Received: 27 October 2021 | Revised: 26 July 2022 | Accepted: 22 September 2022

DOI: 10.1111/jors.12624

RESEARCH ARTICLE

JOURNAL OF REGIONAL SCIENCE WILEY

Digital gravity? Firm birth and relocation patterns of young digital firms in Germany

Vanessa Hellwig

Department of Economics, TU Dortmund University, Dortmund, Germany

Correspondence
Vanessa Hellwig, Department of Economics, TU Dortmund University, Vogelpothsweg 87, D-44227 Dortmund, Germany.
Email: Vanessa.hellwig@tu-dortmund.de

Abstract

This paper analyses the spatial patterns of young (<10 years) digital firms in Germany between 2008 and 2017 on county level. Determinants of firm birth locations as well as relocations are considered jointly to understand differences in location choices within firms' life cycles. I match commercial register data of 107,321 firms with county-level administrative data to capture local characteristics. Using an OLS model with fixed effects, I find that the local knowledge base—that is, universities, research institutes, and colocated incumbents—are significant key determinants of digital firm birth when controlling for a host of local characteristics. My results indicate that for five firms per 1000 inhabitants, there is around one firm birth. Second, using a fixed effects gravity model for the analysis of relocations, I find that the most dominant explanatory factor for firm relocation across specifications is distance, that is, relocation costs. Relocation flows are more than twice as high to neighboring counties relative to other locations which shows that digital firms are not as footloose as their business model may suggest. Jointly, my results reflect economic activity's regional persistence, particularly for new firms. My paper provides evidence for policies targeting homogenous digital clusters based on strong colocation and that digital economic activity is not shifted over long distances, but the regional entrepreneurship capital is crucial for local growth.

KEYWORDS

digital, firm relocation, gravity model, ICT, location choices, mobility, start-ups

1 | INTRODUCTION

Scholars observe a recent trend of young, high-tech, and knowledge-intensive firms to locate in clusters in central districts and neighborhoods in large cities (see e.g., Duvivier et al., 2018; Foord, 2013), seemingly challenging the Silicon Valleys of the Western World. Part of this trend is not only driven by newly found startups but also by firms relocating to cities. For example, Twitter, Uber, and Airbnb have chosen to set up their new headquarters in downtown San Francisco (Duvivier et al., 2018), while Oracle, HP, and Tesla chose to relocate to "Silicon Hills" in Austin, TX. As the big players in the technology industry bring employment, taxes, and a host of other externalities for the region, their relocations exemplify just how important location decisions of both new and existing firms are for regional economic development (Audretsch et al., 2006; Fritsch & Mueller, 2008).

Thus, there are a range of public efforts often using strong subsidisation to support and foster digital economic activity by means of attracting new businesses, incubating startups, developing and/or supporting technology hubs and networks. To this end, understanding the factors which determine locational choices for start-ups and incumbents' relocations are crucial for efficient policy making (Y. Lee, 2008). Initial location choices and the role of regional factors for entrepreneurship are well studied and the literature shows that start-up hotspots are highly persistent over time (Fossen & Martin, 2018). At the same time, young knowledge-intensive firms become ever more likely to change locations (Esteve-Pérez et al., 2018) to grow and gain competitiveness (Guzman, 2019; Stam, 2007).

This paper uses commercial register data on 107,321 young digital firms in German counties (NUTS3) between 2008 and 2017 to analyze firm birth and relocation patterns in the high-tech industry. The main objective is to compare regional preferences in firm births and relocation patterns to assess whether digital start-ups and relocating companies favor the same locational characteristics. If they do not share the same location requirements, more tailored policies towards those two avenues of local economic growth are necessary.

First, I analyze digital firm birth intensity using a linear regression model ordinary least squares (OLS) with county and time fixed-effects and a host of regional control variables. Results show that accessibility of industry-specific knowledge as proxied by the colocation of digital firms is highly conducive to startup activity. For each firm per 1000 inhabitants, there are 0.17 startups. In addition and in line with the literature, universities are a significant factor for new digital businesses. This result indicates that infant digital firms rely on industry as well as institutional know-how and that locational costs outweigh such locally-bound, tacit benefits.

Second, I test spatial patterns of business relocations building on the theory and operationalization of the (fixed-effects) gravity model. Counties with higher agglomeration benefits such as a specialized high-tech labor markets and the potential for IT-specific knowledge spillovers are expected to attract more relocating firms. Additionally, most relocations should occur between geographically proximate or contiguous counties where moving costs are low while access to locally bound factors such as local customers, suppliers, or networks remains relatively low-cost.

The results of this empirical exercise show that significantly more digital businesses relocate to counties with a high density of digital firms. Therefore, digital firms show a strong preference to cluster. I find flows between neighboring counties are more than twice as big as other flows and relocation flows decay with distance. This is consistent with the fact that moving costs as captured by distance play a crucial, deterring role in relocation decisions. These findings are at odds with the common perception that digital businesses are relatively footloose (Esteve-Pérez et al., 2018; Weterings & Koben, 2014). On the contrary, my findings indicate that they tend to stay in their regions of origin. This finding, however, is in line with Knoben's result (2011) that firms with high dependency on outside resources and strong networks do not relocate over long distances. These results are in line with the literature on industries' regional persistence (e.g., Fritsch & Wyrwich, 2014). Moreover, the joint investigation of firm birth, as well as incumbents' relocation patterns, reveals that policies targeting digital firm birth also spill over into neighboring counties in the medium to long run next to the expected initial local benefits. Therefore, an intraregional cooperation strategy to foster firm birth where counties administrations pool their resources would be a promising approach.

This paper contributes to the literature in three ways. First, very few papers allow direct comparison of firm birth locations and relocation (Holl, 2004; I. H. I. Lee, 2020; Manjón-Antolín & Arauzo-Carod, 2011; Stam, 2007) while none—to the best of my knowledge—focuses on the relocation of "infant" firms. Nonetheless, a joint analysis of new and young digital firms is highly informative to understand the development of regional patterns of entrepreneurship in the long run. Moreover, it is new as well as young digital businesses which drive industry growth next to the long-established big players in the industry. Especially young, growing firms are a potential asset for a county's economic development. Thus, understanding this particular neck of the digital industry and its spatial preferences matters greatly in the context of political interventions aimed at alleviating regional disparities and smoothening structural economic transitions by means of financially supporting the digital industry.

Second, many of the existing papers on relocation focus on manufacturing (e.g., Conroy et al., 2016; Holl, 2004; Yi 2018) or compare different sets of industries along the business models' knowledge-intensities (e.g., Kronenberg, 2013; Nguyen et al., 2013; Weterings & Knoben, 2013). Yet, digital business models similarly to other service sector industries have a different cost-structure than manufacturing—labor intensive while sunk costs are low—thus requiring separate consideration. Moreover, digital companies provide broadly applicable technology which affords productivity gains to almost all other sectors and therefore differ from other industries in terms of market reach and locational choices.[1]

Third, I make a methodological contribution to the literature by applying a state-of-the-art gravity model framework to the analysis of firm mobility. This is in contrast to the majority of regional studies that rely on discrete choice models which neither include relocation costs nor mobility's geographic dimension. Moreover, these specifications are rather ad-hoc without theoretical priors on the key decision factors and mechanisms driving the results. In the literature on FDI, however, the gravity model has been (theoretically) established (e.g., Portes & Rey, 2005). Since FDIs are essentially cross-border firm or subsidiary relocations, there is reason to believe that the gravity model is appropriate for modeling firm relocation flows.

As the workhorse model in the migration and trade literature, the gravity model captures the spatial dimension by using the physical distance between origin and destination implicitly capturing costs which in turn are increasing in distance. For example, information costs about a new business site are harder to assess from far away. The key advantage of using the gravity model with a full set of origin and destination fixed effects for the analysis of firm mobility is that it allows for modeling one of the key unobservables, namely implicit and explicit relocation costs, while also controlling for location-specific characteristics and common time trends affecting all locations. The inclusion of these fixed effects is important for identification as a lot of the variation in the data can be explained by these location-specific characteristics. Thus, the remaining significant determinants in the model can be interpreted beyond any time and location-specific confounding factors.

The remainder of this paper proceeds as follows. Section 2 contextualizes the paper in light of the existing literature while Section 3 embeds it in a theoretical framework. Section 4 describes the data in detail and Section 5

[1]While there are many studies on entrepreneurship (e.g., Fossen & Martin, 2018; Fritsch & Mueller, 2008) in Germany, there are hardly any studies about firm relocation in Germany.

the methodology. In Section 6, I present the findings: first my findings on startup patterns, then the results for the gravity models of ICT firms' relocations in Germany. The last section concludes, while also highlighting the key takeaways for policy makers and regional planners.

2 | LITERATURE ON REGIONAL DETERMINANTS OF FIRM BIRTH AND RELOCATION

2.1 | Firm birth

Factors which are conducive to entrepreneurship as measured by high local startup rates have been well studied in the literature (e.g., Buczkowska & de Lapparent, 2014; Fritsch & Storey, 2014; Glaeser et al., 2010, 2015). Bade and Nerlinger (2000) in a study of West-German new technology-based firms (NTBF), find the highest startup rates relative to the labor force in close proximity to core cities, while core cities have the most startups in absolute terms. van Oort and Atzema (2004) find ICT firms to colocate in areas with dense economic activities.

Moreover, Audretsch et al. (2012) find that local employees' propensity to start a business is highest in urban agglomerations and their periphery. In this context, Pijnenburg and Kholodilin (2014) link entrepreneurship capital and knowledge-based startup rates. They find that knowledge spills over from its source to the startup also across NUTS-3 borders. Since urban agglomerations innately offer a diverse economic environment they provide key factors to thrive for digital businesses. I thus expect high startup rates in core cities and their surroundings.

2.2 | Relocation

The focus in this paper lies on location patterns of new digital businesses and firms younger than 10 years. The probability of relocation is very high within the first 10 years of a firm's life: to reduce risk of failure, firms have to innovate and reconsider products, activities, and eventually also their location (Esteve-Pérez et al., 2018; Rossi & Dej, 2019).

Apart from individual firm characteristics, a firm's decision to relocate can be motivated by external location-specific characteristics that “push” the firm away from its current location (van Dijk & Pellenbarg, 2000; van Wissen, 2000) such as steeply rising real estate prices. Relocating firms can be “pulled” into regions offering more suitable location characteristics (Holl, 2004; Kronenberg, 2013).

Agglomeration benefits are strongly identified as pull factors since firms across all sectors—in particular services, are drawn to densely populated municipalities to benefit from higher local demand, stronger, better-educated workforces, and a wider supply of local public amenities (Bodenmann & Axhausen, 2012; Kronenberg, 2013; Nguyen et al., 2013; Rossi & Dej, 2019; Weterings & Knoben, 2013).

Typically, service sector firms want to benefit from the locally bound knowledge and labor pool and thus are attracted to dense, high-quality-of-life municipalities in spite of their high sector-specific wages (Kronenberg, 2013). That is, firms do not necessarily adopt a pure cost minimization strategy but choose locations where they can be certain that high-skilled workers and necessary amenities are available (Rossi & Dej, 2019). Stam (2007) argues that relationships with social networks are especially important in the early stages of a firm's life while cost considerations become more important later, while Knoben (2011) finds firms being dependent on outside resources tend to move short distances. Presumably, knowledge-intensive digital firms do not need as much space as manufacturing firms with constant access to knowledge being more important than a low-cost location.

3 | THEORETICAL FRAMEWORK

3.1 | Agglomeration effects and firms' location decisions

Firms derive advantages from agglomeration externalities especially in cities. The main benefits are the pooled labor market, access to specialized suppliers, and benefits from knowledge spillovers (Armington & Acs, 2002; Krugman, 1991). These agglomeration effects can be divided into localization and urbanization economies which are considered in turn.

Urbanization effects refer to the benefits of diversity and density of amenities such as public infrastructure which cities typically offer (Jacobs, 1969). Next to the size of the labor market, it is in particular the urban density of universities, research institutes, and other knowledge- and research-related activities facilitating knowledge spillovers between firms. Novel knowledge and innovation are closely linked to entrepreneurship through commercialization of knowledge into new firms (Acs et al., 2009). Thus, a vital regional knowledge base is more likely to be bigger in cities rather than regional agglomerations rendering cities particularly attractive locations for young firms.

Localization describes colocation of similar firms in close proximity (Porter, 1990) with firms mainly benefiting from employment advantages such as specialized employees and a lower probability of labor shortages (Krugman, 1991). A particularly important benefit for software firms (Trippl et al., 2009) are knowledge spillovers accruing from interfirm cooperation as well as fluctuation of employees. Moreover, colocation of related industries fosters entrepreneurship by lowering costs of starting a business for individuals and enabling better access to a more diverse range of inputs and complementary goods (Glaeser & Kerr, 2009).

Novel knowledge and innovations are key for profits and growth and thus are particularly relevant for location analysis. General innovations are typically developed and harbored in universities and research institutes while incumbent firms typically hold an advantage in new marketable products. Now, if the expected value of a certain piece of knowledge is higher for an individual than for the decision maker in the institution or the firm this individual will start a new business if costs are low (Pijnenburg & Kholodilin, 2014). There is an interdependence between existing knowledge institutions and industry players which is conducive to entrepreneurship within the spatial reach of such knowledge spillovers (Bade & Nerlinger, 2000; Fritsch & Aamoucke, 2013; Trippl et al., 2009). Furthermore, university graduates are a source of qualified labor supply to local firms. This can be advantageous for relocated firms if the labor market at their origin is insufficient (Armington & Acs, 2002).

3.2 | Firm birth and mobility

The fundamental difference between location and relocation theory is that relocations explicitly substitute one location for another, while newly established firms are not constrained by previous location decisions. In general, a firm moves from its current location if the location is no longer inside the spatial margins of profitability (Brouwer et al., 2004; Ozmen-Ertekin et al., 2007). Relocation decisions are likely to be explained by the differences between origin and destination in terms of profitability as well as relocation costs.

This paper moves away from individual firm movements and relies on an aggregate approach distinguishing between interregional and intraregional migrations. Businesses intra-regional moves amount to industrial suburbanization around larger urban agglomerations (van Dijk & Pellenbarg, 2017). This is referred to as the incubator hypothesis (Leone & Struyk, 1976) which postulates that manufacturing firms are born in central urban areas and they out-migrate to urban peripheries in their growth phase to find expansion space at a location that is easily accessible for clients and suppliers (van Dijk & Pellenbarg, 2017).

Entrepreneurs tend to disproportionally take their hometown as a natural firm birth location and thus exhibit a strong home bias (Figueiredo et al., 2002; Michelacci & Silva, 2007). This local entrepreneurship capital is an element of the region's endogenous economic potential (Fritsch & Storey, 2014; Stam, 2007). Conceivably, home-biased entrepreneurs located in the periphery revise their initial location decision once they have proven viable and move to nearby cities to benefit from agglomeration advantages.

