Impact of Agglomeration Economies on Regional Performance in Germany

Andreas Behr
University of Duisburg-Essen

Christoph Schiwy
University of Duisburg-Essen

Lucy Hong
University of Washington

Received: 08/12/2021
Accepted: 05/23/2022

Abstract

We analyze the impact of agglomeration economies on employment growth rates for German sector-region combinations using firm-level data. We provide a comprehensive overview of the impact of agglomeration economies analyzing all relevant industries on three spatial scales, taking spatial dependence thoroughly into account in the regression models. As cluster formation is generally actively encouraged, we allow for a nonlinear effect of specialization in our regression model, including different industries. On the finest scale, we find that high specialization contributes to employment growth for the two sectors trade and social services, but has a negative direct effect on the manufacturing sector.

1 Introduction

A large body of literature has examined the influence of agglomeration, which refers to the spatial concentration of economic activity, on regional firm performance and growth (e.g., Jacobs, 1969; van Soest et al., 2006; Delgado et al., 2014). In theory, agglomeration can evoke positive effects on firms. These benefits, also referred to as positive externalities, are considered to play a crucial role in explaining the existence of agglomerations and regional growth, e.g., knowledge spillovers.

While there has been a rising interest in agglomeration externalities in the empirical regional research since the early 1990s, due to inconsistency in research results, it remains a conundrum as to which conditions foster the occurrence of positive agglomeration externalities and, therefore, enhance growth and innovations (Beaudry and Schiffauerova, 2009; Burger et al., 2010). In this regard, different research designs have been used to obtain a deeper understanding of the emergence of knowledge spillovers (Beaudry and Schiffauerova, 2009, pp. 48–49). Evidence points to varying extents of agglomeration forces in different sectors (Combes, 2000; Beaudry and Schiffauerova, 2009; van Soest et al., 2006).

As regional economic growth is connected to job creation, a profound understanding of the local economic structure is essential for policymakers wishing to modify local regional development programs (Blien et al., 2006). Over the past decade, a discussion on 'cluster policy' has emerged in Germany, raising the question of which specific regional conditions enhance the extent of agglomeration economies (Illy et al., 2009; Blien et al., 2006). Cluster formation in Germany is actively encouraged on a national level with the 'Leading-Edge Cluster Competition' from the Federal Ministry of Education and Research and a federal state level, e.g., the 'Cluster Offensive Bayern' (Rothgang et al., 2017). While the 'Leading-Edge Cluster Competition' focuses strongly on research and innovation, the 'Cluster Offensive Bayern' is more broadly positioned (Rothgang et al., 2017, p. 3), with fields such as 'Automotive', 'Forest and Wood' or 'Environmental Technology' (StMWi, 2020).

Since policymakers are interested in regional development and the reduction of local unemployment, it is useful to consider regional data (e.g., zip codes or districts) instead of firm-level data. However, regional data is associated with the potential impact of the spatial scale on research results (Burger et al., 2010). This problem is also known as the 'Modifiable Areal Unit Problem' (MAUP) (Openshaw and Taylor, 1979). Firstly, regional boundaries are not designed for (agglomeration) research, which is why they may not display clusters adequately. Secondly, the use of a certain level of aggregation (e.g., administrative districts, counties, or municipals) is arbitrary, and, similar to gerrymandering, can alter empirical results. Although early studies, such as Yule and Kendall (1950) or Openshaw and Taylor (1979), have already revealed that different zoning and scaling affect the correlation between variables, early literature covering agglomeration externalities has regularly neglected this issue.

More recent literature has analyzed and incorporated the spatial dependence between the regions (Garcia-López and Muñiz, 2013) since it corresponds to agglomeration theories which imply that agglomeration externalities and economic growth in one region influence its neighboring regions, requiring specific spatial solutions (Burger et al., 2010). Burger et al. (2010) and Chen (2019) find that the relationship between agglomeration externalities and economic growth or stability, respectively, varies according to whether an ordinary least square (OLS) or spatial regression is used.

