



Digital divide, craft firms' websites and urban-rural disparities—empirical evidence from a web-scraping approach

Anita Thonipara¹ · Rolf Sternberg² · Till Proeger¹ ·
Lukas Haefner²

Accepted: 26 June 2022 / Published online: 21 July 2022
© The Author(s) 2022

Abstract Following the “death of distance” postulate, digitalization may reduce or even eliminate the penalty of firms being located in rural areas compared with those in urban agglomerations. Despite many recent attempts to measure digitalization effects across space, there remains a lack of empirical evidence regarding the adoption of digital technologies from an explicit spatial perspective. Using web-scraping data for a sample of 345,000 small firms in Germany, we analyze the determinants of website prevalence. Comparing urban with rural areas, we show that running a website—as a proxy for the degree of digitalization of the respective firm—is highly dependent on location, whereby firms in urban areas are almost twice as likely to run websites compared with those located in rural areas. Our county-level analysis shows that a high population density, a young population and a high educational level have a positive and significant association with the probability that firms run websites. Surprisingly, we find a negative and significant association of gross domestic product per capita with website prevalence, which is driven by urban regions. There are no differences between urban, semi-urban and rural areas in terms of website up-to-dateness as well as social media prevalence. We conclude that there is a substantial digital divide and discuss policy implications.

Availability of data and material The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable

✉ Lukas Haefner
haefner@wigeo.uni-hannover.de

¹ Institute for Small Business Economics, Georg-August-University Goettingen, Heinrich-Düker-Weg 6, 37073 Göttingen, Germany

² Institute of Economic and Cultural Geography, Leibniz University Hannover, Schneiderberg 50, 30167 Hannover, Germany

Keywords Digital divide · Digitalization · Rural · Small and Medium-Sized Enterprises · Urban · Web-scraping

Digital Divide, Websites von Handwerksunternehmen und städtisch-ländliche Disparitäten – empirische Erkenntnisse aus einer Web-Scraping Analyse

Zusammenfassung Dem Postulat des „Endes der Entfernung“ zufolge kann die Digitalisierung die Nachteile für Unternehmen in ländlichen Gebieten gegenüber Unternehmen in städtischen Ballungsräumen verringern oder sogar beseitigen. Doch trotz einiger jüngerer Versuche räumliche Digitalisierungseffekte zu messen, besteht nach wie vor ein Mangel an empirischer Evidenz bezüglich der Nutzung digitaler Technologien aus einer räumlichen Perspektive. Mithilfe einer Web-Scraping Analyse untersuchen wir die Determinanten der Verbreitung von Websites für 345.000 kleine Unternehmen aus Deutschland. Der Vergleich städtischer und ländlicher Regionen zeigt, dass das Betreiben einer Website, als Näherungswert für das Digitalisierungsniveau eines Unternehmens, stark von dessen Standort abhängt. Unternehmen in Städten betreiben fast doppelt so häufig eine Website wie Unternehmen in ländlichen Regionen. Unsere Analyse auf der Kreisebene zeigt, dass eine hohe Bevölkerungsdichte, eine junge Bevölkerung und ein hohes Bildungsniveau einen positiven und signifikanten Zusammenhang aufweist mit der Wahrscheinlichkeit, dass ein Unternehmen eine Website betreibt. Überraschenderweise finden wir einen negativen und signifikanten Zusammenhang zwischen dem Bruttoinlandsprodukt pro Person mit der regionalen Verbreitung von Websites. Dieses Ergebnis wird maßgeblich durch städtische Regionen bestimmt. Wir finden keine Unterschiede zwischen städtischen, semi-urbanen und ländlichen Regionen bezüglich der Aktualität sowie der Verbreitung von Social Media. Wir folgern, dass es einen erheblichen digital divide gibt und stellen Ansätze für politische Maßnahmen vor.

1 Introduction

There is a public perception of a growing divergence in the adoption of digital technologies among urban and rural firms. Digitalization is considered to be the major technological development driving future business models and thus regional and firm-level competitiveness. Acknowledging a growing divergence in digitalization levels could warrant far-reaching innovation and digitalization support for rural firms to prevent a further polarization in competitiveness. Such support is timely since the adoption of digital technologies can be of a cumulative nature: once a firm lacks the capacity to adapt to new digital developments, further developments are even more difficult to cope with, eventually causing a further growing polarization. This is problematic because the unequal distribution of competitiveness and wealth among firms and regions is already leading to social tensions in many countries (Dijkstra et al. 2020; Rodríguez-Pose 2018). Digitalization thus poses a risk to social cohesion when new digital technologies are primarily adopted in and benefit thriving regions. Additionally, digitalization yields positive and negative impacts at the

same time. For example, workers performing routine tasks face a high risk of substitution through digital technologies, while workers executing complex tasks tend to be complemented (Akerman et al. 2015). This could amplify the threat for social cohesion when thriving regions benefit from digitalization, while at the same time less-wealthy regions suffer from job degradation or job automation. Consequently, policy-makers at many geographical levels have created government programs to improve the quality and quantity of broadband access—the fundamental prerequisite to compete in digitalization—in all regions and all types of regions.¹

Contrary to the threat of causing further polarization in competitiveness among firms and regions, politicians often see digitalization as an opportunity to reduce inter-regional disparities (BMI 2019; Grimes 2005). Especially in regions that have previously largely been excluded from markets or value chains, digital technologies based on the internet could enable the better inclusion of firms (Galloway et al. 2011). The assumption of growing opportunities for these firms was based on the expectation of a decreasing importance of physical proximity, eventually triggering the “death of distance” (Cairncross 1997). Once internet infrastructure was sufficiently available in all regions, it was expected that firms could digitally exchange knowledge, services or even products. Firms in marginalized rural regions could thus overcome disadvantages through integration into previously-inaccessible networks and value chains, whereby these regions could consequently catch up with more advanced regions.

However, such digitalization-induced catch-up processes have hardly been witnessed, prompting some authors to revise their expectations based on a more recent perspective (Cairncross 2018). Even in the digital era, there are “mountains in a flat world”, whereby geographical proximity still matters for the location of economic activity, as Rodríguez-Pose and Crescenzi (2008) put it. One prominent explanation for the absence of catch-up processes of laggard regions is the digital divide among firms in rural and urban regions (Salemink et al. 2017), for which two levels can be distinguished (Büchi et al. 2016; Scheerder et al. 2017). The first-level digital divide refers to access: while nowadays most firms in rural regions have access to the internet, the quality of the connection often trails behind that in urban regions (Briglauer et al. 2019b; Prieger 2013), which thus limits rural firms' opportunities of digital participation. The second-level digital divide refers to usage, whereby the adoption of not only the internet but also other digital technologies is lower in rural firms compared with their urban counterparts. This is not necessarily due to missing availability (i.e. due to the first-level digital divide) but rather to unfavorable socio-demographic characteristics of many rural regions: Research—which is mostly carried out on households rather than firms—indicates that high income, high professional qualification levels and a young population and labor force are central determinants in fostering the regional usage intensity of digital technologies (Blank

¹ See e.g. the Digital Agenda for Europe for an attempt at the supra-national level of the EU (<https://ec.europa.eu/digital-agenda/en>), the “digital agenda” for an attempt at the national level (for Germany, see the Federal Government of Germany 2018) or the “Broadband for Bavaria” initiative for the sub-national level of the German Federal State of Bavaria (https://ec.europa.eu/commission/presscorner/detail/en/ip_19_6630; see also Briglauer et al. 2019a).

et al. 2018; Billon et al. 2016; Prieger 2013; Schleife 2010). Often these indicators are deficient in rural regions.

While an unequal access to the internet (first-level digital divide) is easily observable, it is less straightforward to determine a spatial divide in firms' actual adoption of digital technologies (second-level). However, exactly this adoption is needed to seize the opportunities of digitalization, not only for firms but also for regions and countries as a whole (Whitacre et al. 2014a, b). Determining whether there is a second-level digital divide among urban and rural firms is therefore an important research goal to substantiate or refute political calls for a novel support of rural firms aimed at preventing a decoupling of digitalization dynamics. Empirical contributions face a number of obstacles, particularly in determining and compiling suitable indicators measuring digitalization in firms. Since digitalization is a very heterogeneous and sector-specific development, there are few common indicators that can be used to compare digitalization across industries and regions. To date, most indicators on the spatial digital divide mirror either the first-level or, when capturing the second-level, use household data or only small samples of firm data. Indices on the second-level are introduced for some study areas but most often do not reflect digitalization at a low spatial level and/or are not transferable to other study areas due to matters with data availability.

In this paper, we suggest using a firm's decision to create and promote a website in the yellow pages, its respective up-to-dateness and social media references as a proxy for its degree of digitalization. While these indicators primarily capture online marketing, we assume that its basic nature entails a firm's propensity to conduct further digitalization measures. Put simply, the assumption is that if a firm geared at business-to-consumer (B2C) services refrains from running a website as a near-costless measure of marketing, it is highly unlikely that this firm has invested in digitizing its internal workflows, let alone that it uses advanced digital technologies in its production. We are confident that these indicators capture a firm's basic propensity to conduct digitalization measures and hence enable us to test differences in digitalization across regions. Moreover, it can be assumed that using websites and social media has a fairly similar function for firms, even in very different (consumer-oriented) sectors. Since most other digital technologies are used quite heterogeneously across sectors, website prevalence enables us to compare larger samples of firms.

Using this approach to measuring digitalization, we run a web-scraping algorithm, which gives us roughly 345,000 firm entries of the yellow pages and about 104,000 websites, which are further analyzed regarding social media references and up-to-dateness. Using the firms' postal codes, we link county-level data for various socio-economic indicators. This provides us with evidence regarding the occurrence and potential structural determinants of a digital divide among German SMEs. We more specifically focus the analysis on firms from the German craft sector, which is characterized by smaller, often family-led businesses, which constitute a substantial share of German SMEs and tend to lack behind in digitalization (see e.g. Proeger and Runst 2020).

This article's contribution to the literature is twofold. By harnessing hitherto-untapped but publicly-available data, our first contribution is of a methodological

nature. Research on the second-level digital divide and its spatial effects is suffering from a scarcity of indicators proxying the prevalence of digital technologies in firms, especially aggregated at the regional level for Europe (Billon et al. 2016; Ruiz-Rodríguez et al. 2017). We propose a novel indicator for measuring the digitalization degree of regions that is widely accessible and does not build on the digital infrastructure but rather the actual usage of digital technologies by firms. The indicator comprises a large dataset and a wide range of sectors and is likely to be feasible for many countries without raising measurement and data accessibility problems that are inherent in indices. In providing support for Kinne and Axenbeck's (2020) suggestion that exploiting large-scale website data offers advantages compared to traditional data sources and opens up new paths for answering (economic) research questions, our methodological contribution hence extends beyond research on the digital divide. Second, by applying this indicator, new insights into the spatial digital divide and its regional determinants are presented. While we confirm an urban-rural digital divide in website prevalence, the descriptive statistics do not show a substantial spatial divide in the technological characteristics of these websites. In line with existing literature, we find evidence that population density, the professional qualification of the workforce and a lower population age are important regional determinants that are positively associated with the regional digitalization level. Contrary to existing findings, the wealth of regions (measured by gross domestic product, GDP) is only positively associated with website prevalence in rural areas. In urban areas, GDP is negatively associated with website prevalence. Finally, our results tend to show new evidence that lacking internet access is still linked to lacking digitalization efforts in some rural regions.

