vooral de efficiëntie van innovatieve en kennisintensieve werknemers en bedrijven die zich vaker richten op timing-kwaliteit-differentiatie van niche producten (Brynjolfsson and Hitt, 2000; Autor et al., 2003). Dit ging juist ten koste van routinematige werknemers en bedrijven die zich vaker richten op prijs-differentiatie door middel van gestandaardiseerde massaproductie. Technologische vooruitgang in combinatie met het wegnemen van handelsbarrières hebben ook geleid tot toenemende importconcurrentie uit lagelonenlanden die met name ook concureren op prijs met massaproductie (Bloom et al., 2016). Deze transformatie in hogelonenlanden heeft waarschijnlijk ook geleid tot veranderingen in de relevantie van bepaalde agglomeratiemechanismes. Een andere veelvoorkomende uitleg in de stedelijke economie is de daling van transportkosten die nabijheid minder belangrijk kan maken, zeker met betrekking tot het verplaatsen van goederen (Glaeser and Kohlhase, 2004). Daarom wordt in vraag #2 gekeken in hoeverre veranderingen in importconcurrentie, technologische veranderingen en transportkosten van goederen samenhangen met de (veranderingen in de) agglomeratiekrachten van Marshall.

Industriële revoluties hebben zeer waarschijnlijk ook een grote invloed op arbeidsdeling. De relatie tussen ruimtelijke concentratie en arbeidsdeling is afhankelijk van twee samenhangende componenten; de mate in welke twee personen elkaar moeten ontmoeten om de vruchten van hun arbeidstaken uit te wisselen en de mate in welke arbeidstaken over meerdere mensen verdeeld kan worden (Smith, 1776; Hausmann et al., 2014). De literatuur toont dat mensen meer behoefte hebben elkaar fysiek te ontmoeten als de kennis die uitgewisseld wordt nieuwer en complexer is en vertrouwen moet worden opgebouwd tussen personen (Storper and Venables, 2004). Ook is er de tendens dat met de voortschrijding van kennis personen zich moeten specialiseren een nauwer afgebakend vakgebied en daardoor moeten samenwerken om nieuwe doorbraken te bewerkstelligen omdat het voor één persoon niet meer mogelijk is om al die kennis te hebben (Jones, 2009). Wat suggereert dat complexe economische activiteiten zich steeds meer in grote steden concentreren. Dit wordt uitgezocht in onderzoeksvraag #3.

Industriële revoluties zijn ook perioden waarin het ontwikkelen van activiteiten vitaal is voor regio's omdat als zij gespecialiseerd zijn in activiteiten die niet meer kunnen concurreren er een diepe crisis kan ontstaan en werkgelegenheid en leefbaarheid kan afnemen (Boschma, 2015). Crises kunnen dus ook een factor van economische verandering zijn. De literatuur suggereert daarbij dat steden met een diverse economie bestaande uit veel verschillende industrieën meer combinaties van kennis en vaardigheden kan maken om nieuwe activiteiten te ontwikkelen. Ook is er in casestudies gesuggereerd dat er in dit soort steden minder dominante spelers uit één industrie zijn die nieuwe ontwikkelingen kunnen blokkeren (Grabher, 1993; Neffke et al., 2018). Door te bouwen op grootschalige data-analyses wordt gekeken in hoeverre diverse steden anders diversificeren dan gespecialiseerde steden in onderzoeksvraag #4.

Onderzoeksvraag #1 In hoeverre is de relevantie van elk van Marshalls agglomeratiekrachten veranderd in de loop van de tijd?

In het eerste hoofdstuk, in samenwerking met Hans Koster en Frank van Oort, wordt ingegaan op de drie bronnen van agglomeratievoordelen die Marshall (1890) beschreef: arbeidsmarktdeling, verticale relaties en kennisuitwisseling. Voortbouwend op Ellison et al. (2010) wordt er gekeken per paar van industrieën in hoeverre vestigingen van elk in dezelfde steden voorkomen en dat vergeleken met proxies voor arbeidsmarkt deling en verticale relaties. Deze proxies zijn, respectievelijk, de correlatie in het aandeel werknemers per baancategorie tussen industrieën en het aandeel producten dat de ene industrie aan de ander afneemt/toelevert als proxies voor, respectievelijk, arbeidsmarktdeling en verticale relaties. Voor kennisuitwisseling gebruiken wij geen patencitaties als proxy, zoals de voorgaande literatuur, maar *technological relatedness* omdat deze alleen gebruik maakt van de technologieën die vermeld zijn op de patenten en normaliseert voor de grootte van beide industrieën. Voor deze maatstaf zijn een nieuwe formule en code voor R ontwikkeld, zoals uitgelegd in de bijlage van dit proefschrift.