Interregional moves involve industrial decentralization from economic core areas to peripheral areas. Within this type of movement, firms move to areas with lower land and/or labor costs (van Dijk & Pellenbarg, 2017). The increase in land prices induced by agglomeration of economic activities (Combes et al., 2012) could be a more significant incentive for older, established firms to opt-out of industrial agglomerates as the cost–benefit trade-off becomes unfavorable for the urban location.

Moves between core cities often reflect firms moving from diversified (urbanization economies) into specialized cities (localization economies) (Duranton & Puga, 2001). Businesses start out in diversified cities (urbanization economics) until they find an ideal business process and ultimately relocate into a specialized city (localization economics) when switching to mass production. Systematic comparisons of location patterns of start-ups and relocating firms such as by Holl (2004) and Manjón-Antolín and Arauzo-Carod (2011), for example, indicate that startup activities are highly associated with industrial diversity while firm relocations are not.

3.3 | Theoretical predictions and hypothesis

The conceptual framework of this paper draws on the theoretical and empirical findings on spatial patterns of firm birth laid out above. Agglomeration benefits stand out as a crucial location factor for firm birth and entrepreneurial activity. I thus expect *young digital firms to display a strong preference towards cities (Hypothesis 1)*.

With regard to localization and urbanization, I use service—and industrial ratios to capture the general local economy and its specialization. With increasing colocation of similar firms (digital firms per 1000 inhabitants), theoretical agglomeration mechanisms predict increasing firm birth rates. As local (rent-) prices display a firm's willingness to pay for agglomeration benefits, they are included in the empirical analysis. As a knowledge-intensive industry, firms in the digital economy are dependent on several knowledge providers such as universities and research institutes and their possible labor markets. *Universities and research institutes are expected to have a positive effect on firm births (Hypothesis 2)*.

Applying the gravity model to digital firms' relocation patterns, I expect that agglomeration benefits vary considerably with firm age as well as other firm characteristics and thus individual firms may find agglomerates attractive according to their individual preferences as captured by gravitational force of the associated agglomeration benefits. One such example could be a maturing firm seeking to innovate their products or to broker into a new market. Therefore, I hypothesize that *agglomerated areas receive higher in-flows (Hypothesis 3)*.

Digital firms do not require a lot of physical space neither when starting out nor when growing since scaling up of digital products has different space requirements than manufacturing—the most common sector of investigation in the relocation literature.[2] Therefore, I do not expect firms to select out into peripheral areas due to an unfavorable trade-off between costs and access to knowledge and market, but rather to remain within the agglomeration effects' spatial margins. Thus, consistent with the assumptions of the gravity model, *relocation flows are expected to decrease with distance (Hypothesis 4)*.

[2]Moreover, it is conceivable that digital firms compete through the price channel rather based on scale as is typically the case for service industries (see e.g., Saarenketo et al., 2008).

4 | DATA

The tailor-made data set encompasses a panel of 107,321 digital companies in Germany that stands out for its precise tracking of firms' locations over their lifecycles (Section 4.1) down to the address level. These firm-level data are combined with regional characteristics from several sources with a strong focus on those which are relevant for digital businesses and the knowledge economy more generally.

4.1 | Firm data

The core data set is provided by North Data (2019). The firm-level data originates from statutory publications of German corporations.[3] It encompasses the date of incorporation, date of termination (if applicable), economic field, a description of the company's main business area, and address history. Since address changes were only tracked digitally after 2007, the analysis covers companies that have entered the market between 2008 and 2017.

These data do not include individual firm information such as financials or the number of employees. However, 90% of all ICT firms have 0–9 employees (in 2017), and 7.8% employ 10–49 employees (Destatis,, 2021a). Therefore we assume that the majority of firms in our sample are similarly small in size as is typical for Germany with its high density of SMEs (Destatis, 2021b).

As there is no agreed-upon definition of the digital economy (e.g., Duvivier et al., 2018) for the purpose of this paper, a digital firm is defined as information-technology driven and internet-based. I select firms using NACE codes (similar to Weber et al., 2018): 62.01.0, 62.01.1, 62.01.9, 62.02.0, 62.03.0, 62.09.0, 63.11.0, 63.12.0.[4] Yet, standard industry classification systems have limitations, especially industries that cross over traditional product categories as is the case for digital firms (Oakey et al., 2001).

Since digital business models complement many other sectors, firms may be registered in other NACE codes although running a digital business model. For example, a survey of German startups finds that 31.8% of new businesses in 2020 were registered in ICT but 66% in the sample state that they are operating on a digital business model (BDS 2020). Furthermore, the NACE code is missing for some firms in the data set. To tackle this issue the description of the company's main business area is used. With the help of a word-search selection, firms that are not registered in the ICT sector but operate on a digital business model were added to the data set.[5] The resulting sample encompasses firms which are similar in their requirements in terms of employees as well as knowledge; that is, in terms of the two factors crucial to their competitiveness. In total, 107 321 firms are covered in the data set over the sample period of 9 years.

The location of a firm in a given year is the location as of 31 December. Each firm has a unique location namely their headquarter, possible subsidiaries are not considered. The panel data set consists of firm-year observations. Firm relocation is observed based on changes in the 5 to 6—digit unique county identifier in the two subsequent years. Firms which exit the market are deleted from the firm register and excluded from the panel after the year of deletion. In total, 12.27% of the resulting sample relocated to a different county between 2008 and 2017 amounting to 14,878 moves.[6]

[3]Commercial register, commercial register announcements, insolvency announcements and electronic federal gazette. The data set is not an official data set, but the data is quasi-official by the virtue of its origin. For details in data generation see North Data (2019).

[4]Covering general programming activities, software development, web portals, data processing, and the development of web pages, processing, hosting, and related activities and web portals

[5]First, the description of the identified ICT firms has been analyzed and the most frequently used words related to IT and software has been identified (Software Development, Internet services, IT-services, information technology, and programming). Then, these keywords are used to obtain those firms operating on digital business models with the help of several word combinations. Further, firms that only distribute their products via a webpage have been excluded (Main keyword Online Shop). For those firms, keywords related to "software development" needed to be included. As an example a firm that is registered in "Placement of workers" has been included in the sample, because the objective of the company is "the operation of a social networking platform for skills enhancement and marketing as well as the provision, brokerage and distribution of products and Internet-based services." Here, Internet-based service has been the selected keyword.

[6]In total there are 1571 firms that move more than once during the sample period.

The location is available on point-level but is aggregated to the county level (NUTS3). The analysis of a panel of firms on NUTS3 level and thus moving away from observing individual firms provides insights into regional dynamics. Moreover, meaningful policy implications can be drawn based on the same spatial unit of analysis as relevant for policy makers seeking to foster entrepreneurship.

4.2 | Regional characteristics

The aggregate of firm level data are merged with several other datasets containing regional characteristics for all 401 German counties between 2008 and 2017. The majority of the data are retrieved from the INKAR database published by the German Institute for Research on Building, Urban Affairs and Spatial Development (BBSR).[7]

Population density (BBSR, 2020) is used as a measure of local market potential as well as urbanization economies and agglomeration benefits (Mameli et al., 2014; Rosenthal & Strange, 2008). As a proxy for local price levels, I use an apartment rent real-estate price index originating from Immobilienscout24—the largest German real estate search site, which is provided by the *Leibniz-Institut für Wirtschaftsforschung* (RWI) (Klick & Schaffner, 2021). The RWI's index captures the difference of the counties' mean rent price to the German average price.[8] Housing costs reflect the willingness to pay for agglomeration benefits (Combes et al., 2019) and the price differential proxies high-cost locations, in particular, which can be assumed to offer the highest benefits in line with costs. Lastly, to measure labor costs, the local average gross income (BBSR, 2020) is used.

Capturing specialization and localization mechanisms, an industrial and a services ratio are used. Both capture the percentage of employees per 100 inhabitants of working age in the respective sector (BBSR, 2020). Moreover, to measure regional colocation of same-sector firms (localization), the number of digital firms per 1000 inhabitants that entered the market after 2007 in the respective county is included. This density measure of incumbent firms also is indicative of industry depth as well as regional specialization.

Locally available knowledge and research intensity are proxied by the number of universities and publicly funded research institutions in the respective counties. To this end, the locations of research institutes belonging to the four major German research associations (Fraunhofer Institut, 2019; Helmholz Gesellschaft, 2019; Leibniz Association, 2019; Max-Planck-Institute, 2019), major research institutes funded by federal states as well as the national government (Forschungseinrichtungen des Bundes und der Länder [OEFW], 2016) are included. For universities, locations published in a public register of colleges and universities (Hochschulkompass, 2020) are used.[9] Both, research and university data are time-variate although the variation is not large with only 31 new colleges which amounts to 6%. Table A1 provides an overview on the regional data.

4.3 | Descriptive statistics

4.3.1 | Firm birth

In Table 1, summary statistics for startups and county-level regional characteristics for the period from 2008 to 2017 are presented. There are 25 new businesses each year in the average county, while Berlin registered 1482 new digital firms at its peak in 2015. Figure 1 presents the regional distribution of total firm births. In absolute terms, Berlin is the top location for digital firms. Hamburg and Munich rank second and are almost on par with at

[7]For an overview of data on regional characteristics and sources see Appendix Table A1.

[8]For more information see Klick and Schaffner (2021).

[9]Firm addresses have been geocoded with the geocode command in R.

TABLE 1 Descriptive statistics for firm birth county level and regional characteristics

Statistic	*N*	Mean	St. Dev.	Min	Max
Population density	4010	522.511	681.476	36	4713
Industrial ratio	4010	17.934	8.521	4.900	91.800
Service ratio	4010	35.548	14.608	12.900	96.600
Gross income	4010	2.381	0.381	1.598	4.367
Firm birth/year	4010	24.926	79.199	0	1482
Firms per 1000 inhabitants	4010	0.507	0.439	0.009	5.283
Price index	4010	−10.616	18.151	−84	119
Price index change	4010	6.571	8.644	−13	85
Research institutes	4010	1.402	3.826	0	45
Universities	4010	1.246	2.517	0	33

Note: Sample includes 401 German counties and covers all years from 2008 to 2017. See Table A1 for more information on data sources and definitions.

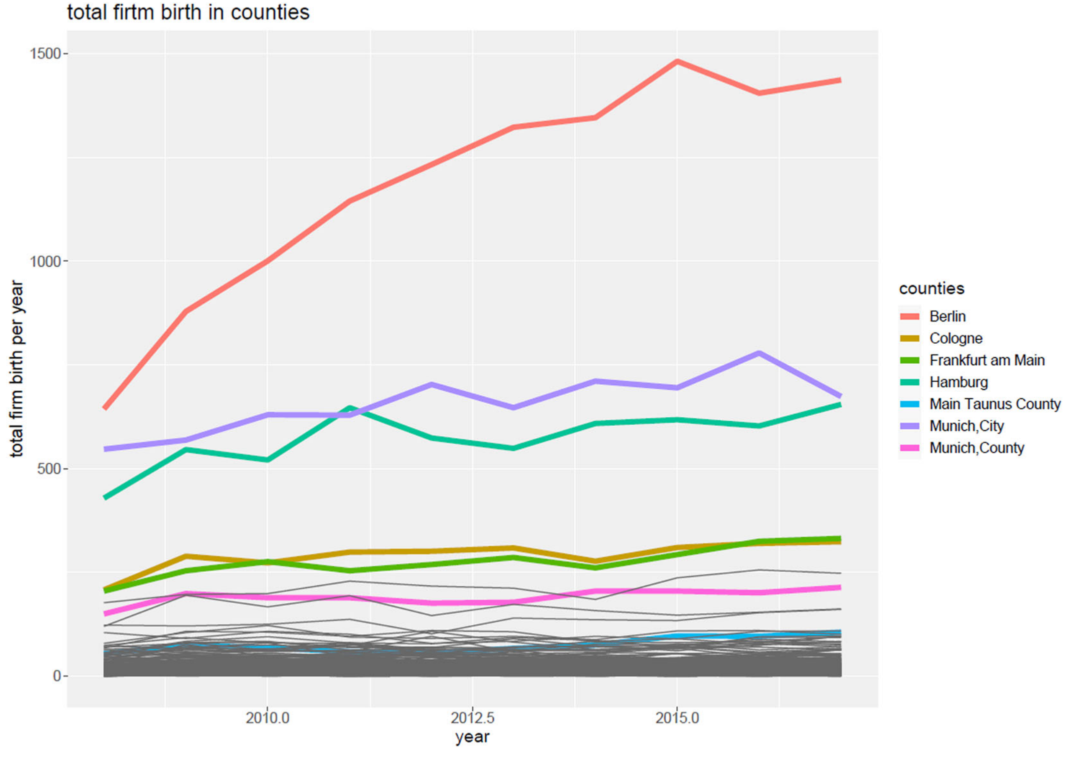

FIGURE 1 Total firm birth of digital firms in counties 2008–2017. All other counties in grey.

655 and 674 new digital firms respectively in 2017, followed by Cologne (324 firms in 2017) and Frankfurt/Main (325 firms in 2017).

It is worth noting that the top six locations for firm births over time are cities with more than 500,000 inhabitants underpinning the fact that start-up culture is most pronounced in urban areas. Figure 2 presents firm births relative to the population. Munich County—neighboring Munich City—has the highest firm birth per 1000 inhabitants. Also, Main-Taunus-County—neighboring Frankfurt/Main—shows high firm births per 1000 inhabitants. The descriptive statistics indicate—consistent with theory and similar to Bade and Nerlinger's (2000) findings—that absolute startup rates are highest in core cities, while firm birth relative to the population can also be high in the core areas' periphery.

4.3.2 | Firm mobility

There are 10,108 out of 1,443,600 origin-destination county pairs with positive relocation counts over the study period (see Table 2), that is 0.7% of observations. When pooled over the whole sample period, there are 3.6% positive flows. Compared to the 12% of firms moving to different counties, this aggregated flow appears to be very small. The reason for this is that the data are highly spatially dispersed as 80% of the origin-destination pairs only have one movement in a given year, accounting for 54.4% of all flows. That means that 20% of the positive observations cover 45% of all relocations. Strikingly, the largest single relocation count is 58 businesses moving from Munich City into Munich County in 2017 firms (29.9% of businesses out-migrating Munich City). To contrast this, there is only one county (Hildburghausen) which has never been a destination of a relocating digital firm within the sample period.