The remainder of this paper is structured as follows. Section two reviews the relevant literature for our research goal, before section three presents our method and dataset. Section four presents our results, while section five discusses our results and section six concludes and presents policy implications.

2 Literature review

2.1 Related research

2.1.1 Definitions of digital divide

In principle, a digital divide refers to the different availability, usage and usability of digital infrastructure, which may lead to gaps between people and groups of people with specific attributes within a society. Such a divide may be related to various joint characteristics of the named people or groups of people, e.g. their age (old vs. young), income (rich vs. poor) or education (university degree vs. no degree). In our paper, we follow the most common interpretation in the academic literature (see Graham 2019), which refers to spatial entities (territories or spatial units) when referring to a digital divide. These territories may be assigned to all possible spatial scales, such as the global (e.g. comparison of continents), national (comparing countries), regional (i.e. comparing sub-national ones like the 50 states in the US or the

sixteen federal states in Germany) or the local scale (e.g. comparing the quarters of a city). In addition to these spatial units based on administrative boundaries, comparisons between urban and rural regions have attracted strong attention among scholars (see Rodríguez-Pose 2018) in recent years, partially driven by specific policy programs and interests in ensuring equal living conditions in all regions of a country (Briglauer et al. 2019a). These comparisons between urban and rural areas may be intra-national (all areas within a country), cross-national (all areas in various or all countries within the same continent) or cross-continental. A plausible—and often-used—proxy for the rural-urban dichotomy is population density, e.g. measuring the number of inhabitants per km², as urban areas exhibit a higher population density compared to rural areas. While this method has some weaknesses concerning comparisons between urban areas, it is a reasonable measure when comparing urban with non-urban areas (like rural ones). In fact, population density is a more precise indicator for detecting urban-rural disparities than the sheer population size of the related spatial entities.

There is an important stream of literature that distinguishes explicitly or—more often—implicitly between urban and rural areas when it comes to some socio-economic or demographic characteristics of the territories under study, i.e. based on variables of the regional context. These characteristics can be summarized by the dimensions of economic prosperity (e.g. measured by GDP per capita), human capital (e.g. measured by the share of inhabitants with an academic degree), and demographics (e.g. measured by the share of inhabitants older than 64 years). Of course, some of the values of several of these categories strongly correlate. As our paper is dedicated to German regions, the following empirical evidence may be of interest: in 2017, urban counties showed a higher annual GDP per capita (43,400 €; rural counties: 31,400 €), a lower share of inhabitants aged 65+ (20.7% in 2017; rural counties: 23.0%), and a higher share of the workforce (place of residence) with an academic degree (18.2%; rural counties: 10.2%) compared with rural counties (all data derived from INKAR²). Of course—and by definition—urban regions show higher population densities than rural ones. Therefore, it makes sense to develop hypotheses that address the aforementioned dimensions, whereby we also implicitly address the digital divide between urban and rural regions. Consequently, we focus our literature overview on the named dimensions in section 2.1.3.

In the following description of the state-of-the-art research on the (spatial) digital divide, we distinguish between the first- and second-level digital divide. While the former has already provoked a lot of empirical and some theoretical research, the latter has not. We will explicitly address urban-rural differences in our overview of the literature.

² INKAR (Indikatoren und Karten zur Raum- und Stadtentwicklung [Indicators and maps on spatial and urban development]) is a free tool provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (see <https://www.inkar.de/>).

2.1.2 First-level digital divide: Digital infrastructure and its spatial effects

There is no doubt that in all countries there remains a gap of digital infrastructure between sub-national regions in terms of quantity (i.e. availability; Prieger 2003 for the US and Wernick and Bender 2017, for Germany) and/or quality (e.g. speed of internet connections, see Riddlesden and Singleton 2014). Moreover, the spatial diffusion process of web-based technologies varies across different geographical regions in most industrialized countries (Papagiannidis et al. 2015). When assessing the potential effects of digital infrastructure in rural areas, it is relevant to note that high-quality digital infrastructure (e.g. super-fast internet) is available more often and earlier in economically strong urban areas in a given country, which—*ceteris paribus*—leads to an increase in inequalities between urban and rural areas. The persistence of this divide is often ascribed to liberalized telecommunications markets and the fact that internet expansion for telecommunication providers is more profitable in urban rather than rural regions (Briglauer et al. 2019a; Grimes 2003; for the role of municipalities in Germany, see Wernick and Bender 2017). In some cases, the spatial digital divide is even growing (Townsend et al. 2013), which is often the case particularly for novel technologies or next generation access, i.e. higher bandwidths (Briglauer et al. 2019b; Prieger 2013). Politicians have recognized this problematic divide and many policies have aimed to narrow this gap (Briglauer et al. 2019a), although they often quickly become outdated by new technological developments (Salemink et al. 2017).

Regarding the spatial effects of digital infrastructure, there is some empirical evidence in the literature if inter-regional differences between sub-national regions are considered (Bertschek et al. 2015; Grimes 2005; Malecki 2003). However, these empirical results are often contradictory regarding the extent, type and direction of such effects (Jung and López-Bazo 2020). This also holds true for the effects on rural vs. urban regions (for a detailed overview of the literature, see Haefner and Sternberg 2020). Some qualitative studies show positive effects of digital infrastructure for rural areas (Townsend et al. 2017). To give an example, in many rural areas in developing countries, digital infrastructure allows time to integrate such regions into global value chains and enables a low-cost and reliable exchange with extra-regional and international customers (Krone and Dannenberg 2018). However, only some quantitative studies find that these effects are also positive in a relative sense, i.e. they are stronger than the positive effects of the infrastructure in favor of urban regions (Fabritz 2013; Bertschek et al. 2015; Ivus and Boland 2015). Such relatively (compared with urban regions) strong effects of digital infrastructure are a necessary condition if digitalization is intended to be used as a regional policy instrument to reduce inequalities between urban and rural areas of a country, i.e. to achieve convergence instead of (further) divergence (Jung and López-Bazo 2020; Celbis and de Crombrughe 2018). If urban and rural areas benefited from digital infrastructure to the same extent, then inter-regional disparities between rural and urban areas would not be reduced. By contrast, many authors argue in favor of (often economically strong) urban regions regarding the positive spatial effects of digitalization (Camagni and Capello 2005; Ciffolilli and Muscio 2018). Digital infrastructure seems to have stronger positive impacts in urban rather than rural areas

in terms of both quantitative as well as qualitative aspects (Malecki 2002; Tranos 2016; WWG 2015). In particular, this seems to be true for young (or even new) and knowledge-intensive firms (see Mack et al. 2011; McCoy et al. 2018) and higher bandwidth levels (Briglauer et al. 2019a).

2.1.3 *Second-level digital divide: Spatial effects and regional determinants*

As shown above, research on whether digitalization (via the internet) mainly benefits urban or rural regions (and thus whether digitalization leads to increased or reduced inter-regional disparities) yields inconclusive results. The availability of digital infrastructure (high quantity and/or high quality) is nowadays only a necessary condition for digitalization's effects on the development of urban and rural areas (Evangelista et al. 2014; Tranos 2012) and an increase or decrease in economic disparities between the two types of areas. If this first-level component meets higher or even the highest demands, the regional economic effects would nevertheless remain very limited if people living and/or working in these areas lack the necessary motivations and skills required for exploiting the potential offered by a very high standard infrastructure. In other words, the second-level component—usage and usage skills—is decisive if a certain standard of the first-level component is provided (Evangelista et al. 2014). Studies confirm this view, showing that increased internet availability does not necessarily have positive (spatial) impacts, but rather that the usage matters (Mack and Faggian 2013; Whitacre et al. 2014a, b; Scheerder et al. 2017). Consequently, in order to answer the question asked above regarding whether digitalization mainly benefits urban or rural regions, it appears more fruitful to determine a spatial digital divide in usage rather than accessibility (Haefner and Sternberg 2020).

A number of studies comparing urban and rural regions find a divide in the usage of digital technologies, which does not stem solely from lacking access in rural regions (Moriset et al. 2012; Prieger 2013; Saleminck et al. 2017; Townsend et al. 2013). Thus, there must be some regional determinants apart from infrastructure availability that explain this second-level digital divide. Studying the impact of these determinants on the degree of digitalization and their regional variation can thus help to better understand the digital divide. So far, the amount of theoretical or empirical research regarding the impact of regional determinants on the urban-rural second-level digital divide is rather small, particularly for firms.

An important regional determinant for technology adoption that has been discussed is population density. Technologies are often developed and first adopted in cities, i.e. in densely populated areas, and subsequently diffuse to less densely populated, more peripheral areas (Saleminck et al. 2017). One reason for the early adoption in cities is the fact that infrastructure expansion for telecommunication providers is more profitable in more densely populated compared to less densely populated areas (Briglauer et al. 2019a; Grimes 2003). Beyond that, the process of (digital) technology diffusion is not only starting earlier in densely populated areas, the speed of diffusion is often expected to be higher, too. Once pioneers adopt a technology, information on this technology can—due to the advantages of direct personal interaction—spill over more easily and frequently from users to non-users

in densely populated areas, leading to even increased adoption rates compared to less densely populated areas (Billon et al. 2016).

Some scholars have addressed the role of employees' qualification for the use of digital technologies. The lack of highly-educated employees with profound digital skills is considered to be one reason for various problems faced by firms in rural areas when trying to exploit the opportunities offered by digital technologies in such areas (Camagni and Capello 2005; Grimes 2005; Prieger 2013; Saleminck et al. 2017). These differences in terms of human capital in general—and digital skills in particular—between the two types of regions result to a significant degree from the higher attractiveness of urban areas for the young well-educated workforce. For example, Blank et al. (2018) stress the key role played by education when trying to assess regional internet usage in the UK, even when controlling for the structural differences in terms of education between urban and rural areas. Billon et al. (2016) identify a specific pattern of ICT use by firms (not by households) that is predominantly determined by educational variables. Mack et al. (2011) emphasize the importance of broadband provision to knowledge-intensive firm locations, but of course such firms search for specific location criteria (at least partially depending on the preferences of their highly-qualified employees). Hereby in fact the educational level of the employees influences internet usage (albeit not internet access), and this also has a spatial dimension. In her spatially-sensitive study on internet usage in German regions, Schleife (2010) shows that on average people who have reached a higher level of education are more likely to use the internet than less-educated individuals. Thus, individuals living in areas with a larger proportion of highly-educated individuals are more likely to be surrounded by internet users.

