Pleticha, Petr

Working Paper

Entrepreneurship in the information age: An empirical analysis of the European regions

IES Working Paper, No. 26/2018

Provided in Cooperation with:
Charles University, Institute of Economic Studies (IES)

Suggested Citation: Pleticha, Petr (2018) : Entrepreneurship in the information age: An empirical analysis of the European regions, IES Working Paper, No. 26/2018, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:
https://hdl.handle.net/10419/203206

Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

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

Petr Pleticha

IES Working Paper: 26/2018

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Pleticha P. (2018): "Entrepreneurship in the Information Age: An Empirical Analysis of the European Regions" IES Working Papers 26/2018. IES FSV. Charles University.

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

Petr Pleticha[a,b]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 21, 110 00, Prague, Czech Republic
[b]CERGE EI, Politickych veznu 7, 11000 Prague, Czech Republic
Email (corresponding author): petr.pleticha@cerge-ei.cz

October 2018

Abstract:
Decelerating productivity in recent years raised questions about technology diffusion in the economy. This study focuses on one particular diffusion channel, entrepreneurship, and inspects the mechanics through which it interacts with digitalization. The composite indicator of digitalization is split into separate components which enables analyzing digitalization's interplay with entrepreneurship as a dynamic process. Based on the econometric analysis of Eurostat regional data covering the period 2008-2015, I find significant links between digitalization and entrepreneurship. Specifically, digitalization is associated with an increase in the rate at which firms are created and with a decrease in their survival rate after 3 years. The paper demonstrates that the interaction is dynamic in its nature as the effects of initial stages of digitalization reverse or vanish in its later phases. A sectoral analysis shows the persistence of the results across industries. Moreover, there is evidence that professional, scientific and technical activities are especially sensitive towards digitalization, experiencing strong, yet short-term shock in the firms' birth, death, and survival rates. Accounting for geographic variation reveals heterogeneity between regions but not large enough to affect the overall results.

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

1. Introduction

Digitalization[1] and entrepreneurship are both focal points of policies aiming to foster economic development. 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 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 a national and subnational 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 economic growth via digitalization and entrepreneurship is widely pursued in the EU and beyond.

In this paper, I aim to provide empirical evidence on the relationship between digitalization and entrepreneurship in the European regions. Extensive literature proves innovation, especially in the ICT (information and communication technology) sector, to be one of the main factors of economic growth (Cardona, Kretschmer, & Strobel, 2013). Similarly, entrepreneurship has been successfully linked to economic progress (Audretsch & Keilbach, 2004). However, it is not clear how technological progress and entrepreneurship affect one another. Is it the entrepreneurs who push technology forward or is it rather the spoils of technology allowing new businesses to grow? Does entrepreneurship serve as a diffusion channel for new technology? And if so, is there any observable dynamics of the relationship between digitalization and entrepreneurship? Specifically, does digitalization dynamically affect the firm creation and firm destruction?

To provide an answer to these questions, it is crucial to perceive digitalization and entrepreneurship as dynamic processes and inspect their interplay in different stages of their lifespan. Early stages of digitalization may provide opportunities to entrepreneurs, but they may not yet cause any disruption among the incumbent businesses. Similarly, with the digitalization reaching its peak, opportunities for new business may already be exhausted, yet the disruptive effect of the new technology still takes its toll among the maladaptive firms. I inspect these dynamics by dividing the digitalization process into three stages and analyzing how the individual stages relate to entrepreneurial activity, specifically to the firms' birth, death and survival rates.

The contribution to the existing literature is twofold. Firstly, I construct a digitalization index capturing the technological transformation, adapt to the needs of regional analysis, and split it to capture different digitalization stages. Secondly, I investigate the relationship between digitalization and entrepreneurship. There is detailed literature describing how business produces and implements innovation (Baden-Fuller & Haefliger, 2013) but there is surprisingly little inquiry into the effects of advancing technology on entrepreneurship. The exception is the recent paper revealing positive impact of mobile broadband on entrepreneurship (Alderete, 2017). This study aims to contribute to this still thin strand of literature by inspecting the dynamism of digitalization's effect on business activity.

[1] This paper uses the term *digitalization* meaning a process of integrating new digital technologies into business models as well as our everyday lives. Digitization, on the other hand, refers to process of converting any information into a digital format (Gartner, 2017).

The rest of the paper is organized in the following way. I first introduce the relevant literature regarding entrepreneurship and digitalization. A theoretical framework putting the empirical analysis into the context of the real economy comes next. I describe the architecture of digitalization index assembled for the purposes of this paper as well as data and methods needed for the analysis. The results section is followed by an inquiry into regional and sectoral heterogeneity. Lastly, I briefly discuss limitations of the study and conclude the paper.

2. Literature review

The productivity puzzle discussion has been a hot topic in economics for decades. The grim notion of secular stagnation predicting long periods of near-zero productivity growth (Gordon, 2014) is countered with the optimistic view of rapid innovation where technology itself is the source of soaring productivity (Brynjolfsson and McAfee, 2014). Brynjolfsson and others claimed that the observed productivity deceleration is caused by mismeasurement; that it is a mere mirage. Recently, Syverson (2017) admitted that a fraction of the productivity gap can be attributed to mismeasurement issues, but illustrates that the gap is too large to be explained by distorted metrics. Dysfunctional diffusion dynamics can be another cause of the gap. Decker et al. (2014) use US data to show that entrepreneurship has been declining in recent decades. Because entrepreneurship is known to be a vessel for innovation implementation (Audretsch et al., 2008) and the research and development efforts are on the rise, it seems that the modern economies could struggle to spread the new ideas across the business sector.

2.1. Digitalization

When addressing the aspects of digitalization, most scholars focus on ICT capital (Cardona et al., 2013; Edquist & Henrekson, 2017; Miller & Atkinson, 2014). Others also analyzed phone coverage (Muto & Yamano, 2009) and broadband penetration (Crandall et al. , 2007; Thompson & Garbacz, 2008). 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.

Katz et al. (2014) fill this gap by proposing a digitalization index measuring the 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 a high level of technological literacy which can be estimated by usage of social networks and exploitation of 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] 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.

2.2. Entrepreneurship and its role in the economy

Entrepreneurship is difficult to quantify and include in models of neoclassical economics, so the economists had simply overlooked it (Baumol, 1968). This had changed over the last 15 years when entrepreneurship experienced its comeback, and databases such as the Global Entrepreneurship Monitor (GEM) were assembled. Despite Baumol (1996) who argued that virtue of entrepreneurship depends on the structure of the economy, empirical studies of developed countries predominantly point to the benign effect of entrepreneurship on economic growth (Erken et al., 2016; Galindo & Méndez, 2014; Holcombe, 1998; McQuaid, 2002).