Voor dit hoofdstuk is een uitgebreide longitudinale dataset met consistente industrieclassificatie en administratieve grenzen samengesteld voor de 363 grootste steden in de V.S. gedefinieerd als Metropolitan Statistical Area (MSA) zoals vastgesteld door het Censusbureau van de V.S. Door met OCR data te verzamelen van gescande documenten uit het verleden is het mogelijk om terug te gaan tot 1970 voor coagglomeratie, Marshalls agglomeratiekrachten, controlevariabelen en instrumentele variabelen. De controlevariabelen meten de gezamenlijke afhankelijkheid van verscheidene geografische inputs buiten de vervaardigingsindustrie die ook coagglomeratie kunnen bepalen. Bijvoorbeeld omdat twee industrieën beiden afhankelijk zijn van dezelfde grondstoffen of infrastructuur als havens. De controlevariabelen worden ook gebruikt om andere controlevariabelen waarvoor geen data is te simuleren zoals voorgesteld door Oster (2019). Instrumentele variabelen zijn nodig omdat coagglomeratie mogelijk niet het gevolg maar de oorzaak kan zijn van Marshalls agglomeratiekrachten. Een vestiging kan bijvoorbeeld dezelfde soort werknemers inhuren nadat ze in de buurt van een andere industrie zijn gaan vestigen in plaats van andersom. Door data te verzamelen over de soort technologieën en werknemers die industrieën gebruiken in regio's waar een andere industrie niet aanwezig is kan geschat worden welk deel van de arbeidsmarktdeling en kenniswisseling onafhankelijk is van de coagglomerende industrie. Deze zogenaamde ruimtelijke instrumenten zijn niet beschikbaar voor verticale relaties omdat data over het aanbod en gebruik van industrieën op lokaal niveau niet beschikbaar is.

De resultaten laten zien dat gemiddeld tussen 1970 en 2014 arbeidsmarktdeling de belangrijkste reden voor coagglomeratie is gevolgd door kennisuitwisseling en verticale relaties. Deze volgorde is omgekeerd ten opzichte van de resultaten van Ellison et al. (2010) die data uit 1987 en een andere proxy voor kenniswisseling gebruikten. De analyses laten ook zien dat *technological relatedness* ook empirisch een sterkere proxy is voor kennisuitwisseling dan patent citaties.

Het interessantst zijn de ontwikkelingen over tijd. Tussen 1970 en 2014 is kennisuitwisseling een steeds belangrijkere factor geworden voor coagglomeratie. Dit is de eerste keer dat deze trend gedocumenteerd wordt terwijl die door een uitgebreide literatuur gesuggereerd wordt als de belangrijkste reden dat geografische nabijheid relevanter is geworden in een tijd dat het digitaal mogelijk is om de hele wereld te bereiken. Tegelijkertijd neemt het belang van arbeidsmarktdeling en verticale relaties af over dezelfde periode.

Onderzoeksvraag #2 Waarom is de relevantie van elk van Marshalls agglomeratiekrachten veranderd over de tijd?

In het eerste hoofdstuk wordt ook ingegaan op de redenen achter de gedocumenteerde trends in Marshalls agglomeratiekrachten. Waar de literatuur eensgezind is over de trend in kennisuitwisseling zijn er verschillende verwachtingen uitgesproken over trends in arbeidsmarktdeling en verticale relaties. Zo verwacht Moretti (2012) dat met het toenemende opleidingsniveau van werknemers arbeidsmarktdeling belangrijker wordt als vestigingsreden terwijl Faggio et al. (2017) in een doorsnedeanalyse laten zien dat kennisintensieve bedrijven juist arbeidsmarktdeling minder belangrijk vinden. Dit komt volgens hun, op basis van de *nursery city* hypothese van Duranton and Puga (2001), omdat arbeidstaken minder gestandaardiseerd zijn en werknemers daarom lastiger in te wisselen zijn met nabijgelegen bedrijven terwijl kennisuitwisseling in die fase erg belangrijk is. De verschuiving naar een kennisintensievere economie zou dan betekenen dat arbeidsmarktdeling minder belangrijk wordt. Verticale relaties zijn in de verwachting van McCann and Fingleton (1996) en Duranton and Storper (2008) belangrijker geworden omdat *just-in-time* leveringen en *face-to-face* contact in handelsrelaties belangrijker zijn geworden. Aan de andere kant suggereert Glaeser and Kohlhase (2004) dat de dalende trend van de transportkosten van goederen ertoe zal leiden dat bedrijven met verticale relaties niet meer dichtbij elkaar hoeven te zitten. Een andere reden waarom verticale relaties in belang afnemen zijn de bevindingen van Faggio et al. (2017) dat verticale relaties minder belangrijk zijn voor kennisintensieve bedrijven geeft.

