POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

# Agglomeration Economies and Urban Structure: A Study on European Countries

Tesi di Laurea Magistrale in
Management Engineering - Ingegneria Gestionale

Author: **Erika Galletta**

Student ID: 10649642
Advisor: Prof. Lucia Tajoli
Academic Year: 2022-23

# Abstract

This thesis explores the concept of agglomeration economies within the context of the European urban structure, focusing on the productivity, modelled as Gross Value Added per capita, across different regions and different sector classification. The study employs a data-driven approach, using NUTS3 geographical level and OECD classification of economic activities based on NACE Rev 2 for sector classification. The research methodology involves the construction of an econometric model to estimate agglomeration economies and their changes over time. The model was applied to data from 23 EU28 countries, plus Albania, Macedonia and Norway. The results reveal evidence of agglomeration economies in several regions, with significant predictors of productivity identified. The study also highlights the changes in agglomeration economies between 2005 and 2020, providing insights into the dynamics of economic productivity in the context of urban agglomeration. However, the analysis is subject to certain limitations, which are discussed in the thesis. The findings of this study contribute to the understanding of New Economic Geography and have implications for policy-making in the European Union.

**Keywords:** Agglomeration Economies, European Urban Structure, Productivity, Econometric Models, New Economic Geography

# Abstract in lingua italiana

Questa tesi esplora il concetto di economie di agglomerazione in diversi Stati europei, dal 2005 al 2020, in particolare concentrandosi sugli anni 2005, 2012 e 2020. Utilizza un dataset definito a livello geografico NUTS3 e una classificazione settoriale dell'OCSE basata sulla NACE Rev 2. Lo studio sviluppa e valida un modello econometria per analizzare la produttività, intesa come Valore Aggiunto Lordo per capita, complessivo e differenziato per settori specifici. Tale modello è implementato su dati raccolti per 23 Stati Membri (EU28), oltre a Albania, Macedonia e Norvegia. I risultati indicano la presenza di economie di agglomerazione in determinati paesi, con notevoli differenze rispetto ai settori presi in considerazione. L'analisi permette, inoltre, di osservare i cambiamenti avvenuti fra il 2005 e il 2020. Lo studio evidenzia anche i limiti dell'analisi e suggerisce aree per ulteriori ricerche e possibili implicazioni nell'elaborazione di politiche a livello Europeo.

**Parole chiave:** Economie di Agglomerazione, Valore Aggiunto Lordo per capita, Modelli Econometrici, Tessuto Urbano Europeo

# Contents

# Introduction

Agglomeration economies are an economic phenomenon that gained traction in the past couple of decades, reaching public debate and policymaking decisions. The concept refers to the benefits enjoyed by firms and industries clustering together in particular geographic areas. These benefits are often summarized as *sharing*, *matching* and *learning* [14][25]. *Sharing* refers to the opportunity of lowering cost of infrastructures by utilizing them more efficiently: a clear example are transportation and telecommunication, as the value of the investment in these kind of assets increases with the number of users. Moreover, being adjacent to customers increases productivity and efficiency. *Matching* refers to the creation of labor market pools in agglomerated urban areas: these realities consistently attract highly skilled workers, which facilitate firms' search for qualified human capital. Finally, *learning* refers to the primary benefit of agglomeration economies, as clustering facilitates the creation of knowledge and the sharing of information. This phenomenon is often referred to as "knowledge spillover", which occurs when highly skilled individuals and firms interact directly. This interaction leads to the creation of new ideas and fosters innovation. In addition to these benefits, agglomerate urban areas, being bigger and denser realities, foster competition, resulting in more productive clusters, both in terms of firms and individuals [25].

Agglomeration economies have been observed across the globe and Europe is no exception. European cities, benefiting from agglomeration effects, attract businesses and talents and foster innovation, constituting a crucial driver for economic growth and development. Measuring agglomeration effects in Italy, Spain, UK, Germany and France, Ciccone [17] finds that the effect size does not differ between countries and the estimations are similar to measurements of agglomeration in US cities. The results, referring to data from 1990 to 2001, suggest that the average productivity of the top 5 most productive Italian, Spanish and French NUTS3 is two-thirds higher than the bottom 5 NUTS3. Even stronger evidence can be found in Germany. Policy makers have recognized the importance of agglomeration effects for European economies and have supported agglomeration economies through Cohesion Policy funding for urban regeneration and infrastructure development. In particular, policies in the first decade of the 21st century have focused exclusively on cities to foster growth, discontinuing supports towards unproductive realities, in line with 2009 Work Bank recommendations [58].

On the other hand, less consistent evidence of agglomeration economies in Europe can be found regarding data from 2000 to 2008. Dijkstra et al. [21] argue that the majority of studies on agglomeration economies are based on data from US and East Asian cities, which present a very different urban structure compared to Europe. Specifically, Europe is deemed to have a more polycentric and less concentrated urban structure – only 7% of EU citizens live in metropolises, while 70% of Europeans live in urban areas of more than 5,000 inhabitants, well-connected to other cities of similar size. Dijkstra et al. [21] find that different agglomeration trends indeed exist within Europe: Central-Eastern Europe and Southern Europe exhibit population decline in non-Metropolitan regions, caused by mobility towards urban centers. These regions therefore seem to be following the agglomeration trend. On the other hand, Western and Northern Europe report evidence of a reverse urban-rural shift, whereby rural and intermediate regions have grown and played a much more important role than cities. While the exact reason for this shift is unclear, a possible explanation could be larger-than-expected negative externalities in cities, such as congestion costs.

Indeed, despite the mixed evidence, agglomeration economies and their effects have become the latest topic of interest in policy making, with many reports and articles being published at the beginning of 2010s discussing the best approach to the phenomenon [12][31]. Based on this, Cohesion Policy 2014-2020 were formulated with a place-based logic, promoting different strategies according to specific territories. In the past decade, events such as Brexit and the rise of populism [24][67] in many European countries have increased interest in the mechanisms behind agglomeration and its effects, especially in terms of negative externalities. It seems, therefore, that the agglomeration trend is back on its track across Europe: regional convergence was still far from achieved in 2018 and regional inequalities within countries remain constant [13]. In particular, Metropolitan regions tend to be very successful, while non-Metropolitan regions' prosperity depends on accessibility to large cities, with isolated areas being disproportionately poor [42]. Consistent with this evidence, the latest talking point in public debate centers around congestion costs, in particular the housing crisis that is plaguing major cities across Europe.

This thesis investigates the issue of agglomeration economies across Europe. The purpose is twofold: first, this thesis aims at verifying the claims from Dijkstra et al. [21] and Roemish [59] concerning different agglomeration trends co-existing within Europe. Subsequently, the analysis estimates agglomeration economies in Europe in manufacturing and in Advanced Tradable Services, to investigate which sector is primarily responsible for the phenomenon in European countries. The paper will be structured as follows: first a literature review regarding the concept of agglomeration economies will be presented. Afterwards, the theory will be applied to the European context and papers presenting its urban structure will be discussed. Thirdly, agglomeration economies in Europe will be measured according to the selected model using data from 2005 to 2020. Lastly, results and limitations of the analysis will be presented and implications will be discussed and evaluated.

# 1 | Agglomeration economies: theory and their existence in Europe

As stated in the introduction, agglomeration economies is a term that describes benefits enjoyed by firms and industries clustering together in a particular geographic areas and the mechanism behind the clustering. According to the literature, a distinction is drawn between localization economies and urbanization economies: while the first result from the concentration of firms from the same industry in one place, the latter arise from dense and diverse economic activity in an area. In this thesis, the focus will be primarily on urbanization economies and I will refer to the term "agglomeration economies" as a synonym. This chapter will be sub-divided in two section: the first will focus on the economic theory behind the concept of agglomeration economies, while the second will discuss its implication in the context of European Integration.

## 1.1. New Economic Geography

Prior to discussing the economic models supporting the concept of agglomeration economies, I begin by providing evidence that agglomeration economies do exist and urban size and density distributions are not simply due to a casual pattern. For start, Spatial Impossibility Theorem [66] predicts that, regardless of agglomeration economies, distribution of economic activity should not be anticipated to be uniform, since space itself is not uniform, as topographical maps show. Just like landscape is heterogenous, regional endowment of natural resources is uneven, with some areas having multiple water streams and others having harsh climate conditions. These locational fundamentals are considered exogenous and they range from amenities to accessibilities [25]. The fundamentals can either be positive, encouraging population and economic activity, or negative: an example of computation of locational fundamentals is the MSA amenity score for US cities [25], which takes into consideration January temperature, July temperature, humidity, hours of sunlight, water surface and topography. Therefore, the existence of these natural advantages could explain the distribution pattern of economic activity observed today. In this regard, Ellison and Glaeser [29] developed an index for spatial concentration and found that regions' endowments of resources only account for 20% of industrial localization. Therefore, a sizable portion of spatial concentration observed today must be explained by other factors, possibly the same indicating agglomeration economies.

Furthermore, economic activity's concentration is predicted by two other economic theories, aside from agglomeration economies. It should be distinguished between relative concentration and absolute concentration. The first refers to industries whose spread of production differ from average spread between countries - high relative concentration implies high specialization at country level. Absolute concentration, instead, refers to unevenly distributed spread of activity between countries: the two measures would coincide for countries of identical size [40]. Traditional Trade Theory, specifically the Heckscher-Ohlin model, assumes resource endowments as determinants of specialization. The core notion of this model is comparative advantage, meaning the ability of a country to produce a good or service for a lower opportunity cost than another country. Hence, traditional trade theory predicts countries to specialize in the production of goods they have a comparative advantage in, thus resulting in relative concentration of activity between countries. Conversely, New Trade Theory put forward the notion of market size effect - a demand bias in favor of a good creates a large home market for this good. The model expects industries to be more concentrated the bigger their home market is. This relates to the concept of scale economies, namely the cost advantages gained by a company with an increase in production: the degree of scale economies characterizing an industry, coupled with the market size effect, conveys information regarding absolute concentration, as the higher the scale economies enjoyed by an industry, the more concentrate it will be, as firms will locate close to the larger markets. In concentrated industries, the higher competition for factors of production will raise the costs for input. New Trade Theory expects a trade-off between market size effect advantages and factor costs: if trade costs are low, it is more cost-effective for firms to locate farther from agglomeration centers, thus absolute concentration will be lower. In contrast, Traditional Trade Theory predicts that lower trade costs will induce countries to further specialize, therefore increasing relative concentration.

Finally, the discipline of New Economic Geography intends to explain the formation of various kind of economic agglomeration across space [33]. The models developed in this context seek to understand the mechanism behind absolute spatial concentration and explain when it changes and what it implies for the overall economy. The key-assumptions behind these models, which differ from neoclassical economics, concern increasing returns, and therefore the existence of monopolistic competition, and the locational movement of production factors and consumers, which in turn make transportation costs of the utmost relevance.

All three theories are at least partially supported by evidence: for example, market size effect is shown to be one of the most important determinant of industrial localization, supporting New Trade Theory [40]. Empirical evidence of agglomeration economies can be found in the distribution of income [57]. Higher wages and rents are observed in cities, along with greater consumption amenities. If firms and workers accept the higher living costs and longer commuting times associated with the urban environment, the concentration of firms and workers in cities could be a sign of an urban advantage. To confirm this, Glaeser and Maré [37] implemented a model controlling for variables that could otherwise explain the income difference observed, such as workers' observable skills, parental background and fixed effects. The result provides evidence of a significant wage premium associated with dense cities, meaning that workers in this environment are indeed more productive, even when controlling for their education and skills. Indeed, Rosenthal and Strange [61] find that doubling city size increases productivity from 3 to 8% for large cities. The results still hold when instrumenting current size or density with historical advantage of location to control for reverse causality – higher output per worker could be the cause of higher density. Instrumenting simply reduces the estimated effects of agglomeration, but these do not disappear [18].

A common counter-hypothesis to agglomeration economies is firm selection: a denser and larger city offers a more competitive environment which drives less productive firms out of the market. Melitz and Ottaviano [49] implement a model on large French cities - comparing log productivity distribution to the expected distribution in case of agglomeration economies and in case of selection effect. In the first case, the distribution should shift rightwards and show an increase in dilation, as all firms become more productive overall and the most efficient firms benefit more from agglomeration effects. In the second a left truncation is expected, as least productive firms exit the market. The authors find that log productivity distribution of firms in large French cities correspond to that of small French cities rightward shifted and dilated. Similar results are found by Duranton [25]: agglomeration effects disproportionately accumulate income to the top earnings/productivity of workers and firms, dilating the income distribution. This consideration is supported by the correlation between city size and density with inequality in economic outcomes. While the size elasticity of the Gini coefficient falls from 0.011 to 0.008 when controlling for the share of college graduates, the correlation persists, consistent with the dilation effect of agglomeration economies on income distribution.

Having presented consistent evidence of spatial agglomeration and its effects, regardless of natural endowments, selection effect or simple casualty, I will now discuss three models of New Economic Geography, to illustrate the mechanism of centripetal and centrifugal forces that underlies agglomeration economies. The first model discussed is the "Core-Periphery" model [44]. In the model, there are two production sectors, two regions and two types of labor – agriculture uses farmers to produce goods with constant return to scale, manufacture employs workers and produces with scale economies. All firms are equally located in between the regions, but farmers are immobile while workers can freely move. Finally, agriculture products are traded without costs, while manufactured goods incur in transport costs in iceberg form, meaning that goods are assumed to be shipped freely and that part of shipment melts in transit, to avoid taking transportation as an industry into account. In this model centrifugal and centripetal forces can be recognized as follow: farmers' immobility are a source of centrifugal forces, as manufacturing firms would like to locate close to final customers to reduce transport costs and farmers are scattered. Centripetal forces, instead, are twofold: the incentive for workers to be close to firms to enjoy higher variety of goods and higher real income, defined as forward linkages, and the backward linkage, the incentive for firms to concentrate close to a larger market to benefit from scale economies and lower transportation costs, called home market effect. The core-periphery pattern occurs when centripetal forces are stronger than centrifugal forces: this is the case when transportation costs are low enough, goods are sufficiently differentiated or expenditure on manufacturers is large enough.

The second model reviewed is the "Urban system" model developed by Fujita and Krugman in the late 1990s [35][32][34]. Based off on the "Core-periphery" model, the "Urban system" tries to model the dynamic process of urbanization following a more realistic approach. In this case, there is a single region, described as a real line: workers are all identical and free to move, land is the immobile factor and goods can be moved, incurring in transportation costs. The model begins at *isolated state*, a single city composed by a cluster of manufacturers and surrounded by agricultural land. As population grow, the city expands until it reaches a size big enough that its limits are too far from the center and it becomes convenient for firms to move away and build a new city. Therefore, by simply growing the population, the model can give rise to a whole system of cities. The next question is which factors make a location attractive enough for a new city to be established. In the literature, the location attractiveness is often indexed as *market potential*, which co-evolves with economic activity: while market potential first determines where economic activity will locate, the location shift of that same economic activity will change the market potential map. In this framework, the initial location pick is based on natural favorable aspects, meaning geography of rivers, harbors etc. Once a new city is established, it grows through a self-reinforcing process, possibly until the initial advantage of the location becomes irrelevant compared to the agglomeration advantages themselves. As in the previous model, multiple stable equilibria exist, along with regularities: once enough cities are created, size and distance between those tend to settle down at a roughly constant level depending on the relative strength of centripetal and centrifugal forces.

The last model examined is the "industrial concentration and trade" model proposed by Krugman and Venables [45]. The earliest "Core periphery" model considered factor mobility as the element creating agglomeration, but real life examples, European cities above all, shows the existence of multiple ultra-specialized cities, where other forces, much harder to model, are at play. The original model is therefore adapted by shifting from agglomeration of resources to clustering of particular industries, which allows international specialization, while labor stay immobile. To model the complexity of trade, the authors distinguish between upstream and downstream firms, both incurring in increasing returns and transport costs. The simplest version of the model assumes that upstream and downstream firms consume the same input and produce the same output. The results are similar to the "Core-periphery" model: backward and forward linkages exist and tend to concentrate upstream and downstream firms in the same spot, similarly to the clustering of workers and farmers. The advanced version, differentiating between the two kinds of firms, results in a more complex input-output matrix in which some combinations result in industrial clusters. This variant allows the authors to explain why certain regions (and which ones) will industrialize. Regardless of this assumption, the general model predicts country level specialization in very different production, because of increasing returns at industry level. This outcome is further confirmed by extensive empirical evidence [48], such as Italian specialization in tile production or British specialization in financial services – indeed, the findings indicate that specialization go beyond country level, as clusters occur in specific location within countries as well. The model is able to show that, while global trade grows and transport costs fall, the world divide spontaneously and arbitrarily in high wage, industrialized region and low wage, primary producing regions, and later the second rise at the expenses of the first.

In conclusion, the centripetal forces are identified by these models in linkages (forward and backward), thick markets and pure external economies, including knowledge spillovers. Likewise, centrifugal forces are associated with immobile factors, land rent, commuting costs, congestion and other pure dis-economies [33]. Having established the existence of agglomeration economies and discussed the most prominent models, I now want to analyze the possible benefits that firms can reap by clustering together. In the literature, these are often summarized as *sharing*, *matching* and *learning* [27].

*Sharing* concerns a wide range of actors. The most immediate sharing advantage refers to the benefits of sharing facilities with large fixed costs, deemed as "sunk", as the larger the user base, the lower the cost per capita. This perk is limited by the fact that too many users cause congestion and crowding, further limiting the growth of the user base itself [57]. The second sharing mechanism concerns suppliers: increasing returns exist at city level production, despite perfectly competitive final market, due to input sharing across wider supplier's base [8]. Evidence of the mechanism is not straightforward: simply measuring input purchase relative to value added as a proxy for input sharing relevance gives weak results that are statistically insignificant [28]. Stronger support to the theory is provided by Overman and Puga [53]: they find that importance of input sharing depends on spatial concentration of supplier industries. This is described as *vertical linkages*: if supplier industries are dispersed, firms do not need to be concentrated as in any location suppliers will be closely available. An analysis of similarities across industries can help predict agglomeration [30]: sectors buying similar intermediary goods are most likely to co-agglomerate and a similar, less strong, effect is seen in industries employing similar workers. Two more sharing gains for agglomerated firms can come from workers: first, more workers concentrated in a given industry increases output more than proportionately, as employees are able to specialize on a narrower set of tasks. Furthermore, one of the biggest source of agglomeration gains come from sharing labor pools: idiosyncratic shock's effects, seen as adjustments in production and employment levels, are greater the more isolated firms are [53]. Therefore, highly volatile industries tend to be spatially concentrated to lower shocks' effects, facilitate labor transfer and improve adaptability [57].

Regarding *matching*, this term also refers to a range of possible actors: the matches implied can be between employers and employees, buyers and suppliers or between business partners. Overall, agglomeration economies improve matches in the market in terms of both quality and quantity. First, if workers incur in re-training costs proportionate to their skill mismatch to firms' requirement, a denser location reduce the average costs of mismatching and the likelihood of no matches working out [57]. Therefore, agglomeration increases chances for matches. Simultaneously, high probability of matches allow firms and workers to be more choosy, increasing quality of matches but reducing probability of matches. This second effect is harder to prove empirically due to issues in quality measurement and definition, although the overall effect of agglomeration on matches has been proved empirically [43][36].

Finally, *learning* refers to the benefits enjoyed in transmission and creation of knowledge. Regarding transmission, it is theorized that in dense environments interactions between experience workers and newcomers help the latter acquire skills and the former share the rents of this learning process [38]. Although promising, this theory has not yet been developed extensively in practice. There is, nevertheless, empirical support concerning the advantages in knowledge transmission: Audretsch and Feldman [11] find that innovative activity, measured as new product launches, tends to cluster more in industries where new economic knowledge plays a relevant role. Moreover, invention and adoption of new technology is more likely in places with more prior adopters and inventors [10][41]. Concerning creation of knowledge, "nursery cities" paradigm theorizes that a diverse urban environment facilitates research and innovation by offering many alternatives to try and implement without need for relocation. Indeed, young firms locate to diverse environment to benefit from these dynamic advantages, while more mature firms locate to specialized towns [26].

As seen, there are roughly three main sources of agglomeration economies: labor market pooling, input sharing, and knowledge spillover, which mirror the three key transport costs proposed by Glaeser [38]. Combined, these three sources account for 30% of variation in spatial concentration. The costs of moving input, or goods, has significantly decreased thanks to transportation improvement, which in turn has also impacted the cost of moving ideas and people. Simultaneously, the Internet and electronic communication have further reduced the costs of sharing ideas. Therefore, the spatial reach of agglomeration economies is increasing. Labor market pooling is one of the strongest agglomeration effects, as Rosenthal and Strange [60] find it to be a relevant variable in explaining agglomeration variation across industries at all geographic levels (state, county, zip code). Although robust, these effects decay rapidly with distance and an even more rapid decay is observed from knowledge spillover, which have a positive effect on zip code level only [62][60]. Information technology, although revolutionary, is fairly recent so its effect on urban structure is not yet clear: up until now, there is evidence of agglomeration economies operating at a very local level, such as neighborhoods, individual buildings and spatial arrangement of workers within buildings [62]. Highly local interactions might also be complementary to more distant interactions now allowed by technology, so that proximate and distant interactions might not be as opposite as they appear. Despite technological progress, I might not see dispersion effects in the future because of the urban amenities offered by cities, which are disproportionately enjoyed and sought after by highly productive workers. Therefore, the gentrification effect could be permanent regardless of how easy it has become to move goods and ideas – highly educated workers will still prefer to live in city centers and highly productive firms will still locate in tall commercial buildings with many amenities. However, gentrification is not the only way forward. As Fujita and Krugman [33] remarks, the crucial implication of the "core-periphery" model is that agglomeration economies do not need to occur: small changes in a critical parameter can be enough to swing the whole economy towards extreme differentiation. These small shifts are called *catastrophic bifurcations* by the authors, indicating moments in history when the economic structure is suddenly tipped off. The authors also emphasize that there are conditions where agglomerations can arise and others where agglomerations must arise. There is therefore the possibility where both clustering (monocentric) and equal divisions (polycentric) could be sustainable equilibria. This would mean, for example, that Europe potentially could cluster "US-style" in one big city such as New York or not. Indeed, Krugman came up with the first model of New Economic Geography while evaluating the consequences of European Integration in the late 1980s: international economy models usually assume immobile factors and costless trade, which hardly reflect the reality of European Integration.

## 1.2. Distribution of economic activities in the EU

So far I have discussed the theory behind the concept of agglomeration economies and the most relevant differences with other established economic theories. I now want to examine the literature that applies these concepts to the European Union. Indeed, investigating agglomeration trends in the EU could provide relevant insights on what factors swing the urban structures towards a monocentric or a polycentric structure and what each equilibrium implies. The European Union, being a conglomeration of national states with similar levels of development but different social, institutional and technological characteristics, appears to leave a much greater role for smaller and medium sized cities: this polycentric system implies that labor mobility and supply chains operated from multiple centers instead of a single large monocentric structure such as the one found in America or in Asia. Indeed, European urban structure is much less concentrated: only 7% of EU citizens live in cities of 5+ millions inhabitants – in fact there are only five cities of such size in Europe! 70% of European population lives in urban areas of over 5000 inhabitants and the Union has experience much lower rates of urbanization than the rest of the world [21]. Moreover, it seems that the advantage shared by highly productive European cities is not related to their size but to their connectivity, as the vast majority of European cities are medium small [12]. In this chapter, the literature on the topic will be presented following a chronological order. There is great variation in terms of geographical unit used in the analysis: the majority of papers use NUTS2 and NUTS3 units, while others refer to Functional Urban Areas and the differentiation between Metropolitan Region, non-Metropolitan Region and Metropolitan Capital Region. In a few cases, Predominantly Urban, Intermediate and Predominantly Rural units are used. The definition and distinction between these measurement units will be discussed in depth in the next chapter, including the implication of using one unit over another for results' interpretation.

Among the first to analyze agglomeration effects in Europe, Ciccone [17] uses NUTS3 data from Germany, France, Italy, Spain and the UK, dating back to 1986-1988. He detects substantial and similar agglomeration economies in all five countries: average productivity of top five most productive NUTS3 in Italy, France and Spain appeared to be two thirds more productive than the bottom five, just above the UK level. Even more pronounced, German top five counties were 140% more productive than the bottom five. These results are aligned with US evidence, showing indeed that higher density in pre-integration European countries leads to higher productivity. Given the complex and ongoing integration process, an analysis of income level trend before and after major economic integration turn points could help us determine the role played by European Integration in urban economy. Breinlich [15] plotted the GDP per capita in 1999 and distance from Luxemburg, used as a proxy for market access: the author finds a strong core-periphery gradient, as the GDP per capita in the 5% richest NUTS2 regions was three plus times higher than the 5% poorest regions, which were located on the fringe of the EU. The same analysis was also applied on panel data from 1975 -1978 and 1992-1997 to capture the immediate effect of the first round of economic integration. While, in the first time frame, crossing a national border reduced trade by 72%, after EU integration the effect declined to 65%, showing that European Integration decreased the importance of distance. In this perspective, the main beneficiaries of European Integration are peripheries, as this decline in direct trade costs allows for catch up effect. Furthermore, higher market access – meaning being located closer to the capital city - was found to influence income level by incentivizing human and physical capital accumulation.

Somewhat opposing these findings, a cross-sectional panel between 1985 and 1992 shows that the importance of intra-industry linkages in predicting economic concentration was positively and significantly affected by EU integration: this result is aligned with New Economic Geography theory, therefore associating EU integration with higher agglomeration and divergence in income levels [40]. In this case, intra-industry linkages were modeled as the sector percentage usage of its own products as intermediates. A similar result is obtained by Mathä [47], which focused on geographical concentration of multinationals as an indicator of presence and importance of agglomeration forces. The author assumes that MNEs, being *footlose*, are able to react faster to the effects of

European Integration. Agglomeration forces are modeled by economies of scale, which increase firms' tendency to concentrate production when trade costs are lowered, and by vertical integration, as MNEs will strive to exploit each country's competitive advantage by locating capital intensive activities in capital intensive countries and vice versa. Dispersion forces are represented by high transaction costs, which induce a third country firm to produce in more countries instead of one single concentrated location. Finally, two hypotheses are proposed: EU integration can either increase market size or market accessibility: in the first case, dispersion of economic activity would be promoted, as firms would locate next to larger markets, the latter would foster concentration. By observing MNEs' location choice in Sweden following entrance into the European Union, Mathä [47] found that EU integration enhanced relevance of agglomeration forces in explaining geographical concentration of MNEs production in Sweden. The author finds a significant reduction in intra EU transaction costs and results consistent with market enlargement hypothesis and increase vertical integration, hence supporting the neoclassical theory that higher integration leads to more specialization of countries.

Further evidence that EU integration acts as a shock to city size distribution can be found analyzing space distribution in EU8, indicating the ten countries – seven of which from the former Easter Bloc - that joined the European Union in 2004. The same structural break observed in EU15 countries between 1996 and 1998 can be found in EU8 around 2004, meaning that city size distribution becomes more uneven, showing positive effect of EU integration on agglomeration forces [65]. The analysis of the geographical patterns of growth in EU28 from 2000 to 2011 turns out to be coherent with these results. There is evidence in support of Neoclassical theory at country level: poorer regions - such as Poland, Romania and Hungary – exhibit higher growth rates than many rich regions, showing an overall convergence trend [55]. Albeit this evidence, the author finds support for cumulative causation and NEG: the five regions with the highest GDP levels in CEECs also have the highest growth rates. Furthermore, these happens to be the capital regions, where economies of scale are able to attract even more activities and growth. The work carried out by Postoiu [55] reveals that out performing regions are usually isolated, surrounded by declining regions, with the exception of very few cases in Germany, North East Europe and Northern Spain. Instead, in most cases, declining regions share the trend with the neighboring clusters, as it's the case in Italy, United Kingdom, Greece and France. Völlmecke et al. [68] interpret the trend observed in CEECs as an indicator of a poverty trap: the authors infer that the rapid growth of capital regions causes an increase in inequalities, as regions accumulate at both ends of the income distribution [35].