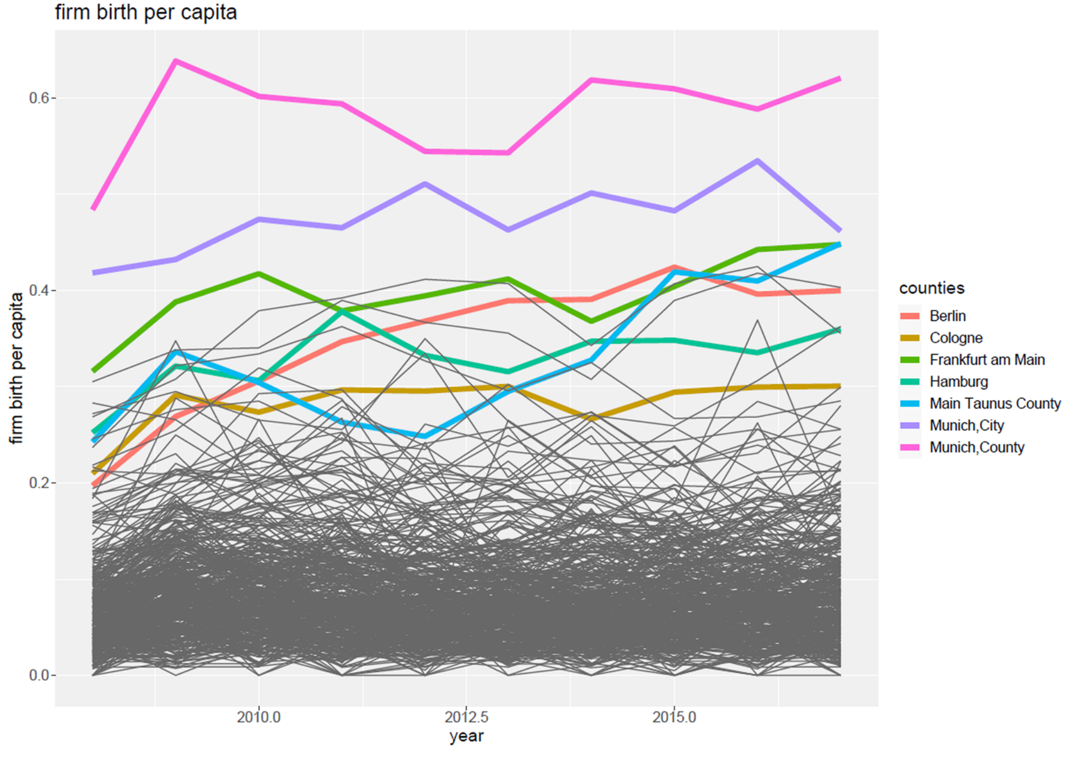

FIGURE 2 Firm birth of digital firms per 1000 inhabitants in counties 2007–2017. All other counties in grey.

TABLE 2 Descriptive statistics for bilateral flows and distance

Statistic	N	Mean	St. Dev.	Min	Max
Bilateral flow	1,443,600	0.01	0.20	0	58
Distance in km	1,443,600	304.40	151.66	1.26	824.48
Bilateral flow only positive	10,108	1.47	1.85	1	58

Note: Sample covers the years 2008–2017.

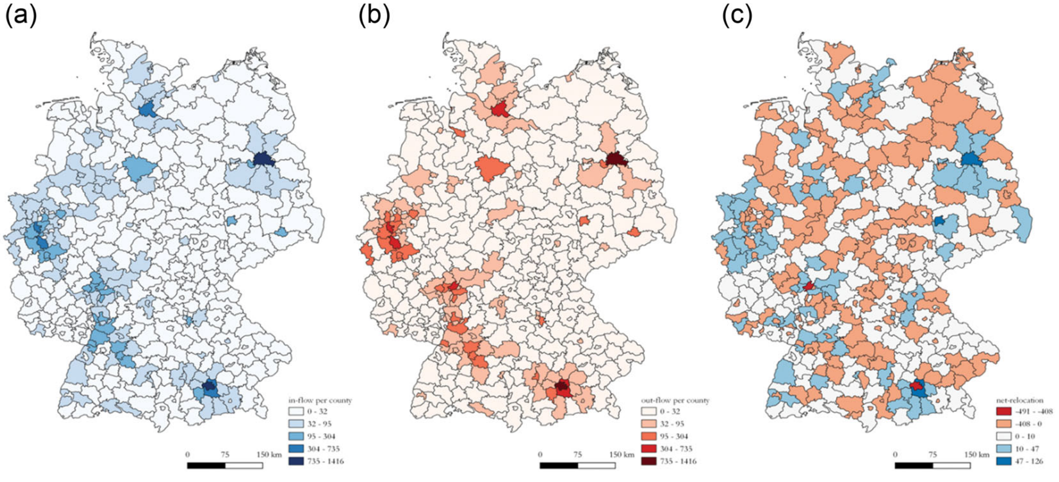

FIGURE 3 Firm migration patterns. (a) displays the inflows of firms per county, (b) the out-flows per county, and (c) the net relocation. All flows have been summed up to all examined years. Basemap: GeoBasis-DE/BKG 2021.

When looking at counties receiving firm inflows, movement between the City of Munich and County Munich and vice versa accounts for the largest relocation flows with 539 firms (3.6%) relocating in total. The third and fourth largest flows occur from firms migrating from Munich to Berlin and from Hamburg to Berlin, respectively. Table 2 presents summary statistics on the bilateral flows and distance.

Figure 3 shows relocation activity geographically, where a shows the absolute inflow of firms in a distinct county for all years, b shows out-flows per county, and c the net relocation. In c, the counties marked in red show a negative net relocation, while blue-colored counties have an overall surplus of in-migrating firms. We see that large cities like Munich, Hamburg, and Berlin have both strong inflows and outflows. However, Berlin is the only metropolitan city gaining a surplus of digital firms while counties in the surroundings of big cities like Munich and Frankfurt/Main seem to attract relocating firms.

In sum, the Figures reveal two relocation patterns. First, firm migration flows into big cities, in particular, big inflows to Berlin, which is indicative of the fact that digital firms behave in accordance with Duranton and Puga's (2001) model; that is, they move to a specialized city with the highest absolute firm birth rates as is the case for Berlin. Second, we see a striking pattern of an urban-core and urban-periphery dynamic.

Moreover, the strong persistence in the data is reflective for a very persistent relocation behavior of firms with only few counties being deemed attractive for relocating firms. As much as this pattern is in accordance with expectations it also poses econometric challenges. In terms of the method used, the gravity model is geared particularly well to such persistence in observations and indeed spatial disparities while in terms of the research question it provides a framework for understanding the factors explaining these high flows into particular counties. The empirical, as well as the robustness section, will deal with this empirical issue in detail.

5 | METHODOLOGY AND EMPIRICAL MODELING

Empirically, three main models are estimated. First, I model the initial location of young digital firms on county level using annual data covering the years 2008 to 2017 as a function of regional characteristics of the firms' birth locations (Section 5.1). Second, to model relocation I use the same set of variables in a gravity model which considers both origin and destination counties, in two specifications (models 2 and 3, as presented in Section 5.2).

5.1 | Firm birth analysis

To explore the initial location choice of new digital businesses, four pooled and fixed-effect OLS models are employed, where models 1.1 and 1.3 are fixed effects models. Model 1.2 and 1.4 are pooled models (without fixed-effects) mainly to benchmark against the literature. The baseline model is presented in (1):

$$\text{lnfirm birth}_{i,t} = lnl_{i,t} + G_i + NG_i + T_t + \gamma_i + \varepsilon_{i,t.} \tag{1}$$

Where the dependent variable is digital firm birth per 1000 inhabitants in location i at time t (model 1.1), and the absolute firm birth in location i at time t (model 1.3). The independent variables enter both models as follows: $lnl_{i,t}$ refers to the set of locational characteristics in location i at time t (population densities, service ratio, industrial ratio, gross income, price index (change), universities, research institutes).

All variables enter the model as laid out in the data section above. G is an indicator variable equal to 1 if the county is a city with more than 500,000 inhabitants. NG is an indicator variable equal to 1 if the counties share a border with a city of more than 500,000 inhabitants capturing spatial externalities of agglomerations. For models 1.1 and 1.3, T_t is a year fixed-effect controlling for common trends, γ_i is the time-invariant county fixed-effect capturing unobservable county-specific and time-invariant factors which are potentially correlated with the number of new firms. $\varepsilon_{i,t}$ is the error term.

5.2 | Models for the analysis of firm relocations—Gravity model

Firm relocation is typically modelled by individual firm location decisions following McFadden's choice model (1973) (e.g., Arauzo-Carod et al., 2010; Bodenmann & Axhausen, 2012; Kronenberg, 2013). Here, the profit-maximizing, fully-informed firm decides to move after having screened profitability prospects in all possible alternative locations. The estimation outcome captures the probability to move to a particular location given the destination's characteristics.

A more precise understanding of the factors motivating relocations, however, requires modeling firms' movements *between* locations, that is, a direct differential between origin and destination characteristics to explain firms' bilateral flows (Conroy et al., 2016; Kohler, 1997). The key underlying assumption is that the new location reflects the firm's "revealed preference" and thus is the result of a curating process of all relevant alternatives. A standard theoretically embedded model for bilateral flows in a monopolistic competition environment as it should be the case for the digital economy is the gravity model.

The gravity model is the workhorse theoretical framework in empirical analysis of bilateral flows in trade as well as their spatial determinants (Head & Mayer, 2014; Yotov et al., 2016). However, it has been applied in many other fields, such as labor migration (Karemera et al., 2000), tourism flows (Morley et al., 2014), or the selection of airline hubs as an example of industry location (Drezner & Drezner, 2001; Redding et al., 2011; see also Head & Mayer, 2014, p. 148f). The key similarity between these literatures and this paper is that production factors and/or facilities are relocated across spatial units.

Closest to the line of analysis in this paper is the literature on FDI (e.g., Burchardi et al., 2019; Egger & Pfaffermayr, 2004; Portes & Rey, 2005) which in essence constitutes a cross-border firm or subsidiary relocation. The model's intellectual baseline is that larger counties are expected to receive greater relocation flows due to gravitational force while two counties further apart have lower relocation flows. Moreover, firm-specific costs are captured in the gravity model's distance variable to overcome the fact that data on costs associated with firm mobility are not available.

Relocating firms in monopolistic competition incur considerable monetary moving costs and, possibly even more importantly, indirect costs such as costs to find new employees or establishing a new network and obtaining information. In contrast to the assumptions of a choice model firms are assumed to have greater knowledge about geographically closer markets and locations. In the firm mobility literature, Conroy et al. (2016) and Pan et al. (2020) use a gravity-like model, albeit without considering the key determinant of the gravity model—the distance. Pan et al. (2020), however, suggest its usage for future studies. Hence, this paper is the first to analyze firms' relocation explicitly modelling the costs using distance as a proxy.

The theoretical justifications for the model have been provided by Anderson (1979) and Anderson and Van Wincoop (2003) among others. In terms of the model's theoretical foundation, a profit-maximizing individual firm k decides to relocate from location i (origin) to location j (destination) if the expected return R in the destination is greater than the expected return in the origin plus the costs of relocation C as a function of distance d:

$$E\left(R_j^k\right) > E\left(R_i^k\right) + C(d_{ij}). \tag{2}$$

When the above condition is satisfied and the firm relocates, the variable M_{ij}^k being equal to 1 (0 otherwise) is defined. By aggregating individual movements by county and employing a general gravity-type model specification, M_{ij}^k can be expressed as:

$$\sum M_{ij}^k = M_{ij} = f(\Delta l_{ij}, d_{ij}), \tag{3}$$

where $i = 1, 2, \ldots, 401$; $j = 1,2, \ldots, 401$ (with $i \neq j$) and l = vector of regional characteristics of the origin i and destination j and d_{ij} is the distance between i and j.

I conduct the classical gravity specification expressed in log-linear form:

$$ln\, M_{ij,t} = ln\, l_{i,t} + ln\, l_{j,t} + N_{ij} + ln\, d_{ij} + T_t + \gamma_i + \varphi_j + \varepsilon_{ij,t}, \tag{4}$$

where $M_{ij,t}$ refers to gross relocation flows from county i to county j in year t. The dependent variable enters the model in log form to smoothen its distribution as well as to allow for the coefficients to be interpreted semi-elasticities. To assess whether the same location factors are important for firm births as well as for relocating firms I use the same location characteristics l as presented in model 1 (population densities, service ratio, industrial ratio, gross income, price index [change], universities, research institutes). The specification for model 2 (Equation 4) captures push and pull factors. Consequently, $lnl_{i,t}$ encompasses location characteristics for origin i (push factors), while $lnl_{j,t}$ are the corresponding location characteristics in destination j (pull factors) in year t.

N_{ij} is a dummy variable capturing contiguity equal to one if i and j are neighbors. As the key variable to proxy costs of relocation d_{ij} is the linear distance in kilometer among the centroids of the counties. T is a year fixed effect that controls for common time shocks and general trends. γ_i and φ_j are time-invariant origin and destination-specific fixed-effects controlling for any unobservable county-specific factors affecting relocation and eliminating biases due to multilateral resistance (Bertoli & Fernández-Huertas Moraga, 2013). $\varepsilon_{ij,t}$ is the error term.

The second gravity model (model 3, Equation 5) is a variant of model 2 using a differential approach between origin and destination counties (Conroy et al., 2016) to investigate the differences in location characteristics. This is a useful exercise as it can be expected that most firms make their decisions by evaluating their current location vis-à-vis other options with the new location reflecting the "winner" of these considerations and therefore capturing the firm's revealed locational preferences.

The empirical representation for the panel data can be expressed as

$$ln\, M_{ij,t} = \sum_{l=1}^{L} \Delta X_{l,ij,t} + N_{ij} + Ind_{ij} + T_t + \gamma_i + \varphi_j + \varepsilon_{ij,t}, \tag{5}$$

where $M_{ij,t}$ refers to the gross relocation flow from county i to county j measured as the log of the count of firms relocating from i into county j in year t. $\Delta x_{l,ij,t}$ is the difference in the location characteristics between origin county i and destination county j (destination minus origin) across the set of l empirical variables in year t. N_{ij}, $ln\, d_{ij}$, T_t, γ_i, φ_j, and$\varepsilon_{ij,t}$ are the same as in model (2).

This paper relies on both OLS as well as Pseudo Poisson Maximum Likelihood (PPML) as estimation methods for the gravity model. Comparing results of these two empirical exercises allows interesting conclusions as to the roles of zero-observations. As highlighted by Silva and Tenreyro (2006), the most common practice in empirical applications of the gravity model has been to take natural logarithms and to estimate the model by using OLS. However, the literature, in particular on trade, has developed several empirical solutions to deal with zero observations, that is nonflows. One possible and commonly used solution is the PPML model as proposed and discussed by Santos Silva and Tenreyro (2006).

In the PPML, possible biases due to the amount of zero observations in the dependent variable are corrected while also accounting for heteroscedasticity. Further, Monte-Carlo simulations show that the estimator performs well in spite a large proportion of zeros (also see Yotov et al., 2016, p. 20) and the validity of the estimator does not depend on very strong assumptions of the distribution of the data as for example would be the case for a zero-inflated model (Silva & Tenreyro, 2021). Overall, the PPML estimator is widely accepted to be very well suited for gravity estimations.