Socio-demographic variables are often found to explain large parts of the regional variation of digital technology usage and thus the second-level digital divide. Age belongs among the most frequently-used indicators among these variables as the first law of the internet is that everything is related to age (Blank et al. 2018), at least nowadays. Older people are less likely to be internet users (Blank and Groselji 2014), which also seems to be true for Germany as Billon et al. (2016) underline that age is an important individual factor for the decision to start internet use in that country. Schleife (2010) shows that new internet users are not only significantly better educated but also younger compared to non-users. However, it remains to be seen whether the age variable will still be relevant for internet usage once the whole population comprises digital natives.

As for income, Mills and Whitacre (2003) stressed in their comparison between US metropolitan and non-metropolitan regions that differences in household attributes—particularly education and income—accounted for 63% of the metropolitan vs. non-metropolitan digital divide 20 years ago. LaRose et al. (2007) emphasize that broadband internet service adoption in rural communities has been caused by income, beside other determinants like prior experience with the internet, the expected outcomes of broadband usage, self-efficacy and age (but not education). Prieger (2013) states that for the US gaps in broadband usage between rural and urban households are greater for low-income households. Schleife (2010) shows that mean disposable household income is an important factor to explain differences

between early internet users and non-users in German regions. Finally, also Billon et al. (2016) stress the influence of GDP on technology usage.

The literature shows that it is not necessarily “rurality per se”, a low population density or geographical distance but also other regional characteristics that hinder extensive digitalization and digitalization-induced catch-up processes of rural regions. Apart from the central economic and socio-demographic determinants described above, other determinants for the regional digitalization degree that have attracted some attention include the sectoral mix of the economy, average firm size and institutional characteristics (Billon et al. 2016; Dengler et al. 2018; Moriset et al. 2012).

2.1.4 *Measuring the (spatial) digital divide*

Measuring the digital divide requires first measuring digitalization. Most indicators aimed at capturing digitalization and determining the digital divide between regions refer to the internet. While initial attempts to capture the digital divide have focused on mere access, research has more recently considered the speed of the available internet infrastructure (Forman et al. 2005a, b; Saleminck et al. 2017; Townsend et al. 2013) or the number of broadband providers (Prieger 2013). An even more important recent shift in the literature on measuring the spatial implications of digitalization refers to usage and skills (second-level digital divide) instead of access to digital infrastructure (first-level digital divide), at least within advanced economies (Büchi et al. 2016; Evangelista et al. 2014). Here, indicators for the usage degree of computers and the internet are considered, with data stemming mostly from surveys (Billon et al. 2016; Prieger 2013; Scheerder et al. 2017). However, it is important to note that these variables usually refer to households rather than firms. Kinne and Axenbeck (2020) analyze website data of German firms and find regional differences in the share of firms running a website, which can be used to examine the digital divide. Another possibility for assessing the regional digitalization degree is to focus on digital skills. Exploiting labor market data, Castellacci et al. (2020) construct an “e-skill task intensity index” of a region’s labor force for a panel of European regions. While both of the latter two approaches for constructing measures of the regional digitalization degree are novel and promising from a methodological perspective, to date they have not been used to rigorously analyze (the determinants of) the digital divide or measure spatial digitalization effects.

Proceeding beyond the adoption of single indicators, some digitalization indices have been elaborated that allow for a geographical digitalization comparison and thus determining a potential digital divide. The Enterprise Digital Development Index developed by Ruiz-Rodríguez et al. (2017) uses data for Spanish regions (NUTS2-level) and focuses solely on the digitalization degree of firms. It comprises four components: ICTs and connectivity (access) to the internet, the use of ICTs, e-commerce, and e-government. The most interesting indicators for this study are the indicators of the “use of ICTs” component, namely enterprises having a website or homepage, the use of social networks, cloud computing services used over the internet, employees using computers, enterprises that pay to advertise on the internet, as well as indicators of the “e-commerce” component, namely enterprises

sending e-invoices B2BG that are suitable for automated processing, and enterprises receiving e-invoices that are suitable for automated processing. These indicators are measured as the percentage of enterprises in a region, except for “persons employed using computers”, which is measured as the percentage of total employment. The European Commission (2020) has developed the Digital Economy and Society Index, which comprises data on countries rather than regions. It comprises five main dimensions: connectivity, human capital, use of the internet, integration of digital technology, and digital public services. The “integration of digital technology” dimension captures the degree of digitalization in firms. Its first sub-dimension of “business digitalization” includes the indicators of electronic information sharing, social media, big data, and cloud. Its second sub-dimension of “e-commerce” includes the indicators of SMEs selling online, e-commerce turnover, and selling online cross-border. Katz and Koutroumpis (2013) have developed a digitalization index that also enables comparisons between countries but not between regions. It comprises six components: affordability, infrastructure reliability, network access, capacity, usage, and human capital. However, it does not comprise any detailed information on firms' digitalization degree.

In this study, we propose using web-scraped data on the share of firms running a website as a proxy for the regional digitalization degree. This indicator has not yet been exploited to assess the spatial digital divide. The indicator does not rely on data on households or infrastructure, but rather firms' actual operation of websites, as a basic and widespread digital application. It is therefore well suited to assess the second-level digital divide. Beyond these content-related advantages, our indicator has more practical advantages compared to more traditional forms of data gathering: The data that we use is already available on a large scale, has a wide coverage of firms and sectors and comes at no cost. The indicator allows easy transferability to other spatial levels and study areas, which is an advantage compared to indices. The web-scraping process also enables the retrieval of other website characteristics of interest and therefore opens up a wide area for future research (see Kinne and Axenbeck 2020). Of course, there are also important downsides of the indicator. First, there are technical challenges (data retrieval, data harmonization, etc). Second—and more importantly—the indicator covers only one specific aspect of digitalization and its representativeness for other facets of digitalization remains unclear. This is a disadvantage compared to indices. Our indicator covers firms' usage of online marketing instruments and therefore represents a very basic form of digitalization. Other digital technologies like big data analytics or robots are more disruptive and supposedly more representative of the public and political perception of the current scope of digitalization. Moreover, our indicator does not capture digital skills. These are the prerequisite for usage, whereby studying skills could offer more straightforward policy implications than studying technology usage.

2.2 Hypotheses

It has been argued that regional determinants strongly influence the adoption of digital technologies and accordingly affect the second-level digital divide. However, especially for Europe, there is a lack of regional data for the usage of digital tech-

nologies in firms (Billon et al. 2016; Ruiz-Rodríguez et al. 2017), which underlines the need for developing new indicators. The regional degree of website prevalence in firms—addressed in our paper—serves as a proxy for the regional degree of digitalization. We expect that it is only partially dependent on access to high-speed internet, whereas other regional determinants also play a major role. We formulate four hypotheses for regional determinants that may explain the regional degree of website prevalence in firms.

H1: Population density is positively associated with website prevalence The first hypothesis argues that new information technologies and specific instruments like websites spread across space via personal interactions. Since a higher population density—as is usual in urban areas compared with rural ones—encourages personal interaction as well as technology supply, new technologies are earlier available and more extensively used in regions with more inhabitants per km² (Billon et al. 2016; Salemink et al. 2017). Network effects should encourage technology adoption because the regional number of users of a technology should positively influence the number of new users of this technology in the same region (Li and Shiu 2012; Schleife 2010).

H2: The level of professional qualification in the workforce is positively associated with website prevalence From a capabilities approach, it can be expected that higher shares of employees with professional qualification will increase website prevalence in firms and demand for online information by customers in a given region. This expectation is in line with previous findings in the literature. At the firm level, a number of studies have shown that professional qualification is a central determinant for the effective use of digital technologies (Akerman et al. 2015; Berger and Frey 2016a, b). From a spatial perspective, digital technologies seem to be used effectively particularly in regions with high levels of skilled labor (Salemink et al. 2017; Schleife 2010), which in turn benefit from digitalization (Atasoy 2013; Hasbi 2020; Mack and Faggian 2013; McCoy et al. 2018).

H3: Increasing average age of a region's inhabitants is associated with lower website prevalence Demographic variables have proven to be strong predictors of the use of digital technologies, not only from a spatial perspective (Büchi et al. 2016; Scheerder et al. 2017). In this regard, the age structure of a region is an important determinant of internet use (Prieger 2013; Schleife 2010; Vicente and López 2011). Older employees in firms tend to be less digitally skilled and less willing to accept organizational changes through digitalization (Billon et al. 2016; Blank et al. 2018; Evangelista et al. 2014). Behavioral patterns associated with digitalization are thus more strongly adopted among younger persons. Therefore, we expect that regions with a higher or increasing average age are less likely to be associated with high website prevalence.

H4: GDP per capita is positively associated with website prevalence In line with previous research, we expect GDP per capita to be a strong predictor of regional digitalization (Billon et al. 2016; Vicente and López 2011). The GDP per capita

of a given region is associated with firm competitiveness, higher education of the workforce and higher overall purchasing power, all of which are likely to affect firm innovativeness, which should be reflected in higher levels of digitalization (website prevalence). Furthermore, as digital technologies are not free of cost, their usage also depends on investment decisions (Schleife 2010) and it should therefore increase with household income and GDP (Li and Shiu 2012).

3 Method

3.1 Web-scraping

To obtain our data, we employ a web-scraping procedure, which was conducted in June 2018. Web-scraping can be described as a simple algorithm that downloads specific information from a predefined number of websites. Once fed with a large number of website links, the algorithm can download the predefined information efficiently and it can be used to build a dataset encompassing the respective variables. Despite still being a rather recent data collection method in regional research, there are already some promising attempts showing that the operation of websites in Germany and their characteristics such as the number of subpages and hyperlinks, text volume, language used differs according to—among others—the location and the innovativeness of the respective firm (see e.g. Kinne and Axenbeck 2020). To obtain information on firm websites, we use the German yellow pages, which provides us with a large set of B2C firm contacts, their respective postal codes and—if entered—the website address. We chose a broad set of industries³ that represent firms from the craft sector, focusing on consumers to prevent analyzing e.g. industry suppliers with a limited number of fix business contracts, which might not have the same incentive to use digital marketing compared with consumer-oriented businesses.