Deze discussie laat zien dat er nog veel onenigheid heerst in het veld met betrekking tot het belang van nabijheid en agglomeratievoordelen in het bijzonder. Uitgangspunt hierbij zijn transportkosten. Als die afwezig zijn is het namelijk niet de moeite waard de hoge kosten van land in steden te betalen. Transportkosten bestaan uit monetaire kosten, zoals de prijs van een vrachtwagen of treinkaartje, en tijdskosten, het productiviteitsverlies van een persoon die reist bijvoorbeeld om *face-to-face* af te spreken. De factoren die het eerste beïnvloeden zijn relatief makkelijk vast te stellen. De factoren die het tweede beïnvloeden zijn een stuk lastiger vast te stellen maar omdat het salaris een goede indicator is van tijdskosten is het hier toepasselijk om ideeën uit de arbeidseconomie toe te passen.

In dat vakgebied is veel aandacht voor de trend in stijgende inkomensongelijkheid tussen hoogopgeleiden en lager opgeleiden sinds het einde van de jaren '70 van de vorige eeuw als gevolg van technologische ontwikkeling en importconcurrentie (Autor et al., 2015). Waar de industriële revolutie omtrent elektriciteit leidde tot massaproductieprocessen die gestandaardiseerd waren en veel laag- en middenopgeleide werknemers voor routinematige taken zocht leidde de computerrevolutie juist tot een groeiende vraag naar hoogopgeleide werknemers die nieuwe producten en productieprocessen konden ontwikkelen en programmeren die weer veel routinematige banen wegautomatiseerden (Goldin and Katz, 1998; Brynjolfsson and Hitt, 2000). Importconcurrentie uit lagelonenlanden erodeerde de concurrentiepositie van arbeidsintensieve laagtechnologische vestigingen in hogelonenlanden verder. Hiermee stegen de lonen juist voor creatieve hoogopgeleide werknemers en aangezien deze *face-to-face* contact nodig hebben om ideeën uit te wisselen is het belangrijk voor deze werknemers om dicht bij anderen te zijn. Technologische vooruitgang en importconcurrentie vormen daarom logische oorzaken voor de gedocumenteerde trends in Marshalls agglomeratiekrachten naast de eerdergenoemde daling in de transportkosten van goederen.

Qua data en maatstaven wordt technologische vooruitgang gemeten door het aandeel werknemers met grotendeels routinematige taken te meten per industrie en over de tijd, zie Autor et al. (2015). Importconcurrentie uit lagelonenlanden wordt gemeten door het aandeel producten geïmporteerd uit lagelonenlanden, gedefinieerd als landen die 15% van het bnp per hoofd van de bevolking hebben van de VS, in het totaal van de import van producten per industrie en over de tijd, zie Bernard et al. (2006). Voor de transportkosten van goederen gebruiken we het aandeel van de uitgaven aan transportbedrijven in het totaal van uitgaven per industrie en over de tijd.

Ook in deze analyse is er risico op wederzijdse causaliteit. Industrieën die namelijk sterk coagglomeren vanwege een bepaalde Marshalliaanse agglomeratiekracht kunnen juist inzetten op technologische investeringen en importsubstitutie. Verder kan een laag aandeel aan transportkosten ook komen doordat industrieën bij elkaar zitten om verticale relaties te onderhouden. Om dit tegen te gaan worden er instrumenten ontwikkeld die, respectievelijk, meten: wat het aandeel routinematige werknemers is in gebieden waar industrieën niet zijn om van een bepaalde agglomeratiekracht gebruik te maken, zoals de ruimtelijke instrumenten eerder; het aandeel import uit lagelonenlanden in andere hogelonenlanden; en de gemiddelde waarde van een ton aan producten van een industrie. De redenering achter dit laatste instrument is dat bij een hoge waarde van een ton transportkosten minder relevant zijn in de totale prijs.