6 | ECONOMETRIC RESULTS

6.1 | Firm birth locations

Econometric results on German firm births between 2008 and 2017 are presented in Table 3, where Column (1) and (2) corresponds to the models based on firm birth per 1000 inhabitants. Model 1.1 shows the results of the fixed-effects estimation and Model 1.2 displays the results of a pooled OLS (estimation without fixed-effects). The results of the pooled OLS mainly serve as a benchmark to assess level effects against the FE-models that capture the sensitivity to changes in county characteristics and to assess which county characteristics are indeed absorbed by the FE.[10] Column (3) and (4) refer to model (1.3 fixed effects and 1.4 pooled OLS) with the total number of firm birth as dependent variable.

[10]The pooled OLS model seems inappropriate to serve as a baseline model. The fixed effects model estimation appears superior because i. it seems unlikely that county characteristics are not randomly distributed and ii. the fixed effects are individually significant and increase the R^2. Results of a Hausman–Test show that fixed effects is chosen over random effects.

TABLE 3 Regression results model (1) firm birth ordinary least squares (OLS)

Dependent variable	Firm birth per 1000 inhabitants(ln) (1.1)	(1.2)	Total firmbirth (ln) (1.3)	(1.4)
Population density (ln)	−0.192 (0.031)***	0.003 (0.001)***	−1.612 (0.486)***	0.083 (0.011)***
Industrial ratio	−0.001 (0.001)	0.0005 (0.0001)***	−0.0002 (0.006)	−0.002 (0.001)*
Service ratio	−0.001 (0.0005)***	0.0003 (0.0001)***	−0.015 (0.005)***	−0.008 (0.001)***
Gross income (ln)	0.005 (0.024)	−0.066 (0.006)***	−0.231 (0.364)	0.031 (0.091)
Firms per 1000 inhabitants (ln)	0.183 (0.010)***	0.157 (0.005)***	2.171 (0.144)***	0.613 (0.058)***
Price index	0.002 (0.0004)***	0.001 (0.00004)***	0.030 (0.005)***	0.007 (0.001)***
Price index change	−0.002 (0.0003)***	−0.002 (0.0001)***	−0.028 (0.005)***	−0.009 (0.001)***
Metropole (G)	0.043 (0.038)	0.019 (0.005)***	2.977 (0.384)***	0.159 (0.038)***
Neighbor is metro (NG)	−0.583 (0.088)***	0.001 (0.001)	−4.598 (1.354)***	0.089 (0.020)***
Firm birth lag (ln)	−0.009 (0.001)***	0.0005 (0.001)	−0.142 (0.019)***	0.591 (0.015)***
Universities	0.005 (0.002)**	0.0004 (0.0004)	−0.043 (0.024)*	0.050 (0.005)***
Research Institutes	0.00002 (0.003)	0.002 (0.0003)***	−0.031 (0.028)	0.007 (0.002)***
County FE	Yes	No	Yes	No
Time Fe	Yes	No	Yes	No
Observations	4010	4010	4010	4010
Adjusted R^2	0.87	0.767	0.889	0.798

Note: Significance levels are: ***p < 0.001; **p < 0.01; *p < 0.05; all standard errors clustered by county (in parentheses next to coefficients); dependent variables as logarithms.

6.1.1 | General location-specific factors

Population density is used to proxy for agglomeration effects (Rosenthal & Strange, 2008) and to partially control for market size and accessibility (Arauzo-Carod 2005), that is why a positive effect of population density on firm birth is expected (see Hypothesis 1 in Section 3.3). With regard to models (1.2) and (1.4), there is a positive and significant effect of agglomeration effects on digital firm birth in line with the literature as well as theory-based expectations laid out above. However, when including time- and county-fixed effects that control for general time trends as well as county-specific characteristics (Models 1.1 and 1.3), the effect of agglomeration benefits remains significant but the coefficient becomes negative. That is, an above-average growth in population density (which implies an above-average growth in population) has a negative effect on firm birth. While this switch in the signs may seem counter-intuitive at first, it overall shows that as shown by the pooled model densely populated counties offer a host of agglomeration externalities conducive to firm birth. The FE model instead reveals that the change in population density—as the effect is identified by difference in the growth of population density does not have an effect on digital firm birth. Thus, above-average growth is probably driven by general labor market effects such as strong industry players that offer attractive jobs or various amenities of dense cities. As reflected in the knowledge spillover theory of entrepreneurship (Acs et al., 2009; Audretsch et al., 2008), new innovative firms as they are part of the employed data set thereby originate in the commercialization of ideas in incumbent firms or universities. These new ideas need time to develop into marketable innovation, what makes it reasonable to expect a time-lag from a growth in population until a growth in firm birth.

Similarly, there are significantly more digital firm births in cities above 500,000 inhabitants than elsewhere (see Models 1.2 and 1.4). However, once controlling for unobservable characteristics of such metropolis (1.1 and 1.3), there still is a positive effect in absolute terms, but not relative to the population. Therefore, an above-average firm birth activity is not solely driven by the quantity of the population, but by unique, unobservable characteristics of metropolitan cities that oftentimes come with greater density like for example a cultural setting or entrepreneurial culture.

These unobservable advantages seem to spill over into neighboring counties, as there is significantly more firm birth (1.4) for neighbors of metropolis cities, but the effect vanishes with the inclusion of the fixed effect (1.1 and 1.3). This is evidence that core cities' networks and knowledge are accessed at lower costs. Moreover, entrepreneurs might be too risk-averse to rent expensive offices in their first years of uncertain income streams.

6.1.2 | Industry-specific factors

Additionally, localization economies have a positive effect on firm birth. For each digital firm per 1000 inhabitants, there is 0.18 digital firm births per 1000 inhabitants, or in other words, five digital firms bring about one additional firm birth. There are two possible explanations. First, colocation lowers the costs of starting a business and allows access to a more diverse range of inputs and complementary goods (Feldman et al., 2005).

With a large number of small-sized firms (we assume this is the case as laid out above), the chance of employees leaving a firm and starting their own business increases (Pijnenburg & Kholodilin, 2014) increases. Therefore, part of the higher firm birth rates with higher co-location potentially occurs from spin-offs. Further, the number of firm births when lagged by 1 year is significant and negative. This means that the average county does not register steadily positive growth rates in the digital sector, but high growth rates is a phenomenon in very few cities.

However, the models also reveal that a higher service ratio is associated with less digital firm birth in absolute and relative terms. This could mean that a very high regional specialization in one service sector (e.g., Frankfurt in banking or Düsseldorf in consulting) does not offer enough industrial diversity for digital firm birth. Considering the fact that the service ratio includes ICT services as well, this might be indicative evidence that digital firms seek proximity to a broad number of recipients which are not other service sector firms. This is in contrast to the fact that there is no such effect for the industrial ratio.

Second, with increasing colocation of similar firms, a large labor pool of specialized workers—crucial for digital businesses—is available. This agglomeration advantage in addition to the local amenities exceed the disadvantages of higher prices (Rossi & Dej, 2019). This is also what I find in my model: in counties with 5% higher prices for apartment rents than German average there is one firm birth per 1000 inhabitants, while a 5% growth in prices (relative to the German average) leads to one less firm birth per 1000 inhabitants. That is, higher price levels are accepted by entrepreneurs while strong price growth has a negative impact on firm birth. However, gross income as a measure of overall wealth does not have a significant effect. In sum, this is indicative that colocation benefits outweigh the costs.

Digital firm birth activity was expected to be high in cities where firms derive advantages from agglomeration benefits that is firms show a strong preference towards cities (Hypothesis 1). Overall, results show a high level of agglomeration benefits is conducive to firm birth, and that digital firm birth activity is mostly driven by strong colocation of similar firms and the regional knowledge base.

Another important finding is that universities contribute significantly to higher firm birth rates providing support for Hypothesis 2 (Section 3.3). For each additional university there is 0.5 firm births per 1000 inhabitants. For example, an average county with a university has 38 digital firm births on average, while a county without a university has only 11. One link between universities and start-up activity is that innovative students or college employees found spin-offs because the expected value of commercializing an idea is higher for the individual than the expected income offered through employment in an established firm (Pijnenburg & Kholodilin, 2014).

Moreover, university-bound knowledge may breed product development particularly by young digital firms while regional knowledge is crucial for start-up rates due to spillovers' spatial limits (Bade & Nerlinger, 2000). The number of research institutes is significant in the pooled model, while there is no significant effect in the fixed-effects model. This shows that a general presence of research institutes is conducive to firm birth, but an increase does not imply more firm birth in the same year. Hypothesis 2 is therefore partly confirmed. This indicates that a mere geographical presence of knowledge-producing institutions is not sufficient for digital firm birth, but rather that transfer channels and potential labor market effects differ between research institutions and universities.

6.2 | Relocation

Tables 4 and 5 show the results for the gravity models laid out in Equations (2) and (3). In model (2), absolute origin and destination characteristics are included. A positive/negative coefficient in the origin/destination characteristic

TABLE 4 Regression results model (2) gravity model (PPML)

	M_{ijt}	
Dependent variable	**(2.1)**	**(2.2)**
Population density (ln), origin	1.19772 (1.05330)	0.12288*** (0.01350)
Population density (ln), destination	0.26170 (0.99375)	0.12734*** (0.01331)
Industrial ratio, origin	0.02284 (0.01556)	−0.03132*** (0.00166)
Industrial ratio, destination	0.01307 (0.01609)	−0.02101*** (0.00165)
Service ratio, origin	0.02151 (0.01238)	−0.00050 (0.00086)
Service ratio, destination	0.00223 (0.01210)	−0.00702*** (0.00092)
Gross income (ln), origin	−1.48138 (0.83224)	2.04666*** (0.11828)
Gross income (ln), destination	0.68121 (0.81643)	1.25158*** (0.11832)
Price index, origin	−0.00473 (0.01103)	0.00997*** (0.00060)
Price index, destination	0.01612 (0.01055)	0.00563*** (0.00063)
Price index change, origin	0.00218 (0.00962)	−0.00952*** (0.00102)
Price index change, destination	−0.01577 (0.00942)	−0.00726*** (0.00108)
Firm birth (ln), origin	−0.07896 (0.04685)	0.16464*** (0.02044)
Firm birth (ln), destination	0.02792 (0.04455)	0.33232*** (0.02124)
Firms per 1000 inhabitants, origin	−0.13080 (0.09183)	0.00020 (0.00011)
Firms per 1000 inhabitants, destination	0.20865* (0.08660)	0.00014 (0.00012)
Research Institutes, origin	−0.00839 (0.04625)	0.00282 (0.00249)
Research Institutes, destination	−0.05027 (0.04794)	−0.01771*** (0.00270)
Universities, origin	0.07386* (0.03258)	0.10703*** (0.00360)
Universities, destination	0.01235 (0.03536)	0.13669*** (0.00385)
Neighbor county	1.21148*** (0.04223)	1.24792*** (0.03039)
Distance (ln)	−1.15491*** (0.01647)	−1.03266*** (0.01164)
Time Fixed Effect	Yes	No
County Fixed Effects	Yes	No
Num. Obs.	1,432,818	1,443,600

Note: Significance levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$, standard errors in parentheses; dependent variable as log-link.

TABLE 5 Regression results Model (3) Gravity model difference approach (PPML)

Dependent variable	M_{ijt} (3.1)	M_{ijt} (3.2)
Population density	0.00013 (0.00029)	−0.00015*** (0.00001)
Industrial ratio	−0.00130 (0.01076)	0.00215* (0.00100)
Service ratio	−0.00733 (0.00791)	−0.00117 (0.00061)
Gross income	0.00025 (0.00022)	−0.00003 (0.00003)
Price index	0.01662 (0.00900)	−0.00651*** (0.00065)
Price index change	−0.01464 (0.00808)	0.00465*** (0.00110)
Firm birth	−0.00021 (0.00034)	−0.00064*** (0.00015)
Firms per 1000 inhabitants	0.17086* (0.06783)	0.09770*** (0.02743)
Research Institutes	−0.02315 (0.03586)	−0.02936*** (0.00286)
Universities	−0.01905 (0.02846)	0.05375*** (0.00467)
Neighbor County	1.21144*** (0.04226)	1.62317*** (0.02775)
Distance (ln)	−1.15495*** (0.01648)	−1.03624*** (0.00980)
Time fixed effects	Yes	No
County fixed effects	Yes	No
Num. Obs.	1,432,818	1,443,600

Note: Significant levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$, year, origin and destination fixed effects included; standard errors in parentheses; variables coded as differences between origin and destination; firm birth and firms per 1000 inhabitants refer to digital firms; dependent variable as log-link.

implies a higher/lower flow of out-migrating/in-migrating firms. In model (3), the variables display the difference between origin and destination. Thus, a positive/negative coefficient implies higher/lower values in the destination than in the origin.

Similar to the results on firm birth, the models are presented with fixed effects (0.1) and without fixed effects (0.2), respectively. In the fixed effects models which controls for location-specific characteristics and common time trends which affect all locations none of the coefficients for regional characteristics have significant explanatory power in comparison to the models without fixed effects that capture level effects. This result in its own right sheds light on the lack of relevance of regional characteristics up and beyond their idiosyncrasies for firms' relocation decisions.

That is, why the fixed effects model employed here generates considerably less significant effects on the vector of regional characteristics explicitly controlled for in the model in contrast to, for example, Kronenberg (2013) who employs a pooled OLS estimation without fixed effects. Nonetheless, the model's remaining significant determinants—strikingly mostly industrial factors—can be interpreted as relevant factors for interregional relocation when controlling for county characteristics. As there are many unobservable regional characteristics we prefer to focus on the fixed-effects model for the interpretation of the results.

6.2.1 | Industry-specific factors

Results show that other than expected, except for digital firms per 1000 inhabitants (in the destination) and universities (in the origin), none of the regional characteristics have a statistically significant effect on relocation.

Therefore, the industrial- and knowledge base is most important. An urban "upgrading" in, for example, amenities is of limited use to foster firm birth, as soft factors are not identified as pull factors.

Both models show a significant "gravitational" effect for localization as measured by firm density (digital firms per 1000 inhabitants) and thus same-sector density can be identified as a relevant pull factor. Model (2) predicts that with one additional digital firm per 1000 inhabitants relocation inflows increase by 23 percentage points. Thus, model (3) partly provides support for Hypothesis 3, that is, firms do not necessarily move.

This result is different from Stam's (2007) and Nguyen et al.'s (2013) investigation of manufacturing firms. In contrast to manufacturing, digital firms are more likely to cluster and seek out competition. Furthermore, there is no significant effect for salaries or housing prices. In line with Rossi and Dej (2019) and Kronenberg (2013) this finding implies that digital firms do not necessarily adopt pure cost minimization, but choose counties where they can benefit from the availability of high-skilled workers. This is comparable to the above finding for digital start-ups, indicating an industry-specific behavior which is independent of firm maturity.