We contribute to the agglomeration literature in various ways. First, we analyze the whole economy on multiple spatial scales. In contrast to the traditional agglomeration literature, in which studies often exclude or highly aggregate sectors, our study considers agglomeration effects in all sectors of the economy, carefully excluding only a few branches in which factors other than agglomeration economies may primarily determine employment. Regarding the MAUP, we consider three different spatial scales: 2-digit zip code, county and independent city, and 3-digit zip code (in order of increasing number of regions). We choose to include a spatial scale, which does not depend solely on zip codes, since zip code digits are a nested version for partitioning space into different regions.

Second, although previous research has employed spatial regression models to estimate agglomeration economies, a large number of studies restricted themselves to using the spatial autoregressive (SAR) or spatial error (SEM) model. Point estimates of spatial models, including the lagged variable of the dependent variable, are often interpreted in order to measure the magnitude and direction of the effect of neighboring regions (Elhorst, 2010). However, such an interpretation may not be appropriate, since a change in an observation may lead to a direct effect on the region, but also to indirect effects on other regions (Lesage and Fischer, 2008, p. 294). Therefore, we follow a model selection approach proposed by Elhorst (2010), combining the bottom-up and top-down approach, and calculate direct and indirect impact measures proposed by LeSage and Pace (2006) to ensure a correct interpretation of spatial parameter estimates and more valid model comparisons with different specifications. If no spillovers are assumed, a simple OLS model is often estimated. We compare this on a global level to a spatial linear mixed model, which allows for a regional error component, hence using the provided information more efficiently.

We allow for a nonlinear effect of specialization in our regression model, since the effect of high specialization in different industries has not been well analyzed so far. On the one hand, high sectoral specialization can lead to the construction of sector-specific infrastructure, from which firms can benefit, but on the other hand, it may lead to congestion effects (Henderson et al., 1995; Li et al., 2019) or amplify the effects of negative sector-specific shocks (Duranton and Puga, 2004).

The paper is structured as follows: theoretical foundations and an overview of the empirical literature are presented in Section 2. Section 3 describes the data source and variable construction. In Section 4, we present the estimating equations and further methodology on dealing with spatial dependence in regression models. Descriptive results are presented in Section 5 and the regression results in Section 6. We conclude in Section 7.

2 Literature review

In this section, we provide a brief theoretical background and discuss the empirical findings in the literature.

2.1 Theoretical foundations

Microfoundations of Agglomeration Economies Regional growth, defined as an increase in economic activity in a region, has traditionally been an important area of research (Acs and Sanders, 2014, p. 194). Many researchers agree that firm performance is affected by factors such as interchanges with customers and suppliers or technological developments (Hoogstra and van Dijk, 2004, pp. 180–181). Endogenous growth theories focus on positive external effects (externalities) of agglomerations, i.e., agglomeration economies, on a third party, instead of advantages arising from a location's natural resources (Fischer and Nijkamp, 2013, p. xxiii).

Marshall (1890) was one of the first to introduce that economies of scale emerging from the spatial concentration of economic activity in the same sector (localization economies), often labeled as specialization, leading to higher productivity and economic gains. Marshall (1890) emphasizes three determinants of agglomeration: labor pooling, input-output linkages, and technological spillovers. In particular, specialization can result in labor market pooling, providing skilled workers with relevant knowledge and reducing search costs for firms. Specialization can attract specialized suppliers or consumers, engendering reduced transport costs for firms. Also, a highly specialized industry may lead to the construction of sector-specific infrastructure, from which firms can benefit. However, since specialization may also be accompanied by congestion and commuting costs (Henderson et al., 1995; Li et al., 2019) as well as make a region less resilient to sector-specific shocks (Duranton and Puga, 2004), the direction of the net effect of sectoral specialization may be ambiguous.

Taking up the aspect of technological knowledge spillovers, Glaeser et al. (1992) extend the theory of Marshall (1890) by combining the three agglomeration theories of Marshall (1890), Arrow (1971), and Romer (1986) to the concept of Marshall-Arrow-Romer (MAR) externalities. They emphasize that the spatial proximity of firms belonging to the same sector eases the communication and transmission of knowledge for firms and experts. The flow of knowledge increases collective learning processes and the labor market pooling of specialized firms and professionals, enhancing both innovation and productivity. Because firms are aware that nearby firms can copy them, they will try to conceal innovations from others or slow down their research and development. However, due to the proximity to other firms and informal relationships between workers, knowledge spillovers can presumingly only be reduced, but not completely prevented. The MAR theory favors local monopoly since the innovation speed of firms can increase again if the number of competitors declines (Glaeser et al., 1992, p. 1131).