For firms that state having a website address, we conduct a further web-scraping procedure, which provides us with information concerning whether the website is actually online, when it was most recently updated and whether a social media reference (Facebook, Twitter, Instagram) is used. There are plausible arguments in favor of a relation between social media and website usage. One could assume that website prevalence is overall decreasing and that firms primarily use social media for digital marketing. Regarding the relevance of websites in comparison to social media, we would argue that—instead of a prevalence or substitution—social media *complements* websites. There may be industries or sectors in which social media has become more important than websites, with hairdressers being an example in our data (low share of websites, high share of social media plugins within the existing websites). However, it is basic digital marketing knowledge that a firm should be present on numerous (social media) platforms and channel visitors from those platforms to the firm's website, on which the actual sale or (pre-sale) contact is organized. Thus, a central website and numerous different platform presences are

³ The full set of industries is documented in Online-Appendix Table 4.

the optimal digital marketing strategy. Obviously, this often exceeds the capacities of smaller firms, which rely on single digital marketing channels. However, our main argument is that more digitalized firms do not cease to have websites but rather build a net of channels leading customers towards their website. Thus, an answer to the potential bias of our data in terms of a decrease in website usage would be that websites remains the central digital marketing channel, which is supplemented by social media channels and thus remains a valid indicator for digital marketing.

The information regarding whether a website have been updated is analyzed using three separate proxies. First, the “last update” variable indicates the year in which the website was last up-dated. This variable is only available for a small fraction of the sample, which is why we use two other proxies, namely “HTML5” (Hypertext Markup Language) and “HTTPS” (Hypertext Transfer Protocol Secure). HTML5 is the most recent language used for programming websites and it stands for a modern appearance of the website. The usage of HTTPS indicates whether a company has recently improved the website’s security as HTTPS has become a standard security protocol only in recent years, despite its first release in 1994. This could hence be considered another indicator of up-to-datedness.⁴

This approach has the strong advantage of bypassing a number of problems associated with direct surveys or interviews regarding the state of digitalization, including low response rates, the self-selection of digital pioneers or larger firms, overly optimistic self-assertions and regional or sectoral biases. Instead, a larger set of information can be used that directly measures firms’ behavior on a large scale across sectors and regions. The disadvantages of using the yellow pages as the source of firms’ websites include the heterogeneous structure of the yellow pages in Germany, historically different cost models⁵ for entering firm information as well as potential errors during the web-scraping procedure itself, potentially resulting in a loss of data. Furthermore, using the yellow pages might exclude digital pioneers that rely entirely on other platforms for finding customers. However, it is important to consider that internet research on (craft) services is not universal in Germany, particularly not among older people, who constitute a substantial share of

⁴ For an example of this procedure, consider a hairdresser in Berlin listed in the yellow pages. In the initial scraping procedure, we scrape all firms that have identified as hairdressers when registering in the yellow pages. Thus, each firm has a industry and postal code registered in the yellow pages. The web-scraping procedure enables us to download the information for all firms with a given occupation. Among them is the exemplary Berlin-based hairdresser. We obtain this firm’s industry, postal code and website address. In the second step, we use the website address to collect additional information on the firm, specifically whether the website is working, when it was last updated, whether there is a social media reference on the website and whether HTML5 and HTTPS are used on the website. Thus, the data obtained for a specific firm would look as follows: industry, postal code, website address, website working?, year last update, Facebook?, Twitter?, Instagram?, HTML5?, HTTPS?. The regional data was merged according to the postal code in a subsequent step and is therefore unrelated to the web-scraping procedure.

⁵ Overall, the costs of being listed in the yellow pages are negligible. The basic firm information including the website is costless, while firms can optionally pay for better search rankings and additional firm information such as links to social media. The costs for these measures are customized to regions and industries. Thus, we see no strong financial reasons for firms not being listed in the yellow pages. Furthermore, a firm could be listed in the yellow pages but not state a website. As there are no reasons for a firm not to state its existing website, we believe that this should only account for a neglectable proportion of the firms in the dataset.

the German population. For them, the printed yellow pages (which is replicated in the online version) remains a central source of information. Firms that choose not to be listed forego these potential customers. The yellow pages claim (based upon market research⁶) that in 2017 slightly more than half of the German population over 16 years has used the yellow pages in its different formats (print, online, smartphone app). This amounts to 32.6 million persons conducting roughly 900 million searches. Craft services are the second most important domain of search requests. Overall, while the market research is a form of marketing by the yellow pages, their numbers substantiate the claim that being listed in the yellow pages continues to make sense for firms. Finally, it can be argued that firms pursuing a modern digital marketing strategy conduct search engine optimization (SEO) to be listed prominently among Google searches. An important aspect of SEO is being listed on numerous platforms, which improves a firm's SEO score, whereby the yellow pages helps to increase this score. Therefore, digitally strong firms similarly have an incentive to be listed on the yellow pages for technical reasons (see e.g. Grappone and Couzin 2011). Nevertheless, we ultimately cannot assert that each firm has a strong incentive to use the yellow pages and enter its own website. Thus, we conclude that our dataset presents a large set of firms, while acknowledging that it might not present a fully representative picture of the craft sector; however, there are strong incentives for firms to be listed on the yellow pages and to state their website on the yellow pages.

3.2 Sample specifics

Our sample comprises craft firms (following the German legal definition of craft services, see e.g. Runst and Thomä 2020). This sector is in itself heterogeneous in terms of industries as it contains a historically-grown, legal set of occupations whose defining feature is its prevalence of manual work. The craft sector is considered to be rather slow in terms of digitalization, at least when compared to other sectors characterized by SMEs. It can thus be assumed that many craft firms continue to rely on the previously-dominant information platform, namely the yellow pages. Therefore, we assume that we reach a broader, more representative sample of craft firms using the yellow pages rather than with a straightforward Google search, which would not yield firms without a website. We chose the craft sector for our analysis of the digital divide because its firms can be found all across Germany and craft firms are also traditionally present in rural regions. Most industries are uniformly distributed across different spatial units (consider e.g. hairdressers) and the firm density primarily relies on population density. Thus, choosing the craft sector is likely to yield a spatially-balanced sample.

Another important reason is that craft firms have a lower technological performance than other sectors in the German economy, and they tend to be more traditional in their firm policies. Therefore, we are likely to witness a sector that is currently undergoing a digital transformation in the domain of digital marketing.

⁶ A summary of the respective study by gfk research from 2020 can be found at the yellow pages: <https://www.gelbeseiten.de/gsservice/Newsroom/Pressemitteilungen/Pressemeldungen-national/Knapp-jeder-fuenfte-Nutzer-von-Gelbe-Seiten-sucht-nach-einem-Handwerker>.

Thus, our indicator is likely to capture marked spatial patterns in the craft sector that similarly occurred and have continually disappeared in other sectors in recent years. Therefore, using a digitally relatively backward sector using the indicator of website prevalence is likely to show us transformative processes that have occurred in other sectors and are occurring in similar sectors. We would not suggest that our sample is representative of all firms in the yellow pages or all firms in Germany, but rather for the craft sector, which is an integral part of the German economy and had (official numbers for 2019) 1 million firms, 5.58 million employees—which translates to about 12% of all German employees—and an overall revenue of 640 billion Euros.⁷ Thus, overall we would suggest that our sample is representative of the craft sector and other sectors with smaller firms addressing private customers and a lower propensity to conduct digitalization measures.

Overall, our sample comprises 346,361 firm entries, of which 104,460 (30.2%) have entered a website in the yellow pages that was online and working in June 2018. We downloaded firm data from 44 different industries with quite heterogeneous numbers of firms and levels of website prevalence. An overview of the industries and respective number of observations is documented in Table 4 in the online appendix. We use 44 industries, which are the relevant craft occupations in the yellow pages. There are small craft industries that are formally part of the craft sector with very few firms in Germany, which have not been included in the analysis, for example ropemakers. Overall, our goal was to represent the craft sector on the yellow pages using the respective industries, which then yields 346,361 firms. The web-scraping procedure further yielded information on the existence of a Facebook/Twitter/Instagram reference in the website, as well as information on the most recent update and the use of HTML5 and HTTPS.

At the regional level, there was no limitation during the web-scraping procedure, and thus all firms of the respective industries were used, irrespective of their regional origin. We used the postal codes associated with the firm entries and matched this information to the respective counties. This approach leads to inaccuracies in some cases, whereby the associated error can be assumed to be uniformly distributed over all counties/postal codes. With firms clustered at the county level, we matched a number of additional county-level regional indicators, namely population density, the share of employees without a professional qualification, the share of employees with an academic education, the revenue of craft firms, annual GDP per capita, rates of regional in-migration and out-migration⁸, increase/decrease of the population

⁷ The data is from the German Federal Statistical Office, and thus our sample includes one-third of all officially-listed firms in German craft chambers (for a summary see <https://www.zdh.de/daten-fakten/kennzahlen-des-handwerks/>, last access: 27.10.2020). However, the official number of 1 million firms should be considered inflated, as non-trivial numbers of factually-inactive firms continue to be listed in the official records. Further, the yellow pages contain firms and industries primarily addressing private customers. This excludes firms, e.g. in the building sector, working as contractors for industry firms or public organizations, which makes the yellow pages irrelevant for these firms as they tend to react to public biddings. While there are no exact numbers to substantiate this claim, we are confident that our sample encompasses all active German craft firms addressing private customer and excludes inactive yet listed firms and those primarily working in B2B contracts or for the public sector.

⁸ As the rate of regional out-migration is highly correlated with the rate of regional in-migration we decide to only use the latter rate in the econometric analyses.

aged 65 and older, the number of employees in craft firms,⁹ and finally the share of households with high-speed internet access.¹⁰

Using the web-scraping data as well as the additional regional indicators, we can analyze the regional disparities in website prevalence, social media prevalence and the up-to-dateness of websites and show potential drivers of the differences using regional indicators.

3.3 Methodological approach/quantitative analysis

In a first step, we use descriptive statistics to analyze the share of firms running websites, social media as well as the up-to-datedness of the websites for each regional category (rural, semi-urban, urban). For this purpose, the dataset obtained by our web-scraping procedure is used. The second part of the analysis comprises a regression analysis. As our dependent variable y (website prevalence) is a binary outcome dependent variable—which is “1” if a company has a website entry in the yellow pages and “0” if not—we apply a linear probability model taking the following form:

$$y = \beta_0 + \beta_1 \bar{X}_1 + u \quad (1)$$

Estimating the equation:

$$\hat{P} = \hat{p}r [y] = 1|x] = x'\beta \quad (2)$$

The linear probability model models the probability of a firm having a website, meaning that website prevalence “ y ” equals “1” in dependence of x . This again is equal to a functional form of $F(x'\beta)$. x is a vector of independent variables. This vector includes the explanatory variables which were introduced in the hypotheses in section 2.3. (population density, share of employees with an academic education, share of employees without formal education, increase/decrease of the population aged 65 years or older, GDP per capita). Besides these variables we include a number of further variables:

- The firm's industry: as there are industries which are more technology affine than others (electrician vs. mason).
- The share of households with access to broadband internet (<50 Mbit/s): if the share of households with less than 50Mbit/s is greater in a region we expect the probability of a firm having a website to be lower.