Moreover, my results give weak indication that digital firms move from diverse into more specialized cities seeking proximity to their competitors. For manufacturing, Duranton and Puga (2001) find firms to innovate in diversified cities and then switching to mass production in localized cities. Although digital firms are usually not producing physical goods and economies of scale are limited in many digital branches, it seems reasonable that digital firms do reach some point of process standardization and stable growth path for which locational diversity is less of a requirement. That is, knowledge bound to diverse cities and their universities can become less relevant for digital firms later in their lifecycle. For every university, the outflow of firms increases by 107% (see model 2). When assuming that universities contribute to a high and diverse output of firm births in different sectors and thereby to a diverse local economy, this finding provides further support for the hypothesis that firms move from diverse into more specialized places.

An alternative explanation may be that firms seek direct competition with others as innovation and knowledge in software, for example, are hard to be legally protected via patents thus rendering spillovers, networking, and shared work and customer flows easier and more attractive. Consequently, while firm birth seeks knowledge bound in universities, relocates seek industry-specific knowledge and competition.

In sum, digital firms show a strong tendency for clustering and co-location with other digital firms. This is a self-reinforcing mechanism that contributes to higher firm birth and inflows of relocating businesses. Therefore, the expectation that agglomerated areas receive higher inflows (Hypothesis 3) cannot necessarily be confirmed since benefits for firms are rather driven by knowledge and industry rather than pure population density.

6.2.2 | Spatial mobility pattern

Models (2) and (3) reveal a very strong regional persistence of digital firms. This supports the indicative findings from the descriptive statistics. In light of the high number of zero-flows, this result is even more striking. Expecting transaction costs increasing with distance, counties being in closer geographical proximity show higher relocation flows than counties further apart. The results show that with a one unit increase in the ln(distance) between county i and j the average sizes of the relocation flow from i to j decreases by the factor of 0.31. Compared to all other relocation flows, flows between neighboring counties are expected to be 236% higher. In other words, counties receive more than twice as many relocates from their neighboring counties than from others. This confirms the expectation from Hypothesis 4 that relocation flows decrease with distance.

As shown in Section 4.3.2, the flow between the City of Munich and Munich County account for the biggest relocation flows. In light of the dominance of these observations in the sample I repeat the empirical exercise above excluding this pair. My results still hold (see Appendix A2 and A3) indicating that regional persistence and industrial path dependence are crucial for all German counties and results are not solely driven by the "Greater Munich effect."

Moreover, the results are consistent with Conroy et al. (2016) who also find a strong neighbor effect for interstate relocation for US manufacturing firms. Furthermore, the results are in line with Knoben (2011) who finds that young ICT firms, which are highly dependent on outside resources remain within the local economy since costs do play a significant, prohibitive role in relocation choices.

6.3 | Robustness

Several robustness test have been conducted. To assess possible multicollinearity issues in the specification, a Correlation Table is presented in Table A2. Due to the high correlation of universities and research institutes (0.8), a variance inflation factor (see Table A3) has been calculated. Results show that none of the values is above 10, which indicates that multicolinearity is not a concern for the regression results (Wooldridge, 2013, p. 98). Moreover, when taking out either of these variables from the model or when aggregating the two variables the estimations yield the same results, and R^2 remains the same. Next to confirming robustness, these results indicate that the findings presented above indeed warrant a differentiation between the different types of knowledge these institutions offer for digital industries.

For the robustness of the gravity models, the key concern certainly are the large number of zero-flows, as in fact only 0.7% amount of observations in my sample are positive. When including these zero observations in the estimation as done above conceptually we are estimating county-pairs with firm mobility relative to nonintegrated counties that is those ones that have no firm mobility between them (the majority of observations). The advantage of this estimation is that it allows us to understand factors which explain why certain counties have firm mobility at all and the determinants which “switch on” firm mobility. This is similar to the extensive margin in the trade literature (e.g. Chaney, 2008; Helpman et al., 2008) where the extensive margin is important for FDI (e.g., Blum et al., 2020), and thus is my preferred specification.

To analyse the relevance of zero-observations for my findings I repeat all gravity estimations as a negative binomial (NB) as an alternative to the PPML and find consistent results. Furthermore, I repeat the empirical exercise of the main specification this time excluding all county-pairs without any flows, thereby effectively estimating the intensity of firm mobility between counties with any firm migration. I conduct this exercise using the three standard methods in the literature PPML, NB, and OLS (see Appendix A4 and A5, denoted as OP). Results can be interpreted as a difference between the extensive and intensive margin in relocation flows for regional characteristics. Results of a specification excluding zero-observations and the one carried out above including all observations allows for re-pivoting the analysis towards focusing on the determinants within “mobility-relevant” counties rather than benchmarking “mobility-relevant” counties against those without any mobility at all.

Across all models, the distance between counties and contiguity are the predominant explanatory factors for the size of relocation flows between counties. Both variables are significant at a 99% level and similar in magnitude to the specification analysed above. This is indicative of the fact that there are structural differences between receiving/sending counties and those which never have any mobility of digital firms which explain mobility which is in line with the persistence seen in the summary statistics. Seeing that both conceptual, as well as methodological approaches, render consistent results yields further support for the analysis presented above.

As elaborated in Section 5.2, the inclusion of origin- and destination-fixed effects in the panel model above controls for multilateral resistance, as commonly done in the trade literature using gravity models. Moreover, usage of origin-time and destination-time fixed effects to fully account for multilateral resistance is commonly employed (e.g., Head & Mayer, 2014; Olivero & Yotov, 2012, p. 151).

In the relocation context, these fixed effects can be seen as a barrier to move that a firm faces with all its possible new locations. In other words, the estimation captures relative changes in counties attractiveness while controlling for all observable and unobservable country-specific variables that vary in the respective dimension (Yotov et al., 2016, p. 19). In this case, the inclusion of origin-time and destination-time fixed effects leaves variation in contiguity and distance. The results in Table A6 show that the size and significance of the coefficients remain consistent with the baseline estimations in Table 4.

Further, binary choice models capturing the probability to move have been estimated (Appendix Tables A7–A11). The advantage of such models over continuous models as used in this paper is that individual firm characteristics are controlled for. That is one of the reasons why they have been used in the literature (e.g., Kronenberg, 2013; Nguyen et al., 2013) so extensively. Results also identify colocation of similar firms as a significant pull factor. I do find that the spatial components of distance and flow into neighboring counties are most significant across all models. Thus, there is reasonable ground to believe that the gravity model with its theory-based spatial dynamic reflects and encapsulates the results obtained by a choice model. Moreover, this robustness check confirms that my findings in spite of a different empirical strategy are comparable to Kronenberg (2013) and Nguyen et al. (2013).

To test the feasibility of the data selection of digital firms, the relocation analysis has been conducted with a reduced sample by only using the firms that have been selected as ICT via the NACE code (Tables A8 and A9). This sample is more restrictive and leaves 3600 origin-destination pairs with positive relocation-flows (instead of 10,108 in the full sample). This selection absorbs some heterogeneity of the firm landscape: I assume that firms are more similar in their business models referring back to the example of 31.8% of new businesses being registered in ICT but 66% state to operate on a digital business model (BDS, 2020). Results show, contrary to the full sample selection, that higher firm birth rates in counties results in higher outflows of ICT firms. This shows that, the more homogeneous the market, that is higher competition as less diversity, the more firms leave. Therefore, a diverse firm landscape of the digital economy can be advantageous for a county to prevent firms from leaving.

Additionally, a complementary analysis has been performed on NUTS2 aggregation (see A10 and A11). This aggregate includes firms that relocate over longer distances and therefore tolerate higher relocation costs (in comparison to NUTS 3) and increases the share of positive flows to 29%. The striking difference between the two geographical aggregations is that the gross income in the origin is significant and negative across all tested specifications and the number of universities in the destination is significant and positive.

Drawing on the difference to the results from the NUTS 3 aggregation that shows a clearer picture on the role of universities and knowledge diffusion: As NUTS 3-units show a higher out-flow of digital firms with more universities, NUTS2-units show a higher in-flow of digital firms. This may indicate that access to knowledge is a striking advantage that firms accept higher costs for, especially in very early stages. Further, firms accepting higher costs of movement relocate into regions with lower gross income that is they follow a price mechanism here.

In sum, the key findings of this paper are robust to various specifications. One caveat is that the model may suffer from endogeneity issues, such as selection into locations based on characteristics. The best we can do to remedy is to employ a range of fixed effects to identify the effect. Moreover, it is reasonable to assume that our results indicate a lower bound for the "real" effect should selection be an issue therefore not posing a substantial threat to the empirical strategy of the paper.

7 | CONCLUSIONS

This study investigates the determinants of firm birth location and relocation choices of young digital firm in Germany from 2008 to 2017. The main objective of the paper is to shed light on the importance of regional characteristics for firm birth and relocation of firms, by asking if both types of location choices are driven by the same determinants. Results reveal that the set of regional characteristics is conducive to firm birth while only very few these regional characteristics are significant when seeking to explain firm relocation. For the latter, the major explanatory factors are distance, which is the cost of relocation, as well as industry-relevant factors, such as density of digital firms. Therefore I conclude that agglomeration effects do in fact matter for digital firms, as firms tend to stay in the region where benefits arising from agglomeration can potentially be accessed.

For digital firm birth in Germany, results are in line with the theoretical prediction of Hypothesis 1 with firms displaying a preference towards cities and van Oort and Atzema (2004) and Trippl et al. (2009). By employing pooled as well as fixed-effects models, the results indicate that high levels of agglomeration benefits are conducive to firm birth in general. Nevertheless, an above average growth that increases the benefits is not conducive to new firms in a given year. However, results also show advantages for digital firm birth arise mostly from industry-specific benefits such as other digital firms. Therefore policy makers should focus on enhancing industry-specific factors. Additionally, local knowledge available through universities and research institutes is conducive to new firm formations (Hypothesis 2), although research institutes play a minor role.

In terms of relocation, Hypothesis 3 is that counties with higher agglomeration benefits such as a specialized high-tech labor markets and the potential for IT-specific knowledge spillovers attract more relocations. The analysis shows that relocation inflows are higher in cities with high digital firm densities. The results suggest that firms expect that negative effects of colocation such as congestion, competition and higher prices do not outweigh the benefits and advantages derived from colocation of similar firms such as access to industry-specific knowledge and a specialized labor pool.

Further, the fourth hypothesis was that most relocations are between geographically proximate or contiguous counties where moving costs are low while access to locally bound factors such as local customers, suppliers, or networks remains relatively low-cost.

Previous research indicates that young digital firms are highly mobile as they are less exposed to incur sunk costs in the relocation process (Esteve-Pérez et al., 2018; Sleutjes & Beckers, 2013). With the usage of the gravity model, this study provides evidence that relocation costs play a significant role for digital firms. Thus, I conclude that digital firms are not as footloose even in their youth as one might expect. The distance between firm birth location and relocation destination is the predominant explanatory factor for relocation while controlling for several regional characteristics in the employed fixed effects model. Therefore digital firms predominantly stay in the region where they were born initially or in the direct vicinity. This is why regional economic patterns remain very persistent over long time spans.

The results of the study are particularly relevant for policy makers trying to foster local economic growth by attracting digital firms. The study reveals that strong colocation of similar firms deepens regional specialization, since industry concentration is very conducive to attracting digital firms as well as relocating businesses. The study also shows that digital entrepreneurship is a regional phenomenon when considering the medium term. That is cities offer a breeding ground for new digital businesses and in the medium run the surrounding counties benefit once these firms decide to leave their birth towns. Moreover, these results imply that firms are willing to access networks and knowledge that is bound in the region after relocation. Thus, competitiveness-improving effects are not limited to a certain region but spread across NUTS 3 borders to a limited extend (Pijnenburg & Kholodilins's, 2014). Once a location decision has been made, it is very likely that the firm will be staying in its region of origin which leads to reinforcing spatial patterns of digital firms. Political measures that target start-up rates can have a positive, long term effect on neighboring counties.

ACKNOWLEDGMENTS

The author would like to thank Christiane Hellmanzik, Lukas Kuld, Maren Kaliske, and Sara Mitchell for their excellent suggestions and comments during the research process. Open Access funding enabled and organized by Projekt DEAL. The author received no specific funding for this study.

DATA AVAILABILITY STATEMENT

The data that support the findings will be available in North Data at https://www.northdata.de/ following an embargo from the date of publication to allow for commercialization of research findings.

ORCID

Vanessa Hellwig http://orcid.org/0000-0001-6171-0530

REFERENCES

Acs, Z. J., Braunerhjelm, P., Audretsch, D. B., & Carlsson, B. (2009). The knowledge spillover theory of entrepreneurship. *Small Business Economics*, *32*(1), 15–30. https://doi.org/10.1007/s11187-008-9157-3

Anderson, J. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, *69*, 106–116.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*(1), 170–192. https://doi.org/10.1257/000282803321455214

Arauzo Carod, J. M. (2005). Determinants of industrial location: An application for Catalan municipalities. *Papers in Regional Science*, *84*(1), 105–120. https://doi.org/10.1111/j.1435-5957.2005.00006.x

Arauzo-Carod, J. M., Liviano-Solis, D., & Manjón-Antolín, M. (2010). Empirical studies in industrial location: An assessment of their methods and results. *Journal of Regional Science*, *50*(3), 685–711. https://doi.org/10.1111/j.1467-9787.2009.00625.x

Armington, C., & Acs, Z. J. (2002). The determinants of regional variation in new firm formation. *Regional Studies*, *36*(1), 33–45. https://doi.org/10.1080/00343400120099843

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

Audretsch, D. B., Falck, O., Feldman, M. P., & Heblich, S. (2012). Local entrepreneurship in context. *Regional Studies*, *46*(3), 379–389. https://doi.org/10.1080/00343404.2010.490209

Audretsch, D. B., Keilbach, M. C., & Lehmann, E. E. (2006). *Entrepreneurship and economic growth*. Oxford University Press.

Bade, F. J., & Nerlinger, E. A. (2000). The spatial distribution of new technology-based firms: Empirical results for West Germany. *Papers in Regional Science*, *79*(2), 155–176. https://doi.org/10.1007/s101100050041

BBSR Bundesinstitut für Bau-, Stadt- und Raumforschung (ed). (2020). Indikatoren und Karten zur Raum- und Stadtentwicklung. INKAR. Bonn.

Bertoli, S., & Fernández-Huertas Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, *102*, 79–100. https://doi.org/10.1016/j.jdeveco.2012.12.001

Blum, B. S., Claro, S., Horstmann, I., & Tombe, T. (2020). The DNA of new exporters: Spin-offs and FDI at the extensive margin of trade. *American Economic Review: Insights*, *2*(3), 397–408. https://doi.org/10.1257/aeri.20190379

Bodenmann, B. R. & Axhausen, K. W. (2012). Destination choice for relocating firms: A discrete choice model for the St. Gallen region, Switzerland. *Papers in Regional Science*, *91*(2), 319–341. https://doi.org/10.1111/j.1435-5957.2011.00389.x

Brouwer, A. E., Mariotti, I., & Van Ommeren, J. N. (2004). The firm relocation decision: An empirical investigation. *The Annals of Regional Science*, *38*(2), 335–347. https://doi.org/10.1007/s00168-004-0198-5

Buczkowska, S., & de Lapparent, M. (2014). Location choices of newly created establishments: Spatial patterns at the aggregate level. *Regional Science and Urban Economics*, *48*, 68–81. https://doi.org/10.1016/j.regsciurbeco.2014.05.001

Bundesverband Deutsche Startups (BDS) (ed.) (2020): Deutscher Startup Monitor 2020. Innovation statt Krise. https://deutscherstartupmonitor.de/wp-content/uploads/2020/09/dsm_2020.pdf

Burchardi, K. B., Chaney, T., & Hassan, T. A. (2019). Migrants, ancestors, and foreign investments. *The Review of Economic Studies*, *86*(4), 1448–1486. https://doi.org/10.1093/restud/rdy044

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, *98*(4), 1707–1721. https://doi.org/10.1257/aer.98.4.1707

Combes, P. P., Duranton, G., & Gobillon, L. (2019). The costs of agglomeration: House and land prices in French cities. *The Review of Economic Studies*, *86*(4), 1556–1589. https://doi.org/10.1093/restud/rdy063

Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, *80*(6), 2543–2594. https://doi.org/10.3982/ECTA8442

Conroy, T., Deller, S., & Tsvetkova, A. (2016). Regional business climate and interstate manufacturing relocation decisions. *Regional Science and Urban Economics*, *60*, 155-168. https://doi.org/10.1016/j.regsciurbeco.2016.06.009

Destatis (Statistisches Bundesamt). (2021a). Umsatzsteuerstatistik (Voranmeldungen). https://www.genesis.destatis.de/genesis/online?operation=ergebnistabelleQualitaetSeparatAUS&levelindex=2&levelid=1633370000001&downloadname=73311-0011abreadcrumb

Destatis (Statistisches Bundesamt). (2021b). Small and medium-sized enterprises. https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Enterprises/Small-Sized-Enterprises-Medium-Sized-Enterprises/_node.html;jsessionid=FC99D6F5B8376C36E439306517CBDA44.live742

van Dijk, J., & Pellenbarg, P. H. (2000). Firm relocation decisions in the Netherlands: An ordered logit approach. *Papers in Regional Science*, *79*(2), 191–219.

van Dijk, J., & Pellenbarg, P. H. (2017). Firm migration. In D. Richardson, N. Castree, M. F. Goodchild, A. L. Kobayashi, W. Liu, & R. A. Marston (Eds.), *International encyclopedia of geography: People, the Earth, Environment and Technology* (pp. 1–12).). American Association of Geographers. https://doi.org/10.1002/9781118786352.wbieg0814

Drezner, T., & Drezner, Z. (2001). A note on applying the gravity rule to the airline hub problem. *Journal of Regional Science*, *41*(1), 67–72. https://doi.org/10.1111/0022-4146.00207

Duranton, G. & Puga, D. (2001): Nursery cities: Urban diversity, process innovation and the life cycle of products. *American Economic Review 91*(5):1454–1477 https://doi.org/10.1257/aer.91.5.1454

Duvivier, C., Polèse, M., & Apparicio, P. (2018). The location of information technology-led new economy jobs in cities: Office parks or cool neighbourhoods? *Regional Studies*, *52*(6), 756–767. https://doi.org/10.1080/00343404.2017.1322686

Egger, P. and Pfaffermayr, M. (2004). Distance, trade and FDI: A Hausman–Taylor sur approach. *Journal of Applied Econometrics*, *19*(2): 227–246. https://doi.org/10.1002/jae.721

Esteve-Pérez, S., Pieri, F., & Rodriguez, D. (2018). Age and productivity as determinants of firm survival over the industry life cycle. *Industry and Innovation*, *25*(2), 167–198. https://doi.org/10.1080/13662716.2017.1291329

Feldman, M., Francis, J., & Bercovitz, J. (2005). Creating a cluster while building a firm: Entrepreneurs and the formation of industrial clusters. *Regional Studies*, *39*(1), 129–141. https://doi.org/10.1080/0034340052000320888

Figueiredo, O., Guimarães, P., & Woodward, D. (2002). Home-field advantage: Location decisions of Portuguese entrepreneurs. *Journal of Urban Economics*, *52*(2), 341–361. https://doi.org/10.1016/S0094-1190(02)00006-2

Foord, J. (2013). The new boomtown? Creative city to Tech City in east London. *Cities*, *33*, 51–60. https://doi.org/10.1016/j.cities.2012.08.009

Fossen, F. M., & Martin, T. (2018). Entrepreneurial dynamics over space and time. *Regional Science and Urban Economics*, *70*, 204–214. https://doi.org/10.1016/j.regsciurbeco.2018.04.004

Fraunhofer Institut. (2019). Institutes and Research Establishments in Germany. https://www.fraunhofer.de/en/institutes/institutes-and-research-establishments-in-germany.html

Fritsch, M., & Aamoucke, R. (2013). Regional public research, higher education, and innovative start-ups: An empirical investigation. *Small Business Economics*, *41*(4), 865–885. https://doi.org/10.1007/s11187-013-9510-z

Fritsch, M., & Mueller, P. (2008). The effect of new business formation on regional development over time: The case of Germany. *Small Business Economics*, *30*(1), 15–29. https://doi.org/10.1007/s11187-007-9067-9

Fritsch, M., & Storey, D. J. (2014). Entrepreneurship in a regional context: Historical roots, recent developments and future challenges. *Regional Studies*, *48*(6), 939–954. https://doi.org/10.1080/00343404.2014.892574

Fritsch, M., & Wyrwich, M. (2014). The long persistence of regional levels of entrepreneurship: Germany, 1925–2005. *Regional Studies*, *48*(6), 955–973. https://doi.org/10.1080/00343404.2013.816414

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, *97*(2), 498–520. https://doi.org/10.1162/REST_a_00456

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics & Management Strategy*, *18*(3), 623–663. https://doi.org/10.1111/j.1530-9134.2009.00225.x

Glaeser, E. L., Rosenthal, S. S., & Strange, W. C. (2010). Urban economics and entrepreneurship. *Journal of Urban Economics*, *67*(1), 1–14. https://doi.org/10.1016/j.jue.2009.10.005

Guzman, J. (2019). Go west young firm: Agglomeration and embeddedness in startup migrations to Silicon Valley. Columbia Business School Research Paper No. 18–49. https://doi.org/10.2139/ssrn.3175328

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 4, pp. 131–195). Elsevier. https://doi.org/10.1016/B978-0-444-54314-1.00003-3

Helmholz Gesellschaft. (2019). Our Research Centers. https://www.helmholtz.de/en/about-us/helmholtz-centers/

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, *123*(2), 441–487. https://doi.org/10.1162/qjec.2008.123.2.441

Hochschulkompass. (2020). Download von Hochschullisten. https://www.hochschulkompass.de/hochschulen/downloads.html

Holl, A. (2004). Startups and relocations: Manufacturing plant location in Portugal. *Papers in Regional Science*, *83*(4), 649–668. https://doi.org/10.1111/j.1435-5597.2004.tb01932.x

Jacobs, J. (1969). *The economies of cities*. Random House.

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, *32*(13), 1745–1755. https://doi.org/10.1080/000368400421093

Klick, L., & Schaffner, S. (2021). Regional real estate price indices for Germany, 2008–2019: RWI–GEO–REDX. *Jahrbücher für Nationalökonomie und Statistik*, *241*(1), 119–129. https://doi.org/10.1515/jbnst-2020-0001

Knoben, J. (2011). The geographic distance of relocation search: An extended resource-based perspective. *Economic Geography*, *87*(4), 371–392. https://doi.org/10.1111/j.1944-8287.2011.01123.x

Kohler HP (1997) The effect of hedonic migration decisions and region-specific Amenities on industrial location: Could silicon valley be in South Dakota? *Journal of Regional Science 37*: 379–394 https://doi.org/10.1111/0022-4146.00060

Kronenberg, K. (2013). Firm relocations in the Netherlands: Why do firms move, and where do they go? *Papers in Regional Science*, *92*(4), 691–713. https://doi.org/10.1111/j.1435-5957.2012.00443.x

Krugman, P. (1991). *Geography and trade*. The MIT Press.

Lee, I. H. I. (2020). Startups, relocation, and firm performance: A transaction cost economics perspective. *Small Business Economics*, *58*(3), 1–20. https://doi.org/10.1007/s11187-020-00406-z

Lee, Y. (2008). Geographic redistribution of US manufacturing and the role of state development policy. *Journal of Urban Economics*, *64*(2), 436–450. https://doi.org/10.1016/j.jue.2008.04.001

Leibniz Association. (2019). Leibniz Institutes. https://www.leibniz-gemeinschaft.de/en/institutes/leibniz-institutes-all-lists.html

Leone, R. A., & Struyk, R. (1976). The incubator hypothesis: Evidence from five SMSAs. *Urban Studies*, *13*(3), 325–331.

Mameli, F., Faggian, A., & Mccann, P. (2014). Estimation of local employment growth: Do sectoral aggregation and industry definition matter? *Regional Studies*, *48*(11), 1813–1828. https://doi.org/10.1080/00343404.2012.756578

Manjón-Antolín, M. C., & Arauzo-Carod, J. M. (2011). Locations and relocations: Determinants, modelling, and interrelations. *The Annals of Regional Science*, *47*(1), 131–146. https://doi.org/10.1007/s00168-009-0360-1

Max-Planck-Institute. (2019). Max Planck Institutes and Experts. https://www.mpg.de/institutes

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

Michelacci, C., & Silva, O. (2007). Why so many local entrepreneurs? *Review of Economics and Statistics*, *89*(4), 615–633. https://doi.org/10.1162/rest.89.4.615

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research*, *48*, 1–10. https://doi.org/10.1016/j.annals.2014.05.008

Nguyen, C. Y., Sano, K., Tran, T. V., & Doan, T. T. (2013). Firm relocation patterns incorporating spatial interactions. *The Annals of Regional Science*, *50*(3), 685–703. https://doi.org/10.1007/s00168-012-0523-3

North Data. (2019). Data Services. https://www.northdata.com/_data

Oakey, R., Kipling, M., & Wildgust, S. (2001). Clustering among firms in the non-broadcast visual communications (NBVC) sector. *Regional Studies*, *35*(5), 401–414. https://doi.org/10.1080/00343400125033

Olivero, M. P., & Yotov, Y. V. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics/Revue canadienne d'économique*, *45*(1), 64–92. https://doi.org/10.1111/j.1540-5982.2011.01687.x

van Oort, F. G., & Atzema, O.A.L.C. (2004). On the conceptualization of agglomeration economies: The case of new firm formation in the Dutch ICT sector. *The Annals of Regional Science*, *38*(2), 263–290. https://doi.org/10.1007/s00168-004-0195-8

Organisationen und Einrichtungen in Forschung und Wissenschaft (OEFW). (2016). Bundesbericht Forschung und Innovation 2016. Ergänzungsband II.

Ozmen-Ertekin, D., Ozbay, K., & Holguin-Veras, J. (2007). Role of transportation accessibility in attracting new businesses to New Jersey. *Journal of Urban Planning and Development*, *133*(2), 138–149. https://doi.org/10.1061/(ASCE)0733-9488(2007)133:2(138)

Pan, Y., Conroy, T., Tsvetkova, A., & Kures, M. (2020). Incentives and firm migration: An interstate comparison approach. *Economic Development Quarterly*, *34*(2), 140–153. https://doi.org/10.1177/0891242420917756

Pijnenburg, K., & Kholodilin, K. A. (2014). Do regions with entrepreneurial neighbours perform better? A spatial econometric approach for German regions. *Regional Studies*, *48*(5), 866–882. https://doi.org/10.1080/00343404.2012.697143

Porter, M. (1990). *The competitive advantage of nations*. Macmillan.

Portes, R., & Rey, H. (2005). The determinants of cross-border equity flows. *Journal of International Economics*, *65*(2), 269–296. https://doi.org/10.1016/j.jinteco.2004.05.002

Redding, S. J., Sturm, D. M., & Wolf, N. (2011). History and industry location: Evidence from German airports. *Review of Economics and Statistics*, *93*(3), 814–831. https://doi.org/10.1162/REST_a_00096

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373-389. https://doi.org/10.1016/j.jue.2008.02.006

Rossi, F., & Dej, M. (2019). Where do firms relocate? Location optimization within and between Polish metropolitan areas. *The Annals of Regional Science*, *64*, 1-26. https://doi.org/10.1007/s00168-019-00948-5

Saarenketo, S., Puumalainen, K., Kyläheiko, K., & Kuivalainen, O. (2008). Linking knowledge and internationalization in small and medium-sized enterprises in the ICT sector. *Technovation*, *28*(9), 591–601. https://doi.org/10.1016/j.technovation.2007.12.003

Santos Silva, J. S., & Tenreyro, S. (2021). The Log of Gravity At 15 (No. 0121). School of Economics, University of Surrey.

Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4):641–658. https://doi.org/10.1162/rest.88.4.641

Sleutjes, B., & Beckers, P. (2013). Exploring the role of the neighbourhood in firm relocation: Differences between stayers and movers. *Journal of Housing and the Built Environment*, *28*(3), 417-433. https://doi.org/10.1007/s10901-012-9317-2

Stam, E. (2007) Why butterflies Don't leave: Locational behavior of entrepreneurial firms. *Economic Geography*, *3*: 27–50. https://doi.org/10.1111/j.1944-8287.2007.tb00332.x

Trippl, M., Tödtling, F., & Lengauer, L. (2009). Knowledge sourcing beyond buzz and pipelines: Evidence from the Vienna software sector. *Economic Geography*, *85*(4), 443–462. https://doi.org/10.1111/j.1944-8287.2009.01047.x

Weber, T., Bertschek, I., Ohnemus, J., & Ebert, M. (2018). Monitoring-report wirtschaft digital 2018. Technicalreport, ZEW-Gutachten und Forschungsberichte.
Weterings, A. (2014). What makes firms leave the neighbourhood? *Urban Studies*, *51*(8), 1613–1633. https://doi.org/10.1177/0042098013498624
Weterings, A., & Knoben, J. (2013). Footloose: An analysis of the drivers of firm relocations over different distances. *Papers in Regional Science*, *92*(4), 791–809. https://doi.org/10.1111/j.1435-5957.2012.00440.x
van Wissen, L. (2000) A micro-simulation model of firms: Applications of concepts of the demography of the firm. *Papers in Regional Science 79*(2):111–134 https://doi.org/10.1111/j.1435-5597.2000.tb00764.x
Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Cengage learning.
Yi, Y. (2018). Firm relocation and age-dependent reliance on agglomeration externalities. *The Annals of Regional Science*, *61*(2), 439–456. https://doi.org/10.1007/s00168-018-0875-4
Yotov, Y. V., Piermartini, R., Monteiro, J. A., & Larch, M. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization. https://doi.org/10.18356/57a768e5-en

How to cite this article: Hellwig, V. (2023). Digital gravity? Firm birth and relocation patterns of young digital firms in Germany. *Journal of Regional Science*, 63, 340–378. https://doi.org/10.1111/jors.12624

APPENDIX A:

TABLE A1 Variable description and sources of regional characteristics used in regressions

Variable	Description	Source
Population density	Population per km^2	BBSR
Industry ratio	Employees at the place of work in industry (WZ 2008) per 100 inhabitants of working age	BBSR
Service ratio	Employees at work in the service sector (WZ 2008) per 100 inhabitants of working age	BBSR
Price index	Difference of the counties mean rent price (in percent) to the German mean price	RWI 2020
Price index change	Change in difference of the counties mean rent price (in percent) to the German mean rent price	RWI 2020
Gross income	Gross monthly earnings of employees in euros	BBSR
Firms per 1000 inhabitants	INKAR; North Data	North Data (2019), BBSR
Universities	Number of universities and universities of applied science in counties	Hochschulkompass (2020)
Research Institutes	Number of publicly funded research institutes in district and institutes of the four biggest research organizations in Germany	Fraunhofer, Helmholtz, Leibniz, Max-Planck-Institutes, Forschungseinrichtungen des Bundes und der Länder
Metropole	Cities with more than 500 000 inhabitants in 2017: Hamburg, Berlin, Munich, Stuttgart, Cologne, Frankfurt am Main, Düsseldorf, Dortmund, Essen	BBSR
Distance	Linear distance in kilometer (as the crow flies)	Own calculation with R geocode command

TABLE A2 Correlation table on regional characteristics

	Bilateral flow	Industrial ratio	Service ratio	Price index	Price index change	Firms per 1000 inhabitants	Research institutes	Universities	Firm birth (ln)	Gross income (ln)	Population density (ln)
Bilateral flow	1	−0.02	0.05	0.08	0.04	0.10	0.10	0.11	0.09	0.06	0.06
Industrial ratio	−0.02	1	0.12	−0.06	0.14	−0.06	−0.17	−0.11	−0.19	0.39	0.01
Service ratio	0.05	0.12	1	0.44	0.30	0.57	0.42	0.44	0.32	0.45	0.66
Price index	0.08	−0.06	0.44	1	0.31	0.70	0.41	0.41	0.65	0.58	0.52
Price index change	0.04	0.14	0.30	0.31	1	0.49	0.20	0.19	0.13	0.48	0.17
Firms per 1000 inhabitants	0.10	−0.06	0.57	0.70	0.49	1	0.50	0.48	0.61	0.63	0.51
research institutes	0.10	−0.17	0.42	0.41	0.20	0.50	1	0.80	0.57	0.25	0.45
Universities	0.11	−0.11	0.44	0.41	0.19	0.48	0.80	1	0.63	0.31	0.51
Firm birth (ln)	0.09	−0.19	0.32	0.65	0.13	0.61	0.57	0.63	1	0.44	0.56
Gross income (ln)	0.06	0.39	0.45	0.58	0.48	0.63	0.25	0.31	0.44	1	0.54
Population density (ln)	0.06	0.01	0.66	0.52	0.17	0.51	0.45	0.51	0.56	0.54	1

TABLE A3 Variance inflation factor for regression 1.4 (pooled OLS total firm birth)

Population density	**2.769**
Industrial ratio	1.843
Service ratio	2.626
Gross income	3.798
Firms per 1000 inhabitants	4.223
Price index	2.232
Price Index change	1.731
Metropole	2.080
Neighbor is metro	1.209
lag(firmbirth)	3.263
Universities	3.918
Research Institutes	3.068

TABLE A4 Gravity model (2) robustness

	Dependent variable: M_{ijt}									
	PPML without Munich -Pair		Negative binomial		PPML only positive		Negative binomial only positive		OLS only positive	
Population density (ln), origin	1.63213	(1.3885)	2.83420*	(1.42778)	0.02343	(0.83008)	0.07375	(0.97812)	0.08120	(0.41348)
Population density (ln), Dest.	0.71251	(0.98430)	−0.48506	(1.37770)	0.05778	(0.79220)	0.10875	(0.93831)	0.27016	(0.39645)
Industrial ratio, origin	0.01880	(0.01519)	0.01525	(0.02163)	−0.01548	(0.01075)	−0.01530	(0.01451)	−0.00856	(0.00546)
Industrial ratio, destination	0.00919	(0.01583)	0.00304	(0.02129)	0.00439	(0.01091)	0.00406	(0.01462)	0.00501	(0.00582)
Service ratio, origin	0.02295	(0.01209)	0.01360	(0.01674)	0.00729	(0.00868)	0.00710	(0.01125)	0.00473	(0.00457)
Service ratio, destination	0.00711	(0.01204)	−0.00410	(0.01664)	−0.00175	(0.00959)	−0.00163	(0.01134)	−0.00010	(0.00471)
Gross income (ln), origin	−1.45613	(0.82506)	−2.34002*	(1.14170)	−0.45356	(0.57452)	-0.44300	(0.77407)	−0.25914	(0.32720)
Gross income (ln), destination	0.70904	(0.81127)	0.58105	(1.12996)	0.34709	(0.59761)	0.33541	(0.76130)	−0.01262	(0.31153)
Price index, origin	−0.00139	(0.01023)	0.01339	(0.01518)	−0.02030*	(0.01005)	−0.01980*	(0.00876)	−0.00843	(0.00477)
Price index, destination	0.02161*	(0.01031)	0.01787	(0.01486)	−0.00237	(0.01068)	-0.00297	(0.00881)	−0.00381	(0.00466)
Price index change, origin	−0.00165	(0.00899)	-0.01426	(0.01336)	0.01762	(0.00970)	0.01699*	(0.00743)	0.00718	(0.00441)
Price index change, destination	−0.02168*	(0.00925)	-0.01845	(0.01311)	0.00213	(0.01009)	0.00257	(0.00748)	0.00298	(0.00431)
Firm birth (ln), origin	−0.08206	(0.04667)	−0.02441	(0.06051)	−0.00115	(0.02821)	−0.00135	(0.04651)	−0.00409	(0.01575)
Firm birth (ln), destination	0.02143	(0.04442)	−0.00645	(0.05785)	0.00596	(0.02775)	0.00595	(0.04428)	−0.00413	(0.01485)
Firms per 1000 inhabitants, origin	−0.08646	(0.08722)	−0.08917	(0.12763)	0.01308	(0.08749)	0.00675	(0.07311)	−0.01731	(0.04150)
Firms per 1000 inhabitants, Dest.	0.25127**	(0.08703)	0.33920**	(0.12973)	0.10950	(0.10334)	0.09478	(0.07114)	0.02160	(0.04525)
Research Institutes, origin	−0.01354	(0.04253)	−0.08155	(0.07059)	−0.02616	(0.03787)	−0.02441	(0.03426)	−0.01744	(0.02027)
Research Institutes, destination	−0.06343	(0.04540)	−0.13990	(0.07578)	0.00495	(0.04893)	0.00543	(0.03692)	0.00529	(0.02327)
Universities, origin	0.06766*	(0.03000)	0.05790	(0.05153)	0.06622*	(0.02720)	0.06647**	(0.02546)	0.03583*	(0.01508)
Universities, destination	0.00625	(0.03309)	0.02285	(0.05370)	0.01333	(0.03513)	0.01541	(0.02600)	0.00503	(0.01690)

TABLE A4 (Continued)

	Dependent variable: M_{ijt}									
	PPML without Munich -Pair		Negative binomial		PPML only positive		Negative binomial only positive		OLS only positive	
Neighbor county	1.19925***	(0.04095)	1.40058***	(0.05799)	0.28267***	(0.02196)	0.28259***	(0.03116)	0.19787***	(0.01400)
Distance (ln)	−1.20744***	(0.01584)	−1.40641***	(0.02122)	−0.26496***	(0.01533)	−0.25571***	(0.01253)	−0.11935***	(0.00617)
Num. obs.	1,432,800		1,432,818		10,108		10,108		10,108	

Note: Significant levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; time, origin and destination fixed effects included, standard errors in parentheses; dependent variable as log-link.

TABLE A5 Gravity Model (3) (difference approach) robustness

	Dependent variable: M_{ijt}				
	PPML without Munich-Pair	Negative binomial	PPML only positive	Negative binomial only positive	OLS only positive
Population density	0.00011 (0.00027)	0.00023 (0.00045)	−0.00032 (0.00032)	−0.00029 (0.00022)	−0.00004 (0.00001)
Industrial ratio	−0.00108 (0.01077)	0.00300 (0.01524)	0.00530 (0.00677)	0.00545 (0.01028)	0.00591 (0.00005)
Service ratio	−0.00582 (0.00778)	0.00104 (0.01114)	−0.00754 (0.00636)	−0.00697 (0.00726)	−0.00292 (0.00004)
Gross income	0.00026 (0.00004)	0.00032 (0.00030)	0.00014 (0.00015)	0.00013 (0.00020)	0.00004 (0.00000)
Price index	0.01850 (0.00866)*	0.01122 (0.01143)	0.00762 (0.00879)	0.00734 (0.00718)	0.00201 (0.00004)*
Price index change	−0.01624 (0.00778)*	−0.01086 (0.01015)	−0.00629 (0.00835)	−0.00599 (0.00624)	−0.00176 (0.00004)*
Firm birth	−0.00029 (0.00032)	−0.00027 (0.00055)	0.00002 (0.00034)	0.00002 (0.00025)	0.00004 (0.00001)
Firms per 1000 inhabitants	0.17277 (0.06527)**	0.15518 (0.09082)	0.05985 (0.08113)	0.05381 (0.05042)	0.01784 (0.00069)
Research Institutes	−0.02444 (0.03320)	−0.02914 (0.05313)	0.02234 (0.03264)	0.02103 (0.02482)	0.01261 (0.00071)
Universities	−0.01512 (0.02654)	−0.00456 (0.04770)	−0.01982 (0.02329)	−0.01954 (0.02254)	−0.01658 (0.00065)
Neighbor county	1.19930 (0.04097)***	1.40335 (0.05831)***	0.28140 (0.02209)***	0.28118 (0.03117)***	0.19704 (0.00263)***
Distance (ln)	−1.20746 (0.01584)***	−1.40692 (0.02131)***	−0.26403 (0.01550)***	−0.25460 (0.01253)***	−0.11926 (0.00020)***
Num. obs.	1432800	1432818	10108	10108	10108

Note: Significant levels at at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; time, origin and destination fixed effects included, standard errors in parentheses; dependent variable as log-link.

TABLE A6 Regression results model (2) gravity model (PPML): multilateral resistance county-time FEs

Dependent variable	M_{ijt}
Neighbor County	1.21401 (0.04061)***
Distance (ln)	−1.16234 (0.01617)***
Time-county fixed effect	Yes
Num. obs.	770,120

Note: Significant levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

TABLE A7 Binary regressions push and pull factors

Push factors				Pull factors		
	OLS	*logistic*	*Probit*	OLS	*logistic*	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Firmage	−0.003 (0.0001)***	−0.117 (0.004)***	−0.050 (0.002)***	0.001 (0.0001)***	0.025 (0.003)***	0.011 (0.001)***
Population density	−0.00000 (0.00000)	−0.0001 (0.0002)	−0.00003 (0.0001)	−0.00001 (0.00000)*	−0.0003 (0.0002)*	−0.0001 (0.0001)*
Industrial ratio	0.0002 (0.0002)	0.002 (0.006)	0.001 (0.003)	0.0003 (0.0002)	0.009 (0.008)	0.004 (0.003)
Service ratio	−0.0003 (0.0003)	−0.011 (0.008)	−0.005 (0.004)	0.0001 (0.0003)	0.004 (0.010)	0.001 (0.004)
Gross income	−0.003 (0.008)	−0.048 (0.232)	−0.030 (0.102)	0.004 (0.007)	0.111 (0.262)	0.052 (0.112)
Stock of firms	0.00000 (0.00000)**	0.00002 (0.00003)	0.00001 (0.00001)	−0.00000 (0.00000)	−0.00002 (0.00003)	−0.00001 (0.00001)
Firms per 1000 inhabitants	−0.003 (0.002)	−0.003 (0.055)	−0.006 (0.025)	0.004 (0.002)*	0.112 (0.066)*	0.047 (0.029)*
Price Index	0.0002 (0.0002)	0.001 (0.007)	0.001 (0.003)	0.0003 (0.0002)	0.011 (0.008)	0.005 (0.004)
Price Index change	−0.0002 (0.0002)	−0.001 (0.006)	−0.001 (0.003)	−0.0003 (0.0002)*	−0.011 (0.007)	−0.005 (0.003)
Research Institutes	−0.001 (0.001)	−0.034 (0.032)	−0.015 (0.014)	−0.001 (0.001)	−0.025 (0.037)	−0.012 (0.016)
Universities	−0.0005 (0.001)	−0.00003 (0.029)	−0.0002 (0.013)	0.0001 (0.001)	0.007 (0.034)	0.002 (0.014)
Firm birth (lag)	−0.00001 (0.00001)	−0.0002 (0.0003)	−0.0001 (0.0001)	−0.00001 (0.00001)	−0.0002 (0.0004)	−0.0001 (0.0002)
Metropole	0.006 (0.004)	0.142 (0.137)	0.060 (0.058)	0.006 (0.003)*	0.289 (0.146)**	0.121 (0.061)**
Neighbor is metro	−0.005 (0.012)	−0.085 (0.317)	−0.039 (0.141)	−0.018 (0.011)	−0.554 (0.332)*	−0.244 (0.146)*
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	552,384	552,384	552,384	552,384	552,384	552,384
Adjusted R^2	0.006			0.005		
Log Likelihood		−75,148.870	−75,157.100		−67,201.860	−67,199.900

Note: Significant levels at $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$, cluster robust standard errors in parentheses.