However, similar analysis implemented post-EU8 integration to EU15 countries find a different pattern in the countries that joined the EU in 1995. Dijkstra et al. [21] indeed found that Central and Eastern European countries (CEEC) follow a growth rate aligned with NEG expectation concerning population growth and productivity, concentrated in Metropolitan Regions. On the contrary, Western and Northern Europe exhibits higher growth in Non Metropolitan Regions: ever since 2001, predominantly rural and intermediate regions in the area have played a much more relevant role than predominantly urban cities, showing support to Neoclassical economics prospected convergence. One of the proposed explanation for this shift in trend observed in EU15 refers to improvements in public service provision and infrastructure quality across regions: these changes might have weakened the advantages held by large cities, fostering dispersion. Otherwise, it could be a sign of declining marginal returns to productivity of labor in cities or a sign of increased congestion costs and negative externalities, meaning that expansion might become less efficient. Overall, Dijkstra et al. [21] clearly shows a change in trends around the Millennium, after which large cities in most of Europe no longer play a primary role, replaced by middle cohort regions. Indeed, the two contrasting tendencies are attested by Monastiriotis [50]: the author finds internal interregional divergence in Central and Eastern Europe, experiencing major growth of the capital city compared with the rest of the country, and internal interregional convergence or stability in Western and Northern European economies. A possible explanation for the observed decline in agglomeration economies generally in Western European countries and increase in CEECs is proposed by Roemish [59]: the level of agglomeration economies and its change is interpreted to be to some extent correlated to GDP per capita and GDP growth. Indeed, the weak negative correlation between agglomeration economies and GDP per capita and the strong correlation between high GDP growth and increasing agglomeration economies create a U-shaped pattern. Agglomeration economies may increase more in early stages of development, which is also correlated to an increase in inequality as per the Kuznet's curve [9], and then decrease at more advanced stages.

The observed process appears to be reversed after the economic crisis in 2008, which caused severe contraction in global multinational investments and intra EU flows. The long term effect of the crisis greatly affected labor market flows: while flows from CEECs to Western Europe declined, streams from Southern Europe to Northern Europe and out of the EU increase. No convergence is observed since the 2008 crisis and greater interregional divergence in productivity and employment is seen across Europe [20]. Dijkstra et al. [22] finds conflicting results, with no clear pattern of divergence or convergence: generally proximity to cities appears to have helped rural regions, while urban regions have been underperforming since 2004. The authors conclude that cities have the power to determine the fate of an economy: in robust economies, cities drive national growth, while during crisis they can drag down the whole economy. Therefore, agglomeration economies may magnify both negative and positive effects: this could be due to the fact that large cities have stronger connection to international markets, meaning that a collapse in FDI and trade could have a greater effect on Metropolitan regions, while smaller cities could prove more resilient. Nevertheless, the observed centrality of cities during the 2008 global crisis could be circumstantial: much of the initial crisis was linked to the real estate sector, which is heavily dominated by urban regions.

More recent analysis confirms that the trend observed in the first decade of the millennium, if maintained over time, would lead to further divergence in production, employment and population [13]. In the prospected scenario, future growth rate are modeled on the basis of past growth rates for Gross Value Added and Employment per sector and it is assumed that by 2150 key demographic parameters such as fertility, mortality and migration rates will be identical across Europe. Albeit conservative, as no significant structural change to the regional economies is considered, the model provides evidence of an overall increase in agglomeration economies across Europe. Aligned with this projection, Pina and Sicari [54] find that in 2018 the average GDP per capita in the richest European regions was three times higher than in the poorest regions. The authors argue that the decline in regional disparities in GDP evident up until the global crisis was actually due to CEECs catching up with European average. Older Member States were not improving and the crisis simply strengthened pre-existing divergences: Metropolitan and Capital Regions tend to be very successful, non-Metropolitan Regions with good accessibility to large cities are able to maintain their rank, while remote non-Metropolitan Regions appear to be disproportionately poor. Similarly, Iammarino et al. [42] identifies four income categories in its regional inequality analysis, using panel data from 2000 to 20014: the most economically successful are Very High-income regions, namely large Metropolitan and Capital city Regions, characterized by highly urbanized realities attracting population. High-income regions follows suit, although less dynamic demographically and less innovative. In the Medium-income category, the authors identify two types of regions. The first is characterized by declining employment, low population growth and mostly manufacturing jobs, as it's the case in declining industrial peripheries in France and Northern Italy. The second features, instead, growing population, demand for non-tradeable local services but limited skill development. Lastly, the Low-income category is overall identified by low employment rates. Again, the authors differentiate between Low-Income and Low-Growth regions: the first are mainly located in Central and Eastern Europe, stuck into a vicious cycle of high skilled out-migration, resulting in low levels of innovation and limited participation in cross-EU value chain. The Low-Growth regions, instead, are located in Southern Europe, endowed with better infrastructure but suffering a long standing incapacity to produce and assimilate innovation, both due to skill shortages and weak institutions.

Iammarino et al. [42] identifies one of the possible driving forces behind regional inequalities in 1970's technological progress, which favors large Metropolitan areas and draw large number of skilled workers, and the subsequent reduction in employment caused by automation, which enhanced high skilled roles at the expenses of low-skilled workers. Furthermore, the interaction between divergent and convergent forces described by agglomeration economies has been concentrating wealth in High- and Very High-income regions. Part of the missing convergence can be explained by the decline in low-skilled migration, partially due to changing nature of skills: in the past decade, a rise in experience based skills occurred, usually identified with social and collaborative skills as opposed to formal skills – in this perspective, location disadvantages poorer regions even when they are successfully educating workers. Furthermore, returns to education increasingly diverge, with STEM professions being rewarded much more than the rest, meaning that low-skilled workers face economical barriers to moving into prosperous regions. Most labor flow is concentrated towards Metropolitan regions, while inter-regional mobility remains limited, slowed down by shrinking wage premium in urban setting and increasing urban living costs [54]. This causes urban settings to be flooded by human capital in-flow, while rural regions are stuck with low-skilled human capital. Indeed, labour mobility is imperfect: barriers to mobility are usually assumed to be related to the inelastic housing supply in cities, but workers often decide to remain due to emotional attachment, age and insufficient skills [58]. Finally, physical connectivity does not translate into automatic spillover, instead it often results into annexation to the hinterland: this spillover deficiency is also partially due to the difference between knowledge creation, which breeds inequalities, and knowledge spillover, achieved through strong organizational channels such as multinational corporations and firms [42]. These players can be a key diffusion force when they locate high- and low-tier activities across regions, supporting spillover and assimilation of new technologies in lower-income areas. These diffusion forces often skip EU less developed regions in favor of developing countries, as European lagging behind regions are incapable of exploit the advantage of backwardness due to labor cost pressure. Pina and Sicari [54] also point out that non-Metropolitan Regions cannot sustain the competition against emerging economies. The dependency of these regions on less productive and less innovative sectors is at the root of the observed inequality, threatened greatly by automation and digitalization. In this regard, de-industrialization appears to go hand in hand with an increase in income inequality [67] and spatial concentration in Europe is expected to intensify in the next decade.

Before moving onto the analysis, I present a brief discussion on how agglomeration economies affect different sectors. Regardless of whether European countries are experiencing agglomeration or dispersion, the trend may not be consistent across all sectors, just as it seems to be inconsistent across countries. Oqubay and Lin [52] investigate convergence and divergence changes between 2000-2002 and 2014-2016, distinguishing between three macro regions - Southern Europe, Northern Europe and Central Eastern Europe. The authors refer, in the context of their analysis, to de-industrialization when highly industrialized areas in a region are the ones suffering the trend of de-industrialization the most, while industrialization is seen as a sign of agglomeration, as most industrialized areas in a region are the ones gaining further share in manufacturing. The latter is the case for low income regions in CEECs, while the strongest de-industrialization activity appears to be in low and middle income regions of Southern Europe, meaning Italy, Greece, Spain and Portugal. At first glance, Northern Europe seems to be subjected to neither trend, but further examination reveals that some regions are undergoing further agglomeration, while others are de-industrializing. The first are those areas belonging to the "central European manufacturing core", namely Germany, Switzerland and Austria. The manufacturing "belt" extends to other central European economies such as Poland, Czech Republic, Slovenia and Slovakia. This manufacturing hub exports more than 50% above their shares in European GDP would predict and it spills over to Western Romania, Northern Italy, Southern Sweden and Finland. Outside of the regions composing this strong manufacturing core, European middle and high income regions' involvement in manufacturing follows a U-shaped curve, as the higher the income level, the higher the importance of advanced tradable services in the region's economy. This trend is also observed in capital regions in CEECs, which are undergoing de-industrialization as they move towards service-oriented economies. Aligned with these observations, De-Miguel-Molina et al. [19] demonstrate that European regions' wealth is related to the presence of agglomerations of highly technological manufacturing industries, in accord to OECD classification, and of creative industries as well. The authors define "creative industries" as sectors where intellectual labor is predominant and output is related to intellectual property - often referred to as knowledge-intensive sectors (KIS). These industries are usually located in major urban areas, often concentrated in wealthiest regions [56]. Indeed, regions classified as more creative display higher share of high tech industries than other regions, supporting the claim that regional industrial structure has a relevant influence on regions' wealth. The observed trend is endorsed by Marrocu et al. [46], which find that the type of agglomerations experienced by a region varies according to its development stage, as in the case between EU15 and EU8, and the industry specialization pattern - low-tech manufacturing (LTM) vs knowledge-intensive sectors (KIS). Part of the variation is also explained by the region's territorial features in terms of urbanization. The dualistic landscape emerging from their analysis, as shown in Figure 1.1, supports the "nursery cities" theory from Duranton and Puga [26], presented in the previous Section 1.1. Moreover, Marrocu et al. [46] examine the different roles played by agglomeration economies, namely industrial specialization and diversity, on local industry productivity growth. They find that diversity externalities have a positive effect on productivity growth in EU15 countries, enhancing KIS in very densely populated urban areas, they negatively affect EU8 countries, regardless of sector and urban structure. Specialization externalities, instead, appear to positively influence all sectors in "New Europe", especially regions with high shares in low tech manufacturing activities. On the contrary, specialization in LTM sectors has negative effects on growth in "Old Europe", while high value added sectors are less affected, as they are able to exploit specialization to increase productivity.

TABLE 1: Sectoral Employment Shares (% Over Total Employment)

| | Low-Tech Manufacturing | | Knowledge-Intensive Services | |
|---|---|---|---|---|
| | 1999 | 2007 | 1999 | 2007 |
| Old Europe: EU15, Norway, Switzerland | 11.7 | 9.8 | 19.5 | 21.8 |
| New Europe: 12 new accession countries | 17.7 | 16.5 | 13.3 | 15.8 |
| Whole Europe | 12.9 | 11.1 | 18.3 | 20.6 |

Figure 1.1: Sectoral employment shares in EU15 and EU8, Marrocu et al. [46]

Broadly, Europe seems to be enjoying lower rates of growth compared to the US, which could be limiting the potential benefits of agglomeration economies. This could be the result of weaknesses in innovation, limited R&D spending or cooperation failures in innovation development. Furthermore, evidence from the US shows that productivity spillover could reach as far as 300 km [51], however, European under-performing regions are often adjacent to urban agglomeration [54]. Therefore, it may be that European urban structure fails to generate the substantial agglomeration economies needed to produce productivity growth. The observed inequalities may be worsened by the Covid-19 pandemic: Southern Europe was hit the most, given its dependence on tourism and small firms, while Central and Eastern Europe has suffered a much smaller loss. Considering the presence of weaker institutions and a less diversified economy, it can be expected that poorer regions will struggle more in the aftermath of the pandemic.

# 2 | Description of the dataset and exploratory analysis

The evidence presented so far indicates that agglomeration economies in Europe do occur, although the trend is more unstable and possibly more fragmented than theory predicts. A clear swing in trend is detected in EU15 around the Millennium, indicating convergence was on its way. On the other hand, EU8 seemed to be characterized by a strong core-periphery urban structure. In the aftermath of the 2008 financial crisis, the majority of European Member States appeared to be again dominated by agglomeration economies, heading towards an increase in regional inequalities. Findings from data from the 1990s to 2000s can be considered established given the abundance of analysis and papers in support. Results on the past decade are, instead, not as consolidated: Iammarino et al. [42] implement their analysis using NUTS2 as geographical units and their results only pertain data until 2015. Similarly, Roemish [59] account for data up to 2012, using NUTS3 units. Thus both analysis do not reflect the colossal changes occurred in communication technologies in the last ten years nor they include completely the reprise from the 2008 crisis. It could also be argued that negative externalities may have only recently reached a size relevant enough to potentially overcome agglomeration forces. Given that occurrences such as urban cost of living and congestion increase as cities grow, European urban structure may have been dominated by centripetal forces for long enough to now start showing signs of cities' un-sustainability. Moreover, NUTS2 units are coarse-grained, meaning that key details in economic activity distribution may have been missed. Therefore, I can contribute to the literature by developing a model to estimate agglomeration effects using data from 2005 to 2020 and improving granularity of analysis. Furthermore, after implementing the model to assess agglomeration economies' effect on a general productivity indicator, I will repeat the analysis focusing on different sectors, to verify whether the overall trend affects all sectors equally. This chapter will discuss the choices made in terms of granularity of analyses, variable selection and sector classification. Then, once the dataset is established, descriptive statistics will be presented.

# 2.1. Dataset definition

## 2.1.1. Geographical level

I shall discuss the granularity of the analysis: the majority of papers discussed up to now uses NUTS3 as geographical units. The NUTS zoning system is widely used within the European union. It consists in a hierarchical system of territorial units articulated in NUTS0 (countries), NUTS2 (regions) and NUTS3 (provinces or districts). The system is comparable to local administrative units (LAU) such as municipalities, which are the finest level of granularity [13]. All these units are defined nationally by administrative or legal boundaries. Due to inconsistencies in the definition of municipalities, districts or even cities across countries, the size of local units can be very different and hardly comparable (Figure 2.1). For these reasons, a EU-OECD definition was developed: the first concept to indicate cities refers to the space covered by an area of high population density and minimum population size. The second notion builds on the functional and economic extent of cities, thus including low density areas that surround cities and are closely linked to the cities' economic activity. The two notions combined describe a functional urban area (FUA), indicating the city and its surroundings that participate in the city labor market [23]. FUAs make a powerful tool to compare socioeconomic and spatial characteristics between cities, better capturing agglomeration economies. FUAs can also be classified according to their dimension: Metropolitan regions, FUAs of at least 250000 inhabitants, which can be further differentiated between Metropolitan Capital regions and just Metropolitan regions, and non-Metropolitan. This classification allows FUAs to incorporate the regional typologies used in OECD territorial classification scheme. The OECD scheme distinguishes units in Predominantly Urban (PU), Intermediate Regions (IN) and Predominantly Rural, separated between Relatively Close to Cities (PRC) and Relatively Remote from Cities (PRR). The classification is primarily based on population density of municipalities, closely tied to NUTS3 definition in most European countries. By implementing the classification between Metro and non-Metro regions, FUAs can partially include the logic behind the OECD scheme: Metro regions may account for PU and IR, as well as for PRC, while PRR always correspond to non-Metro regions. The FUAs' distinction is virtually better suited for European urban structure, as almost 60% of EU28 population lives in a Metro-region, although not necessarily in strictly Predominantly Urban areas [22].

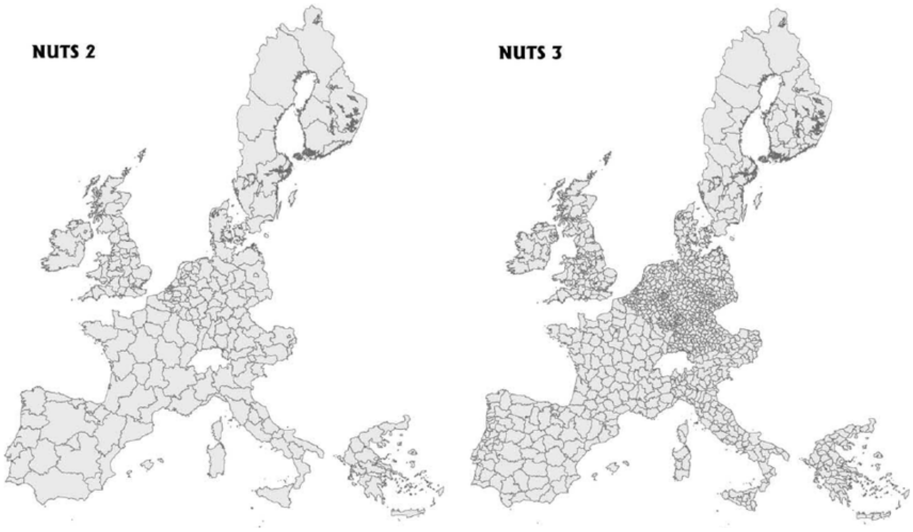



Figure 2.1: NUTS2 and NUTS3 [39]

Functional Urban Units represent a cohesive and comprehensive classification system that aims at replacing and unifying the current miscellaneous landscape of geographical units. I now present how FUAs are constructed in detail, to fully understand what they represent. Functional Urban Areas are constructed iteratively, using a clustering technique: first an urban center is identified, then the corresponding city is defined and lastly the relative commuting zone, or hinterland. A FUA is the combination of city with its commuting zone. To begin, an urban center is defined as a concentration of population in space, independently of administrative or political boundaries. Using a fine population grid cell, an iterative algorithm is used: first all grid cells with a density of more than 1500 inhabitants per square kilometer are selected. Then the contiguous high-density cells are clustered and subsequently only clusters with at least 50000 inhabitants are kept. Finally, looking at the selected cells, gaps in each cluster are filled. Cities are then defined as one or more local units, i.e municipalities and communes, with at least 50% of their population in an urban center. The smaller the local unit considered, the better the match obtained between urban center and city. It should be noted that it is usually complicated to retrieve data at this level of granularity. The commuting zone is instead identified as all the local units with at least 15% of their employed residents working in the city. A *polycentricity check* is carried out, to establish if two or more cities belong to the same Functional Urban Area - this is the case if more than 15% of the population of one city commutes to work to the other (Figure 2.2).

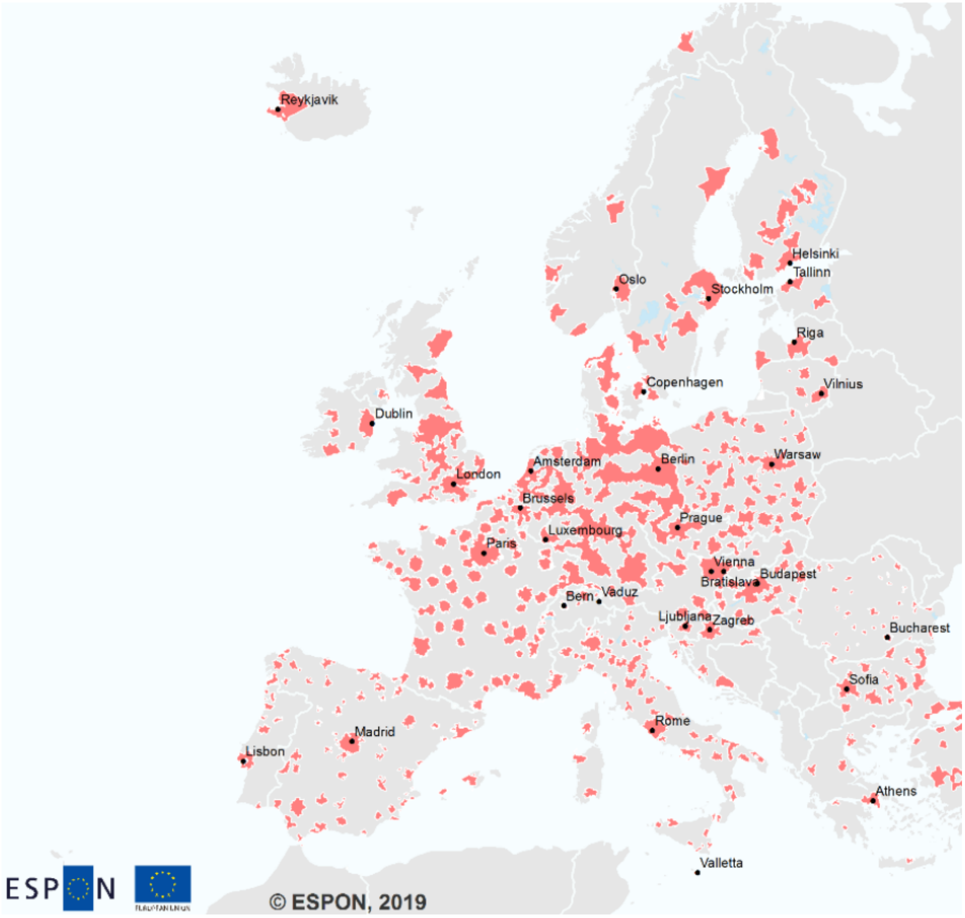



Figure 2.2: Functional Urban Areas across Europe, ESPON Database

In contrast with FUAs, NUTS3 are not constructed with the final goal of representing the full extent of cities' reach. This is a considerable drawback that could bias results and should be accounted for. Regardless, the data available at FUAs' level is scarce and usually estimated from NUTS3 data. NUTS 2 data are collected on a bi-annual basis using representative samples, whereas the NUTS 3 data is census information collected on a decade basis. As NUTS2 and NUTS3 are established measurement units that local and national governments recognize and implement, data collection at NUTS level is much more extensive and fitting to this analysis' goal. For these reasons, the analysis will be done using NUTS3 data, so to keep the finest granularity possible while accessing reliable and consolidated data. Using NUTS3 level will allow me to measure agglomeration economies across a wide range of sectors and extend the analysis up to 2020, in contrast with data at FUA level.

### 2.1.2. Sectors classification

Economic activities in the European Union are classified by NACE Rev 2, a system that first sort sectors in 21 categories, from A to U. Manufacturing, for example, is category C, while M is "Professional, scientifical and technical activities". Each letter category is then segmented in other subsections, so that M is composed by seven other sections: 69) Legal and Accounting activities, 70) Activities of head offices; management consultancy activities, 71) Architectural and engineering activities; technical testing and analysis, 72) Scientific research and development and so on. These sections can also be further sub-divided to more specific categories. The OECD proposed a classification of economic activities distinguishing Manufacturing in High Technology, Medium-High Technology, Medium-Low Technology and Low Technology and Service in High Technology Knowledge Intensive Sectors (HTKIS), Other Knowledge Intensive Sectors (OKIS) and Low Knowledge Intensive Sectors (LKIS). Each class is composed by activities from various categories, such as the example below:

**High Technology Knowledge Intensive Sector**

| Class | Code | Activity |
|---|---|---|
| J | 59 | Motion Picture, Video and Television production [...] |
| J | 60 | Programming and Broadcasting Activities |
| J | 61 | Telecommunication |
| J | 62 | Computer programming, consultancy and related activities |
| J | 63 | Information service and activities |
| M | 72 | Scientific research and development |

Table 2.1: OECD classification of economic activities based on NACE Rev 2

A simpler and more straightforward version of the OECD classification is that proposed by Oqubay and Lin [52]: the authors only differentiate between Manufacturing activities and Advanced Tradable Services (ATS). While Manufacturing is simply category C of NACE Rev 2, ATS is composed by Information and Communication (J), Financial services and insurance (K), Professional, scientifical and technical activities (M) and Administrative and support service activities (N), excluding any activity related to Real Estate. This second classification is better suited for this analysis, as data at NUTS3 level is available aggregated per NACE Rev 2 categories, thus it is not possible to split, for example, Class M into multiple categories as OECD does. Therefore, the classification used in this thesis will be as follows:

**Aggregation of economic activities based on NACE Rev 2**

| Category | Class | Activity |
|---|---|---|
| Manufacturing | C | Manufacturing activities |
| Advanced Tradable Services | J | Information and Communication |
| Advanced Tradable Services | K | Financial services and insurance |
| Advanced Tradable Services | M | Professional, scientifical and technical activities |
| Advanced Tradable Services | N | Administrative and support service activities |

Table 2.2: Selected aggregation of economic activities

### 2.1.3. Variables definition

This thesis estimates agglomeration economies at the NUTS3 level of regions for 26 countries. These include 23 countries from EU28, plus Albania, Macedonia and Norway. Five EU28 countries are not included in the analysis. Ireland and Estonia were excluded due to missing data, while Luxemburg, Malta and Cyprus were not accounted for because of their size, since the selected model, which will be discussed in the next chapter, requires at least two NUTS3 per country. For all countries data for the analysis is taken from ESPON Database Portal and covers the years 2005 to 2020. Variables considered were population at first of January [7] and gross value added (GVA) at basic prices. GVA is defined as the value of output less the value of intermediate consumption, representing the contribution of labour and capital to the production process. GVA data was collected twice: the first dataset refers to total GVA per NUTS3 [5], while the second differentiates between manufacturing GVA (NACE activity C) [1] and Advanced Tradable Services GVA (NACE activities J, K, M, N) [2][3][4]. Furthermore, NUTS3 are coded by Urban Type (Predominantly Urban, Intermediate, Predominantly Rural), proximity to coasts (On coast, Coastal with more than 50% of population living within 50 km of the coastline, Non-coastal) and Topography [6]. Finally, the analysis refers to NUTS 2016 classification due to data being available in this format.

## 2.2. Exploratory Data Analysis

Now that the dataset has been defined, a first exploration of the data is carried out. Before we begin, note that five German NUTS3 do not have data, often coded as

"0" in the output: these are concentrated in the German State of Saarland, which is therefore not represented by the analysis. I begin by checking the severity of regional inequalities in terms of Gross Value Added, choosing data from 2019 to avoid biases due to the production alt caused by Covid-19 in 2020. From the histogram, we can see a very unequal pattern, with the graph dominated by regions having GVA between 0 and 44735.4 millions of euros while very few regions reach up to 223677 millions of euros in 2019 (Figure A.1). Using the Equal Intervals map, it is clear that the vast majority of NUTS3 are concentrated in the lowest bin (1323 regions), while only two regions make up the first bin - the metropolitan region of Madrid and the very core of Paris - six compose the second and five the third (Figure 2.3).

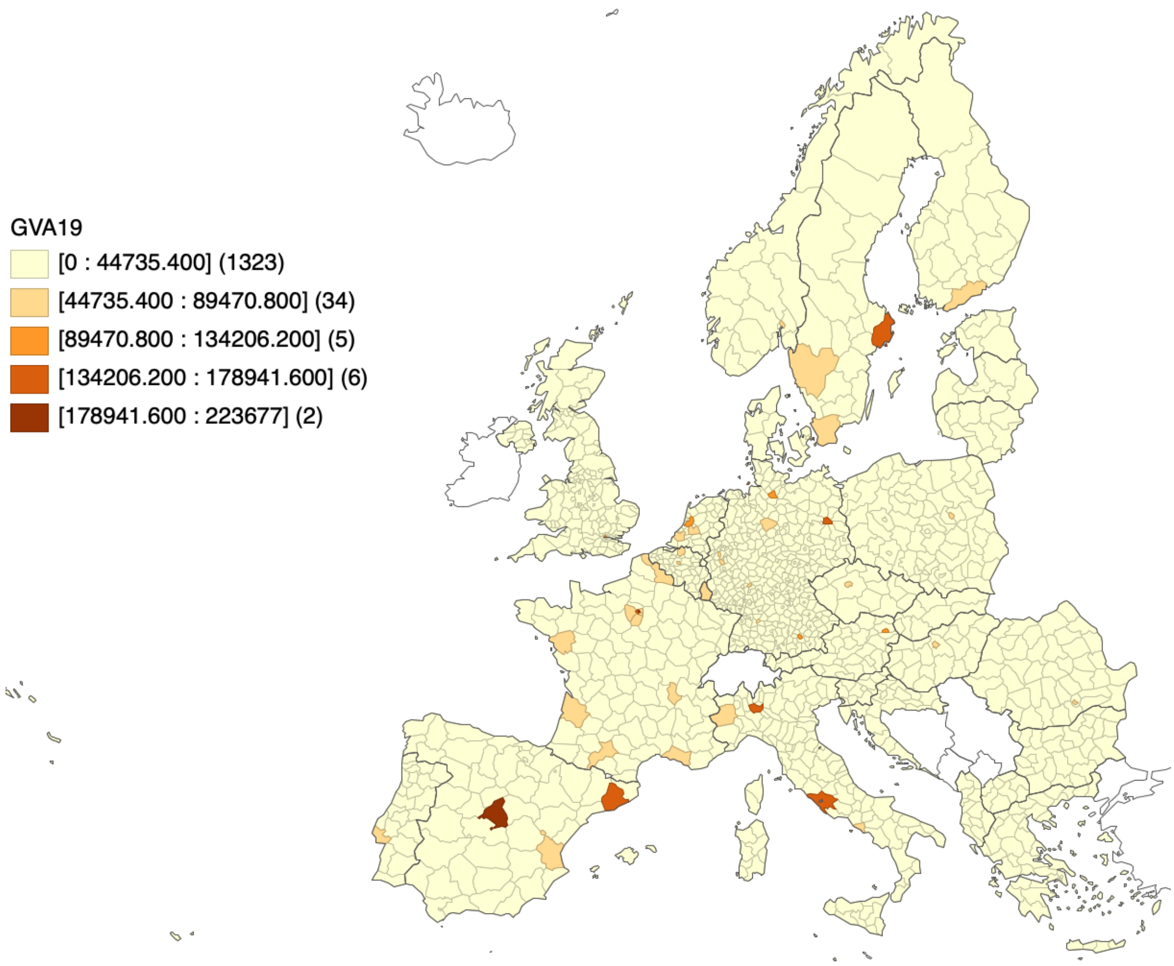



Figure 2.3: Gross Value Added in 2019 at NUTS3 level

Thus, the relevant regional inequalities detected by Iammarino et al. [42] are confirmed.

To better visualize the categorical variables with which NUTS3 are classified, I present two maps: the first shows the urban structure of European Countries (Figure 2.4). Type 1 indicates the Predominantly Urban regions, type 2 those Intermediate and type 3 the predominantly Rural areas.

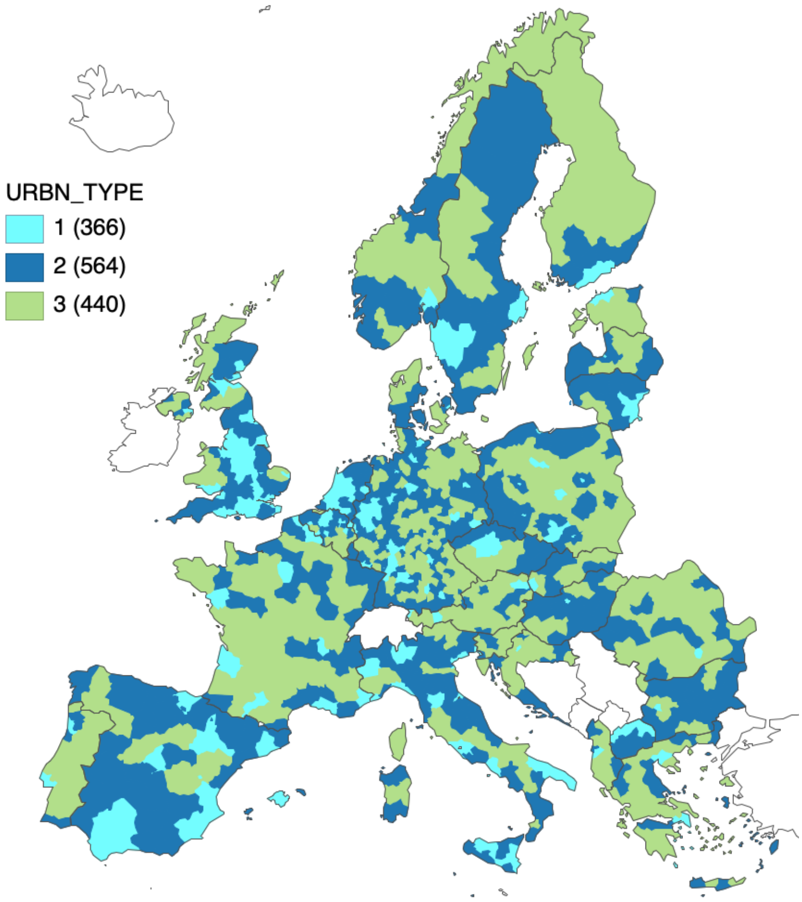



Figure 2.4: Urban Type classification of NUTS3 2016

The second map shows the topography of European Countries, classified as having either more than 50% of population living in mountain areas (type 1), more that 50% of surface in mountain areas (type 2) or more than 50% of population and 50% of surface in mountain areas (type 3). Else they are listed as "Other regions" (type 4) (Figure 2.5).

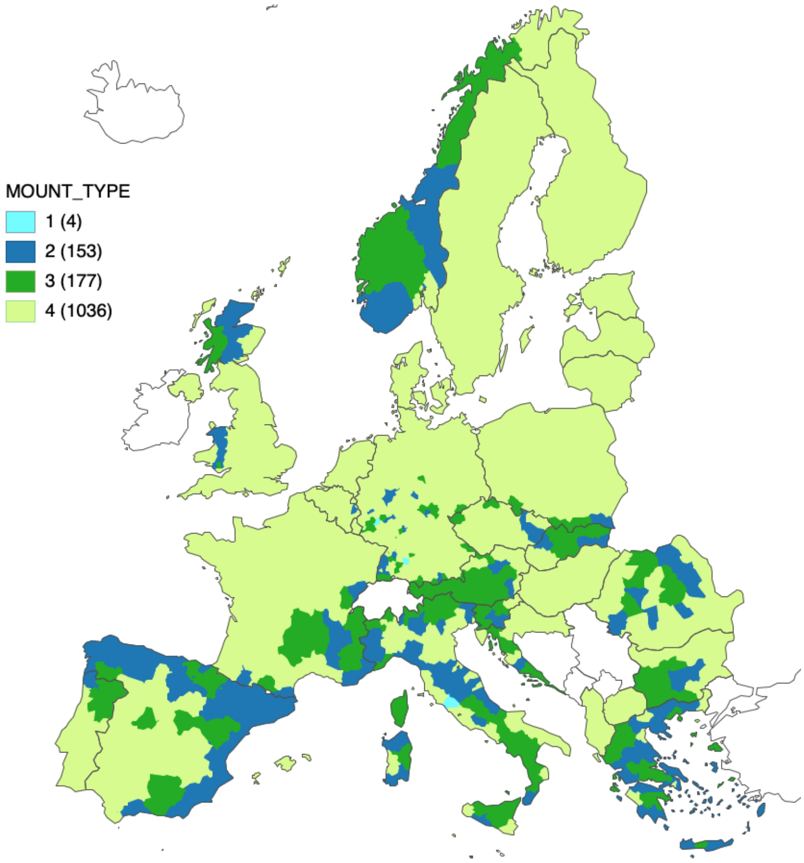
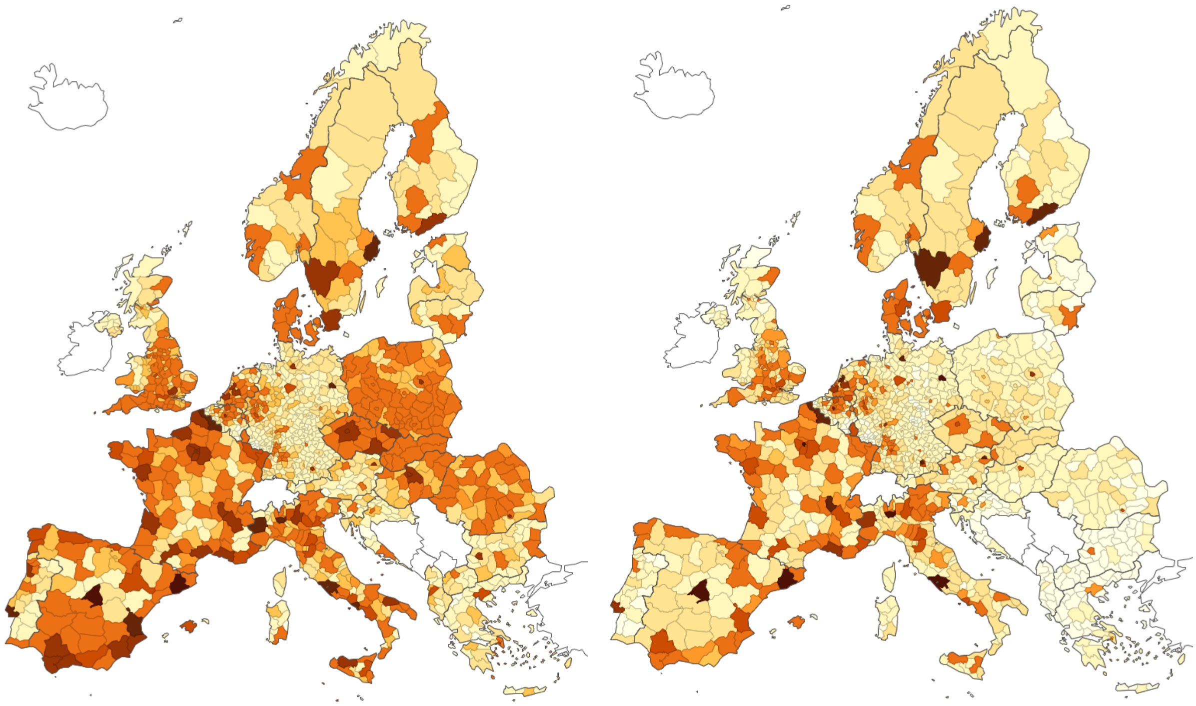



Figure 2.5: Mountain Type classification of NUTS3 2016

Having better understood the geographical characteristics that each variable represents, we can now start to investigate how population and overall GVA are distributed across European Countries. To do so, I use a clustering method called "Natural Breaks", a nonlinear algorithm that group observations such that the within-group homogeneity is maximized. I run the clustering on both population data and GVA data referring to 2019, breaking the data into 10 clusters. Below, the two outputs are compared (Figures 2.6a, 2.6b). To double-check the algorithm's outcome, I also run a K-means clustering algorithm, which instead attempts to minimize the variance within each cluster and assign each data point to the cluster with the closest mean (centroid). The results are sufficiently aligned with the Natural Breaks algorithm's (Figures A.2a, A.2b). In two figures below, a visual correlation between population and GVA can be identified, at least in EU15 countries: the most populous regions are also consistently the ones with the highest value added. Instead, Poland, Romania, Czechia and Hungary stand out, with the majority of regions being moderately populated, yet very low value adding capacities. An odd result is shown across German NUTS3: the majority appears to be both under-populated and under-performing compared to the rest of Europe.



(a) Population in NUTS3 in 2019 (b) Gross Value Added in NUTS3 in 2019

Figure 2.6: Comparison of population and GVA distribution across NUTS3 in 2019

A possible explanation for this bizarre finding can be traced back to the disparity in NUTS3 definition across EU Member States, previously discussed. Indeed, German NUTS3 are much smaller than the rest, often representing single cities or such fine-grained units. Since these analysis are performed on overall population and GVA, such small units appear to be much less populated and performing than other much bigger NUTS3, such as those found in Spain and France. To verify this hypothesis, I implement the Natural Breaks Algorithm on per capita productivity, modeled as GVA/population. Indeed, the resulting map corrects the previous misrepresentation and the result is aligned with findings from Ciccone and Hall [18] and Oqubay and Lin [52] (Figure 2.7). This issue regarding NUTS3 size inconsistency should be noted and taken into consideration when evaluating the results from the model estimating agglomeration economies in the next chapter.

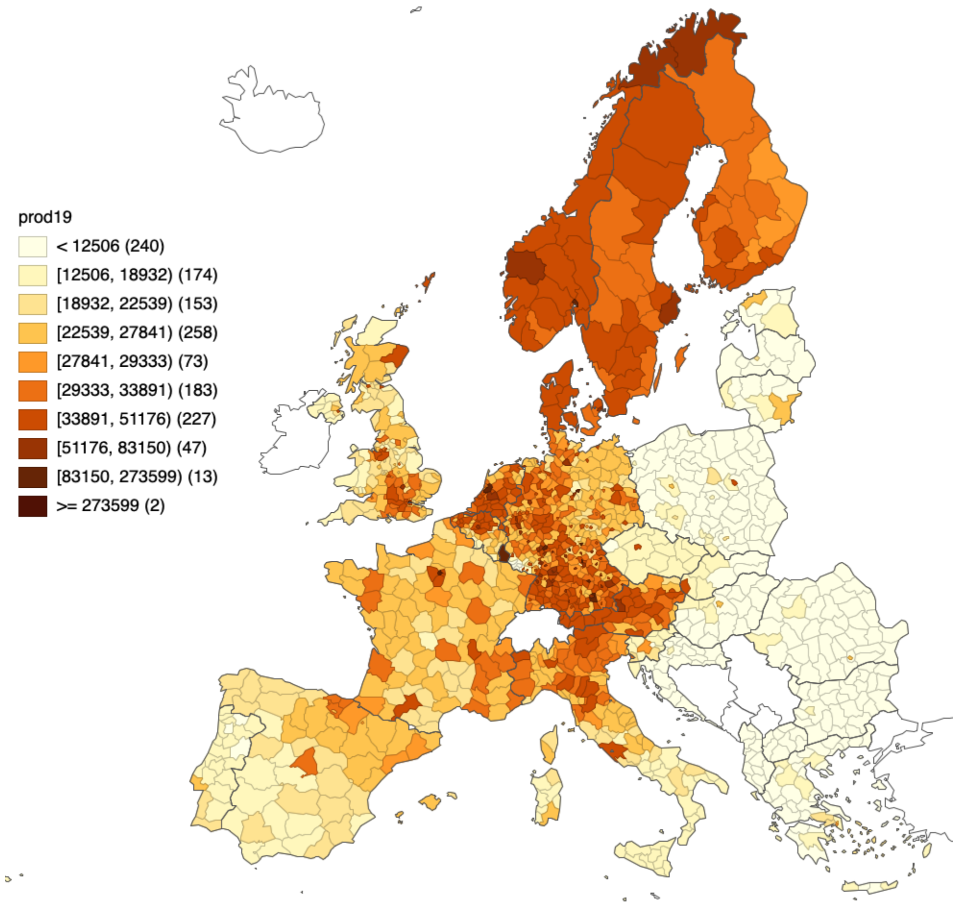



Figure 2.7: Per capita productivity in NUTS3 in 2019

Simultaneously, German NUTS3, being so fine-grained, offer us the chance to spot a pattern that could potentially constitute evidence of agglomeration economies. This is the most evident when mapping the distribution of the standard deviation of productivity in 2019. Standard deviation map can be used as a criterion to identify outliers: after transforming the productivity variable to a standardized variable with mean 0 and standard deviation 1, units are classified depending on how distant their productivity is from the mean. In Figure 2.8, a close-up of the output, the dark red units are more than two standard deviation above the European mean, similarly dark blue units are more than two standard deviation below. These could be interpreted as the outlier units, so regions much more productive or much less productive than average. Since major German cities are directly represented through NUTS3 units, the figure allows us to identify the vast majority of dark red outliers in Germany with cities. This interpretation is also supported by the fact that these smaller units are encircled by larger NUTS3, such as a city's periphery or larger metropolitan region. The same phenomenon is observed in other European countries, such as in the case of Paris, London, Copenhagen, Brussels and

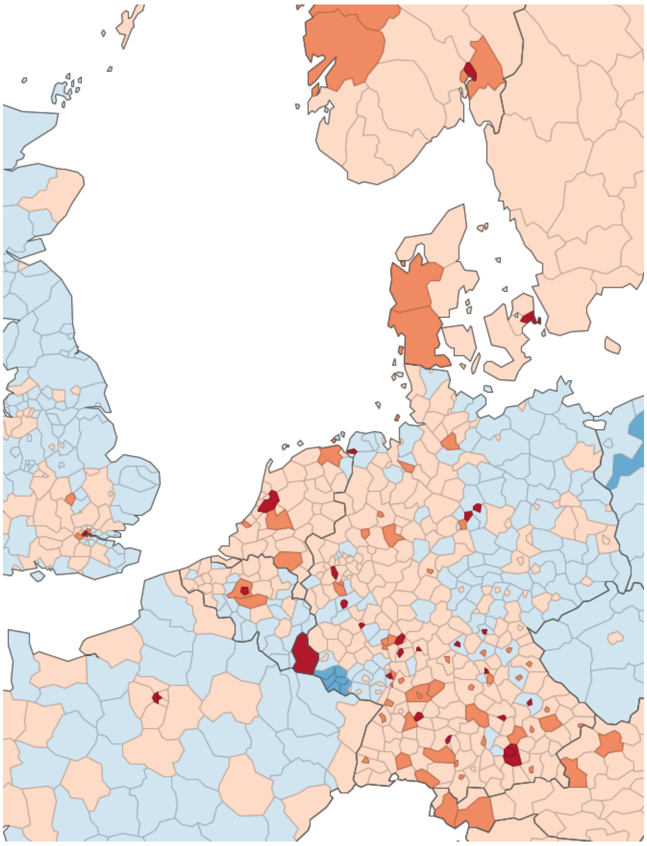
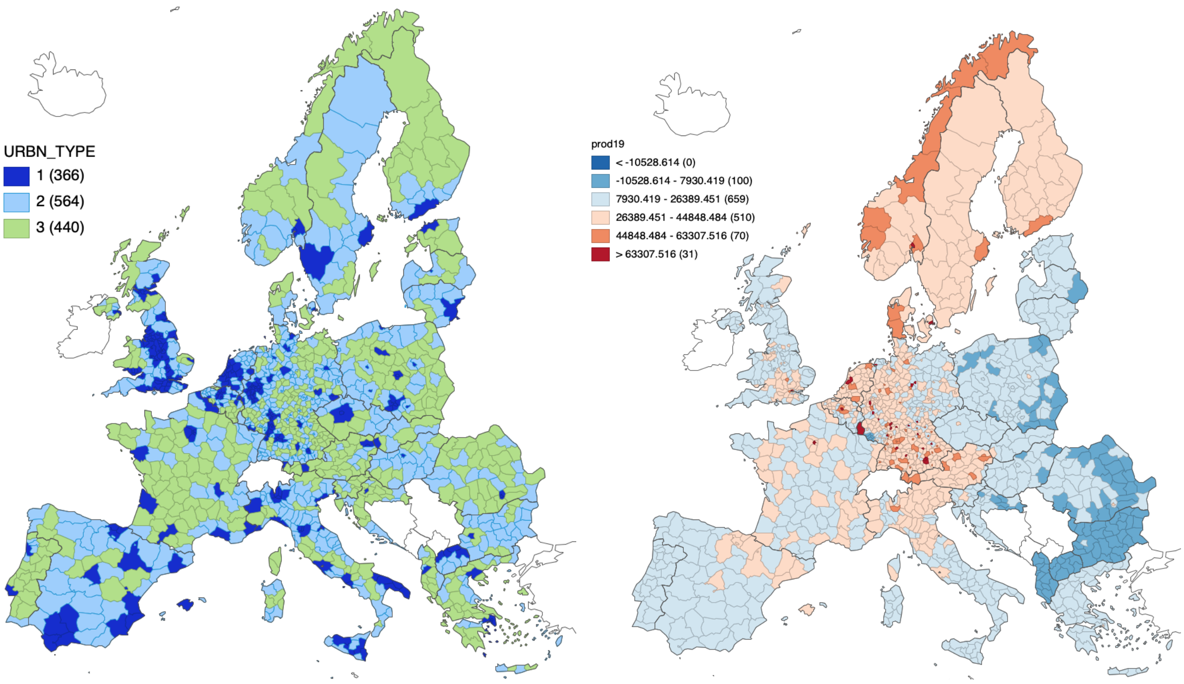


Figure 2.8: Standard deviation of productivity in 2019, detail

Oslo. Similarly, though less evidently, the larger region of The Hague-Rotterdam exhibits an agglomerating pattern. These outputs support this thesis' investigation, confirming the existence of agglomeration. In order to validate this intuition, I compare the standard deviation of productivity in 2019 with the urban classification of NUTS3, previously presented. The two maps are shown in Figures 2.9a and 2.9b: while the interpretation above is partially supported, as all dark red units are indeed predominantly urban areas and the vast majority of light orange and orange units are either urban or intermediate, agglomeration economies are not consistent across all countries. Central and Eastern European countries are systematically below standardized mean, regardless of the urban type of the single unit. This pattern could partially be explained by taking into account the fact that the considered mean is computed on all European countries, which are of course very different in terms of productivity. Therefore, Figure 2.9b provides evidence of divergence patterns in terms of European national economies, with Western Europe being much more productive than CEECs. It can potentially also constitute evidence of agglomeration economies at European level, if Western Europe is taken as the centre, thus the core, agglomerating economic activity, and the rest as periphery. What remains unseen is whether individual countries are experiencing agglomerating forces: if Romania

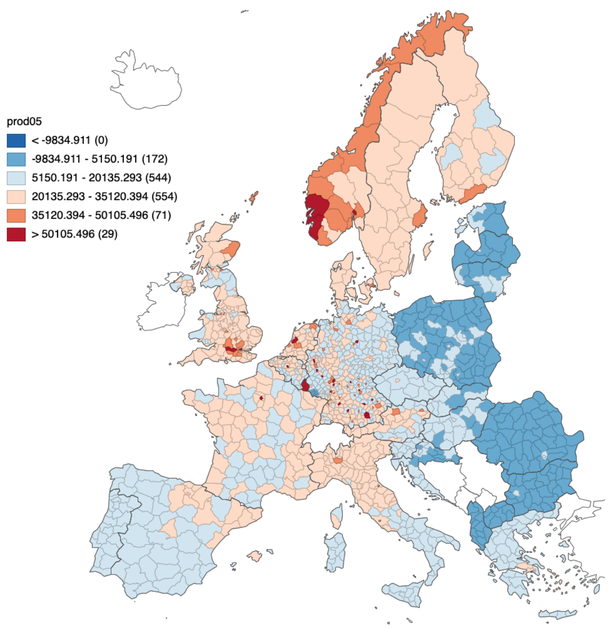



(a) Urban type of NUTS3 2016 (b) Standard deviation of productivity in 2019

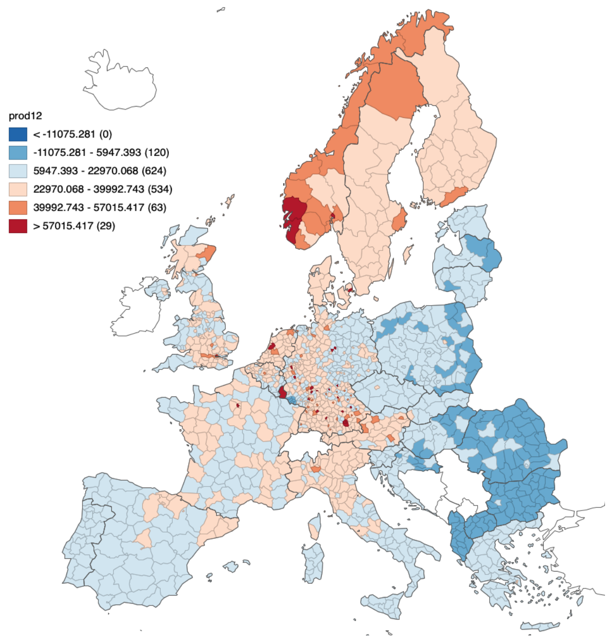
Figure 2.9: Comparison between NUTS3 productivity in 2019 and their urban type

were undergoing agglomeration economies, concentrating wealth in Bucharest, an analysis on such a coarse-grained level would not be able to catch it. Therefore, it is still to be determined whether agglomeration economies are on the rise and whether these patterns are common across all countries and sectors.

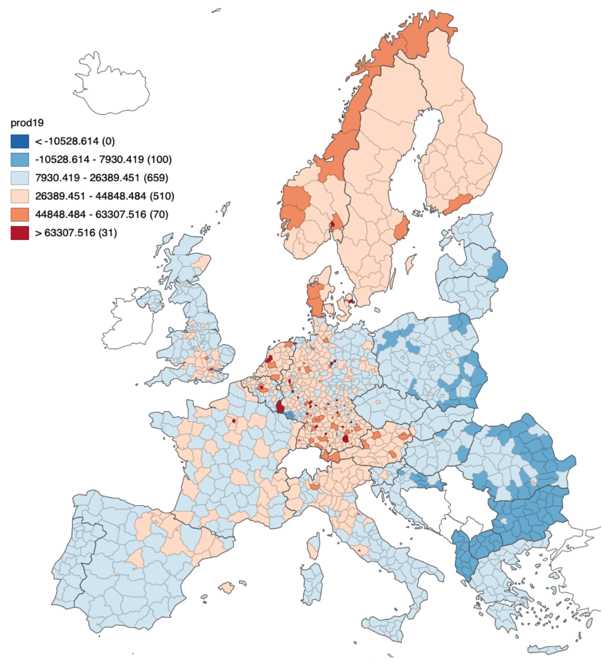
Finally, I look at how agglomeration patterns changed throughout the past 15 years, mapping standard deviation of productivity in 2005, 2012 (post 2008 crisis) and 2019 (Figures 2.10a, 2.10b and 2.10c). Improvement in CEECs' productivity can be observed, while the picture in Western Europe is less clear: the presence of regions one standard deviation above mean appear to decrease in Italy and France, while increasing in Germany. Norwegian productivity seems to decline as well, although none of these observations can be interpreted as a sign of agglomeration. The only case where this could be possible is the UK, which sees a sharp and stark decline in regional productivity across multiple NUTS3. The decline is observed the most outside of the capital region, showing a hint of possible agglomeration economies. To evaluate these observations analytically, I will implement an econometric model, which i will present in the next chapter.




(a) NUTS3 Productivity in 2005




(b) NUTS3 Productivity in 2012




(c) Standard deviation of productivity in 2019

Figure 2.10: Comparison between NUTS3 productivity in 2005, 2012 and 2019

# 3 | Design of Analysis: selection of the model

In the following subsections, I will discuss different econometrical models that can be used to estimated agglomeration economies. The selected model will then be discussed thoroughly.

## 3.1. Review of models

The first model to be presented is the one implemented by Haaland et al. [40]. The goal of the paper is to capture the importance of factors predicted by Traditional Trade Theory, New Trade Theory and New Economic Geography in the degree of geographical concentration in 35 industries in 13 countries within the European Union. The analysis is computed on a cross sectional panel of data between 1985 and 1992 to estimate the effect of market integration on the importance of different factors over time. The authors develop two models, one estimating relative concentration of trade and one estimating absolute concentration of trade. Both dependent variables are modeled as size of industry, measured in three different ways: output, Value Added and employment. New Economic Geography refers to changes solely in absolute concentration of economic activity, therefore I will discuss only the second model:

$$S_i{}^A = \alpha + \beta_1 EXPEN_i{}^A + \beta_2 SCECON_i + \beta_3 IO_i + \beta_4 NTB$$

$S_i^A$ indicates absolute trade concentration of industry $i$. $EXPEN_i^A$ refers to the market size effect predicted by New Trade Theory: it is measured in terms of absolute concentration of production, with:

$$EXPEN_i^A = \sqrt{\frac{1}{c}\sum_j \frac{e_{ij}}{\sum_j e_{ij}}}$$

$e_{ij}$ considers both final consumption and intermediate goods in country $j$ and industry $i$:

$$e_{ij} = Production_i + Imports_i - Export_i$$

$SCECON_i$ measures scale economies as the percentage reduction in average costs for each percentage increase in output. The authors note issues of possible mismeasurement or multicollinearity related to this variable, as it may be measuring potential scale economies instead of the target. $IO_i$ is used to indicate intra-industry linkages as the percentage usage of an industry's own product as intermediate good. Both $IO_i$ and $SCECON_i$ were included to account for reinforcement effect between the two variables, as predicted by New Economic Geography. Finally, NTB stands for Non Tariff Barriers, as no tariff on manufacturing production exists within the EU. The authors implement the classification developed by Buigues et al. [16], adjusted to fit industry classification available on EUROSTAT. The authors use data from OECD STAN database and EUROSTAT at NUTS0 level. Reverse causality was taken into consideration as concentration of production could be caused by trade concentration patterns. The potential endogeneity was accounted for by implementing 2SLS and using lagged variables of $EXPEN_i^A$ as IV instruments. The model provides appealing results, with market size being the most important determinant of industrial localization. The relative influence of market size effects on industrial concentration shows strong support for the presence of agglomeration economies, as well as the relevance of intra-industry linkages. Nevertheless, it proves to be a limited instrument as scale economies appear to be mis-measured and provide inconsistent results. Furthermore, the model is unnecessarily complex for the scope of this thesis, which only target the measurement of agglomeration economies, disregarding other trade theories.

The second model to be discussed is implemented by Batista et al. [13]. The report focuses on measuring convergence trends across European Member States and projecting observed growth rates in the future to estimate whether divergence or convergence will be achieved. This is obtained by estimating sigma-convergence: this convergence occurs when the dispersion of income per capita across different economies falls overtime, resulting in a reduction of disparities [63]. To assess the sigma-convergence, the authors choose to compute the coefficient of variation (CV) over time for a set of European regions: an higher CV means higher dispersion of distribution, therefore higher regional disparities. Other possible indexes available for this computation are the Gini coefficient, Atkinson index, Theil index and mean logarithmic deviation. The coefficient is computed using data from NUTS3 units, covering from 2000 to 2011: basic demographic and economic variables were taken from the EUROSTAT, while output variables considered were total population, GDP per capita, GDP, employment and productivity, modeled as GDP/employment. It should be noted that GDP was expressed in PPS, or euro/PPP coefficient, an artificial currency used in the European Union to reflect the price differences observed across countries and allow comparison across Europe. The model provides a reliable estimator of convergence, as well as the possibility of using a diverse range of indexes to enhance results.

However, estimating the presence or absence of convergence would not reveal whether agglomeration forces are occurring across Europe. Indeed, agglomeration economies could occur if economies are converging if the main drivers of catch-up in poorer countries are larger cities. Therefore, this model does not provide the tools our analysis needs.

The last relevant model is the one developed by Roemish [59]: the core intuition relies on Zipf's law, often deemed as a reliable and robust estimator for city size distribution. Zipf's law draws a log-linear relationship between city size and its rank in the hierarchy for size of cities within a country:

$$P_i = \frac{A}{r_i^b} \; ln(P_i) = ln(A) - bln(r_i)$$

P is the size of region *i*, modeled as population, A is a constant closely represented by the population of the largest region, r is the rank of region *i* in the hierarchy of regions ranked by population size, with 1 being the largest. Finally, b is the Pareto exponent that determines the shape of the distribution: if it follows Zipf's law, b=1, which means that the biggest region is *i* times as big as the *i*-est region. If b<1, the population follows a more even distribution: thus b can be interpreted as a measure of agglomeration - the higher b, the higher the weight of the largest regions in the total population of a country. By using b as distribution parameter, the analysis can be independent from the number of regions in a country: this is especially relevant in the European context. With other indexes, such as the mean logarithmic deviation and Herfindahl index, the size of the index depends on the number of observation and the distribution parameter – therefore when comparing two countries with same distribution but different number of regions, the indexes might not be a reliable measure in cross country comparison of agglomeration. Starting from this, Roemish [59] argue that, assuming Zipf's distribution of population, if agglomeration economies do not occur, then the distribution of economic activity should mirror that of distribution. If, instead, agglomeration economies occur, then economic activity should be higher in more populated regions, as these regions would be more productive per capita. The model developed, therefore, takes productivity as $Y_i = E_i/P_i$, with $E_i$ measuring the economic activity. If no agglomeration occurs, then $Y_1 = Y_2 = Y_3$, else I should find that $Y_1 > Y_2 > Y_3$. $E_i$ is assumed to follow a Zipf's distribution:

$$E_i = \frac{B}{r_i^c}$$

B represents the economic activity in the largest region, c is the Pareto coefficient specific to $E_i$. Therefore, assuming that regions *i* only differ for population size and nothing else:

$$Y_i = \frac{D}{r_i^(c-b)} \text{ with } D = \frac{B}{A}$$

$$ln(Y_i) = ln(D) - (c-b)ln(r_i)$$

The log-linearized equation is a simple yet effective test for existence, importance and size of agglomeration economies. A similar model is implemented by Ciccone [17] to test agglomeration effect in European countries, achieved by regressing productivity on employment density and adding some control variables. Indeed, the coefficient for employment density represented the entity of agglomeration economies.

To estimate agglomeration economies, the author models Y as employment to population ratio and global value added to population ratio, using as independent variable the rank for population. The results show that distribution of population is indeed well explained by Zipf's distribution, meaning that the fundamental assumption of the model is respected. Moreover, findings from the implementation of the model on data from 23 countries of EU28, from 2000 to 2012 at NUTS3 level, are well aligned with similar estimations of agglomeration economies referring to the same period. Therefore, this model provides a strong theoretical structure for our analysis, while offering a relatively straightforward instrument to measure agglomeration economies, whose results can be easily read and interpreted.

## 3.2. Model construction

The model I will implement is therefore the one presented by Roemish [59]: as the one presented above relies on the assumption that regions only differ in population size, the enhanced version will now be presented. In the extension, control variables X can be added to represent differences between regions such as skill endowment, sectoral structure etc:

$$Y_i = \frac{DX^a}{r_i^(c-b)} \text{ with } D = \frac{B}{A}$$

In case the control variable could be subject itself to agglomeration effects, Roemish (2015) propose to use an auxiliary equation to obtain an unbiased estimator for X and for r. The auxiliary equation models X as $X_i = \frac{Z}{r_i^x}\varepsilon_i$, with x representing the agglomeration coefficient of $X_i$ and $\varepsilon_i$ the variation over the data generating the process. By log-linearizing the auxilary equation, I can distinguish between the agglomeration effect $ln(r_i)$ and the region's characteristics $ln(\varepsilon_i)$:

$$ln(X_i) = ln(Z) - xln(r_i) + ln(\varepsilon_i)$$

Then the unbiased estimator can be obtained by substituting in the main equation residuals from the auxiliary regression:

$$ln(Y_i) = ln(D) - (c - b)ln(r_i) + \beta ln(\varepsilon_i) + u$$

To avoid autocorrelation between the Xs and endogeneity due to reverse causality, I will not include any socio-economic control variable, as it is very likely to be affected by agglomeration economies. Instead, to maintain endogeneity at minimum, I will only include fixed geographical effects, which can not be influenced by agglomeration effects. Examples of these variables can be found in Chapter 2, such as the geographical units' typography or their proximity to coastal borders.

# 4 | Implementation of the model and Analysis

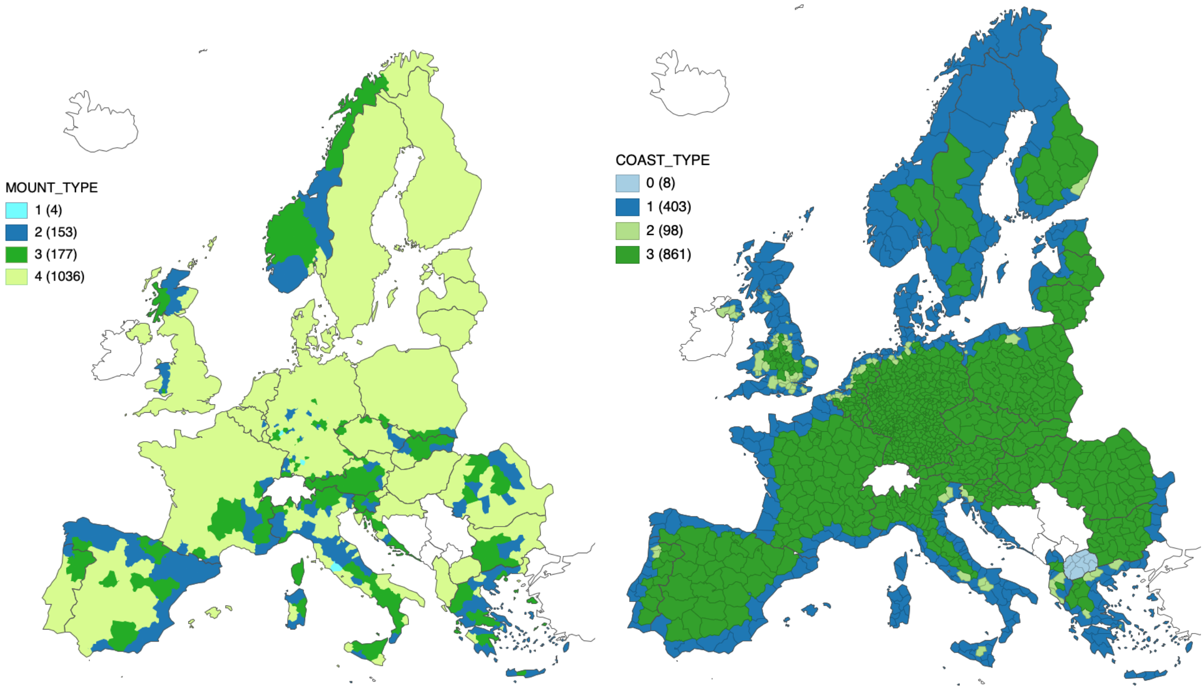
## 4.1. Model 1 - overall productivity

### 4.1.1. Data

The first model is implemented on 2016 NUTS3 regions for 23 countries from EU28, plus Albania, Macedonia and Norway. For all countries, data for the analysis is taken from ESPON Database Portal and covers the years 2005 to 2020. Variables considered were population at 1st of January and overall gross value added (GVA) at basic prices. Furthermore, NUTS3 are coded by proximity to coasts and topography of territory.

### 4.1.2. Model

The first model implemented for the analysis takes as dependent variable Y productivity as total GVA/population. The predictor used is $log(rank_i)$, as explained in Chapter 3. Two possible control variables are taken into consideration, $mount_i$ and $coast_i$. The first classifies the topography of the region in four types (Figure 4.1a) - having either more than 50% of population living in mountain areas (type 1), more that 50% of surface in mountain areas (type 2) or more than 50% of population and 50% of surface in mountain areas (type 3). Else they are listed as "Other regions" (type 4). The second variable classifies NUTS3 based on their proximity to the coast line: On coast (type 1), Coastal with more than 50% of population living within 50 km of the coastline (type 2) and Non-coastal (type 3) (Figure 4.1b). Three models were created. A different model was used on each country after assessing whether the control variables were significant or not - in some cases it was taken into consideration the fact that some countries were landlocked or all NUTS3 were classified the same in mount-type, as it is the case for Belgium, Albania and the Netherlands. The first model includes both control variables and was implemented on data referring to Portugal and Slovenia, as both variables appeared to be significant




(a) Mountain Type classification (b) Coast Type classification

Figure 4.1: Control variables: $mount_i$ and $coast_i$

predictors of productivity modelled as GVA/population.

$$ln(productivity_i) = ln(D) - (c - b)ln(rank_i) + \beta_1 mount_i + \beta_2 coast_i + u$$

The second model only includes the variable $coast_i$ and was implemented on data from Belgium, Finland, France and Italy. In these cases, the variable $mount_i$ was not found to be a significant predictor.

$$ln(productivity_i) = ln(D) - (c - b)ln(rank_i) + \beta_1 coast_i + u$$

The last model is the most basic, excluding all control variables: this is the most implemented model, used for Albania, Austria, Bulgaria, Czech Republic, Germany, Denmark, Estonia, Greece, Spain, Croatia, Hungary, Latvia, Lithuania, Macedonia, the Netherlands, Norway, Poland, Romania, Sweden, Slovakia and the UK.

$$ln(productivity_i) = ln(D) - (c - b)ln(rank_i) + u$$

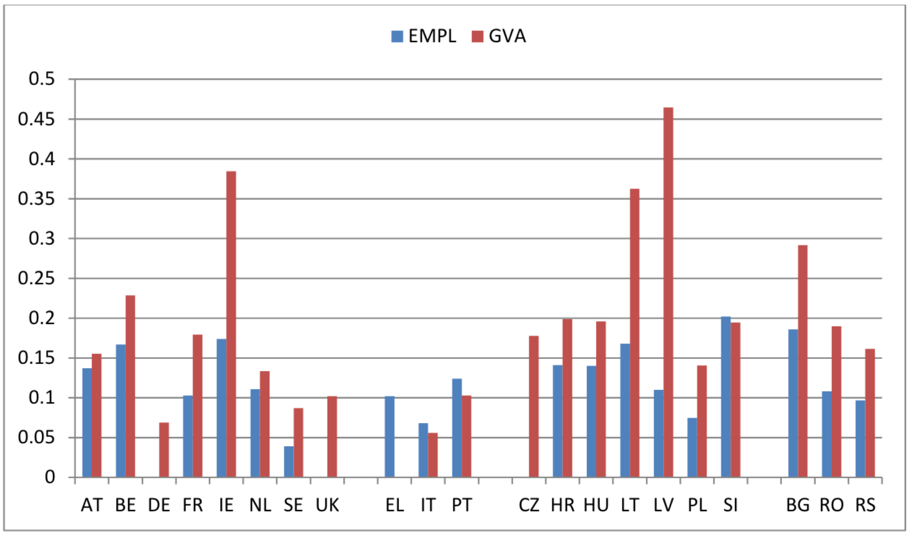
### 4.1.3. Results

Model's results for Spain, Latvia, Macedonia, Portugal, Slovakia, Denmark and the UK are not significant (i.e., the coefficient $(c - b)$ is not significantly different from zero), meaning that the model does not detect any agglomeration economies through out the

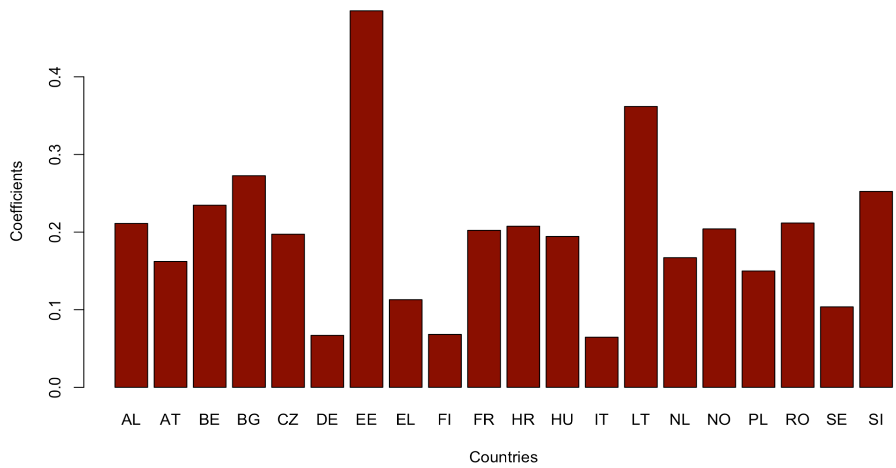
16 years horizon for these seven countries. Albania, Finland, Hungary and Estonia have significant coefficients for $log(rank_i)$ with p $<$ 0.05, while the rest of the countries are all significant with p $<$ 0.02 or p $<$ 0.01. The variance explained by the model can be found in Table B.1 and Table B.2 for all years. The adjusted R-squared varies significantly between countries, with Romania, Hungary and Greece have an adjusted R-square between 10% to 20% and the greatest proportion of variance is explained by the model in Slovenia, as the adjusted R-square is often over 90%. Germany, conversely, have the lowest adjusted R-square value, at only 4%. This is aligned with the estimated coefficient for agglomeration economies in Germany, which is very close to zero. This result could seem conflicting with the findings of Figure 2.8: this could be explained by considering the fact that the first model's results only refer to German productivity, while the analysis in Chapter 2 considers the whole Europe. Therefore, while German cities are much more productive than European average and show signs of agglomeration forces, they are not significantly more productive than German average, thus meaning that GVA productivity in Germany is evenly distributed.

The countries with the highest coefficient are Estonia, with 0.59 in 2016, and Lithuania, with 0.42 in 2020. According to the model construction, agglomeration economies can be understood to occur when the coefficient $(c-b)$ is equal to 1. Thus, generally speaking, no European country shows clear and definite agglomeration economies, although some, as in the cases of Estonia and Lithuania, as well as Slovenia, Romania and Bulgaria, have stronger agglomeration economies than others. Results from the model are aligned with Roemish [59], except for differences in the countries considered in the analysis, as shown in Figures 4.2a and 4.2b.

I plotted the coefficient for $log(rank_i)$ across time for the countries for which the analysis is significant and three major patterns can be detected. Albania, Belgium, Bulgaria, Czech Republic, France, Lithuania, the Netherlands and Poland all show a clearly growing trend, although the strength of agglomeration economies differs in each country (Figures A.3a, A.3c, A.3d, A.4a, A.5b, A.6b, A.6c, A.7a). The second pattern detected shows a declining trend in agglomeration economies, observable in Austria, Germany, Norway and Slovenia. Again, the size of this decrease depends on the country: the most remarkable is shown in Figure A.6d, referring to Norway, which sees agglomeration economies going from 0.22 in 2005 to 0.125 in 2020. The trend in Austria and Slovenia, instead, shows a smaller decrease (Figures A.3b and A.7d). In Germany, although significant and decreasing, the estimated agglomeration economies are very small, barely reaching 0.08 at their peak (Figure A.4b). Finally, the last group is composed by countries showing an oscillating pattern in estimated agglomeration economies. Some show agglomeration economies




(a) Agglomeration economies in 2012, Roemish [59]




(b) Agglomeration economies in 2012

Figure 4.2: Estimated agglomeration economies in 2012, comparison between Roemish [59] and this thesis' results

to be growing and then declining between 2005 and 2020, such as Sweden, Romania, Estonia and Croatia (Figures A.7c, A.7b, A.4c, A.5c). For others, on the contrary, agglomeration economies decrease and then grow again, namely Finland, Greece, Hungary (Figures A.5a, A.4d, A.5d). Italy has a very irregular graph, opening up the possibility that 2020 data should be considered an outlier and excluded from analysis: regardless, agglomeration economies are very small, similar to Germany and, apart for 2020, show a rather constant pattern (Figure A.6a).

By looking at the change observed in agglomeration economies from 2005 to 2020 (Figure 4.3, only significant values reported), it is possible to see both the direction and the size of the change in each country. Note that, given the outlier detected, the change for Italy was computed between 2005 to 2019. It can be observed how agglomeration economies decreased or just slightly increased for European countries with higher level of economic development (with the exception of France), while larger increases were observed across CEECs, SEE countries and Lithuania, while Estonia shows a small negative change.

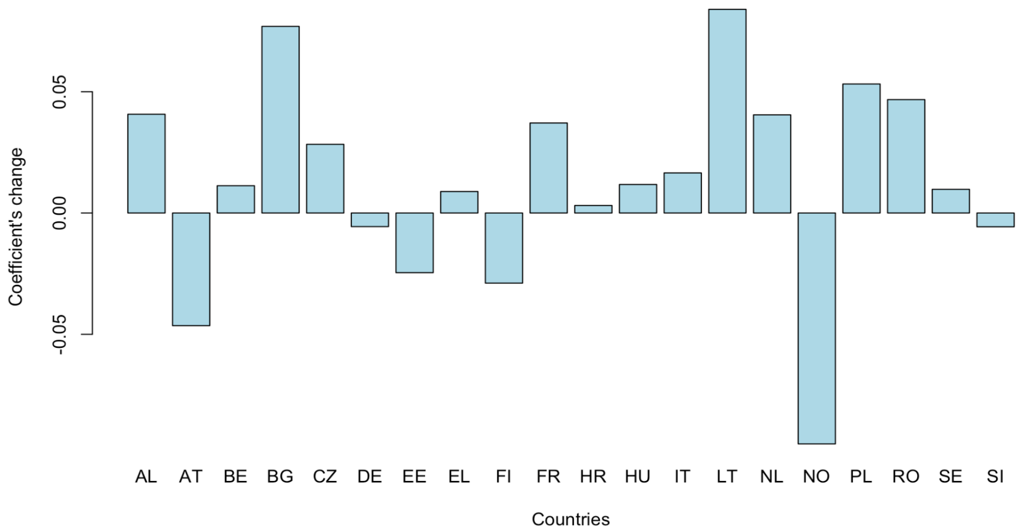



Figure 4.3: Change in agglomeration economies between 2005 and 2020

To investigate this further, I examined the estimated agglomeration economies in 2005, 2012 and 2020: the strongest effect is observed in Baltic countries and Slovenia, while Italy, Finland and Germany are among the least affected by agglomeration economies. From Figures **??** 4.2b and A.9 the variation detected between 2012 and 2020 seems weaker than the variation between 2005 to 2012. For better clarity, figures 4.4 and 4.5 show the coefficient's change between 2005 and 2012 and 2012 to 2020. Provided that we agree to interpret the detected variation as partially influenced by the 2008 financial crisis, it seems that countries being affected the most by agglomeration forces are Bulgaria, Czech Republic and Romania, followed by Albania, France and Croatia. Instead, Finland shows the strongest decrease, although it should be noted that such variation is about -0.04. Conversely, the countries showing the biggest change between 2005 and 2012 do not show a similar pattern between 2012 and 2020: the model detects a minimal change in coefficient in these countries, with the exception of Croatia, which instead shows a decrease. In figure 4.5, Poland and Lithuania both show the highest increase, about +0.06. The countries reporting a consistent pattern in both time frames are Austria, with agglomeration economies decreasing of about -0.02, and Norway, which goes from a limited decrease of -0.01 in figure 4.4 to -0.08 in figure 4.5. Finally, the three countries showing the strongest agglomerations, Estonia, Lithuania and Slovenia, do not have similar growth trends in coefficient's change. Estonia shows a persistent, although weak, decreasing trend. Instead, Lithuania reports a positive change, stronger in the second time frame (almost +0.06). Slovenia sits in between the extremes, as its coefficient increases in the aftermath of the 2008 financial crisis and then decrease in the following decade. A specific geographical pattern is not easily recognized, posing the need for a more in-depth analysis.

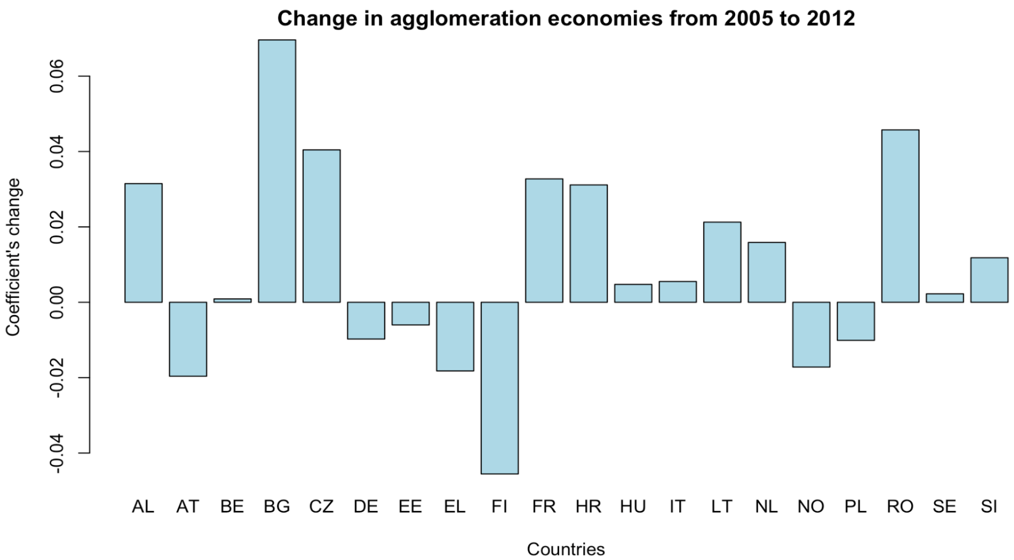



Figure 4.4: Change in agglomeration economies between 2005 and 2012

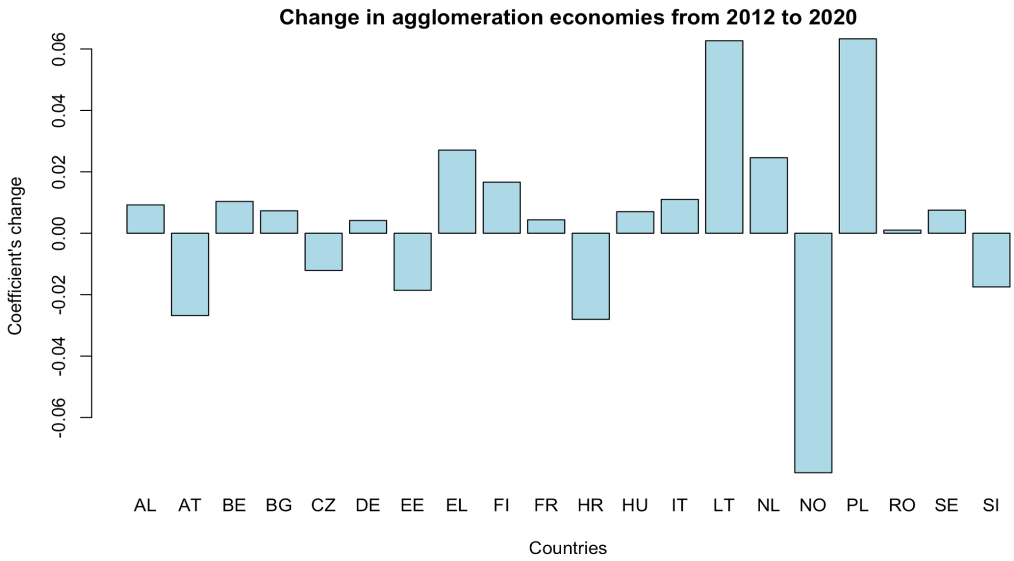



Figure 4.5: Change in agglomeration economies between 2012 and 2020

## 4.2. Model 2.a and 2.b - specific sectors' productivity

### 4.2.1. Data

Model 2.a is implemented on 2016 NUTS3 regions for 22 countries from EU28, plus Albania and Macedonia. Estonia, the United Kingdom and Croatia are not taken into the analysis due to inconsistent data referring to manufacturing GVA. Model 2.b is implemented on 2016 NUTS3 regions for 22 countries from EU28, plus Albania and Macedonia. Estonia, Croatia, Austria, the United Kingdom, Spain, Poland and Germany are not taken into the analysis due to lack of data referring to Advanced Tradable Services GVA. ATS is composed by Information and Communication (J), Financial services and insurance (K), Professional, scientifical and technical activities (M) and Administrative and support service activities (N), excluding any activity related to Real Estate. The only data available for Romania referred to sector J, thus the country was excluded from the analysis as well. Two Belgian NUTS3 (BE223 and BE323) were missing GVA data on all four sectors for the years between 2007 and 2017, thus interpolation was used.

For all countries data for the analysis is taken from ESPON Database Portal and covers the years 2005 to 2020. Variables considered were population at 1st of January and gross value added (GVA), differentiating between manufacturing GVA (NACE activity C) and Advanced Tradable Services GVA (NACE activities J, K, M, N) at basic prices. Furthermore, NUTS3 are coded by proximity to coasts and topography of territory.

### 4.2.2. Model

The second model implemented for the analysis takes as dependent variable Y two measures of productivity - manufacturing GVA/population $Y_a$ and ATS GVA/population $Y_b$. The predictor used is $log(rank_i)$, as explained in Chapter 3. Two possible control variables are taken into consideration, $mount_i$ and $coast_i$. As for the previous model, the first classifies the topography of the region in four types (Figure 4.1a) - having either more than 50% of population living in mountain areas (type 1), more that 50% of surface in mountain areas (type 2) or more than 50% of population and 50% of surface in mountain areas (type 3). Else they are listed as "Other regions" (type 4). The second variable classifies NUTS3 based on their proximity to the coast line: On coast (type 1), Coastal with more than 50% of population living within 50 km of the coastline (type 2) and Non-coastal (type3) (Figure 4.1b).

As in the previous Section 4.1, three models were created for each dependent variable. Again, a different model was used on each country after assessing whether the control variables were significant or not. Regarding model 2.a, with $Y_a$= manufacturing GVA/population, the model which includes both control variables was implemented on data referring to Spain and France, as both variables appeared to be significant predictors.

$$ln(Y_{ai}) = ln(D) - (c - b)ln(rank_i) + \beta_1 mount_i + \beta_2 coast_i + u$$

The second model only includes the variable $coast_i$ and was implemented on data from Germany, Italy and the Netherlands. In these cases, the variable $mount_i$ was found to be non-significant predictors.

$$ln(Y_{ai}) = ln(D) - (c - b)ln(rank_i) + \beta_1 coast_i + u$$

The last model is the most basic, excluding all control variables: this is the most implemented model, used for Albania, Austria, Belgium, Bulgaria, Czech Republic, Denmark, Greece, Hungary, Finland, Latvia, Lithuania, Macedonia, Poland, Portugal, Romania, Sweden, Slovenia and Slovakia.

$$ln(Y_{ai}) = ln(D) - (c - b)ln(rank_i) + u$$

Regarding model 2.b, no country was found to have both categorical variables as significant predictors of $Y_b$=ATS GVA/population. The model including only $mount_i$ as control variable was implemented on Greek data.

$$ln(Y_{bi}) = ln(D) - (c - b)ln(rank_i) + \beta_1 mount_i + u$$

The model including $coast_i$, instead, was used for Belgium, Finland, France, Italy and Slovenia.

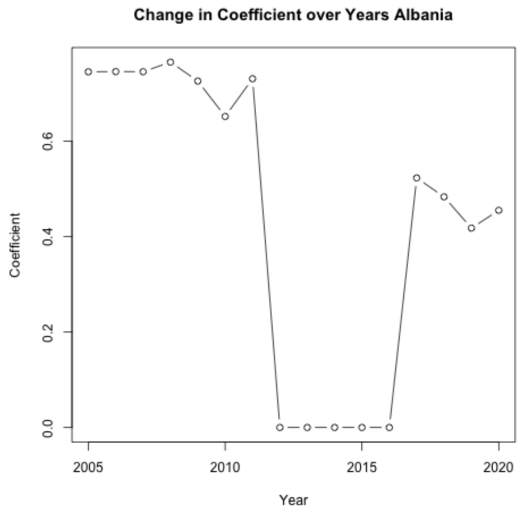
$$ln(Y_{bi}) = ln(D) - (c - b)ln(rank_i) + \beta_1 coast_i + u$$

Finally, the simplest model, including only the predictor variable of $ln(rank_i)$, was used for Albania, Bulgaria, Czech Republic, Denmark, Hungary, Lithuania, Latvia, Macedonia, the Netherlands, Portugal, Sweden and Slovakia.

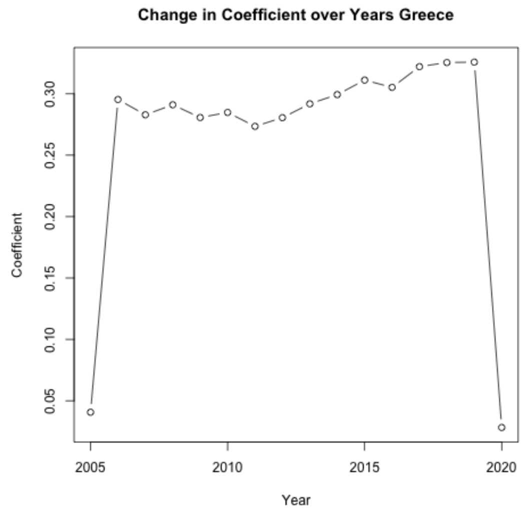
$$ln(Y_{bi}) = ln(D) - (c - b)ln(rank_i) + u$$

### 4.2.3. Results

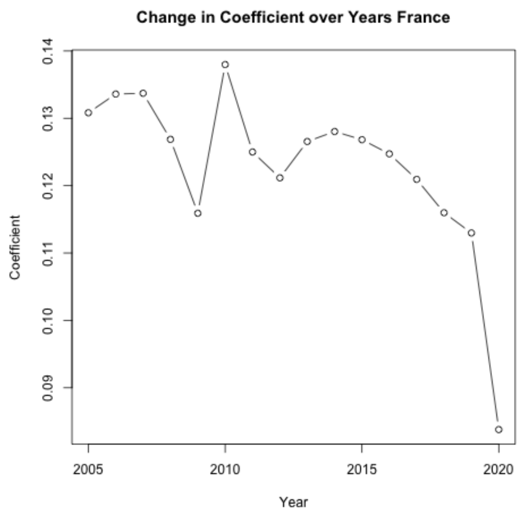
I start by discussing the results of the model with productivity $Y_a$: the great majority of the countries analyzed gives non-significant results, meaning that the null hypothesis that agglomeration economies coefficient is zero cannot be rejected. Thus, agglomeration economies have weak or null effects on manufacturing productivity for almost all countries, with the exception of Albania, Greece, France and Romania. In these instances, the variable $log(rank_i)$ is a significant predictor with $p < 0.05$ or $p < 0.02$. I plotted the coefficient for $log(rank_i)$ across time for these countries. As it can be observed in Figures 4.6a, 4.6b, 4.6c and 4.6d, the $log(rank_i)$ coefficient for both Albania and Romania becomes insignificant through the years. Greece and France, instead, have consistent results in terms of significance. Finland also gives significant results for the first four years of the analysis, as shown in Figure A.10. Agglomeration economies in manufacturing are, in most cases, weaker than $coast_i$ and $mount_i$ effects on productivity: indeed, the model for Germany, Spain and Italy, despite non-significant $log(rank_i)$, shows very strong significance of the geographical variables as predictors. For Italy the average difference in $Y_a$ between each coastal categories is between 0.40 and 0.57 through out the years, meaning that the further away the NUTS3 is from the coast, the higher the manufacturing productivity predicted. For Germany, the difference through out the years is positive but slightly less pronounced, oscillating between 0.30 and 0.37. In the case of Spain, both geographical variables are significant, but their coefficients have opposing signs: as for Germany and Italy, being further away from the coast increases the predicted productivity for each category, with the estimated coefficient growing from 0.4 to 0.56 from 2005 to 2020. Instead, the average difference in $Y_a$ between the categories of $mount_i$ is negative, meaning that moving from type 1 to type 2 to type 3 and so on decreases the predicted productivity of 0.40 to 0.50 from 2005 to 2020. Finally, the model finds that both categorical variables are significant predictors of productivity in France: in this case both variables positively affect the prediction, with going from type 1 to type 2 of $mount_i$ increasing $\hat{Y_a}$ of 0.15 and moving away from the coast increasing $\hat{Y_a}$ of around 0.18. Thus, in the French model, agglomeration economies positively affect the predicted productiv-



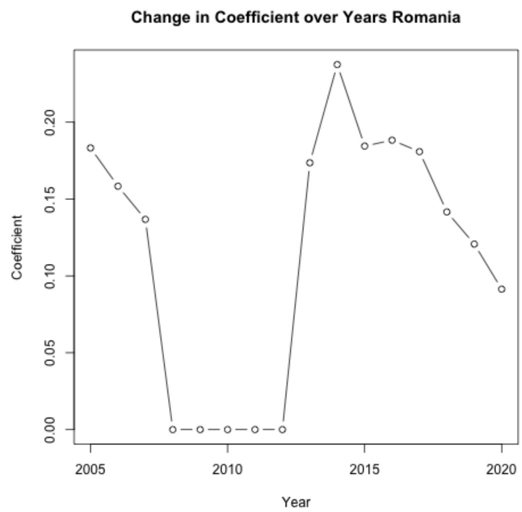

(a) model 2.a results Albania




(b) model 2.a results Greece




(c) model 2.a results France




(d) model 2.a results Romania

Figure 4.6: Model 2.a results for Albania, Greece, France and Romania

ity of around the same amount as $coast_i$ and $mount_i$ do. Overall, agglomeration effects in manufacturing productivity are limited and isolated, thus no meaningful and sound interpretation of results can be proposed.

Conversely, the results for model 2.b show generally much stronger significance of $log(rank_i)$ as a predictor of $Y_b$. The coefficient $(c - b)$ for Czech Republic, Denmark and Slovakia is not significantly different from zero, thus no agglomeration economies were detected through the 16 years horizon. Agglomeration economies in Albania are null until 2012, after which $log(rank_i)$ becomes a positive predictor of ATS productivity with $p < 0.05$. The opposite pattern is observed in Macedonia, where agglomeration economies are present with $p < 0.05$ until 2011 and then their predicting power decrease until becoming non-significant in 2016. Belgium, Bulgaria, Greece, Finland, France, Italy, Lithuania, the Netherlands, Portugal, Sweden, Slovenia, Hungary and Latvia, instead, all have agglomeration economies as significant predictors of ATS productivity, with $p < 0.05$ for Latvia, $p < 0.02$ for Hungary and $p < 0.01$ for all the others. The adjusted R-square tables can be found in B.3 and B.4.

The estimated coefficients by model 2.b are overall much higher than those estimated by the first model. Figure 4.7 shows the estimated agglomeration forces in ATS productivity in 2020: with the exception of Italy, which has an estimated coefficient of around 0.2, the rest of the reported countries have coefficients of at least 0.4. In comparison, the highest estimated agglomeration forces in total productivity in 2020 (Figure A.9) are Estonia, with 0.59 in 2016, and Lithuania, with 0.42 in 2020. Thus, the agglomeration effects in Advanced Tradable Services productivity is much stronger and relevant. Furthermore, provided that agglomeration economies occur when the coefficient $(c - b)$ is equal to 1, Lithuania, Latvia and Bulgaria reveal compelling evidence of agglomeration economies. Strong agglomeration economies can be seen in 2005 and 2012 as well, especially in Macedonia (Figures A.11 and A.12).

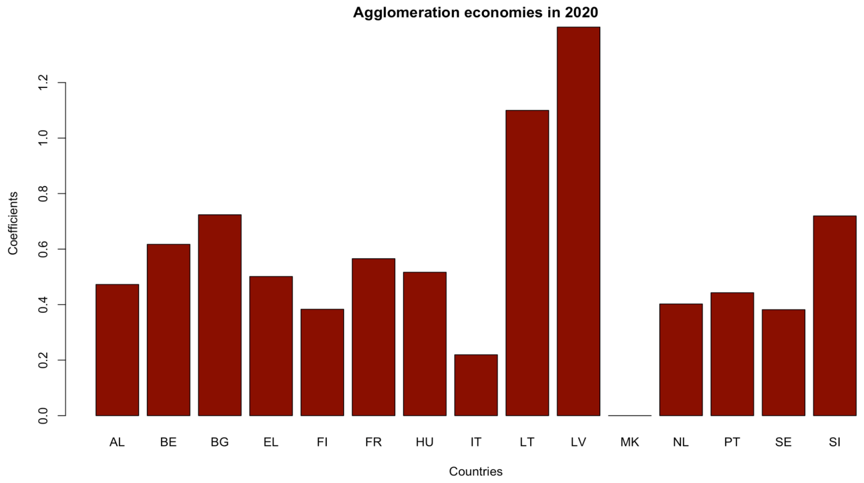



Figure 4.7: Agglomeration economies in 2020 (for Italy, 2019 is considered)

I plotted the change observed in agglomeration economies from 2005 to 2012 (Figure 4.8) and from 2012 to 2020 (Figure 4.9). Only countries where $log(rank_i)$ was found to be a significant predictor of productivity are reported and Macedonia was excluded in Figure 4.9 as the drop from significant to null coefficient biased the plot. For the same reason Albania is not reported in Figure 4.8. Again a possible outlier was detected in the 2020 results for Italy, thus the Italian coefficient's change was computed based on 2019 results.

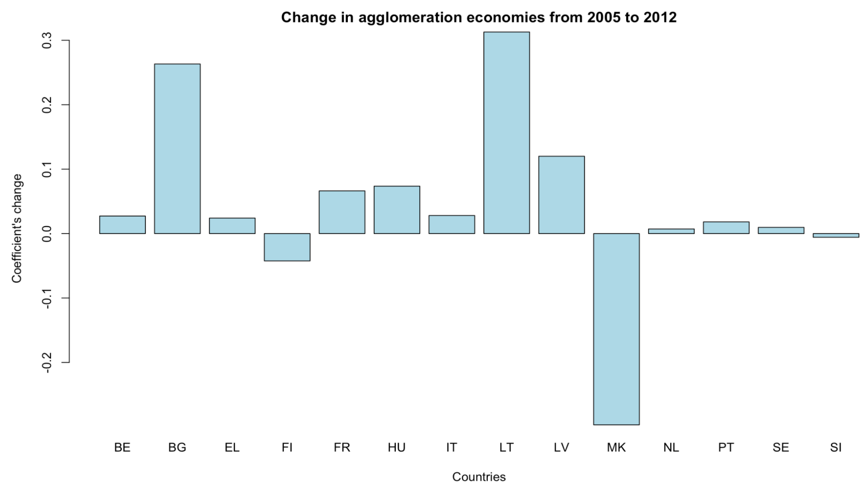



Figure 4.8: Change in agglomeration economies between 2005 and 2012

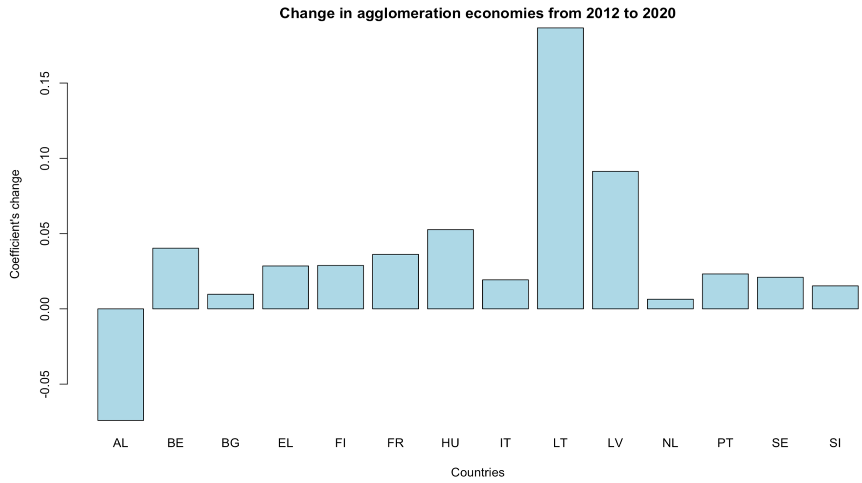



Figure 4.9: Change in agglomeration economies between 2012 and 2020

Change is consistently positive for all countries in which agglomeration economies were detected, with the exception of Macedonia, where significance of the predictor disappear after 2016. Albania also shows a slight decrease from 2012 to 2020 of around -0.05. Bulgaria, Latvia and Lithuania are the countries where agglomeration economies grew the most in the 2005-2020 period, although the biggest changes are concentrated in the years from 2005 to 2012 for all reported countries.

I plotted the coefficient for $log(rank_i)$ across time for the countries for which the regressor is a significant predictor. Albania, Belgium, Bulgaria, France, Lithuania, Latvia, the Netherlands, Portugal and Sweden show a growing trend (Figures A.13a, A.13b, A.13c, A.14b, A.14c, A.15a, A.15b, A.15d, A.16a, A.16b). For some of these countries, the growing trend appears throughout the time frame analyzed: this is the case for Belgium, France, Hungary, Lithuania, the Netherlands, Portugal and Sweden. Albania, Bulgaria and Latvia, instead, show a sudden jump and then the coefficient remains more stable.

Greece, Finland and Slovenia have more unstable graphs, with the estimated coefficient fluctuating much more: in these cases, while agglomeration economies might decrease in certain period, the overall change detected between 2005 and 2020 stays positive (Figures A.13d, A.14a, A.16c). Finally Macedonia is the only country with a decreasing trend, as the coefficient loses significance in 2016 (Figure A.15c). Italy has an irregular graph, opening up the possibility that 2020 data should be considered an outlier and excluded from analysis: regardless, agglomeration economies are the smallest among the group and, apart for 2020, show a rather stable pattern (Figure A.14d).

I now investigate whether the findings of Dijkstra and Kommission [20], Monastiriotis [50] and Roemish [59] can be supported by my analysis. The authors find evidence of different trends between Central and Eastern Europe, experiencing major growth in the capital city compared with the rest of the country, and Western and Northern European economies, showing internal inter-regional convergence or stability. Roemish [59] proposes the hypothesis that agglomeration economies increase more in early stages of development and then decrease at more advanced stages, creating a U-shaped pattern. To inquire this hypothesis, I plot the estimated coefficients of $log(rank_i)$ for all countries and ordered the results by ascending order. The countries for which the analysis resulted not significant have coefficient equal to zero. Given that agglomeration economies estimated by the first model are quite small, I only investigate the results of the model 2.b as interpretation would be more robust. The plots referring to the first model can be found in Appendix A ( A.17a, A.17b and A.17c). Starting from results referring to 2005, a few "exceptions" to Roemish's hypothesis can be seen, both at the bottom of the distribution - Czech Republic, Slovakia and Hungary are among the countries with lowest agglomeration economies - and at the top, with France and Belgium ranking between Bulgaria and Lithuania, although their coefficient is significantly lower than the top three. 2012 results seem to be slightly more aligned with Roemish's predictions: agglomeration economies in France and Belgium are no longer between the strongest in Europe, but some exceptions still exist at the bottom. A similar situation is found in the 2020 graph: France and Belgium still rank quite high compared to other EU15 countries and Macedonia and Slovakia find themselves at the bottom, with no agglomeration economies detected. Considering the few and regular "outliers" to the theory detected, my results seem to provide supporting evidence to Roemish's hypothesis, although the difference between agglomeration economies in Northern and Western Europe and CEECS is not very pronounced. It should be noted that, according to Oqubay and Lin [52], Advanced Tradable Services' importance in regions' economy is not homogeneous and ATS are usually more concentrated in wealthier areas. Therefore, countries where wealth distribution is more uneven between regions should show a concentration of ATS in the richest areas, thus accentuating signs of agglomeration. This is, indeed, what is observed in my analysis' results. Instead, no clear difference is detected between Northern and Western Europe and Southern European countries. Finally, it should be noted that while CEE countries seem to be more agglomerated than Northern and Western Europe in regards of ATS productivity, agglomeration economies in the EU15 countries do not seem to be decreasing, at least according to results from model 2.b.

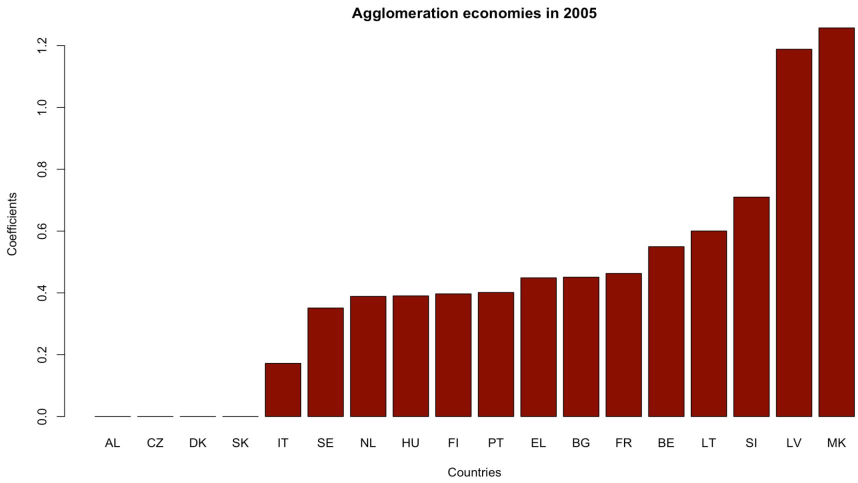



Figure 4.10: Agglomeration economies in 2005, ordered

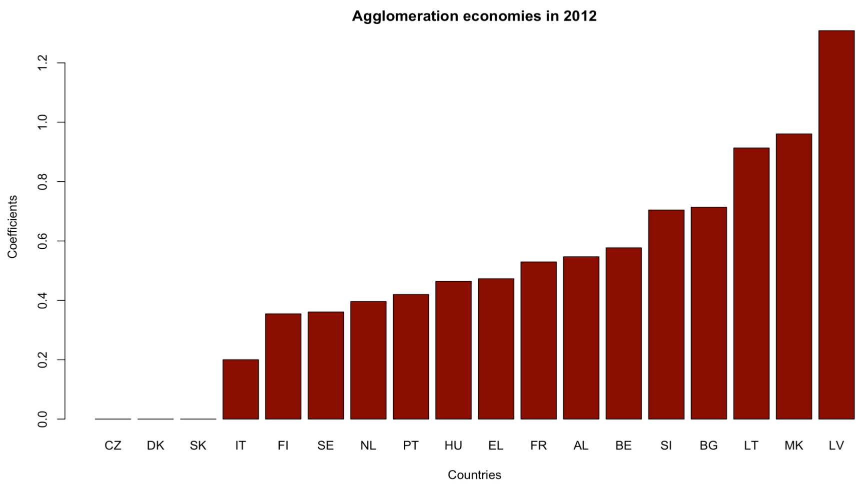



Figure 4.11: Agglomeration economies in 2012, ordered

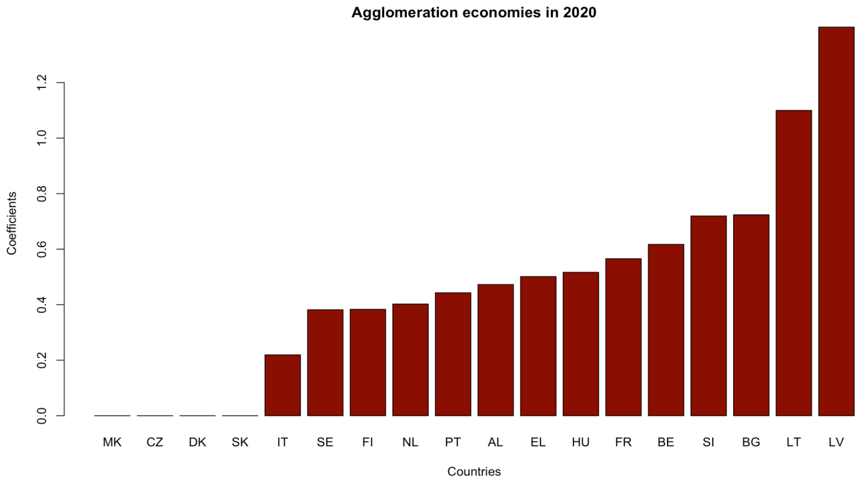



Figure 4.12: Agglomeration economies in 2020, ordered

# 5 | Results' Interpretation, Implications and Limitations

## 5.1. Discussion

The analysis gives markedly different results for model 2.a and 2.b. Manufacturing productivity in Europe is not as affected by agglomeration forces as Advanced Tradable Services productivity is. The only two countries where both analyses were significant, not including countries for which the 2.b model could not be implemented due to data scarcity, are France and Albania: in both cases, agglomeration economies in ATS productivity were stronger than those observed in manufacturing, although not substantially bigger. Thus, provided that we classify manufacturing activities as less knowledge- and technology-intensive than ATS activities, my analysis points to the latter as the main factor driving agglomeration in Europe. Instead, manufacturing seems to have a weaker role and in some instances it seems to defuse agglomeration forces, as observed in the first model's results. Furthermore, these findings could shed some light on which of the three main benefits of agglomeration economies is the most important actor in European agglomerations. As presented in Chapter 1, possible benefits that firms can reap by clustering together are often described as *sharing*, *matching* and *learning*. While all three contributors concern both manufacturing and ATS activities, *learning* has been shown to strongly affects technology- and knowledge-intensive sectors, where innovation plays a key role [11][10][41]. Such is the case for Advanced Tradable Services, which include Information and Communication (NACE Rev.2 J), Financial services and insurance (NACE Rev.2 K), Professional, scientifical and technical activities (NACE Rev.2 M), Administrative and support service activities (NACE Rev.2 N). Rosenthal and Strange [60] found that knowledge spillovers seemed to have effect only at zip-code level, rapidly attenuating across space. According to my analysis, this might not be the case in Europe, given that evidence of knowledge spillovers could be found at NUTS3 level: this could be due to better connectivity between cities and municipalities or greater geographical proximity compared to the U.S. It could also be a consequence of widespread and improved information technology, although the time frame does not fully catch the most recent uptake in the use of digital platforms, boosted by the Covid-19 Pandemic.

It should be noted that my results clearly and strongly detects agglomeration economies in Latvia and Lithuania, while the rest of the countries have estimated coefficients that are positive but smaller than 1: in these cases, the distribution of productivity is more uneven than when $(c-b)=0$ but the most populous NUTS3 is not $i$ times as productive as the $i$-est NUTS3. These findings are consistent with the European urban system, which leaves much greater role for smaller and medium sized cities, and support evidence for a polycentric system, where labor mobility and supply chains operates from multiple centres instead of a large monocentric structure like the one observed in the US. As stressed by Barca et al. [12], many highly productive cities' advantage is not their size, as they are often medium small, but their high connectivity.

## 5.2. Implications for EU policies

This thesis's results can have implications with regards to European policies. The 2009 World bank report and 2004 Sapir Report advocate for the advantages of agglomeration effects in large cities and predict uneven growth, suggesting that development policies focused on non-agglomerated places undermine growth and prosperity. The Sapir Report argues that EU cohesion policies should encourage productivity factors, such as workers, to move where they are most productive, meaning cities [12][64]. In the first decade of the Millennium, policies considered subsidies to unproductive places as a waste of resources and regarded "lagging behind" regions as areas with no economic potential [58]. Yet, it is not clear whether Europe should strive for greater agglomeration economies. As illustrated in Chapter 1, the benefits of agglomeration are strongly related to growth, innovation and economic development. However, there are some unseen negative externalities that should be taken into consideration. The left-behind regions, areas that have witnessed long period of economic decline and have seen better times, such as the peripheries of Metropolises, are found to breed anti-EU discontent [24]. Rodríguez-Pose [58] finds that, in Europe, persistent and low level of growth is at the root of political instability, and the same can be said for the US. Such evidence is often referred to as geography of discontent. Such feeling, as if there were no hope and no future, finds soothing in the populist ideology, which in Europe takes the form of anti-establishment movement, striving for de-globalization, and political resentment which reflects the urban-regional divide. This is well exemplified by Brexit, where large cities consistently voted to remain in the European Union while industrial declining and disadvantaged rural areas voted to leave. The same pattern can be found in France as well: the main cities - Lille, Metz, Nancy, Amiens, Reims - opposed the populist trend and supported Emmanuel Macron. New Economic Geography predicts negative externalities from agglomeration economies such as high land rent, congestion and pollution, but it also predicts the benefits to largely outweigh the disadvantages. The theory does not account for this unexpected source of externality: non agglomerated areas suffering distress from de-agglomeration.

The consequences of this are twofold. Firstly, the populist revolt emerging from this territorial divide will negatively affect the areas that are pushing it forward. Less economic stability and more inefficient governments would primarily affect lower-income regions and jobs for the medium- and low-skilled. Brexit will affect London's growth prospects in the same way it will negatively impact the economic perspectives of the North and North East of England. Secondly, this poses the risk of shaking the foundations of the European Union, calling into question its existence as a whole. Moreover, applying the 2009 World bank suggestions to the EU reality would mean abandoning most middle- and small-sized cities, which are the core of European growth. Enhancing agglomeration economies in Europe, regardless of the territorial inequalities that this entails, could exacerbate the current populist sentiment, as Tomaney and Pike [67] shows in the case of the United Kingdom. Yet cities drive national growth and are often hubs of innovation and national wealth, thus resisting agglomeration forces could slow-down national economy as a whole.

Indeed, results of traditional strategies are often underwhelming due to poor quality institutions and implementation of projects that promote individual interests. In developed countries, the main strategy has been transferring wealth from cities to rural realities and sustain employment through the public sector. In the past decade, a new approach emerged, called "place-based", which aims at maximizing development potential, without focusing on convergence [12]. Although, according to my analysis, most European countries are not affected by strong agglomeration effects, territorial inequalities and agglomeration are widely discussed topics in the public debate. In the current political and economic climate, low-income and low-density regions are often markedly opposed to highly productive and innovative areas. Even if results from economic analysis do not show such a stark contrast between metropolis and periphery, congestion costs and negative externalities deriving from agglomeration are becoming a burden for European citizens, as the housing crisis demonstrates. Perfect or almost perfect labour mobility is a general assumption in urban economics, which consider it preferable to move where opportunities are rather than to stay in a place where chances of finding a job are limited. According to theory, the main barrier to mobility is usually associated to be in-elastic housing supply in cities. In reality, people often decide not to move due to emotional attachment, age, insufficient qualifications and skills.

Therefore, relieving congestion costs, such as easing restrictions on housing in cities, does not solve the issue of perceived inequality and feeling left behind. Policies should simultaneously promote growth and innovation in highly productive areas, while taking into consideration the struggles faced by low-income regions. Such an approach is what Iammarino et al. [42] present: while very high- and high-income regions face risks related to knowledge spillovers spreading out innovation and routinizing high wage activities, middle- and low-income regions struggle due to limited skills and assets, high labor costs for low-skilled activities and under-productivity compared to rich regions. Different realities need different policies: very high- and high-income regions must keep pushing the edge of innovation and exploit the benefits of agglomeration economies. Middle income regions require re-skilling and institutional improvements to overcome the risk of middle-income trap. Low-income regions, instead, need to focus on increasing labor force participation, education quality, skill matching and university-industry linkages. Thus, Europe's best option is fostering the diversity and uniqueness of its urban structure, which has allowed economic development to translate into extremely high quality of life.

## 5.3. Limitations

The analysis carried out in this thesis presents some limitations. First, agglomeration economies as presented in Chapter 1 are always understood in the context of cities, at a municipality level. The model in this thesis was implemented on data at NUTS3 level, which are more coarse-grained. As discussed in Section 2.1.1, data at a more detailed level is very difficult to collect for many European countries and data referring to FUA is not sufficiently available yet. Data at municipality level, aside for availability issues, also poses the issue of comparability among European countries, as each Member State individually legislates the definition of municipality. This issue is also present in NUTS3, as seen in the size difference between German NUTS3 and the rest (Section 2.2), but it's is a well-established system, tried and tested in many other studies, and allows for extensive analysis due to abundance of data. Nevertheless, the validity of the results may be reduced because of this and it should be taken into consideration when evaluating the results. Possibly, this limitation could have been tested by examining the results from the model implementation in Germany, as German NUTS3 classification distinguishes between cities and their peripheries (Figure 2.8). Unfortunately, missing data on Advanced Tradable Services GVA in Germany prevented the implementation of model 2.b. Nevertheless, the first model could be implemented and it showed very weak agglomeration economies, similar to Italy. Thus, model 2.b could have confirmed this, as in the Italian case, therefore signaling absence of agglomeration economies, or contradict the previous finding. Regardless, the use of NUTS3 data is an important limitation to the analysis' validity and interpretation. Detecting agglomeration economies could be interpreted as evidence of the presence of the phenomenon operating at regional level, thus implicating that agglomeration externalities are bigger than previously understood. At the same time, the models may not be able to catch the severity of activity concentration within the richest regions.