The weaker impact of entrepreneurship in developing countries (van Stel et al., 2005; Wong et al., 2005) should not be taken as a proof of rent-seeking entrepreneurship. It could 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 (N. Bosma & Schutjens, 2011).

There is also a substantial variation between qualitatively similar nations and regions. Explaining the variation, 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. This paper follows their framework when building the empirical model.

3. Theoretical framework

The perception of both entrepreneurship and digitalization being agents of economic progress stems from their embedment in economic theory. Technology is a part of virtually all growth models and is considered as the decisive factor in long-term growth. Entrepreneurship is also closely linked to commercial activity in the economic theory. The traditional Schumpeterian view presents entrepreneurs as disrupters of economic equilibria, participating in creative destruction. The Schumpeterian entrepreneur introduces new technology, and by doing so, she pushes the economy off its equilibrium, only to be moving it swiftly towards the new and also more productive steady state (Schumpeter, 1961; Holcombe, 1998). In contrast, the Kirznerian entrepreneur exploits opportunities which are available in the economy taking the technology progress for granted. The entrepreneurial activity has a correcting effect on the economy given the new level of technology (Kirzner, 1997). Following this school of thought, advancing technology creates opportunities for entrepreneurs to exploit, suggesting a one-sided causal relationship. Hence, digitalization would contribute to economic growth via entrepreneurship as its diffusion channel.[3]

The main distinction is the direction of the causal relationship between technology and entrepreneurship. In this study, I focus solely on the Kirznerian entrepreneur as I inspect the effects of digitalization on entrepreneurship. Digitalization's stimulating (or stiffening) impact on entrepreneurs is transmitted through several diffusion channels. Firstly, the new technology enables the 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 digital revolution. An entrepreneur is naturally more aggressive in

[3] 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).

introducing the new products and services to the market which translates into an increased birth rate of companies (N. S. Bosma & Levie, 2010).

Secondly, current products and services are improved based on the new technology. Incumbent firms can defend their market share against the new entrants by embracing new technology and improving their products. They have a competitive advantage of in-house expertise and means for investment. Garcia-Macia et al. (2016) indeed found that most of the incumbents' innovation happens through improvement rather than the introduction of new product varieties. This possibly deters a portion of eager entrepreneurs from entering the market. It also lowers 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.[4] In either case, the survival rate of the recently created firms falls, and the death rate rises.

Thirdly, it is not only the firm's product which requires improvement to maintain its competitiveness. As a general purpose technology, digitalization is used in process innovation, positively affecting company's efficiency (Evangelista & Vezzani, 2010). It unlocks the economies of scale enabling companies to grow and acquire others. Increasing efficiency of the incumbents works as a hurdle for the market entrants' success because entrepreneurs may fail to foresee the change in other firms and underestimate the competition. This leads to the reduced survival rate of the newly created companies.

Lastly, digitalization decreases transaction costs.[5] Whether it is the cost of sharing information or rapidly decreasing transportation costs (Hummels, 2007), a decline in these costs results in new supply chains. It is unprecedentedly easy to access a distant market without having the knowledge of or representation on the market. Such reorganization of whole industries puts a great strain on those firms which are not adaptive enough, and 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 result in a higher business turnover.

Using this framework, I lay out three major hypotheses which can be subject to statistical testing:

Hypothesis 1: Digitalization *increases* the rate at which new firms are being founded.

Hypothesis 2: Digitalization *increases* the rate at which firms cease their operations.

Hypothesis 3: Digitalization *decreases* the rate at which newly created firms survive.

4. Data Description

The single source of the data is Eurostat's regional database. The NUTS 2 classification provides data on 292 comparable territorial units which is a desirable and unique level of granularity. The shortcoming of the database is the relatively limited time dimension and numerous missing values. Some countries, for

[4] However, not all mergers are result of technological progress. The topic of mergers and acquisitions is further discussed in section 4.2.

[5] I focus on inter-firm transaction costs. Lower intra-firm transaction costs would actually lead to stiffer competition as the position of the incumbents would strengthen. However, the fragmentation of value chains in recent years suggest that the inter-firm effect is the dominant one (Baldwin & Lopez-Gonzalez, 2015).

example, do not report data on business demographics at the regional level. Those countries have been eliminated from the dataset. The remaining regional data cover the time period from 2008 to 2015.

Certain observations 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).[6] This effective method yields satisfying results for as much as 50% of missing values in the dataset (Raghunathan et al., 2001; Schenker & Taylor, 1996). Following these scholars, I use only regions with less than 50% of missing values which reduces the number of covered NUTS 2 regions to 174.