⁹ More specific information on the variables is presented in online appendix Table 5.

¹⁰ To determine a region's access to high-speed internet, we use a dataset from the TÜV Rheinland and the German Federal Ministry of Transport and Digital Infrastructure at the county level. The dataset gives us the share of households that can access an internet speed of 50Mbit/s, which we use as the broadband internet variable. For urban counties, there is essentially no variation in the data as essentially all households can access 50Mbit/s, there is little variation for semi-urban counties and strong variation for rural counties.

- Sales revenue in the craft sector: as regions with higher sales revenues in the craft sector could possibly have a higher level of digitalization.
- Increase/decrease of birth rate: increasing birth rates may be an indicator for a younger population which would explain a higher level of digitalization as younger people tend to use digital devices and communicate digitally.
- Regional in-migration (as well as regional out-migration used in the robustness tests): we assume that regions with higher regional in-migration have a higher level of website prevalence as the influx of people to a region goes along with a certain degree of anonymity. Hence, firms would have to invest into a website in order to be recognized by new inhabitants.
- Share of employees in the craft sector in a region's labor force: this variable could be used as a proxy for competitiveness. If there are more employees in the craft sector in a region firms have to compete for customers but also for qualified workers. Hence, employees in the craft sector could have a positive effect on website prevalence.
- Firm size: One could argue that larger firms are more likely to have a higher level of digitalization. If larger firms are more likely to be situated in urban areas, the effects we find would only reflect the effects of firm size on the website prevalence instead of urban/semi-urban/rural regions (Kinne and Axenbeck 2020). We decide to include only firm size as a variable in the model as firm sales correlate with firm size. Data on firm size is not available on firm level as the firms do not report the number of employees on their website. Instead, we use data on average firm sizes by industries and by county from the sixteen Statistical Offices of the German Federal States.¹¹

A list of the variables used, their description as well as descriptive statistics can be found in Tables 4, 5 and 6 in the online appendix. We specify robust standard errors clustered by the German counties, which capture their unobserved heterogeneity.

We have also applied a logit model using a logistic distribution function as well as a probit model using a standard normal distribution function, whereby the results (and marginal effect sizes) in all three models yield similar results and effect sizes. We decide to report the results of the linear probability model estimated by ordinary least squares which allows effect sizes to be directly interpreted.

We run five separate specifications: (I) using all firms in all regions while excluding broadband access, (II) using all firms and all variables, using the sub-samples for (III) firms in rural counties, (IV) firms in semi-urban counties and (V) urban firms including all variables. We decide to split the dataset in order to run the last three specifications instead of only using dummy variables for each region type as we are particularly interested in the region type specific effect each variable has on website prevalence.

¹¹ In the rare cases that an industry-county combination comprises only very few firms, the average firm sizes have often been omitted by the Statistical Offices due to data privacy reasons. In these cases we proxy the omitted values by inserting the average size of the industry in Germany and adjust it by the relative average firm size of all industries in the respective county.

To confirm whether our results are robust, we conduct a number of robustness tests. First of all, Table 15 in the online appendix presents a correlation matrix of the main variables. The correlation matrix including all industries is not reported for space reasons. However, their correlation coefficients are smaller than 0.01 in the majority of cases and can hence be neglected. In our robustness tests, we drop step by step variables that show higher correlation coefficients. Overall, we conduct six robustness test specifications, each time using the different regional sub-samples (see Tables 9–14 in online appendix). The results remain robust throughout all robustness tests.

By matching regional data with firm-specific data received from web-scraping, the model is likely to suffer from the modifiable areal unit problem (MAUP). We acknowledge this problem and therefore run several robustness tests on different aggregation levels (county and regional planning level)¹² for each regional category (whole sample, rural, semi-urban, urban) to ascertain whether the results remain robust. As Table 16 in the online appendix shows, the results remain robust regardless of the aggregation level chosen. Hence, we are confident that MAUP does not confound our results.¹³

4 Results

4.1 Descriptive results

The basic finding regarding website prevalence is shown at the county-level applying the respective classification of all German counties into rural, semi-urban and urban ones.¹⁴ Fig. 1 shows how website prevalence differs between the three types of regions, with urban counties showing roughly a double share of website prevalence compared with rural counties. One can clearly see that rural firms have a substantially lower propensity to conduct digital marketing by establishing a website compared with urban firms. Thus, there is a clear and strong digital divide in terms of website prevalence, which warrants further analysis.

Thus, there is a clear and strong digital divide in terms of website prevalence, which warrants further analysis. However, as shown in Fig. 2, different from indicators of economic prosperity or economic growth, there is no clear east-west pattern or differences across Federal states. Urban areas, more or less, show higher values than the rural ones, independent from the Federal state.

Furthermore, by comparing firms' propensity to update their websites, we can use the broad classification of counties in rural, semi-urban and urban counties. Recall

¹² Using a smaller regional level is not possible due to data availability of the variables used.

¹³ Another possibility to circumvent this problem is using a multilevel approach as used in Tranos and Stich (2020) or López-Bazo and Motellón (2018). However, multilevel modeling is not applicable in our case as we do not have any firm-level explanatory variables except for the variable controlling for the firm's industry.

¹⁴ This classification is based upon population density and the role of large cities in the respective region. See <https://www.inkar.de/documents/Erlaeuterungen%20Raumbeuege19.pdf> for details. Last access 2021/11/24.

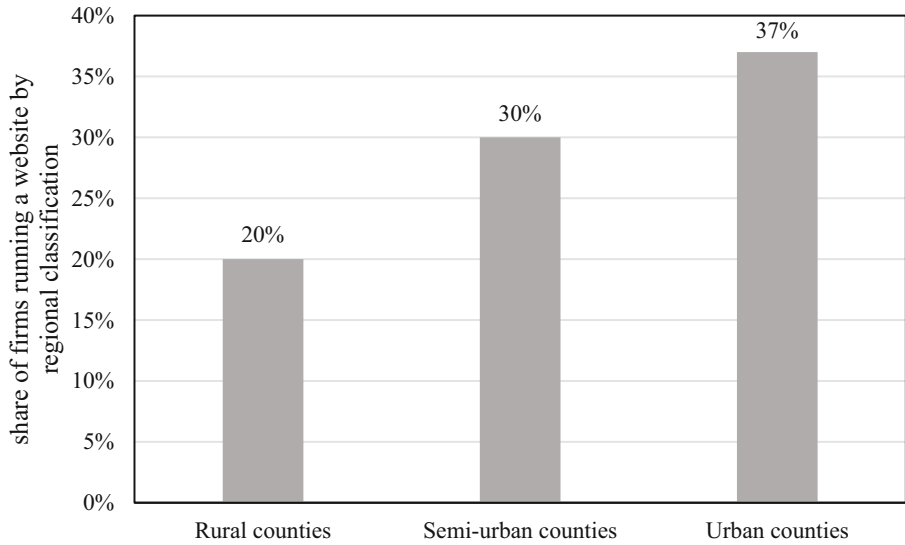


Fig. 1 Share of firms running websites classified by rural, semi-urban and urban counties

that the “last update” variable is common among older website formats and that it can only be obtained from a low share of websites. Therefore, the two proxies HTML5 and HTTPS can be used to comment on the up-to-dateness of websites. They can both be interpreted as follows: HTML5 has been introduced roughly four years prior to the data collection process and has become the respective standard in website design. If a firm has conducted a substantial technical update of their website to the current format in the past four years, it has most likely used the format. HTTPS shows that the owner of a website has invested into the communication security of the website in the past years and hence it can be regarded as another proxy of up-to-dateness. Table 1. provides an overview of the share of websites featuring the respective variable according to the classification of the respective county.

Interestingly, unlike in website prevalence itself, there are very small differences between urban, semi-urban and rural regions in terms of up-to-dateness of the websites. While urban firms tend to have slightly higher shares on the three indicators, the differences are weak at best. Thus, rural firms are less likely to run websites, although such firms tend to keep their website similarly up-to-date as urban firms.

Table 1 Website update indicators

Region type (no. of firms)	Update 2018	HTML5	HTTPS
Urban	0.247 (10,040)	0.573 (48,164)	0.296 (48,164)
Semi-urban	0.236 (5780)	0.558 (29,059)	0.287 (29,059)
Rural	0.239 (3721)	0.551 (17,561)	0.291 (17,561)

The number reported in brackets represents the number of firms for which data could be retrieved for the specific variable. Hence, the results have to be read as follows: we were able to retrieve 10,040 observations on the last update of a firm’s website in urban regions. 24.7% of these firms updated their website in 2018. Differences in the number of firms for different variables are due to data availability

Fig. 2 Website prevalence in German counties

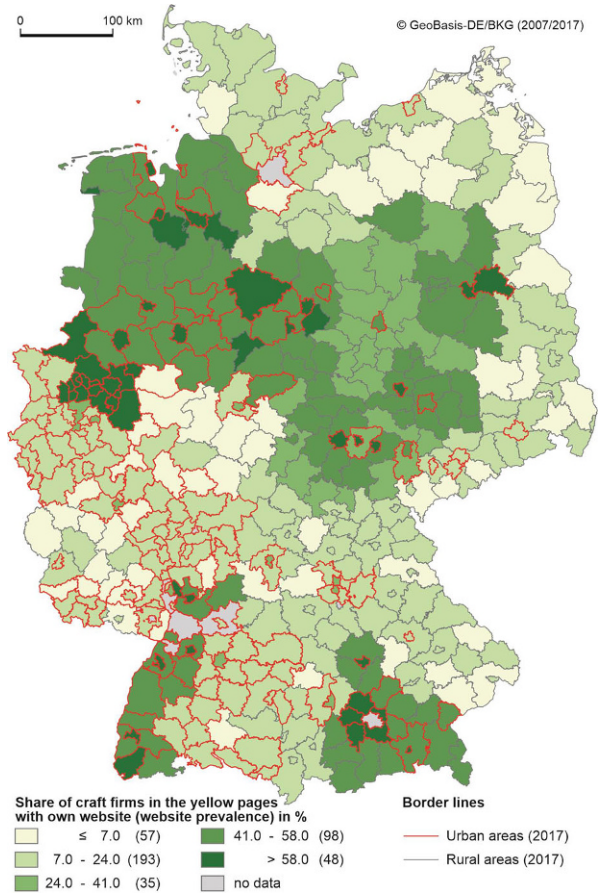


Table 2 Share of social media plugins in the websites

Region type (no. of firms)	Facebook	Instagram	Twitter
Urban (48,164)	0.309	0.059	0.179
Semi-urban (29,059)	0.297	0.045	0.171
Rural (17,561)	0.311	0.047	0.177

A similar descriptive analysis was conducted regarding social media prevalence, as shown in Table 2. Urban firms tend to have slightly higher shares of social media plugins, yet no substantial effects were discovered.

Thus, our core descriptive result is that firms in rural counties have a substantially lower propensity to run websites. However, those rural firms running websites show a similar disposition to keep their websites updated and connected to social media. Building upon these descriptive results, we run several regressions aiming to determine the driving factors behind the strong difference in website prevalence. As online annex Table 4 reveals, there are some differences in terms of the share of craft firms with and without websites across industries, the range covers values

between 18.3% and 51.6%. However, the standard deviation is not extraordinarily large and there are no correlations between the absolute number of firms for particular industries on the one hand and the respective share of website availability on the other.

4.2 Regression analysis

The linear probability regression model analyzes which regional variables increase the probability of a firm having a website. The results of the different regression specifications are presented in Table 3.

Turning to the interpretation of our results, we identify some regional determinants that are significantly associated with website prevalence. We find that population density has a significantly positive association with website prevalence for the overall sample, both including and excluding broadband access. In terms of effect sizes, an increase of one inhabitant per square kilometer increases the webpage prevalence by 0.01%, which can be considered a moderate effect. Population density can therefore be seen as one determinant for the second-level digital divide between rural and urban firms and the respective markets (Fig. 1). However, this does not warrant a simple interpretation of this result. The positive association with population density could either be due to the greater availability of the technology and network effects in cities or it could be interpreted as a rational business decision

Table 3 Regression output (short) to explain website prevalence

Independent variables	All variables, except broadband internet	All variables	Rural	Semi-urban	Urban
Broadband internet (over 50 Mbit/s in 2017)	–	0.0004*	0.0040***	–0.0017***	–0.00471***
Population density	0.0001***	0.0001***	–0.0001***	0.00003***	0.0001***
Share of employees with acad. Education	0.0074***	0.0071***	0.0064***	0.0135***	–0.0021*
Share of employees without formal educ	–0.0112***	–0.0116***	–0.0231***	–0.0003	–0.0562***
Sales revenue in craft sector	0.0022***	0.0022***	0.0030***	–0.001***	0.0024***
GDP per capita	–0.0030***	–0.0030***	0.0013***	–0.0013***	–0.0023***
Birth rate	0.0032***	0.0031***	–0.0004*	–0.0024***	0.0085***
Rate of reg. in-migration	0.0001***	0.0001*	–0.0005***	0.0007***	0.0010***
Increase/Decrease of inhabitants aged 65+	–0.0157***	–0.0158***	0.0110***	–0.0125***	–0.0484***
Employees in craft sector overall	–0.0028***	–0.0024***	0.0006	0.0128***	–0.0109***
Firm Size	0.0016***	0.0016***	–0.0001	0.0023***	0.0013**
Constant	0.161***	0.183***	–0.291***	0.286***	1.373***
Observations	296,228	296,228	73,948	93,376	128,901
R ²	0.07	0.07	0.10	0.06	0.11

The full regression output including industries is reported in Table 7 in the online appendix

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

such that rural firms prefer a more direct, non-digital interaction with their customers and vice versa.

Result 1: There is evidence in favor of H1, implying a digital divide between urban and rural firms.

Regarding the level of professional qualification among employees at the place of residence in the regional workforce, we find that higher shares of academically-trained employees are associated with higher website prevalence, and vice versa higher shares of employees without professional qualification are associated with lower website prevalence. Both effect sizes are fairly strong given that a one-percent increase of the respective share leads to a 0.7%/1% in(de)crease in webpage prevalence. This hints at the knowledge-based nature of the current technological change through digitalization, whereby employees, firms and regions with a stronger emphasis on higher education and knowledge transfer are more innovative, while those without are threatened by digitalization.

Result 2: Higher levels of professional qualification in the regional workforce are associated with a higher website prevalence, which provides evidence in favor of H2.

Connected to this result is the regions' development of the number of inhabitants of pension age (65+ in Germany). We find that counties with a high increase of inhabitants in pension age have overall lower levels of website prevalence, which can be interpreted as resulting from the different affinity to using digital technologies among older generations, in the private as well as the occupational context. All effect sizes are fairly strong, with a one-percent increase in the share of inhabitants 65+ leading to around one percent decrease in webpage prevalence. However, when looking at the differentiation between rural, semi-urban and urban counties, we find that a higher increase of older persons in rural regions is associated with higher website prevalence, meaning that the overall effect is driven by semi-urban and urban counties. Thus, while the overall result corresponds to previous results, the evidence is mixed when employing a regional differentiation.

Result 3: The development of the number of inhabitants aged 65+ is negatively associated with website prevalence, which provides evidence in favor of H3.

Further, it has been hypothesized that GDP per capita has a positive association with website prevalence, which can be explained from a capability as well as an affordability perspective regarding firms and customers. However, the results of our analysis provide evidence of a negative association of GDP per capita in the overall sample and particularly in urban regions with moderate effect sizes. Hence, even less-wealthy regions can have high levels of website prevalence. Nonetheless, in rural areas the effect is the reversed, as GDP per capita is positively associated with the probability of running a website.

Result 4: GDP per capita is positively associated with website prevalence in rural areas only. In urban areas, GDP per capita is negatively associated with website prevalence. Consequently, there is mixed evidence regarding H4.

Building upon these central results, our analysis addresses a number of additional variables and their association with website prevalence. The results suggest that other regional factors may partially explain the likelihood of having a website. First, we ask whether regional in-migration drives website prevalence, whereby this

variable proves to be significantly positive for the overall sample. This result seems plausible as the majority of in-migrants within advanced countries are young adults who migrate for tertiary education or in the early stages of their careers, as is the case for this article's study area. This group tends to have profound digital skills compared to older persons, particularly through life-long learning ("digital natives"). Regions with a higher share of in-migrants should thus have a stronger propensity to use digital technologies like websites for digital marketing. On the opposite side, counties with a high regional out-migration are likely to lose younger inhabitants with digital skills, thus making firm-level digitalization less relevant. Our results support this interpretation as the rate of out-migration in several robustness checks is negatively associated with website prevalence.

Second, we find an overall positive association of the development of the birth rate with website prevalence. This, again, points at the importance of a young population (the parents) for the regional level of digitalization. However, this result is driven by urban regions, where an increased birth rate is associated with higher website prevalence. We find a negative significant association in semi-urban and no significant association in rural areas.

Third, we controlled for the availability of high-speed internet. While the speed of 50Mbit/second is not needed to run a website, the availability of high-speed internet is likely to positively affect a firm's willingness to run a website by making investments in digitalization more rewarding overall for firms, which should also affect website prevalence. For customers, it might increase internet usage overall and thus foster a stronger demand for online information on firms, thus prompting them to provide such information. For the entire sample, there is a weak positive effect for broadband access. However, when looking at rural regions, there is a clear and significant positive effect, while it is significantly negative in semi-urban and urban counties. This result can be partly traced back to the respective broadband variable: there is only substantial variation in terms of households' access to internet with 50Mbit/second for rural counties. For semi-urban and urban counties in Germany, there are very few cases with an internet speed of less than 50Mbit/second. Thus, only the rural sub-sample should be interpreted. Accordingly, rural counties with good high-speed internet infrastructure are significantly more likely to have high shares of websites in craft firms than rural counties with poor internet infrastructure. It goes without saying that this result says little about causality; rather, there might be a causal link between the existence of larger and innovative firms in particular rural counties, whose political pressure has led to broadband access and whose existence leads to the internal migration of high-skilled employees who require non-relational information on craft firms.

Last but not least, the county's average firm size has a significantly positive effect on the probability of having a website. Hence, the larger a firm, the more likely it has a website. Now, if larger firms are more likely to be situated in urban areas, our results would only reflect the effects of firm size differences across region types on the website prevalence. We can rule out this argument by using a robustness test and presenting additional evidence. First of all, the effect size of the variable firm size is small and the fit increase is also comparatively low (see Table 17 in online appendix). Secondly, when looking at the region type specific effects, firm

size has only significantly positive effects in urban and sub-urban regions whereas firm size does not have a significant effect in the rural sub-sample. We also analyze the average firm sizes for each industry depending on the type of region (Fig. 3 in the online appendix). The average firm size in rural counties is 8.3 and 9.6 in urban counties. The difference is even smaller when using the median (6.4 in rural counties and 7.0 in urban counties). Only two industries (baker and building cleaner) show major differences in firm sizes between urban and rural counties. We therefore run another regression specification in which we drop the two industries with the largest differences in firm sizes between urban and rural counties (baker and building cleaner, Table 8 in online appendix). The results remain robust. Hence, taking this evidence together, the suspicion that larger firms in rural counties drive the results can be rejected.

One important limitation which should not be neglected is the goodness of fit of our model. The R^2 -value ranges between 0.06 in the semi-urban sub-sample, 0.07 in the overall sample and 0.10 in the rural or 0.11 in the urban sub-sample. This means that we can explain only around 7–11% of why a firm uses a website. This is admittedly low, however, it is due to the nature of the data we use which has many previously discussed advantages but also the disadvantage of lacking firm level data. The significance of the variable coefficients could be driven by the large number of our sample. We therefore calculated the fit increase of each variable which we report in Table 17 in the online appendix. One can see that each variable increases the goodness of fit. In particular, the firm's industry (0.0345) followed by population density (0.0247), the share of employees with an academic education (0.0182) as well as broadband internet availability (0.0114) explain whether a firm uses a website or not.

5 Discussion

The hypotheses on the regional determinants of the digital divide that we have addressed are well established in the literature, as are the indicators representing these determinants (the independent variables). The novelty in our study lies in the measurement of the dependent variable. Despite the profound differences to previously-used indicators, we largely confirm their results. However, due to the indicator used, our results include aspects of digitalization and the digital divide that have been missing in other studies. Our data represents rational business decisions of firms, i.e. running a website, and does not represent supply (internet access) or private consumption decisions (household usage of digital technologies, for example in Prieger 2013). Compared to other studies, these firms cover a wide range of different sectors and are relatively even distributed in space (dependent on population distribution) at a low spatial level (NUTS3 compared to NUTS1 for example in Billon et al. 2016 or NUTS2 in Ruiz-Rodriguez et al. 2017). Due to the basic nature of the digital technology in question, our analysis therefore suggests evidence of a—in contrast to previous studies—very fine-grained spatial divide in basic digitalization efforts in a wide range of industries. However, our study does not reflect usage skills, which are supposed to be growing in importance in the future. Castellacci et al. (2020) have

introduced a promising indicator but have not yet applied it to assess the second-level digital divide and its regional determinants. Moreover, as discussed in section 2.2, our indicator only represents certain aspects of digitalization. It would be interesting to see whether the indicator is highly correlated with other indicators and indices proxying the digitalization degree of regions. Even if we cannot be perfectly certain regarding this point, in replicating many previous results, our empirical results suggest that this is the case. We therefore think that this web-scraping approach on firms' websites is also promising in other research contexts, since it can be used to retrieve a variety of website characteristics (Kinne and Axenbeck 2020). For instance, in research on the digital divide it can be used to retrieve information on the prevalence of other digital technologies by applying text-mining procedures.