Ook in deze analyse wordt gecontroleerd voor onmeetbare ontbrekende variabelen in navolging van de methode van Oster (2019) worden er namelijk twee variabelen toegevoegd die de kapitaalintensiteit en uitgaven aan R&D meten. Deze variabelen kan men als *proxy controls* zien, zie Angrist and Pischke (2008), omdat ze deels het effect oppakken van technologische ontwikkeling en importconcurrentie. Ook hier wordt de data verzameld van 1970 tot 2014 en waar nodig omgezet naar dezelfde industriële classificatie voor consistentie.

Qua methodologie wordt een vernieuwende twee-staps-methode toegepast. In de eerste stap wordt per industrie en tijdsperiode een coëfficiënt geschat voor elk van de agglomeratiekrachten. Met acht periodes en 140 industrieën levert dit 1.120 observaties op per agglomeratiekracht die in de tweede stap worden gebruikt als afhankelijke variabele waarbij de maatstaven voor technologie, handel en transport de onafhankelijke variabelen zijn. Om standaardfouten te berekenen wordt gebruik gemaakt van *boot-strapping*.

De resultaten tonen dat technologische vooruitgang en importconcurrentie van sterke invloed zijn op arbeidsmarktdeling in negatieve zin en kennisuitwisseling in positieve zin. Dit suggereert dat de opkomst van (computer)technologie en handelcompetitie sinds de jaren '70 sterk verbonden is met de neergang van arbeidsmarktdeling en de opkomst van kennisuitwisseling als coagglomeratiemotief. Dit is in lijn met de verwachtingen op basis van de resultaten over kennisintensieve industrieën van Faggio et al. (2017). Voor verticale relaties blijken, in tegenstelling tot de verwachtingen van Glaeser and Kohlhase (2004), transportkosten van goederen geen belangrijke factor te zijn. De enige variabele die een statistisch en economisch significante invloed heeft op verticale relaties is importconcurrentie. Dit lijkt echter niet te komen omdat dit tot meer kennisintensieve industrieën leidt, zoals op basis van Faggio et al. (2017) verwacht kan worden, omdat in aanvullende analyses andere variabelen voor kennisintensieve eigenschappen, zoals R&D uitgaven en vaardigheid van werknemers, net als technologische vooruitgang ook geen rol spelen. Wel blijkt dat vooral importcompetitie invloed heeft op toeleverende verticale relaties. Dit suggereert dat het belang van coagglomeratie voor verticale relaties afneemt omdat lokale toeleveranciers worden vervangen door toeleveranciers uit lagelonenlanden.

Onderzoeksvraag #3: Concentreren complexe economische activiteiten zich in grote steden?

In dit hoofdstuk, in samenwerking met Pierre-Alexandre Balland, Cristian Jara-Figueroa, Sergio Petralia, David Rigby en César Hidalgo, staat arbeidsdeling als agglomeratievoordeel en voortschrijdend menselijk inzicht als factor van economische verandering centraal. Arbeidsdeling wordt meestal geassocieerd met Smith (1776) maar de eerste noties ervan waren al bekend bij Plato en Aristoteles in de Griekse oudheid. Door arbeidstaken over een groter aantal personen te verdelen kan ieder zich specialiseren in een kleinere subset van taken en stijgt de efficiëntie van de productie. Door de vruchten van ieders taken uit te wisselen lukt het één ieder om meer producten/diensten te hebben dan als één ieder zelfstandig zou produceren. Als voor deze uitwisseling fysieke nabijheid nodig is kan de verhoogde productiviteit degenen compenseren voor hogere locatiekosten. Zo vergeleek Smith (1776) bewoners van de Schotse hooglanden waar ieder huishouden zowel brouwer, slager als bakker was met stadsbewoners waar dit zelfstandige beroepen zijn.

De enorme groei van steden van de afgelopen decennia suggereert dat fysieke nabijheid in arbeidsdeling alleen maar belangrijker is geworden. Dit terwijl de transport- en communicatiekosten dusdanig zijn afgenomen dat ook de huidige bewoners van de Schotse hooglanden zich kunnen specialiseren en niet meer alles zelf hoeven te produceren.

De belangrijkste reden hiervoor is volgens een uitgebreide literatuur dat *face-to-face* contact niet vervangen kan worden door telecommunicatie om complexe informatie over te brengen en vertrouwen tussen personen op te bouwen en dat deze twee aspecten steeds belangrijker zijn geworden (Storper and Venables, 2004; McCann, 2008; Glaeser, 2011). Om de groei in steden te begrijpen is het dus nodig om te begrijpen waarom beide redenen voor face-to-face contact belangrijker zijn geworden. Een groot deel van de literatuur in de stedelijke economie doet dit door kwantitatieve maatstaven als het aantal patenten en R&D uitgaven te meten en te tonen dat deze activiteiten geografisch meer geconcentreerd zijn dan andere menselijke activiteiten, zowel in de economische geografie/stedelijke economie, zie Carlino and Kerr (2015) als in de literatuur gebaseerd op machtswetten (*power laws*) zie Bettencourt et al. (2007). Echter is er een grote verscheidenheid in de complexiteit van kennis per patent of R&D activiteit en biedt dit geen theoretische fundering om te begrijpen waarom beide redenen voor *face-to-face* contact belangrijker worden.