Countering the idea that externalities arise from specialization, other theories focus on diversity, that is, the spatial concentration of firms belonging to diverse sectors. Firms in a diversified environment can benefit from resources, such as access to qualified labor with diversified skills, well-developed infrastructures, or proximity to a variety of suppliers and clients (urbanization economies). Jacobs (1969) states that competition increases technology adoption and innovations. Although competition could decrease the returns to firms, they are pressured to innovate so as to prevent being driven out of the market. The pressure to innovate is assumed to be more important than the potential decrease in returns. Therefore, a monopoly hampering the innovation ability of firms is not regarded as socially desirable (Glaeser et al., 1992, p. 1130).

Porter (1990) regards the concentration of firms in the same sector, with interacting value-added chains, as beneficial to growth. Firms can benefit primarily from well-developed infrastructures, specialized suppliers, and knowledge spillovers, which promote the productivity of the firms in a cluster. However, agreeing with Jacobs (1969), Porter (1990) argues that local competition pressures firms to innovate and, therefore, is more desirable than a local monopoly. Concerning the labor market, a larger market can enhance the matching process between vacant jobs of firms and workers, while increasing net wage per worker and firm size (Kim (1989)). By contrast, Helsley and Strange (1990) point out that local competition can also lead to negative externalities, rising marginal social costs of workers and negative employment effects.

The New Economic Geography (NEG) addresses the role of clustering and dispersion forces generating uneven distributions of economic activity. Krugman (1991b) develops a 'core-periphery' model which relies on the assumption of labor mobility. A high concentration of workers creates a large labor market for firms, and the entry of firms attracts workers, leading to clustering. Illustrating the history of location for the U.S. "manufacturing belt", Krugman (1991a) shows that firms locate where demand is high, i.e., consumers are concentrated, creating economies of scale and lowering transportation costs. Once clusters are formed, the advantages are locked in, which is why the U.S. "manufacturing belt" persists.

More recently, advocators of the evolutionary economic geography (EEG) (cf. Frenken et al., 2007; Pessoa, 2014; Caragliu et al., 2016; Boschma, 2017) suggest a refinement to divide the traditional Jacobs externalities into the effects of related and unrelated variety on local growth. The pure diversification effect is labeled "related variety" and it is expected that being located in a region with a related variety of sectors leads to higher knowledge spillovers than in regions with an unrelated variety of sectors. However, regions with an unrelated variety of sectors are better protected against asymmetric sector-specific shocks (portfolio effect) (Caragliu et al., 2016, pp. 89–91). Boschma (2017) demonstrates the need to go beyond the simple division of related and unrelated variety. Furthermore, more clarity is needed on the relatedness of products and the distinction between related and unrelated diversification.

Modifiable Areal Unit Problem Observed spatial data is mostly aggregated to polygon entities with defined boundaries, which primarily are not designed for research but other purposes, such as administration (municipalities, counties and independent cities) or postal deliveries (zip codes) (Bivand et al., 2013). Most traditional spatial econometric studies use Metropolitan Statistical Areas (MSAs), counties, or zip codes to define boundaries (see Glaeser et al., 1992; van Soest et al., 2006). However, the use of a certain level of aggregation is arbitrary and mainly owing to data limitations.

Changes to areal unit boundaries and aggregation levels may affect a model's results (Bivand et al., 2013; Wong, 2009). Openshaw and Taylor (1979) first considered the potential issues of zoning and scaling effects, referring to them as the 'Modifiable Areal Unit Problem' (MAUP).

2.2 Empirical findings in the literature

The empirical literature[1] is rather heterogeneous regarding the impact of agglomeration externalities on economic growth (Burger et al., 2010). Most authors agree that the effects of externalities arising from specialization and diversity do not mutually exclude each other in most models (Beaudry and Schiffauerova, 2009). For instance, Glaeser et al. (1992), who analyze the employment growth of large industries in 170 U.S. cities with regard to three types of agglomeration externalities arising from specialization, diversity, and competition, find evidence of knowledge spillovers across sectors. On the contrary, Illy et al. (2009) examine German employment growth between 2003 and 2007, and find a negative effect of spillovers arising from sector diversity when estimating a regression in the services sector for free cities. Combes (2000), who estimates separate regressions for the industry and manufacturing sectors in France, obtains positive diversity effects for the services sector, but negative effects for the industry sector, showing that the effects of agglomerations are also inconsistent across different sectors. Stavropoulos et al. (2020) find that regional-related variety, as well as sectoral specialization, have a positive impact on firm productivity, particularly in high-tech and high-services sectors. According to Li et al. (2019), positive externalities arising from specialization, more precisely localization, are stronger if the sector is not dominated by a few large firms.