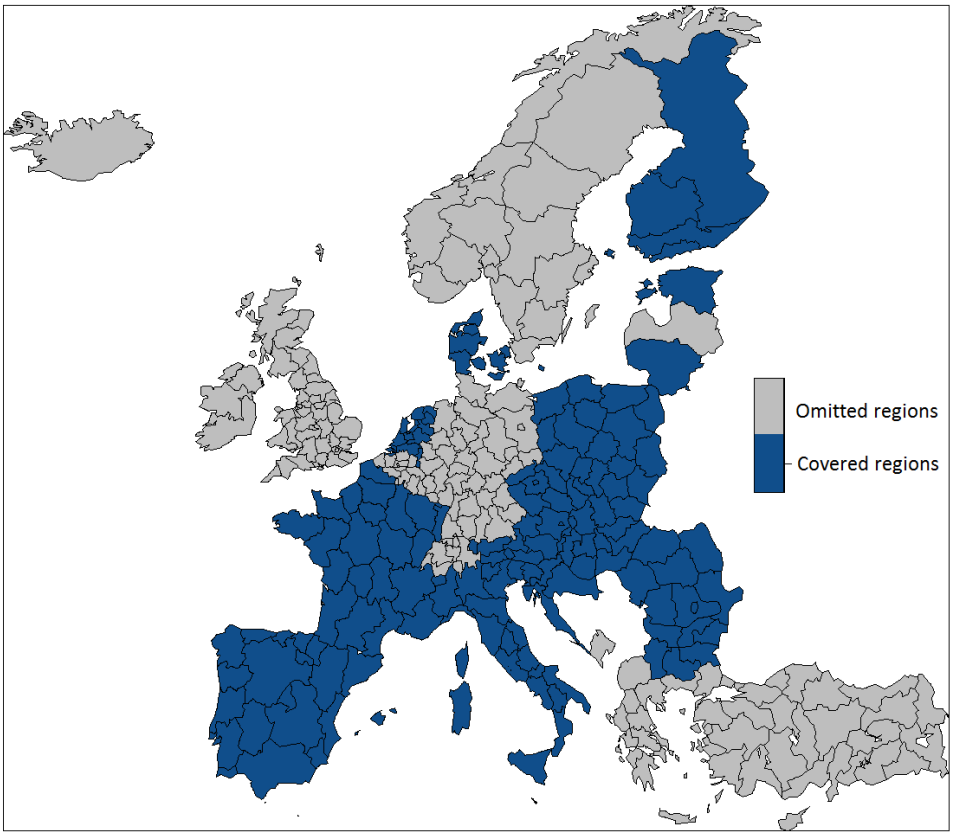

Figure 1: Covered regions

4.1. Digitalization

To analyze digitalization as a dynamic process, I define its three stages: opportune environment, easy access, and widespread adoption. Although the composite indicator of digitalization focuses on its dynamics and its logic is different than that of Katz, it needs to be stressed that the individual components overlap with Katz' indicator and it is mainly their grouping which is different from his 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

[6] The MICE algorithm with predictive mean matching is described in the appendix.

estimate its effects on business demographics in time. Table 1 provides a description of the digitalization variables.

ICT labor	percentage of knowledge-intensive labor employed in ICT sector
ICT investment	gross capital formation in the ICT sector per GDP per capita
Broadband penetration	percentage of households having access to a broadband connection
Internet penetration	percentage of households having access to an internet connection
E-commerce	percentage of the population who ordered goods or services online last year
E-government	percentage of the population who used internet for communication with public authorities during last 12 months

Table 1: Digitalization variables

The variables are grouped into separate stages based on economic intuition and factor analysis. A simple analysis with three factors and varimax rotation reveals three distinct groups of variables. Moreover, each factor is important for explaining the variance within the data as the sums of squared loadings higher than one suggest. Adding one more factor yields an increase in explained variance by 1% with sum of squared loadings 0.05 which is further evidence for having 3 relevant factors in the data. Table 2 below reports the results of the factor analysis.

		Factor 1	Factor 2	Factor 3
Stage 1: Environment	ICT labor	0.18	0.19	**0.67**
	ICT investment	0.00	0.00	**0.74**
Stage 2: Access	Broadband penetration	**0.86**	0.37	0.18
	Internet penetration	**0.82**	0.47	0.19
Stage 3: Adoption	E-commerce	0.49	**0.77**	0.11
	E-government	0.39	**0.85**	0.17
SS loadings		1.91	1.71	1.13
Cumulative variance		0.31	0.61	0.82

Table 2: Digitalization variables, factor loadings

Stage 1 consists of ICT investment and ICT labor indicators. Investment in digitalization is difficult to measure. To overcome the lack of data, I use overall gross fixed capital formation in the ICT sector. The implicit assumption is that digitalization investment is strongly correlated with that in the ICT sector and the latter can thus be used as a viable proxy. Further, I scale the investment data to the local GDP per

capita, and I use the harmonized index of consumer prices to adjust the data to inflation.[7] Apart from capital formation, the competent labor force is another prerequisite for developing and implementing new technology. To control for the level of the labor dedicated to digitalization, I use percentage of knowledge-intensive labor employed in the ICT sector.

The next stage makes the technology widely accessible. Because the variables capturing internet, computer and mobile penetration are highly correlated, I use only internet and broadband penetration. The former reflects baseline access to the technology which is sufficient for basic digital services such as email or web browsing. Broadband penetration, on the other hand, reflects more advanced access to the new technology, making, for instance, streaming or cloud services available.

Even if widely accessible, has digitalization impacted significant aspects of people's behavior? I measure the degree to which digitalization permeates our everyday lives by focusing on two variables. Firstly, the proportion of the population which ordered goods or services online during the last 12 months reveals shifts in consumer behavior. Secondly, the percentage of people who engaged with the government agencies using the web reflect the shift in the public sector which tend to be more resilient towards new technology than the private sphere.

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 a more detailed glance into the underlying relationship between digitalization and entrepreneurship, it misses the big picture. Policies enhancing digitalization hardly allow for cherry-picking consequences of digitalization. For example, a market economy with broad internet access and educated population naturally shifts a 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 after that 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 variables 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 in analyzing the effect of the prices themselves, a simple national dummy (or fixed effects estimator) controls for the 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 national aggregates would not benefit the analysis as they fall into the trap of national dummies.

4.2. Entrepreneurship and control variables

Previous studies used various variables such as measures of business demographics (Wong et al., 2005) or percentage of working population engaged in an entrepreneurial activity (Freytag & Thurik, 2007; Simón-Moya et al., 2014) to describe entrepreneurship. This study uses the former because of the

[7] 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 effect on entrepreneurship. Such issues can conceivably be controlled for in the future by using, for instance, hedonic price index.

availability of the data at regional level. Hence, I use birth, death, and survival rate after three years as proxies for entrepreneurship. Changes in those rates can be induced by different underlying factors, such as economic turbulences, gradual cultural transformation, or novelties in corporate law.[8] 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.

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 suggesting they are in a good position to leave their mark on the market. Whereas death rate is a proxy for the destructive element of the Schumpeterian entrepreneur, the survival rate is the long-term creative component of entrepreneurship, cleansed from possible whims and failures of both investors and entrepreneurs.

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. Economic factors appropriate for regional analysis are gross domestic product and unemployment. Further, I use public investment as a proxy for the quality of public infrastructure which is a way to control for the state of formal institutions (Afonso, Schuknecht, & Tanzi, 2005). The implicit assumption is that higher public investment is associated with environment benign to entrepreneurial activity. I scale the data as provided by Eurostat by the population of the region and I use 2005 prices.

Quantifying regional culture affecting entrepreneurship is problematic but conceivable. Cultural heritage represented by community trademarks weighted by GDP per capita is tracked by the EU and is thus a possible proxy for entrepreneurial ingenuity and originality of the region. It does not capture the culture per se; it rather reflects its consequences. In a culturally rich environment for entrepreneurship, more innovative and unique products are developed, and thus more trademarks are registered.

