STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic Year 2016-17

Entrepreneurship in the Information Age: An Empirical Analysis of the European Regions

Petr Pleticha (40897)

Abstract

This thesis analyzes the relationship between digitalization and entrepreneurship using European regional data. Decelerating productivity in recent years raised questions about technology dissemination in the economy. This study inspects one of the dissemination channels, entrepreneurship, and links the empirical findings to Kirzner's and Schumpeter's theory of the entrepreneur. Based on econometric analysis, I find a significant relationship between digitalization and entrepreneurship. Specifically, digitalization increases the rate at which firms are created and it decreases their survival rate after 3 years. This influence seems to be dynamic in its nature as the effects of initial stages of digitalization reverse in its later phases. Moreover, the results are not uniform across Europe. The impact of digitalization on entrepreneurship varies among regions with the Nordic countries being especially responsive to the first stage of the digitalization process. The results suggest that digital technology uses new business creation and destruction as a dissemination channel. Public policies as well as corporate strategies may thus consider the complementarity of digitalization and entrepreneurship.

Keywords: Digitalization, Entrepreneurship, Innovation economics, Technology dissemination
JEL: L16, L26, O33, R11

Supervisor: Örjan Sjöberg
Date submitted: May 14, 2017
Date examined: May 29, 2017
Examiner: Johanna Wallenius
Discussant: Anne Beck

Acknowledgements

I wish to express my gratitude to Harald Edquist and Örjan Sjöberg who guided me through the research project with their helpful comments and inspiring discussions. I also want to thank my family who supported me and put their trust in me during my research efforts. This study would not be possible without them.

Contents

List of Figures

List of Tables

1 Introduction

Digitalization, entrepreneurship, ICT investment. These concepts have become popular buzzwords often used to cause frantic nodding of the audience. They have also become leitmotifs in the policy makers' rhetoric. The digital agenda in the EU's strategy Europe 2020 contains the aim to address overdue investments in telecom infrastructure and to promote skills required for a successful participation in the labor market. At the same time, The Entrepreneurship 2020 Action Plan of the European Commission aims to provide entrepreneurial education and to reignite the entrepreneurial culture. Analogous strategies are observed at national level as well. For instance, the Czech Republic's target is to provide 100 Mbps broadband access to at least 50 percent households and enterprises by 2020, and business incubators, funded by both the private and public sector, sprout in every major European city. The goal of spurring growth via digitalization and entrepreneurship is widely pursued in the EU and beyond.

Although there is no consensus on how entrepreneurs increase productivity and promote economic growth, there are both theoretical and empirical foundations stating they do so (Karlsson et al., 2004). The same applies to innovation, whether it is digitalization or any other kind. Extensive literature proves innovation (especially in the ICT sector) to be one of the main factors of economic growth (Cardona et al., 2013). However, it is not clear how technological progress and entrepreneurship affect one another and what the exact channels are through which both phenomena enhance economic activity. Is it the entrepreneurs who push technology forward or is it rather the spoils of technology which allow new businesses to grow? Does entrepreneurship serve as a dissemination channel for new technology? And if so, is there any observable dynamics of the relationship between digitalization and entrepreneurship? Specifically, does digitalization dynamically affect firm creation and firm destruction? These are the questions I attempt to address in this paper.

Interpretations of Kirzner and Schumpeter provide two distinct narratives of the role of entrepreneurs in the economy. The Kirznerian view states that entrepreneurs exploit opportunities which are available in the economy, essentially participating in arbitrage. It is those unused opportunities which hold the economy back from its equilibrium. The entrepreneurial activity has thus a correcting effect for the economy, bringing it back to the theoretical equilibrium (Kirzner, 1997). Following this school of thought, advancing technology creates more opportunities for entrepreneurs to exploit, suggesting one-sided causal relationship. Hence, digitalization would positively affect economic growth (among other channels) via promoting entrepreneurship.

On the contrary, the traditional Schumpeterian view presents entrepreneurs as disrupters of economic equilibria, participating in creative destruction. Schumpeterian entrepreneurs are not necessarily a source of technological progress, but they are the channel through which the economy adapts to new technology. Thus they constantly bring the economy off the equilibrium, only to be moving it swiftly towards the new steady state (Schumpeter, 1961; Holcombe, 1998). Although both narratives successfully describe how technological progress is implemented, they do not elaborate on its origin. Apart from researchers in both privately and publicly funded institutions, could the entrepreneurs be also the possible suspects? Then it is small businesses who push technology forward and the whole economy with it. This suggests a more complicated interplay between entrepreneurship and digitalization.[1]

Because many policies aim to foster entrepreneurship and speed up the digitalization process, it is vital to understand causal relationships between those variables. In this study, panel data regression methods are used to establish empirical dependency between digitalization and entrepreneurship. Focusing on the Kirznerian entrepreneur, the main model employs fixed and random effect panel data techniques to reveal digitalization's effect on entrepreneurship. Apart from creating new arbitrage opportunities, digitalization, in general, reduces transaction costs and cost of information for people as well as for businesses. The question of interest is whether the reduced costs affect existing businesses more than the potential entrants. In other words, whether the transaction costs within enterprises diminish at faster rate than the costs among enterprises (including costs of market entry). This could either lead to more entrepreneurial opportunities and new businesses entering the market, thus stiffening the competition, or it could make the position of the incumbents even stronger through enhanced profitability.

The contribution to the existing literature is threefold. Firstly, the paper builds on the previous literature on digitalization indices trying to capture the technological transformation and adapts it to the needs of regional analysis. Thanks to the composite digitalization index, future analysis of regional data becomes much easier. Secondly, it investigates the relationship between digitalization and entrepreneurship. There is a detailed literature describing how business produces and implements innovation (Baden-Fuller and Haefliger, 2013) but there is no inquiry into the effects of advancing technology on entrepreneurship. This study aims to close the gap in the literature. Thirdly, I seek to test universality of the results with respect to spatial arrangement of the regions.

The analysis proves there is a significant interplay between digitalization and entrepreneurship. If a region moves up by 10 percentile points in the digitalization ranking, it can expect decreased rate at which the newly created firms survive by 1.7%. GMM estimation even points towards causality of the effects which digitalization has on entrepreneurship causally. It increases the rate at which companies are founded but decreases their chance of survival. This suggests that digitalization fosters dynamic business environment. However, such impacts are not permanent, as the effects of initial stages of digitalization reverse in its later phases. The first two stages of digitalization spur firm creation; gain of 10 percentile points in both stages

[1] Although the role of the Kirznerian and Schumpeterian entrepreneur in the economy differs substantially, they are not mutually exclusive. Kirzner himself acknowledges that, aspiring to reconcile with Schumpeter's entrepreneur (Kirzner, 1999).

of digitalization is associated with 1.2% increase in birth rate. However, the same performance in stage 3 is linked to a shrinkage in the company birth rate by almost 1%. Moreover, the results are not uniform across Europe. The effects of digitalization on entrepreneurship vary among regions, with the Nordic countries being especially responsive to the first stage of the digitalization process. For example, whereas an "average" region's rise by 10 percentile points in digitalization's initial stage ranking relates to mere 1.2% increase in the birth rate, the same change in relative digitalization standing is associated with a rise by 4% in the Nordic regions.

In the paper, I firstly introduce the relevant literature regarding entrepreneurship, digitalization, and productivity growth. A theoretical framework putting the empirical analysis into the context of real economy comes next. I then describe the data and methods used in the analysis. The results section is followed by discussion of endogeneity. In the robustness section, I conduct several tests to see whether the results hold under different circumstances. Then, I briefly summarize the results, discuss limitations of the study and suggest areas of future research. Finally, I conclude the paper.

2 Literature Review

2.1 Entrepreneurship and its role in the economy

Although public policy often points towards entrepreneurship as to the engine of economic growth, empirical economic literature has long neglected it as an important factor of economic development. Entrepreneurship is difficult to quantify and include in models of neo-classical economics and thus economists simply overlooked it (Baumol, 1968). However, entrepreneurship has made its comeback and has been extensively analyzed in span of the last 15 years. Thanks to the Global Entrepreneurship Monitor (GEM), comprehensive data on the national level is now available fueling research on this topic. The economic theory is surprisingly in line regarding entrepreneurship. There is a debate on what entrepreneurship means, how it should be defined and what proxy captures it the best, but no matter what the opinion of an economist is and whether she follows the Kirznerian or the Schumpeterian school of thought, she always concludes that entrepreneurship indeed spurs innovation and is a vital ingredient of economic growth (McQuaid, 2002; Holcombe, 1998).

This is true even considering a possibility of harmful role of the entrepreneur. Baumol (1996) argues that virtue of entrepreneurship depends on the structure of the economy. If the rules are not set properly, the entrepreneur engages in rent-seeking rather than technology dissemination. However, empirical literature shows that in contemporary Western economies, the institution quality is sufficient for the entrepreneurial paragon to dominate. Indeed, entrepreneurship positively affects both total factor productivity (Erken et al., 2016) as well as directly economic growth (Galindo and Méndez, 2014). Moreover, economic growth stimulates entrepreneurship putting a virtuous circle in place. These optimistic findings are partially undermined by looking below the aggregate data. Based on the GEM data, van Stel et al. (2005) have found that entrepreneurship contributes to economic development mostly in already developed countries. Wong et al. (2005) independently confirm these results by showing that only entrepreneurship associated with high growth potential (in contrast with entrepreneurship caused by sheer necessity) stimulates economic progress.

This might not come as a surprise though. Both van Stel et al. and Galindo and Méndez take total entrepreneurship activity (TEA) provided by GEM as their entrepreneurship proxy but TEA is constructed as a share of individuals who are either nascent entrepreneurs or take part as an owner-manager in their new business. It thus might be the lack of other conventional opportunities which drives the entrepreneurial activity of most. Entrepreneurship then coincides with poor economic opportunities, not a vibrant business environment (Ács, 2006). Higher opportunity costs of entrepreneurship in developed countries only amplify this effect (Bosma and Schutjens, 2011). Moreover, the data come from a survey and are therefore self-reported which further adds to the distortions across countries caused by different perception of individual economic conditions.

Defining entrepreneurship as self-employment has also been widely criticized since it is the rapid growth of firms which provides the entrepreneurial dynamics to the economy (McQuaid, 2002). Bjuggren et al. (2012) also criticised such metrics pointing out that the rise in entrepreneurship in Sweden in 1987 was solely due to changed definition of self-employment, not due to suddenly flourishing entrepreneurship in the country. Although TEA, essentially a measure of start-up participation, is a significant step forward, there are several issues with the metric. It is strongly correlated with the blunt self-employment rate which magnifies the concerns about self-reporting bias. Identifying this shortcoming, Henrekson and Sanandaji (2014) propose to define entrepreneurship as number of high impact entrepreneurs, self-made billionaires who accumulated wealth by founding a new business. As ingenious as this method might be, it is unfortunately not fit for regional analysis.

No matter what definition and measure of entrepreneurship one decides to follow, there is a substantial variation between nations and regions (Masuda, 2006). The explanations for the variation is needed when designing and evaluating policies intended to spur entrepreneurship. What are the determinants of nascent business activity? Simón-Moya et al. (2014) define four sets of drivers of entrepreneurship: economic, institutional, cultural, and educational. They manage to prove that low GDP per capita, high unemployment, and high income inequality constrain start-up rate. Strong institutions and developed human capital, on the other hand, correspond with vibrant start-up sector. Lastly, the significance of the cultural drivers has not been proven. Wennekers et al. (2007) explored the effect of uncertainty avoidance on entrepreneurship and identified additional possible determinants of entrepreneurship such as female labor participation and demographic structure. Grilo and Thurik (2008) analyzed a survey of the US and 18 European states, inspecting, among other things, the perception of administrative complexity, availability of financing, risk tolerance, and preference for being self-employed. They found that high perception of administrative complexities indeed discourages potential entrepreneurs.

2.2 Digitalization

There is a confusion in the literature when using the terms digitalization and digitization. Katz and Koutroumpis (2013) and Jacobsen et al. (2011) clearly talk about the same phenomenon, yet their terminology differs. To be clear and concise, I will adhere to the following definition in this study which is loosely based on Gartner (2017), the research and advisory company:

> Digitalization is a process of integrating new digital technologies into business models as well as our everyday lives. Digitization, on the other hand, is a process of converting any information into a digital format.

When addressing the aspects of digitalization, most scholars focus on ICT capital (Miller and Atkinson, 2014; Cardona et al., 2013; Edquist and Henrekson, 2017). Others have also analyzed phone coverage (Muto and Yamano, 2009) and broadband penetration (Thompson and Garbacz, 2008). Gruber and Koutroumpis (2011) use a more general model trying to describe the mutual relationship of mobile penetration and gross domestic product per capita. Separately estimating supply of and demand for mobile infrastructure as well as production functions for mobile infrastructure and the whole economy, they managed to control for the causal effects via three-staged least squares estimation. They found that mobile penetration indeed positively affects GDP growth. Although this approach accounts for more aspects of digitalization (specifically mobile penetration, mobile monthly subscriptions, and revenue of the mobile industry), there is no broad variety of literature attempting to cover the whole phenomenon of digitalization.

Katz et al. (2014) fill this gap by proposing a digitalization index measuring holistic impact of ICT. The index tries to capture not only ICT penetration but also the degree to which households and businesses adopt the new technology. It surpasses the Digital Opportunity Index calculated annually by International Telecommunication Union (ITU) as it includes additional features of digitalization which the Digital Opportunity Index omits.[2] In total, Katz' index comprises six equally weighted components: affordability, infrastructure reliability, network access, capacity, usage, and human capital. The index explores technical characteristics beyond network penetrations. The impact of digitalization is indeed expected to progress faster in societies with high level of technological literacy which can be estimated by usage of social networks and exploiting e-government and e-commerce services. Therefore, affordability, reliability, and sufficient capacity are all included in the index as they are essential for digitalization being assimilated into our daily lives.

2.3 Productivity growth discussion

There is an extensive literature discussing productivity growth caused by advancing ICT and IT in general. Although it is nowadays widely accepted that ICT investment leads to increased productivity, it was a disputed topic back in 1990's. Robert Solow had famously mocked the technological progress in *The New York Times Book Review* (Solow, 1987): "You can see the computer age everywhere but in the productivity statistics." Indeed, empirical literature could not prove linkage between intensive investing into ICT and productivity. Brynjolfsson (1993) called it a paradox of information technology and productivity but claimed that it is the fault of mismeasurement and that the productivity growth is merely not visible in the data.

One of the possible explanations for the paradox is based on the reasonable assumption that new technology makes everyday tasks less repetitive and tedious and thus the whole work life better. Because the quality of work life is a form of output distributed to the employees,

[2]Digitalization Opportunity Index is based on number of subscriber per 100 inhabitants for fixed telephone, mobile cellular, internet, and broadband services. Usage of online services or human capital is not considered at all. See ITU for further detail.

ceteris paribus, we would falsely observe stagnant productivity, when in fact it has increased substantially. In general, qualitative improvements in services are extremely difficult to measure and because ICT progress affects largely the service sector, the productivity estimates might be misleading (Baily and Gordon, 1989; Noyelle, 1990). Moreover, ICT investment is of rather long-term character, so it was simply premature to make any conclusions based on the early studies. Brynjolfsson et al. (1994) found it takes two to three years until the technology investments starts to yield organizational results.

David (1990) also noted that it takes time for the new technology to deliver productivity increases. He presented a clear analogy between ICT and invention of electricity. Although both inventions were revolutionary, it took time for other industries to implement them. Bresnahan and Trajtenberg (1995) coined the term "general purpose technologies" including ICT, electricity, and the combustion engine. Such technologies offer a huge productivity growth in wide spectrum of industries but require time and significant investments into infrastructure and both human and physical capital. Jovanovic and Rousseau (2005) focused on the two main general purpose technologies: electricity and ICT. They have analyzed periods spanning over 35 years to capture the long-term development of the two technologies. Electricity was adopted more rapidly and uniformly across industries, but it lagged behind ICT in terms of new patents and trademarks. Both general purpose technologies, however, spurred creative destruction in the markets. More businesses were created, more died and even more mergers and acquisition took place in these periods.

But only with new datasets enabling analysis on firm level, a positive effect of ICT on productivity was proven (Brynjolfsson and Hitt, 1995; Dewan and Min, 1997). At the dawn of new millennium, Oliner and Sichel (2000) showed that ICT investment gives a significant boost to productivity also based on aggregate macroeconomic data. Similarly, Jorgenson et al. (2008)[3] identify ICT as the driving factor of productivity acceleration whose effect peaked during the ICT investment boom in late 1990's and slightly diminished in early 2000's because of the dot-com crash. With this issue resolved, other puzzles quickly materialized. The effect of ICT investment seemed to vary on both sides of the Atlantic Ocean. Europe had been climbing toward US productivity throughout the post-war period but the gap began to widen again in 1995. Addressing the phenomenon, van Ark et al. (2008) attributed the European productivity slowdown to lower ICT investment compared to the United States. On the same note, Gordon (2004) found that the US productivity revival was mostly due to ICT and that the European sluggishness could be accounted for by the southern states being slow at adopting the new technology, hence the differential. Cardona et al. (2013) also linked the widening productivity gap lasting until the mid-2000's to ICT investment and the respective technology implementation.

[3]Jorgenson et al. (2008) also predicted 2.5% productivity growth in the next 10 years. Such prediction turned out to be overally optimistic.

A more institutional point of view is the one of Phelps. He attributes the sudden departure of the European and American productivities to the different varieties of capitalism. Whereas the liberal market economies such as the US, the UK, or Canada are dynamic and open to creative destruction and innovation, the coordinated market economies of continental Europe rely more on corporatism designed to curtail such dynamism. In liberal economies the new businesses get financed through venture capital rather than banking system which is a generally less agile and risk-taking way of financing new projects (Phelps, 2006). Following this reasoning, coordinated market economies are well-suited for catching up with the productivity leader, but they find it difficult to claim the productivity leadership. Moreover, liberal economies should experience higher level of small business activity than the coordinated ones.

The debate about productivity is far from settled. The grim notion of secular stagnation predicting long periods of near-zero productivity growth (Gordon, 2014) is countered by the optimistic view of secular innovation where technology itself is the source of soaring productivity (Brynjolfsson and McAfee, 2014). However, if one does not dismiss the reliability of productivity data straight away, an explanation of the decelerating productivity is necessary. One such explanation is diffusion dynamics, a mechanism of spreading new technology to other regions, companies, and sectors. This paper aims to add to this discussion by shedding light on how entrepreneurship serves digitalization as the diffusion mechanism.

3 Theoretical Framework

Both Kirzner and Schumpeter admit that technology is a crucial factor in entrepreneurial activities. No matter whether the entrepreneurs are solely looking for opportunities or are driving the technological change themselves, digitalization plays an important part in their endeavor. Because ICT is a general purpose technology, it does not affect businesses only in one industry; it affects the whole economy. So digitalization offers an arbitrage opportunity to the Kirznerian entrepreneur and the Schumpeterian entrepreneur creates the opportunity itself using the same technology. Either way, technology spurs business activity, so we can abstract from the schism in entrepreneur's definition and focus only on the relationship itself. Moreover, because ICT has been identified as a general purpose technology, it is possible to focus on aggregate data. Naturally, productivity and business activity are affected more in certain sectors, but the focus of this study is to analyze the overall effect on the economy.

I identify five channels through which digitalization disseminates via entrepreneurship and thus affects its dynamics. Firstly, the new technology enables creation of brand new products and services. Industries such as computer gaming, social networks, and many others would not exist, had not we experienced the ICT revolution. These opportunities invite the pioneering spirit of entrepreneur as she has a comparative advantage over the existing players of not being tied by the industry structure. She can thus afford to be more aggressive and disruptive as she does not have any stake in the status quo. This translates mostly into increased birth rate of companies. The more products and services are developed, the more companies are founded.

The brand new products provide many other opportunities. For example, the invention of combustion engine enabled development of the whole oil and gas industry. Similarly, inventions in the ICT sector create new possibilities of adjacent products and services which are dependent on the original one. The whole cyber security industry provides value to the users only with respect to their consumption of other ICT services. The products and services themselves are of no value. This phenomenon also creates opportunities for new market entrants, increasing the rate at which companies are born.

Thirdly, current products and services can be improved based on new technology. Although one might use Microsoft Office on the HP laptop as one did 20 years ago, the quality of both the software and the hardware has increased substantially. Indeed, incumbent firms can defend their market share against the new entrants by embracing new technology and improving their product. They have a competitive advantage of having much in-house expertise and means for investment. This can possibly deter a portion of the eager entrepreneurs but it certainly affects the probability of their survival. If the incumbents are aware of the necessity of keeping up with progressing technology, they will either try to outperform the entrepreneurs or simply buy them. However, not all mergers are result of technological progress. The topic of mergers and acquisitions is further discussed in section 4.2. In either case, the survival rate of the recently created firms falls and the death rate rises.

It is not only the firm's product which requires improvement to maintain its competitiveness. As a general purpose technology, ICT can be used in numerous different ways to increase company's efficiency. Reorganizing its internal processes makes a firm more competitive. It increases the economies of scale which enable companies to grow or acquire others. As the result of creative destruction, the death rate of businesses rises. Increasing efficiency also works as an additional hurdle for the market entrants. Competitors in year t might be substantially less productive than competitors in year t+3. Entrepreneur may fail to foresee the change in other firms and thus underestimate her competition. This effect thus also reduces the survival rate of the newly created companies as it benefits the incumbents.

Lastly, ICT decreases transaction costs. Whether it is the costs of sharing information or rapidly decreasing transportation costs (Hummels, 2007), decline in these costs result in new supply chains. It is now extremely easy to use portals such as Alibaba to order goods from Southeast Asia without having knowledge of or representation in the market. Such reorganization of whole industries puts a great strain on those firms which are not adaptive enough but it also creates new gaps to fill for entrepreneurs. Moreover, diminishing costs of founding a business encourage trial and error methods in entrepreneurship. When it is easy to establish a business, more people will attempt to become entrepreneurs, more businesses fail and less of the nascent ones survive. Thus, lower transaction costs results in higher birth and death rates and lower survival rate.

4 Data Description

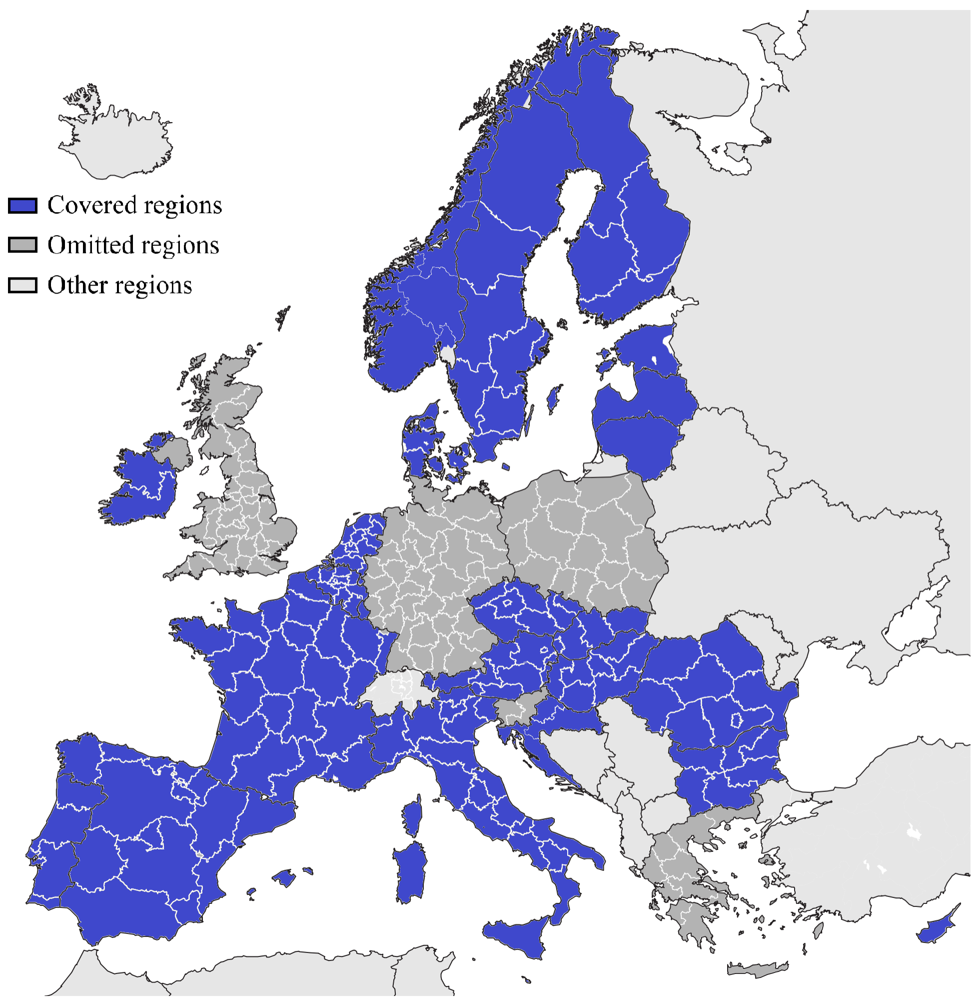
Both entrepreneurship and digitalization are clearly defined in theory, but the definitions are not suitable for empirical analysis. It is therefore necessary to use proxies. In case of entrepreneurship, previous studies used various variables such as number of newly founded firms (Wong et al., 2005) or percentage of working population engaged in entrepreneurial activity (Simón-Moya et al., 2014; Freytag and Thurik, 2007). Although both metrics are plausible proxies, this study uses the former. This is because the latter can be easily distorted by the number of self-employed workers. It is often the tax design rather than the entrepreneurial nature which makes people prefer being self-employed to having traditional employment contract. Moreover, data about business demographics captures the shifts in the industrial organization.

The case of digitalization is even more complicated. There is no single variable capturing this complex phenomenon. Some studies focusing on innovation in general take patents as their proxy (Wong et al., 2005). Others look at penetration of wireless or broadband technologies (Jensen, 2007; Crandall et al., 2007). Cardona et al. (2013) provide an overview of empirical literature regarding ICT and productivity and conclude it is usually ICT capital which serves as a proxy for advancing digital technology. However, neither proxy captures other aspects of digitalization. For example, digital literacy, network accessibility, or intensity with which digitalization transforms everyday lives are all left out, even though they contain valuable information about the transformative process. It is therefore conceivable to construct a composite index which can assess holistic impact of ICT transformation. Realizing that such effort can never be perfect and that there will always be important, yet neglected variables, a digitalization index is an attempt to capture as many aspects of digitalization as possible.

The single source of the analyzed data is Eurostat's regional database. This vast dataset ventures beneath the national level by creating comparable territorial units within national states. Instead of 28 observations, the NUTS2 classification provides data from 276 regions and hence substantially increases the inductive power of empirical analysis. Thus, all the data used in this study is region-specific. The shortcoming of the database is the relatively limited time dimension and numerous missing values. In an attempt to optimize the tradeoff between missing values and covered time period, a dataset spanning from 2008 to 2014 is used. It minimizes the proportion of missing values and at the same time exceeds the length of a business cycle. But it also captures a very specific time period as it reflects the global financial crisis triggered by the collapse of Lehman Brother in September 2008. Regionally, the data focuses on period of European debt crisis. Such exceptional time period is, among other things, expressed by low and sometimes even sub-zero interest rates. The low interest rates and their effect on the analysis is further discussed in section 9.

Unfortunately, such a dataset still contains many missing values which must be handled. Certain variables contain up to 25% of missing values, so case deletion is not feasible as the remaining dataset would be simply too small. However, the missing values can be imputed using multivariate imputation by chained equations algorithm (MICE) which employs predictive mean matching technique (PMM).[4] This method is effective and yields satisfying results for as much as 50% of missing values in the dataset (Schenker and Taylor, 1996; Raghunathan et al., 2001). Following these scholars, I use only regions with less than 50% of missing values. Such measure results in omitting three major countries (i.e. the UK, Poland, and Germany) and reduces the number of followed NUTS 2 regions to 167. The lack of data for those three countries is problematic, as they represent three distinct regions in Europe: the Anglo-Saxon liberal element, the German manufacturing powerhouse, and the largest soaring post-communist economy. Nonetheless, the regional dataset constructed in this manner still dwarfs datasets based on the national data as it can be seen in Figure 1 below.

Figure 1: Map of NUTS2 regions



Source: The map is an adapted version of Eurostat's blueprint for NUTS2 maps of Europe.

[4]The MICE algorithm with predictive mean matching is described in section 5.1.

4.1 Digitalization

The broad definition of digitalization provided in section 2.2 is not very suggestive on which components it is reasonable to analyze. The goal is to evaluate conditions in which digitalization may progress, the progress itself, and the results it eventually delivers. Therefore, I define three stages of successful digitalization process: fertile environment, widespread access, and adoption by users. ICT investment, widespread computer literacy, and human capital describe favorable breeding grounds for digitalization. Broadband, internet and mobile penetration capture the accessibility of the technology. Lastly, engagement in e-commerce, e-government, and frequency of internet usage picture the current state of digitalization.

This approach is introduced for the first time in this paper. Although it focuses on the dynamics of digitalization and its logic is different than that of Katz, it needs to be stressed that the individual components overlap and it is mainly their grouping which is different from Katz' approach. The benefit of the dynamic framework is that it depicts digitalization as a complex process whose components could not be easily disentangled. It is also possible to analyze different stages of digitalization and thus estimate its effects on business demographics in time. However, a composite indicator constructed out of the aforementioned components still closely mirrors digitization[5] index developed by Katz et al. (2014).

Digitalization is a complex process whose take-off requires opportune conditions. Firstly, it is vital that businesses invest in ICT. Without substantial investment into the communication networks, the digitalization process can hardly begin. Although it is only a proxy for the ICT investment across sectors, due to data availability, I use gross fixed capital formation in the ICT sector. Such measure includes non-ICT investment as well, but I assume that it nonetheless reflects the momentum of the ICT sector which, in turn, mirrors the overall ICT investment. In the analysis, I scale the investment in the ICT sector to local population and I use the harmonized index of consumer prices to adjust the data to inflation.[6] Secondly, labor force must have skills to develop and implement the advancing technology. Developing new digital solutions requires dedicated researchers and implementing these solutions is a competency of engineers. To control for the level of human capital, I use percentage of population employed in science and technology and proportion of labor force active in high-tech sectors. Lastly, the environment also depends on potential market size. People who feel very distant from technology are unlikely to be susceptible to new technological advancements. Computer penetration controls for the potential market size scaled by the total population: it measures percentage of individuals who have never used a computer.

[5]This is an example of confusion in the digitization vs. digitalization terminology. According to the definition provided by this paper, Katz' index would be called digitalization index.

[6]Because the ICT characteristics and capabilities change rapidly, using standard deflators such as CPI leads to underestimation of capital input. This could potentially cause overestimation of the ICT investment's effect on entrepreneurship. Such issues can conceivably be controlled for in the future by using, for instance, hedonic price index.

Having the technology in place is only the first step. The next one is to ensure it is widely accessible. To reflect different spheres of our life influenced by digitalization, I include three distinct measures of accessibility: broadband penetration, internet penetration, and mobile internet access. Internet penetration is the most rudimentary metric of access to the digital world. It does not capture quality of the connection nor its availability in every second of an ordinary day. On the other hand, broadband, a faster and more stable internet connection, can reflect the standards of the internet access. Similarly, mobile connection takes into account the ubiquity of internet connection.

Even if widely accessible, has digitalization really impacted significant aspects of people's behavior? I measure the degree to which digitalization permeates our everyday lives by focusing on three variables. Firstly, the proportion of population which ordered goods or services online during the last 12 months reveals shifts in consumer behavior. Secondly, the percentage of people who used internet for personal, civic, commercial, or political purposes in the last week captures the intensity with which we embrace the new tools. Lastly, if even the less flexible institutions adopt new technological solutions, it is reasonable to claim that one of the waves of digitalization has been completed. Therefore, the proportion of individuals who used internet to engage with public authorities in the last year serves as a proxy for government's attitude towards digitalization. The framework is presented in Figure 2.

Figure 2: Digitalization framework

Digitalization framework	
Stage 1: Environment	• ICT investment (gross capital formation in ICT sector per capita) • R&D labor (as % of population employed in science and technology) • Skilled labor (as % active population employed in high-tech sectors) • Computer access (as % of population)
Stage 2: Access	• Broadband penetration (as % of households) • Internet penetration (as % of households) • Mobile internet access (as % of population)
Stage 3: Adoption	• Internet usage (as % of population who used internet last week) • E-commerce (as % population who ordered goods or services online last year) • E-government (as % of population who used internet for communication with public authorities during last 12 months)

Source: The digitalization framework for regional analysis is original to this paper.

When estimating effects of digitalization, it is possible to either include its separate components or construct a composite index. Although inspecting separate variables gives us more detailed glance into the underlying relationship between digitalization and entrepreneurship, it misses the big picture. Policies enhancing digitalization hardly allow to cherry-pick phenomena caused by digitalization. For example, a market economy with broad internet access and educated population naturally shifts portion of its retail sector online because people increasingly engage in e-commerce, a clear business opportunity. It is thus advisable to look initially at the overall impact of digitalization and only thereafter analyze its components.

There are certainly other variables suitable for analysis of digitalization. Indeed, Katz proposed to include, among others, fixed line costs, international bandwidth, or broadband speed. The decision to omit these indicators is based on two reasons: suitability of the indicators for regional analysis and their availability. Although prices are a relevant factor, they are not region-specific. Data plans are usually uniform across countries or specific for every street. If there is no particular interest of analyzing the effect of the prices themselves, a simple national dummy controls for the quantifiable price differences. Other variables, such as bandwidth or speed of connection, might differ across regions, but to my knowledge, the data are not available at the regional level. Including the national aggregates would not benefit the analysis as even these components easily fall into the trap of national dummies. Moreover, using fixed effect panel data regression model controls for region-specific effect eliminating even the need to include national dummies.

4.2 Entrepreneurship

The focus of this study lies on the dynamic relation between entrepreneurship and digitalization. It is therefore reasonable to consider business dynamics rather than static, structural characteristics of the economy. For this purpose, a fitting approach is to mimic GEM total entrepreneurship activity metric and use business demographics indicators such as start-up rate, death rate, and survival rate after three years as proxies for entrepreneurship. This method avoids the issues with self-reporting which is inherent to the GEM survey data. Eurostat gathers all the metrics; it provides regional data about business demographics containing annual business deaths, business formation, and the number of survivals (companies still in business after 3 years).

Trying to describe shifts in entrepreneurial environment, one cannot miss the rate at which firms are created. The rate is obtained by taking the total of new businesses founded within one year and dividing it by their total. The variations in birth rate can be induced by different underlying factors, such as changing economy, gradual cultural transformation, or novelties in corporate law. But it is also advancing technology which creates opportunities for new businesses to rise and that is the effect which this study attempts to uncover.

Dynamic business environment also entails firms which cannot cope with the change and thus go out of business. I measure the rate at which firms depart from the market by taking the total of businesses having ceased their operations within one year and dividing it by the total of active businesses. Apart from natural turnover, death rate captures varying intensity of creative destruction. A natural objection is that mergers and acquisitions distort such measures. High number of mergers would artificially inflate death rate suggesting consolidation of industry rather than creative destruction. However, a brief look at the data reveals that it is not the case. In 2013, more than 2.3 million European businesses went under whereas there were only around 15,000 mergers taking place in Europe (Eurostat and IMAA statistics). It is impossible to include them in the analysis because, to my knowledge, no comprehensive regional data on mergers and acquisitions exists. Fortunately, its effect is very small; the total of mergers and acquisition in Europe is a mere quarter of standard deviation of European business deaths during the measured period. Although mergers and acquisitions are most likely not uniformly distributed across regions, they are linked to the overall entrepreneurial activity and thus are likely to represent a comparable share of total businesses within a region. Hence, it should not significantly affect the analysis.

Similarly, birth rate and survival rate can be artificially inflated by strategic splitting of existing firms and spin-offs. Although regional or aggregate data on spin-offs are not available, I do not suspect this metric's impact to be considerably greater then the number of mergers and acquisition. Moreover, a portion of spin-offs is caused by advancing technology, so it would not be advisable to exclude them anyway. Restructuring of a business which translated into its splitting might be a consequence of advancing digitalization. Spin-offs, on the other hand, are often an entrepreneurial effort of larger companies. As such, they both fall into this study's analytical scope.

Birth and death rate can be combined into net birth rate revealing the general trend of the explanatory variable's contribution. Does digitalization lead to industry consolidation or rather industry fragmentation? Interesting as this question may be, a regional analysis using short time series cannot provide definite answers. There is a substantial noise in the data caused by volatility of short-term, unsuccessful projects which often come in waves due to economic bubbles. To avoid seasonality affecting the analysis, longer time span would have to be analyzed. More importantly, such estimates are likely to be dependent on industry which makes an aggregate analysis obsolete. Therefore, this study does not include churn rate as a dependent variable and leaves it for future research.

Spikes in birth and death rates suggest extraordinary buzz on the start-up scene but they do not have to lead to structural changes. Valuations of some nascent businesses and the amount of venture capital available create a huge incentive for all newcomers. But are the new businesses successful? Do they transform the economy as they outgrow the incumbents? Although the impact of individual firms and the rate at which they grow is not measured, I include the number of recently born firms which successfully survived the first three years.

Having survived the crucial first three years, a firm is in a good position to operate on and affect the market. Whereas death rate is a proxy for the destructive element of the Schumpeterian entrepreneur, survival rate is the long-term creative component of entrepreneurship, cleansed from possible whims and failures of both investors and entrepreneurs. Furthermore, I assume survival rate not to be affected by herd behavior of the aforementioned agents.

4.3 Control variables

Estimating digitalization's effect on business demographics requires controlling for other causes of business creation and destruction. I consider economic factors, formal institutions, culture, and education as Simón-Moya et al. (2014) have proven all these aspects to be significant drivers of entrepreneurship. It is important to keep in mind that many metrics normally serving well as controls are uniform within a country and thus can be, among other factors, captured by country-specific dummy or fixed effect regression. Hence they are not suitable for the regional analysis which is the focus of this paper.

Economic factors appropriate for regional analysis are gross domestic product per capita and unemployment rate.[7] Because the interest of this study lies solely with Europe, we can only consider high growth potential entrepreneurship. Entrepreneurship in Europe has rarely the characteristics of necessity-driven entrepreneurship because of Europe's widespread social security and her high level of economic development. That enables me to follow Simón-Moya et al. (2014) and assume that both regional employment and regional GDP per capita positively affect entrepreneurial activity within the respective region which means they are both viable controls representing the regional economic conditions.

Formal institutions within regions are difficult to measure. I use public investment as a proxy for government quality and number of submitted patents to reveal ingenuity of the region. Public investment per capita is relatively straightforward way to control for the state of formal institutions (Afonso et al., 2005). The implicit assumption is that higher public investment is associated with legislative environment benign to entrepreneurial activity. The number of submitted patents weighted by the active population in millions then reflects how well the intellectual property rights are enforced. It is only when the innovator believes that her rights will be protected, that she undergoes often arduous process necessary for a patent being granted to her. However, patent submissions also capture the entrepreneurial spirit of a region. I scale the data as provided by Eurostat by population of the region and I use 2005 prices.

Quantifying regional culture affecting entrepreneurship is problematic but conceivable.

[7] Eurostat follows the guideline of International Labour Organization defining unemployed person as someone aged 15 to 74, without work during the reference week, available to start within the next two weeks, and actively having sought employment at some time during the last four weeks. The unemployment rate is the number unemployed people as fraction of the labor force.

Cultural heritage represented by community trademarks weighted by GDP in EUR billion is tracked by the EU, and is thus a possible proxy for cultural environment of the region. It does not capture the culture per se; it rather tracks its consequences. In a culturally rich environment for entrepreneurship, more innovative and unique products are developed and thus also more trademarks are registered. Culture is not easily changed and if the region displayed extraordinary entrepreneurial spirit in the past resulting in various trademarks, it is likely to have kept such cultural trait.

Lastly, I use proportion of individuals who have completed secondary education to control for education as the last pillar of factors of entrepreneurship. Completion of secondary or tertiary education is often a prerequisite for obtaining certificate allowing participation in certain vocations. But secondary education, compared with tertiary education, is also less correlated with measures of skilled and R&D labor, thus bringing more information into the analysis. It is also helpful for identifying opportunities on the market (Clercq and Arenius, 2006). Moreover, following the Kirznerian perspective, higher level of education increases the pool of possible entrepreneurs who are able to identify not just opportunities on the labor market, but also business opportunities.

Table 1 presents summary statistics of all used variables. These statistics have been calculated before the logarithmic transformation of the data and before the data imputation.

Table 1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Broadband penetration (in %)	984	63.15	17.98	9	97
Internet penetration (in %)	992	69.39	16.72	17	99
Mobile access (in %)	451	43.65	20.26	3	85
Computer access (in %)	1,000	76.70	15.25	37	100
E-commerce (in %)	999	36.32	22.35	1	84
E-government (in %)	601	46.05	21.20	3	89
Internet usage (in %)	999	66.05	18.06	22	98
R&D labor (in %)	1,155	3.35	1.59	1.00	9.50
Skilled labor (in %)	1,110	3.44	1.85	0.60	10.10
ICT investment (in EUR per capita)	834	2.08	2.69	0.01	18.25
Birth rate (in % of total firms)	808	9.90	3.22	4.83	26.11
Death rate (in % of total firms)	642	8.96	4.91	4.14	65.02
Survival rate (in % of total firms)	804	5.51	1.32	2.91	11.82
GDP per capita (in EUR)	1,141	25,636.81	13,450.19	3,100	87,600
Unemployment (in %)	1,169	9.61	5.82	1.80	37.00
Public investment (in EUR per capita)	834	9.05	6.33	0.41	39.33
Submitted patents (per population)	810	179.20	202.09	0.40	1,399.16
Cultural trademarks (per GDP)	1,129	4.56	4.05	0.05	46.45
Secondary education (in %)	1,169	70.51	15.05	18.00	97.30

Source: The data comes from the Eurostat's regional database. Further computations are conducted by the author.

5 Methods

5.1 Data imputation

Even after deletion of regions with more than 50% of missing values, the occurrence of missing data points is still high. To maintain as much information in the data as possible, I do not delete any other observations. Indeed, Little and Rubin (1987) suggest that if more than 5% of the data is missing, imputation of the missing values is advisable. The data imputation is based on the observed data using multivariate imputation. Although it is also possible to use mean, median, or Monte Carlo simulation, mean and median imputation is quite crude and leads to substantial bias in the imputed data and Monte Carlo requires processing power which is beyond what the author of this paper possesses. The multivariate imputation by chained equation method (MICE) algorithm, on the other hand, exploits interdependencies in the data and predicts missing values based on the known parameters in each observation and is feasible to be run on a desktop computer. Specifically, predictive mean matching (PMM) estimates the coefficient describing linear dependency of the imputed and all the other variables. This is done using only complete observations. Based on the coefficient, for each missing value, a set of observed values is constructed such that their predicted values are close to the predicted value for the case with missing data. From this set, one value is randomly chosen as a substitution for the missing one.

The method has several benefits. Because it draws the imputations from the actual observations, one does not have to worry about predictions being out of feasible range. Its random element also enables to repeat the process several times and pool the individual estimates into one estimate robust with respect to the random element of the data imputation process. Indeed, many studies have shown validity of PMM and its comparable advantage with respect to other methods of data imputation. The actual data imputation algorithm used in this paper is MICE developed specifically for R by van Buuren and Groothuis-Oudshoorn (2011). The algorithm can be described in 4 steps:

1. A variable to be imputed as the first one is chosen. The missing values in the rest of the dataset are simply imputed by the mean values.

2. Using imputation model of choice (PMM in this case), the first variable is imputed based on the rest of the dataset.

3. Using the variable which was imputed as the first as independent variable, all the other missing values in the dataset are imputed. Hence both the observed and imputed data points are used in subsequent imputations.

4. Keeping these data points, steps 1–3 are repeated number of times (80 times in this case). This process of numerous iterations ensures robustness of the MICE method.

Only the last iteration is saved as one imputed dataset.

One of the benefits of MICE is its ability to preserve both the relations and the uncertainty in data. The process itself is also very efficient; it often achieves satisfying convergence only after 10 iterations (van Buuren et al., 2006). Although several problems may arise when employing MICE with PMM, including circular dependence and unfeasible imputations, the structure of the dataset allows to deal with such shortcomings. In this setting, it is therefore plausible to use the PMM method for every variable even though the design of MICE allows for variation in methods for each imputed variable.

Following van Buuren and Groothuis-Oudshoorn (2011), all used variables are included in the imputation process. High number of predictors counterintuitively decreases bias and it also makes the MAR assumption (missing at random) more plausible. The only risk is multicollinearity which is not acceptable for multivariate imputation. Variables causing multicollinearity must be omitted in the imputation process. However, in the working dataset, such issues have not occurred and thus no variable had to be eliminated. Although the MICE algorithm deals with both MAR and NMAR (not missing at random) data, NMAR characteristic increases the probability of explosive behavior. When imputing the data, the algorithm creates mutual causality which yields explosive results as we increase the number of iterations. That is why it is necessary to investigate the convergence of the imputed variable. To avoid this issue, an adjusted prediction matrix was used such that internet access is not a predictor of broadband access, e-government is not a predictor of e-commerce and broadband access is not predictor of internet usage. Removing strong predictor from the regression prevents the explosive behavior (van Buuren and Groothuis-Oudshoorn, 2011) since high correlation between the pairs can cause undesirable explosiveness. Using the adjusted prediction matrix, the behavior of the variables follows the expected pattern resembling the white noise. Certain variables converge only after 20 iterations which is to be expected as the dataset is quite large and contains many missing values (van Buuren and Groothuis-Oudshoorn, 2011). In the subsequent iterations, we observe neither trend, nor strong path-dependency. Indeed, the 5 distinct imputation processes intermingle with each other and follow a stationary pattern (see Figures 5, 6 and 7 in the appendix for visualization of the imputation process).

Imputation of missing values can be extremely useful but there are shortcomings of which an imputer should be aware. It creates an additional layer of complexity which can often undermine credibility of the results. It also operates with the MAR assumption. Because we do not observe the counterfactual, this assumption cannot be tested. However, including many variables increases the likelihood of MAR assumption's sensibility. Another pitfall is the choice of the imputation model. Although I choose PMM because it yields reasonable results for variables which are expected to be within certain interval (proportional penetrations are measured in percentages) and it also provides uncertainty in the imputation process, the choice is arbitrary. There is no consensus among the researchers about the superiority of certain model over others.

There are cases in which the MICE algorithm simply cannot be used. For instance, if a categorical variable should be used to impute a continuous one and vice-versa, MICE cannot yield sensible results. That is because the joint distribution of these variables does not exist and hence it always depends on which variable is imputed as the first one. Fortunately, the imputation process in this study does not include categorical variable and this trap is therefore avoided. Still, a decision about the number of imputations must be made. White et al. (2010) show that increased number of imputations make the estimates more accurate. However, the decision about the number of imputations also depends on the subsequent analysis, so the rule of thumb which White et al. developed is not generally applicable. Most studies use less than 30 imputations (Raghunathan et al. (2001) use 25, Rubin (1996) sticks with 20). To be conservative, I use 80 imputations in this study.

Acknowledging the potential dangers of data imputations, the method is nonetheless of great value when analyzing the regional data. It is the only way how to proceed with the study given the dataset. Analyzing the European regions comes also at a price of reduced flexibility in the choice of variables. Hence I must take the Eurostat's database as given and draw as much information from it as possible. The MICE method keeps all the information in the data and that is why it was chosen. The alternative would be partial data deletion. However, because of (to some degree) uniformly distributed missing values, there would be no complete observation to work with given the current variable setup. Limiting the number of variables would also be possible but I would not be able to holistically capture digitalization and only randomly choose proxies of the process. The last alternative would be to scrap the regional analysis altogether and work with national data. However, I believe there is a great value in delving beneath the national level and I thus proceed with the data imputation.

5.2 Index construction

Having dealt with the missing values in the raw dataset, the next step is to analyze the digitalization data and construct the digitalization index. Composite indices such as digitalization index enable aggregation of complex phenomena, which are then easier to interpret than a set of many distinct variables.[8] However, their construction is subject to many arbitrary decisions of the researcher such as choice of variables and weight-setting. The OECD *Handbook on Constructing Composite Indicators* (Nardo et al., 2005) offers thorough guidance for construction and use of composite indicators and addresses challenging obstacles to composite index creation.

[8] The discussion about suitability of composite indicators in economic research and policy-making is far from settled. For an overview of the major arguments, see Saltelli et al. (2005)

After choosing the set of variables included in the composite indicator, the researcher must analyze the multivariate structure of the variables. It reveals hidden interrelationship between them and enables to identify redundant variables. Multivariate analysis has also significant implications for weighting and aggregation (Nardo et al., 2005). Principal component analysis is one of the tools fit for such task; it transforms the correlated variables into several principal and uncorrelated components. The pitfall of the principal component method is that it ignores real influence of the variables. Even two highly correlated variables might have both distinct influence on the underlying phenomenon and thus should be separately included in the index. Moreover, such method is not suitable for small datasets.[9] However, the digitalization dataset is quite large and the choice of variables is based on their economic rather than statistical significance.

Although most indices rely on equal weighting, it is mostly the lack of consensus on alternatives than its superiority which makes it wide-spread (Nardo et al., 2005). Its implicit assumption of equal importance is as arbitrary as any other set of chosen weights. With highly correlated variables, equal weights also cause certain factors to be counted multiple times. Advanced methods such as multiple regression or benefit of doubt method (Cherchye et al., 2007) could be employed for assigning weights, but more comprehensive and more common way is the principal component analysis. Weights based on principal components group the variables, so that overlapping information is not counted more than once (Nardo et al., 2005). When constructing the digitalization index as well as indices for each stage of digitalization, weights are thus based on the principal component analysis. However, it should be noted that the weights obtained by the principal components do not substantially vary from equal weights.

Principal component analysis reveals that only two components explain 87.5% of the variance in digitalization index. Choosing two principal components adheres to Kaiser criterion which suggests to include components with eigenvalue greater than one,[10] as well as to variance explained criterion which advises to include components explaining at least 80% of variance (Nardo et al., 2005). These criteria serve as a mere rule of thumb when deciding on the number of included principal components, so I perform a robustness analysis with equal weights in section 8 of this paper. Luckily, the results stay largely intact.

Construction of the index also requires normalization of the data. The optimal method depends of the data structure. For example, ranking technique deals well with outliers but absolute values get lost. The Min-Max method, on the other hand, normalizes the data into range [0,1]. As this method fits the data structure into small interval, it amplifies its overall effect compared to the ranking technique. Consequently, outliers cause severe distortion within

[9] There is no scientific measure of what small dataset is. Rules of thumb are often used such as the rule of 10 (10 observations for each variable) or 3:1 ratio (the observations-to-variables ratio is higher than 3). For more details see Nardo et al. (2005). The dataset used in this study meets all mentioned requirements.