Hierom worden in dit hoofdstuk meetbare maatstaven voorgesteld om de *kwalitatieve* aspecten van kennis te meten gebaseerd op ideeën uit innovatie studies en complexiteitstheorie. Zo tonen Breschi and Lissoni (2001) aan dat *face-to-face* contact met name belangrijk is als de organisatie van taken, terminologie van informatie en vertrouwensrelaties nog relatief onbekend zijn. Verder laat Jones (2009) zien dat over tijd met de voortschrijding van menselijk inzicht binnen het onderzoek het onmogelijk is voor één persoon om van alle (sub)disciplines alles te leren om op de grens van de menselijke kennis te zitten. Om toch tot nieuwe uitvindingen te komen specialiseren uitvinders zich in een steeds beperktere set van (sub)disciplines en werken ze samen met steeds grotere teams om alle puzzelstukjes van kennis bij elkaar te leggen. In de complexiteitstheorie brengen Fleming and Sorenson (2001) deze beide componenten samen door te laten zien dat patenten innovatiever zijn als ze meer verschillende stukken kennis samenbrengen, N, en stukken kennis die minder vaak samen gebruikt zijn, K. Hausmann et al. (2014) gebruiken een vergelijkbare denkwijze en bouwen voort op arbeidsdeling door te stellen dat welvarendere landen niet per se slimmere mensen hebben maar een complexer netwerk tussen personen waardoor deze zich specifieker kunnen specialiseren en er een grotere hoeveelheid puzzelstukjes kennis op vernieuwendere wijze bij elkaar kunnen worden gebracht.

Om de universaliteit van deze principes en diens relatie tot de ruimtelijke concentratie van activiteiten in kaart te brengen, ontwikkelen wij maatstaven voor vier verschillende activiteiten: werk; de productie van goederen/diensten; het publiceren van wetenschappelijke artikelen; en het doen van gepatenteerde uitvindingen. De maatstaven die we hierbij gebruiken zijn: het aantal jaren onderwijs dat een werknemer heeft genoten voor de activiteiten werk en productie en het aantal teamleden voor wetenschap en uitvindingen. Daarnaast worden in aanvullende analyses voor patenten ook de NK-maatstaf van Fleming and Sorenson (2001) gebruikt en het jaar dat een subklasse is geïntroduceerd. Voor al deze maatstaven geldt dat de hoger deze is de meer de benodigde kennis complexer is en over meer personen verdeeld dient te worden die *face-to-face* contact nodig hebben om vertrouwen op te bouwen en kennis uit te wisselen.

Qua methodologie wordt de machtswet aanpak van Bettencourt et al. (2007) gevolgd. Dit houdt in dat er regressies worden gedraaid om het verband tussen de natuurlijke logaritme van de omvang van een bepaalde activiteit en de natuurlijke logaritme van de omvang van de lokale bevolking te meten voor 363 steden in de V.S., weer gedefinieerd als MSA. Als de coëfficiënt op bevolking groter dan 1 is dan houdt dit in dat de activiteit superlineair schaalt. *I.e.* een stad met twee keer zoveel inwoners heeft meer dan twee keer zo'n grote activiteit en is daarom ruimtelijk geconcentreerd in grote steden.

In een eerste stap tonen we, net als voorgaande literatuur, dat het aantal werknemers, het bnp, het aantal patenten en wetenschappelijke artikelen superlineair schaalt. Voor werk is dit slechts een factor $1,04$ maar voor wetenschappelijke artikelen is dit zelfs 1.54. In een tweede stap wordt gekeken hoe de schaalfactor varieert ten opzichte van de complexiteit binnen elk van deze activiteiten. De resultaten laten zien dat banen en industrieën waarvoor meer jaren onderwijs nodig is en patenten en wetenschappelijke artikelen waarvoor grotere teams zijn de schalingsfactor hoger is. Ditzelfde geldt voor patenten met een hogere NK-maatstaf of met technologische subklassen met een recenter jaar van introductie. Dit toont duidelijk aan dat complexere economische activiteiten zich meer concentreren in grote steden.