Lastly, I use a proportion of individuals who have completed secondary education to control for education as the last pillar of entrepreneurship. A secondary or tertiary education degree is often a prerequisite for certain professions. But secondary education, compared with tertiary education, is also less correlated with the measures of skilled labor, thus bringing more information to the analysis. Moreover, a higher level of education increases the pool of possible entrepreneurs who can identify not just opportunities in the labor market, but also business opportunities.

Table 3 presents summary statistics of all used variables. These statistics have been calculated before the data imputation.

[8] Mergers, acquisitions and spin-offs can also distort such measures. Unfortunately, reported M&A deals are likely only a fraction of all companies merged. The total of reported European deals in 2013 were only around 15,000 (IMAA statistics), but the grand total is probably much higher. Moreover, this study would need regional distribution of M&A's. Such distribution does not exist even for the reported deals.

	N (1069 without missing values)	Mean	St. Dev.	Min	Max
Broadband penetration	806	59.96	17.04	9.00	95.00
Internet access	806	65.62	15.14	17.00	96.00
E-commerce	814	30.38	19.26	1.00	83.00
E-government	508	38.55	18.43	3.00	88.00
ICT labor	930	2.24	1.52	0.60	9.20
log(ICT investment)	924	-4.93	1.31	-8.74	-1.11
log(GDP per capita)	1063	9.84	0.64	8.07	10.96
Unemployment	1050	10.47	6.14	1.90	37.00
log(Public investment)	924	-2.99	0.85	-7.34	-1.01
log(Community trademarks)	1000	-6.7	0.93	-10.14	-4.81
Secondary education	1050	70.03	16.17	18.00	97.30
Birth rate	1050	10.12	3.22	4.83	30.32
Death rate	861	8.66	3.27	3.78	18.79
Survival rate	1019	5.51	1.27	0.00	11.82

Table 3: Summary statistics

5. Methods

Even after deleting regions with more than 50% of missing values, the occurrence of missing data points is still high (see table 3). To maintain as much information in the data as possible, I do not delete any other observations. Little and Rubin (1987) suggest that if more than 5% of the data is missing, imputation of the missing values is advisable. Although it is also possible to use mean or median, such imputation is quite crude and leads to substantial bias in the imputed data. 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 data points. Specifically, predictive mean matching (PMM) estimates the coefficients describing linear dependency of the imputed and all the other variables. This is done using only complete observations. Based on the coefficients, 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 is explained in the appendix.

Once the full dataset is at hand, it is possible to construct the digitalization index. The construction itself is, however, subject to many arbitrary decisions of the researcher such as choice of variables and weights. The OECD Handbook on Constructing Composite Indicators (Nardo, Saisana, Saltelli, & Tarantola, 2005) is therefore used to adhere to the best practices and of composite indicator construction. The construction of the index requires normalization of the data. Standardization of the data to distribution with zero mean and standard deviation of one sustains the inherent data structure in a better way than alternatives and is thus used in this study. When setting the weights of the composite indicators, equal weighting is chosen for simplicity. Factor analysis reveals that their relative explanatory power is very similar anyway, so by equal weighting, I avoid an additional layer of complexity.

As hinted before, many omitted variables have national-specific characteristics. Therefore, I include a set of country dummies in the explanatory variable set to control for those effects. 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 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.

The regression equations for birth rate as explained variable are thus:

$$birth\,rate_{it} = \beta_0^1 + \beta_1^1 \cdot Digitalization_{it} + \gamma^1 \cdot Controls_{it} + e_{it}$$

$$birth\,rate_{it} = \beta_0^2 + \beta_1^2 \cdot Stage\_1_{it} + \beta_2^2 \cdot Stage\_2_{it} + \beta_3^2 \cdot Stage\_3_{it} + \gamma^2 \cdot Controls_{it} + e_{it}$$

The regression equations for death and survival rates are analogical. Having all the variables rescaled to standard normal distribution, the interpretation is straightforward. By moving up by one standard error in the digitalization index, we expect the region to move β_1^1 percentage points in the business demographics measure. The same applies to the stage variables.

Estimating the model with imputed data might lead to imprecise standard errors. Therefore, following Rubin (1996), I construct 100 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:

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

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

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

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

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

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

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

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

6. Results

The digitalization index yields foreseeable results. The Netherlands, Denmark, and Finland are the forerunners of digitalization in Europe. Surprisingly, the Southern states are virtually parring with Central Europe and Estonia. Not so surprisingly, Romania and Bulgaria (and to some extent Hungary) represent the laggards in digitalization. The country of origin notwithstanding, regions containing capital cities score significantly better than the rest (e.g., Madrid, Paris, or Prague). Figure 3 shows the detailed digitalization ranking in 2014.