Discussing our results in further detail, the positive association of website prevalence with population density underlines the importance of population density for the urban-rural second-level digital divide. Firms in cities seem to perceive the advantages of and implement action towards online marketing more often than firms in rural regions. A higher adoption of (digital) technologies in urban rather than rural regions is often explained by the diffusion theory, stating that technologies are first developed and adopted in cities and then trickle down to more peripheral regions (see the seminal work of, e.g., Hägerstrand 1966 and Richardson 1973, or more recent attempts referring to internet domains (Sternberg and Krymalowski 2002) or internet technologies in general (Forman et al. 2005a)). If our results for website prevalence are interpreted from this perspective, we see a slow diffusion process even for this very basic, cheap and low-barrier technology. This gives rise to concerns, since more advanced digital technologies in Europe are primarily being developed and adopted in the economic centers (e.g. enabling technologies for Industry 4.0, Ciffolilli and Muscio 2018). These high-barrier technologies should trickle down at even slower speed or not diffuse at all and could thus exacerbate the urban-rural divide in technological competitiveness and—in the long term—living conditions.

Besides population density, other well-known regional determinants for the digital divide have been confirmed in our study, including qualification level and age, as characteristics of attractive labor market centers. The positive association of in-migration with website prevalence confirms this interpretation. Regions with a high attractiveness for the highly qualified therefore seem to be associated with higher digitalization levels. Since these characteristics (high qualification level, low average age, high in-migration) usually represent urban regions (see 2.1.1), these results can be interpreted as explanatory factors for the urban-rural digital divide. However, we find a negative association of GDP per capita with website prevalence. One possible explanation for this result might be that a number of German cities—in particular Berlin and in the Ruhr area—have a below-average GDP per capita despite being densely populated. Due to the high number of inhabitants, there are many craft firms reaching their urban customers with websites. Due to the relative weight of these cities in the overall sample, their high share of websites and low GDP might have caused the overall effect witnessed in the regression.

While the overall pattern of an urban-rural technological divide is not new, we suggest that the disruptive effects of digitalization with substantial changes to market structures change the picture. Our results are not limited to specific sectors that un-

dergo technological change, but rather they affect all sectors characterized by smaller firms alike, which means that their impact on regional economic structures is likely to be substantially more grave than previous technological change. Thus, regional disparities between rural and (semi-)urban regions might become substantially more pronounced if rural firms fail to be included into the technological dynamics of digitalization. Since market pressure from the respective demographics only facilitates digitalization in specific regions, we argue that there is a growing demand for mechanisms of technology transfer to firms in rural regions. This need is even enhanced by the fact that websites represent a very basic digitalization effort. It is plausible to assume that the divide in more advanced technologies is much stronger. However, if rural firms cannot even compete in terms of websites, it is hardly imaginable that they can catch up in advanced technologies without intervention.

Furthermore, it should be emphasized that the roll-out of broadband would hardly eliminate the second-level digital divide. In line with previous findings, our results show that other regional characteristics—notably professional qualification, population density, age structure and migration—play a major role in regional digitalization. Socio-demographically “weak” regions would still tend to benefit less from broadband internet than socio-demographically “strong” regions, as differences in these characteristics are persistent and can hardly be changed. Given that these regional characteristics seem to determine the adoption of websites, which are technologically fairly simple, an even stronger influence of these characteristics must be expected for more advanced technologies like big data analyses or cloud computing. In order to spread positive effects of digital technologies and prevent the rise in inter-regional disparities, policy measures must aim to diffuse more advanced technologies beyond the socio-demographically strong regions.

Different forms of digital training might be an adequate policy instrument for regions that lack some of the characteristics that “automatically” drive regional digitalization, especially since some studies have shown that (informal) life-long learning is even more important for the usage of digital technologies than professional qualification (Billon et al. 2016; Evangelista et al. 2014). Policy measures could thus promote the advancement of digital skills in socio-demographically “weak” regions. Digital skills should be more easily obtainable than professional qualifications, and in many regions they should be essential for technological diversification, potentially helping to develop new technological specialization patterns (Castellacci et al. 2020).

6 Conclusion

In this paper, we have introduced a novel indicator for measuring the regional digitalization degree, which comprises data on website prevalence in craft firms in Germany and is constructed using a web-scraping algorithm. We use this indicator to assess the regional determinants of the second-level digital divide, i.e. spatial differences in the usage of digital technologies. While we argue that this indicator differs from previously-used indicators for assessing the second-level digital divide, our results largely follow previous results. We therefore argue that the web-scraping

procedure is a promising tool for future studies in research on digitalization from a spatial perspective.

Our results show a positive association between population density and website prevalence, which hints at the conclusion that population density is an important determinant for the second-level digital divide. Equally in line with previous research, we find a positive association of formal education levels with website prevalence, illustrating the knowledge-based nature of the current technological change. A young population and high in-migration rates are also associated with a higher digitalization degree. Since these website-promoting characteristics are usually found in cities, we find a profound urban-rural second-level digital divide that is not only rooted in population density but also in other socio-demographic regional characteristics. In contrast to recent studies, we find a negative association of GDP per capita with regional website prevalence and discuss interpretations. This result is driven by urban regions, while the opposite is true for rural regions. Our descriptive results show no substantial urban-rural differences in website characteristics (up-to-datedness, social media references) warranting no detailed evaluation of these characteristics. Further results show no overall association of broadband availability with website prevalence, but a positive association in rural regions.

We find promising avenues for future research in applying the proposed web-scraping procedure. Future studies could transfer our empirical approach to other countries or spatial scales and integrate further potentially-relevant regional characteristics that were unavailable in our sample but have been used in previous research, such as regional industry and sector structures and regional institutional factors (Billon et al. 2016; Dengler et al. 2018; Moriset et al. 2012).

Supplementary Information The online version of this article (<https://doi.org/10.1007/s10037-022-00170-5>) contains supplementary material, which is available to authorized users.

Funding The results of this study are based on the project “Digital Transformation of SMEs in the Crafts Sector of Southern Lower Saxony” by the Institute for Small Business Economics. The project was funded by the European Regional Development Fund as well as by the state of Lower Saxony (grant number ZW 6-85017668).

In addition, we acknowledge the funding from the Lower Saxony Ministry of Science and Culture within the Lower Saxony “Vorab” of the Volkswagen Foundation (grant number ZN3492) and the support by the Center for Digital Innovations (ZDIN).

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Conflict of interest A. Thonipara, R. Sternberg, T. Proeger and L. Haefner declare that they have no competing interests.

References

- Akerman A, Gaarder I, Mogstad M (2015) The skill complementarity of broadband internet. *Q J Econ* 130:1781–1824. <https://doi.org/10.1093/qje/qjv028>
- Atasoy H (2013) The effects of broadband Internet expansion on labor market outcomes. *Ind Labor Relat Rev* 66(2):315–345
- Berger T, Frey CB (2016a) Structural transformation in the OECD: digitalisation, deindustrialisation and the future of work. OECD Social, Employment and Migration Working Papers 193. OECD Publishing, Paris
- Berger T, Frey CB (2016b) Did the Computer Revolution shift the fortunes of U.S. cities? Technology shocks and the geography of new jobs. *Reg Sci Urban Econ* 57:38–45. <https://doi.org/10.1016/j.regsciurbeco.2015.11.003>
- Bertschek I, Briglauer W, Hüschelrath K, Kauf B, Niebel T (2015) The economic impacts of broadband Internet: a survey. *Rev Netw Econ* 14(4):201–227
- Billon M, Lera-Lopez F, Marco R (2016) ICT use by households and firms in the EU: links and determinants from a multivariate perspective. *Rev World Econ* 152:629–654. <https://doi.org/10.1007/s10290-016-0259-8>
- Blank G, Groselj D (2014) Dimensions of internet use. *Inf Commun Soc* 17(4):417–435. <https://doi.org/10.1080/1369118X.2014.889189>
- Blank G, Graham M, Calvino C (2018) Local geographies of digital inequality. *Soc Sci Comput Rev* 36(1):82–102. <https://doi.org/10.1177/0894439317693332>
- BMI (Bundesministerium des Innern, für Bau und Heimat—Federal Ministry of the Interior, Building and Community; Ed.) (2019) Unser Plan für Deutschland – Gleichwertige Lebensverhältnisse überall. <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/heimat-integration/schlussfolgerungen-kom-gl.html>. Accessed 6 June 2020
- Briglauer W, Dürr NS, Falck O, Hüschelrath K (2019a) Does state aid for broadband deployment in rural areas close the digital and economic divide? *Inf Econ Policy* 46:68–85. <https://doi.org/10.1016/j.infoecopol.2019.01.001>
- Briglauer W, Dürr NS, Gugler K (2019b) A retrospective study on the regional benefits and Spillover effects of high speed broadband networks: evidence from German counties. ZEW—Centre for European Economic Research Discussion Paper No. 19–026. <https://doi.org/10.2139/ssrn.3423864>
- Büchi M, Just N, Latzer M (2016) Modeling the second-level digital divide: a five-country study of social differences in Internet use. *New Media Soc* 18(11):2703–2722. <https://doi.org/10.1177/1461444815604154>
- Cairncross F (1997) The death of distance. How the communications revolution Will change our lives. Harvard Business School Press, Boston
- Cairncross F (2018) The death of distance—has distance died? *RSAI Newsl New Ser* 17:8–9
- Camagni R, Capello R (2005) ICTs and territorial competitiveness in the era of internet. *Ann Reg Sci* 39:421–438. <https://doi.org/10.1007/s00168-005-0244-y>
- Castellacci F, Consoli D, Santoalha A (2020) The role of e-skills in technological diversification in European regions. *Reg Stud* 54(8):1123–1135. <https://doi.org/10.1080/00343404.2019.1681585>
- Celbis MG, de Crombrughe D (2018) Internet infrastructure and regional convergence: evidence from Turkey. *Pap Reg Sci* 97(2):387–409. <https://doi.org/10.1111/pirs.12244>
- Ciffolilli A, Muscio A (2018) Industry 4.0: national and regional comparative advantages in key enabling technologies. *Eur Plan Stud* 26(12):2323–2343. <https://doi.org/10.1080/09654313.2018.1529145>
- Dengler K, Matthes B, Wydra-Sommaggio G (2018) Regionale Branchen- und Berufsstrukturen prägen die Substituierbarkeitspotenziale. *IAB-Kurzberichte* 22/2018
- Dijkstra L, Poelman H, Rodríguez-Pose A (2020) The geography of EU discontent. *Reg Stud* 54:737–753. <https://doi.org/10.1080/00343404.2019.1654603>
- European Commission (2020) Digital Scoreboard—Data & Indicators. <https://digital-agenda-data.eu/>. Accessed 17 Aug 2020
- Evangelista R, Guerrieri P, Meliciani V (2014) The economic impact of digital technologies in Europe. *Econ Innov New Technol* 23(8):802–824. <https://doi.org/10.1080/10438599.2014.918438>
- Fabritz N (2013) The impact of broadband on economic activity in rural areas: evidence from German municipalities. Ifo Working Paper No. 166.
- Forman C, Goldfarb A, Greenstein S (2005a) Geographic location and the diffusion of Internet technology. *Electron Commer Res Appl* 4(1):1–13. <https://doi.org/10.1016/j.elerap.2004.10.005>