In een derde stap kijken we naar ontwikkelingen in de schalingsfactor over de tijd door de patentdata te gebruiken die teruggaat tot 1850. Deze analyse laat zien dat complexe activiteiten zich steeds meer in grote steden zijn gaan concentreren en dat dit proces met name versnelt gedurende industriële revoluties. Waarschijnlijk omdat dan de hoeveelheid nieuwe combineerbare kennis toeneemt door de radicale veranderingen in de economie. Ook valt op dat na de derde industriële revolutie, bijgenaamd de computerrevolutie, de minst complexe patenten juist steeds meer verspreid over de ruimte worden aangevraagd in plaats van ook te concentreren in grote steden. Dit kan komen doordat communicatietechnologieën wel een goede vervanger zijn voor het uitwisselen van minder complexe informatie die voor dit soort patenten nodig zijn, zoals ook voorspeld door Leamer and Storper (2001).

Voor beleidsmakers toont de toenemende complexiteit van innoveren en de vraag voor grote steden die daaruit voorkomt hoe ingewikkeld het is geworden voor kleinere regio's om ook complexe activiteiten aan te kunnen trekken en daarmee de groeiende kloof van ruimtelijke ongelijkheid te dichten.

Onderzoeksvraag #4: In hoeverre vertonen diverse steden ander diversificatiepatronen ten opzichte van gespecialiseerde steden tijdens crises?
In dit hoofdstuk, in samenwerking met Pierre-Alexandre Balland, David Rigby en Ron Boschma, staat het dynamische aspect van agglomeratievoordelen en het effect van lokale crises gedurende periodes van technologische verandering centraal. Het ontwikkelen van nieuwe activiteiten, het zogenaamde diversificeren, als oude specialisaties in een regio werkgelegenheid verliezen is een centraal thema in het evolutionaire perspectief op regionale weerbaarheid (Boschma, 2015).

De belangen rondom dit onderwerp zijn groot zoals bijvoorbeeld de tegengestelde groeipaden van Boston en Detroit laten zien. Waar Detroit sinds 1970 duizenden banen en inwoners verloor toen de dominante auto-industrie neerging wist Boston zich te ontwikkelen tot een innovatieve stad toen daar de lokale vervaardigingsindustrie aan werkgelegenheid verloor, zie Glaeser (2005); Hill et al. (2012). Tegelijkertijd lukt het steden als Parijs, Amsterdam en New York om gedurende eeuwen de grootste stad van het land te zijn ondanks grote crises en veranderingen in de economie. Wat is hun geheim?

In de literatuur is bekend dat regio's vaak voortbouwen op kennis en vaardigheden die in vorige activiteiten nuttig waren, zoals het eerdergenoemde ontstaan van de financiële sector in Amsterdam uit de zakelijke dienstverlening rondom havenactiviteiten uit het verleden. Verder is er anekdotisch bewijs, zie Vernon (1960) en Grabher (1993), dat in gespecialiseerde regio's nieuwe ontwikkelingen worden geblokkeerd doordat spelers uit de dominante industrie grote invloed hebben op beleid en ondernemerschap. Wat betreft crises is bekend dat diverse regio's vaak minder hard geraakt worden omdat er meer sectoren zijn die werknemers uit geraakte sectoren over kunnen nemen. Het is echter nog de vraag in hoeverre regio's minder diversificeren gedurende crises of de focus dan meer op gerelateerde activiteiten ligt en in hoeverre diverse steden hierbij verschillen.

Het feit dat dit onderwerp nog niet systematisch onderzocht is komt doordat het tot voor kort moeilijk was om een kwalitatief gegeven als gerelateerd en ongerelateerd te definiëren en dat er nog niet zoveel data hiervoor beschikbaar was. Door te bouwen op het concept *relatedness*, waarvoor in de bijlage van dit proefschrift een nieuwe formule en R code is ontwikkeld, kan meetbaar gemaakt worden hoe gerelateerd twee technologieën aan elkaar zijn. De formule vergelijkt of twee technologieklassen vaker in dezelfde gepatenteerde uitvinding gebruikt worden in vergelijking met een willekeurige verdeling van technologieklassen over patenten. De doorbraak in databeschikbaarheid komt voort uit de HISTPAT-database (Petralia et al., 2016) waar uit gescande patentdocumenten de locatie is gewonnen. Hierdoor is er data sinds 1836 beschikbaar en kan de focus gelegd worden op de drie grootste crises die de V.S. gekend heeft: de lange depressie (1873-1879), de grote depressie (1929-1934) en de eerste oliecrisis (1973-1975).