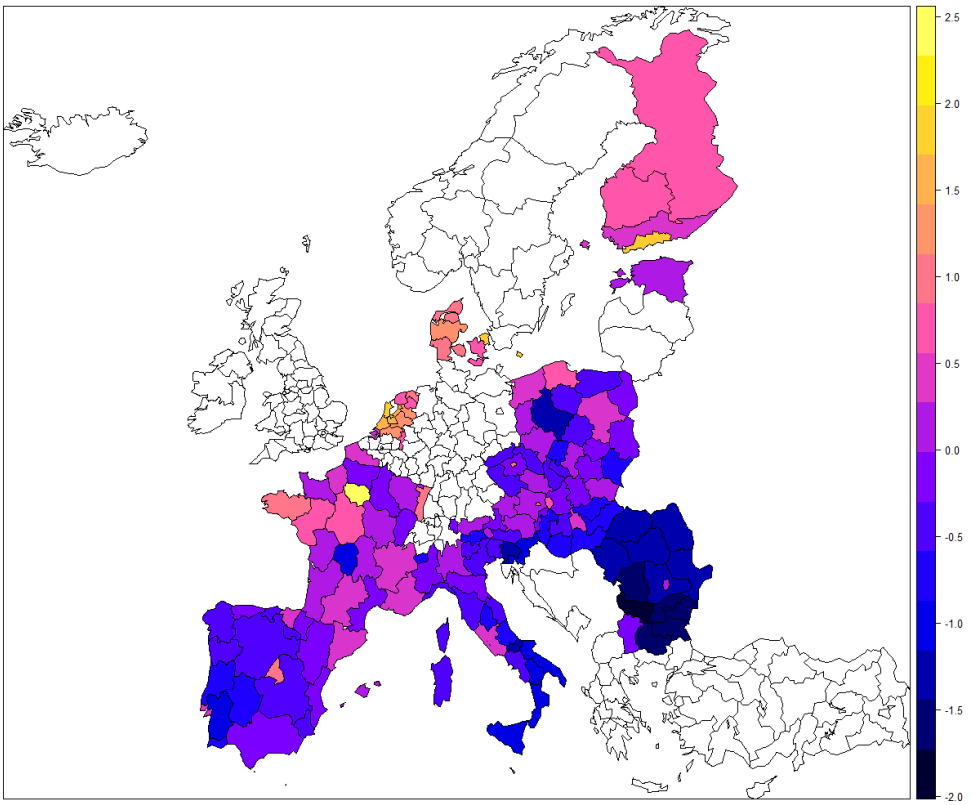

Figure 2: Regional digital development, 2014

Based on the Chow test for poolability, Breusch-Pagan test for individual/time effects and Hausman test for fixed/random effects, I present fixed effects and Generalized Method of Moments (GMM) models. Because all the tests can be run only within each regression (i.e., before the pooling), I pool the test statistics in accordance with Meng & Rubin (1992).[9] Fixed effects model seems to be the preferred one from the fixed/random effects and pooled model family. GMM is an attempt to counter endogeneity in the data and provide a robustness check with respect to the statistical method. Serial correlation and heteroscedasticity are countered with standard errors robust with respect to both.

Firstly, a holistic effect of digitalization on birth, death, and survival rates of businesses is estimated. Each component might affect entrepreneurial activities in a different way which undermines the statistical

[9] The results are robust with respect to the random element of the imputation. In other words, if the imputation is run with different sets of random numbers, the results stay intact.

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 4 shows the regression results based on 100 independently imputed datasets aggregated by Rubin's method.

	Dependent variable					
	Birth rate		Death rate		Survival rate	
	FE (1)	GMM (2)	FE (3)	GMM (4)	FE (5)	GMM (6)
Digitalization	**0.83***	**0.62****	**-0.03**	**-0.44***	**-0.37***	**-0.13**
	(0.34)	(0.23)	(0.31)	(0.22)	(0.16)	(0.13)
GDP per capita	-1.54	0.57	-2.03	-0.28	0.12	-0.44
	(1.23)	(1.42)	(1.18)	(1.22)	(0.52)	(0.46)
Unemployment	0.03	-0.04	0.06	0.16***	-0.05	-0.04
	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.04)
Public investment	0.12	-0.06	-0.7**	-0.49	-0.23*	-0.21*
	(0.25)	(0.19)	(0.26)	(0.25)	(0.11)	(0.08)
Cultural trademark	0.00	0.01	-0.00	-0.01	0.00	-0.00
	(0.00)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)
Secondary education	0.04	0.04	-0.02	0.02	0.03	0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Sargan test (p-value)	-	0.01	-	0.03	-	0.00
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1069	1069	1069	1069	1069	1069
Imputations	100	100	100	100	100	100

Table 4: Digitalization index regressions

As hypothesized, there is some evidence of a positive relationship between digitalization and the birth rate. One standard deviation increase in digitalization index can be associated with an increase in the birth rate of nascent businesses by 0.83 percentage points. The interpretation of such effect is difficult, but it is apparent that the economic significance of the effect is considerable. 4-sigma difference between regions in digitalization index (the difference between forerunners and laggards) is associated with 3.3 percentage point difference in the birth rate of companies. Taking the average business population of a NUTS2 region as the reference point, this difference translates into 4,200 more business solely due to advances in digitalization. Such difference in digitalization can be, for illustration, observed between Paris and any Romanian region except for Sofia.

The results of panel data regressions show strong links between digitalization and entrepreneurship, but they do not prove a causal relationship. That is because the data might suffer from endogeneity. The obvious ad hoc solution would be to use a 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, Masaki, & Pepinsky, 2015).

I try to control for endogeneity using GMM estimation. GMM instruments 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. Conducting Sargan test for suitability of the instruments used by GMM, it is clear that the lagged values of explained and explanatory variables are not persuasively good instruments. But due to the specificity of the data, it was unfortunately not possible to find better instruments. Tables 4 and 5 show the GMM estimates are mostly in line with the simple fixed effects regressions. Because of the weak instruments, that is hardly a surprise. The GMM analysis should thus be taken with a pinch of salt.

Focusing on fixed effect estimation, a negative link to the survival rate is observed. This is also in line with the stated hypotheses; indeed, the spike of new entrants caused by digitalization produces disproportionately more failed businesses. The race for reaping the spoils of digitalization produces many losers. Interestingly, there is no evidence of an immediate interplay between digitalization and the death rate of businesses. The GMM estimation even suggests that the effect is negative. Such finding can be interpreted as a piecewise evidence of weak disruption effect of digitalization on the business landscape. It is mainly the new entrants who fail (hence the drop in survival rate), but the overall business population remains largely unaffected.

The weak overall statistical significance of the models can be attributed to the composite characteristics of the index – there are simply different forces within digitalization pushing against each other. Therefore, the next model inspects different stages of digitalization.

	Dependent variable					
	Birth rate		Death rate		Survival rate	
	FE (1)	GMM (2)	FE (3)	GMM (4)	FE (5)	GMM (6)
Stage 1	**0.47**	**0.32**	**0.58**	**0.44**	**-0.15**	**0.02**
	(0.36)	(0.36)	(0.36)	(0.34)	(0.16)	(0.17)
Stage 2	**0.61****	**0.94*****	**0.21**	**0.02**	**-0.71*****	**-0.59*****
	(0.19)	(0.17)	(0.19)	(0.21)	(0.08)	(0.09)
Stage 3	**-0.04**	**-0.31**	**-0.26**	**-0.22**	**0.55*****	**0.48*****
	(0.21)	(0.19)	(0.21)	(0.2)	(0.09)	(0.09)
GDP per capita	-2.17*	-1.94	-2.68***	-2.22**	-0.45	-0.67*
	(0.93)	(1.02)	(0.77)	(0.83)	(0.32)	(0.33)
Unemployment	0.01	-0.06	0.04	0.09	-0.04**	-0.03
	(0.03)	(0.05)	(0.03)	(0.05)	(0.01)	(0.03)
Public investment	0.09	0.06	-0.77**	-0.58	-0.14	-0.11
	(0.25)	(0.26)	(0.25)	(0.24)	(0.11)	(0.1)
Cultural trademark	0	0.01	0	-0.02	0	0.01
	(0)	(0.03)	(0)	(0.02)	(0)	(0.01)
Secondary education	0.02	-0.36**	-0.04	-0.47**	0.01	0.43***
	(0.03)	(0.11)	(0.02)	(0.18)	(0.01)	(0.05)
Sargan test	-	0.02	-	0.00	-	0.03

Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,069	1,069	1,069	1,069	1,069	1,069
Imputations	100	100	100	100	100	100

Table 5: 3 stages of digitalization

Table 5 shows the outcome of regressions where the digitalization index is split into three temporal components. The statistical significance of important digitalization stages starts to be convincing. The increase in business creation happens mostly in the access stage of digitalization. This is intuitive. Only after the infrastructure for the new technology is in place, entrepreneurs begin to smell the new opportunity. On the other hand, once the market is saturated (the adoption stage), the new prospects begin to vanish.