- Forman C, Goldfarb A, Greenstein S (2005b) How did location affect adoption of the commercial Internet? Global village vs. urban leadership. *J Urban Econ* 58(3):389–420. <https://doi.org/10.1016/j.jue.2005.05.004>
- Galloway L, Sanders J, Deakins D (2011) Rural small firms' use of the internet: from global to local. *J Rural Stud* 27:254–262. <https://doi.org/10.1016/j.jrurstud.2011.05.005>
- Graham M (2019) Digital economies at global margins. MIT Press, Cambridge
- Grappone J, Couzin G (2011) Search engine optimization (SEO): an hour a day. John Wiley & Sons
- Grimes S (2003) The digital economy challenge facing peripheral rural areas. *Prog Hum Geogr* 27(2):174–193. <https://doi.org/10.1191/0309132503ph421oa>
- Grimes S (2005) How well are Europe's rural businesses connected to the digital economy? *Eur Plan Stud* 13(7):1063–1081. <https://doi.org/10.1080/09654310500242105>
- Haefner L, Sternberg R (2020) Spatial implications of digitization: State of the field and research agenda. *Geogr Compass*. <https://doi.org/10.1111/gec3.12544>
- Hägerstrand T (1966) Aspects of the spatial structure of social communication and the diffusion of information. *Pap Reg Sci Ass* 16:27–42
- Hasbi M (2020) Impact of very high-speed broadband on company creation and entrepreneurship: empirical evidence. *Telecomm Policy*. <https://doi.org/10.1016/j.telpol.2019.101873>
- Ivus O, Boland M (2015) The employment and wage impact of broadband deployment in Canada. *Can J Econ* 48(5):1803–1830. <https://doi.org/10.1111/caje.12180>
- Jung J, López-Bazo E (2020) On the regional impact of broadband on productivity: the case of Brazil. *Telecomm Policy*. <https://doi.org/10.1016/j.telpol.2019.05.002>
- Katz RL, Koutroumpis P (2013) Measuring digitization: A growth and welfare multiplier. *Technovation* 33(10–11):314–319. <https://doi.org/10.1016/j.technovation.2013.06.004>
- Kinne J, Axenbeck J (2020) Web mining for innovation ecosystem mapping: a framework and a large-scale pilot study. *Scientometrics* 125:2011–2041
- Krone M, Dannenberg P (2018) Analysing the effects of information and communication technologies (ICT) on the integration of East African farmers in a value chain context. *Z Wirtschgeogr* 62(1):65–81. <https://doi.org/10.1515/zfw-2017-0029>
- LaRose R, Gregg JL, Strover S, Straubhaar J, Carpenter S (2007) Closing the rural broadband gap: promoting adoption of the internet in rural America. *Telecomm Policy* 31(6–7):359–373. <https://doi.org/10.1016/j.telpol.2007.04.004>
- Li R, Shiu A (2012) Internet diffusion in China: a dynamic panel data analysis. *Telecomm Policy* 36(10–11):872–887. <https://doi.org/10.1016/j.telpol.2012.06.004>
- López-Bazo E, Motellón E (2018) Innovation, heterogeneous firms and the region: evidence from Spain. *Reg Stud* 52(5):673–687. <https://doi.org/10.1080/00343404.2017.1331296>
- Mack E, Faggiani A (2013) Productivity and broadband: the human factor. *Int Reg Sci Rev* 36(3):392–423
- Mack E, Anselin L, Grubestic TH (2011) The importance of broadband provision to knowledge intensive firm location. *Reg Sci Policy Pract* 3(1):17–35. <https://doi.org/10.1111/j.1757-7802.2011.01026.x>
- Malecki EJ (2002) The Economic Geography of the Internet's Infrastructure. *Econ Geogr* 78(4):399–424
- Malecki EJ (2003) Digital development in rural areas: potentials and pitfalls. *J Rural Stud* 19:201–214
- McCoy D, Lyons S, Morgenroth E, Palcic D, Allen L (2018) The impact of broadband and other infrastructure on the location of new business establishments. *J Reg Sci* 58(3):509–534. <https://doi.org/10.1111/jors.12376>
- Mills BF, Whitacre BE (2003) Understanding the non-metropolitan-metropolitan digital divide. *Growth Change* 34(2):219–243. <https://doi.org/10.1111/1468-2257.00215>
- Moriset B, Brette O, Chaireire O, Grasland L, Richaud A, Thivaut E (2012) The “death of distance” fifteen years on. Information technology and knowledge-based service firms in rural areas. HAL archives-ouvertes, halshs-00728338
- Papagiannidis S, Gebka B, Gertner D, Stahl F (2015) Diffusion of web technologies and practices: a longitudinal study. *Technol Forecast Soc Change* 96:308–321. <https://doi.org/10.1016/j.techfore.2015.04.011>
- Prieger JE (2003) The supply side of the digital divide: is there equal availability in the broadband Internet access market? *Econ Inq* 41(2):346–363. <https://doi.org/10.1093/ei/cbg013>
- Prieger JE (2013) The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecomm Policy* 37(6–7):483–502. <https://doi.org/10.2139/ssrn.2258112>
- Proeger T, Runst P (2020) Digitization and knowledge spillover effectiveness—Evidence from the “German Mittelstand”. *J Know Econ* 11(4):1509–1528
- Richardson HW (1973) Regional growth theory London

- Riddlesden D, Singleton AD (2014) Broadband speed equity: a new digital divide? *Appl Geogr* 52:25–33. <https://doi.org/10.1016/j.apgeog.2014.04.008>
- Rodríguez-Pose A (2018) The revenge of the places that don't matter (and what to do about it). *Camb J Reg Econ Soc* 11:189–209. <https://doi.org/10.1093/cjres/rsx024>
- Rodríguez-Pose A, Crescenzi R (2008) Mountains in a flat world: why proximity still matters for the location of economic activity. *Camb J Reg Econ Soc* 1(3):371–338
- Ruiz-Rodríguez F, Lucendo-Monedero AL, González-Relaño (2017) Measurement and characterisation of the Digital Divide of Spanish regions at enterprise level. A comparative analysis with the European context. *Telecomm Policy* 42:187–211. <https://doi.org/10.1016/j.telpol.2017.11.007>
- Runst P, Thomä J (2020) Does occupational deregulation affect in-company vocational training?—Evidence from the 2004 Reform of the German Trade and Crafts Code. *J Econ Stat* 240(1):51–88
- Salemink K, Strijker D, Bosworth G (2017) Rural development in the digital age: a systematic literature review on unequal ICT availability, adoption, and use in rural areas. *J Rural Stud* 54:360–371. <https://doi.org/10.1016/j.jrurstud.2015.09.001>
- Scheerder A, von Deursen A, van Dijk J (2017) Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telemat Inf* 34(8):1607–1624. <https://doi.org/10.1016/j.tele.2017.07.007>
- Schleife K (2010) What really matters: Regional versus individual determinants of the digital divide in Germany. *Res Policy* 39:173–185. <https://doi.org/10.1016/j.respol.2009.11.003>
- Sternberg R, Krymalowski M (2002) Internet domains and the innovativeness of cities/regions—Evidence from Germany and Munich. *Eur Plan St* 10(2):251–274
- The Federal Government of Germany (2018) Shaping Digitalization. Implementation strategy of the Federal Government. The Federal Government of Germany, Berlin
- Townsend L, Sathiaselalan A, Fairhurst G, Wallace C (2013) Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. *Local Econ* 28:580–595. <https://doi.org/10.1177/0269094213496974>
- Townsend L, Wallace C, Fairhurst G, Anderson A (2017) Broadband and the creative industries in rural Scotland. *J Rural Stud* 54:451–458. <https://doi.org/10.1016/j.jrurstud.2016.09.001>
- Tranos E (2012) The causal effect of the Internet infrastructure on the economic development of European city regions. *Spat Econ Anal* 7(3):319–337. <https://doi.org/10.1080/17421772.2012.694140>
- Tranos E (2016) The internet: its geography, growth and the creation of (digital) social capital. In: Shear-mur R, Carrincazeaux C, Doloreux D (eds) *Handbook on the geographies of innovation*. Edward Elgar, Cheltenham, pp 356–369
- Tranos E, Stich C (2020) Individual internet usage and the availability of online content of local interest: a multilevel approach. *Comput Environ Urban Syst*. <https://doi.org/10.1016/j.compenvurbsys.2019.101371>
- Vicente MR, López AJ (2011) Assessing the regional digital divide across the European Union-27. *Telecomm Policy* 35(3):220–237. <https://doi.org/10.1016/j.telpol.2010.12.013>
- Wernick C, Bender CM (2017) The role of municipalities for broadband deployment in rural areas in Germany: an economic perspective. *Digiworld Econ J 1st Quart* 105:91–110
- What works Centre for local economic growth (WWG) (2015) Evidence review 6—broadband
- Whitacre B, Gallardo R, Strover S (2014a) Does rural broadband impact jobs and income? Evidence from spatial and first-differenced regressions. *Ann Reg Sci* 53:649–670. <https://doi.org/10.1007/s00168-014-0637-x>
- Whitacre B, Gallardo R, Strover S (2014b) Broadband's contribution to economic growth in rural areas: moving towards a causal relationship. *Telecomm Policy* 38:1011–1023. <https://doi.org/10.1016/j.telpol.2014.05.005>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.