De data worden voorbereid door tijdsperioden te onderscheiden in groei- en crisisperioden voor steden in de V.S., weer gedefinieerd als MSA, die een voldoende aantal patenten hebben om als technologisch actief te worden gezien. Dit wordt gedaan door het aantal patenten per stad per jaar te tellen en dan met de *boom-bust* cyclus algoritme van Harding and Pagan (2002) onder te verdelen in afwisselende perioden van groei en crisis. Als een lokale crisis voorkomt buiten de genoemde drie grote crises dan wordt deze verwijderd om alleen lokale crises over te houden die zeer waarschijnlijk met algemene verzwakte economische omstandigheden te maken hebben. Per stad-tijdsperiode wordt gekeken welke technologieën in een technologisch portfolio aanwezig zijn en welke niet. In de volgende tijdsperiode van dezelfde stad wordt dan gekeken of de niet aanwezige technologieën nu wel in het technologisch portfolio zijn opgenomen. Als dit het geval is krijgt de variabele binnenkomst de waarde één en anders de waarde nul. De diversiteit van een stad wordt gemeten door de *Relative Diversity Index* (RDI) van Duranton and Puga (2000).

Qua methodologie wordt dan een *entry* model gebruikt om de dichotome variabele voor binnenkomst te regresseren op de gerelateerdheid ervan ten opzichte van de technologische portfolio van de stad in de vorige periode geïnteracteerd met een dummy variabele voor crisis en de RDI plus controlevariabelen. Deze controlevariabelen zijn de aanwezigheid van gerelateerde technologieën in nabije steden, de centraliteit van een stad in het uitvindersnetwerk en de totale bevolking van de stad. Om deze logistische regressies uit te voeren is een nieuwe package voor R geschreven, *fastlogitME*, die sneller en efficiënter dan vorige packages de marginale effecten kan berekenen bij grote datasets zoals in dit hoofdstuk.

De resultaten laten zien dat steden gedurende crises minder diversificeren en wanneer ze dat wel doen zijn het vooral gerelateerde technologieën die binnenkomen. Dit suggereert dat er tijdens een crisis minder fondsen beschikbaar zijn om nieuwe en ongerelateerde activiteiten te ontplooien. Dit komt overheen met de *demand-pull* hypothese in de lange golf literatuur, zie Schmookler (1966) en Freeman et al. (1982). Met betrekking tot diversiteit laten de resultaten zien dat diverse steden meer diversificeren, ook gedurende crises. Dit komt boven op het voordeel dat diverse steden al hebben namelijk dat er meer potentiële technologieën gerelateerd zijn omdat de technologische portfolio's diverser zijn. Dit bevestigt de vermoedens van eerdere studies dat er waarschijnlijk meer openheid en draagvlak voor nieuwe ideeën is in diverse regio's (Grabher, 1993; Neffke et al., 2018). Deze voordelen van diversiteit doen een licht schijnen op huidige diversificatiestrategieën zoals de Horizon 2020 programma's van de EU omdat het bevorderen van diversiteit geen deel hiervan uitmaakt terwijl het regio's wel wendbaarder kan maken in tijden van crises.

Conclusie

In dit proefschrift zijn er op basis van bijdragen aan de theorie, data, methodologie, R code en empirie vier grote onderzoeksvragen omtrent de dynamiek in agglomeratievoordelen en de factoren van economische veranderingen beantwoord. De belangrijkste inzichten zijn hierbij dat uit hoofdstuk 1 blijkt dat arbeidsmarktdeling en verticale relaties minder belangrijke coagglomeratiemotieven zijn geworden en kennisuitwisseling juist belangrijker. De eerste en laatste lijken vooral het resultaat van het kennisintensiever worden van de economie door technologische verandering en importconcurrentie. Verticale relaties lijken vooral beïnvloed te worden doordat lokale leveranciers vervangen worden door leveranciers uit lagelonenlanden. Uit hoofdstuk 2 blijkt dat complexe activiteiten steeds meer in grote steden plaatsvinden een belangrijke reden hiervoor lijkt de steeds grotere arbeidsdeling in een steeds complexere economie en de daaruit voortvloeiende behoefte voor *face-to-face* contact. Uit hoofdstuk 3 blijkt dat crises ervoor zorgen dat regio's minder nieuwe activiteiten ontplooien en wanneer dat wel gebeurt dit vooral gerelateerde activiteiten zijn. Diverse steden tonen daarbij meer capaciteit te hebben om te diversificeren.