Another possible limitation could be found in the dependent variable used for productivity: Roemish [59] finds that agglomeration economies tend to be stronger in terms of GVA than in terms of employment, thus my results' relevance could decrease if labour productivity was used. The choice of GVA productivity as Y variable was due to data availability, especially regarding distinction in NACE Rev.2 sectors. The analysis could be expanded by considering, instead, labour productivity and evaluating the differences in findings.

Finally, the control variables included in the model may not be sufficient to take into account regional differences. The models implemented in this study were intentionally simple to allow straightforward and solid interpretation of results. They could be expanded by including a variable that takes into consideration the distance of the NUTS3 from the capital city. The categorical variable $Urbantype_i$, presented in chapter 2, could also be considered as a potential regressor. It was not used in this study as its inclusion would pose a risk of endogeneity.

# 6 | Conclusions

This thesis has provided a comprehensive analysis of agglomeration economies across different regions in Europe, from 2005 to 2020, focusing, in particular, on the years 2005, 2012, and 2020. The main research question regarding how agglomeration economies in European countries change within this time frame was addressed through the use of a dataset defined at the NUTS3 geographical level and sectors' classification based on the OECD's NACE Rev 2. The results indicate that agglomeration economies are present in certain countries, with notable changes observed between years and between sectors. In terms of the industries driving agglomeration economies in European countries, I found that Advanced Tradable Services play a significant role. This finding is aligned with the literature reviewed and offers a new insight into the composition of agglomeration forces, suggesting that knowledge- and technology-intensive industries can benefit much more than others from agglomeration economies due to factors such as knowledge spillovers.

However, the research process was not without its limitations. The use of the NUTS3 geographical level and the OECD's NACE Rev 2 sectors' classification may not capture all the nuances of agglomeration economies. Furthermore, the computational models used in the analysis may not fully account for other regional differences that could influence productivity, such distance from the capital city. Despite its limitations, this work has contributed with new knowledge to the field by providing a detailed analysis of agglomeration economies in Europe over a 15-year period, differentiating between manufacturing and ATS productivity. The findings of this study could be used to inform economic policies and strategies aimed at enhancing productivity in different regions.

For future work, it could be beneficial to expand the dataset to investigate the matter on a finer-grained level such as Functional Urban Areas. This would allow for a more comprehensive understanding of agglomeration economies. Additionally, exploring other sectors' classifications and implementing the model using employment productivity could provide further insights into the dynamics of agglomeration economies. In conclusion, this thesis has shed light on the concept of agglomeration economies in Europe, providing valuable insights that could inform future research and policy-making in this area.

# Bibliography

[1] Indicator: Gross value added: NACE activity: C | ESPON Database Portal, . URL https://database.espon.eu/indicator/2896/.

[2] Indicator: Gross value added: NACE activity: J | ESPON Database Portal, . URL https://database.espon.eu/indicator/2902/.

[3] Indicator: Gross value added: NACE activity: K | ESPON Database Portal, . URL https://database.espon.eu/indicator/2903/.

[4] Indicator: Gross value added: NACE activities: M-N | ESPON Database Portal, . URL https://database.espon.eu/indicator/2899/.

[5] Indicator: Gross Value Added (GVA) at basic prices | ESPON Database Portal, . URL https://database.espon.eu/indicator/319/.

[6] NUTS - GISCO - Eurostat. URL https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts.

[7] Indicator: Population (total) on 1st January | ESPON Database Portal. URL https://database.espon.eu/indicator/3047/.

[8] H. Abdel-Rahman and M. Fujita. Product Variety, Marshallian Externalities, and City Sizes. *Journal of Regional Science*, 30:165–183, July 2006. doi: 10.1111/j.1467-9787.1990.tb00091.x.

[9] D. Acemoglu and J. A. Robinson. The Political Economy of the Kuznets Curve. *Review of Development Economics*, 6(2):183–203, June 2002. ISSN 1363-6669, 1467-9361. doi: 10.1111/1467-9361.00149. URL https://onlinelibrary.wiley.com/doi/10.1111/1467-9361.00149.

[10] J. Y. Angela. Agglomeration Effects in the Diffusion of Advanced Manufacturing Technologies. 2003. URL https://www.semanticscholar.org/paper/Agglomeration-Effects-in-the-Diffusion-of-Advanced-Angela/21ca5b81d35049214562b368e9a0074da4696c01.

[11] D. Audretsch and M. P. Feldman. R&D Spillovers and the Geography of Innovation and Production. *American Economic Review*, 86(3):630–40, 1996. URL `https://econpapers.repec.org/article/aeaaecrev/v_3a86_3ay_3a1996_3ai_3a3_3ap_3a630-40.htm`. Publisher: American Economic Association.

[12] F. Barca, P. McCann, and A. Rodríguez-Pose. The case for regional development intervention: place-based versus place-neutral approaches. *Journal of Regional Science*, 52(1):134–152, Feb. 2012. ISSN 00224146. doi: 10.1111/j.1467-9787.2011.00756.x. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9787.2011.00756.x`.

[13] F. Batista, L. Dijkstra, P. Martinez, and C. Lavalle. *Regionalisation of demographic and economic projections: trend and convergence scenarios from 2015 to 2060.* European Commission. Joint Research Centre., LU, 2016. URL `https://data.europa.eu/doi/10.2788/458769`.

[14] K. Bolter and J. Robey. Agglomeration Economies: A Literature Review. *Reports*, Sept. 2020. URL `https://research.upjohn.org/reports/252`.

[15] H. Breinlich. The spatial income structure in the European Union—what role for Economic Geography? *Journal of Economic Geography*, 6(5):593–617, Nov. 2006. ISSN 1468-2710, 1468-2702. doi: 10.1093/jeg/lbl018. URL `http://academic.oup.com/joeg/article/6/5/593/904506/The-spatial-income-structure-in-the-European`.

[16] P.-A. Buigues, F. Ilzkovitz, and J.-F. Lebrun. *The Impact of the Internal Market by Industrial Sector: The Challenge for the Member States*. Office for Official Publications of the European Communities, 1990. ISBN 978-92-826-1818-9. Google-Books-ID: lLBBAQAAIAAJ.

[17] A. Ciccone. Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227, 2002. URL `https://EconPapers.repec.org/RePEc:eee:eecrev:v:46:y:2002:i:2:p:213-227`.

[18] A. Ciccone and R. Hall. Productivity and the Density of Economic Activity. Technical Report w4313, National Bureau of Economic Research, Cambridge, MA, Apr. 1993. URL `http://www.nber.org/papers/w4313.pdf`.

[19] B. De-Miguel-Molina, J.-L. Hervas-Oliver, R. Boix, and M. De-Miguel-Molina. The Importance of Creative Industry Agglomerations in Explaining the Wealth of European Regions. *European Planning Studies*, 20(8):1263–1280, Aug. 2012. ISSN 0965-4313. doi: 10.1080/09654313.2012.680579. URL `https:`

//doi.org/10.1080/09654313.2012.680579. Publisher: Routledge _eprint: https://doi.org/10.1080/09654313.2012.680579.

[20] L. Dijkstra and E. Kommission, editors. *Investment for jobs and growth: promoting development and good governance in EU regions and cities*. Number 6.2014 in Report on economic, social and territorial cohesion. Publ. Off. of the Europ. Union, Luxembourg, 2014. ISBN 978-92-79-35489-2.

[21] L. Dijkstra, E. Garcilazo, and P. McCann. The Economic Performance of European Cities and City Regions: Myths and Realities. *European Planning Studies*, 21(3):334–354, Mar. 2013. ISSN 0965-4313, 1469-5944. doi: 10.1080/09654313.2012.716245. URL http://www.tandfonline.com/doi/abs/10.1080/09654313.2012.716245.

[22] L. Dijkstra, E. Garcilazo, and P. McCann. The effects of the global financial crisis on European regions and cities. *Journal of Economic Geography*, 15(5):935–949, Sept. 2015. ISSN 1468-2702, 1468-2710. doi: 10.1093/jeg/lbv032. URL https://academic.oup.com/joeg/article-lookup/doi/10.1093/jeg/lbv032.

[23] L. Dijkstra, H. Poelman, and P. Veneri. The EU-OECD definition of a functional urban area. OECD Regional Development Working Papers 2019/11, Dec. 2019. URL https://www.oecd-ilibrary.org/urban-rural-and-regional-development/the-eu-oecd-definition-of-a-functional-urban-area_d58cb34d-en. Series: OECD Regional Development Working Papers Volume: 2019/11.

[24] L. Dijkstra, H. Poelman, and A. Rodríguez-Pose. The geography of EU discontent. *Regional Studies*, 54(6):737–753, June 2020. ISSN 0034-3404, 1360-0591. doi: 10.1080/00343404.2019.1654603. URL https://www.tandfonline.com/doi/full/10.1080/00343404.2019.1654603.

[25] G. Duranton, editor. *Handbook of regional and urban economics. Vol. 5A / ed. by Gilles Duranton.* Number 7 in Handbooks in economics. Elsevier, North-Holland, Amsterdam Heidelberg, 2015. ISBN 978-0-444-59517-1.

[26] G. Duranton and D. Puga. Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle Of Products. *American Economic Review*, 91:1454–1477, Dec. 2001. doi: 10.1257/aer.91.5.1454.

[27] G. Duranton and D. Puga. Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2063–2117, 2004. URL https://ideas.repec.org//h/eee/regchp/4-48.html. Publisher: Elsevier.

[28] G. Ellison and E. L. Glaeser. Geographic Concentration in U.S. Manufacturing Indus-

tries: A Dartboard Approach. *Journal of Political Economy*, 105(5):889–927, 1997. ISSN 0022-3808. doi: 10.1086/262098. URL `https://www.jstor.org/stable/10.1086/262098`. Publisher: The University of Chicago Press.

[29] G. Ellison and E. L. Glaeser. The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review*, 89(2): 311–316, May 1999. ISSN 0002-8282. doi: 10.1257/aer.89.2.311. URL `https://pubs.aeaweb.org/doi/10.1257/aer.89.2.311`.

[30] G. Ellison, E. L. Glaeser, and W. R. Kerr. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(3):1195–1213, June 2010. ISSN 0002-8282. doi: 10.1257/aer.100.3.1195. URL `https://www.aeaweb.org/articles?id=10.1257/aer.100.3.1195`.

[31] T. Farole, S. Goga, and M. Ionescu-Heroiu. Rethinking lagging regions - using Cohesion Policy to deliver on the potential of Europe's regions. *World Bank*, 2018.

[32] M. Fujita and P. Krugman. When is the economy monocentric?: von Thunen and Chamberlin unified. *Regional Science and Urban Economics*, 25(4):505–528, 1995. URL `https://ideas.repec.org//a/eee/regeco/v25y1995i4p505-528.html`. Publisher: Elsevier.

[33] M. Fujita and P. Krugman. The new economic geography: Past, present and the future. *Papers in Regional Science*, 83(1):139–164, Oct. 2003. ISSN 1056-8190, 1435-5957. doi: 10.1007/s10110-003-0180-0. URL `http://doi.wiley.com/10.1007/s10110-003-0180-0`.

[34] M. Fujita and T. Mori. Structural stability and evolution of urban systems. *Regional Science and Urban Economics*, 27(4):399–442, Aug. 1997. ISSN 0166-0462. doi: 10.1016/S0166-0462(97)80004-X. URL `https://www.sciencedirect.com/science/article/pii/S016604629780004X`.

[35] M. Fujita, P. Krugman, and T. Mori. On the evolution of hierarchical urban systems1. *European Economic Review*, 43(2):209–251, 1999. ISSN 0014-2921. URL `https://econpapers.repec.org/article/eeeeecrev/v_3a43_3ay_3a1999_3ai_3a2_3ap_3a209-251.htm`. Publisher: Elsevier.

[36] L. Gan and Q. Li. Efficiency of Thin and Thick Markets. *NBER Working Papers*, Oct. 2004. URL `https://ideas.repec.org//p/nbr/nberwo/10815.html`. Number: 10815 Publisher: National Bureau of Economic Research, Inc.

[37] E. Glaeser and D. Maré. Cities and Skills. *Journal of Labor Economics*, 19(2):316–

342, Apr. 2001. ISSN 0734-306X, 1537-5307. doi: 10.1086/319563. URL https://www.journals.uchicago.edu/doi/10.1086/319563.

[38] E. L. Glaeser. Are Cities Dying? *Journal of Economic Perspectives*, 12(2):139–160, June 1998. ISSN 0895-3309. doi: 10.1257/jep.12.2.139. URL https://www.aeaweb.org/articles?id=10.1257/jep.12.2.139.

[39] A. Gobin, R. Jones, M. Kirkby, P. Campling, G. Govers, C. Kosmas, and A. R. Gentile. Indicators for pan-European assessment and monitoring of soil erosion by water. *Environmental Science & Policy*, 7(1):25–38, Feb. 2004. ISSN 1462-9011. doi: 10.1016/j.envsci.2003.09.004. URL https://www.sciencedirect.com/science/article/pii/S1462901103001369.

[40] J. I. Haaland, H. J. Kind, K. H. M. Knarvik, and J. Torstensson. What determines the economic geography of europe? 1999.

[41] R. Henderson, A. Jaffe, and M. Trajtenberg. Geographic Localization of Knowledge Spillovers As Evidenced By Patent Citations. *The Quarterly Journal of Economics*, 108:577–98, Feb. 1993. doi: 10.2307/2118401.

[42] S. Iammarino, A. Rodriguez-Pose, and M. Storper. Regional inequality in Europe: evidence, theory and policy implications. *Journal of Economic Geography*, 19(2): 273–298, Mar. 2019. ISSN 1468-2702, 1468-2710. doi: 10.1093/jeg/lby021. URL https://academic.oup.com/joeg/article/19/2/273/4989323.

[43] M. Kahn and D. Costa. Power Couples: Changes In The Locational Choice Of The College Educated, 1940-1990. *The Quarterly Journal of Economics*, 115:1287–1315, Feb. 2000. doi: 10.1162/003355300555079.

[44] P. Krugman. Increasing Returns and Economic Geography. *JOURNAL OF POLITICAL ECONOMY*, 1999.

[45] P. Krugman and A. J. Venables. Globalization and the Inequality of Nations. Technical report, Centre for Economic Performance, LSE, 1995. URL https://cep.lse.ac.uk/_new/publications/abstract.asp?index=7206.

[46] E. Marrocu, R. Paci, and S. Usai. Productivity Growth in the Old and New Europe: The Role of Agglomeration Externalities. *Journal of Regional Science*, 53(3):418–442, 2013. ISSN 1467-9787. doi: 10.1111/jors.12000. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jors.12000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jors.12000.

[47] T. Mathä. European Integration and Geographical Concentration of Swedish Multinationals. 1999.

[48] F. McDonald and G. Vertova. Geographical concentration and competitiveness in the European Union. *European Business Review*, 13(3):157–165, June 2001. ISSN 0955-534X. doi: 10.1108/09555340110391275. URL `https://www.emerald.com/insight/content/doi/10.1108/09555340110391275/full/html`.

[49] M. J. Melitz and G. I. P. Ottaviano. Market Size, Trade, and Productivity. *Review of Economic Studies*, 75(1):295–316, Jan. 2008. ISSN 0034-6527, 1467-937X. doi: 10.1111/j.1467-937X.2007.00463.x. URL `https://academic.oup.com/restud/article-lookup/doi/10.1111/j.1467-937X.2007.00463.x`.

[50] V. Monastiriotis. Regional Growth and National Development: Transition in Central and Eastern Europe and the Regional Kuznets Curve in the East and the West. *Spatial Economic Analysis*, 9(2):142–161, Apr. 2014. ISSN 1742-1772. doi: 10.1080/17421772.2014.891156. URL `https://doi.org/10.1080/17421772.2014.891156`. Publisher: Routledge _eprint: https://doi.org/10.1080/17421772.2014.891156.

[51] OECD. The Metropolitan Century: Understanding Urbanisation and its Consequences | READ online, 2015. URL `https://read.oecd-ilibrary.org/urban-rural-and-regional-development/the-metropolitan-century_9789264228733-en`.

[52] A. Oqubay and J. Y. Lin, editors. *The Oxford Handbook of Industrial Hubs and Economic Development*. Oxford University Press, 1 edition, July 2020. ISBN 978-0-19-885043-4 978-0-19-188551-8. doi: 10.1093/oxfordhb/9780198850434.001.0001. URL `https://academic.oup.com/edited-volume/40436`.

[53] H. G. Overman and D. Puga. Labour pooling as a source of agglomeration: An empirical investigation. 2010.

[54] A. Pina and P. Sicari. Enhancing regional convergence in the European Union. OECD Economics Department Working Papers 1696, OECD Economics Department Working Papers, Dec. 2021. URL `https://www.oecd-ilibrary.org/economics/enhancing-regional-convergence-in-the-european-union_253dd6ee-en`. Series: OECD Economics Department Working Papers Volume: 1696.

[55] C. Postoiu. Regional Growth Patterns in the European Union. *Procedia Economics and Finance*, 30:656–663, 2015. ISSN 22125671. doi: 10.1016/S2212-5671(15)01284-8. URL `https://linkinghub.elsevier.com/retrieve/pii/S2212567115012848`.

[56] D. Power and T. Nielsén. Priority Sector Report: Creative and Cultural Industries. 2010.

[57] D. Puga. The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1):203–219, Feb. 2010. ISSN 00224146, 14679787. doi: 10.1111/j.1467-9787.2009.00657.x. URL `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9787.2009.00657.x`.

[58] A. Rodríguez-Pose. 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, Mar. 2018. ISSN 1752-1378, 1752-1386. doi: 10.1093/cjres/rsx024. URL `https://academic.oup.com/cjres/article/11/1/189/4821289`.

[59] R. Roemish. Estimating agglomeration in the EU and the Western Balkan regions. *The Vienna Institute for International Economic Studies*, wiiw Balkan Observatory Working Papers(117), 2015.

[60] S. S. Rosenthal and W. C. Strange. The Determinants of Agglomeration. *Journal of Urban Economics*, 50(2):191–229, Sept. 2001. ISSN 00941190. doi: 10.1006/juec.2001.2230. URL `https://linkinghub.elsevier.com/retrieve/pii/S0094119001922302`.

[61] S. S. Rosenthal and W. C. Strange. Chapter 49 Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2119–2171. Elsevier, 2004. ISBN 978-0-444-50967-3. doi: 10.1016/S1574-0080(04)80006-3. URL `https://linkinghub.elsevier.com/retrieve/pii/S1574008004800063`.

[62] S. S. Rosenthal and W. C. Strange. How Close Is Close? The Spatial Reach of Agglomeration Economies. *Journal of Economic Perspectives*, 34(3):27–49, Aug. 2020. ISSN 0895-3309. doi: 10.1257/jep.34.3.27. URL `https://pubs.aeaweb.org/doi/10.1257/jep.34.3.27`.

[63] X. X. Sala-i Martin. Regional cohesion: evidence and theories of regional growth and convergence. *European Economic Review*, 40(6):1325–1352, June 1996. ISSN 00142921. doi: 10.1016/0014-2921(95)00029-1. URL `https://linkinghub.elsevier.com/retrieve/pii/0014292195000291`.

[64] A. Sapir, editor. *An agenda for a growing Europe: the Sapir report*. Oxford Univ. Press, Oxford, 1. publ edition, 2004. ISBN 978-0-19-927148-1 978-0-19-927149-8.

[65] E. Sorhun. Is Economic Integration a Historical Shock to City-size Distribution?

*Comparative Economic Research. Central and Eastern Europe*, 21(1):83–100, Apr. 2018. ISSN 2082-6737, 1508-2008. doi: 10.2478/cer-2018-0005. URL https://czasopisma.uni.lodz.pl/CER/article/view/2971.

[66] D. Starrett. Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1):21–37, Feb. 1978. ISSN 00220531. doi: 10.1016/0022-0531(78)90120-5. URL https://linkinghub.elsevier.com/retrieve/pii/0022053178901205.

[67] J. Tomaney and A. Pike. Brexit, devolution and economic development in 'left-behind' regions. *Welsh Economic Review*, 26(0):29, Dec. 2018. ISSN 2397-8716. doi: 10.18573/wer.231. URL https://wer.cardiffuniversitypress.org/article/10.18573/wer.231/.

[68] D. Völlmecke, B. Jindra, and P. Marek. FDI, human capital and income convergence—Evidence for European regions. *Economic Systems*, 40(2):288–307, June 2016. ISSN 09393625. doi: 10.1016/j.ecosys.2015.11.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0939362516300140.

# A | Appendix A

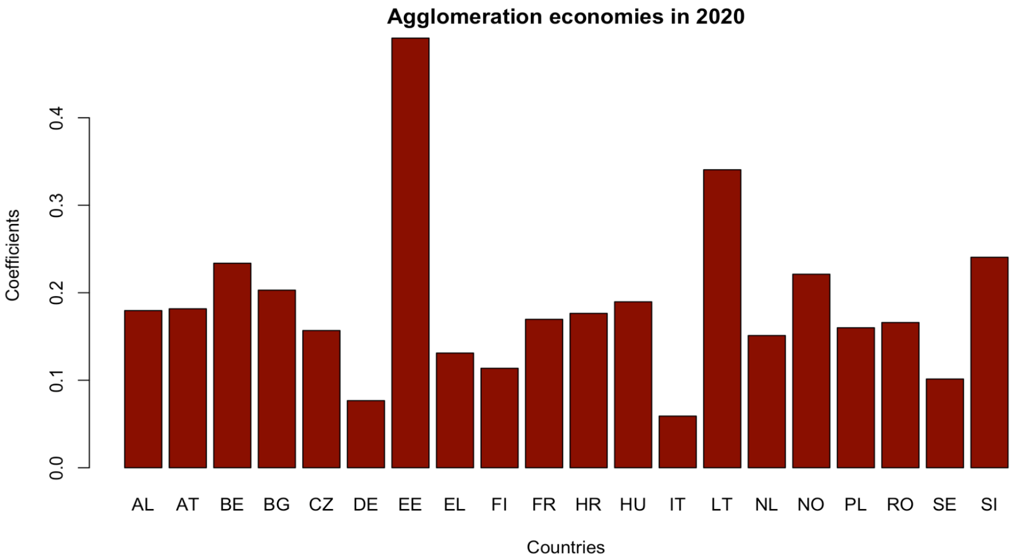



Figure A.8: Agglomeration economies in 2005

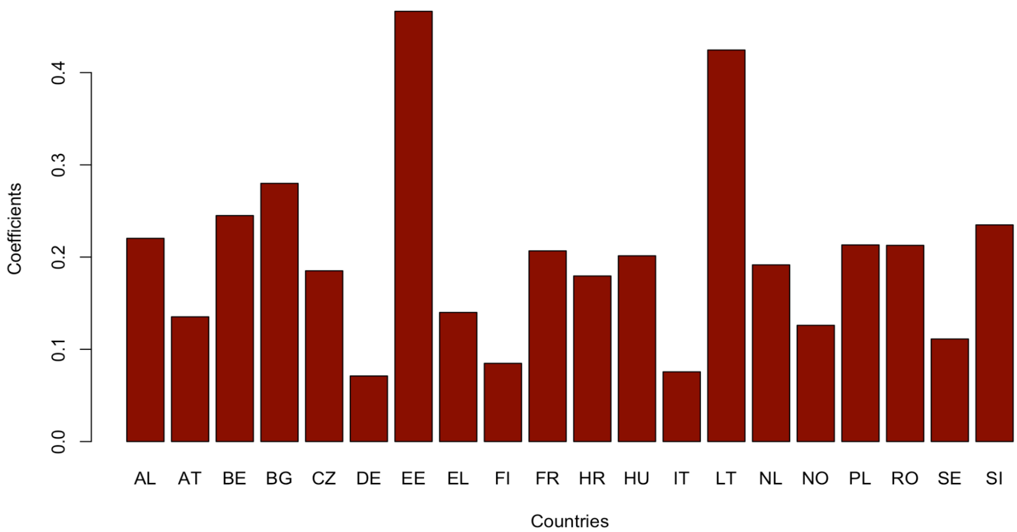



Figure A.9: Agglomeration economies in 2020 (for Italy, 2019 is considered)

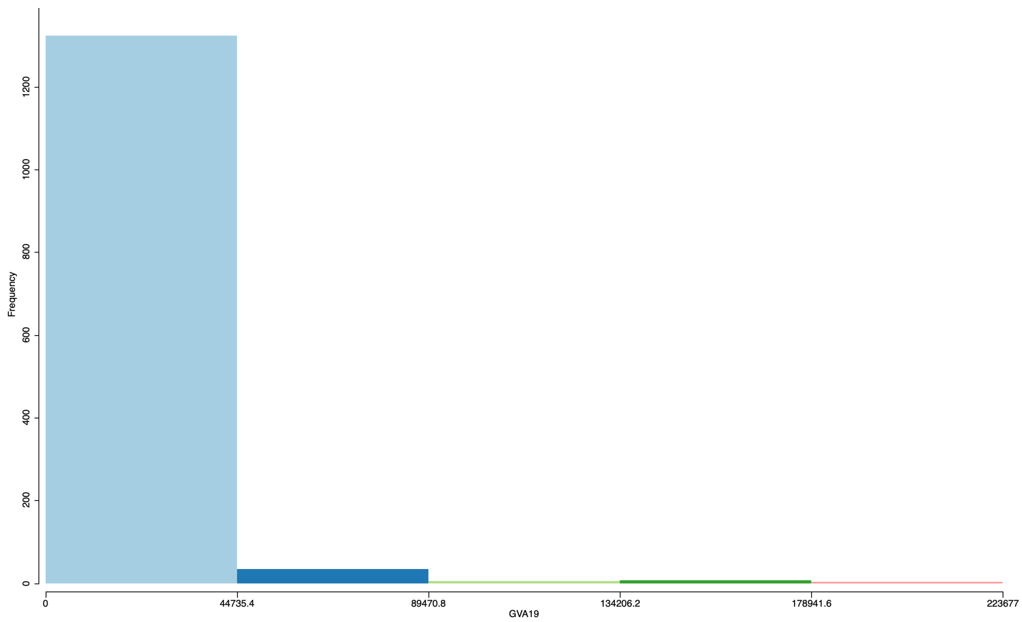
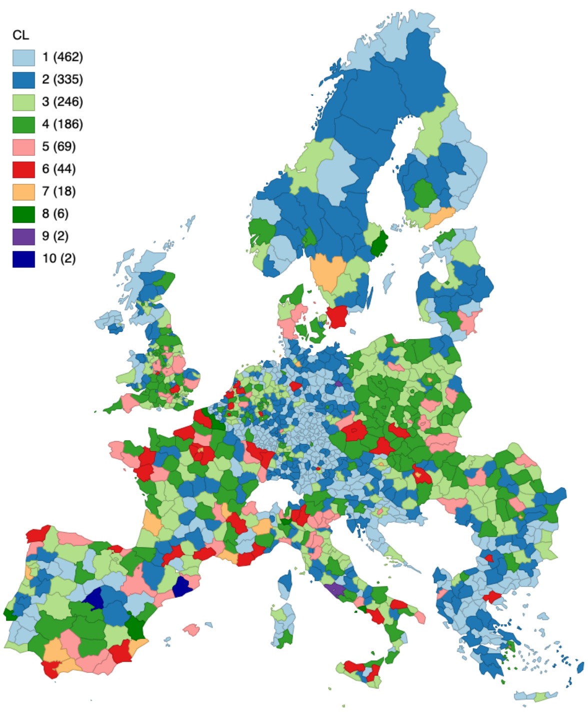
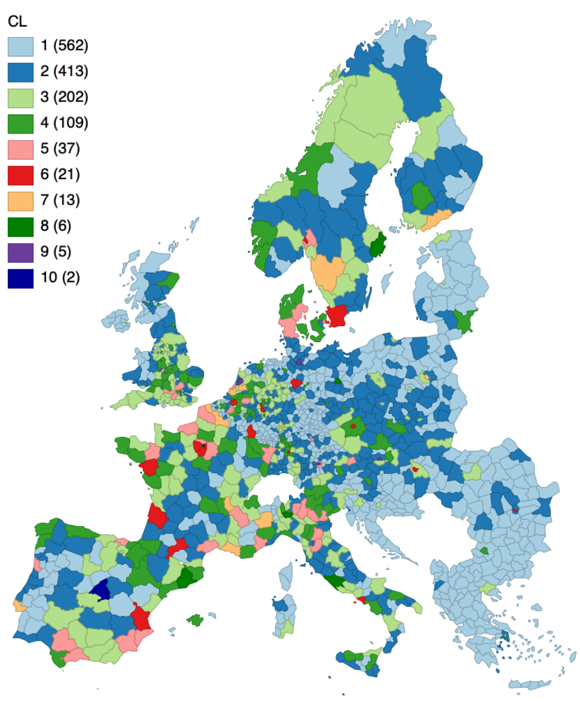



Figure A.1: Gross Value Added in 2019 at NUTS3 level

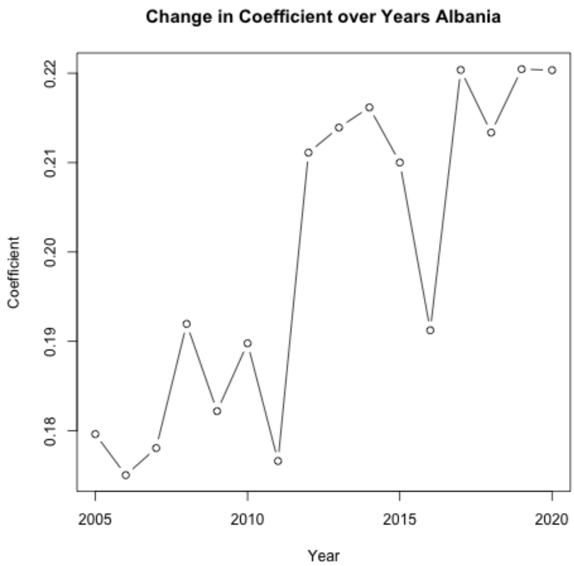



(a) K-means clustering of population in 2019

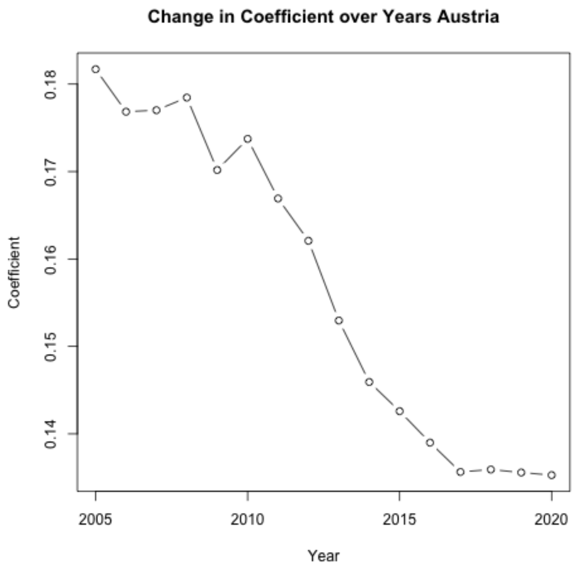



(b) K-means clustering of GVA in 2019

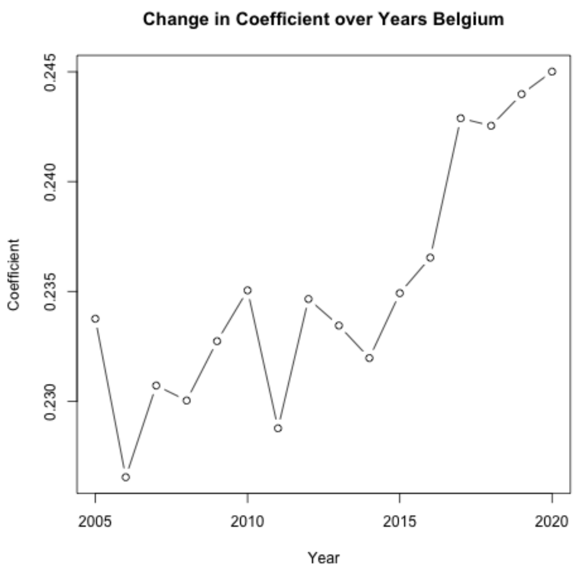
Figure A.2: K-means clustering output



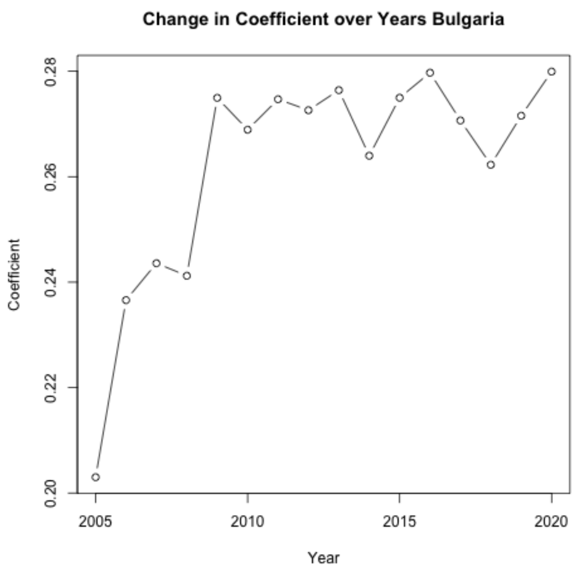

(a) model 1 results Albania




(b) model 1 results Austria

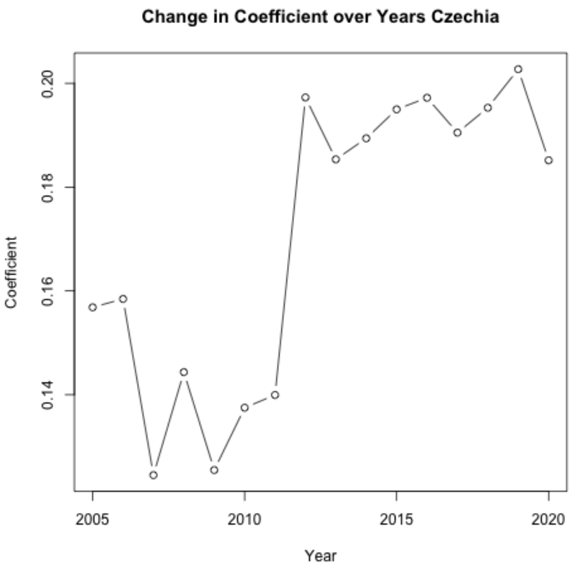



(c) model 1 results Belgium

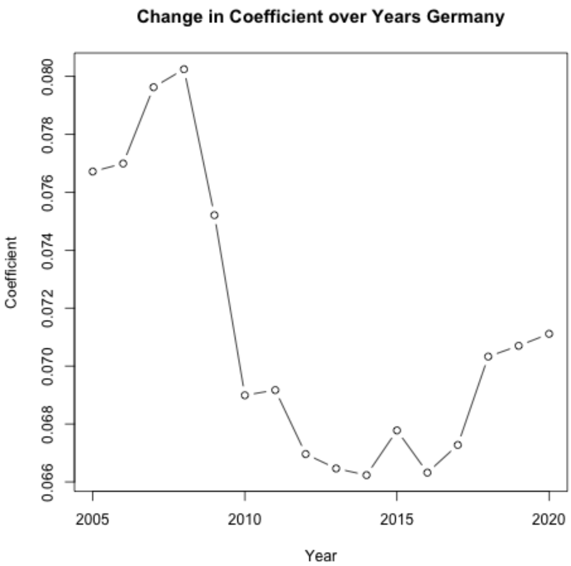



(d) model 1 results Bulgaria

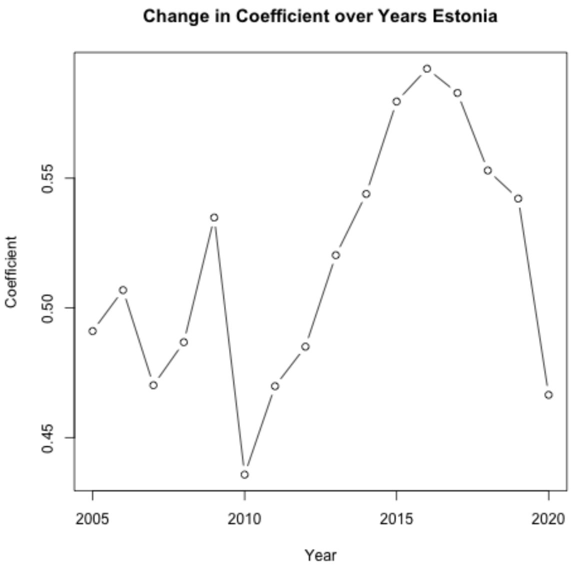
Figure A.3: Model 1 results for Albania, Austria, Belgium and Bulgaria



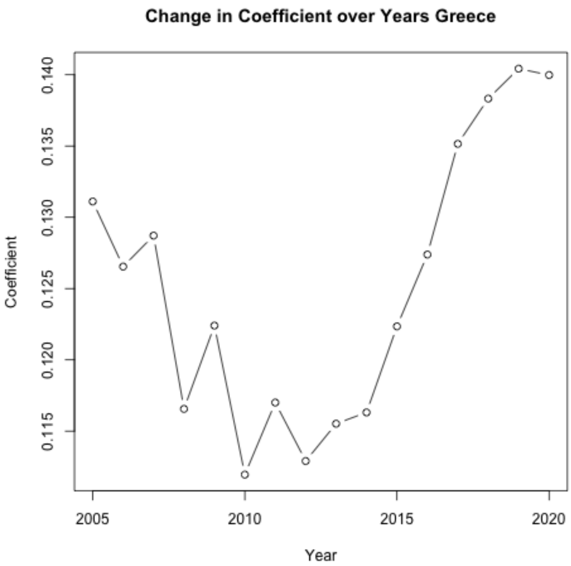

(a) model 1 results Czech Republic




(b) model 1 results Germany

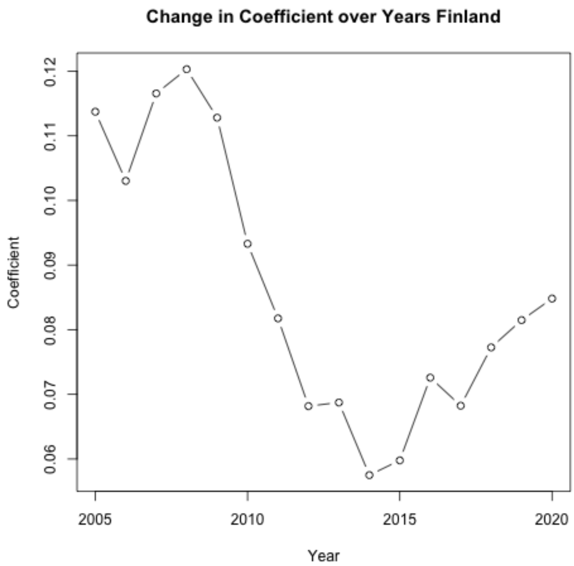



(c) model 1 results Estonia

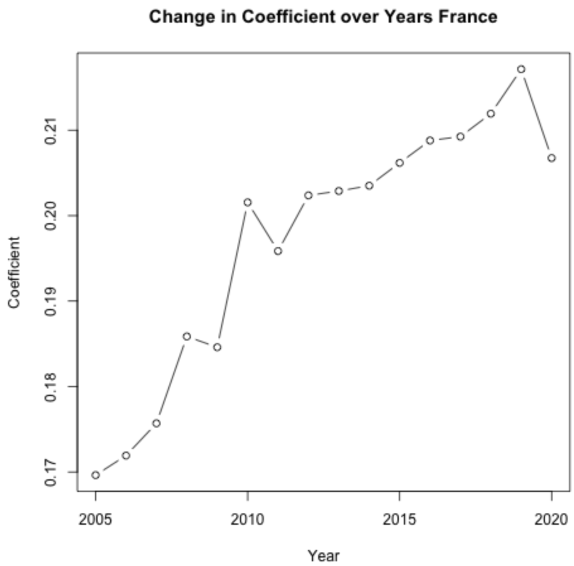



(d) model 1 results Greece

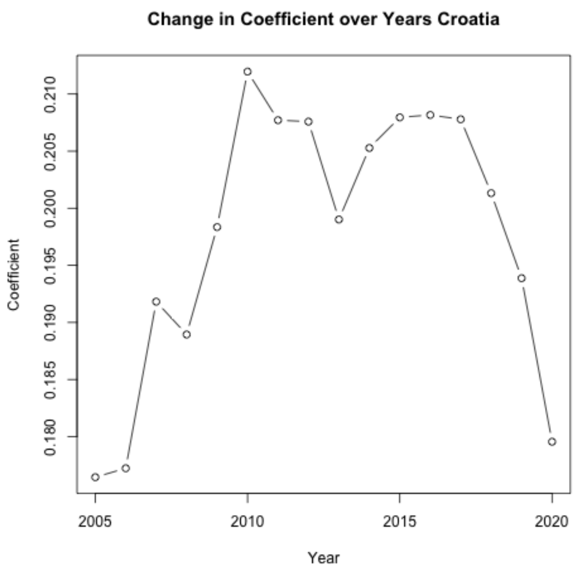
Figure A.4: Model 1 results for Czech Republic, Germany, Estonia and Greece



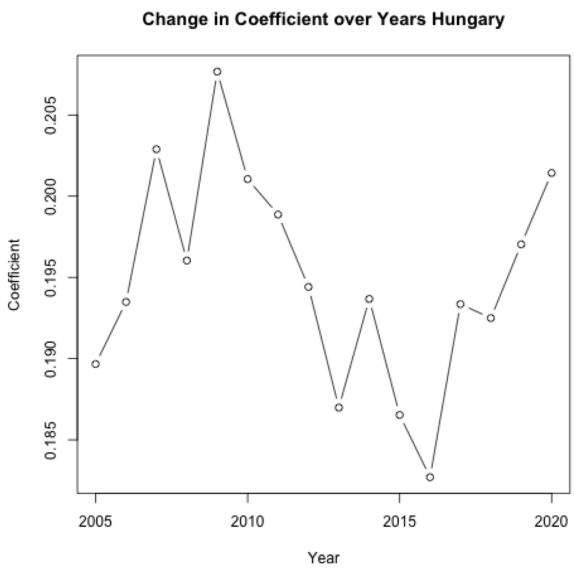

(a) model 1 results Finland




(b) model 1 results France

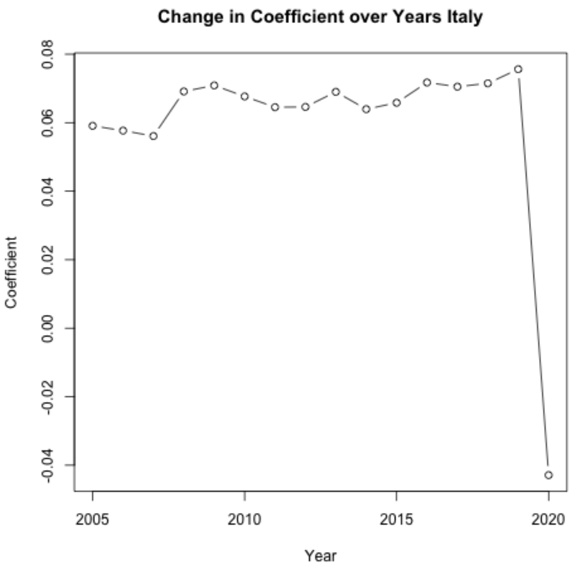



(c) model 1 results Croatia

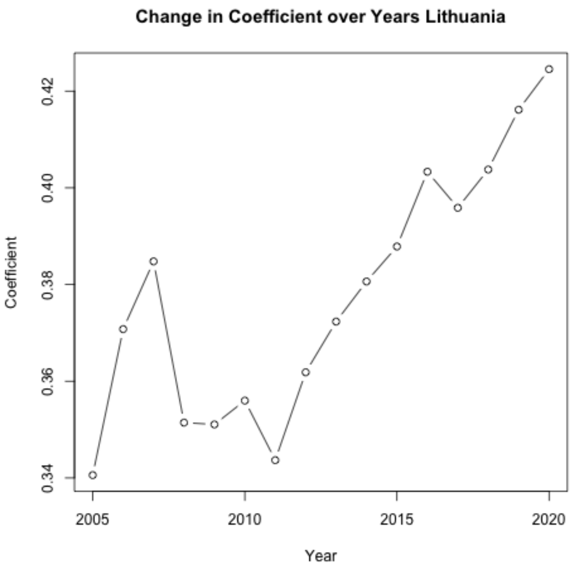



(d) model 1 results Hungary

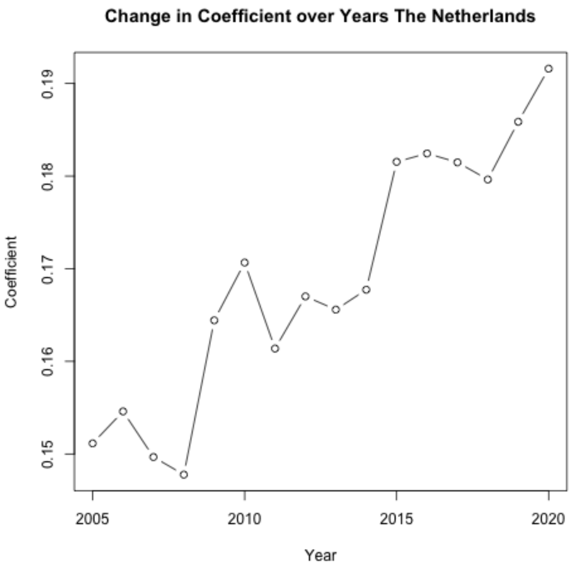
Figure A.5: Model 1 results for Finland, France, Croatia and Hungary



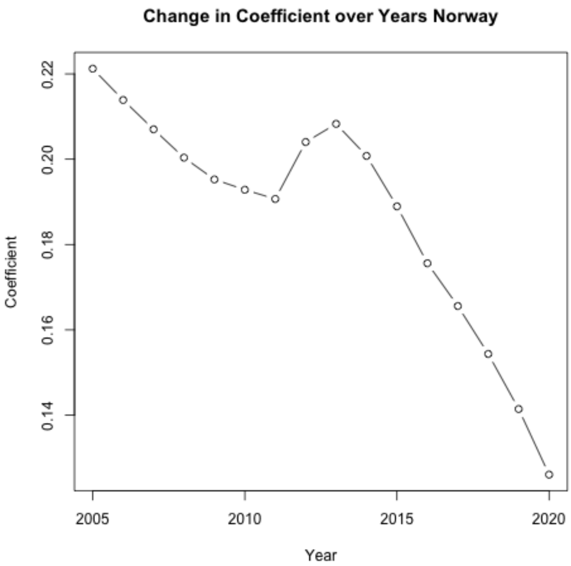

(a) model 1 results Italy




(b) model 1 results Lithuania

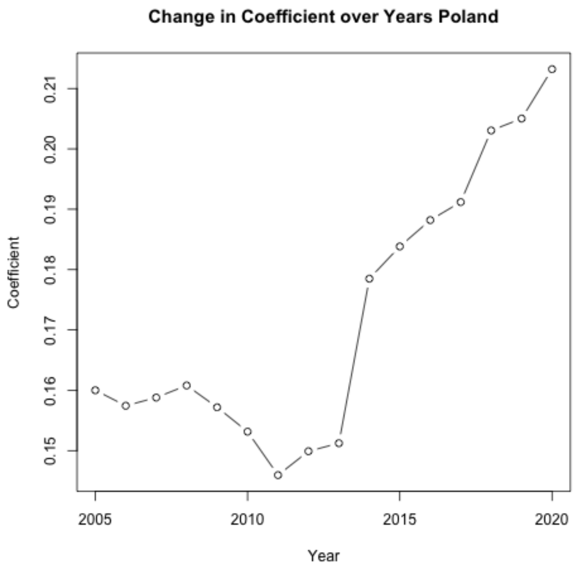



(c) model 1 results The Netherlands

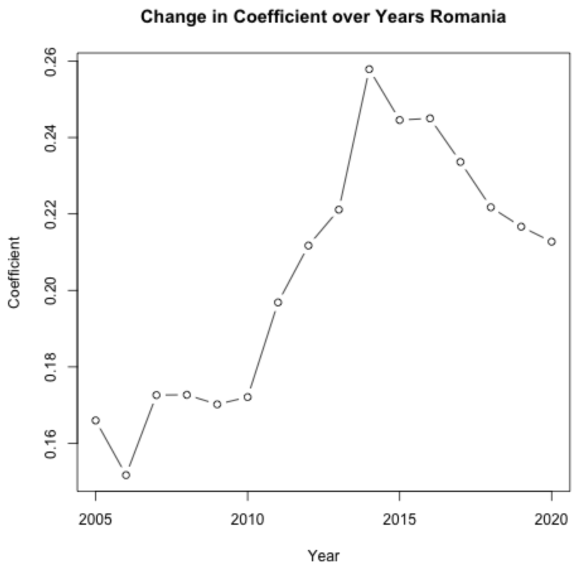



(d) model 1 results Norway

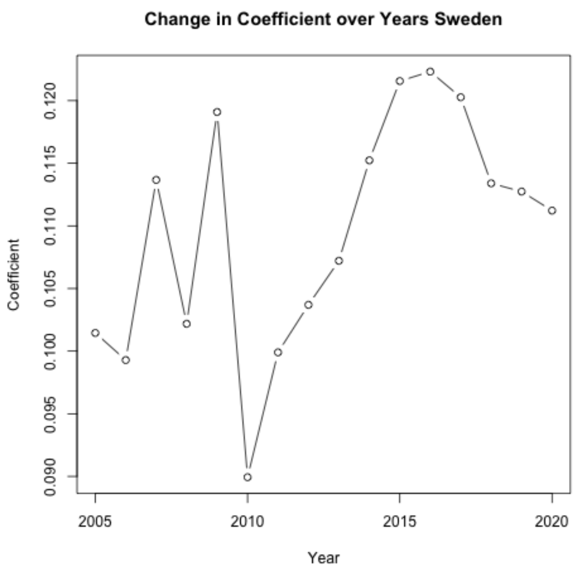
Figure A.6: Model 1 results for Italy, Lithuania, The Netherlands and Norway



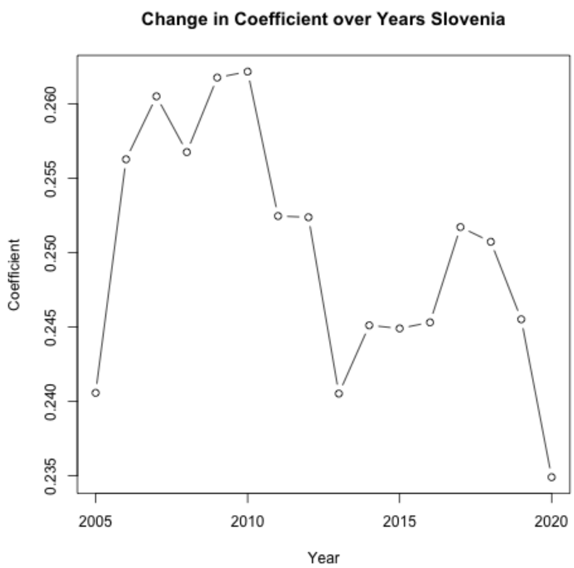

(a) model 1 results Poland




(b) model 1 results Romania




(c) model 1 results Sweden




(d) model 1 results Slovenia

Figure A.7: Model 1 results for Poland, Romania, Sweden and Slovenia

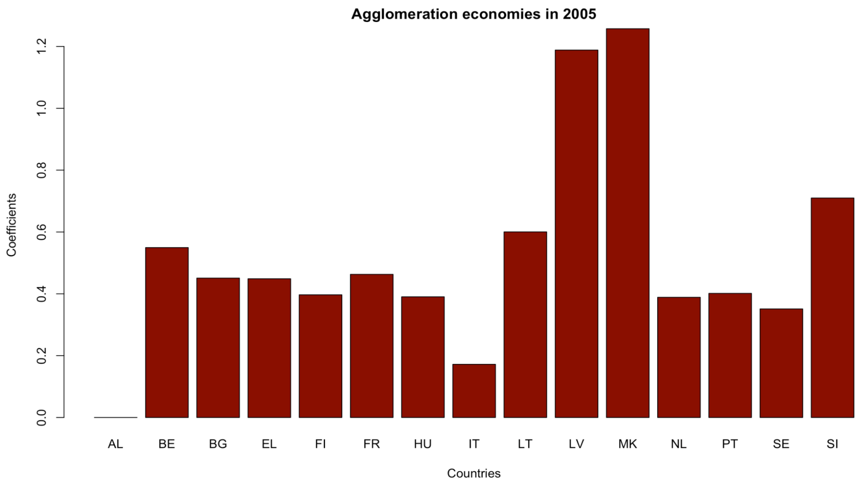



Figure A.11: Agglomeration economies in 2005

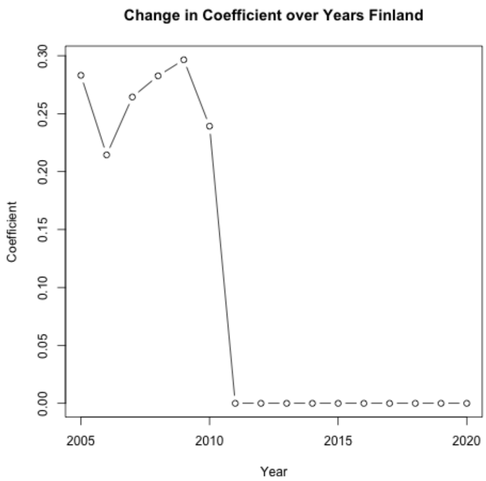



Figure A.10: model 2 results Finland

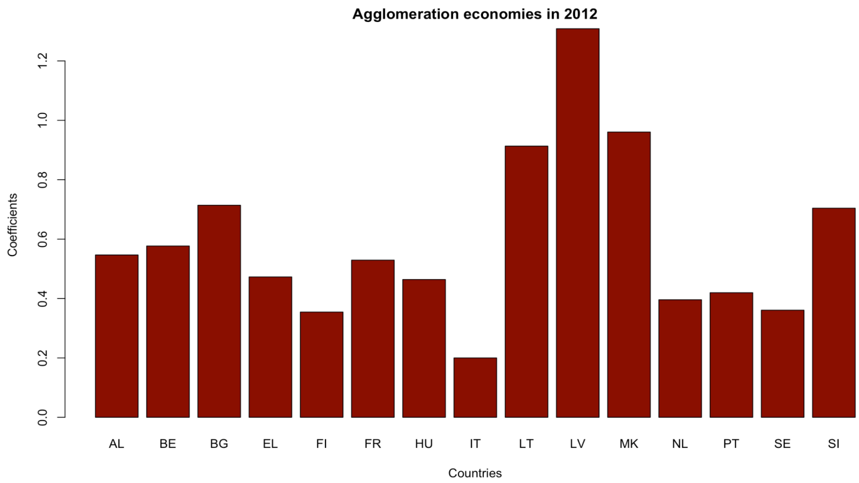
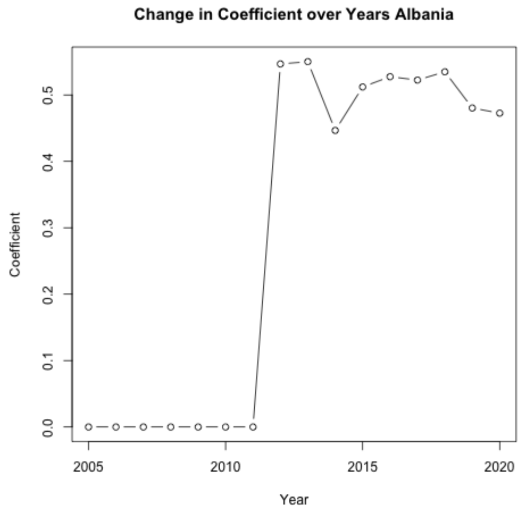
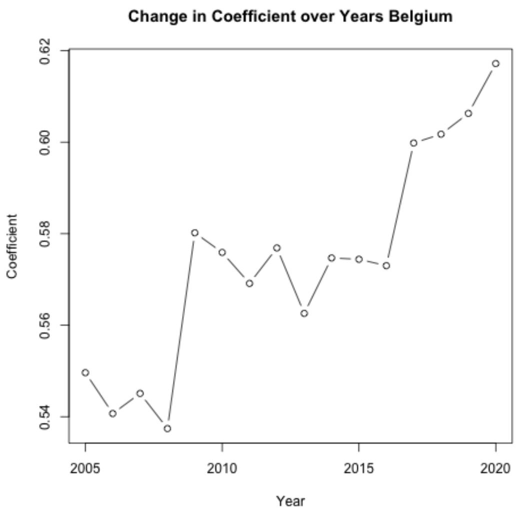


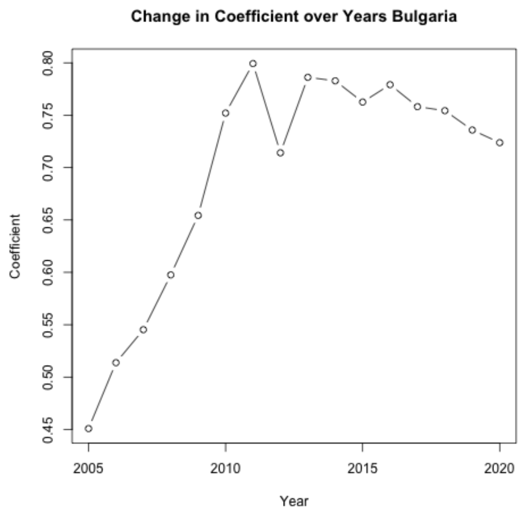

Figure A.12: Agglomeration economies in 2012



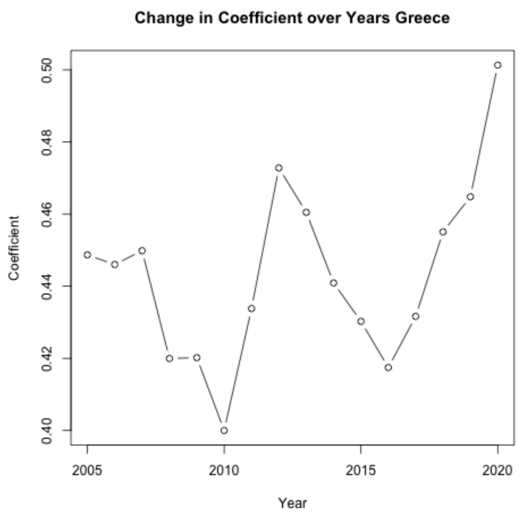

(a) model 2.b results Albania




(b) model 2.b results Belgium

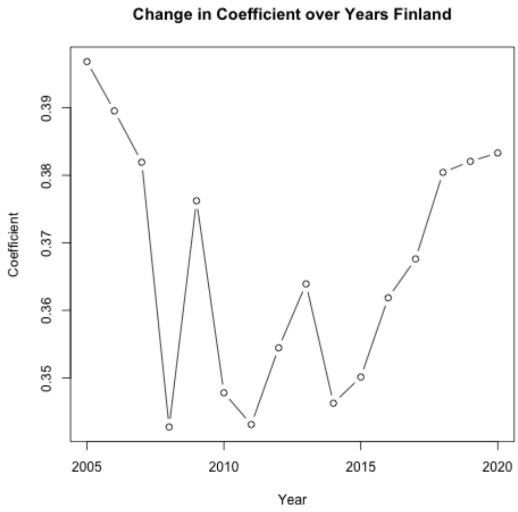



(c) model 2.b results Bulgaria

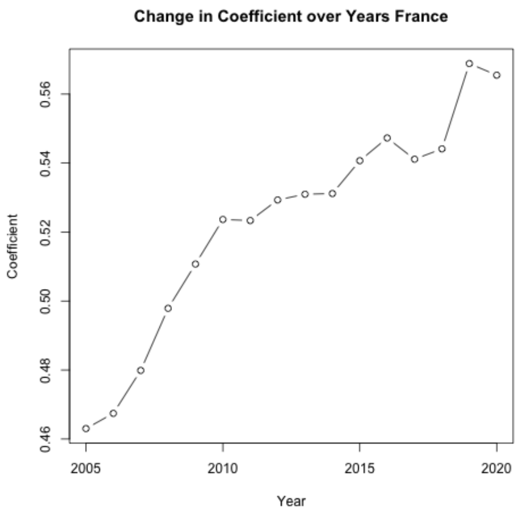



(d) model 2.b results Greece

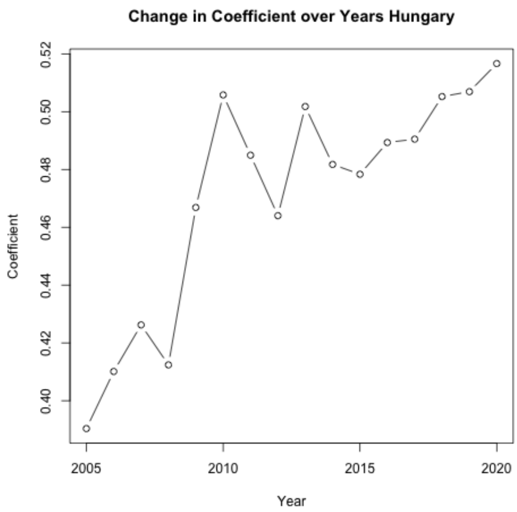
Figure A.13: Model 2.b results for Albania, Belgium, Bulgaria and Greece



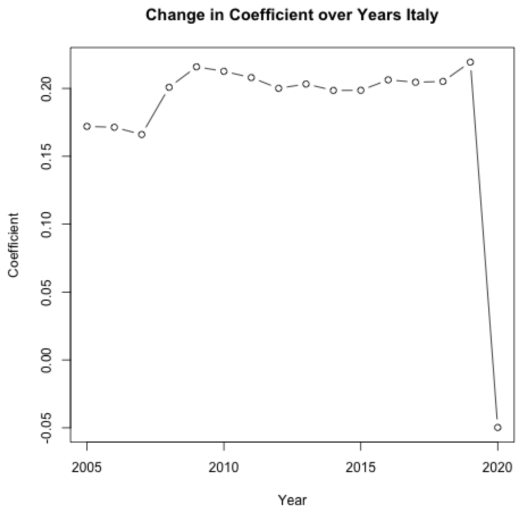

(a) model 2.b results Finland




(b) model 2.b results France



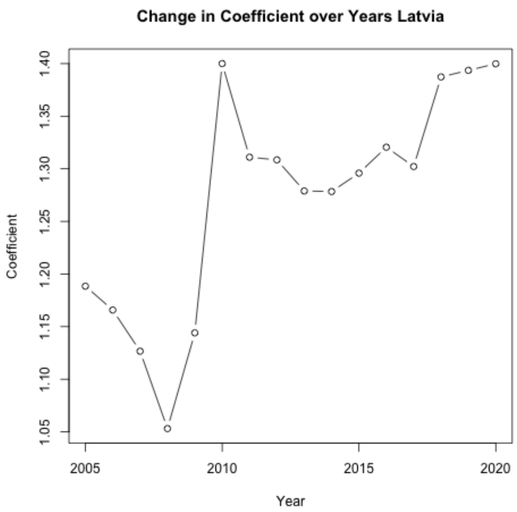

(c) model 2.b results Hungary



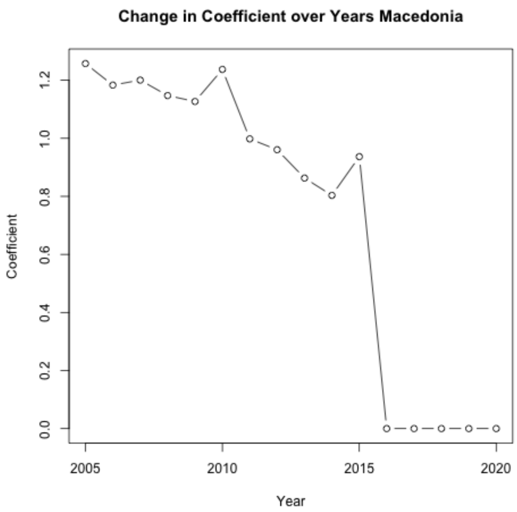

(d) model 2.b results Italy

Figure A.14: Model 2.b results for Finland, France, Hungary and Italy

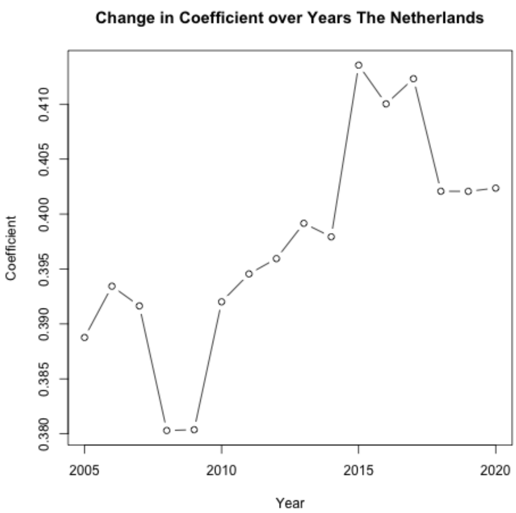

(a) model 2.b results Lithuania




(b) model 2.b results Latvia

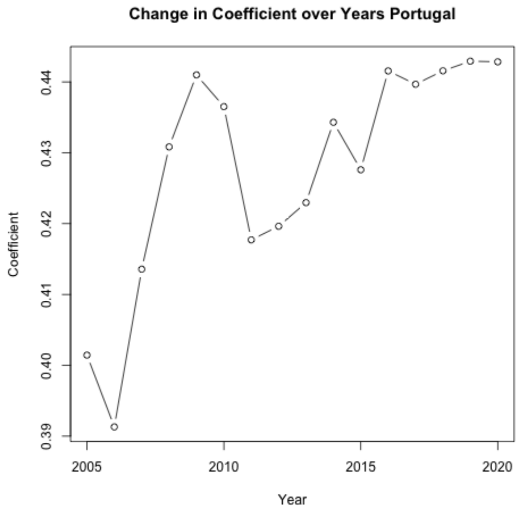



(c) model 2.b results Macedonia

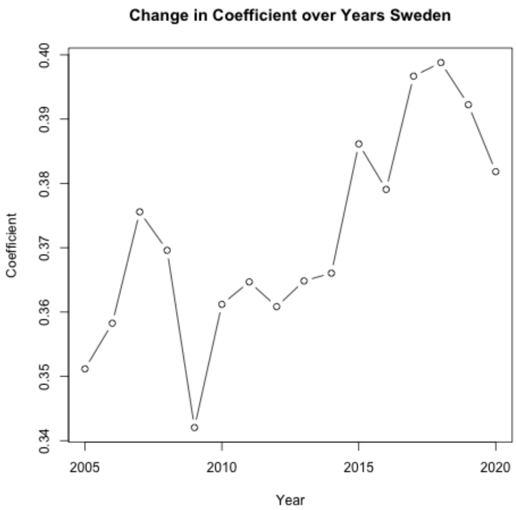



(d) model 2.b results the Netherlands

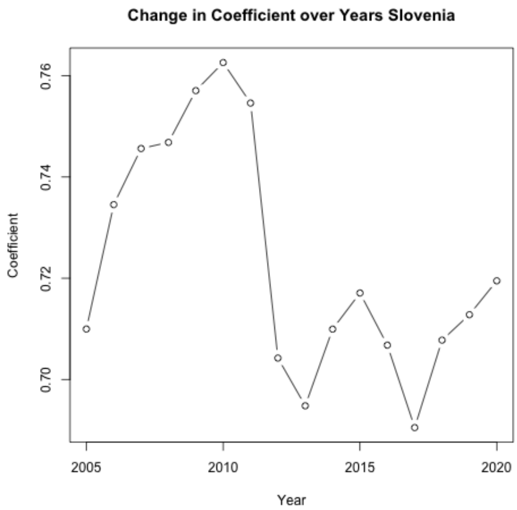
Figure A.15: Model 2.b results for Lithuania, Latvia, Macedonia and the Netherlands




(a) model 2.b results Portugal

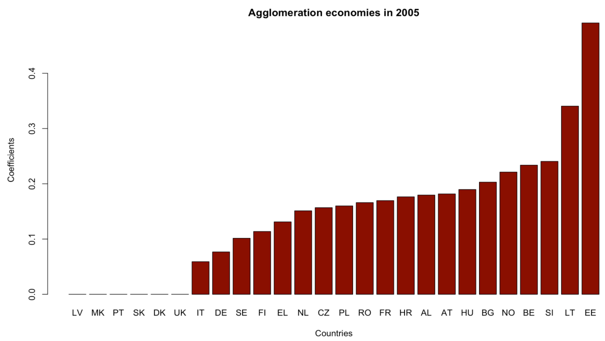



(b) model 2.b results Sweden

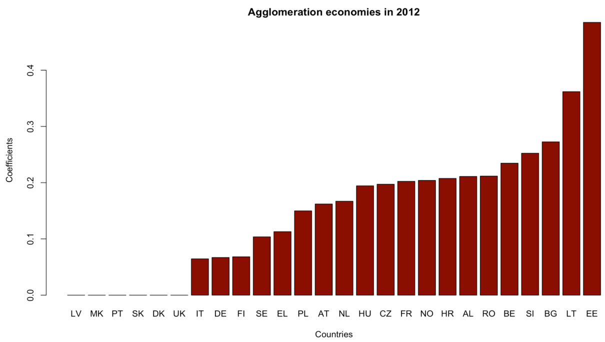



(c) model 2.b results Slovenia

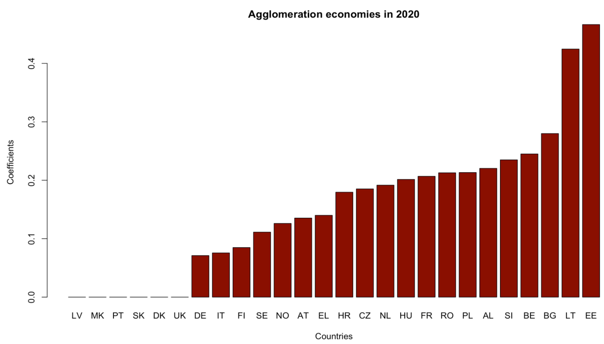
Figure A.16: Model 2.b results for Portugal, Sweden and Slovenia




(a) Agglomeration economies in 2005, model 1




(b) Agglomeration economies in 2012, model 1




(c) Agglomeration economies in 2020, model 1

Figure A.17: Agglomeration economies, model 1, ordered

# B | Appendix B

Table B.1: Adjusted R-square, first model, years 2005-2012

| Country | y05 | y06 | y07 | y08 | y09 | y10 | y11 | y12 |
|---|---|---|---|---|---|---|---|---|
| AL | 0.2995661 | 0.2503450 | 0.2657007 | 0.3557066 | 0.3361200 | 0.4363099 | 0.2639825 | 0.361194566212064 |
| AT | 0.3038914 | 0.2884505 | 0.3070154 | 0.3158679 | 0.2964941 | 0.3192816 | 0.2987790 | 0.278182002526098 |
| BE | 0.5254093 | 0.5085696 | 0.5161218 | 0.5336984 | 0.5343905 | 0.5323086 | 0.5362465 | 0.546805966306371 |
| BG | 0.5077523 | 0.5442533 | 0.5103652 | 0.4800393 | 0.5535749 | 0.5441011 | 0.5453237 | 0.530417905977385 |
| DE | 0.0405413 | 0.0405912 | 0.0436807 | 0.0465728 | 0.0414110 | 0.0340470 | 0.0331528 | 0.0321673480281011 |
| EE | 0.6236918 | 0.6197981 | 0.6088321 | 0.7489289 | 0.7994048 | 0.6086170 | 0.6699257 | 0.646557390416695 |
| EL | 0.1887413 | 0.1533604 | 0.1687190 | 0.1366301 | 0.1575667 | 0.1340221 | 0.1459049 | 0.136623531103481 |
| FI | 0.5341091 | 0.6179327 | 0.6273123 | 0.6385301 | 0.5362520 | 0.5132780 | 0.4643265 | 0.404005936015858 |
| FR | 0.4585021 | 0.4688567 | 0.4782006 | 0.5008012 | 0.4963872 | 0.5152483 | 0.5055021 | 0.514575521780715 |
| HR | 0.2327773 | 0.2491456 | 0.2790115 | 0.2933026 | 0.3000074 | 0.3085831 | 0.3023079 | 0.294574833764612 |
| HU | 0.1881981 | 0.1863680 | 0.1922074 | 0.1785929 | 0.2078602 | 0.1888735 | 0.1900345 | 0.1758044 |
| IT | 0.3445827 | 0.3492575 | 0.3474900 | 0.3342147 | 0.3155094 | 0.3257510 | 0.3130138 | 0.31293646919443 |
| LT | 0.6161496 | 0.6786294 | 0.6706133 | 0.6747460 | 0.6551502 | 0.7577932 | 0.7686663 | 0.835561862859171 |
| NL | 0.3017586 | 0.3147474 | 0.3097587 | 0.2921055 | 0.3815392 | 0.3962372 | 0.3887545 | 0.392305330061723 |
| NO | 0.5120131 | 0.4915588 | 0.4769919 | 0.5477945 | 0.5457272 | 0.4667852 | 0.4635165 | 0.516433840542319 |
| PL | 0.2158181 | 0.2013875 | 0.2036389 | 0.2148750 | 0.1901957 | 0.1718032 | 0.1534529 | 0.158600011127305 |
| RO | 0.1616979 | 0.1281381 | 0.1644518 | 0.1741031 | 0.1688841 | 0.1823630 | 0.2147335 | 0.288392518538474 |
| SE | 0.3998720 | 0.3941391 | 0.5051300 | 0.3929702 | 0.4582496 | 0.2720035 | 0.3284958 | 0.372768224901276 |
| SI | 0.9117153 | 0.9221520 | 0.9158491 | 0.9219362 | 0.9415815 | 0.9378129 | 0.9085470 | 0.876210920923217 |

Table B.2: Adjusted R-square, first model, years 2013-2020

| Country | y13 | y14 | y15 | y16 | y17 | y18 | y19 | y20 |
|---|---|---|---|---|---|---|---|---|
| AL | 0.4009563 | 0.3970830 | 0.4141294 | 0.3832579 | 0.4811803 | 0.4677876 | 0.4517450 | 0.4061987 |
| AT | 0.2631592 | 0.2419224 | 0.2180314 | 0.2202686 | 0.2257370 | 0.2195388 | 0.2245289 | 0.2270544 |
| BE | 0.5466586 | 0.5337093 | 0.5369016 | 0.5292288 | 0.5481388 | 0.5387725 | 0.5243216 | 0.5208971 |
| BG | 0.5577416 | 0.4732501 | 0.5012519 | 0.4953050 | 0.4668296 | 0.4562555 | 0.4979896 | 0.5143516 |
| DE | 0.0325291 | 0.0329971 | 0.0351101 | 0.0324618 | 0.0341432 | 0.0378631 | 0.0393179 | 0.0402960 |
| EE | 0.7333053 | 0.7497228 | 0.7875505 | 0.8171731 | 0.8071485 | 0.7984540 | 0.7407419 | 0.5585838 |
| EL | 0.1390348 | 0.1381870 | 0.1629708 | 0.1789002 | 0.1914837 | 0.2007080 | 0.2056372 | 0.1994375 |
| FI | 0.4189998 | 0.4130723 | 0.4436246 | 0.4813757 | 0.5214397 | 0.5762667 | 0.5589599 | 0.5236253 |
| FR | 0.5109322 | 0.5179271 | 0.5237192 | 0.5171565 | 0.5226047 | 0.5148578 | 0.5307123 | 0.4945494 |
| HR | 0.2853591 | 0.2794089 | 0.2991098 | 0.3080443 | 0.3157070 | 0.3169725 | 0.3147989 | 0.3717523 |
| HU | 0.1665970 | 0.1802016 | 0.1722974 | 0.1576552 | 0.1944330 | 0.2028897 | 0.2143184 | 0.2499894 |
| IT | 0.3257815 | 0.3207531 | 0.3193735 | 0.3189593 | 0.3254842 | 0.3310233 | 0.3391075 | 0.2962546 |
| LT | 0.8510543 | 0.8836658 | 0.8922835 | 0.9094106 | 0.9158653 | 0.9109933 | 0.9335216 | 0.9495746 |
| NL | 0.3758894 | 0.3948781 | 0.4432628 | 0.4640040 | 0.4599455 | 0.4530245 | 0.4746702 | 0.4840041 |
| NO | 0.5171242 | 0.4901207 | 0.4855439 | 0.4416876 | 0.3726358 | 0.2735100 | 0.1749251 | 0.0876644 |
| PL | 0.1671132 | 0.2483476 | 0.2551718 | 0.2663783 | 0.2784542 | 0.3077148 | 0.3058379 | 0.3239323 |
| RO | 0.3341125 | 0.4128236 | 0.3769127 | 0.3656946 | 0.3570546 | 0.3566136 | 0.3426325 | 0.3053191 |
| SE | 0.3880724 | 0.4302809 | 0.4543738 | 0.4791839 | 0.4457999 | 0.4119308 | 0.3767144 | 0.3365934 |
| SI | 0.8716096 | 0.8464703 | 0.8097659 | 0.7988614 | 0.7920558 | 0.7750058 | 0.7765752 | 0.7918604 |

Table B.3: Adjusted R-square, 2.b model, years 2005-2012

| Country | y05 | y06 | y07 | y08 | y09 | y10 | y11 | y12 |
|---|---|---|---|---|---|---|---|---|
| AL | 0.0917400 | 0.1128717 | 0.1122883 | 0.1094285 | 0.1295444 | 0.2178107 | 0.1379823 | 0.3621185 |
| BE | 0.7154467 | 0.7049564 | 0.7119050 | 0.7036572 | 0.7258318 | 0.7242585 | 0.7179635 | 0.7133659 |
| BG | 0.5783497 | 0.5886624 | 0.5901279 | 0.6550372 | 0.6047670 | 0.6977767 | 0.7229997 | 0.6585332 |
| EL | 0.5325835 | 0.5507650 | 0.5403617 | 0.4862357 | 0.4828932 | 0.4739294 | 0.5122420 | 0.5250220 |
| FI | 0.6345201 | 0.6438695 | 0.6657333 | 0.5916588 | 0.6302508 | 0.5727315 | 0.5436464 | 0.5694635 |
| FR | 0.5619472 | 0.5643272 | 0.5784772 | 0.5906742 | 0.5803238 | 0.5959753 | 0.6040645 | 0.5924913 |
| HU | 0.3316065 | 0.3317699 | 0.3472225 | 0.3342845 | 0.3755824 | 0.4165385 | 0.4043964 | 0.3625122 |
| IT | 0.3438201 | 0.3458497 | 0.3395479 | 0.3591526 | 0.3760337 | 0.3656344 | 0.3420082 | 0.3383356 |
| LT | 0.8781406 | 0.9054465 | 0.9017273 | 0.8629051 | 0.8459490 | 0.8482334 | 0.8546459 | 0.8655911 |
| LV | 0.6118980 | 0.5832039 | 0.6745415 | 0.5931628 | 0.6147073 | 0.7884126 | 0.8339640 | 0.7222954 |
| MK | 0.7037940 | 0.6140686 | 0.5865814 | 0.5863142 | 0.5973224 | 0.6290562 | 0.4732770 | 0.4201937 |
| NL | 0.4468704 | 0.4523526 | 0.4540312 | 0.4516769 | 0.4721144 | 0.4732823 | 0.4780479 | 0.4938544 |
| PT | 0.3978097 | 0.3938956 | 0.4149325 | 0.4555240 | 0.4949596 | 0.5025057 | 0.4655828 | 0.5057091 |
| SE | 0.3909833 | 0.4045563 | 0.4580974 | 0.4599444 | 0.3960844 | 0.4386895 | 0.4598204 | 0.4320977 |
| SI | 0.8112330 | 0.8184568 | 0.8015493 | 0.8253978 | 0.8316284 | 0.8471417 | 0.8668372 | 0.8412788 |

Table B.4: Adjusted R-square, 2.b model, years 2013-2020

| Country | y13 | y14 | y15 | y16 | y17 | y18 | y19 | y20 |
|---|---|---|---|---|---|---|---|---|
| AL | 0.3656628 | 0.3057310 | 0.3399088 | 0.4005848 | 0.3732253 | 0.4525749 | 0.3373239 | 0.298779310668224 |
| BE | 0.7193668 | 0.6997119 | 0.6960576 | 0.6912635 | 0.7208163 | 0.7142227 | 0.6971039 | 0.698875162231879 |
| BG | 0.7171655 | 0.7037567 | 0.6740324 | 0.6883779 | 0.6723407 | 0.6806106 | 0.6315284 | 0.579928712229299 |
| EL | 0.5388971 | 0.4944825 | 0.5070364 | 0.5024997 | 0.5178549 | 0.5359267 | 0.5392341 | 0.567454465691256 |
| FI | 0.6019288 | 0.5942702 | 0.5518369 | 0.5605152 | 0.5939300 | 0.6029468 | 0.5849153 | 0.553850728243514 |
| FR | 0.5913678 | 0.5964663 | 0.6052281 | 0.6018963 | 0.5996827 | 0.6072542 | 0.6119364 | 0.598131234013013 |
| HU | 0.3955634 | 0.3698847 | 0.3665812 | 0.3720289 | 0.3806150 | 0.4060795 | 0.4281796 | 0.424089509963304 |
| IT | 0.3535061 | 0.3456492 | 0.3438824 | 0.3424125 | 0.3559505 | 0.3578240 | 0.3719136 | 0.194019072890268 |
| LT | 0.8740382 | 0.8759693 | 0.8571404 | 0.8858482 | 0.9075539 | 0.9005532 | 0.8703728 | 0.835936969803581 |
| LV | 0.7142500 | 0.7355547 | 0.7316776 | 0.7415618 | 0.7173637 | 0.7248076 | 0.7549112 | 0.781221164415972 |
| MK | 0.4107489 | 0.3434126 | 0.4409811 | 0.2081723 | 0.1286161 | 0.0953330 | 0.0648841 | 0.0332783 |
| NL | 0.4947665 | 0.4880355 | 0.5112948 | 0.5059534 | 0.4913628 | 0.4747124 | 0.4715824 | 0.46404844424319 |
| PT | 0.5205734 | 0.5241823 | 0.5258908 | 0.5522919 | 0.5248510 | 0.5214708 | 0.5495995 | 0.564329045777034 |
| SE | 0.4097360 | 0.4119487 | 0.4231411 | 0.4096938 | 0.4599616 | 0.4249060 | 0.4078935 | 0.380462600984211 |
| SI | 0.8408112 | 0.8530129 | 0.8388747 | 0.7922435 | 0.8004660 | 0.8164233 | 0.8359157 | 0.820672847054823 |

# List of Figures

# List of Tables

# List of Acronyms

| Acronym | Description |
|---|---|
| EU28 | EU Member States as of december 2019 |
| EU15 | pre-2004 European Member states |
| EU8 | States that joined the EU January 2004 |
| NUTS (0,1,2,3) | Nomenclature of territorial units for statistics |
| FUA | Functional Urban Areas |
| CEEC | Central and Eastern European Country |
| SEE | South-East Europe |
| NEG | New Economic Geography |
| GVA | Global Value Added |
| LTM | Low-Technology manufacturing |
| KIS | Knowledge-intensive sectors |
| HTKIS | High Technology- and Knowledge-intensive sectors |
| OKIS | Other Knowledge-intensive sectors |
| LKIS | Low Knowledge-intensive sectors |
| PU | Predominantly Urban |
| IN | Intermediate regions |
| PRC | Relatively close to cities |
| PRR | Relatively remote from cities |
| NACE Rev.2 | Statistical classification of economic activities in the European Community |
| ATS | Advanced Tradable Services |

# Acknowledgements

I would like to express my sincere appreciation to my supervisor, professor Lucia Tajoli, for her support, guidance, and expertise throughout the entire process of completing this master's thesis. I want to express my gratitude to the wider academic community, researchers, and scholars whose work has shaped my understanding and provided a foundation for this thesis. Their contributions have inspired and influenced my research endeavors.

I am grateful to my friends in Pareto who have provided a supportive and stimulating environment. I want to thank the friends I met abroad, both at HEC Lausanne and Solvay Brussels School, for their intellectual discussions, shared experiences, and collaboration. I would like to acknowledge the QTEM program, in particular professor Sergio Terzi, which provided the opportunity of meeting ambitious and skilled students and alumni that inspired and motivated me. The people I met along the way, at Polimi, HEC, SBS and QTEM, have contributed to my personal and academic development, for which I am grateful.

I extend my heartfelt appreciation to my family for their endless love, encouragement, and unwavering belief in my abilities. Their constant support and understanding have been a source of motivation and strength throughout my academic journey. Lastly, I am thankful for the invaluable help of Andrea, who has proof-read my drafts and evaluated with me the model and its implications. His insightful feedback, valuable suggestions, and encouragement have been crucial for this research, as well as for my personal growth.

Although it is challenging to mention everyone by name, I am truly thankful to all those who have supported me directly or indirectly during this journey. Your encouragement, assistance, and belief in my abilities have been invaluable.

Thank you all for being a part of my master's journey and for making this achievement possible.