[10] Dropping components with eigenvalues below one means to omit components explaining less variance than a single average variable.

the Min-Max technique. Standardization of the data to distribution with zero mean and standard deviation of one sustains the inherent data structure in better way than the previous methods, but causes variables with more outliers to have a greater effect on the composite index (Nardo et al., 2005). Nonetheless, to maintain the data structure in the aggregation process, normalization of the data into z-scores is used in this study.

The last step in composite index construction is aggregation. Linear aggregation is suitable for variables with the same units of measurement and it also allows compensability, i.e. low score in one variable can be easily compensated with higher score in another variable. Geometric aggregation is not so favorable toward compensability which leads to high marginal utility from addressing lagging factors of the composite index. Country, region, or any subject of the composite index would then be keener to focus on those aspects to improve its ranking. Multi-criteria approach allows non-compensability but is computationally very demanding, especially with high number of countries, regions, or other subjects of interest (Munda and Nardo, 2009). Because the aim of this index is not to analyze marginal contributions of its components, linear aggregation is used to ensure transparency and understandability to the reader.

5.3 Regression analysis

For the purposes of regression analysis, rescaling of the data is advisable. This way, the estimated coefficients can be easily interpreted. Hence, I use logarithmic transformation of all dependent and control variables. Unfortunately, it is not possible to use logarithmic transformation of the digitalization index, as it is normalized to standard normal distribution and thus contains negative values. A different transformation is necessary. Keeping in mind that there is no transformation which would be scale invariant, I choose such transformation, so that even the index has an intuitive interpretation.

I rank all the regions based on the digitalization index and assign percentiles to each of them. This way a regression coefficient obtains a straightforward interpretation. If a region increases its relative digitalization score so that its ranking increases from 50^{th} percentile to 51^{st} percentile, we expect 100 times beta percent increase in the explained variable given it has been a subject of logarithmic transformation. Keeping the percentile measure in interval [0,1], the coefficient suggests beta percent increase in explained variable with each gained percentile point. Moreover, for easier readability, the data presented in tables is scaled up by factor of two. Hence, as presented in the tables, increase by 100 percentile points is associated with beta percent increase in the explained variables. The same transformation is used for different stages of digitalization. Lastly, in regressions where no composite indicators take place of explanatory variables, natural logarithm of all the variables is taken.

As hinted before, many omitted variables have national-specific characteristic. Therefore, I include a set of dummies in the explanatory variable set to control for those effects. To avoid the dummy variable trap, a dummy associated with Belgium is contained in the intercept. Naturally, those dummies are omitted in the fixed effect model. However, they must be present in both pooling model and random effect model, as their omission would dramatically decrease the power of the models.

In order to test the suspected relationships, I specify several regression models. Birth, death, and survival rates are always the dependent variables as I aim to reveal the effect of digitalization on entrepreneurship. The set of controls introduced in section 4 is present in every single regression model. What varies, however, are the explanatory variables. In the first step, only the aggregate digitalization index is used as an explanatory variable revealing the overall, long-term effect of digitalization. In the second step, the digitalization index is split into three stages to shed light on the dynamics of digitalization's effect on entrepreneurship. In the third, last step, the index is completely disaggregated, so that significance of each component can be established and evaluated.

The regression equations are thus:

$$\begin{aligned} log(birth\ rate) = {} & \beta_0^1 + \beta_1^1 \cdot Digitalization + \gamma_1^1 \cdot log(GDP\ per\ capita) \\ & + \gamma_2^1 \cdot log(Unemp.) + \gamma_3^1 \cdot log(Public\ inv.) + \gamma_4^1 \cdot log(Sub.\ patents) \\ & + \gamma_5^1 \cdot log(Cult.\ trademarks) + \gamma_6^1 \cdot log(Secondary\ edu.) \end{aligned}$$

$$\begin{aligned} log(birth\ rate) = {} & \beta_0^2 + \beta_1^2 \cdot Stage\ 1 + \beta_2^2 \cdot Stage\ 2 + \beta_3^2 \cdot Stage\ 3 \\ & + \gamma_1^2 \cdot log(GDP\ per\ capita) + \gamma_2^2 \cdot log(Unemp.) \\ & + \gamma_3^2 \cdot log(Public\ inv.) + \gamma_4^2 \cdot log(Sub.\ patents) \\ & + \gamma_5^2 \cdot log(Cult.\ trademarks) + \gamma_6^2 \cdot log(Secondary\ edu.) \end{aligned}$$

$$\begin{aligned} log(birth\ rate) = {} & \beta_0^3 + \beta_1^3 \cdot log(ICT\ inv.) + \beta_2^3 \cdot log(R\&D\ lab.) + \beta_3^3 \cdot log(Skilled\ lab.) \\ & + \beta_4^3 \cdot log(Computer\ acc.) + \beta_5^3 \cdot log(Broadband\ acc.) \\ & + \beta_6^3 \cdot log(Internet\ acc.) + \beta_7^3 \cdot log(Mobile\ acc.) \\ & + \beta_8^3 \cdot log(E\text{-}commerce) + \beta_9^3 \cdot log(E\text{-}government) \\ & + \beta_{10}^3 \cdot log(Int.\ usage) + \gamma_1^3 \cdot log(GDP\ per\ capita) \\ & + \gamma_2^3 \cdot log(Unemp.) + \gamma_3^3 \cdot log(Public\ inv.) + \gamma_4^3 \cdot log(Sub.\ patents) \\ & + \gamma_5^3 \cdot log(Cult.\ trademarks) + \gamma_6^3 \cdot log(Secondary\ edu.) \end{aligned}$$

The regression equations for death rate and survival rate are analogical.

Estimating the model with imputed data might lead to imprecise standard errors. Therefore, following Rubin (1996), I construct 80 in parallel imputed datasets, run the regressions with each of those datasets and then pool the results into a single estimate. Using Rubin's notation, let us assume $\hat{Q}_m$ is the estimate of parameter Q computed from the m^{th} imputation. In this case, Q is simply a regression coefficient. Then the repeated-imputation estimate of Q is:

$$\bar{Q}_M = \frac{1}{M} \sum_{m=1}^{M} \hat{Q}_m \tag{1}$$

Variance T_M of $\bar{Q}_M$ is:

$$T_M = \bar{U}_M + (1 + M^{-1}) B_m \tag{2}$$

where U is the within-imputation variability (taken from variance-covariance matrix of the estimation) and B is the between-imputation variability:

$$\bar{U}_M = \frac{1}{M} \sum_{m=1}^{M} U_m \qquad and \qquad B_M = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{Q}_m - \bar{Q})^2 \tag{3}$$

As we increase the number of imputations, Q converges in distribution to:

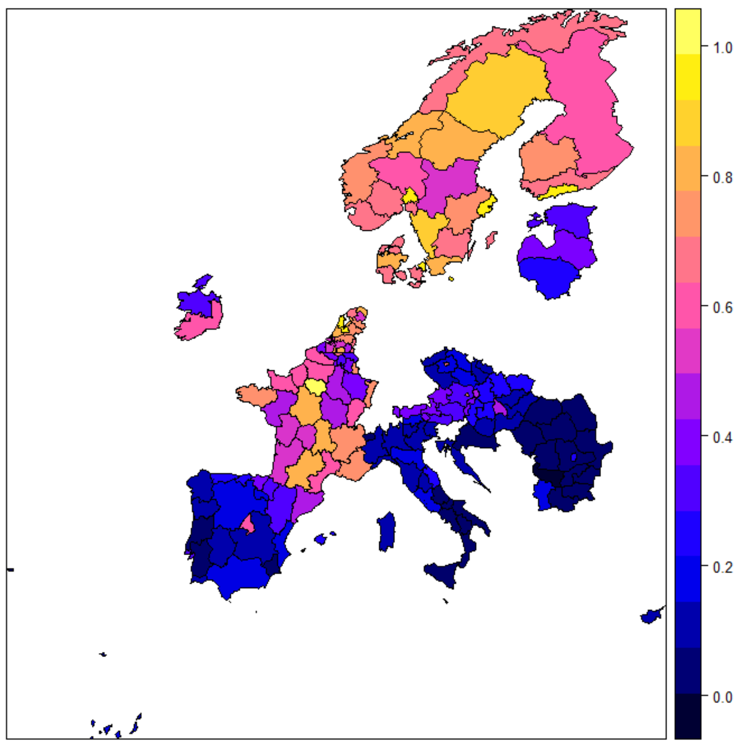
$$Q \sim N(\bar{Q}_\infty, T_\infty) \tag{4}$$

P-values of the estimates can thus be calculated accordingly.

6 Results

The digitalization index construction yields foreseeable results. The Benelux, the Nordic countries, and France are the forerunners of digitalization in Europe. Surprisingly, the Southern states are virtually parring with Central Europe and the Baltic countries. The Southeastern Europe, as suspected, represents the laggards in digitalization. Additionally, regions containing capital cities score significantly better then the rest, e.g. Madrid, Paris, or Prague. Figure 3 below shows the detailed digitalization ranking in 2014. Figure 8 in the appendix presents average performance of individual countries.

Figure 3: Relative ranking in digitalization, 2014



Source: The map is produced by the author using Eurostat's data.

Due to panel characteristics of the data, I choose from pooling, fixed effect, and random effect models. Firstly, the Chow test for the poolability of the data is conducted. This F-test tests stability of coefficients between pooling and fixed effect model. Had not the hypothesis of stability of coefficients been rejected, a simple pooling model would be used for estimation. However, I was able to reject the stability hypothesis in every single case. Hence I move on to a Lagrange multiplier test (the Breusch-Pagan version) to determine whether there is individual and/or time effect in the data. Having properly defined the fixed effect model, the Hausman test decides whether to use the fixed effect model or the random effect model. It should be noted that time dummies are included in all fixed effect models except those with death rate as the explained variable. Including time dummies in these models (which is not suggested by the Breusch-Pagan test) only slightly affects the results and does not alter the statistical significance of the estimates.[11]

As a starting point, I estimate a holistic effect of digitalization of birth, death, and survival rates on businesses. Each component might affect entrepreneurial activities in different way which undermines the statistical significance of the composite indicator. Nonetheless, it is important to evaluate digitalization's overall effect, because, in practice, it is not possible to pick and choose only certain aspects of digitalization. Table 2 shows the regression results based on 80 independently imputed datasets. The tests were not conclusive about superiority of any particular model, so I present results from both fixed effect and random effect regressions to provide consistency throughout the paper. However, the model suggested by the Hausman test at 5% significane level is marked by bold font in the tables. When the hypothesis of insignificant time effect could be rejected, I have included time dummies in the regression. The regressions were run 80 times with 80 imputed datasets to ensure robust and statistically coherent results.

[11] Including time dummies in the random effect models is not possible because the estimated variance is negative. The random effect models thus rather serve as a robustness check in this analysis even though the Hausman test occasionally suggests their validity.

Table 2: Digitalization index regressions

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	FE (3)	**RE** (4)	FE (5)	**RE** (6)
Digitalization	13.1	13.12**	-21.87	-31.34	-17.57**	-16.93***
	(9.66)	(6.45)	(70.27)	(66.05)	(8.68)	(5.57)
GDP per capita	-25.44**	-0.82	67.45	102.18*	20.38*	11.29***
	(11.71)	(5.03)	(106.2)	(59.4)	(11.7)	(4.23)
Unemployment	0.15	6.13**	31.95	48.81**	-8.03**	2.45
	(4.21)	(2.39)	(29.52)	(24.42)	(3.59)	(1.92)
Public investment	4.05	4.87**	-24.02	-21.31	0.44	3.7*
	(3.14)	(2.34)	(25.94)	(23.69)	(2.86)	(1.93)
Submitted patents	-2.55	-3.22**	18.95	23.12*	-2.62	-1.73
	(1.97)	(1.54)	(14.39)	(13.25)	(1.61)	(1.15)
Cult. trademarks	-0.67	-0.25	22.27	22.52	0.13	1.12
	(1.88)	(1.35)	(17.18)	(14.4)	(1.76)	(1.11)
Secondary edu.	26.71*	-12.62	-70.48	-55.48	-17.55	-18.93***
	(16.23)	(8.06)	(152.84)	(108.91)	(14.18)	(6.25)
Time dummies	Yes	No	No	No	Yes	No
Observations	1,169	1,169	1,169	1,169	1,169	1,169
Imputations	80	80	80	80	80	80

Note: *p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect
Source: The calculations are done by the author, the data comes from the Eurostat's regional database.

As hypothesized, digitalization has negative effect on survival rate. An increase in percentile ranking of a region by 10 points can be associated with decrease in survival rate of nascent businesses by 1.7%. The economic significance of the estimate is considerable. Based on the estimates, we would expect the least digitalized region in Europe to have 17% lower survival rate compared to the most digitalized region. For an average NUTS2 region in term of its business population, this effect translates into more than 1,800 firms not surviving the first three years due to digitalization. The estimated effects on birth rate is positive as hypothesized but it is not statistically significant. The effect of digitalization on death rate is not conclusive at all. The weak statistical significance can be attributed to the composite characteristics of the index: there are simply different forces within digitalization pushing against each other. I thus run a second regression where I look into different stages of digitalization (see section 4.1 for description of the stages).

Table 3: 3 stages of digitalization

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	FE (3)	**RE** (4)	FE (5)	**RE** (6)
Stage 1	4.86**	9.34***	8.43	18.46	4.5***	4.62***
	(2.45)	(2.63)	(17.9)	(16.57)	(1.49)	(1.61)
Stage 2	8.46***	6.09**	9.11	1.91	-10.53***	-10.57***
	(2.85)	(2.47)	(20.57)	(20.81)	(2.24)	(2.11)
Stage 3	-9.04**	-13.16***	-32.38	-36.03	-3.37	-3.31
	(4.33)	(3.85)	(30.67)	(30.24)	(3.31)	(3.42)
GDP per capita	-27.63**	-5.4	67.82	89.69	9.84**	8.94**
	(11.63)	(5.34)	(106.71)	(60.07)	(4.02)	(4.09)
Unemployment	-0.16	6.6***	34.82	55.56**	5.41***	6.39***
	(4.2)	(2.29)	(31.75)	(25.75)	(1.97)	(1.85)
Public investment	4.75	4.53**	-24.44	-23.36	1.56	2.08
	(3.18)	(2.3)	(26.03)	(23.72)	(1.89)	(1.82)
Submitted patents	-2.56	-3.6**	20.38	24.27*	-1.96*	-1.86*
	(1.98)	(1.5)	(14.34)	(13.16)	(1.05)	(1.02)
Cultural trademarks	-0.82	-0.06	22.56	23.52	2.03*	2.35**
	(1.86)	(1.28)	(17.38)	(14.5)	(1.04)	(1.01)
Secondary education	32.79*	-10.85	-55.85	-43.6	-11.94**	-12.25**
	(16.84)	(8.16)	(147.3)	(103.9)	(5.93)	(5.79)
Time dummies	Yes	No	No	No	Yes	No
Observations	1,169	1,169	1,169	1,169	1,169	1,169
Imputations	80	80	80	80	80	80

Note: *p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect
Source: The calculations are done by the author, the data comes from the Eurostat's regional database.

Table 3 shows the outcome of regressions where the digitalization index is split into three components describing its dynamics. The first stage confirms my hypotheses regarding birth rate and death rate; both are positively affected by environment fit for digitalization. However, the initial positive effect on survival rate is counter-intuitive. One plausible explanation is that the new firms are faster at adopting the new technology which gives them, at least in the initial stage, a competitive advantage over the incumbents. But, as digitalization progresses, these effects reverse or vanish. This dynamic phenomenon explains the diminished significance of digitalization as a whole and also indicates short-term characteristics of the digitalization process, especially with respect to birth and survival rates.

Death rate does not seem to be affected and shows no significant dynamic behavior. This suggests that digitalization does not accelerate the process of Schumpeterian creative destruction, or at least its destructive part. Since the rise in birth and survival rates in the initial stage of digitalization is not accompanied by rising death rate, digitalization seems to initially inflate the market rather than intensify the competition. It produces more opportunities but does not necessarily make the current business models obsolete.

Having newly available technology and suitable environment for its development attracts new entrepreneurs who seek to come up with new products and services. That is one of the possible explanations of why rising business birth rate can be observed in the initial stage of digitalization. Surprisingly, this is not accompanied by steep rise in death rate as active businesses implement the new technology to improve their products and reorganize their internal processes, which intensifies competition both among the incumbents and the new entrants. This suggests an overall fragmentation of industries facing the conditions of highly skilled labor, technically literate population and ICT investment.

As the technology gets widely accessible, more firms continue to get born, but the effect on death rate is still insignificant. Thus, with more potential customers, the net growth in business population is positively influenced by the "access" stage of digitalization. However, the effect on survival rate inverses with respect to the previous stage, fitting the theoretical framework presented in section 3. The reaction of the incumbents seems to be strong enough to defend them against the new entrants. The new firms, on the other hand, may succumb to herd behavior and start businesses too optimistically and without proper rationale.

The third stage of digitalization describes maturing of the technology as more people adopt it and benefit from it. The results suggest that with the final stage, there is a general tendency of offsetting previous changes induced by the technology. Indeed, with saturation of the market, birth rate drops to partially compensate for its initial gain. Death and survival rates tend to follow similar pattern, but they lack statistical significance. This general trend is attributable to the fact that the most accessible benefits of digitalization have been exhausted. Indeed, as the low-hanging fruit is gathered, the dynamics of entrepreneurship brings it back to its state before the digital transformation.

The analysis per stage presented in Table 3 also confirms the results from the composite index analysis. Digitalization has a long-lasting effects. If a region performs similarly in each stage compared to other regions, we expect aggregate effects not dissimilar to the results of the first regression. But because of the complexity of weighting based on principal components, it is not possible to simply add up the effects of each digitalization stage. What is encouraging though, is the statistical significance of the variables after the split. That indicates that its statistical insignificance can indeed be caused by its components cancelling each other out.

Abstracting from any form of aggregation, Table 4 presents results of regressions including all the explanatory variables. All the variables are logarithmically transformed, hence the interpretation: $\%\Delta y = \beta \cdot \%\Delta x$. The results are again in line with the outcome of previous regressions. However, because the variables are not aggregated, it hints at what the magnitude of the effects might be in absolute terms (compared with the previously presented coefficients which were tied to the relative ranking of a region). The economic significance is often substantial. For example, 1% increase in internet access is associated with 3% decrease in death rate.

Table 4: Individual components of digitalization

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	**FE** (3)	RE (4)	**FE** (5)	RE (6)
ICT investment	0.07***	0.07***	0.05	0.05	0.01	0.01
	(0.02)	(0.01)	(0.15)	(0.13)	(0.01)	(0.01)
R&D labor	-0.05	-0.04	0.36	0.36	0.06**	0.04
	(0.04)	(0.03)	(0.4)	(0.35)	(0.03)	(0.03)
Skilled labor	0.07*	0.08***	0.31	0.5*	0.04**	0.05**
	(0.04)	(0.02)	(0.38)	(0.29)	(0.02)	(0.02)
Computer Access	-0.46**	-0.33*	1.91	2.18	-0.16	-0.18
	(0.22)	(0.19)	(1.99)	(1.91)	(0.15)	(0.15)
Broadband Access	0.2***	0.11*	1.5**	1.66***	-0.21***	-0.22***
	(0.07)	(0.06)	(0.63)	(0.6)	(0.05)	(0.05)
Internet Access	-0.26*	-0.05	-2.82**	-3.28***	0.26**	0.3***
	(0.14)	(0.12)	(1.24)	(1.19)	(0.1)	(0.1)
Mobile Access	0.01	0.03	-0.05	-0.06	0.04	0.03
	(0.04)	(0.04)	(0.3)	(0.31)	(0.03)	(0.03)
Internet usage	0.34*	0.16	-1.17	-1.28	-0.42***	-0.39***
	(0.18)	(0.16)	(1.59)	(1.51)	(0.13)	(0.13)
E-commerce	0.06	-0.01	0.13	0.11	-0.06**	-0.07**
	(0.04)	(0.03)	(0.33)	(0.31)	(0.03)	(0.03)
E-government	-0.1***	-0.09***	-0.27	-0.26	0.09***	0.09***
	(0.04)	(0.04)	(0.23)	(0.24)	(0.03)	(0.03)
GDP per capita	-0.31**	-0.15***	0	0.28	0.06	0.06
	(0.12)	(0.05)	(1.23)	(0.69)	(0.04)	(0.04)
Unemployment	0	0.05**	0.38	0.55**	0.03	0.04**
	(0.04)	(0.02)	(0.32)	(0.26)	(0.02)	(0.02)
Public investment	0.05	0.04	-0.18	-0.17	0.01	0.01
	(0.03)	(0.02)	(0.27)	(0.25)	(0.02)	(0.02)
Submitted patents	-0.02	-0.04**	0.2	0.21	-0.02**	-0.02**
	(0.02)	(0.02)	(0.15)	(0.14)	(0.01)	(0.01)
Cultural trademarks	-0.01	0	0.24	0.24*	0.02	0.02**
	(0.02)	(0.01)	(0.18)	(0.15)	(0.01)	(0.01)
Secondary education	0.36**	-0.1	-0.62	-0.73	-0.1*	-0.1*
	(0.17)	(0.08)	(1.59)	(1.11)	(0.06)	(0.06)
Time dummies	Yes	No	No	No	Yes	No
Observations	1,169	1,169	1,169	1,169	1,169	1,169
Imputations	80	80	80	80	80	80

Note: *p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect

Source: The calculations are done by the author, the data comes from the Eurostat's regional database.

Hence one can conclude that internet access pushes against the forces of creative destruction. Influence of other variables is milder but that is expectable as each digitalization component is only one of the many driving factors of entrepreneurship. Still, several components can be labeled as those driving the process. Those are ICT investment, skilled labor, computer access, broadband access, internet access, internet usage, and e-government. Although variables of all stages seem important for the business demographics, the "access" stage variables clearly dominate the effect of the whole digitalization process.

I present only the coefficient estimates and their standard errors because it is only those estimates which can be pooled according to Rubin's procedure. Other parameters of the models must be drawn from individual regressions. R-squared and especially adjusted R-squared of fixed effect models is extremely low (around 0.1 and even negative, respectively), but that should not be of major concern as we expect the fixed effect model to have such trait. Much of the variability is explained by the individual dummies which decrease the predictive power of other variables. Moreover, in comparison with other models, fixed effect estimation eliminates national dummies, so the model seemingly loses even more power. Nonetheless, the model always fares better than a simple constant model as F-statistic for joint significance of coefficients is always significant.

7 Endogeneity

Interesting as the presented results may be, they merely suggest association between digitalization and entrepreneurship; they do not show causal relationship. That is because the data might suffer from endogeneity. Since there are no empirical studies trying to disentangle the dynamics of digitalization and entrepreneurship, I cannot rely on previous literature. The obvious ad hoc solution would be to use lagged version of digitalization. Although endogeneity is a potential problem in the estimation, lagging the explanatory variable does not solve the issue but only replaces one assumption with another. In this case, the assumption of exogeneity is traded for assuming serial correlation of the endogenous variable and no serial correlation among the unobserved sources of endogeneity. The latter assumption is not even possible to be tested (Bellemare et al., 2015; Reed, 2015). Even if the lags are not a solution for endogeneity, the effect of digitalization itself might be lagged. Unfortunately, there is no literature on what the right lag might be. Identifying the suitable lag with limited time span is not feasible and it is beyond the scope of this study. Hence, I abstain from employing lags altogether and explore other methods of dealing with endogeneity.

7.1 Instrumental variables

Using instrumental variables for digitalization would be a way of exploring its causal effect on entrepreneurship. The quest for a suitable instrument is always challenging but it is not possible in the context of available regional data. Working with NUTS2 regions limits the scope of useful data to the Eurostat's regional database and the database simply does not contain data which could be perceived as driving factors of entrepreneurship, but only through the digitalization channel. In general, it is difficult to find instruments for a composite indicator as qualities of a good instrument overlap with the reasoning for including the variable in the index. It might be easier to instrument single stages of the digitalization process or even each component of digitalization. Unfortunately, even this task is extremely difficult at the regional level. For purposes of future research, however, I identify conceivable instruments for each stage of digitalization even if the regional data is not available. After all, if the causal relationship can be established on any geographical level, one can expect it to hold universally.

Instrumenting the environment stage of digitalization requires finding a phenomenon which is strongly associated with skilled, digitally literate labor force whose substantial proportion works in R&D. Assuming educated people who work in high-tech sectors or in research profess greater interest in science, it is possible to use scientific magazine subscriptions as an instrument for the first stage of digitalization. Geographical units where magazines such as *Nature*, *Scientific American* or other local scientific publications are frequently read are likely to be forerunners in the digitalization process. Moreover, higher proportion of people reading certain journals can hardly cause shifts in business demography. Thus the journal subscription may serve well as an instrument for the environment stage. Its drawback is that it is not likely to be associated with ICT investment. Nonetheless, such instrument may still yield considerable predictive power for the digitalization's first stage.

A similar logic can be applied when instrumenting the access stage of digitalization. As people have wider, better, and more ubiquitous access to the internet connection, they use more online services. This is reflected in the number of subscriptions to services such as Netflix or Spotify or search traffic via search engines such as Google or Yahoo. Regions where such services enjoy high popularity are likely to have a good access to digital technology, as it is a prerequisite for streaming audio or audiovisual content and browsing through the web. At the same time, it is unlikely that popularity of such services has a significant effect on entrepreneurship.

The last stage, adoption, tries to capture the point where digitalization reaches its maturity. The mature stage of technology manifests itself by infiltrating various business processes but the change in the way firms function requires investment. It is therefore only after there is a great confidence in the technology that the enterprises adopt it. A potential instrument for this stage can thus be the proportion of corporate users who exploit services such as SharePoint, Skype for Business, or Salesforce. These metrics reflect the maturity of the technology but are unlikely to be directly linked with swings in entrepreneurial activity.

Such data might prove difficult to gather as one would have to rely on the good will of the agents in question. Even with the data at our disposal, the challenge of assigning the data to different geographical units so that they stay comparable remains. That is why this paper does not attempt to address the question of endogeneity by instrumental variables. However, I believe it is a useful starting point for future research.

7.2 GMM estimation

Because instrumental variable approach is not feasible in the setting of this study, I try do control for endogeneity using General Method of Moments estimation (GMM). GMM deals with endogeneity by instrumenting the explanatory variables with the lagged explained variable as well as other explanatory variables. For this purpose, I use birth, death, and survival rates and all the control variables. I do not use lagged digitalization variables as this would not help to identify causality, it would merely suggest a lagged effect. The GMM estimation is done using robust covariance matrix and time dummies in each regression to remain consistent in the analysis. Performing Sargan test to test for suitability of the instruments used by the GMM method, it is clear that the lagged values of explained and explanatory variables are not persuasively good instruments. Unfortunately, it was not possible to find better ones. Considering that, the GMM analysis should be taken with caution.

Table 5: Digitalization index, GMM estimation

	Dependent variable:		
	Birth rate	Death rate	Survival rate
	(1)	(2)	(3)
Digitalization	4.23	-17.94	-11.21
	(9.7)	(79.39)	(9.6)
GDP per capita	25.94	283.12	24.07
	(27.87)	(218.94)	(20.71)
Unemployment	2.02	72.75	-13.16
	(6.59)	(46.07)	(5.41)
Public investment	-6.5	-27.9	1.63
	(6.82)	(44.03)	(5.19)
Submitted patents	2.09	31.7	-1.71
	(5.64)	(29.95)	(3.72)
Cultural trademarks	-1.79	31.44	-4.18
	(3.71)	(30.19)	(3.79)
Secondary education	98.22***	-25.16	-30.04
	(27.85)	(244.07)	(21.98)
Sargan test (p-value)	0.04	0.08	0.03
Time dummies	Yes	Yes	Yes
Observations	835	835	835
Imputations	80	80	80

Note: *p<0.1; **p<0.05; ***p<0.01
Source: The calculations are done by the author, the data comes from the Eurostat's regional database.

Table 5 shows the GMM estimates of the effects of digitalization. Although the signs as well as the magnitudes of the coefficient estimates are in line with the simple fixed and random effect estimations, GMM does not provide convincing statistical significance. Hence, the relatively strong statistical significance of the negative effect which digitalization has on survival rate was not proven to be causal by the GMM estimation.

Considering the dynamics of digitalization, GMM confirms some results from the previous regressions and suggests the relationship to be causal. The positive effect of the two initial stages of digitalization on birth rate as well as the negative effect of the intermediate stage on survival rate are in line the original results. The dynamics of digitalization, however, holds only in the case of birth rate. The dynamic character of the digitalization's effect on survival rate was not possible to replicate and thus should not be viewed as having causal qualities. Table 6 shows the results in detail.

Table 6: 3 stages of digitalization, GMM estimation

	Dependent variable:		
	Birth rate	Death rate	Survival rate
	(1)	(2)	(3)
Stage 1	4.75**	9.52	-1.42
	(2.33)	(15.19)	(1.64)
Stage 2	17.08***	13.25	-14.94***
	(2.9)	(23.28)	(2.81)
Stage 3	-11.89**	-52.82	5.82
	(4.88)	(32.12)	(4.56)
GDP per capita	-12.08	266.71	35.54**
	(21.98)	(222.44)	(20.27)
Unemployment	0.08	66.85	-10.86**
	(5.93)	(47.48)	(5.44)
Public investment	2.31	-29.27	-2.39
	(5.11)	(45.21)	(4.63)
Submitted patents	-2.96	38.64	0.24
	(4.78)	(29.58)	(3.17)
Cultural trademarks	-0.53	31.58	-3.24
	(3.5)	(29.62)	(3.53)
Secondary education	75.51***	10.56	-34.35
	(21.63)	(242.91)	(21.94)
Sargan test (p-value)	0.05	0.07	0.03
Time dummies	Yes	Yes	Yes
Observations	835	835	835
Imputations	80	80	80
Note:	*p<0.1; **p<0.05; ***p<0.01		
Source:	The calculations are done by the author, the data comes from the Eurostat's regional database.		

GMM analysis of individual components confirms majority of results obtained by standard methods. The "access stage" indeed seems the most important for business demography. Additionaly, individual components affect birth rate the most which suggests that digitalization has causal effect mostly on the creative part of Schumpeter's creative destruction. The results are presented in Table 7. In summary, GMM estimation partially confirms the estimates from fixed and random effect regressions but the dynamic characteristics of the digitalization process in case of survival rate were not replicated. This suggest a two-way causal relationship between digitalization and entrepreneurship. Further research is necessary for establishing precise causal links between the two phenomena.

Table 7: Individual components of digitalization, GMM estimation

	Dependent variable:		
	Birth rate	Death rate	Survival rate
	(1)	(2)	(3)
ICT investment	0.06***	0.04	0
	(0.02)	(0.16)	(0.01)
R&D labor	0.03	0.35	0.03
	(0.05)	(0.46)	(0.05)
Skilled labor	0.03	0.33	-0.05
	(0.05)	(0.42)	(0.04)
Computer Access	-0.83***	2.01	-0.2
	(0.23)	(1.98)	(0.2)
Broadband Access	0.4***	1.8**	-0.27***
	(0.08)	(0.8)	(0.07)
Internet Access	-0.21	-3.15**	0.1
	(0.14)	(1.42)	(0.14)
Mobile Access	-0.04	-0.08	0.05
	(0.04)	(0.34)	(0.04)
Internet usage	0.53***	-1.51	-0.19
	(0.2)	(1.63)	(0.18)
E-commerce	0.08**	0.03	-0.04
	(0.04)	(0.3)	(0.03)
E-government	-0.15***	-0.42*	0.11***
	(0.04)	(0.25)	(0.03)
GDP per capita	-0.27	1.16	0.42*
	(0.25)	(2.42)	(0.23)
Unemployment	0.02	0.48	-0.12**
	(0.06)	(0.46)	(0.05)
Public investment	0.05	-0.22	-0.05
	(0.05)	(0.45)	(0.05)
Submitted patents	-0.01	0.36	0
	(0.05)	(0.31)	(0.03)
Cultural trademarks	0.01	0.37	-0.04
	(0.03)	(0.3)	(0.03)
Secondary education	0.8***	0.49	-0.21
	(0.22)	(2.6)	(0.22)
Sargan test (p-value)	0.03	0.05	0.03
Time dummies	Yes	Yes	Yes
Observations	835	835	835
Imputations	80	80	80

Note:	*p<0.1; **p<0.05; ***p<0.01
Source:	The calculations are done by the author, the data comes from the Eurostat's regional database.

8 Robustness

GMM estimation has showed that the results are robust with respect to statistical method. However, there are other parts of the study which require a robustness check such as its robustness to the specific dataset. Although it would be conceivable to run the same analysis with data at the national level as another robustness check, this is is beyond the scope of this study. Moreover, such analysis would be more beneficial, if it was adjusted for the specifics of aggregate data. That would lead to a different analysis. Hence, I only check the study's robustness with respect to the composite index construction, data imputation, and the modifiable unit problem (MAUP).

8.1 Composite indicator weighting

The composite index weighting can be done in different ways. Although principal component analysis controls for structure of the data and that is the reason why I used the method in the main analysis, many indicators are nonetheless constructed using equal weights. Since this can significantly affect the analysis, Table 8 and 9 in the appendix present coefficient estimates of composite indicators constructed with equal weights. The overall impact of digitalization is robust with respect to the weighting method. However, the effect of the "environment" stage changes. Weights based on principal components assign low weight to skilled labor as it is highly correlated with R&D labor. Because skilled labor is an important driver of entrepreneurship, equal weights inflate the effect of the first stage of digitalization. The rest of the analysis, however, remains intact.

8.2 Modifiable areal unit problem

Breaking down the national data into regions gives us more insights as well as statistical power, but it also contains the issues of spatial interdependence. Firstly, although the NUTS2 regions are constructed in a way so that the individual regions are coherent units, it is not possible to rule out spatial dependence between regions with close proximity. In other words, had Eurostat arranged the regions in slightly different way, we might observe different outcomes. This phenomenon is the aggregation problem of MAUP and it is widely known in the US as gerrymandering where this property of spatially arranged data is used to affect election results.[12]

Secondly, the scale problem of MAUP reflects the dependence of the statistical inference on the spatial resolution. Because the data is available only at one level of spatial resolution, I do not address the scale problem of MAUP in this study and focus solely on the aggregation

[12]The term was coined in 1812 by an American federalist newspaper. The editors likened the shape of newly defined districts in Massachusetts to the mythical creature, salamander. The Democratic-Republican governor of the state was Elbridge Gerry, hence "gerrymandering" (Cox and Katz, 2002).

problem. However, the spatial dependence in any form leads to both inefficient and biased estimates (Jelinski and Wu, 1996).

To deal with the aggregation problem, I perform geographically weighted regressions (GWR). GWR weights the observations based on the a kernel function where the weight falls with the geographical distance of the data point. Using each observation as a center for one regression, GWR yields one estimate for each geographical unit. Hence it allows us to observe variation in the estimates among regions and establish whether the outcomes are dependent on its spatial position (Brunsdon et al., 1998). The geographical location of a region is represented by a centroid calculated for each region.

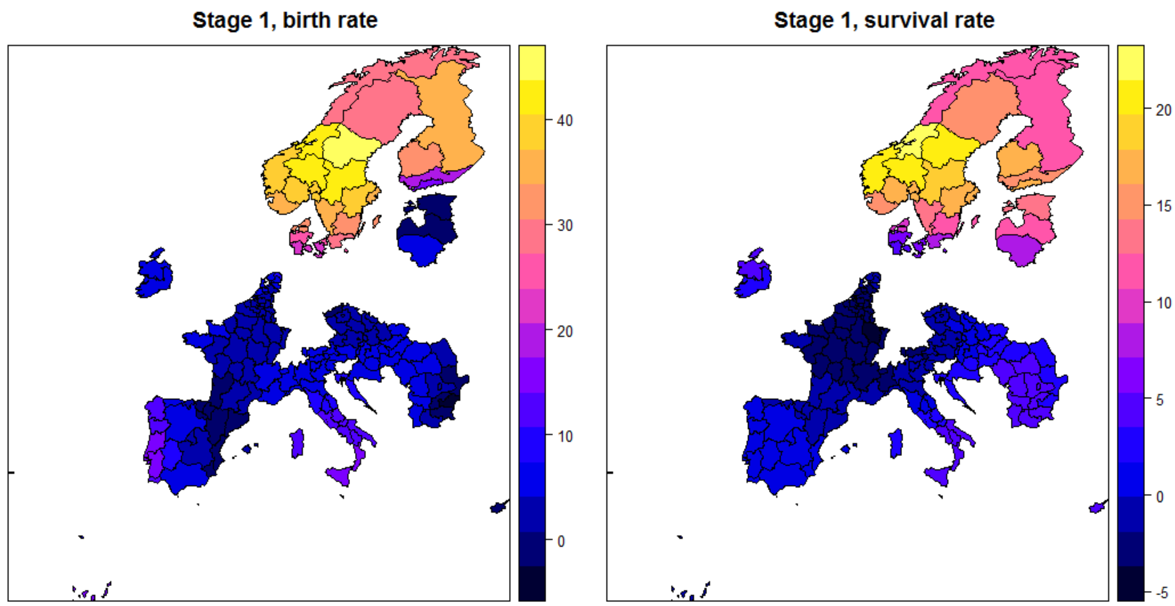
To conduct the spatial analysis, I followed Bivand et al. (2015). The corrected cross-validation method for establishing the bandwidth minimizes the mean square prediction error for GWR. The parameter describes the range of influence exerted on a region, hence, as bandwidth increases, GWR approaches regular regression. I estimate regressions for birth, death, and survival rates using the composite index as well as its stages. A unique bandwidth is estimated for each regression. The results of all GWR regressions are presented in set of maps in Figure 9 in the appendix.

The regional analysis sheds some additional light on the links between digitalization and entrepreneurship. Much of the variability is due to regional characteristics which are not captured by the fixed effect regressions. Indeed the statistical significance of, for example, the first stage of digitalization on birth rate can be partially attributed to the level of similarity of European regions. The mean coefficient estimate is 8.9, with values ranging from -2.6 to 43.7. These results enable us to look at the significant estimates of the fixed effect regression of 4.9 with greater level of confidence.

The most striking finding of the spatial analysis is the exceptionality of the Nordic countries with respect to the "environment stage" of digitalization. Denmark, Finland, Norway, and Sweden seem to be highly receptive towards fertile digitalization environment compared to their continental peers. Figure 4 below shows the exceptionality of the Nordics in the first digitalization stage. Interestingly enough, this responsiveness is reflected in the next stages when the effects on birth and survival rates reverse (see Figure 9 in the appendix). An obvious question about the origin of the Nordic exeptionality arises. However, further research is needed to address this question.

One should note that the presented results suffer from distortion due to omission of several major countries (Germany, the UK, and Poland). The analysis also does not take into account continuity outside of the EU. Hence regions which are handled as peripheral in the analysis might be, in reality, not peripheral at all. To deal with this issue, better and complete data is required.

Figure 4: Regional estimates of the effect of "environment stage"



Source: The map is produced by the author using Eurostat's data.

8.3 Data imputation

Unfortunately, such data is currently not available. The used dataset provides the most elaborate regional data for Europe. This is why I had to employ techniques not standard in the realm of economics and used the MICE algorithm with PMM method to impute the missing values. Although data imputation used in this paper has a strong statistical rationale, it adds an additional layer of complexity and thus raises question marks about its impact on the results.

Because composite indicators cannot be constructed using data with missing values, it is not possible to validate the overall effect of digitalization and avoid data imputation at the same time. Although it is feasible to run a regression with all the components of digitalization, it uses data from mere 60 regions and 2 years due to the missing values. Because of that, the statistical power of the results is not substantial. I was able to replicate the positive effect of ICT investment and skilled labor on birth rate, and broadband access and e-government actually yield contradicting results. The conflicting results can be attributed to the usage a mere third of the available regions. As it was shown, there is a significant geographical variability in the data which might have distorted the coefficient estimates. The regression might have been also affected by the limited time span of 2 years. Table 10 in the appendix shows the detailed results of the regressions.

9 Discussion

This paper's focus is to disentangle the relationship between digitalization and entrepreneurship. Because they are both positively linked to productivity and economic growth, understanding their interplay is of both theoretical and practical interest. I attempted to aggregate the whole process of digitalization by a composite indicator and then I split the process into 3 stages. Such an approach is unique for technology dissemination analysis. Although the Eurostat's regional dataset contains many missing values, the novel method of data imputation enabled the study to proceed forward.

The results show a strong relationship between digitalization and entrepreneurship. Although the statistical significance is often diminished by the aggregation process, the data nonetheless yields both statistically and economically significant results. Confirming my hypothesis, digitalization negatively affects the survival rate of nascent businesses which can be attributed to incumbents' successful market share defense and to the trial and error approach of potential new entrepreneurs. In addition, the regional analysis shows it is mostly the Northern and Eastern countries which drive this significant effect.

The regressions suggest an interesting dynamic interplay between digitalization and entrepreneurship. Birth rate is positively affected as the digitalization process takes off, but these gains are partially compensated for as digitalization finishes its adoption stage. The same dynamics is observed in case of survival rate. Interestingly, death rate seems to be very resilient towards the force of digitalization. The analysis suggests that the Schumpeterian creative destruction is not significantly affected by the digital technology dissemination as the regressions do not provide any robust results.

These results cannot be interpreted as causal because there is no rationale for thinking that entrepreneurship cannot directly or indirectly influence digitalization as well. Although I was not able to address the issue using instrumental variables, I was able to identify suitable instruments for further research. GMM approach deals with the issue of endogeneity but it builds on the problematic assumptions of exogeneity of lagged variables. Still, GMM yields significant results corresponding with, but not perfectly mirroring the original results. This suggests there is a causal interrelatedness between digitalization and entrepreneurship. However, the implications of the mutual influence are important for the theory of entrepreneur: it supports both Schumpeterian and Kirznerian definitions of entrepreneur. After all, those two narratives are not mutually exclusive and so both can be represented in the real economy.

To inspect the regional differences, I used geographically weighted regressions. They revealed considerable variability in the coefficient estimates with respect to geographical location. Indeed, we observe "Nordic exceptionality" in how these regions' birth rate responds to the first stage of digitalization. Although the regional analysis provided many clues which can be pursued in further research, the geographical variability is suspected to have made the original estimation less efficient.

This study has attempted to address entrepreneurship by using business demography as a proxy. One should be aware, however, that entrepreneurship is a complicated phenomenon and that business demography captures only one of its aspects. I disregarded the nature of the new businesses and circumstances of their creation as well as nature and circumstances of those firms ceasing their operations. The data also does not capture firms' size in terms of employees or revenue. These lacking data points do not make the results invalid though; they rather point to future research areas.

The digitalization index used here has been designed for purposes of regional analysis. Hence its construction reflected the needs of distinguishing different regions as well as the availability of data. However, it is conceivable to build a digitalization index for similar purposes in an alternative fashion. Using such index would then potentially yield different results. Therefore, when replicating the results of this paper, it is crucial to keep in mind that deviating from the method of how one captures digitalization (different variables, focus on countries instead of regions, etc.) might substantially influence the results.

The final dataset upon which this paper is based has been subject to rather intensive dicing. In order to have dataset containing only so much missing data that it can be imputed, I had to limit the data both spatially and in terms of time. Consequently, I analyzed only the post-crisis period in which low interest rates were a norm in Europe. Low interest rates exert a downward pressure on cost of capital which increases the profitability of firms and thus the threshold for their survival. This can affect birth, death, and survival rate. However, the interconnectedness of capital markets does not make incorporating interest rates in the analysis as straightforward as it may seem at the first glance. Implications of changing interest rates on entrepreneurship in the context of digitalization can be investigated by future studies.

Although the data imputation process was carefully executed and documented, its suitability for spatially arranged data is not thoroughly investigated. For example, a conceivable extension would be to assign greater probability to geographically close data points in the PMM algorithm. This would reflect the spatial dependence which I uncovered in the data. However, such methodological contribution falls into a scope of technique-developing rather than empirical study, so it was not done in this paper.

The study provides reliable results but they are by no means definite. The knowledge of the interplay between digitalization and entrepreneurship remains limited. The next step in endeavors of understanding this relationship is to establish and quantify causal links between the phenomena. It would be also beneficial to inspect the stages of the digitalization process and include lagged variables in the analysis. There is a strong rationale for lagged effect and this could also distort the current investigation. Lagged effect of the first stage could be interpreted as the effect of the second stage. In other words, including lagged variables would help to deal with omitted variable bias. However, longer time span, i.e. better and more voluminous data, is necessary for such analysis.

I have shown there is a considerable variability of the coefficient estimates among geographical units. However, the results might differ also when one slices the data across different dimensions. A possible extension is to split the data into different industries. This would provide us with insights into how digitalization links to entrepreneurship across different sectors of the economy. Common sense suggests some industries are more exposed to digitalization than others, but whether that translates into increased entrepreneurial activity within these sectors is far from certain.

10 Conclusion

This paper shows that there is indeed significant interaction between digitalization and entrepreneurship. Although I was able to show that the effects of digitalization on entrepreneurship are, to some degree, causal, the relationship is likely to be mutual. Dividing the digitalization process into 3 different stages, a dynamic pattern emerged showing that the initial positive effects on birth and survival rates are offset in the later stages. Moreover, there is a significant regional variability in the estimates with the Nordic countries being especially sensitive to environment suitable for a commencement of digitalization.

Proving that digitalization is linked to entrepreneurship has both theoretical and practical consequences. It gives an additional insight into the dissemination mechanism of technology in the economy. Such knowledge is crucial for understanding the decelerating productivity growth in the developed economies. It shows that technology as such is not enough to transform the economy. Distribution of the technological progress is needed as well and this study has identified entrepreneurs as potential bearers of the technological advancement. This has also serious policy implications. Policies open to digitalization and those focusing on entrepreneurship support are complementary. Both are necessary for economic prosperity.

Although digitalization has some long-lasting effects, it is essentially a dynamic process and thus it should be treated as such. Its dynamics reveals strong effects on entrepreneurship in the initial stages of digitalization but as the digital transformation peaks, the effects reverse. This suggests digitalization is a mere wave of technological progress and once it permeates the whole economy, its effect becomes invisible, just as one does not observe merits of electrification in the developed countries' statistics. It had simply become a part of generally available technology. It is safe to assume that digitalization has the same destiny.

I have analyzed the European regions hoping to produce results applicable to all developed economies. However, the results often vary substantially based on a geographical unit within Europe. This diminishes the universal relevance of this study's outcomes but it provides valuable insights into differences among European countries and even their regions. The Nordic countries proved to be especially sensitive to the initial stage of digitalization. Indeed, those countries are associated with progressive stance to the digital technology which is supported by the analysis. Their policies can serve as an inspiration for continental Europe if it wants to follow the Nordic technological lead.

To conclude, this study confirmed the potential relevance of several hypothesized channels through which digitalizaton affects entrepreneurship. Those are new products and services, improvement of current products and services, reorganization of internal corporate processes, and transaction costs reduction. It is these channels, among others, through which digitalization reshapes the business landscape and thus affects the whole economy.

References

Ács, Z. (2006). How is entrepreneurship good for economic growth? *Innovations 1*(1), 97–107.

Afonso, A., L. Schuknecht, and V. Tanzi (2005). Public sector efficiency: an international comparison. *Public Choice 123*(3), 321–347.

Baden-Fuller, C. and S. Haefliger (2013). Business models and technological innovation. *Long Range Planning 46*(6), 419–426.

Baily, M. N. and R. J. Gordon (1989). The productivity slowdown, measurement i, and the explosion of computer power. Working Paper 1199, National Bureau of Economic Research.

Baumol, W. J. (1968). Entrepreneurship in economic theory. *American Economic Review 58*(2), 64–71.

Baumol, W. J. (1996). Entrepreneurship: Productive, unproductive, and destructive. *Journal of Business Venturing 11*(1), 3–22.

Bellemare, M. F., T. Masaki, and T. B. Pepinsky (2015). Lagged Explanatory Variables and the Estimation of Causal Effects. SSRN Scholarly Paper ID 2568724, Social Science Research Network, Rochester, NY.

Bivand, R., D. Yu, T. Nakaya, M.-A. Garcia-Lopez, and M. R. Bivand (2015). Package 'spgwr'. *R Package Version 0.6–19. Available online: http://cran.r-project.org/web/packages/spgwr/spgwr.pdf (accessed on 124 April 2017).*

Bjuggren, C. M., D. Johansson, and M. Stenkula (2012). Using self-employment as proxy for entrepreneurship: some empirical caveats. *International Journal of Entrepreneurship and Small Business 17*(3), 290–303.

Bosma, N. and V. Schutjens (2011). Understanding regional variation in entrepreneurial activity and entrepreneurial attitude in Europe. *Annals of Regional Science 47*(3), 711–742.

Bresnahan, T. F. and M. Trajtenberg (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics 65*(1), 83–108.

Brunsdon, C., S. Fotheringham, and M. Charlton (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician) 47*(3), 431–443.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM 36*(12), 66–77.

Brynjolfsson, E. and L. Hitt (1995). Information technology as a factor of production: the role of differences among firms. *Economics of Innovation and New Technology 3*(3-4), 183–200.

Brynjolfsson, E., T. W. Malone, V. Gurbaxani, and A. Kambil (1994). Does information technology lead to smaller firms? *Management Science 40*(12), 1628–1644.

Brynjolfsson, E. and A. McAfee (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton, New York.

Cardona, M., T. Kretschmer, and T. Strobel (2013). ICT and productivity: conclusions from the empirical literature. *Information Economics and Policy 25*(3), 109–125.

Cherchye, L., W. Moesen, N. Rogge, and T. V. Puyenbroeck (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research 82*(1), 111–145.

Clercq, D. D. and P. Arenius (2006). The role of knowledge in business start-up activity. *International Small Business Journal 24*(4), 339–358.

Cox, G. W. and J. N. Katz (2002). *Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution*. Cambridge University Press, Cambridge.

Crandall, R. W., W. Lehr, and R. E. Litan (2007). *The effects of broadband deployment on output and employment: A cross-sectional analysis of U.S. data*. The Brookings Institution.

David, P. A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. *American Economic Review 80*(2), 355–361.

Dewan, S. and C.-k. Min (1997). The substitution of information technology for other factors of production: a firm level analysis. *Management Science 43*(12), 1660–1675.

Edquist, H. and M. Henrekson (2017). Do R&D and ICT affect total factor productivity growth differently? *Telecommunications Policy*.

Erken, H., P. Donselaar, and R. Thurik (2016). Total factor productivity and the role of entrepreneurship. *Journal of Technology Transfer*, 1–29. DOI: 10.1007/s10961-016-9504-5.

Freytag, A. and R. Thurik (2007). Entrepreneurship and its determinants in a cross-country setting. *Journal of Evolutionary Economics 17*(2), 117–131.

Galindo, M. and M. T. Méndez (2014). Entrepreneurship, economic growth, and innovation: Are feedback effects at work? *Journal of Business Research 67*(5), 825–829.

Gartner (2017). Gartner IT Glossary. Accessed March 30, 2017. http://www.gartner.com/it-glossary/.

Gordon, R. J. (2004). Why was Europe Left at the Station When America's Productivity Locomotive Departed? Working Paper 10661, National Bureau of Economic Research. DOI: 10.3386/w10661.

Gordon, R. J. (2014). The demise of us economic growth: restatement, rebuttal, and reflections. Technical report, National Bureau of Economic Research.

Grilo, I. and R. Thurik (2008). Determinants of entrepreneurial engagement levels in Europe and the US. *Industrial and Corporate Change 17*(6), 1113–1145.

Gruber, H. and P. Koutroumpis (2011). Mobile telecommunications and the impact on economic development. *Economic Policy 26*(67), 387–426.

Henrekson, M. and T. Sanandaji (2014). Small business activity does not measure entrepreneurship. *Proceedings of the National Academy of Sciences 111*(5), 1760–1765.

Holcombe, R. G. (1998). Entrepreneurship and economic growth. *Quarterly Journal of Austrian Economics 1*(2), 45–62.

Hummels, D. (2007). Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives 21*(3), 131–154.

Jacobsen, J., A. Sørensen, and M. Junge (2011). Digitalization and productivity. *Centre for Economic and Business Research, Copenhagen Business School*.

Jelinski, D. E. and J. Wu (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology 11*(3), 129–140.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quarterly Journal of Economics 122*(3), 879–924.

Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2008). A retrospective look at the us productivity growth resurgence. *The Journal of Economic Perspectives 22*(1), 3–24.

Jovanovic, B. and P. L. Rousseau (2005). General Purpose Technologies. In *Handbook of Economic Growth*, Volume 1, 1181–1224. Elsevier, Amsterdam.

Karlsson, M., L. Trygg, and B.-O. Elfström (2004). Measuring R&D productivity: complementing the picture by focusing on research activities. *Technovation 24*(3), 179–186.

Katz, R., P. Koutroumpis, and F. Martin Callorda (2014). Using a digitization index to measure the economic and social impact of digital agendas. *info 16*(1), 32–44.

Katz, R. L. and P. Koutroumpis (2013). Measuring digitization: A growth and welfare multiplier. *Technovation 33*(10), 314–319.

Kirzner, I. M. (1997). Entrepreneurial discovery and the competitive market process: an austrian approach. *Journal of Economic Literature 35*(1), 60–85.

Kirzner, I. M. (1999). Creativity and/or alertness: A reconsideration of the Schumpeterian entrepreneur. *The Review of Austrian Economics 11*(1-2), 5–17.

Little, R. J. and D. B. Rubin (1987). *Statistical Analysis with Missing Data.* John Wiley, New York.

Masuda, T. (2006). The determinants of latent entrepreneurship in Japan. *Small Business Economics 26*(3), 227–240.

McQuaid, R. W. (2002). Entrepreneurship and ICT industries: Support from regional and local policies. *Regional Studies 36*(8), 909–919.

Miller, B. and R. D. Atkinson (2014). Raising European productivity growth through ICT. *The Information Technology and Innovation Foundation, Washington D.C.*.

Munda, G. and M. Nardo (2009). Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics 41*(12), 1513–1523.

Muto, M. and T. Yamano (2009). The impact of mobile phone coverage expansion on market participation: Panel data evidence from Uganda. *World development 37*(12), 1887–1896.

Nardo, M., M. Saisana, A. Saltelli, and S. Tarantola (2005). Tools for composite indicators building. *European Comission, Ispra.*

Noyelle, T. J. (1990). *Skills, Wages, And Productivity In The Service Sector.* Boulder, CO: Westview Press.

Oliner, S. D. and D. E. Sichel (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? SSRN Scholarly Paper ID 233139, Social Science Research Network, Rochester, NY.

Phelps, E. S. (2006). Dynamic capitalism. *Wall Street Journal, October 10 10*, A14.

Raghunathan, T. E., J. M. Lepkowski, J. Van Hoewyk, and P. Solenberger (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology 27*(1), 85–96.

Reed, W. R. (2015). On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics 77*(6), 897–905.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association 91*(434), 473–489.

Saltelli, A., M. Nardo, J. Saisana, and S. Tarantola (2005). Composite indicators: the controversy and the way forward. *Statistics, Knowledge and Policy*, 359.

Schenker, N. and J. M. G. Taylor (1996). Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis 22*(4), 425–446.

Schumpeter, J. A. (1961). *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Oxford University Press, Oxford.

Simón-Moya, V., L. Revuelto-Taboada, and R. F. Guerrero (2014). Institutional and economic drivers of entrepreneurship: An international perspective. *Journal of Business Research 67*(5), 715–721.

Solow, R. M. (1987). We'd Better Watch Out. *New York Times Book Review, July 12, 36*.

Thompson, H. and C. Garbacz (2008). Broadband impacts on state gdp: direct and indirect impacts. Paper presented at the 17th Biennial Conference of the International Telecommunications Society (ITS). Montreal, June 24-27.

van Ark, B., M. O'Mahony, and M. P. Timmer (2008). The productivity gap between Europe and the United States: trends and causes. *Journal of Economic Perspectives 22*(1), 25–44.

van Buuren, S., J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation 76*(12), 1049–1064.

van Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software 45*(3).

van Stel, A., M. Carree, and R. Thurik (2005). The effect of entrepreneurial activity on national economic growth. *Small Business Economics 24*(3), 311–321.

Wennekers, S., R. Thurik, A. van Stel, and N. Noorderhaven (2007). Uncertainty avoidance and the rate of business ownership across 21 OECD countries, 1976–2004. *Journal of Evolutionary Economics 17*(2), 133–160.

White, I. R., R. Daniel, and P. Royston (2010). Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Computational Statistics and Data Analysis 54*(10), 2267–2275.

Wong, P. K., Y. P. Ho, and E. Autio (2005). Entrepreneurship, innovation and economic growth: Evidence from GEM data. *Small Business Economics 24*(3), 335–350.

11 Appendix

Table 8: Digitalization index regressions, with equal weights

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	FE (3)	**RE** (4)	FE (5)	**RE** (6)
Digitalization	16.71	14.51**	-24.34	-32.48	-19.66**	-18.27***
	(10.67)	(6.8)	(73.71)	(68.77)	(9.27)	(5.72)
GDP per capita	-25.25***	-1.45	68.43	102.74*	20.33*	11.89***
	(11.68)	(5.11)	(106.91)	(60.15)	(11.67)	(4.3)
Unemployment	0.06	5.87**	32.4	49.08**	-7.93**	2.59
	(4.22)	(2.42)	(30.05)	(24.74)	(3.6)	(1.95)
Public investment	4.07	4.93**	-24.04	-21.36	0.43	3.63*
	(3.13)	(2.34)	(25.97)	(23.72)	(2.86)	(1.93)
Submitted patents	-2.52	-3.2**	18.91	23.08**	-2.65*	-1.76
	(1.97)	(1.54)	(14.38)	(13.24)	(1.61)	(1.15)
Cultural trademarks	-0.75	-0.32	22.42	22.6	0.2	1.17
	(1.88)	(1.35)	(17.27)	(14.45)	(1.77)	(1.11)
Secondary education	26.36	-13.2	-68.82	-54.85	-17.28	-18.49***
	(16.26)	(8.14)	(151.56)	(108.4)	(14.17)	(6.27)
Time dummies	Yes	No	No	No	Yes	No
Observations	1,169	1,169	1,169	1,169	1,169	1,169
Imputations	80	80	80	80	80	80

Note: *p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect

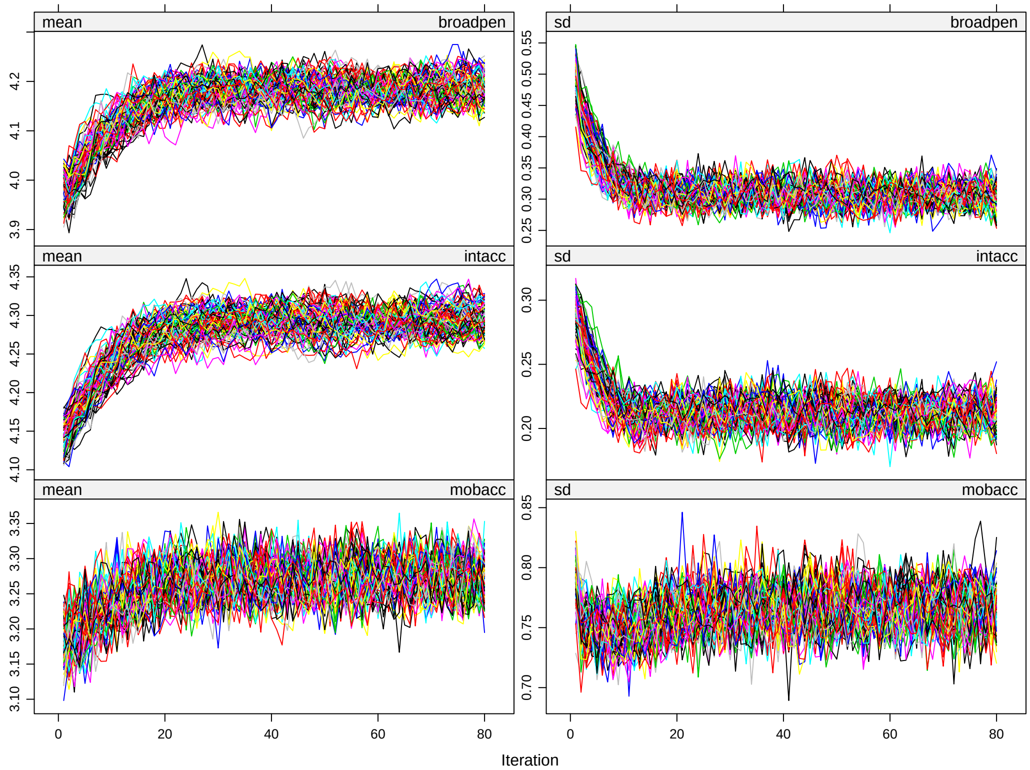
Source: The calculations are done by the author, the data comes from the Eurostat's regional database.

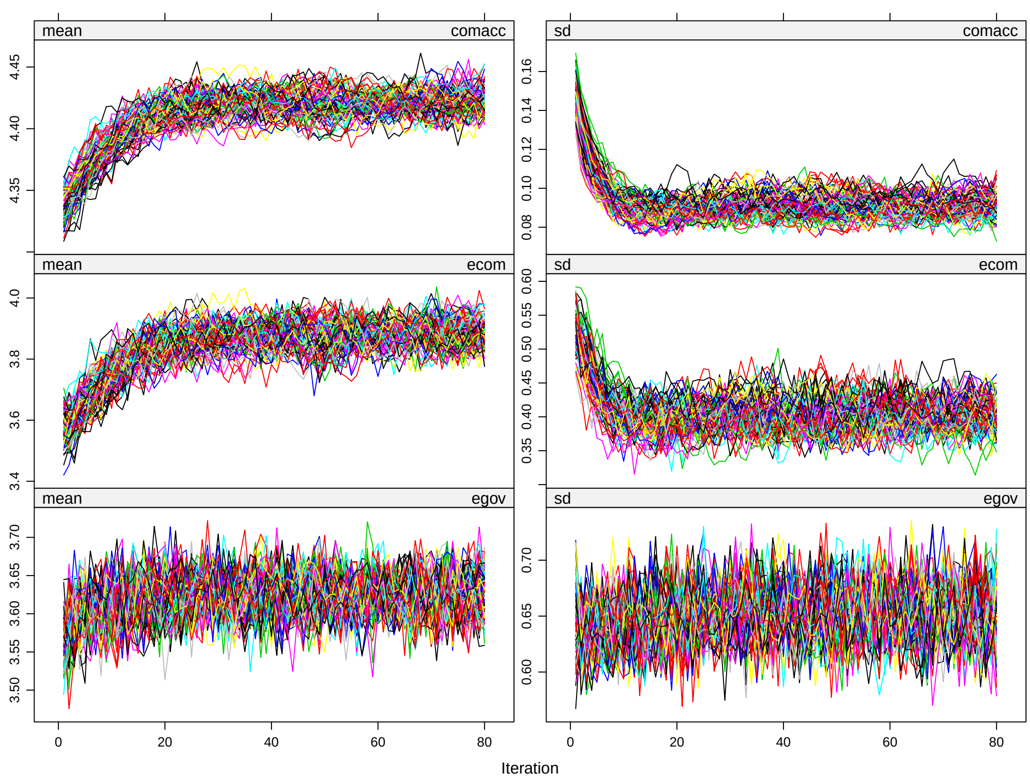
Table 9: 3 stages of digitalization, with equal weights

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	FE (3)	**RE** (4)	FE (5)	**RE** (6)
Stage 1	11.58***	15.82***	46.58	61.25**	8.91***	8.5***
	(3.94)	(2.2)	(39.86)	(29.18)	(1.93)	(1.96)
Stage 2	7.8***	5.58**	5.22	-1.56	-11.24***	-10.91***
	(2.83)	(2.4)	(20.77)	(20.87)	(2.27)	(2.15)
Stage 3	-9.95**	-13.86***	-38.75	-42.44	-3.79	-3.69
	(4.41)	(3.88)	(31.38)	(30.49)	(3.3)	(3.4)
GDP per capita	-31.07***	-13.54***	38.61	37.55	4.6	4.17
	(11.73)	(5.11)	(113.28)	(69.57)	(4.33)	(4.34)
Unemployment	-0.49	5.27**	33.59	51.46**	4.14**	5.48***
	(4.2)	(2.29)	(31.72)	(25.87)	(1.99)	(1.86)
Public investment	4.61	3.9*	-24.84	-25.4	1.05	1.62
	(3.2)	(2.28)	(26.08)	(23.8)	(1.91)	(1.84)
Submitted patents	-2.57	-4.07***	20.47	23.04*	-2.34**	-2.21**
	(1.98)	(1.51)	(14.3)	(13.1)	(1.04)	(1.01)
Cultural trademarks	-0.88	-0.41	22.34	22.09	1.75**	2.15**
	(1.87)	(1.28)	(17.37)	(14.49)	(1.03)	(1.01)
Secondary education	28.53*	-18.59**	-89.21	-89.06	-17.52***	-16.88***
	(16.71)	(7.74)	(150.95)	(107.42)	(6.1)	(5.95)
Time dummies	Yes	No	No	No	Yes	No
Observations	1,169	1,169	1,169	1,169	1,169	1,169
Imputations	80	80	80	80	80	80

Note:	*p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect
Source:	The calculations are done by the author, the data comes from the Eurostat's regional database.