Even when delving beyond the veil of aggregation, digitalization does not seem to force many firms out of the market. Efficiency gains associated with digitalization apparently do not lead to industry consolidation, at least at the aggregate economy level. The data rather point to industry fragmentation as more firms enter the market in the access stage of digitalization. This could be partially caused by decreased transaction costs among businesses. More elaborate value chains invite new firms from the whole world to participate in the production process, which leads to firms having a global reach with their network of subsidiaries. This could also inflate the observed firms' birth rate. However, no evidence of a link between digitalization and the development of the overall business population has been found.

This does not mean its long-term effect on the economic structure is negligible. Digitalization creates a more competitive environment for nascent businesses by reshaping the industry, if not by adding more firms in aggregate. Anecdotic evidence for this phenomenon is the tale of tech companies. We can see that the promising firms are getting quickly bought by the industry behemoths. So the firm has a long-lasting effect even if it ceases to exist relatively shortly after its creation. Indeed, the mindset of some entrepreneurs changed. Where the ultimate goal used to be building the business, it is nowadays selling the business.

The development of the survival rate reveals an interesting dynamic pattern. Complementarily to the behavior of birth rate, the survival rate experiences a drop in the access stage of digitalization. Indeed, the entering firms fail quickly. This drop is, however, offset by a subsequent increase during the adoption stage. Such results indicate an abrupt correction in the decreased survival rate; in other words, the wave caused by digitalization is severe but does not last long. This is, however, not the case when it comes to birth rate. A notion of a permanent increase in the entry rate of businesses is surely far-fetched, so a gradual fading of the effect seems more plausible. Hence the initial wave of new entries bring in a lot of 'gold diggers' doomed to fail, but once this wave is over, firms with 'normal' expected survival rate keep coming until the new opportunities brought about by digitalization are exhausted.

7. A step beyond aggregation

The aggregated data pose a challenge for the external validity of the analysis. The relationship between digitalization and entrepreneurship may vary with respect to the industry or location. Because the data can be disaggregated along these axes, I inspect the robustness of the results with respect to the sectoral classification and spatial distribution.

7.1. Sectoral breakdown

Although digitalization seems to interact with entrepreneurship at the aggregate level, the relationship could be very different in sectors which essentially facilitate digitalization compared to the traditional and more conservative sectors such as finance. It is possible to distinguish between industries which gain from digitalization in terms of the growing market (e.g., ICT sector) and those which are mainly disrupted by the new technology.

Such distinction is apparent in the data; however, the interpretation is not straightforward (see Table 6 below). Whereas the interplay of digitalization (and its stages) with survival rate is present in all industries, birth and death rates are more variant. It is only the ICT industry and professional, scientific, and technical activities whose birth and death rate are responsive to the digital transformation. Indeed, those are industries which, thanks to digitalization, managed to enter other market domains, mostly in the service sector. It is thus no wonder that the new entrants attempted to use the opportunity, especially if the opportunity was short-lived as the drop in birth rate in the last stage suggests. In case of the professional, scientific and technical activities, the death rate has the same dynamics as the birth rate, which suggests re-organization of the industry.

All measures of entrepreneurship are surprisingly inert to the first stage of digitalization. Generally, it is hardly surprising; without the access to the technology, there is no opportunity for the entrepreneurs to exploit. In case of ICT industry, however, the results are striking. It is this particular industry which creates the environment for the digital transformation, so how come there is no buzz as digitalization takes off? A possible explanation is that entrepreneurs respond with a delay. So the environment for digitalization in terms of investment and new hires takes place mostly within the incumbents of the industry. Only after the initial push, the new entrants join the digital rush.

Sector		Birth rate		Death rate		Survival rate	
		Aggregate	Stages	Aggregate	Stages	Aggregate	Stages
Manufacturing, mining, and quarrying	Digitalization	0.12	-	0.07	-	-0.2	-
		(0.08)	-	(0.1)	-	(0.13)	-
	Stage 1	-	0.07	-	0.04	-	-0.02
		-	(0.07)	-	(0.08)	-	(0.09)
	Stage 2	-	0.09	-	-0.07	-	-0.33***
		-	(0.05)	-	(0.06)	-	(0.07)
	Stage 3	-	-0.03	-	0.12	-	0.22***
		-	(0.05)	-	(0.07)	-	(0.08)
Information and communication technologies	Digitalization	0.4**	-	-0.09	-	-0.42**	-
		(0.12)	-	(0.11)	-	(0.13)	-
	Stage 1	-	0.02	-	-0.08	-	-0.09
		-	(0.09)	-	(0.1)	-	(0.11)
	Stage 2	-	0.47***	-	-0.15*	-	-0.48***
		-	(0.07)	-	(0.07)	-	(0.07)
	Stage 3	-	-0.21**	-	0.13	-	0.21**
		-	(0.08)	-	(0.08)	-	(0.09)
Financial, insurance, and real estate activities	Digitalization	0.04	-	0	-	-0.26	-
		(0.1)	-	(0.11)	-	(0.16)	-
	Stage 1	-	0.1	-	0.18*	-	-0.24*
		-	(0.08)	-	(0.09)	-	(0.12)
	Stage 2	-	0.04	-	-0.03	-	-0.54***
		-	(0.06)	-	(0.06)	-	(0.08)
	Stage 3	-	-0.05	-	-0.03	-	0.50***
		-	(0.06)	-	(0.07)	-	(0.1)
Professional, scientific, and technical activities	Digitalization	0.17	-	0.08	-	-0.35*	-
		(0.11)	-	(0.11)	-	(0.15)	-
	Stage 1	-	-0.02	-	0.01	-	-0.11
		-	(0.08)	-	(0.09)	-	(0.09)
	Stage 2	-	0.28***	-	0.28***	-	-0.87***
		-	(0.06)	-	(0.07)	-	(0.07)
	Stage 3	-	-0.20**	-	-0.29***	-	0.85***
		-	(0.07)	-	(0.08)	-	(0.09)
Controls		Yes	Yes	Yes	Yes	Yes	Yes
Time dummies		Yes	Yes	Yes	Yes	Yes	Yes
Observations		1,069	1,069	1,069	1,069	1,069	1,069
Imputations		100	100	100	100	100	100