Dit proefschrift laat hierbij zien dat er veel mogelijkheden zijn om kennis uit stedelijke economie en evolutionaire economische geografie te combineren om vooruitgang te boeken. Dit terwijl de twee vakgebieden minder vaak samenwerken dan men zou verwachten gezien de overlap in interesses. In de aanbevelingen voor vervolgonderzoek wordt ingegaan op de respectievelijke voordelen en vooroordelen bij elk vakgebied, waarbij stedelijke economie vooral voordelen biedt om empirische identificatiestrategiën te verbeteren en beleid te evalueren en evolutionaire economische geografie meer oog heeft voor hoe regionale economieën veranderen en een bredere theoretische invalshoek heeft.

In de aanbevelingen voor vervolgonderzoek wordt er ook ingegaan op het heroriënteren van het onderzoek in economische geografie en stedelijke economie om beter de groeiende ongelijkheid, de aanleiding van dit proefschrift, te begrijpen. Het vakgebied boekt namelijk veel vooruitgang bij het begrijpen van factoren van groei, zoals agglomeratievoordelen, maar het heeft een blinde vlek wat betreft de verdeling van de vruchten van die groei over verschillende groepen in de maatschappij. Door meer te differentiëren naar wie er toegang heeft en profiteert van agglomeratievoordelen kan er in de toekomst meer vooruitgang worden geboekt in het begrijpen van de stijgende ongelijkheid tussen en binnen regio's en kunnen de consequenties van beleidsaanbevelingen per bevolkingsgroep beter worden begrepen.

Al met al blijken agglomeratievoordelen alleen maar relevanter te worden in een tijd waarin steeds meer activiteit op het internet plaatsvindt en leidt dit tot een groeiende ongelijkheid tussen regio's. Dit proefschrift laat zien dat dit waarschijnlijk met name komt door het toenemende belang van fysiek contact voor het uitwisselen van ideeën dat belangrijker wordt in een alsmaar competitievere en complexere economie waar kennis en vaardigheden van een steeds groter aantal mensen samen moet worden gebracht om te innoveren en produceren. De kennis en vaardigheden die regio's daarbij hebben uit hun verleden spelen een belangrijke rol in hun mogelijkheden om nieuwe activiteiten en werkgelegenheid te creëren, waarbij diverse regio's een voordeel hebben ten opzichte van gespecialiseerde regio's.

Curriculum Vitae

Mathieu Steijn was born in Amsterdam, the Netherlands. He holds a *gymnasium* high school diploma from the Sint Ignatius Gymnasium Amsterdam, a Bachelor's degree in earth & economics from the Vrije Universiteit Amsterdam, a Master's degree in urbanism and planning from the Institut d'Urbanisme de Paris, and a Master's degree in Spatial, Transport, Regional, and Environmental Economics from the Vrije Universiteit Amsterdam.

At the end of his studies in 2014, he became a guest researcher at the department of spatial economics at the Vrije Universiteit Amsterdam and in 2015 he became a Ph.D. candidate at the department of human geography and spatial planning at Utrecht University. In 2019 he became a junior assistant professor at that department. Here he also obtained his teaching qualification (BKO). He also held visiting research fellow positions at Universidad EAFIT Medellín in 2015, the Harvard Business School in 2018, and Northeastern University in 2020.

He has presented his research at numerous conferences, such as the UEA, AAG, GEOINNO, and ERSA and at numerous research institutes, such as the NBER, Harvard Kennedy School, MIT Media Lab, and AQR Barcelona. He has published scientific articles in academic journals such as the *Journal of Urban Economics*, see Chapter 2 and *Nature Human Behaviour*, see Chapter 3. He has also published R packages, accessible through his github profile. On his personal website one can find datasets and R code that originate from his research and teaching activities.

At the time of publication of this thesis, he is affiliated to the Vrije Universiteit Amsterdam as a post-doctoral researcher on social mobility and public facilities in the Amsterdam area.

Cities are always changing. Phases of growth or decline are not evenly distributed over time nor over space. Recent growth in spatial inequality puts understanding the underlying changes in the advantages of cities, known as economies of agglomeration, and the channels of economic change that influence them high on the research agenda. This doctoral dissertation explores changes in agglomeration mechanisms, both static and dynamic, and relates them to changes in trade competition, technological progress, transportation costs, the complexity of knowledge and economic crises. Most of the focus is on the effects of the computer revolution since 1970 but parts of the analysis cover time series since the 1830s.

Doctoral dissertation by Mathieu P.A. Steijn (Amsterdam, 1989).