Various studies have analyzed the extent of the impact of local characteristics. For instance, Rosenthal and Strange (2003) examine the effects of agglomeration economies using three different concentric-circle variables, which allow them to measure the spatial employment concentration inside a region and its surroundings. They conclude that localization economies attenuate quickly with distance.

More recent studies (cf. van Soest et al., 2006; van Oort, 2007; Burger et al., 2010; Tanaka and Hashiguchi, 2020) demonstrate the importance of taking spatial dependencies between neighboring regions into account. As the extent of agglomeration externalities may change depending on the scale, the empirical results may be strongly affected by the choice of model and spatial scale. Furthermore, economic activity in one region may influence its neighboring regions, since regional labor markets, specialized urban networks, knowledge

[1] A more extensive overview of literature concerning agglomeration externalities can be found in Table A5.

spillovers, or input-output linkages are not restricted by borders. Externalities arising from knowledge spillovers or matching effects are assumed to be most prevalent at short distances, whereas impacts of market access for goods are greater on a larger scale (Combes and Gobillon, 2015, p. 294).

Van Soest et al. (2006) study employment growth and firm-birth using distance-weighted agglomeration variables based on census data of establishments in a Dutch province. They calculate classic regressions and spatial lag models, but find that the spatial lags of the dependent variable are not significant at the 5% level. Similar results are found by van Oort (2007), who analyzes the effects of spatial agglomeration on employment growth for various sectors using a municipal Dutch dataset. After testing for spatial dependence using various contiguity and inverse distance weight matrices, they employ the spatial lag model with the first-order spatial contiguity weight matrix, since it has the highest significance in measuring the spatial dependence. Van Oort (2007) introduces spatial lags of the explanatory (agglomeration) variables and finds that spatial dependence, measured by spatially lagged versions of explained and explanatory (agglomeration) variables, has a limited impact on economic growth. Moreover, concerning the local pattern of growth, local (spatially lagged) and regional agglomeration indicators often have diametrically opposing effects.

Burger et al. (2010) deal with both the spatial dependence between regions and the MAUP by studying the scale dependency of agglomeration externalities across different sectors in the Netherlands, employing OLS and spatial Durbin models (SDMs). They create their SDM's weight matrix using the row-standardized reciprocals of the distance between pairs of spatial units. Similar to the results of van Oort (2007), their spatial regression results show that the effects of local and spatially lagged versions of agglomeration externalities are often different or even diametrically opposed. Comparing the results of the OLS and spatial Durbin models at the municipal level, they find that MAUP is highly relevant to research on agglomeration externalities, since the relationship between agglomeration externalities arising from different local economic structures and local employment growth varies considerably, depending on the level of aggregation.[2]

There is also a large body of literature that uses panel models to examine the long-run effects of agglomeration on employment growth (cf. Blien et al., 2006; Dauth, 2013; Cieślik et al., 2018; de Araújo et al., 2019). Agglomeration may have dynamic effects, meaning that past growth and economic activity may affect local growth. For Germany, Blien et al. (2006) and Dauth (2013) determine significant long-run effects of agglomerations on the employment growth of local industries.

Recently advocaters of the EEG argue that the traditional dichotomic relationship of externalities arising from specialization and diversity is ill-defined (Caragliu et al., 2016, p. 89). Jacobs externalities stem from related and unrelated variety. A spatial concentration of firms from related sectors may increase knowledge spillovers, whereas a spatial concentration of firms from unrelated sectors may be less affected by a sector-specific shock. Frenken et al. (2007) consider productivity, employment, and unemployment growth using a Dutch dataset at the NUTS-3 level. They employ the spatial lag model (SAR) by including the spatial lag of productivity growth and detect that this lag is significantly negative, signifying those regions with 'highly productive' neighbors that tend to grow slowly and vice versa. They find that related variety is positively related to employment growth, while unrelated variety has a negative effect on unemployment growth, supporting the argument that unrelated variety reduces the effects of negative shocks on individual sectors. Caragliu et al. (2016) find for the European Union that the impact of MAR and Jacobs externalities differ, depending on the region's density of activity. MAR externalities have a large effect in less dense areas, whereas the impacts of Jacobs externalities are stronger in dense environments. Concerning the impacts of related and unrelated variety, the authors could only find a positive relationship between unrelated variety and regional growth.