Table 6: Sectoral breakdown, fixed effects estimation

7.2. Regional variation

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. The NUTS2 regions are constructed so that the individual regions are coherent units, but it is not possible to rule out spatial dependence between regions with close proximity. To deal with the problem, I perform geographically weighted regressions (GWR). GWR weights the observations based on a kernel function where the weight falls within the geographical

distance of the data point. Using each observation as a center for one regression, GWR yields one estimate for each geographical unit. It allows observing variation in the estimates among regions and establishing whether the outcomes are dependent on its spatial position (Brunsdon et al., 1998). The geographical location is represented by centroid calculated for each NUTS2 region.

The regional analysis sheds additional light on the links between digitalization and entrepreneurship. The generalized results obtained by fixed effects and GMM regressions are too blunt because the regions might be very heterogeneous. Looking beneath the aggregates, the coefficients estimates often vary substantially. This variation is, however, not translated into the conventional estimations and their standard errors, and thus represents an additional robustness test.

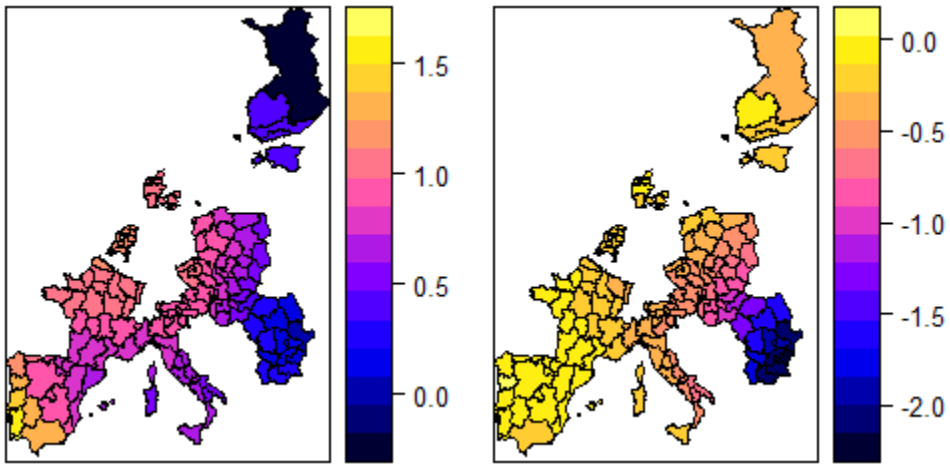

Figure 3: Coefficient estimates of digitalization and birth (left) and survival (right) rates from GWR regressions

Figure 3 illustrates this. Both regular and GMM regressions suggest that the birth rate of firms is positively (and survival rate negatively) associated with digitalization. But the inspection of the regional variation of the estimates shows the effect is not identical across regions. A similar level of heterogeneity can be observed in other coefficient estimates. The results suggest that whereas the positive effect on birth rate is less pronounced in south-eastern Europe, Finland, and Estonia than in the rest of Europe, the drop in survival rate in Romania and Bulgaria is a multiple of that in the rest of Europe. Despite the regional variance, however, the results stay intact even if Romania and Bulgaria are removed from the analysis.

GWR has significant shortcomings. The spatial distribution of the regions and their interdependence is only notional because of the omitted regions. Hence regions which are handled as peripheral in the analysis might be, in reality, not peripheral at all. Those naturally profess different behavior than the more central regions. That being said, reducing the bandwidth (i.e., the measure of interdependence), does not alter the results in a significant fashion.

8. Discussion and Conclusion

The digitalization index has been designed for purposes of regional analysis. Hence its construction reflects the need for distinguishing different regions as well as the availability of data. Data constraints also led to analyzing only the post-crisis period which might have led to results specific to this time period. For instance, low interest rates which were a norm in Europe exerted downward pressure on the cost of capital, which artificially increased the profitability of firms and thus the threshold for their survival. However, the accessibility of capital for smaller businesses was low in post-crisis years, so the specific time period could also affect birth, death, and survival rates. Implications of changing interest rates on entrepreneurship in the context of digitalization can be investigated by future studies.

Another issue is the data imputation process. It was carefully executed and documented, yet its suitability for spatially arranged data is not thoroughly investigated. For example, a conceivable extension would be to assign a greater probability to geographically close data points in the PMM algorithm. This would reflect the spatial dependence which I uncovered in the data. To the knowledge of the author, such methods are unfortunately not developed to this date. Such a methodological contribution might be a subject of future technique-developing study.

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 access stage of digitalization, but as the digital transformation peaks, the effects reverse or vanish. This along with the absence of the link to overall business population suggests that digitalization is a mere wave of technological progress and once it permeates the whole economy, its effect becomes largely 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.

The individual channels of digitalization diffusion, however, remain unaddressed and thus are subject to future research. I was able to show that the interplay between digitalization and entrepreneurship is significant, but it is not clear whether this interaction occurs via new products, adjacent products, product and service improvement, enhanced internal efficiency, or reduced transaction costs. I merely point to the direction of the channels, leaving quantification of each of those channels for the future research.

To conclude, this paper illustrates a significant relationship between digitalization and entrepreneurship. It also provides some evidence on the causality of the digitalization's effects on entrepreneurship. However, more data is needed to answer the causal question definitely. Dividing the digitalization process into 3 different stages, a dynamic pattern emerged showing that the negative effect on survival rate is offset in the later stages. The effect on the birth rate, on the other hand, seems to be a shock which only slowly fades away as digitalization matures. As disruptive as technology may be, there is no evidence that it pushes firms out of business.

References

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

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

Alderete, M. V. (2017). Mobile Broadband: A Key Enabling Technology for Entrepreneurship? *Journal of Small Business Management*, *55*(2), 254–269.

Audretsch, D. B., Bönte, W., & Keilbach, M. (2008). Entrepreneurship capital and its impact on knowledge diffusion and economic performance. *Journal of Business Venturing*, *23*(6), 687–698.