TABLE A8 Gravity model (2) robustness—subsample only ICT firms

	PPML	NB	OLS OP	PPML OP	NB OP
Population density (ln), origin	0.45225 (1.93014)	3.11375 (2.28018)	0.18712 (0.81547)	−0.85236 (1.38927)	−0.81696 (1.38434)
Population density (ln), destination	−2.96359 (1.98166)	0.05174 (2.35171)	−0.97619 (0.69926)	−2.15182 (1.15194)	−2.13519 (1.14870)
Industrial ratio, origin	−0.00094 (0.02507)	−0.00517 (0.02936)	−0.01042 (0.00807)	−0.01219 (0.01450)	−0.01224 (0.01443)
Industrial ratio, destination	−0.01794 (0.02467)	−0.02089 (0.02818)	−0.00048 (0.00940)	0.00648 (0.01507)	0.00623 (0.01502)
Service ratio, origin	0.01564 (0.01951)	0.02259 (0.02187)	0.00433 (0.00783)	0.00166 (0.01287)	0.00169 (0.01282)
Service ratio, destination	−0.03691 (0.02126)	−0.01703 (0.02405)	−0.00423 (0.00756)	−0.00326 (0.01265)	−0.00338 (0.01262)
Gross income (ln), origin	2.20674 (1.56482)	−1.18291 (1.78579)	0.85534 (0.55553)	1.45821 (0.90124)	1.45335 (0.89835)
Gross income (ln), destination	−2.35208 (1.51402)	−2.79974 (1.75435)	0.49265 (0.53697)	0.30016 (0.86886)	0.30794 (0.86476)
Price index, origin	−0.05542 (0.01994)**	−0.02861 (0.02478)	−0.01773 (0.00849)*	−0.02699 (0.01424)	−0.02693 (0.01418)
Price index, destination	0.00743 (0.02021)	0.03359 (0.02450)	−0.00775 (0.00874)	−0.00985 (0.01422)	−0.01000 (0.01420)
Price index change, origin	0.04724 (0.01735)**	0.02281 (0.02172)	0.01469 (0.00779)	0.02419 (0.01325)	0.02407 (0.01319)
Price index change, destination	−0.01082 (0.01766)	−0.03464 (0.02132)	0.00625 (0.00793)	0.00778 (0.01286)	0.00791 (0.01285)
Firmbirth ICT (ln), origin	−0.17063 (0.05925)**	−0.19524 (0.06770)**	−0.03447 (0.01854)	−0.08933 (0.02839)**	−0.08865 (0.02830)**
Firmbirth ICT (ln), destination	−0.05265 (0.05720)	−0.03699 (0.06506)	−0.00399 (0.01927)	0.00554 (0.03061)	0.00544 (0.03046)
Firms per 1000 inhabitants, origin	−0.35879 (0.26393)	−0.33030 (0.32578)	−0.02109 (0.12798)	0.18684 (0.20201)	0.18257 (0.20182)
Firms per 1000 inhabitants, destination	0.67789 (0.26916)*	0.90800 (0.33666)**	0.05192 (0.12507)	0.27037 (0.20564)	0.26430 (0.20554)
Research Institutes, origin	−0.06812 (0.06999)	−0.21143 (0.09443)*	−0.00152 (0.02982)	−0.00399 (0.04769)	−0.00408 (0.04744)
Research Institutes, destination	−0.11007 (0.07554)	−0.25765 (0.09871)**	0.02129 (0.03279)	0.01338 (0.05508)	0.01305 (0.05478)
Universities, origin	−0.05135 (0.06088)	0.06329 (0.08726)	0.00603 (0.02586)	−0.01576 (0.03450)	−0.01528 (0.03443)
Universities, destination	−0.06391 (0.06644)	−0.02075 (0.09226)	0.01330 (0.02584)	0.01424 (0.03859)	0.01462 (0.03853)

(Continues)

TABLE A8 (Continued)

	PPML	NB	OLS OP	PPML OP	NB OP
Neighbor County	1.06446 (0.06724)***	1.19518 (0.07392)***	0.11101 (0.01930)***	0.13765 (0.02835)***	0.13805 (0.02825)***
Distance (ln)	−1.25790 (0.02828)***	−1.44787 (0.03163)***	−0.09906 (0.01008)***	−0.21064 (0.01746)***	−0.20901 (0.01749)***
Num. obs.	820,890	820,890	3537	3537	3537

Note: Significant levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$, year, origin and destination fixed effects included; standard errors in parentheses; variables coded as differences between origin and destination; firm birth and firms per 1000 inhabitants refer to digital firms; dependent variable as log-link. The subsample contains only ICT firms selected via NACE code.

TABLE A9 Gravity model (3) (difference approach) robustness—subsample only ICT firms

	PPML	NB	OLS OP	PPML OP	NB OP
Population density	−0.00057 (0.00055)	−0.00078 (0.00072)	−0.00046 (0.00026)	−0.00089 (0.00044)*	−0.00088 (0.00044)*
Industrial ratio	−0.00539 (0.01769)	−0.00536 (0.02093)	0.00218 (0.00575)	0.00387 (0.00940)	0.00384 (0.00936)
Service ratio	−0.02315 (0.01379)	−0.01935 (0.01578)	−0.00578 (0.00508)	−0.00688 (0.00922)	−0.00682 (0.00917)
Gross income	−0.00093 (0.00037)*	−0.00060 (0.00042)	0.00004 (0.00013)	−0.00005 (0.00021)	−0.00005 (0.00021)
Price index	0.02802 (0.01548)	0.02856 (0.01783)	0.00175 (0.00585)	0.00300 (0.00995)	0.00292 (0.00989)
Price Index change	−0.02508 (0.01374)	−0.02570 (0.01579)	−0.00108 (0.00544)	−0.00285 (0.00951)	−0.00274 (0.00946)
Firm birth	0.00084 (0.00100)	0.00036 (0.00139)	0.00040 (0.00056)	0.00142 (0.00099)	0.00141 (0.00098)
Firms per 1000 inhabitants	0.58614 (0.20703)**	0.71947 (0.25372)**	0.03783 (0.10024)	0.05266 (0.18772)	0.04983 (0.18659)
Research Institutes	−0.02493 (0.05147)	−0.01936 (0.07030)	0.01151 (0.02285)	0.01459 (0.04223)	0.01421 (0.04187)
Universities	0.00533 (0.04681)	−0.02566 (0.06261)	0.01019 (0.01850)	0.02640 (0.02498)	0.02616 (0.02493)
Neighbor county	1.06457 (0.06736)***	1.19435 (0.07390)***	0.11181 (0.01929)***	0.14177 (0.02831)***	0.14203 (0.02821)***
Distance (ln)	−1.25799 (0.02835)***	−1.44694 (0.03163)***	−0.09842 (0.01001)***	−0.20645 (0.01790)***	−0.20497 (0.01789)***
Num. obs.	820,890	820,890	3537	3537	3537

Note: Significant levels at *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$, year, origin and destination fixed effects included; standard errors in parentheses; variables coded as differences between origin and destination; firm birth and firms per 1000 inhabitants refer to digital firms; dependent variable as log-link. The subsample contains only ICT firms selected via NACE code.

TABLE A10 Gravity model (2) robustness—NUTS 2 aggregate

	OLS	PPML	NB	OLS OP	PPML OP	NB OP
Population density (ln), origin	−0.00499 (0.46466)	−0.76959 (2.68611)	−0.32974 (2.63817)	−2.06923 (1.59631)	−2.54100 (2.29354)	−2.38491 (2.18228)
Population density (ln), destination	0.16195 (0.46319)	1.96055 (2.58177)	3.55779 (2.49504)	1.18215 (1.61529)	0.50148 (2.22609)	0.75347 (2.12256)
Industrial ratio, origin	0.00075 (0.00781)	0.01130 (0.03929)	0.00717 (0.04025)	−0.01894 (0.02425)	−0.00335 (0.03122)	−0.00694 (0.03014)
Industrial ratio, destination	−0.00468 (0.00793)	0.00407 (0.03797)	0.00121 (0.03780)	0.01950 (0.02242)	0.02930 (0.03020)	0.02877 (0.02901)
Service ratio, origin	−0.00351 (0.01047)	0.00548 (0.05473)	0.04261 (0.05441)	−0.00670 (0.03363)	−0.05766 (0.04622)	−0.04401 (0.04440)
Service ratio, destination	0.00143 (0.01053)	0.03576 (0.05310)	0.04383 (0.05309)	0.01014 (0.03312)	0.00950 (0.04554)	0.01179 (0.04358)
Gross income (ln), origin	−1.43326 (0.41605)***	−3.93634 (2.02906)	−5.42891 (2.09358)**	−3.42787 (1.26540)**	−3.80443 (1.68390)*	−4.19796 (1.63587)*
Gross income (ln), destination	−1.05697 (0.41537)*	0.00903 (2.02828)	−0.52292 (2.04743)	0.70540 (1.32529)	−0.66257 (1.75547)	−0.87978 (1.69538)
Price index, origin	−0.01080 (0.00654)	0.00346 (0.02036)	−0.00776 (0.02142)	−0.02134 (0.01475)	−0.01633 (0.01858)	-0.02130 (0.01801)
Price index, destination	−0.00913 (0.00657)	−0.00655 (0.02093)	−0.00648 (0.02130)	−0.01750 (0.01530)	−0.02452 (0.01951)	−0.02427 (0.01886)
Price index change, origin	0.00969 (0.00631)	0.00015 (0.02148)	0.01904 (0.02233)	0.02273 (0.01553)	0.01036 (0.01965)	0.01747 (0.01901)
Price index change, destination	0.00773 (0.00622)	−0.00884 (0.02257)	−0.00405 (0.02233)	−0.00064 (0.01594)	−0.00214 (0.02081)	-0.00058 (0.02012)
Firmbirth (ln), origin	−0.02450 (0.03729)	0.09521 (0.20348)	−0.10791 (0.20913)	0.12112 (0.11539)	0.31933 (0.16293)	0.26120 (0.15754)
Firmbirth (ln), destination	0.05162 (0.03755)	0.40652 (0.19797)*	0.38658 (0.19940)	0.29857 (0.12178)*	0.40437 (0.16959)*	0.35582 (0.16160)*
Firms per 1000 inhabitants, origin	0.19808 (0.06525)**	0.09573 (0.22292)	−0.13053 (0.23002)	0.11375 (0.14808)	0.40739 (0.19203)*	0.31195 (0.18529)
Firms per 1000 inhabitants, destination	0.10075 (0.06559)	0.01102 (0.22888)	−0.19011 (0.23190)	0.26824 (0.15585)	0.37976 (0.20351)	0.34314 (0.19700)
Research Institutes, origin	0.02181 (0.01180)	−0.04041 (0.03880)	−0.02375 (0.03816)	0.01132 (0.02653)	−0.02360 (0.03424)	−0.02103 (0.03321)
Research Institutes, destination	0.01547 (0.01144)	−0.02870 (0.03652)	−0.03115 (0.03788)	−0.01964 (0.02617)	−0.01476 (0.03056)	−0.01450 (0.03027)

TABLE A10 (Continued)

	OLS	PPML	NB	OLS OP	PPML OP	NB OP
Universities, origin	0.01778 (0.00884)*	0.00372 (0.03018)	−0.01385 (0.03093)	0.00689 (0.01915)	0.05244 (0.02456)*	0.04349 (0.02384)
Universities, destination	0.04778 (0.00877)***	0.07830 (0.03157)*	0.08383 (0.03172)**	0.02584 (0.02049)	0.08376 (0.02576)**	0.07074 (0.02482)**
Neighbor County	0.28837 (0.02829)***	0.62898 (0.06108)***	0.55653 (0.06145)***	0.31860 (0.04237)***	0.56386 (0.04543)***	0.52520 (0.04531)***
Distance (ln)	−0.31075 (0.01002)***	−0.79693 (0.03365)***	−1.01996 (0.03069)***	−0.44925 (0.01794)***	−0.47760 (0.01846)***	−0.48146 (0.01856)***
Num. obs.	12,654	11,340	11,340	3291	3291	3291

Note: Significant levels at $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$, year, origin and destination fixed effects included; standard errors in parentheses; variables coded as differences between origin and destination; firm birth and firms per 1000 inhabitants refer to digital firms; dependent variable as log-link.

TABLE A11 Gravity model (3) (difference approach) robustness—NUTS 2 aggregate

	OLS	PPML	NB 1	OLS OP	PPML OP	NB op
Population density	0.00001 (0.00029)	−0.00006 (0.00053)	−0.00016 (0.00052	0.00028 (0.00042)	0.00019 (0.00050)	0.00019 (0.00050)
Industrial ratio	−0.00369 (0.00539)	−0.00858 (0.02587)	−0.01107 (0.02643)	0.02044 (0.01529)	0.01922 (0.01963)	0.01922 (0.01963)
Service ratio	0.00165 (0.00662)	0.01063 (0.03062)	−0.00810 (0.03002)	0.01457 (0.01971)	0.03397 (0.02568)	0.03397 (0.02568)
Gross income	0.00007 (0.00011)	0.00028 (0.00055)	0.00061 (0.00054)	0.00015 (0.00033)	−0.00004 (0.00045)	−0.00004 (0.00045)
Price index	0.00082 (0.00214)	−0.00289 (0.00930)	0.00136 (0.00896)	−0.00790 (0.00583)	−0.00856 (0.00753)	−0.00856 (0.00753)
Price Index change	−0.00085 (0.00198)	−0.00304 (0.00943)	−0.00786 (0.00908)	0.00084 (0.00584)	0.00121 (0.00772)	0.00121 (0.00772)
Firm birth	0.00002 (0.00014)	−0.00024 (0.00029)	−0.00008 (0.00031)	−0.00006 (0.00024)	−0.00019 (0.00028)	−0.00019 (0.00028)
Firms per 1000 inhabitants	−0.04450 (0.04288)	0.01866 (0.13466)	0.03418 (0.13197)	0.00611 (0.08725)	−0.00912 (0.10458)	−0.00912 (0.10458)
Research Institutes	−0.00443 (0.00811)	0.00703 (0.02796)	−0.00484 (0.02749)	−0.01172 (0.01855)	0.00823 (0.02253)	0.00823 (0.02253)
Universities	0.01505 (0.00614)*	0.04858 (0.02206)*	0.05551 (0.02157)*	0.01740 (0.01363)	0.02348 (0.01754)	0.02348 (0.01754)
Neighbor county	0.28837 (0.02862)***	0.62907 (0.06131)***	0.55606 (0.06138)***	0.31860 (0.04270)***	0.51797 (0.04664)***	0.51797 (0.04664)***
Distance (ln)	−0.31075 (0.01035)***	−0.79681 (0.03369)***	−1.01876 (0.03073)***	−0.44900 (0.01844)***	−0.48345 (0.01965)***	−0.48345 (0.01965)***
Num. obs.	12,654	11,340	11,340	3291	3291	3291

Note: Significant levels at ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$, year, origin and destination fixed effects included; standard errors in parentheses; variables coded as differences between origin and destination; firm birth and firms per 1000 inhabitants refer to digital firms; dependent variable as log-link.