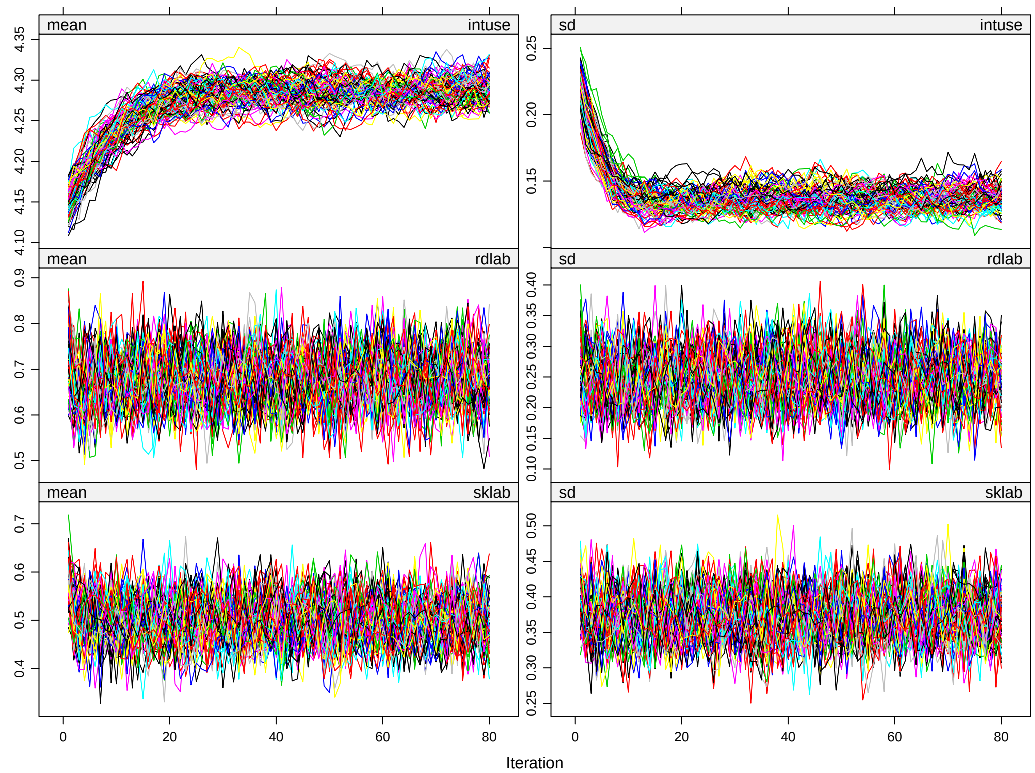
Figure 5: Imputation convergence, part 1

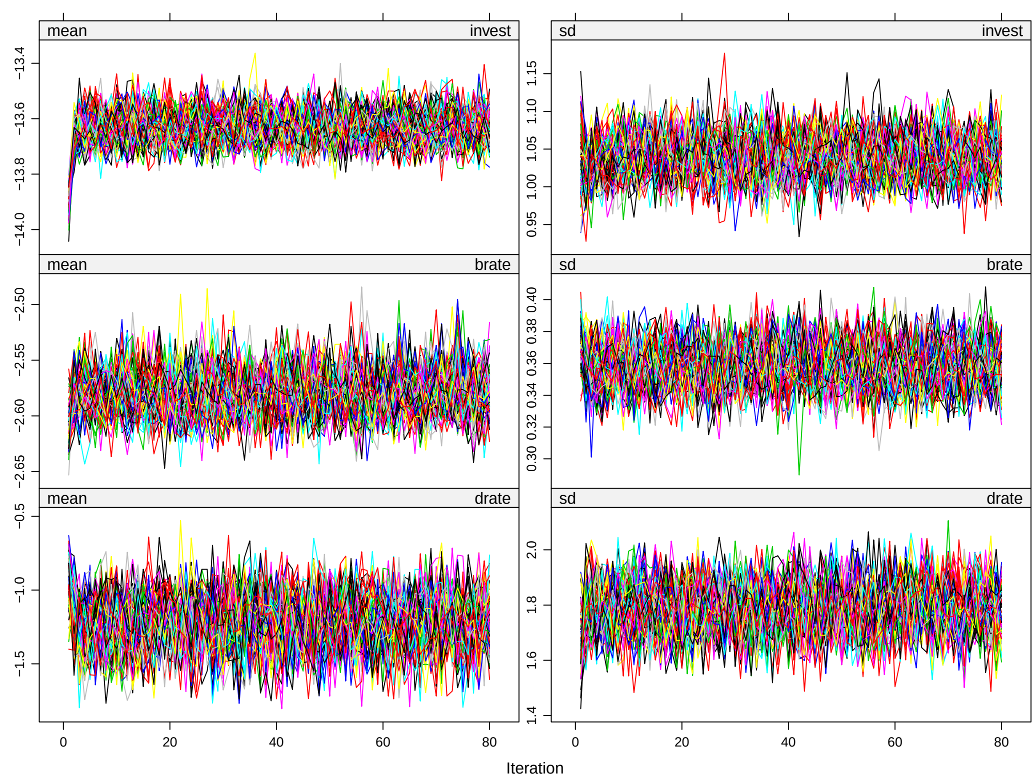




Source: The graphs are based on the author's calculations.

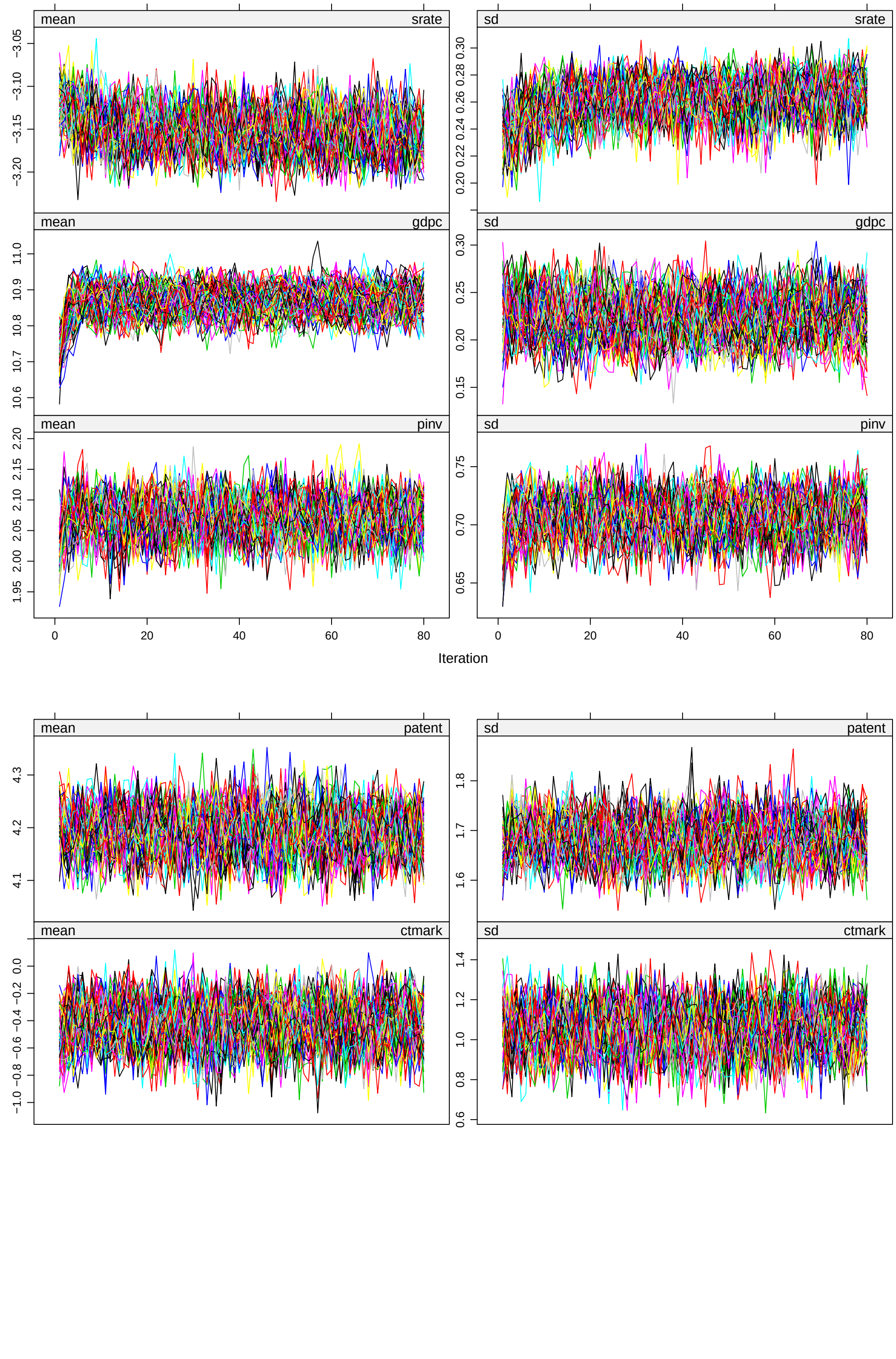
Figure 6: Imputation convergence, part 2





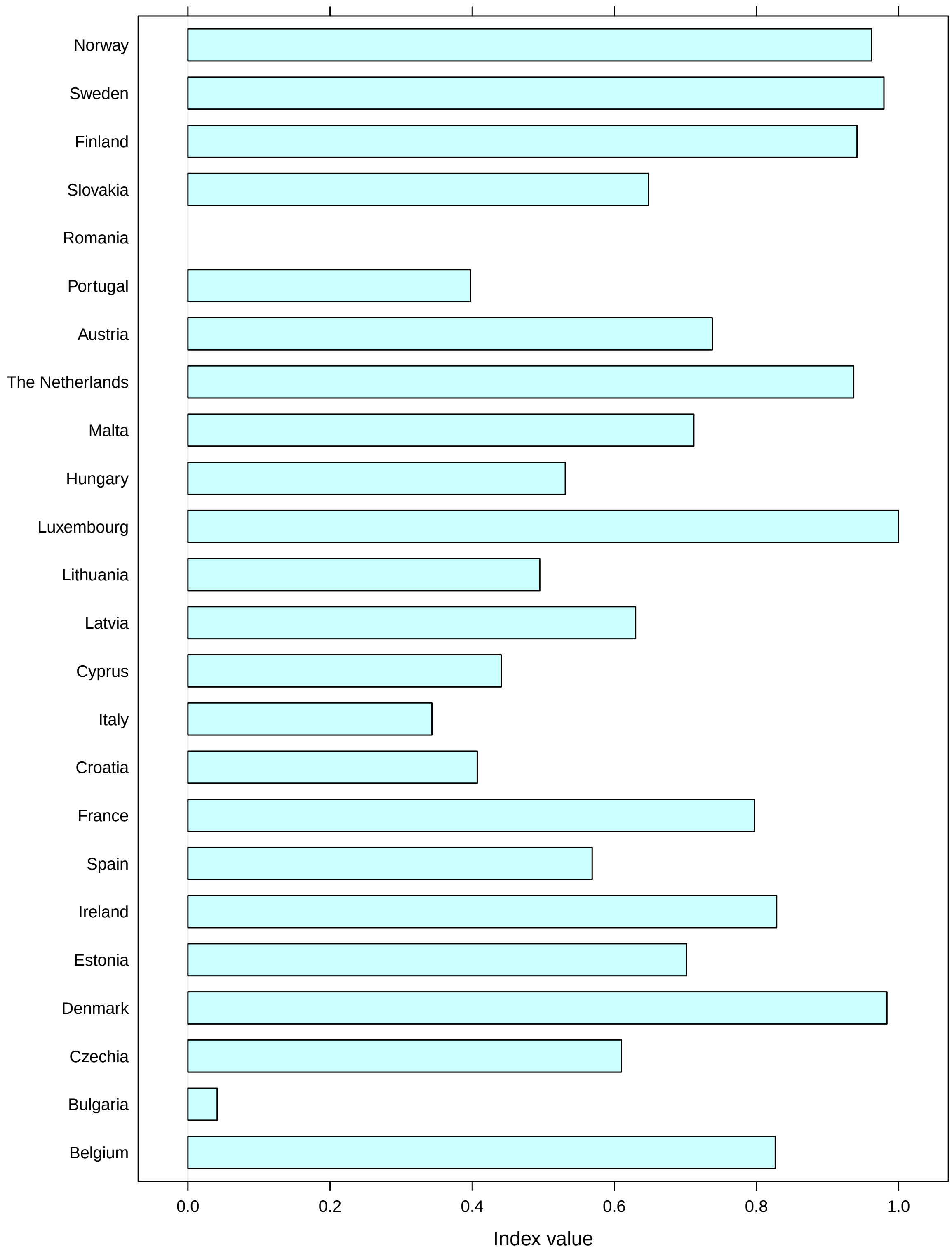
Source: The graphs are based on the author's calculations.

Figure 7: Imputation convergence, part 3



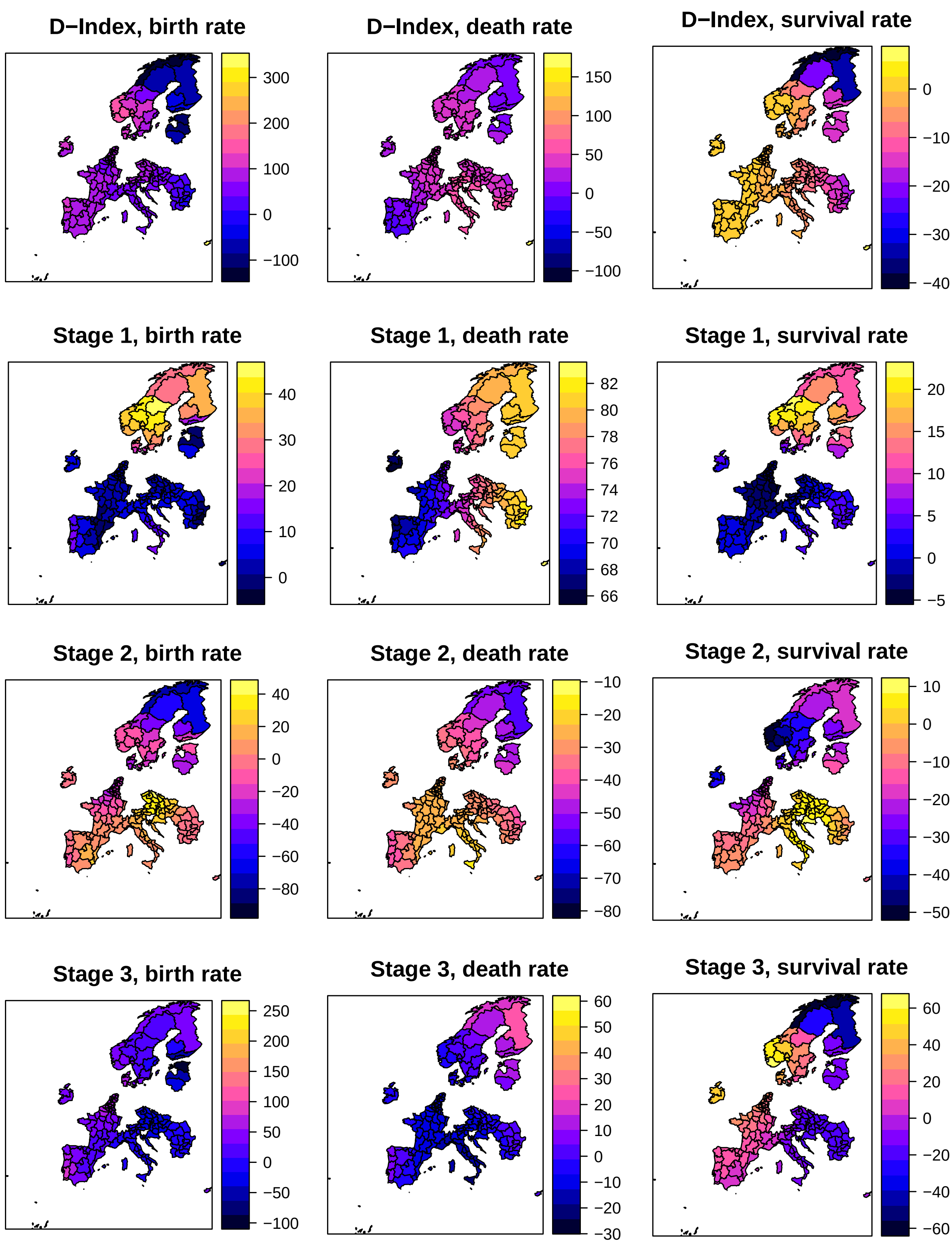
Source: The graphs are based on the author's calculations.

Figure 8: Country ranking in digitalization



Source: The ranking was produced by the author using Eurostat's data.

Figure 9: Sensitivity of regions



Source: The map is produced by the author using Eurostat's data.

Table 10: Individual components of digitalization, with no data imputation

	Dependent variable:					
	Birth rate		Death rate		Survival rate	
	FE (1)	RE (2)	FE (3)	**RE** (4)	**FE** (5)	RE (6)
ICT investment	0.08*	0.03	0.03	0	0.07	0.06***
	(0)	(0)	(0)	(0)	(0)	(0)
R&D labor	0.33**	0.17**	0.19	0.2	-0.18	0.03
	(0.03)	(0)	(0.04)	(0.05)	(0.03)	(0)
Skilled labor	0.18	0.08**	0.21	0.2	0.13	0.04
	(0.01)	(0)	(0.02)	(0.02)	(0.01)	(0)
Computer Access	-0.29	-0.84**	1.5*	1.61	-0.9	-0.1
	(0.68)	(0.14)	(0.72)	(0.98)	(0.68)	(0.14)
Broadband Access	-0.6***	-0.47***	-1.27***	-1.17***	0.33	-0.02
	(0.05)	(0.02)	(0.07)	(0.08)	(0.05)	(0.02)
Internet Access	0.96	1.12***	0.24	0.1	0.68	0.3
	(0.44)	(0.12)	(0.4)	(0.54)	(0.44)	(0.13)
Internet usage	0.37	-0.01	-0.34	-0.45	0.16	-0.15
	(0.35)	(0.13)	(0.47)	(0.63)	(0.35)	(0.14)
E-commerce	-0.03	-0.04	0.08	0.1	0.01	0.02
	(0)	(0)	(0.01)	(0.01)	(0)	(0)
E-government	0.08	0.1**	-0.13*	-0.14*	-0.21***	-0.1**
	(0)	(0)	(0.01)	(0.01)	(0)	(0)
GDP per capita	0.16	-0.18*	2.03**	1.43**	-0.86	-0.09
	(0.42)	(0.01)	(0.67)	(0.39)	(0.47)	(0.01)
Unemployment	0.4***	0.15***	0.53***	0.53***	0.06	0.09**
	(0.02)	(0)	(0.03)	(0.04)	(0.03)	(0)
Public investment	-0.07	0.01	0.22*	0.24*	-0.18*	-0.02
	(0.01)	(0)	(0.02)	(0.02)	(0.01)	(0)
Submitted patents	0.1**	-0.02	0.03	0.02	0.08	-0.01
	(0)	(0)	(0)	(0)	(0)	(0)
Cultural trademarks	0.01	-0.02	0.02	0.01	0.02	0
	(0)	(0)	(0)	(0)	(0)	(0)
Secondary education	-1.74*	0.07	-2.93**	-2.91**	-0.25	-0.33*
	(0.94)	(0.04)	(1.74)	(1.76)	(1.21)	(0.04)
Observations	120	120	120	120	120	120

Note: *p<0.1; **p<0.05; ***p<0.01; FE: fixed effect, RE: random effect

Source: The calculations are done by the author, the data comes from the Eurostat's regional database.