Audretsch, D. B., & Keilbach, M. (2004). Entrepreneurship and regional growth: an evolutionary interpretation. *Journal of Evolutionary Economics*, *14*(5), 605–616.

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

Baldwin, R., & Lopez-Gonzalez, J. (2015). Supply-chain Trade: A Portrait of Global Patterns and Several Testable Hypotheses. *The World Economy*, *38*(11), 1682–1721.

Bellemare, M. F., Masaki, T., & Pepinsky, T. B. (2015). *Lagged Explanatory Variables and the Estimation of Causal Effects* (SSRN Scholarly Paper No. ID 2568724). Rochester, NY: Social Science Research Network.

Bosma, N. S., & Levie, J. (2010). Global Entrepreneurship Monitor 2009 Executive Report. [Report].

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

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

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

Decker, R., Haltiwanger, J., Jarmin, R., & Miranda, J. (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives*, *28*(3), 3–24.

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

Erken, H., Donselaar, P., & Thurik, R. (2016). Total factor productivity and the role of entrepreneurship. *Journal of Technology Transfer*, 1–29.

Evangelista, R., & Vezzani, A. (2010). The economic impact of technological and organizational innovations. A firm-level analysis. *Research Policy*, *39*(10), 1253–1263.

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

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

Garcia-Macia, D., Hsieh, C.-T., & Klenow, P. J. (2016). *How destructive is innovation?* National Bureau of Economic Research.

Gartner. (2017). *Gartner IT Glossary*. Retrieved from http://www.gartner.com/it-glossary/

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.

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

McQuaid, R. W. (2002). Entrepreneurship and ICT Industries: Support from Regional and Local Policies. *Regional Studies*, *36*(8), 909–919.

Meng, X.-L., & Rubin, D. B. (1992). Performing Likelihood Ratio Tests with Multiply-Imputed Data Sets. *Biometrika*, *79*(1), 103–111.

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

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

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for Composite Indicators Building. *European Comission, Ispra*.

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

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

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

Thompson, H., & Garbacz, C. (2008). Broadband impacts on state GDP: direct and indirect impacts.

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

Van Buuren, Stef, & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3).

Van Stel, A., Carree, M., & Thurik, R. (2005). The Effect of Entrepreneurial Activity on National Economic Growth. *Small Business Economics*, *24*(3), 311–321.

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

9. Appendix

Data imputation

The PMM 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. The actual data imputation algorithm used in this paper is MICE developed specifically for R by Van Buuren & 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 an 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 (100 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.

The process itself is very efficient; it often achieves satisfying convergence only after 10 iterations (van Buuren et al., 2006). Following Van Buuren and Groothuis-Oudshoorn (2011), all used variables are included in the imputation process. A high number of predictors counter-intuitively 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. When imputing the data, the algorithm can create 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, adjusted prediction matrix was used such that the most correlated variables are not mutual predictors; only one direction of the prediction was kept. Removing strong mutual predictors from the regression prevents explosive behavior. Using the adjusted prediction matrix, the behavior of the variables follows the expected pattern resembling the white noise.

IES Working Paper Series

2018

1. Karel Janda, Martin Strobl: *Smoking Czechs: Modeling Tobacco Consumption and Taxation*
2. Karel Janda, Michaela Koscova: *Photovoltaics and the Slovak Electricity Market*
3. Simona Malovana, Dominika Kolcunova, Vaclav Broz: *Does Monetary Policy Influence Banks' Perception of Risks?*
4. Karolina Vozkova: *Why Did EU Banks Change Their Business Models in Last Years and What Was the Impact of Net Fee and Commission Income on Their Performance?*
5. Jan Malek, Lukas Recka, Karel Janda: *Impact of German Energiewende on Transmission Lines in the Central European Region*
6. David Svacina: *Devaluation with Exchange rate Floor in a Small Open Economy*
7. Ladislav Kristoufek: *Are the Crude Oil Markets Really Becoming More Efficient over Time? Some New Evidence*
8. Karel Janda, Zuzana Lajksnerova, Jakub Mikolasek: *A General Equilibrium Model of Optimal Alcohol Taxation in the Czech Republic*
9. Nicholas Tyack, Milan Scasny: *Estimating the Value of Crop Diversity Conservation Services Provided by the Czech National Programme for Agrobiodiversity*
10. Laure de Batz: *Financial Impact of Regulatory Sanctions on French Listed Companies*
11. Matej Opatrny: *Extent of Irrationality of the Consumer: Combining the Critical Cost Eciency and Houtman Maks Indices*
12. Mojmir Hampl, Tomas Havranek: *Foreign Capital and Domestic Productivity in the Czech Republic*
13. Miroslav Palansky: *The Value of Political Connections in the Post-Transition Period: Evidence from the Czech Republic*
14. Karel Janda: *Earnings Stability and Peer Selection for Indirect Valuation*
15. Ondrej Tobek, Martin Hronec: *Does the Source of Fundamental Data Matter?*
16. Stefan Schmelzer, Michael Miess, Milan Scasny, Vedunka Kopecna: *Modelling Electric Vehicles as an Abatement Technology in a Hybrid CGE Model*
17. Barbora Malinska, Jozef Barunik: *Volatility Term Structure Modeling Using Nelson-Siegel Model*
18. Lubomir Cingl, Vaclav Korbel: *Underlying Motivations For Rule-Violation Among Juvenile Delinquents: A Lab-in-the-Field Experiment*
19. Petr Jansky, Marek Sedivy: *Estimating the Revenue Costs of Tax Treaties in Developing Countries*
20. Yao Wang, Zdenek Drabek, Zhengwei Wang: *The Predicting Power of Soft Information on Defaults in the Chinese P2P Lending Market*
21. Matej Kuc: *Cost Efficiency of European Cooperative Banks*

22. Dominika Kolcunova, Tomas Havranek: *Estimating the Effective Lower Bound for the Czech National Bank's Policy Rate*
23. Petr Jansky, Markus Meinzer, Miroslav Palansky: *Is Panama Really Your Tax Haven? Secrecy Jurisdictions and the Countries They Harm*
24. Petr Jansky, Marek Sedivy: *How Do Regional Price Levels Affect Income Inequality? Household-Level Evidence from 21 Countries*
25. Mojmir Hampl, Tomas Havranek: *Central Bank Capital as an Instrument of Monetary Policy*
26. Petr Pleticha: *Entrepreneurship in the Information Age: An Empirical Analysis of the European Regions*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz