

Velez-Ospina Jorge A (Orcid ID: 0000-0002-5584-2358)

Hill Inge (Orcid ID: 0000-0002-2231-0187)

## **Determinants of rural creative microclustering: Evidence from web-scraped data for England**

### **Abstract**

This study aims to compare the drivers of clustering of rural and urban creative industries in England, UK. We use pre-pandemic web-scraped data from 154,618 creative industry organisations in England, and use a novel technique to identify 71 distinct rural creative ‘microclusters’ of geographically proximate creative firms. We then consider the role of place-based assets and agglomeration in the presence of microclusters at a micro-level geography and find that the determinants of microclustering are generally consistent between rural and urban areas. On that basis we argue that policies to support creative clusters may drive rural regional development.

**JEL:** R39, Z10

**Keywords:** Microclusters, rural creative industries, location, agglomeration, local development.

### **1. Introduction**

The creative industries (CIs) play an increasingly vital role in the global economy. Prior to the COVID-19 pandemic, they were the fastest-growing industry sector (UNCTAD, 2018); in the EU, in 2019 cultural enterprises had a combined turnover of over 401 billion euros (Eurostat, 2021). The sectors that make up the CIs (and related terms and framings such as ‘creative and cultural industries’ and ‘creative economy’) have multiple, contested definitions<sup>1</sup> but are generally characterised by having high levels of human creativity as inputs, containing symbolic meaning for users, and potentially containing an element of intellectual property (Throsby, 2008).

From a regional policy perspective, CIs are particularly interesting because they are characterised by high levels of agglomeration and clustering (Berg & Hassink, 2014; Bloom et al., 2020; Gong & Hassink, 2017), which can help to drive regional economic performance (Boix et al., 2015; Boix-Domenech & Soler-Marco, 2017; Crociata et al., 2018; OECD, 2019). In line with the broader literature on agglomeration and cities (Duranton & Puga, 2004), creative clusters are most often

---

<sup>1</sup> For more detail see Higgs and Cunningham (2008) and Throsby (2008). Generally speaking, creative industries and cultural industries are defined relatively similarly, while the creative economy (or, relatedly, ‘creative trident’, per Higgs and Cunningham, 2008) is a broader term that includes creative industries businesses as well as workers in creative occupations in other sectors. For reasons we discuss in detail in Section 3, we use the term creative industries, and we adopt a bespoke variation of the UK’s DCMS definition (which in its full form includes advertising and marketing; architecture; crafts; design; film and TV; IT and software; museums and libraries; performing arts and music; and publishing).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pirs.12754

associated with urban areas (e.g. Berg & Hassink, 2014; Boix et al., 2015; Coll-Martinez et al., 2019; Lazzeretti et al., 2008; Lorenzen & Frederiksen, 2008). Despite this urban reporting, there is also a growing body of evidence that rural areas can benefit from the presence of CI activities (Bell & Jayne, 2010; Hill et al., 2021; Mahon et al., 2018; Townsend et al., 2017).

Rural firms are impacted by their locations: rural areas are characterised by a lower population, business and infrastructure density, which affects access to local customers, appropriately skilled staff, and business support (Bosworth & Turner, 2018; Hill & Mole, 2022; Lee & Cowling, 2015). In light of these challenges, neo-endogenous development models in rural areas (e.g. Ray, 2001, 2006) have argued that local place-based assets have the potential to drive the development of rural areas (Naldi et al., 2021) and 'culture economies' more broadly (Argent, 2019; Phillip & Williams, 2019). The potential contribution of rural CIs to economic development has prompted calls to rethink our conceptualisation of creative clusters to include rural, as well as urban, areas (Escalona-Orcao et al., 2016; Harvey et al., 2012).

This article aims to investigate the nature of creative clustering in rural areas. It does so by adopting the concept of the 'microcluster' (Boix et al., 2015; Siepel et al., 2020), which proposes that agglomerations can manifest on a smaller scale of 50 or more businesses in a proximate area. We apply the microcluster concept to rural CIs in England and use the resulting clusters to compare the determinants of clustering in rural and urban settings, with particular attention to the role of local assets as the basis for clustering.

We identified these microclusters using a novel technique whereby we used data scraped from 154,618 CI industry websites in England, UK, which we then mapped down to the street level. We used a density-based clustering method to identify clusters<sup>2</sup> of geographically close rural creative firms. We then aggregated the number of firms in microclusters in fine, granular geographies and estimated a series of regression models to identify the determinants of firms' being based within a microcluster. We identified 71 rural microclusters, representing 38% of all rural organisations in our sample. Our analysis shows that, generally speaking, the determinants of microclustering are broadly consistent between rural and urban areas. We observe strong associations between cultural amenities and clustering, but find limited and subsector-specific association between natural amenities and clustering at the fine geographical level. We also find a weaker association between networking activities in rural areas compared to urban areas.

This article makes two contributions to the literature: first, it presents a novel way of mapping microclusters using scraped web data, which we employ on a larger geographical area (e.g. an entire nation) than has been done previously to our knowledge. Second, using the clusters derived from this analysis we show that drivers of clustering are generally consistent between rural and urban areas. We use our findings to argue that policies to support creative clusters should include rural areas, and that efforts to support formation of rural creative networks might pay dividends.

The article has the following structure. In Section 2 we provide some previous evidence. Section 3 briefly explains the data and the methodology. Section 4 presents and discusses the estimation results. Finally, in Section 5 we conclude.

---

<sup>2</sup> This type of agglomeration represents one type of industrial agglomeration. Regional specialisation is another mechanism suggested in the literature. Our measure captures regions achieving a higher concentration of a particular type of economic activity, regardless of the specific geographical concentration within that region.

## **2. Literature background: rural creative industries and creative clustering**

### **2.1 Creative clusters, rural spaces and agglomeration**

The CIs are widely recognised to be highly clustered, and companies in these sectors are bound to the locations in which they operate (Berg & Hassink, 2014; Boal & Herrero, 2017; Bloom et al., 2020; Domenech et al., 2011; Lazzeretti et al., 2008; Mateos-Garcia & Bakhshi, 2016). The large body of literature on creative clusters (see reviews in Berg & Hassink, 2014; Bloom et al., 2020; and Gong & Hassink, 2017) generally characterises creative clusters as effectively urban phenomena, in line with broader literature identifying agglomeration as largely taking place in cities (see Duranton & Puga, 2004).

An implicitly urban-focused conceptualisation of creative clusters neglects the possibility that CIs, and creative clusters specifically, may play a role in driving rural economic development (Bell & Jayne, 2010; Darchen, 2016; Harvey et al., 2012). A growing literature shows that creative businesses play a significant role in rural economies and societies (Balfour et al., 2018; Hill et al., 2021; Mahon et al., 2018; Townsend et al., 2017). Moreover, there is also evidence that creative clustering can be important in rural settings (Escalona-Orcao et al., 2016; Harvey et al., 2012; Roberts & Townsend, 2016) and that interventions to co-locate creative activity in rural areas, such as creative hubs, can play an important role in mimicking the spatial clustering seen in urban spaces (Hill, 2021; Hill et al., 2021; Merrell et al., 2021; Pratt, 2021).

If rural areas are to benefit from creative clusters, it is also important to recognise the distinctive socio-spatial context in rural areas (Bosworth & Turner, 2018; Hill & Mole, 2022) to which firms must adapt. Rural areas have a lower population density, and hence smaller local/regional numbers of consumers and choice of appropriately skilled employees (Lee & Cowling, 2015). Similarly, the lower density of businesses means less local competition but also fewer business customers, which then necessitates an earlier engagement with exporting activities out of the region and internationally (Dubois et al., 2012; Lee and Cowling, 2015; Mole et al., 2022). Rural businesses face a business support infrastructure that is less dense, poorer access to public transport (Bosworth & Turner, 2018) and inconsistent internet connectivity (Whitacre et al., 2014), all of which can impact firm performance and contribute to – or motivate – clustering in rural areas. We aim here to explore how local place-based assets, as well as agglomeration spillover effects, impact creative clustering in rural areas.

### **2.2 Place-based assets, creative clustering and agglomeration spillover effects**

There is established evidence on the importance of institutional factors (cultural and human capital) in the formation of creative firms and the location of clusters (e.g. Boix et al., 2013; Cerisola & Panzera, 2022; Cooke & Lazzeretti, 2008; Lazzeretti et al., 2012), specifically in the rural context (McGranahan et al., 2010; Naldi et al., 2015). Moreover, there is also evidence suggesting that local factors, such as public transport access, infrastructure, and access to natural spaces, make locations more attractive (Gottlieb, 1995; Naldi et al., 2021) and that these location-specific factors may play a stronger role in driving rural economic development than they do in urban areas (Naldi et al., 2021). This points to a way in which local rural assets may prove to be a basis for growth. These rural assets may serve as an alternative basis for economic development instead of more traditional endogenous growth factors in urban regions (e.g. Duranton & Puga, 2004), as well as Jacobs-style economic diversity. Location-specific factors may then provide a basis for attracting and absorbing outside knowledge (Trippi et al., 2015) that might otherwise be lacking. All of these factors are particularly

applicable in the case of the CIs, where (for instance) 'creative class' workers might seek to trade city 'buzz' for rural life (McGranahan et al., 2011; Verditch, 2010). We now discuss some of these local factors and how they might be expected to relate to creative clustering in rural (and urban) areas.

### **Neighbourhood supply of cultural amenities**

Cultural amenities such as museums, public galleries, libraries and performing arts venues have been found to have implications for both creativity and local economic development (Bakhshi et al., 2014; Cooke & Lazzarotti, 2008; Knudsen et al., 2008). As location-specific factors described above, they may also prove to be both an attractor for the rural creative class and a means of ensuring the production of arts output for the surrounding community (Wojan & Nichols, 2018). On this basis, we expect that cultural amenities will be associated with creative clustering in both rural and urban settings.

*H1: Creative clustering is likely to be associated with presence of cultural amenities in both rural and urban areas.*

### **Local social capital**

Another potential but distinct factor that could drive clustering is local social capital, proxied by the existence of community organisations and venues that facilitate informal social interactions and networking, such as community centres in a neighbourhood, meetup groups, co-working spaces and coffee shops (as used in Hoyman & Faricy, 2009). These organisations, while being location-specific factors, represent an alternate means of building and maintaining creative and cultural networks (Harvey et al., 2012), while also generating a sense of community belonging (Andres & Round, 2015). This result was found by Hoyman and Faricy (2009) to be closely associated with other means of human and intellectual capital, as well as wage growth in US metropolitan areas. Given the association between social capital and economic development (e.g. Iyer et al., 2005; Woodman, Sawyer and Griffin, 2006), we would expect this finding to be associated with creative clustering in both rural and urban settings.

*H2: Creative clustering is likely to be associated with presence of venues facilitating informal social networking in both rural and urban areas.*

### **Natural amenities**

There is a mixed set of evidence about the importance of natural amenities (or nature-based amenities) for rural economic development (Power, 2005). Deller et al. (2008) found a positive relationship between amenities and regional economic growth, and Naldi et al. (2021) found a positive relationship between natural amenities and new firm formation in Sweden. In contrast, using UK data, Agarwal et al. (2009) did not find significant relationships between a natural beauty index and the economic performance of English rural local authorities. However, given the importance placed on outdoor amenities as a quality-of-life attribute for the rural creative class (McGranahan & Wojan, 2007; McGranahan et al., 2010), we would expect natural amenities to be

associated with the presence of creative clusters. In contrast, we would not expect natural amenities to be associated with urban clusters (following, for instance, the finding in Naldi et al., 2021).

*H3: Creative clustering is likely to be associated with natural amenities in rural areas, but not urban areas.*

### **Local knowledge environment**

Many universities and colleges cooperate with local businesses to ensure their offerings meet the skills needs of their respective region. Valero and Van Reenen (2019) also show a positive spillover effect from universities to their closest neighbouring regions. These institutions may also be a source of ideas or cultural amenities (Combes, Duranton & Gobillon, 2011), although companies in clusters may have a negative or ambivalent view about the impact of universities in supporting their activities (see also Chapain et al., 2010). Despite this, there is evidence that universities can be a source of knowledge spillovers in urban as well as rural settings, although in the latter case the spillover effects may be more localised (Andersson et al., 2009). Agglomeration benefits from universities may be generated by infrastructure such as innovation parks (Rosenthal & Strange, 2020), as well as entrepreneurship from graduates (Kitagawa et al., 2022). In non-urban areas, Kitagawa et al. (2022) found that social sciences and humanities graduates are more likely to start businesses near their universities. It is not unreasonable, therefore, to presume that this could be a factor associated with creative microclustering.

*H4: Creative clustering is likely to be positively associated with proximity to universities in urban areas and rural areas.*

### **Agglomeration spillovers**

Beyond these above structural factors, there is some limited evidence about the nature of agglomeration economies in rural settings. Two crucial aspects of agglomeration economies relate to industry specialisation and diversity and, more recently, to the concepts of related diversity and unrelated diversity (Frenken et al., 2007). Related diversity in the regional context refers to the presence of sectors in a region that have related or complementary capabilities, assets or knowledge. Unrelated diversity, on the other hand, refers to industries that are not related or complementary to each other. The general argument is that knowledge spillovers depend on firms being in close cognitive proximity or relatedness (manifested by the homogeneity of capabilities, skills, and knowledge base) – that is, similar sectors are more likely to have higher knowledge spillovers. This type of proximity is assumed to generate an interactive learning environment where firms can discover, interact, learn and innovate (Boschma, 2005; Boschma, 2017; Boschma et al., 2015). Related diversity and clustering occur when different industries within a region are complementary and supportive of each other. For instance, the presence of a strong technology sector can support the growth of other industries such as finance and professional services, leading to a related diversity of industries. This related diversity creates a supportive ecosystem that enables knowledge spillovers and economies of scope, leading to increased competitiveness and innovation.

Unrelated diversity and clustering, on the other hand, occur when different industries within a region are not complementary or supportive of each other. For instance, the presence of a large manufacturing sector in a region may not support the growth of a biotechnology or financial services sector, leading to unrelated diversity. Similarly, the clustering of unrelated industries can lead to competition for resources, rather than collaboration and sharing of knowledge, and thus to reduced innovation and competitiveness. The impact of related and unrelated diversity in rural (as opposed to urban) settings is still emerging in the literature. In the context of rural areas, Naldi et al. (2021) recently found that both related and unrelated diversity were associated with new firm formation. We would therefore anticipate a similar positive relationship between both related and unrelated diversity and the presence of creative clusters.

*H5: Creative clustering is likely to be positively associated with related diversity in urban and rural areas.*

*H6: Creative clustering is likely to be positively associated with unrelated diversity in urban and rural areas.*

### **3. Methodology and data**

Identifying clusters is not a trivial task (see Bergman & Feser, 1999), with methods including qualitative identification alongside index base indicators such as location quotients (LQs), concentration indexes and input–output analyses.<sup>3</sup> More recent applications use spatial statistics, in which the analysis of agglomeration puts a great deal of emphasis on space, distance and spatial dependence (van Oort, 2017). Our approach relies on the use of spatial statistics, as these statistics offer some added advantages compared to other methods. First, they induce measurement improvements in the exact definition of agglomeration as the distance becomes more functional in character. Second, they offer a finer spatial scale than metropolitan areas (cities, commuting zones, local authorities), shedding light on intra-urban dependency (van Oort, 2017; Wallsten, 2001) and, for our purposes, allowing us to elucidate levels of clustering in rural areas, where skewness of population centres means that average measures such as LQs may smooth out activity hotspots.

Our approach in this paper draws upon this spatial approach to clusters by adopting the concept of the creative ‘microcluster’ (Boix et al., 2016; Siepel et al., 2020) as the basis for our analysis. The ‘microcluster’ concept is based on the idea that agglomeration dynamics in the CIs occur within the first kilometre of a business’s location (Arzaghi & Henderson, 2008; Coll-Martinez, 2019; Coll-Martinez et al., 2019). On this basis, examining microclusters, which have previously been defined as consisting of 50 or more proximate businesses (Boix et al., 2016; Siepel et al., 2020), has the potential to provide a granular means of identifying localisation economies at a fine level of geography. This approach can therefore allow us to capture smaller agglomerations, not just in rural areas (where these might be smoothed out using LQs) but also in urban areas, where microclusters can identify specific neighbourhood-level agglomerations.

---

<sup>3</sup> Early exercises studied clustering using LQs, for example Florida, Mellander, and Stolarick (2008) for the music industry in the US; Capone (2008), Lazeretti et al. (2008) and Domeneche et al. (2011) for the case of Italy, France and the UK. For the UK, De Propris et al. (2009) used firm-level data.

Our approach to identifying these microclusters draws upon recent studies that have explored clustering at a fine geographical level using data scraped from company websites (e.g. Papagiannidis et al., 2018; Rammer et al., 2020; Siepel et al., 2020; Stich et al., 2022). These studies generally use the addresses provided on company websites to identify and inductively map clusters of activity, often in a way that differs or is more insightful than standard SIC codes (as in Papagiannidis et al., 2018). Our approach differs from previous studies in that, whereas other studies looking at clustering using web data have looked at clustering in an individual neighbourhood or city (Rammer et al., 2020; Stich et al., 2022) or one region (Papagiannidis et al., 2018), we attempt to map across the whole of England, and we try to use this data to identify clusters in rural areas. To our knowledge, neither of these has been done previously.

### 3.1 Data

The data we used was collected by the analytics company Glass.ai in 2019. From a sample of all UK-based websites (approximately 2,690,395 in total), we sought to identify CI websites that provided physical addresses, and from those to identify those in rural areas. We then used very fine-grained geographical units to identify local assets and to model the relationship between creative microclustering and the assets discussed above. Figure 1 shows our empirical approach.

#### Defining the creative industries

There are a number of terms that are widely used as underlying creative clusters: creative industries, cultural and creative industries and creative economy, among others, can all be used in association with the concept of creative clusters, and all have slightly distinct meanings (see for instance Higgs & Cunningham, 2008). Our interest, for the purposes of the analysis we conducted, was a bespoke extension of the UK creative industries definition by the Department for Culture, Media and Sport (DCMS), which identifies nine subsectors: advertising and marketing; architecture; crafts; design; film, TV and radio; IT and software; publishing; museums and libraries; and music and performing arts (DCMS, 2016), which is the de facto definition used in the UK. Given the nature of the questions we are asking, particularly relating to local cultural amenities, a definition of CIs that includes cultural organisations risks double-counting these organisations on both sides of any model. Therefore, to address this issue, we adopted the DCMS definition but excluded organisations in the museums and libraries, music and performing arts sectors. This selection means that we excluded some key cultural elements in the 'creative and cultural industries' definition; however, the remaining sectors allowed us to explore clustering in related sectors.

Operationalising this definition is also potentially a challenge, as the DCMS definition is based upon four-digit Standard Industry Classification (SIC) codes, and our data is SIC code-agnostic. The scraped web data that we used is inductively classified into 109 broad sectors. Our definition of 'creative industries' is, therefore, all of the broad sectors from the web data that map onto one of the seven subsectors from the DCMS definition: advertising and marketing (the 'marketing and advertising' and 'public relations and communications' categories in the scraped data); architecture ('architecture and planning'); crafts ('arts and crafts'); design ('design' and 'apparel and fashion' (excluding fashion retail)); film, TV and radio ('animation', 'broadcast media', 'media production',

'motion picture and film' and 'photography'); IT and software ('computer games', 'computer software'); publishing ('newspapers and magazines', 'online media', 'publishing', 'translation' and 'writing and editing').

### **Identifying websites with addresses**

Applying this definition to our data resulted in a sample of 361,459 websites that had text indicating participation in one of the sectors named above. But, importantly, not all websites had addresses on them. Businesses may choose to list their address on their website if they want their customers to find them easily, but equally businesses may choose not to list their address (for instance if they do not have an office, or work from home, or do not wish to be visited by members of the public). When we removed the websites that did not list an address, we were left with a working sample of 154,618 creative organisations with postcodes in England. This meant that 42.8% of websites listed an address. While this is considerably higher than the 24% of websites containing addresses in Papagiannidis et al.'s (2018) study, it is possible that this represents a bias in our results, and to this end Table 1 summarises the count of businesses in the overall sample as well as the sample with addresses. Generally speaking, the percentage composition of the samples is remarkably consistent between the two samples. Where there is variation (notably design, film and TV, although some other subsectors as well), we hypothesise that these are in sectors characterised by high levels of self-employment and freelance work, in which workers might not keep consistent offices. Measuring the location and activities of freelancers is a substantial problem for any effort to map CIs (see Paneels et al., 2021, for a summary of the issues). While it would obviously be preferable for all websites to list an address, the coverage we have in this dataset is clearly unique and is likely superior in terms of coverage of freelance workers compared to other forms of business registration.

### **Identifying rural businesses**

We identified 21,124 creative companies located in rural settings as characterised by the UK Office for National Statistics definition of rural areas for England (ONS, 2016), based on the postcodes provided in the analysis. These companies represented 14% of the total sample. Figure 2 demonstrates that about half of the sample firms were located in rural villages and dispersed, and 43% operated within rural towns and the periphery in England.

### **Identifying microclusters**

From the geo-located data, we determined whether a firm was in a microcluster (i.e. a small concentration or group of firms that are close to each other). We implemented a self-adjusting (HDBSCAN) clustering method (Campello et al., 2013; McInnes et al., 2017) to detect areas where companies were concentrated and where their location was based in sparsely populated or empty areas. The clustering method employs an unsupervised machine-learning clustering algorithm to identify a range of distances to separate clusters of varying densities from sparser noise. The algorithm computes hierarchical estimates and scores the outlierness of each data object, extracting



local clusters based on a cluster tree.<sup>4</sup> The algorithm optimally creates the most stable clusters that incorporate as many firms as possible without incorporating noise. Appendix 1 gives further details behind the clustering method.

To identify the threshold of values of what constitutes a 'microcluster', the algorithm requires only one input – the analysis to select the minimum size per cluster. We set the condition of the number of firms in the cluster ranging from 50 firms as the minimum cluster size to N. The threshold of a minimum of 50 firms has been used in previous microcluster studies, including Boix et al. (2016) and Siepel et al. (2020). This threshold could reasonably capture effects at an immediately proximate area as being creative clusters. Looking at the number of neighbours at different radii (Table 2), we see that, up to one kilometre, the average number of neighbours was 14 firms. Up to five kilometres, the average number of neighbours was 44. Our threshold of 50 firms per cluster is, therefore, a relatively conservative measure in capturing hotspots of rural firms in a radius of about one to five kilometres. Previous evidence for the CIs has shown that agglomeration dynamics in the CIs occur within the first 500 metres and kilometre (Coll-Martinez, 2019; Coll-Martinez et al., 2019). Tables A1a and A1b in the appendix show a summary of a sensitivity analysis that we carried out, testing different threshold measures. We see that within the same radius the median and the average number of microclustered firms remain relatively consistent even as the minimum size of the cluster changes.

Through the application of the density-based clustering method, we identified 71 rural creative microclusters across England. Overall, about 38% of rural firms in the sample were in a microcluster (see Table 3). The fraction of firms in microclusters by subsector in Table 3 varied across sectors, with the IT/software sector the most likely to be located in a microcluster. We explore potential explanations for this finding in Section 4.2. Figure 2 displays the clusters identified. Looking at the microclusters map, we can see that some clusters were on the periphery of large cities as those surrounding London, Manchester or Birmingham, while others, such as some in the north-west of England, were near national parks like the Lake District National Park.

### 3.2 Regression analysis

With the mapping complete, the next step was to identify the determinants of clustering. We estimated a set of regression models, intending to analyse how certain local factors were associated with the location of rural–urban creative microclusters. We explain the selection of the variables in more detail below.

#### Dependent variable

We summed the number of creative firms in microclusters across a Lower Super Output Area (LSOA). The sum corresponds to the total number of firms in microclusters at each LSOA within one kilometre of the LSOA centroid. An LSOA is a census dissemination unit that represents homogeneous neighbourhoods of 1,500 residents on average, and it is the smallest geographical

---

<sup>4</sup> For further details on the cluster method see Campello et al. (2013).

unit used by the UK Office of National Statistics. Using LSOAs as a measure has the advantage of capturing microclustering dynamics at a very granular level (i.e. close to a neighbourhood), while offering the opportunity to use centroids to aggregate headcount information at a radius of one kilometre of the LSOA centroid to construct our control variables.<sup>5</sup> Previous empirical analyses have also shown that creative firms only benefit from localisation economies within the first kilometre (Arzaghi & Henderson, 2008; Coll-Martinez, 2019; Coll-Martinez et al., 2019).

## Explanatory variables

Our interest in this empirical exercise is to test the role of different types of location-based amenities and the learning aspect of agglomeration economies, with a focus on rural creative microclusters. Our selection of explanatory variables drew on prior studies of CI location and general firm location studies. We geocoded data from several databases to identify the following set of regressors.

*Neighbourhood supply of cultural amenities.* We defined cultural amenities as including museums, public galleries, heritage sites, libraries, archives and science centres. We calculated the number of cultural institutions within a one-kilometre radius of the centroid of each LSOA (*ln\_cultural\_inst*). The data used was based on the comprehensive Culture24 database<sup>6</sup> of UK cultural organisations, which covered 11,304 listings, for which 10,571 places were geocoded and merged to our main data.

*Local cultural and creative industry-related networks.* We used this to capture opportunities for networking through informal social connections through social networks, fairs and venues. To account for this, we used the Culture24 dataset to identify the number of organisations dedicated to exhibitions, campaigns and initiatives, festivals, cultural and scientific meetings within a one-kilometre radius of the centroid of each LSOA<sup>7</sup> (*ln\_cultural\_network*).

*Neighbourhood supply of nature-based amenities.* Place-bound resources may be exploited by firms if they locate nearby. Naldi et al. (2021), for instance, showed that both urban and rural firms derive a positive benefit from natural amenities such as natural areas and parks. Our measure of natural amenities (*ln\_nature*) summarises the number of gardens, environmental and ecological centres, national parks and areas of outstanding natural beauty within a one-kilometre radius of the centroid of each LSOA.

*Local knowledge environment and local labour pool.* To proxy the local knowledge environment, we controlled for the distance (in kilometres) to the nearest higher education institution (universities and further education colleges) from the LSOA centroid (*ln\_distance\_HEI*).<sup>8</sup>

---

<sup>5</sup> There are 34,753 LSOAs in England and Wales. While the full dataset contains this number of LSOA, some of the crucial control variables that we employ in our models are not available for all LSOAs. Therefore, in subsequent analyses we present data on the sample of LSOAs for which information on all relevant variables is available.

<sup>6</sup> Data was drawn from Culture24, a private organization that operates in the UK and has the rights to the most complete data of cultural amenities available in the country. We thank Culture24 for providing an API to access the data. Culture24 does not bear any responsibility for the analysis of the data.

<sup>7</sup> Data also collected from Culture24. See footnote 6 for further details.

<sup>8</sup> Data retrieved from the Historic England 'Register of Historic Parks and Gardens of special historic interest in England' available at: <https://www.data.gov.uk/dataset/88cfe0de-85cd-431f-9836-2bee841d8165/registered-parks-and-gardens-gis-data>

*Agglomeration spillovers.* Evolutionary economic geography argues that both cognitive proximity and geographical proximity are important in the flow of knowledge through regions (Boschma & Martin, 2007). Drawing on this notion, we introduced three measures usually applied in the empirical literature on industrial location. First, we controlled for the industry composition by computing two diversity indexes of all industry sectors at the LSOA, following Frenken et al.'s (2007) and Wixe and Andersson's (2017) approach. The first index refers to unrelated diversity (*UD*), which determines the extent to which firms operate in different industries that share a number of similarities within the local area. Operationally, UD measures the distribution of employees in the neighbourhood between two-digit industries.<sup>9</sup> The second index, related diversity (*RD*), captures the extent to which firms operate in different industries that share few or limited similarities. The index measures the distribution of employees between five-digit industries within each two-digit sector. The concepts of related diversity and unrelated diversity reflect the level of regional/local specialisation: a low level of regional specialisation could be an indication of a high level of related or unrelated industrial diversity (Aarstad et al., 2016).<sup>10</sup>

The third measure captures neighbourhood specialisation using employment-based LQs at the LSOA level. LQs are computed for CIs, manufacturing, services-based activities and knowledge-based activities (following the use of similar measures in Arauzo et al., 2010; Cruz & Teixeira, 2021; Lazzeretti et al., 2012).

We also tried to control for the presence of agglomeration and urbanisation economies. We first used population density (*ln\_pop*) as densely populated areas display more interactions between economic agents (Rodriguez-Pose & Hardy, 2015). We measured population density as the number of inhabitants per square kilometre at the LSOA level.<sup>11</sup> Second, we controlled for the lack of affordability of local offices and spaces, which is also considered a proxy for agglomeration economies (Andres & Round, 2015; Cruz & Teixeira, 2021; Drennan & Kelly, 2011).<sup>12</sup> We proxied this by including rateable value per square metre at the LSOA level as a control variable (*ln\_rate*). Rateable value is a measure collected by the UK Valuations Office Agency as an indicator of the value of business premises for rental purposes, for the purposes of assessing business tax. We also included the squared term of *ln\_rate* in our models to account for non-linearity. Another control refers to the distance of each LSOA to the main city (*distance2city*). For this purpose, we used the major towns and cities statistical geography, which provides a precise definition of the most important cities and towns in England.<sup>13</sup> Being closer to a core city or area may bring potential economic size benefits (Hanson, 2001). Furthermore, to account for regional economic aspects that

---

<sup>9</sup> We use the Standard Industrial Classifications (SICs). Employment data obtained from the business register data is available on the NOMIS platform.

<sup>10</sup> The measures we employ in this paper are not the only ones that can capture related/unrelated variety. Existing measures also include the co-occurrence of products within firms, input–output linkages, and the intensity of labour relocation (Boschma & Gianelle 2013).

<sup>11</sup> Data drawn from ONS Lower layer Super Output Area population density: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareapopulationdensity>

<sup>12</sup> Data obtained from the Evaluation Office Agency in the UK. For further information on the methodology, see: <https://www.gov.uk/government/statistics/non-domestic-rating-stock-of-properties-2020>

<sup>13</sup> This data, drawn from the ONS, corresponds to 112 major towns and cities in England and Wales. The ONS used a population size threshold of 75,000 usual resident population or workday population to define these cities as at the 2011 census.

can drive clustering, we controlled for the level of unemployment at the district level.<sup>14</sup> As shown by Duranton and Puga (2004) and Glaeser et al. (2015), firms prefer to locate in areas with enough workers.<sup>15</sup> We also controlled for the geography of relative affluence in different LSOAs (*rel\_affluence*), using the UK Index of Multiple Deprivation, which ranks all LSOAs by relative affluence and attributes each LSOA to the decile corresponding to its rank. We use this decile measure for our affluence control.<sup>16</sup> Finally, we controlled for the presence of connectivity infrastructure by including broadband speed at the local level (*ln\_internet\_speed*).

The final model specification, in Equation [1], estimates the number of firms in microclusters as a function of specific local characteristics.

$$\begin{aligned}
 mc\_stock_i = & \beta_0 + \beta_1 \ln\_cultural\_inst_i + \beta_2 \ln\_cultural\_network_i + \\
 & \beta_3 \ln\_nature_i + \beta_4 \ln\_distance\_HEI_i + \beta_5 RD_i + \beta_6 UD_i + \beta_7 LQ_{is} + \\
 & \beta_8 \ln\_internet\_speed_i + \beta_9 \ln\_pop_i + \beta_{10} unemp_i + \beta_{11} distance2city_i + \\
 & \beta_{12} rel\_affluence_i + \beta_{13} \ln\_rate_i + \beta_{14} \ln\_rate_i^2 + \alpha_i + u_i \quad [1]
 \end{aligned}$$

where *mc\_stock<sub>i</sub>* is the number of creative firms in microclusters in area *i*. Two samples were used to estimate this model using the same set of explanatory variables and controlling for dummies for the commuting hinterland where most people work (also known as travel-to-work areas, TTWA) and represented by  $\alpha_i$ . The first sample corresponds to 4,702 rural areas located in England. The second sample covers 18,765 urban areas. We separated these two samples to investigate the reasons for apparent differences in creative microcluster locations in rural and urban areas. For estimation purposes, the reference period for our dependent variable is the year 2019, whereas all explanatory variables are the years 2018 and 2017, where possible, to avoid problems of simultaneity. The data we used was mainly cross-sectional, reducing the possibility of controlling for sources of endogeneity. For instance, the location patterns of creative microclusters could be explained by the innovativeness embedded in places and regions. As we discuss above, we controlled for the presence of universities and colleges as a means of controlling for possible knowledge or innovation spillovers arising from universities. Despite this, knowledge spillovers can come from different sources, for instance dominant technologies being developed in the region, or innovation hubs that attract and support cultural organisations or creative businesses. Nevertheless, the nature of our key variables makes the use of time variation redundant to some extent, as natural and culture-based amenities seldom change over time (and so are time-invariant).

Table 4 displays the variables used and summary statistics. The distribution of our dependent variable (*mc\_stock*) has two features that are worthy of attention. First, the variance is larger than

<sup>14</sup> There are a total of 309 districts in England. They are a level of subnational division of England and determine the structure of local government.

<sup>15</sup> Data obtained from the Annual Population Survey (2018); population data corresponds to 2018 mid-year population estimates by the ONS.

<sup>16</sup> Details on the methodology behind the index of multiple deprivation are available at <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>

the mean, implying that the data is over-dispersed. In addition, the variable refers to the number (or count) of firms in microclusters. For these reasons, we estimated a negative binomial regression (NBR), which can model the dispersion by adding an extra parameter into the model. The NBR is a generalisation of Poisson regression. Table A3 in the appendix reports a correlation matrix. None of the correlations reported appears to be particularly high.<sup>17</sup>

## 4. Results

### 4.1 Rural vs urban determinants of microclustering

We present the results of estimating models by using two different subsamples, split by rural and urban location (Table 5). For each sample, we evaluated four separate models, testing different specifications. Overall, the results are generally robust across all estimations (i.e. coefficients do not change markedly when adding additional controls). We also tested non-linear effects for the office rateable value per square metre (*ln\_rate*). The interpretation that follows is based on our preferred specification, displayed in columns 4 and 8. All models have mean variance inflation factors below 4, and the Hausman test on TTWA fixed effects is also statistically significant (i.e. the model with TTWA fixed effects is preferred to the one without TTWA fixed effects).

Regarding our main explanatory variables, the regressions show a positive sign for cultural amenities when the model is estimated using both subsamples. Regarding the magnitude of the effects, the effect size of cultural amenities is appreciably stronger for rural areas than that for urban areas (0.45 vs 0.28)<sup>18</sup>. This result corroborates the fact that the provision of culture-associated activities is an essential aspect in the dynamics of creative microclusters in urban and rural areas. In other words, the accumulation of culturally led facilities could stimulate clustering, providing a common resource base that brings identity and aesthetic values (Throsby, 2008). These results are parallel to those of Lazeretti et al. (2012) for Italian urban areas, where the presence of cultural and artistic heritage influences the presence of heritage-dependent CIs. On this basis, we find H1 to be supported.

Concerning the role of creative and social networks, the results show that this variable is positively associated with the number of firms in microclusters in local urban and rural areas, but the results are highly significant for urban areas and (weakly) significant for rural areas, only providing partial support for H2. The effect size in urban areas is relatively higher compared to rural areas (0.19 vs 0.11). This finding suggests that these relationships are important, but the density and proximity of informal social networks make this effect stronger in core urban zones, while the relationship with cluster activity in rural areas is potentially weaker owing to smaller and more displaced networks. Even though the variable used in our analysis is an imperfect proxy of creative networking, this result opens the avenue for policy to encourage the formation of informal networks to connect business and people to places, in both rural and urban areas.

We find a complex picture regarding the association of nature-based amenities and clustering. On the surface we do not see a significant association between immediate proximity to nature-based

<sup>17</sup> Variance inflation factors are below the threshold of 9, providing no indication of strong multicollinearity.

<sup>18</sup> By multiplying the coefficient by the standard deviation of the variable, effect sizes can be calculated, enabling a more nuanced comparison of the magnitudes of effects between rural and urban areas.

amenities and creative microclustering in either the urban or rural samples. While this is consistent with previous results for urban areas (Naldi et al., 2021), the lack of a result for rural areas may be explained by the limited variation in the natural amenity measure, especially in urban areas (Table 4 shows that the mean count of urban LSOAs containing natural amenities is 0.08 and, of our urban sample, 93% of urban LSOAs have no natural amenities). One possible explanation is that, in rural areas, those places with high numbers of gardens and other natural amenities are unlikely to also have any substantial (non-agricultural) business activity within a one-kilometre radius. In this case, our results would suggest that these natural amenities themselves do not necessarily appear to attract co-location of creative businesses, and a negative association between natural amenities and microclusters should be expected. However, it is possible that this represents a composition issue in our sample in which there is variation in subsectoral effects, which we explore in detail in the following section. On this basis, we consider H3 to be only partially supported.

Universities play an important role in microclusters in rural and urban areas. The coefficient associated with the log of the distance to the nearest university is negative and statistically significant. This means that microclusters located closer to universities in rural regions exercise an attractive attraction pole that promotes the agglomeration of CIs and perhaps that staff/students at these institutions are active in business offering creative services, or perhaps that graduates of these universities are more likely to start businesses in proximity to these universities, following Kitagawa et al. (2022). We explore this issue in more detail in the next section. On this basis we consider H4 to be accepted.

When examining the relationship between agglomeration, spillovers and the number of microcluster firms in the local area, the coefficients obtained suggest a strong and positive correlation between related diversity, unrelated diversity and microclustering for urban and rural areas. For the case of related variety, which measures the distribution of employees between five-digit industries within each two-digit sector, regions with a high degree of related activities are more likely to have companies located as a part of a microcluster.<sup>19</sup>

Another important finding is that the coefficient of unrelated diversity, which captures the degree to which firms operate in different industries with few similarities, is also positive and statistically significant across models for both samples. This finding, while preliminary, supports the idea that regions that host a highly diverse productive structure (including creative and non-creative industries) are also more likely to host microclusters. In other words, local interactions and spillovers outside the industry may drive agglomeration processes even in rural areas and offer regional resilience to sectoral economic shocks, unlike in single sector regions. With this said, the effect size for related diversification (0.21) is very similar to the effect size for unrelated diversification (0.22). This combination of findings supports the conceptual premise that the process of agglomeration in rural areas is driven by both the co-agglomeration of similar industries (Marshall externalities) but also by a diverse set of unrelated industries (Jacobs externalities). We therefore consider both H5 and H6 to be supported.

We also observe some differences between rural and urban areas regarding industry specialisation, measured by LQs, and the presence of firms in creative microclusters. For the case of rural areas, we find a weakly statistically significant negative association between concentration in services and creative microclustering, and no significant association with manufacturing, CIs, or knowledge-

---

<sup>19</sup> This correlation holds even after including population density, which is thought to have a high correlation with these factors.

intensive sectors. While CI businesses are often in the supply chain of non-tradable sectors like services, it may be that non-tradable sectors reliant on local demand do not find it advantageous to locate in rural settings where the demand they can pull is too low (see Goffette-Nagot & Schmitt, 1999). For urban areas, we find that the presence of creative microclusters is negatively associated with the specialisation of manufacturing and knowledge-based firms. The negative association with manufacturing firms in urban areas may reflect the relative concentration of manufacturing in industrial parks, which we would not expect to also host creative microclusters. The negative association with knowledge-based firms is more perplexing, although – similar to manufacturing – this may also reflect differences between businesses requiring specialised facilities (e.g. R&D labs) that in urban areas might not be collocated with creative businesses.

We find that the area's population density is associated with the number of microclustered firms in both rural and urban areas (see coefficients of  $\ln\_pop$  in columns 1–3 and 5–7 in Table 5). However, the effect of this is absorbed when including the entropy measures (related and unrelated diversity) (see coefficients in columns 4 and 8). Table 5 indicates that the distance coefficient to the city is negative for rural areas and positive for urban areas, and it is statistically significant in both cases. This suggests that creative microclusters tend to thrive best in areas close to cities but not necessarily in the city centre, such as suburbs or exurbs. It is possible that costs may be too high for creative microclusters to thrive within cities, while demand may be too low for them to be successful in more remote areas. When we consider unemployment and microclustering, we find a negative and statistically significant result, in line with previous research showing that CIs firms are less likely to be located in more deprived places with higher rates of unemployment. This finding is also supported by our measure of relative affluence, which finds that creative rural firms in affluent areas are more likely to be associated with microclustering, though as we explore below this has some sectoral composition effects. It is worth noting that, in urban settings, we observe the opposite result, where greater affluence is linked to a lower number of creative microclusters. It is likely that this is the result of a complex interplay of factors, including higher rent and property costs and greater competition, which may mean that consumer-facing creative businesses are successful but which may discourage businesses selling to other businesses, thus deterring the formation of creative microclusters. Another possible explanation is that, as seen in the burgeoning literature on gentrification (Behrens et al., 2022), less-affluent, low-cost locations may be more attractive to creative businesses and thus may be more likely to be the centres of creative microclusters.

The regression results also show a surprising negative association of creative microclusters with internet speed in urban areas. This could be due to a number of factors, including urban areas already having relatively high-speed internet, thus negating further locational pull when even faster internet is made available (Tranos & Mack, 2016); the rise of remote working (even before COVID-19; see also Florida, 2017), which could potentially mitigate drivers for physical co-location; or the possibility that higher speeds are correlated to areas (such as industrial parks) where CIs are unlikely to be located.

Finally, our findings show that the relationship between creative microclusters and office space value is decreasing at an increasing rate. This is evident in the negative coefficient associated with rateable value per square metre and the positive coefficient of rateable value squared. The F-test also provides evidence that the quadratic term is not equal to zero. However, the minimum of the quadratic relationship occurs outside the range of values in the data, suggesting a monotonically decreasing relationship. This means that the slope of the relationship becomes less negative as the rateable value increases.

Overall, these results may reflect general trends across the CIs, such as the increasing prevalence of remote work and digitalisation, and the growing demand for affordable places to do business (Florida, 2017).

## 4.2 Sector-specific determinants of microclustering in rural areas

As shown in Table 3, the percentage of firms in microclusters varies across subsectors. To explore this variation in more detail, we ran separate models for each of the seven DCMS creative subsectors we were considering. All models are estimated using an NBR as in the previous subsection (Table 6). The specification corresponds to models shown in columns 4 and 8 in Table 5, in which all key explanatory variables and controls are included. The dependent variable (*mc\_stock*) corresponds to the number of firms in microclusters that belong to each of the subsectors.

When we consider individual sectors as opposed to ‘creative industries’ as a group, we find relatively few differences, but some that prove to be insightful. Several key results, such as the significance of cultural institutions and proximity to HE institutions, as well as the related diversity and unrelated diversity measures, are all significant across each of the subsectors, indicating that these effects appear to be consistent. But the natural amenities variable shows substantial variation – architecture and crafts are highly significantly associated with natural amenities, and design and film and TV are weakly significant. By contrast, advertising and marketing, IT and software, and publishing do not have an association with natural amenities. The cultural networks variable, which is weakly significant across the whole sample, is only significant – albeit weakly – for the film, TV and radio subsector.

The results in Table 6 also show interesting subsectoral variation around agglomeration and affluence, with advertising and marketing, architecture, and IT and software being more likely to be clustered in more affluent areas. This appears to drive our results across the whole rural sample, and also shows how IT and software (which is associated with other IT businesses but only weakly associated with cultural amenities) and advertising and marketing (which is associated with population density) may have different spatial configurations from other creative sectors (such as crafts, where clustering is associated with lower broadband speed and greater distance to nearest cities).

## 4.3 Robustness checks

The clustering algorithm used in this paper requires the analyst to select the minimum number of firms in a microcluster. As discussed in Section 3.1, we selected 50 firms as a minimum threshold, following previous empirical evidence. To check that our results are not dependent on the selected minimum number of firms in a microcluster, we re-estimated our regression model using 25 firms instead of 50 (resulting in 216 rather than 71 rural microclusters, according to Table A1). We find that the results have broadly similar estimates by using a different threshold (Table A4 in the appendix).

### Changing the geographical unit of analysis

In our previous regressions, we used a finer geographical area – the LSOA – which represents a geographical area containing 1,500 residents on average. However, by using this level of granularity,



we could potentially ignore dynamics that extend beyond this geographical grid. To test for this granularity, we now change the unit of analysis to a higher level of geographical aggregation, corresponding to Middle Layer Super Output Areas (MSOAs). MSOAs are built from groups of contiguous LSOAs, capturing 7,200 inhabitants on average. Table A5 in the appendix reports the regression estimates from Table 5 but now using MSOAs as the unit of analysis. All variables measure the natural logarithm of  $X$  in the MSOAs, where  $X$  corresponds to firms or amenities depending on the variable type. The estimates obtained broadly confirm our previous results, once a higher level of geographical aggregation is accounted for. Note that the coefficient of unrelated diversity ( $UD$ ) becomes statistically insignificant. One possible interpretation is that these factors may not influence the level of microclustering beyond a certain geographical distance. We also note that the negative coefficient for LQ for knowledge-intensive sectors ( $LQ\_know$ ) in rural areas becomes statistically significant in this specification. This negative relationship could be due to the relative absence of specialist (e.g. R&D) facilities in larger rural areas in places where there are otherwise creative microclusters, or possibly different labour markets or limited connectivity to knowledge networks.

### Controlling for spatial dependence

Given that the dynamics of microclustering could extend beyond the local area, we needed to control for the potential influence of neighbours on the location of creative microclusters. In other words, one could expect that geographical areas hosting microclusters could exercise influence on their neighbours (spillovers across geographical units). This type of influence generates spatial dependence across geographical units, which could cause an omitted variable bias (Paelinck, 2000). To correct for this, we estimate four spatial regression models. The first model (Equation [2]) estimates a spatial autoregressive model (SAR).<sup>20</sup> This model is estimated using a contiguous spatial weights matrix ( $W$ ), which measures who the neighbours are. The vector of explanatory variables is represented by  $X$  and  $e$  corresponds to the residuals, which are assumed to be independent and identically distributed (*i.i.d*). We also estimate a spatial error model, in which the dependent variable  $y$  is regressed on a spatially correlated error term ( $We$ ), a vector of explanatory variables and an error term  $v$ , which is *i.i.d* (Equation [3]). The third model in Equation [4] estimates a spatial and error lag model. Finally, we estimate a spatial Durbin model to account for both direct and indirect effects of the explanatory variables on the dependent variable, and the indirect effects are spatially dependent. In this case, the dependent variable is regressed on its lagged values, the lagged values of the neighbouring observations, and the explanatory variables (Equation [5]).

$$y = \rho Wy + X'\beta + e \quad [2]$$

$$y = X'\beta + \lambda We + v \quad [3]$$

$$y = \rho Wy + X'\beta + \lambda We + v \quad [4]$$

$$y = \rho Wy + X'\beta + WX\theta + e \quad [5]$$

<sup>20</sup> We use a contiguous spatial weight matrix. Before estimating models, we checked global spatial dependence by means of Geary's  $c$  and Getis and Ord's  $G$  tests. The null hypothesis of no spatial dependence was rejected ( $p$ -value = 0.000).

We used as many covariates as from the previous regressions as we could, although we had to drop some variables (*ln\_nature*, *LQ\_manuf*, *unemp*, and *ln\_rate2*) due to high levels of correlation. The dependent variable (*y*) corresponds to the number of microcluster firms in the MSOA (*mc\_stock*). This methodology has been previously employed in a recent study by Arauzo-Carod et al. (2023).

Table A6 in the appendix reports the main results for both samples (rural and urban locations). Columns 1 and 5 report regression estimates for a spatial lag autoregressive model (i.e. the spatial lag of the dependent variable enters as an explanatory variable). We can see that the coefficient is statistically significant, confirming that the number of creative firms in microclusters behaves with a spatial structure. The models in columns 2 and 6 estimate an autoregressive model with a spatially autocorrelated error term (*e.mc\_stock*). This variable is statistically significant. A model that combines a spatial and error lag model is estimated in columns 3 and 7. The final model (in columns 4 and 8) corresponds to a mixed regressive-spatial autoregressive model with spatial autocorrelation in the independent variables (spatial Durbin model). All tests on the spatially autocorrelated error terms (*e.mc\_stock*) are statistically significant. The variable associated with the spatially autocorrelated error term is statistically significant across these models. Considering the final model (columns 4 and 8), spatial autocorrelation in the independent variables is statistically significant only for urban areas. Overall, the results from the spatial models support our general findings. Specifically, the coefficients for cultural institutions, cultural networks, and distance to universities exhibit the same direction and similar level of statistical significance in both samples. However, it should be noted that the coefficient for social capital is only weakly significant in the rural sample. Additionally, the coefficient for unrelated diversity is statistically insignificant in the rural sample. The LQ for services businesses in urban areas switches signs to become positive and becomes significant in the model, which could potentially reflect collocation in urban areas where MSOAs are comparatively smaller in area and hence might capture collocation. The LQ for rural knowledge-intensive businesses becomes weakly significant in this specification.

## 5. Conclusions

It is well-established that businesses and workers in the creative industries benefit from agglomeration and co-location. But, to date, most research on creative clusters has assumed this co-location to be a largely urban phenomenon. This paper draws upon growing evidence that clustering can be equally important in rural settings (Harvey et al., 2012; Merrell et al., 2021) to capture the extent of ‘microclustering’ – that is, smaller spatial agglomerations (Boix et al., 2016; Siepel et al., 2020) – of creative firms in rural areas. Using a novel mapping technique based on spatial analysis of scraped company website data for the whole of England, this study has mapped these microclusters in rural and urban settings. It then aimed to understand whether the determinants of these microclusters are different in rural and urban areas, particularly with respect to local assets such as cultural institutions and networks, natural amenities and agglomeration economies. We answered this question using geocoded data from 154,618 websites of CI businesses and organisations in England. We used a clustering algorithm to identify 71 rural ‘microclusters’, each of which had 50 or more creative businesses and which collectively made up 38% of the rural creative industry firms in our sample. We then explored the determinants of clustering and analysed the differences between rural and urban microclusters, using fine geographies to identify these determinants.

Our primary finding is that determinants of clustering appear to generally be similar between rural and urban areas. For instance, in both settings we find that microclustering is associated with heritage and culture-led facilities and a diverse set of local industries that share a number of

similarities (that is, related diversity). Where we find differences (for instance relating to the role of natural amenities), this appears to be driven by variation between the creative subsectors in our sample. We find that creative microclustering is associated with proximity to universities in both urban and rural contexts, and is positively associated with related and unrelated diversification. This paper makes two main contributions to the literature. First, it introduces a novel technique for identifying 'microclusters' that uses scraped web data and an inductive clustering algorithm that allows us to identify clusters across and distinct from geographical boundaries. This method provides a novel technique distinct from the location quotients widely used in the literature elsewhere (Escalona-Orcao et al., 2016). Second, drawing upon this technique, we identify rural creative microclusters and then compare the determinants of clustering in these microclusters to those in urban areas. We therefore contribute to the longstanding literature on firm location by addressing not only the determinants of location for creative firms (Coll-Martinez et al., 2019), but how these vary in different contexts, specifically rural (Naldi et al., 2021) contexts. Our finding that rural determinants of microclustering are very similar to those in urban areas, applying a high level of territorial disaggregation, suggests that the sectoral trends outweigh the urban/rural distinction. Bringing these insights together then allows us to draw greater insights about drivers of agglomeration more broadly, comparing rural and urban areas.

The policy implications are twofold. Our findings regarding the association between cultural institutions and associated creative clustering support the possibility for culture-led regeneration and placemaking in rural areas, creating what Ray (2001) refers to as 'culture economies' and Naldi et al. (2015) discuss with 'smart rural development', where CIs are used as a basis for local revitalisation (Wolman & Hincapie, 2015). We have shown factors that may promote such microclusters but not why that matters or should matter to policymakers. Future research can take up that challenge.

Second, this research suggests that policymakers would benefit from recognising the distinctive features of rural CIs outside of cities, and consider a more nuanced place-based approach. In particular, efforts to support CIs should therefore not overlook or otherwise exclude rural clusters in favour of cities. Indeed, targeted support to develop firms in microclusters and clusters wherever they are will help to unlock the potential of the CIs, in urban and rural contexts. Where rural-specific interventions are designed, efforts that support the development of informal networks between businesses and between rural people and organisations that stimulate local demand seem to be one promising approach (Hill & Mole, 2022).

To conclude, as with all research, our analysis has some limitations. First, although we use a novel data set, its structure is cross-sectional, giving little room to control for potential sources of endogeneity coming from omitted variables and reverse causality. As a result, our findings only demonstrate partial associations and cannot establish causal relationships. Nonetheless, we took several measures to address potential factors that could undermine causal interpretations. For example, we used earlier time periods for all explanatory variables, excluded certain creative sectors from the definition of microclusters to prevent inducing correlations between key explanatory variables and the dependent variables, employed TTWA fixed effects to control for unobserved differences across labour markets affecting outcomes, and controlled for a significant number of confounding variables. These steps helped to strengthen the validity of our results and provide a more robust foundation for future research on this topic. Second, our data reflects spatial distribution, but we cannot control for firm-level characteristics such as age, size and type of organisation, as well as the presence of actors such as freelancers who may not have a web presence. These issues would deserve further research. Third, because we are capturing spatial

clustering, our findings in this article do not allow us to make a statement about the existence of agglomeration economies per se in the identified rural microclusters. Finally, the relationships we identify are based on pre-COVID-19 data. The impact of COVID-19 on rural microclusters is important in multiple ways – in terms of both the changing spatial distributions of business activities as a result of the pandemic (for instance through relocations as creative workers move from urban to rural areas because of the pandemic) and the resilience of rural microcluster businesses. Further research needs to establish the impact of COVID-19 on these clustering patterns. In addition, further evidence about the nature and impact of agglomeration in rural creative clusters could be very helpful in setting policy and research agendas to help rural areas to unlock the socio-economic potential of creative clusters.

### Acknowledgement

We would like to thank the many people who have helped us to develop this research: First and foremost, we are grateful to NICRE, Research England, and AHRC through the PEC for providing funding and support to develop this research. We would like to thank the three anonymous reviewers and the editor of the journal for taking the time and effort necessary to review our manuscript. Their valuable comments and suggestions have been instrumental in enhancing the quality of the paper, and we truly appreciate their constructive feedback. We also like to thank Bruce Tether, Ben Kulka and Eliza Easton at the PEC and Ulrike Hotopp at NICRE for providing a thorough revision of a preliminary version of the paper. We also like to thank the people who attended presentations of this study at the Creative Rural Livelihoods conference (Newcastle University) and at the Creative Industries mobiliser group Seminar (University of Sussex) and who provided constructive comments. The usual disclaimers apply.

### References

- Aarstad, J., Kvitastein, O. A., & Jakobsen, S. E. (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. *Research Policy*, 45(4), 844–856.
- Andersson, R., Quigley, J., & Wilhelmsson, M. (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics*, 66(1), 2–15. <https://doi.org/10.1016/j.jue.2009.02.004>
- Andres, L., & Round, J. (2015). The creative economy in a context of transition: A review of the mechanisms of micro-resilience. *Cities*, 45, 1–6.
- Agarwal, S., Rahman, S., & Errington, A. (2009). Measuring the determinants of relative economic performance of rural areas. *Journal of Rural Studies*, 25, 309–321.
- Arauzo-Carod, J. M., Coll-Martínez, E., & Turcu, C. (2023). Location attributes explaining the entry of firms in creative industries: evidence from France. *The Annals of Regional Science*, 1–32. (last accessed 12-7-23, <https://doi.org/10.1007/s00168-022-01196-w>).
- 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.
- Argent, N. (2019). Rural geography III: Marketing, mobilities, measurement and metanarratives. *Progress in Human Geography*, 43(4), 758–766.
- Arzaghi, M., & Henderson, J. V. (2008). Networking off Madison Avenue. *The Review of Economic Studies*, 75(4), 1011–1038.

- Bakhshi, H., Lee, N., & Mateos-Garcia, J. (2014). *Capital of culture? An econometric analysis of the relationship between arts and cultural clusters, wages and the creative economy in English cities*. National Endowment for Science, Technology and the Arts.
- Balfour, B., Fortunato, M-P., & Alter, A. (2018). The creative fire: An interactional framework for rural arts-based development. *Journal of Rural Studies* 63, 229–239.
- Behrens, K., Boualam, B., Martin, J., & Mayneris, F. (2022). Gentrification and pioneer businesses. *Review of Economics and Statistics*, 1–45. [https://doi.org/10.1162/rest\\_a\\_01154](https://doi.org/10.1162/rest_a_01154)
- Bell, D., & Jayne, M. (2010). The creative countryside: Policy and practice in the UK rural cultural economy. *Journal of Rural Studies*, 26(3), 209–218.
- Berg, S. H., & Hassink, R. (2014). Creative industries from an evolutionary perspective: A critical literature review. *Geography Compass*, 8(9), 653–664.
- Bergman, E. M., & Feser, E. J. (1999). *Industrial and regional clusters: Concepts and comparative applications*. Reprint. Edited by Scott Loveridge and Randall Jackson. WVU Research Repository, 2020.
- Bloom, M., Camerani, R., Casadei, P., Masucci, M., Siepel, J., & Velez-Ospina, J. (2020). *Evolution and trends of creative cluster research: A systematic literature review and future research agenda*. Creative Industries Policy and Evidence Centre. PEC. Nesta. Evidence Review: 2020/02.
- Boal, I., & Herrero, L. C. (2018). Where are the artists? Analysing economies of agglomeration in Castile and León, Spain. *Papers in Regional Science*, 97(4), 995–1016.
- Boix, R., Lazzeretti, L., Capone, F., De Propriis, L., & Sanchez, D. (2013) The geography of creative industries in Europe: comparing France, Great Britain, Italy and Spain. In: Lazzeretti, L. (ed.) *Creative industries and innovation in Europe: concepts, measures, and comparative case studies*, pp. 23–44. London: Routledge.
- Boix, R., Hervás-Oliver, J. L., & De Miguel-Molina, B. (2015). Micro-geographies of creative industries clusters in Europe: From hot spots to assemblages. *Papers in Regional Science*, 94(4), 753–772.
- Boix-Domenech, R., & Soler-Marco, V. (2017). Creative service industries and regional productivity. *Papers in Regional Science*, 96(2), 261–279.
- Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1), 61–74.
- Boschma, R., & Gianelle, C. (2013). *Regional branching and smart specialization policy*. JRC technical reports (06/2104).
- Boschma, R., & Martin, R. (2007). Editorial: Constructing an evolutionary economic geography. *Journal of Economic Geography*, 7(5), 537–548.
- Boschma, R. A. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, 51(3), 351–364.
- Boschma, R. A., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24(1), 223–250.
- Bosworth, G., & Turner, R. (2018). Interrogating the meaning of a rural business through a rural capitals framework. *Journal of Rural Studies*, 60, 1–10.
- Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in knowledge discovery and data mining* Lecture Notes in Computer Science(), vol 7819, (pp. 160–172). Springer.

- Capone F. (2008) Mapping and analysing creative systems in Italy (1991-2001). In: Cooke P., & Lazzeretti L. (eds) *Creative Cities, Cultural Clusters and Local Economic Development*, pp. 338-365. Cheltenham: Edward Elgar Publishing.
- Cerisola, S., & Panzera, E. (2022). Cultural cities, urban economic growth, and regional development: The role of creativity and cosmopolitan identity. *Papers in Regional Science*, 101(2), 285–302.
- Chapain, C., Cooke, P., Propriis, L. D., MacNeill, S., & Mateos-Garcia, J. (2010). *Creative clusters and innovation*. NESTA. [https://media.nesta.org.uk/documents/creative\\_clusters\\_and\\_innovation.pdf](https://media.nesta.org.uk/documents/creative_clusters_and_innovation.pdf).
- Coll-Martínez, E., Moreno-Monroy, A. I., & Arauzo-Carod, J. M. (2019). Agglomeration of creative industries: An intra-metropolitan analysis for Barcelona. *Papers in Regional Science*, 98(1), 409–431.
- Coll-Martínez, E. (2019). Creativity and the city: Testing the attenuation of agglomeration economies in Barcelona. *Journal of Cultural Economics*, 43(3), 365–395.
- Combes, P. P., Duranton, G., & Gobillon, L. (2011). The identification of agglomeration economies. *Journal of economic geography*, 11(2), 253-266.
- Cooke, P. N., & Lazzeretti, L. (eds.) (2008). *Creative cities, cultural clusters and local economic development*. Edward Elgar Publishing.
- Crociata, A., Agovino, M., Russo, A., & Quagliari Domínguez, A. (2018). Creative workforce and economic development in precrisis Europe: Main trends and causality relationships. *International Regional Science Review*, 41(4), 448–479.
- Cruz, S. C. S., & Teixeira, A. A. (2021). Spatial analysis of new firm formation in creative industries before and during the world economic crisis. *The Annals of Regional Science*, 67(2), 385–413.
- Darchen, S. (2016). ‘Clusters’ or ‘communities’? Analysing the spatial agglomeration of video game companies in Australia. *Urban Geography*, 37(2), 202–222.
- DCMS, Department for Culture, Media and Sport. (2016). *Creative industries economic estimates*. Methodology. DCMS.
- Deller, S. C., Lledo, V., & Marcouiller, D. W. (2008). Modeling regional economic growth with a focus on amenities. *Review of Urban & Regional Development Studies: Journal of the Applied Regional Science Conference*, 20(1), 1–21.
- Domenech, R. B., Lazzeretti, L. ; Hervas, J. L., De Miguel-Molina, & B., Ruiz, B. T. (2011). Creative clusters in Europe: a microdata approach, 51st Congress of the European Regional Science Association: "New Challenges for European Regions and Urban Areas in a Globalised World", 30 August - 3 September 2011, Barcelona, Spain, European Regional Science Association (ERSA), Louvain-la-Neuve. ([https://www.econstor.eu/bitstream/10419/120002/1/ERSA2011\\_0471.pdf](https://www.econstor.eu/bitstream/10419/120002/1/ERSA2011_0471.pdf), last accessed 13-7-23)
- Drennan, M. P., & Kelly, H. F. (2011). Measuring urban agglomeration economies with office rents. *Journal of Economic Geography*, 11(3), 481–507.
- Dubois, A., Copus, A., & Hedstrom, M. (2012). Local embeddedness and global links in rural areas: Euclidean and relational space in business networks. In C. Hedbert & R. M. Do Carmo (Eds.), *Translocal ruralism. Mobility and connectivity in European rural spaces*. Springer.
- Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In V. Henderson & J. V. Thisse (Eds.), *Handbook of regional and urban economics. Cities and Geographies* (pp. 2063–2117). Elsevier.

- Escalona-Orcao, A. I., Escolano-Utrilla, S., Sáez-Pérez, L. A., & García, B. S. V. (2016). The location of creative clusters in non-metropolitan areas: A methodological proposition. *Journal of Rural Studies*, *45*, 112–122.
- Eurostat. (2021). *Cultural statistics – cultural enterprises. Structural business statistics*. Brussels. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture\\_statistics\\_-\\_cultural\\_enterprises#Structural\\_business\\_statistics\\_.28SBS.29](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture_statistics_-_cultural_enterprises#Structural_business_statistics_.28SBS.29).
- Florida, R. (2017). *The new urban crisis: Gentrification, housing bubbles, growing inequality, and what we can do about it*. Oneworld.
- Florida, R., Mellander, C., & Stolarick, K. (2010). Music scenes to music clusters: the economic geography of music in the US, 1970-2000. *Environment and Planning A: Economy and Space* *42*(4), 785-804.
- Frenken, K., van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697.
- Glaeser, E., Kerr, S., & Kerr, W. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, *97*(2), 498–520.
- Goffette-Nagot, F., & Schmitt, B. (1999). Agglomeration economies and spatial configurations in rural areas. *Environment and Planning A*, *31*(7), 1239–1257.
- Gong, H., & Hassink, R. (2017). Exploring the clustering of creative industries. *European Planning Studies*, *25*(4), 583–600.
- Gottlieb, P. D. (1995). Residential amenities, firm location and economic development. *Urban Studies*, *32*(9), 1413–1436.
- Hanson, G. H. (2001). Scale economies and the geographic concentration of industry. *Journal of Economic Geography*, *1*(3), 255–276.
- Harvey, D. C., Hawkins, H., & Thomas, N. J. (2012). Thinking creative clusters beyond the city: People, places and networks. *Geoforum*, *43*, 529–539.
- Higgs, P., & Cunningham, S. (2008). Creative industries mapping: Where have we come from and where are we going? *Creative Industries Journal*, *1*(1), 7–30.
- Hill, I. (2021). Spotlight on UK artisan entrepreneurs’ situated collaborations: through the lens of entrepreneurial capitals and their conversion, *International Journal of Entrepreneurial Behavior & Research*, *27*(1), 99–121.
- Hill, I., & Mole, K. (2022). *State of the art review. Supporting rural businesses*. National Innovation Centre for Rural Enterprise. SOTA Review 4. Newcastle. (Last accessed 14-11-22, <https://nicre.co.uk/media/1wmhonux/nicre-sota-no-4-july-2022-supporting-rural-businesses.pdf>)
- Hill, I., Manning, L., & Frost, R. (2021). Rural arts entrepreneurs’ placemaking – how ‘entrepreneurial placemaking’ explains rural creative hub development during COVID-19 lockdowns. *Local Economy*, *36*(7–8), 627–649.
- Hoyman, M., & Faricy, C. (2009). It takes a village: A test of the creative class, social capital, and human capital theories. *Urban Affairs Review*, *44*(3), 311–333.
- Iyer, S., Kitson, M., & Toh, B. (2005). Social capital, economic growth and regional development. *Regional Studies*, *39*(8), 1015–1040.
- Kitagawa, F., Marzocchi, C., Sanchez-Barrioluengo, M., & Uyarra, E. (2022). Anchoring talent to regions: The role of universities in graduate retention through employment and entrepreneurship. *Regional Studies*, *56*(6), 1001–1014. <https://doi.org/10.1080/00343404.2021.1904136>
- Knudsen, B., Florida, R., Stolarick, K., & Gates, G. (2008). Density and creativity in US regions. *Annals of the Association of American Geographers*, *98*(2), 461–478.

- Lazzeretti, L., Boix, R., & Capone, F. (2008). Do creative industries cluster? Mapping creative local production systems in Italy and Spain. *Industry and Innovation*, 15(5), 549–567.
- Lazzeretti, L., Capone, F. and Boix, R. (2012). Reasons for clustering of creative industries in Italy and Spain. *European Planning Studies* 20(8), 1243–1262.
- Lee, N & Cowling, M. (2015). Do rural firms perceive different problems? Geography, sorting and barriers to growth in UK SMEs. *Environment and Planning C – Government and Policy*, 33, 25–42.
- Lorenzen, M., & Frederiksen, L. (2008). Why do cultural industries cluster? Localisation, urbanization, products and projects. In P. Cooke & L. Lazzeretti (Eds.), *Creative cities, cultural clusters and local economic development* (pp. 155–179). Edward Elgar.
- Mahon, M., McGrath, B., Laoire, L. O., & Collins, P. (2018). Artists as workers in the rural: precarious livelihoods, sustaining rural futures. *Journal of Rural Studies*, 63, 271–279.
- Mateos-Garcia, J., & Bakhshi, H. (2016). *The geography of creativity in the UK*. Nesta.
- McGranahan, D., & Wojan, T. (2007). Recasting the creative class to examine growth processes in rural and urban counties. *Regional Studies*, 41(2), 197–216.
- McGranahan, D. A., Wojan, T. R., & Lambert, D. M. (2011). The rural growth trifecta: Outdoor amenities, creative class and entrepreneurial context. *Journal of Economic Geography*, 11(3), 529–557.
- McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205.
- Merrell, I., Rowe, F., & Cowie, P. (2021). ‘Honey pot’ rural enterprise hubs as micro clusters: Exploring their role in creativity-led rural development. *Local Economy*, 36(7–8), 589–605.
- Mole, K., Hill, I., Nguyen, T., & Maoili, S. (2022). *Rural family businesses and exporting behaviour*. Research report No. 5. National Innovation Centre for Rural Enterprise. Newcastle. <https://nicre.co.uk/media/m3jlul4e/nicre-research-report-no-5-february-2022-rural-family-businesses-and-exporting-behaviour.pdf>
- Naldi, L., Nilsson, P., Westlund, H., & Wixe, S. (2015). What is smart rural development? *Journal of Rural Studies*, 40, 90–101.
- Naldi, L., Nilsson, P., Westlund, H., & Wixe, S. (2021). Amenities and new firm formation in rural areas. *Journal of Rural Studies*, 85, 32–42.
- OECD. (2019). *The value of culture and the creative industries in local development. Summer Academy of Cultural and Creative Industries and Local Development handbook*.
- ONS, Office for National Statistics. (2016). *2011 Rural/urban classification*. Office of National Statistics. <https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification>
- Paelinck, J. H. (2000). On aggregation in spatial econometric modelling. *Journal of Geographical Systems*, 2(2), 157–165.
- Paneels, I., Terras, M., Jones, C., Helgason, I., & Komorowski, M. (2021). Plugging the data gap: freelance workers in the creative industries. London: Policy and Evidence Centre for the Creative Industries. (last accessed <https://pec.ac.uk/blog/plugging-the-data-gap-freelance-and-self-employed-workers-in-the-creative-industries>).
- Papagiannidis, S., See-To, E. W. K., Assimakopoulos, D. G., & Yang, Y. (2018). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age? *Computers & Operations Research*, 98, 355–366.
- Phillip, L., & Williams, F. (2019). Remote rural home-based businesses and digital inequalities: Understanding needs and expectations in a digitally underserved community. *Journal of Rural Studies*, 68, 306–318.



- Power, T. M. (2005). The supply and demand for natural amenities: An overview of theory and concepts. In: Green, G. P., Deller, S. C. & Marcouiller, D.D. (eds) *Amenities and rural development: Theory, methods and public policy*, 63-77. Cheltenham: Edward Elgar Publishing.
- Pratt, A. C. (2021). Creative hubs: A critical evaluation. *City, Culture and Society*, 24(1), article 100384. <https://doi.org/10.1016/j.css.2021.100384>
- De Propris, L, Chapain, C., Cooke, P., MacNeill, S., & Mateos-Garcia, J. (2009) *The geography of creativity. Interim report: August 2009*. London: NESTA.
- Ray, C. (2001). *Culture economies*. Centre for Rural Economy. <https://www.ncl.ac.uk/media/wwwnclacuk/centreforruraleconomy/files/culture-economy.pdf>
- Ray, C. (2006). Neo-endogenous development in the EU. In *The handbook of rural studies* (pp. 278–292)., London: Sage plc..
- Rammer, C., Kinne, J., & Blind, K. (2020). Knowledge proximity and firm innovation: A microgeographic analysis for Berlin. *Urban Studies*, 57(5), 996–1014.
- Roberts, E., & Townsend, L. (2016). The contribution of the creative economy to the resilience of rural communities: Exploring cultural and digital capital. *Sociologia Ruralis*, 56(2), 197–219.
- Rodríguez-Pose, A., & Hardy, D. (2015). Cultural diversity and entrepreneurship in England and Wales. *Environment and Planning A: Economy and Space*, 47(2), 392–411.
- Rosenthal, S., & Strange, W. (2020). How close is close? The spatial reach of agglomeration economies. *Journal of Economic Perspectives*, 34(3), 27–49.
- Siepel, J., Camerani, R., Masucci, M, Velez-Ospina, J. A., Casadei, P, & Bloom, M. (2020). *Creative industries radar: Mapping the UK's creative clusters and microclusters*. Creative Industries Policy and Evidence Centre.
- Stich, C., Tranos, E., & Nathan, M. (2023). Modeling clusters from the ground up: A web data approach. *Environment and Planning B: Urban Analytics and City Science*, 50(1), 244-267.
- Tranos, E., & Mack, E. (2016). Broadband provision and knowledge-intensive firms: A causal relationship? *Regional Studies*, 50(7), 1113–1126. <https://doi.org/10.1080/00343404.2014.965136>
- Trippl, M., Grillitsch, M., Isaksen, A., & Sinozic, T. (2015). Perspectives on cluster evolution: Critical review and future research issues. *European Planning Studies*, 23, 2028–2044.
- Townsend, L., Wallace, C., Fairhurst, G., & Anderson, A. (2017). Broadband and the creative industries in rural Scotland. *Journal of Rural Studies*, 54, 451–458.
- Throsby, D. (2008). Modelling the cultural industries. *International Journal of Cultural Policy*, 14(3), 217–232.
- UNCTAD. (2018). *Creative economy outlook. Trends in international trade in creative industries 2002–2015*. New York.
- Valero, A., & Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68, 53-67.
- Van Oort, F. G. (2017). *Urban growth and innovation: Spatially bounded externalities in the Netherlands*. Routledge.
- Verditch, M. (2010). Creative migration? The attraction and retention of the ‘creative class’ in Launceston, Tasmania. *Australian Geographer*, 41(1), 129–140.
- Wallsten, S. J. (2001). An empirical test of geographic knowledge spillovers using geographic information systems and firm-level data. *Regional Science and Urban Economics*, 31(5), 571–599.

- Whitacre, B., Gallardo, R., & Strover, S. (2014). Does rural broadband impact jobs and income? Evidence from spatial and first-differenced regressions. *Annals of Regional Science*, 53, 649–670.
- Wixe, S. (2018). Neighbourhood related diversity, human capital and firm innovation. *Papers in Regional Science*, 97(2) 217–252.
- Wixe, S. and Andersson, M. (2017). Which types of relatedness matter in regional growth? Industry, occupation and education. *Regional Studies* 51(4), 532-536.
- Wojan, T. R., & Nichols, B. (2018). Design, innovation, and rural creative places: Are the arts the cherry on top, or the secret sauce? *PLOS ONE*, 13, e0192962.
- Wolman, H., & Hincapie, D. (2015). Clusters and cluster-based development policy. *Economic Development Quarterly*, 29(2), 135–149.
- Woodman, Richard, Sawyer, JE and Griffin, R.W (2006). Toward a Theory of Organizational Creativity. *Academy of Management Review* 18(2), pp. 293-321.

## Appendix

### Clustering method

We applied the HDBSCAN (hierarchical density-based spatial clustering for applications with noise; Campello et al. 2013) as our clustering method to identify microclusters. This method requires one parameter to tune in the HDBSCAN (the minimum cluster size). This parameter is endogenously determined by the algorithm. The HDBSCAN comprises of five steps according to McInnes et al. (2017):

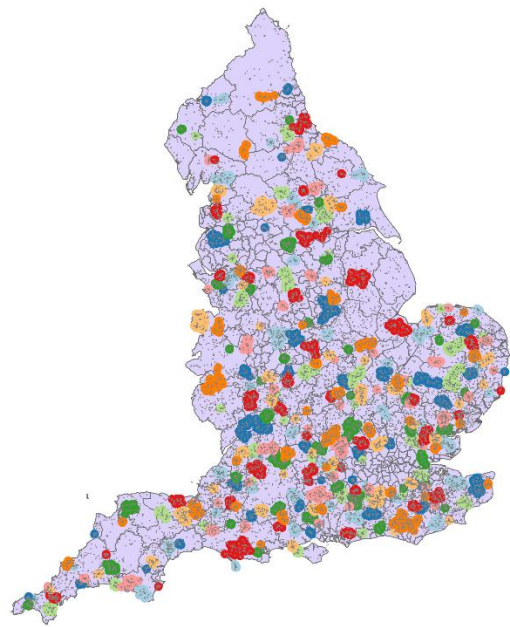
1. Compute the local density measure (number of firms) by defining a ‘mutual reachability’ distance, which is a new distance metric between firms;
2. Build a minimum spanning tree to implement single-linkage clustering or complete-linkage clustering; construct a cluster hierarchy of connected firms by iteratively sorting the edges of the tree by distance in an increasing order. The result can be viewed as a dendrogram that shows where robust single linkage stops;
3. Identify clusters in the hierarchy using a density threshold that is computed based on the density of firms in the sample;
4. Extract clusters that persist and are robust and label any remaining unclustered firms as noise.

Operationally, the HDBSCAN finds various clusters of different density levels and using a bottom-up fashion (single linkage), groups clusters, obtaining a tree that optimally represents the most significant clusters in terms of stability of the results.

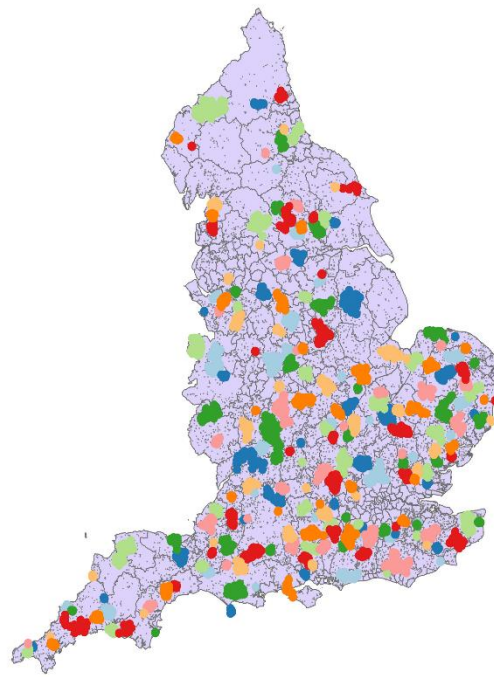
The only parameter to be specified is the minimum cluster size. Figure below presents examples of outputs when changing in the parameter. We can see that is the minimum cluster size is small, more microclusters are identified but there are also greater numbers of points labelled as noise (not part of any clusters); if the cluster size is larger (e.g, 500, 1000), more robust results emerge, although clusters are significantly larger, causing problems in interpretation. Following the literature and after multiple experimental results, 50 firms were used as the value for the minimum cluster size parameter, ensuring a higher number of microclusters but also being cognisant of smaller clusters that may be of relevance.

**Figure A1.** Outputs from sensitivity analysis

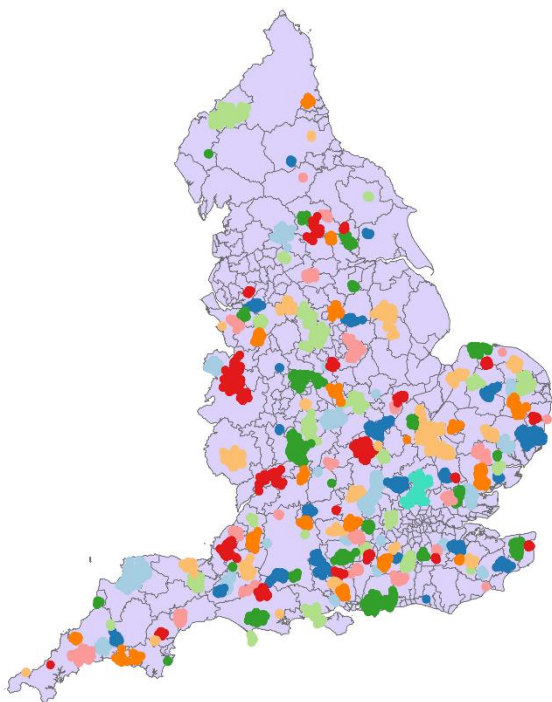
A) Min microcluster size = 20



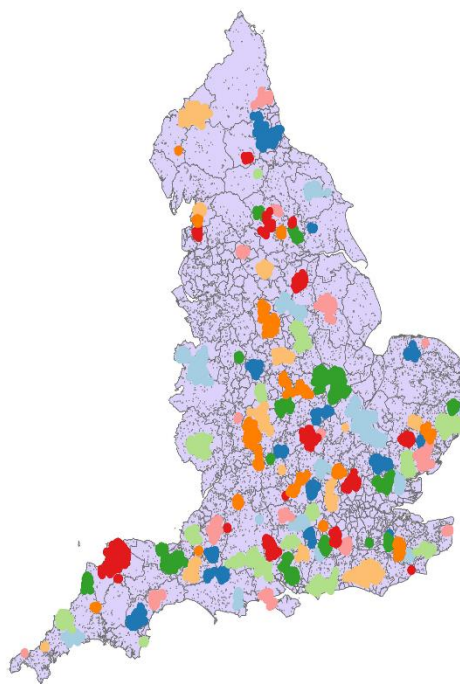
B) Min microcluster size = 25



C) Min microcluster size = 30



D) Min microcluster size = 40



Colours indicate the location of different microclusters

**Table A1a.** Sensitivity analysis (rural microclusters): Counts of microclusters and firms in microclusters when firm size threshold is varied

Minimum size	Threshold (Minimum number of firms to be a microcluster)				
	20 firms	25 firms	30 firms	40 firms	50 firms
Count (number of clusters)	307	216	170	105	71
Count (number of firms in microclusters)	10,166	10,100	9,411	10,803	9,634
% of all rural creative firms in microclusters	40%	40%	37%	42%	34%

**Table A1b: Sensitivity analysis (rural microclusters): Median and average number of companies within radii of companies, for different clustering thresholds**

		<b>Within a 1 Km radius</b>				
<b>Threshold (Minimum number of firms in a cluster)</b>		<b>20</b>	<b>25</b>	<b>30</b>	<b>40</b>	<b>50</b>
Median number of neighbours		10	9	8	7	7
Average number of neighbours		14	14	14	13	14
		<b>Within a 3 Km radius</b>				
<b>Threshold (Minimum number of firms in a cluster)</b>		<b>20</b>	<b>25</b>	<b>30</b>	<b>40</b>	<b>50</b>
Median number of neighbours		23	24	24	22	23
Average number of neighbours		26	27	28	26	28
		<b>Within a 5 Km radius</b>				
<b>Threshold (Minimum number of firms in a cluster)</b>		<b>20</b>	<b>25</b>	<b>30</b>	<b>40</b>	<b>50</b>
Median number of neighbours		36	38	39	40	42
Average number of neighbours		42	43	44	44	46
		<b>Within an 8 Km radius</b>				
<b>Threshold (Minimum number of firms in a cluster)</b>		<b>20</b>	<b>25</b>	<b>30</b>	<b>40</b>	<b>50</b>
Median number of neighbours		68	69	69	68	69
Average number of neighbours		77	75	78	78	77

**Table A2.** Variable definition

Variable	Description	Unit of observation	Period
<b>Dependent variables</b>			
<i>Stock</i>	Microcluster stock: Number of creative firms in microclusters.	LSOA	2019
<b>Amenities</b>			
<i>ln_cultural_inst</i>	Number of museums, public galleries, libraries, archives, heritage site and science centres within 1km	LSOA	2017
<i>ln_cultural_network</i>	The number of organisations dedicated to exhibitions, campaigns and initiatives, festivals, cultural and scientific meetings within 1km	LSOA	2017
<i>ln_nature</i>	Number of environmental and ecological centres, and agricultural sites within 1km	LSOA	2017
<i>ln_distance_HEI</i>	Distance to the nearest university or college	LSOA	2017
<b>Agglomeration, industry, and size</b>			
<i>RD</i>	Related diversity: $-\sum_{j=1}^J E_j \ln H_j$ where $H_j = -\sum_{i=1}^I E_{ij} \ln E_{ij}$ and $E_{ij}$ denotes the share of employees that work in each 5-digit SIC industry $l$ , where the share is measured within each 2-digit SIC industry $j$ .	LSOA	2017
<i>UD</i>	Unrelated diversity: $-\sum_{j=1}^J E_j \ln \ln E_j$ , where $E_j$ denotes the share of total employment in each LSOA that belongs to the same 2-digit SIC code industry $j$ .	LSOA	2017
<i>LQ_Cls</i>	Employment-based Location quotient for Creative Industries. $LQ_{s,r} = \frac{e_{s,r}}{e_s/e}$ where $e_{s,r}$ denotes the number of employees in each LSOA $r$ and industry $s$ . $e_r$ denotes all employees in the LSOA, $e_s$ the number of employees in industries $s$ in England, and $e$ the total number of employees in England	LSOA	2017
<i>LQ_manuf</i>	Employment-based Location quotient for Manufacturing industries	LSOA	2017
<i>LQ_serv</i>	Employment-based Location quotient for Service industries	LSOA	2017
<i>LQ_know</i>	Employment-based Location quotient for knowledge-based industries.	LSOA	2017
<i>ln_internet_speed</i>	Average broadband download speed	LSOA	2017

Variable	Description	Unit of observation	Period
<b>Other controls</b>			
<i>ln_pop</i>	Population density: People per kilometre square KM2	LSOA	2017
<i>unemp</i>	Unemployment rate: Annual rate (Sep-Sep) in %	District	2017
<i>rel_affluence</i>	Wealth decile (multiple deprivation index)	LSOA	2017
<i>distance2city</i>	Distance to the main city in miles.	LSOA	2017
<i>ln_rate</i>	Rateable value per m <sup>2</sup>	LSOA	2017
<i>ln_rate2</i>	Square of the rateable value per m <sup>2</sup>	LSOA	2017
<i>Fixed effects</i>	TTWA: Travel to work areas fixed effects. 228 Dummies.	TTWA	

Notes: LSOA refers to Lower Super Output Area. MSOA refers to Middle Super Output area. Location quotients are based on employment figures from the Business register and employment survey. Sectors were classified using SIC codes as follow: Creative firms: DCMS classification (3212, 5811, 5812, 5813, 5814, 5819, 5821, 5829, 5911, 5912, 5913, 5914, 5920, 6010, 6020, 6201, 6202, 7021, 7311, 7312, 7111, 7410, 7420, 7430, 8552, 9001, 9002, 9003, 9004, 9101, 9102); Manufacturing industries (10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 36, 37, 38, 39, 41, 42, 43); Service activities (45, 46, 47, 49, 50, 51, 52, 53, 55, 56, 61, 64, 65, 66, 68, 69, 70, 7112, 7120, 7490, 75, 77, 79, 80, 81, 82, 84, 86, 87, 88, 92, 9311, 9313, 9319, 94, 95, 96, 97, 98, 99); knowledge based activities (85).



**Table A3.** Pairwise correlations (Full sample: rural and urban LSOAs)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) ln_pop	1.00										
(2) unemp	-0.01	1.00									
(3) distance2capital	0.05	-0.15	1.00								
(4) rel_affluence	0.00	-0.22	-0.30	1.00							
(5) ln_rate	-0.06	-0.23	-0.16	0.23	1.00						
(6) ln_rate2	0.08	-0.03	0.10	-0.08	0.11	1.00					
(7) ln_distance_HEI	0.05	-0.15	0.53	-0.20	-0.13	0.11	1.00				
(8) ln_cultural_network	0.04	-0.02	0.04	0.00	0.02	0.15	0.06	1.00			
(9) ln_nature	0.09	-0.09	0.05	-0.03	0.00	0.15	0.02	0.04	1.00		
(10) RD	0.19	-0.08	0.01	-0.01	0.24	0.20	0.05	0.07	0.05	1.00	
(11) UD	0.22	-0.15	-0.01	0.09	0.02	0.08	0.06	0.02	0.13	0.32	1.00
(12) LQ_ci	-0.02	-0.09	-0.15	0.21	0.13	0.00	-0.08	-0.02	0.07	-0.02	0.13
(13) LQ_serv	-0.01	-0.01	0.02	-0.04	0.23	0.11	0.02	0.04	0.00	0.16	0.06
(14) LQ_manuf	0.04	0.04	0.05	-0.08	-0.32	-0.11	0.04	-0.04	-0.07	-0.10	0.10
(15) LQ_know	-0.02	0.02	-0.05	0.06	0.00	-0.03	-0.04	-0.01	-0.01	-0.06	-0.27
(16) ln_internet_speed	0.01	-0.05	-0.13	0.22	0.16	0.02	-0.14	0.00	-0.05	0.01	-0.04

Accepted Article

**Table A4.** Negative binomial Regressions explaining the number of firms in microclusters (at LSOA), microcluster minimum size of 25 firms

	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
ln_cultural_inst	0.567*** (0.0662)	0.547*** (0.0670)	0.548*** (0.0679)	0.427*** (0.0651)	1.146*** (0.0418)	1.047*** (0.0432)	1.032*** (0.0435)	0.620*** (0.0505)
ln_cultural_network		0.470** (0.169)	0.476** (0.171)	0.439** (0.167)		0.929*** (0.116)	0.902*** (0.116)	0.699*** (0.151)
ln_nature			0.0799 (0.0911)	0.0631 (0.0907)			0.245* (0.122)	0.151 (0.136)
ln_distance_HEI			-0.340*** (0.0837)	-0.342*** (0.0783)			-0.184*** (0.0455)	-0.199*** (0.0530)
RD				2.225*** (0.167)				2.101*** (0.105)
UD				0.328*** (0.0813)				0.931*** (0.0617)
LQ_cis				0.0535 (0.0306)				0.0186 (0.0218)
LQ_serv				-0.644* (0.269)				-0.258 (0.331)
LQ_manuf				-0.0661 (0.0578)				-0.155* (0.0684)
LQ_know				-0.105** (0.0396)				-0.140** (0.0452)
ln_internet_speed				0.0300 (0.0776)				-0.251** (0.0867)
ln_pop	0.700*** (0.152)	0.693*** (0.152)	0.632*** (0.150)	0.189 (0.149)	0.656*** (0.148)	0.661*** (0.148)	0.640*** (0.148)	0.0232 (0.169)
unemp	-0.0164 (0.0618)	-0.0141 (0.0619)	-0.0182 (0.0616)	0.0120 (0.0575)	-0.0843** (0.0297)	-0.0850** (0.0296)	-0.0945** (0.0296)	-0.0489 (0.0277)
distance2city	-0.256*** (0.0604)	-0.260*** (0.0603)	-0.124 (0.0697)	-0.114 (0.0644)	0.278*** (0.0344)	0.285*** (0.0343)	0.349*** (0.0377)	0.323*** (0.0404)
rel_affluence	0.0669***	0.0665***	0.0632***	0.0792***	-0.0915***	-0.0905***	-0.0850***	-0.0519***

	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
	(0.0172)	(0.0172)	(0.0173)	(0.0164)	(0.00979)	(0.00979)	(0.00987)	(0.0102)
In_rate	-0.575	-0.673	-0.667	-1.507*	0.350	0.365	0.462	-0.779
	(0.795)	(0.792)	(0.793)	(0.758)	(0.335)	(0.339)	(0.347)	(0.483)
In_rate2	0.199*	0.212*	0.212*	0.281**	0.0481	0.0421	0.0269	0.138*
	(0.0992)	(0.0989)	(0.0989)	(0.0938)	(0.0384)	(0.0388)	(0.0398)	(0.0538)
N	4702	4702	4702	4702	18765	18765	18765	18765
In_rate, In_rate2 = 0	52.73***	48.93***	50.09***	53.31***	76.76***	60.84***	60.28***	34.73***
TTWA=0 (Chi2)	60.50***	58.20***	56.24***	121.07***	123.08***	233.76***	543.49***	41.60***
chi2	53669.4	57107.5	57376.5	47726.9	10664.4	11378.8	10733.0	11341.0
Mean VIF	3.76	3.75	3.73	3.76	3.67	3.63	3.57	3.62
r2_p	0.09	0.14	0.14	0.17	0.1385	0.1394	0.10	0.12

Notes: As Table 6

**Table A5.** Negative binomial Regressions explaining the number of firms in microclusters (at MSOA)

	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
ln_cultural_inst	0.825*** (0.158)	0.789*** (0.161)	0.810*** (0.164)	0.777*** (0.168)	1.127*** (0.0601)	0.923*** (0.0665)	0.895*** (0.0662)	0.546*** (0.0706)
ln_cultural_network		0.394 (0.366)	0.365 (0.357)	0.617 (0.363)		1.087*** (0.125)	1.043*** (0.126)	0.741*** (0.122)
ln_nature			-0.428 (0.332)	-0.375 (0.340)			0.194 (0.165)	0.306 (0.165)
ln_distance_HEI			-0.147** (0.193)	-0.179*** (0.204)			-0.121 (0.0753)	-0.200* (0.0779)
RD				3.406*** (0.830)				1.995*** (0.280)
UD				-0.468 (0.478)				0.613*** (0.182)
LQ_cis				-0.0443 (0.294)				0.190* (0.0838)
LQ_serv				-4.785* (1.967)				-0.557 (0.996)
LQ_manuf				-0.734 (0.398)				-0.344 (0.209)
LQ_know				-1.149*** (0.297)				-0.518*** (0.132)
ln_internet_speed				0.312 (0.602)				-0.113 (0.190)
ln_pop	0.851*** (0.170)	0.866*** (0.171)	0.770*** (0.171)	0.869*** (0.194)	0.133 (0.0743)	0.0976 (0.0743)	0.0960 (0.0759)	0.250*** (0.0752)
unemp	-1.062*** (0.201)	-1.078*** (0.201)	-1.131*** (0.207)	-1.057*** (0.214)	-0.132* (0.0660)	-0.143* (0.0663)	-0.132* (0.0654)	-0.163** (0.0605)
distance2city	-0.179** (0.046)	-0.173** (0.046)	-0.214** (0.057)	0.115** (0.056)	0.613*** (0.0698)	0.616*** (0.0708)	0.670*** (0.0773)	0.689*** (0.0763)
rel_affluence	0.396*** (0.0911)	0.397*** (0.0908)	0.385*** (0.0898)	0.395*** (0.0905)	-0.0477 (0.0256)	-0.0486 (0.0259)	-0.0458 (0.0261)	-0.00871 (0.0261)

In_rate	-10.11*	-10.97*	-11.59*	-11.37*	2.704*	2.574	2.732*	-1.060
	(4.644)	(4.806)	(4.834)	(4.528)	(1.291)	(1.334)	(1.343)	(1.410)
In_rate2	1.557**	1.658**	1.759**	1.673**	-0.173	-0.172	-0.192	0.173
	(0.590)	(0.612)	(0.615)	(0.561)	(0.147)	(0.153)	(0.154)	(0.158)
N	1160	1160	1160	1160	5362	5362	5362	5362
In_rate, In_rate2 = 0	34.04***	31.36***	36.19***	26.51***	76.94***	61.82***	60.37***	9.90***
TTWA=0 (Chi2)	31.80***	31.79***	32.31***	55.81***	86.25***	86.55***	89.12***	183.07***
chi2	127.3***	120.5***	134.3***	128.8***	1285.5***	1532.3***	1949.2***	1916.3***
Mean VIF	4.37	5.29	5.26	5.40	4.37	5.34	5.36	5.65
r2_p	0.0820	0.0822	0.0832	0.0930	0.260	0.281	0.282	0.443

Notes: Dependent variable: Number of creative firms in a cluster at the MSOA. Robust standard errors in parenthesis. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The number of microclusters, social capital, natural and cultural amenities were winsorised. Treatment of outliers consisted of transforming the continuous variables to  $\ln(x+1)$ .

Accepted Article

**Table A6.** Total effects from spatial regression models explaining the number of firms in microclusters (at MSOAs)

	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
ln_cultural_inst	0.239*** (0.096)	0.157*** (0.0544)	0.311** (0.129)	0.094*** (0.031)	0.552*** (0.044)	0.395*** (0.0275)	0.621*** (0.065)	0.394*** (0.085)
ln_cultural_network	0.272* (0.160)	0.157 (0.0986)	0.322* (0.190)	0.399** (0.108)	0.649*** (0.083)	0.524*** (0.0510)	0.741*** (0.096)	0.263*** (0.151)
ln_distance_HEI	-0.153** (0.077)	-0.0899** (0.386)	-0.159** (0.055)	-0.093** (0.039)	-0.0221 (0.034)	-0.0957*** (0.0307)	-0.021 (0.041)	-0.078*** (0.045)
RD	0.720* (0.381)	0.464** (0.233)	0.936** (0.497)	0.340** (0.146)	1.049*** (0.153)	0.914*** (0.0857)	1.561*** (0.177)	2.033*** (0.257)
UD	-0.125 (0.211)	-0.0521 (0.136)	-0.166 (0.262)	-0.124 (0.140)	0.185** (0.078)	0.187*** (0.0524)	0.234*** (0.085)	0.167*** (0.054)
LQ_cis	0.068 (0.094)	0.0774 (0.0620)	0.099 (0.117)	0.006 (0.170)	0.044 (0.030)	0.069*** (0.0216)	0.059* (0.033)	0.033*** (0.010)
LQ_serv	-0.639* (0.380)	-0.518** (0.235)	-0.896* (0.497)	-0.457* (0.246)	0.726*** (0.182)	0.408*** (0.123)	0.722*** (0.199)	1.281*** (0.330)
LQ_know	-0.161** (0.0800)	-0.108** (0.0490)	-0.201** (0.101)	-0.105* (0.052)	-0.095*** (0.028)	-0.0724*** (0.0186)	-0.106*** (0.031)	-0.064 (0.052)
ln_internet_speed	-0.344**	-0.224*	-0.385*	-0.379*	0.132	0.0849	0.154	0.111

This article is protected by copyright. All rights reserved.

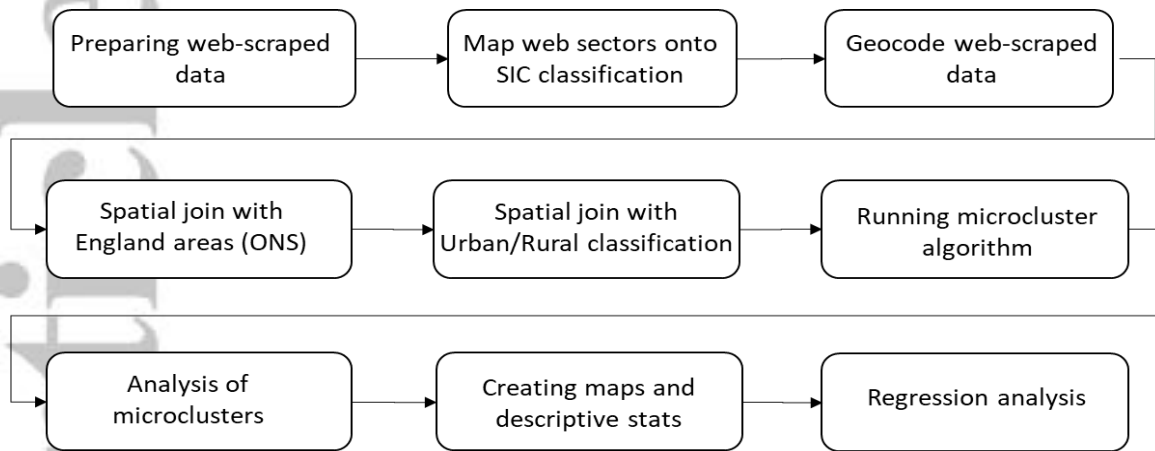
	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
	(0.165)	(0.125)	(0.218)	(0.214)	(0.081)	(0.0696)	(0.095)	(0.096)
Ln_pop	0.305***	0.155***	0.403***	0.231**	0.166***	0.136***	0.192***	0.256***
	(0.094)	(0.0474)	(0.140)	(0.1105)	(0.034)	(0.0242)	(0.039)	(0.055)
	(0.061)	(0.0480)	(0.079)	(0.079)	(0.0137)	(0.0183)	(0.024)	(0.024)
distance2city	0.182**	0.169**	0.186***	0.082**	0.287***	0.232***	0.316***	0.243***
	(0.0342)	(0.0410)	(0.088)	(0.034)	(0.031)	(0.0245)	(0.037)	(0.0252)
ln_rate	0.768***	0.603***	0.977***	0.685***	-0.046	0.115**	0.024	0.121**
	(0.219)	(0.133)	(0.314)	(0.286)	(0.065)	(0.0497)	(0.073)	(0.0506)
rel_affluence	0.074*	0.0583**	0.085*	0.090**	-0.016	-0.0219**	-0.026*	-0.043**
	(0.039)	(0.0282)	(0.009)	(0.046)	(0.012)	(0.00912)	(0.013)	(0.019)
<i>N</i>	1160	1160	1160	1160	5362	5362	5362	5362
Estimation								
$\rho W y \neq 0$	Yes	No	Yes	Yes	Yes	No	Yes	Yes
$\lambda W e \neq 0$	No	Yes	Yes	No	No	Yes	Yes	No
$WX\theta \neq 0$	No	No	No	Yes	No	No	No	Yes
r2_p	0.0911	0.0947	0.0779	0.113	0.218	0.224	0.208	0.252
Mean VIF	1.37	1.37	1.37	1.37	1.72	1.72	1.72	1.72
chi2	151.3***	117.7***	152.7***	122.3***	1808.8***	1648.3***	1777.1***	1677.1***
chi2_c	9.819***	117.3***	103.9***	125.4***	47.71***	689.2***	531.1***	714.1***

This article is protected by copyright. All rights reserved.

	Rural				Urban			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock	mc_stock
TTWA=0 (Chi2)	61.33***	108.22***	34.37***	138.38***	49.98***	41.97***	157.50***	145.48***
AIC	3690.75	3685.4	3683.48	3680.15	17107.48	17029.88	17031.84	16927.43
BIC	3776.71	3771.35	3771.15	3745.83	17219.46	17141.86	17150.4	17125.04

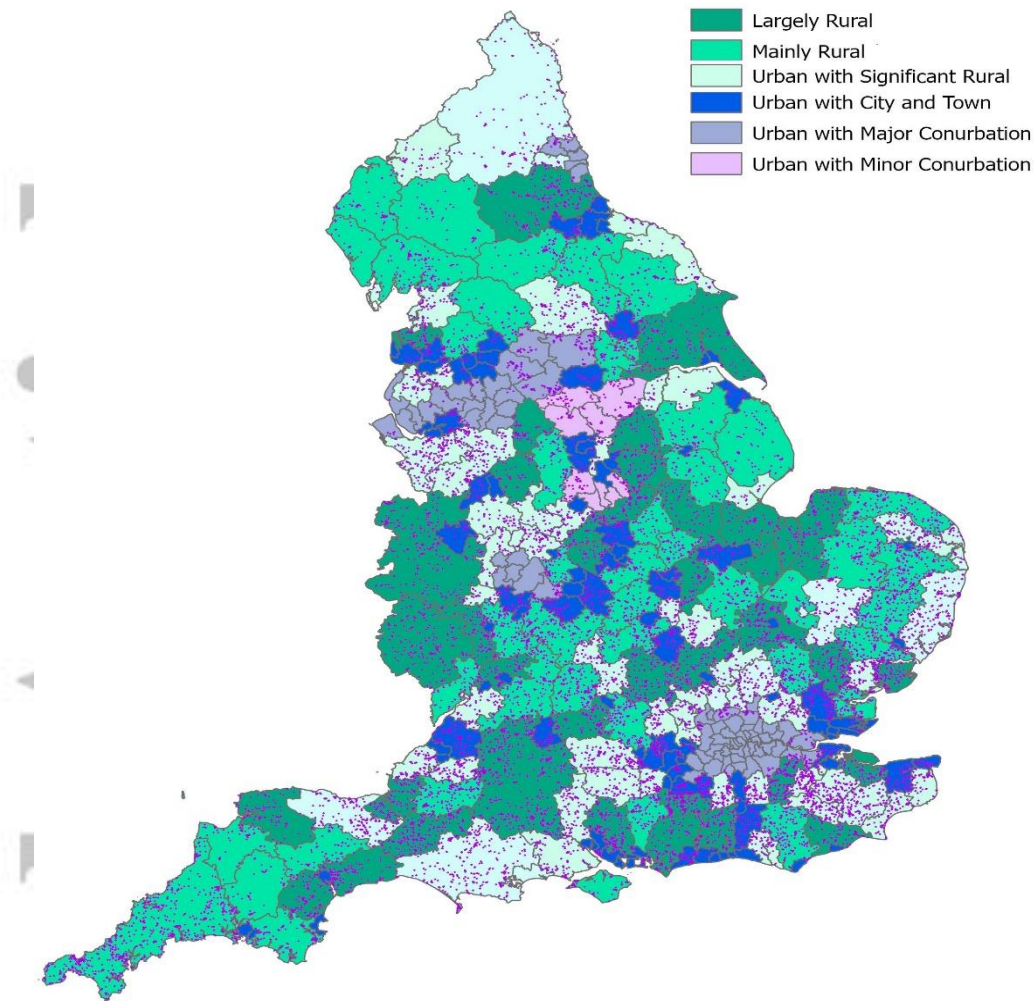
Notes: Delta method standard errors in parenthesis. TTWA fixed effects included in all models \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.010$





**Figure 1:** Analytical framework for web-scraped data and spatial analysis

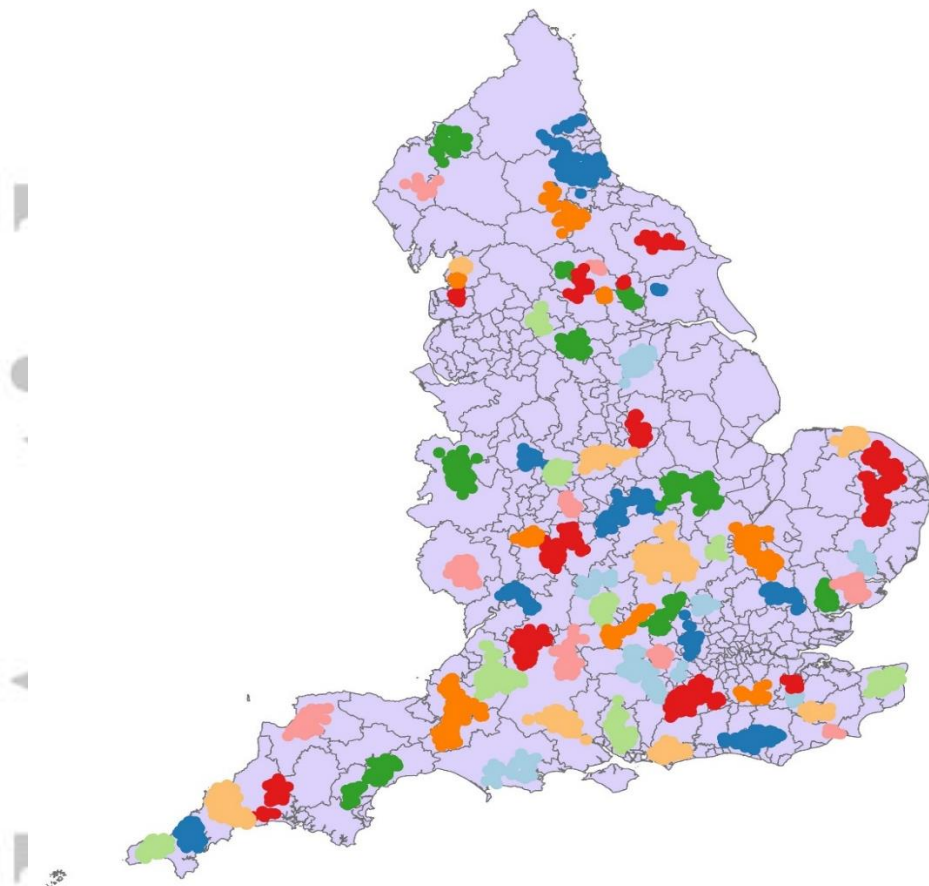
Accepted Article



**Figure 2:** Map of rural creative industries

Note: Map boundaries correspond to local authority districts in England. Each dot represents one of the 21,124 companies. The categories used within the classification of urban and rural correspond to ONS 2011 Rural-Urban classification of local authority districts in England. Further details: <https://www.gov.uk/government/statistics/2011-rural-urban-classification-of-local-authority-and-other-higher-level-geographies-for-statistical-purposes>

Accepted



**Figure 3:** Map of microclusters identified

Note: Map boundaries correspond to local authority districts in England. Each colour represents a visually distinct cluster.

Accepted

**Table 1:** Distribution of creative websites with addresses, by sector category and DCMS subsector

	Count websites with address	% share all websites with address	Count websites	% share all websites	% of all websites with address
Marketing and advertising	17,702	9%	36,454	10%	49%
Public relations and communications	7,083	4%	14,591	4%	49%
<b>All advertising</b>	<b>24,785</b>	<b>13%</b>	<b>51,045</b>	<b>14%</b>	<b>49%</b>
<b>Architecture</b>	<b>22,143</b>	12%	<b>44,726</b>	12%	<b>50%</b>
<b>Crafts</b>	<b>13,754</b>	<b>7%</b>	<b>33,488</b>	<b>9%</b>	<b>41%</b>
Design	18,754	10%	56,330	15%	33%
Apparel and fashion	17,599	9%	39,830	11%	44%
<b>All design</b>	<b>36,353</b>	<b>19%</b>	<b>96,160</b>	<b>26%</b>	<b>38%</b>
Animation	489	0%	2,355	1%	21%
Broadcast media (TV, radio)	3,635	2%	9,071	2%	40%
Media production	5,150	3%	21,547	6%	24%
Motion pictures and film	2,548	1%	10,957	3%	23%
Photography	6,761	4%	29,061	8%	23%
<b>All film, TV, radio and photography</b>	<b>18,583</b>	<b>10%</b>	<b>72,991</b>	<b>19%</b>	<b>25%</b>
Computer games	2,022	1%	8,637	2%	23%
Computer software	19,098	10%	33,023	9%	58%
<b>All IT and software</b>	<b>21,120</b>	<b>11%</b>	<b>41,660</b>	<b>11%</b>	<b>51%</b>
Publishing	10,144	5%	19,217	5%	53%
Newspapers and magazines	3,636	2%	5,218	1%	70%
Translation	2,217	1%	4,600	1%	48%
Writing and editing	1,267	1%	6,120	2%	21%
Online media	616	0%	1,447	0%	43%
<b>All publishing</b>	<b>17,880</b>	<b>10%</b>	<b>36,602</b>	<b>10%</b>	<b>49%</b>
<b>TOTAL</b>	<b>187,872</b>		<b>376,672</b>		<b>43%</b>

**Table 2:** Median count of rural creative firms at different radii

<b>Radius (km)</b>	<b>Average count</b>	<b>Median count</b>	<b>Minimum</b>	<b>Maximum</b>
1	14	8	1	62
3	27	23	1	134
5	44	39	1	182
8	78	69	1	272

Note: Hotspot analysis was carried out to estimate the number of neighbours at different distance bands.

Accepted Article

**Table 3:** Rural and urban creative firms in clusters by DCMS sector classification

DCMS	Rural creative firms			Urban creative firms		
	Firms in microclusters	Total firms in sample	Per cent firms in microclusters	Firms in microclusters	Total firms in sample	Per cent firms in microclusters
Advertising and marketing	1079	2754	39%	10678	22031	48%
Architecture	1381	3627	38%	7652	18516	41%
Crafts	906	2515	36%	4825	11239	43%
Design	1879	5141	37%	13172	31212	42%
Film, TV, video, radio and photography	984	2717	36%	7473	17888	42%
IT, software, video games <sup>a</sup>	931	2011	46%	8147	17087	48%
Publishing and translation	936	2359	39%	6819	15521	44%
<b>Grand total</b>	<b>8,096</b>	<b>21,124</b>	<b>38%</b>	<b>58,766</b>	<b>133,494</b>	<b>44%</b>

Note: <sup>a</sup> Subsector also includes computer services.

**Table 4:** Descriptive statistics of main variables

Var	Description	Rural = 4,702				Urban = 18,765			
		Mean	Std.D	Min	Max	Mean	Std.D	Min	Max
<i>Dependent variable</i>									
mc_stock	Stock of firms in microclusters (#)	1.45	4.2	0	60	2.39	12.92	0	607
<i>Amenities</i>									
cultural_inst	Number of cultural institutions within 1km	.44	.96	0	6	1.1	1.45	0	33
ln_cultural_inst	Cultural institutions within 1km (log)	.24	.44	0	1.95	.16	.42	0	3.49
Cultural_networks	Number of social organisations within 1km (#)	.06	.24	0	1	.07	.37	0	6
ln_cultural_network	Social organisations (log)	.04	.17	0	.69	.04	.19	0	1.79
Nature	Number of nature-based amenities within 1km (#)	.29	.63	0	8	.08	.3	0	6
ln_nature	Nature-based amenities within 1km (log)	.18	.35	0	2.2	.05	.19	0	1.95
distance_HEI	Distance to the nearest university (miles)	12.4	7.27	0.56	52.5	4.65	4.92	.01	52.36
ln_distance_HEI	Distance to the nearest university (log)	2.46	.54	.05	3.98	1.47	.69	.01	3.98
<i>Agglomeration, industry, and size</i>									
RD	Related diversity	.33	.18	-.14	1.13	.33	.23	-.08	1.22
UD	Unrelated diversity	2.37	.44	.23	3.32	2.07	.55	0	3.44
LQ cis	Location quotient of creative industries	.87	1.27	0	22.85	1.08	1.67	0	28.14
LQ serv	Location quotient of services-based industries	.89	.25	0	1.42	.94	.31	0	1.57
LQ manuf	Location quotient of manufacturing industries	1.37	1.08	0	6.98	.98	1.07	0	7.13

This article is protected by copyright. All rights reserved.

Var	Description	Rural = 4,702				Urban = 18,765			
		Mean	Std.D	Min	Max	Mean	Std.D	Min	Max
LQ know	Location quotient of knowledge-based activities	1.24	1.43	0	9.03	1.54	2.17	0	11.09
Internet_speed	Average download speed (megas)	34.76	31.44	2.65	684.3	52.18	16.5	4.08	196.9
In_internet_speed	Average download speed (log)	3.46	.42	1.29	6.53	3.92	.35	1.63	5.29
Pop	People per square km	1639.2	337.5	995	2741	1625.5	264.9	983	2737
In_pop	People per square km (log)	7.38	.2	6.9	7.92	7.38	.16	6.89	7.91
unemp	Unemployment rate (%)	3.49	.94	2	7.9	4.57	1.39	2	8.6
rel_affluence	Local area wealth decile	6.81	2.09	1	10	5.48	2.88	1	10
distance2city	Distance to the major town/city kms)	16.22	14.87	0	110.6	4.61	10.13	0	101.6
distance2city	Distance to the major town/city (log)	2.51	.85	0	4.72	.85	1.17	0	4.63
Rate	Office rateable value per m <sup>2</sup> (£)	52.56	26.75	7	342	87.95	55.87	0	427
In_rate	Office rateable value per m <sup>2</sup> (log)	3.88	.44	2.08	5.84	4.32	.58	0	6.06

Note: Table A2 in the Appendix provides further definitions of these variables



**Table 5:** Negative binomial regressions explaining the number of firms in microclusters (within 1km of the LSOA centroid)

	Rural				Urban			
	(1) mc_stoc k	(2) mc_stoc k	(3) mc_stoc k	(4) mc_stock	(5) mc_stoc k	(6) mc_stock	(7) mc_stock	(8) mc_stock
ln_cultural_inst	0.628*** (0.0897)	0.611*** (0.0911)	0.617*** (0.0920)	0.485*** (0.0858)	1.383*** (0.0478)	1.197*** (0.0511)	1.153*** (0.0517)	0.684*** (0.0512)
ln_cultural_netw ork		0.475 (0.244)	0.537* (0.253)	0.628* (0.248)		1.479*** (0.135)	1.393*** (0.138)	1.029*** (0.123)
ln_nature			0.0684 (0.118)	-0.0103 (0.115)			0.319* (0.161)	0.143 (0.133)
ln_distance_HEI			-0.648*** (0.112)	-0.701*** (0.107)			-0.415*** (0.0571)	-0.366*** (0.0581)
RD				1.120*** (0.239)				2.150*** (0.113)
UD				0.512*** (0.105)				0.921*** (0.0607)
LQ_cis				0.0580 (0.0344)				0.0149 (0.0224)
LQ_serv				-0.735* (0.337)				-0.510 (0.343)
LQ_manuf				-0.0534 (0.0737)				-0.299*** (0.0708)
LQ_know				-0.0844 (0.0494)				-0.126** (0.0471)
ln_internet_spee d				-0.0434 (0.124)				-0.273** (0.0892)
ln_pop	0.689*** (0.202)	0.674*** (0.201)	0.580** (0.200)	0.215 (0.202)	0.375* (0.173)	0.384* (0.176)	0.371* (0.176)	-0.159 (0.162)
unemp	-0.285*** (0.0821)	-0.280*** (0.0819)	-0.288*** (0.0800)	-0.247** (0.0764)	-0.131** (0.0421)	-0.143*** (0.0421)	-0.165*** (0.0414)	-0.115** (0.0354)
distance2city	-0.314*** (0.0864)	-0.312*** (0.0862)	-0.291*** (0.047)	-0.282*** (0.059)	0.344*** (0.0408)	0.352*** (0.0405)	0.494*** (0.0458)	0.464*** (0.0479)
rel_affluence	0.0437*** (0.014)	0.0418*** (0.013)	0.0443** (0.0153)	0.0537** (0.0221)	-0.108** (0.0115)	-0.104*** (0.0115)	-0.0945** (0.0118)	-0.0516*** (0.0109)
ln_rate	-1.710 (0.996)	-1.885 (0.990)	-1.811 (0.988)	-2.741** (1.012)	-0.518 (0.408)	-0.390 (0.417)	-0.293 (0.442)	-1.416** (0.488)
ln_rate2	0.363** (0.125)	0.384** (0.125)	0.376** (0.124)	0.478*** (0.127)	0.180*** (0.0474)	0.157** (0.0485)	0.139** (0.0511)	0.214*** (0.0542)
N	4702	4702	4702	4702	18765	18765	18765	18765
ln_rate, ln_rate2 = 0	149.3***	147.5***	160.4***	124.4***	290.5***	245.8***	232.8***	67.23***

TTWA=0 (Chi2)	46.91 <sup>***</sup>	59.64 <sup>***</sup>	95.61 <sup>***</sup>	33.28 <sup>***</sup>	168.94 <sup>**</sup> *	166.8 <sup>***</sup>	228.9 <sup>***</sup>	178.6 <sup>***</sup>
chi2	33666.3	43566.3	67653.5	58421.9	31390.0	32044.0	32690.39	30995.35
r2_p	0.154	0.238	0.248	0.260	0.112	0.189	0.208	0.237
Mean VIF	3.76	3.75	3.74	3.77	3.67	3.63	3.59	3.65

Notes: Dependent variable: Number of creative firms in a cluster at the LSOA. Robust standard errors in parenthesis. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The number of microclusters, social capital, natural and cultural amenities were winsorised at 2.5% of non-zero values to mitigate the leverage of outliers. Treatment of zeros consisted of transforming the continuous variables to  $\ln(x+1)$ .

**Table 6:** Explaining the number of firms in microclusters by DCMS sectors: Rural LSOAs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Advertising and marketing	Architecture	Crafts	Design	Film, TV, radio	IT, software	Publishing
ln_cultural_inst	0.351*** (0.103)	0.342*** (0.0946)	0.659*** (0.105)	0.285** (0.0881)	0.326** (0.101)	0.247* (0.124)	0.515*** (0.106)
ln_cultural_network	0.247 (0.256)	0.277 (0.209)	0.456 (0.276)	0.0608 (0.219)	0.548* (0.259)	0.469 (0.332)	0.208 (0.291)
ln_nature	0.130 (0.134)	0.423*** (0.119)	0.422** (0.148)	0.228* (0.111)	0.320* (0.133)	0.00802 (0.159)	0.175 (0.140)
ln_distance_HEI	-0.436*** (0.116)	-0.603*** (0.0990)	-0.487*** (0.122)	-0.545*** (0.0934)	-0.549*** (0.105)	-0.478*** (0.123)	-0.511*** (0.130)
RD	1.465*** (0.278)	1.215*** (0.264)	1.512*** (0.295)	1.657*** (0.235)	0.793** (0.273)	1.894*** (0.326)	0.933** (0.347)
UD	0.335* (0.141)	0.854*** (0.150)	0.716*** (0.164)	0.506*** (0.118)	0.538*** (0.141)	0.667*** (0.171)	0.667*** (0.171)
LQ_cis	0.0346 (0.0335)	0.0382 (0.0346)	0.0263 (0.0405)	0.0274 (0.0287)	0.0173 (0.0371)	0.147*** (0.0363)	0.0179 (0.0369)
LQ_serv	-0.767 (0.401)	-0.863* (0.406)	-1.254* (0.521)	-0.491 (0.353)	-0.816* (0.403)	-1.155** (0.419)	-0.905 (0.468)
LQ_manuf	-0.0934 (0.0908)	-0.0784 (0.0886)	-0.197 (0.110)	0.00783 (0.0776)	-0.166 (0.0895)	-0.0293 (0.0943)	-0.153 (0.104)
LQ_know	-0.214*** (0.0620)	-0.147* (0.0626)	-0.0818 (0.0750)	-0.194*** (0.0541)	-0.191** (0.0608)	-0.209** (0.0655)	-0.143* (0.0722)
ln_internet_speed	-0.126 (0.143)	-0.0332 (0.117)	-0.299* (0.145)	-0.108 (0.108)	-0.0250 (0.128)	-0.102 (0.138)	-0.238 (0.147)
ln_pop	0.682** (0.236)	0.406 (0.223)	0.0869 (0.276)	0.392 (0.202)	0.446 (0.233)	0.567 (0.293)	0.416 (0.272)
Unemp	-0.205* (0.0904)	-0.145 (0.0868)	-0.307** (0.0973)	-0.271*** (0.0753)	-0.126 (0.0788)	-0.328** (0.107)	-0.0486 (0.0902)
distance2city	-0.0904 (0.0929)	-0.119 (0.0853)	0.240** (0.100)	0.0622** (0.0195)	-0.0685 (0.0854)	-0.127 (0.102)	0.124*** (0.094)
rel_affluence	0.0361*** (0.023)	0.0388*** (0.008)	-0.0272 (0.010)	0.0405 (0.0221)	0.0290 (0.0263)	0.0318** (0.005)	0.0596 (0.0335)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Advertising and marketing	Architecture	Crafts	Design	Film, TV, radio	IT, software	Publishing
ln_rate	-1.643 (1.177)	-3.903*** (1.152)	-2.192 (1.566)	-1.488 (1.082)	-0.742 (1.127)	-4.787*** (1.433)	-1.935 (1.288)
ln_rate2	0.325* (0.142)	0.588*** (0.142)	0.361 (0.190)	0.280* (0.134)	0.167 (0.139)	0.718*** (0.173)	0.336* (0.159)
N	4702	4702	4702	4702	4702	4702	4702
ln_rate, ln_rate2 = 0	61.82***	64.59***	33.18***	53.29***	28.52***	70.46***	36.67***
TTWA=0 (Chi2)	48.67***	73.92***	161.95***	382.68***	101.11***	130.01***	49.89***
Pseudo R <sup>2</sup>	0.161	0.168	0.154	0.144	0.140	0.186	0.155
Mean VIF	3.76	3.76	3.76	3.76	3.76	3.76	3.76
Chi2	11574.4	129843.2	95557.7	130717.2	81112.8	82115.6	114563.9

Notes: As Table 5