3 Data source and variable construction

Firm-level balance sheet and income statement information was taken from the Orbis database generated by 'Bureau van Dijk' (BvD).[3]

[2] A more extensive overview of literature concerning the MAUP can be found in Table A4.

[3] Some mild selection critera have been applied which are listed in Table A1 in the appendix. We restricted the dataset to firms that do not exceed a number of employees of 2000, as we are interested in the number of employees who are actually present at the geographical location of the firm's address in the Orbis databank (similar to Duschl et al. (2015, pp. 1826–1827)).

Table 1: Sector distribution (in shares) of firms and employees for 2013 and 2017 after excluding some branches. The column 'NACE' contains the first two NACE digits, thus describing the 'divisions'. Note that index j, is in line with Equation (7) and the regression results in section 6.

j	Sector	NACE	firms 2013	firms 2017	empl. 2013	empl. 2017
3	Manufacturing	10–33	0.2506	0.2449	0.2818	0.2719
7	Energy	35–39	0.0309	0.0306	0.0567	0.0565
2	Infrastructure	41–43	0.1604	0.1596	0.0621	0.0637
6	Trade	45–56	0.3281	0.3275	0.2413	0.2482
1	F. & C.	58–68	0.0870	0.0897	0.1100	0.1083
4	Services	69–77, 79–82	0.1349	0.1391	0.2353	0.2386
5	Social	92–94	0.0082	0.0085	0.0129	0.0128
	Sum (absolute)		19397	22597	1139354	1344149

Three different area sizes are used for this analysis. In Germany, zip codes consist of five digits, which correspond to different area sizes. Since the zip code digits are a nested version used to partition space into different regions, we use the 'official municipality key' (Amtlicher Gemeindeschlüssel (AGS)) to analyze the problem at county and independent city level ('Kreis und kreisfreie Städte'), as a different approach to partition space.[4]

Sector information in the Orbis dataset is provided as the NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) code (Eurostat, 2008). We compare our data with census data provided by Regionaldatenbank Deutschland (2020). Census data is classified according to WZ 2008, the structure of which corresponds to that of NACE according to Statistisches Bundesamt (2008, pp. 47–48). According to Tables A2 and A3, employment shares per sector and federal state are similarly distributed for both datasets.

As the employment in some sectors may be determined primarily by factors other than agglomeration economies, we follow Illy et al. (2009, p. 14) and exclude the following branches based on NACE divisions: agriculture (01–02), fishing (03), mining (05–09), public sector (78, 84–91), private households (95–98), exterritorial organizations (99). The remaining branches are classified into seven groups based on all divisions (note that in NACE, some values are not assigned, e.g., there is no division 04) provided by Eurostat (2008): manufacturing (10–33), energy (35–39), infrastructure (41–43), trade (45–56), finance and communication (58–68), services (69–77, 79–82), and social (92–94).

The dataset after preparation consists of 22666 unique German firms. An overview of the sectoral structure is given in Table 1. The dominant sectors are trade and manufacturing, accounting for more than half of all firms and employees. While the social sector accounts for only 13–14% of the firms (for both years), its share in the number of employees is around 24%. For the infrastructure sector, we find a reverse relationship.

3.1 Variable construction

This subsection describes all considered variables for the three different spatial scales. Explanatory variables are from 2013, if not indicated otherwise. Moreover, the index r represents the region (2-digit zip codes, counties and independent cities, or 3-digit zip codes).

Relative Regional Sectoral Employment Growth Using job creation as the general objective from a socio-economic perspective, in line with Illy et al. (2009), we use the employment growth[5] indicator

$$GR_{rs} = \frac{emp_{rs}^{17}/emp_{rs}^{13}}{emp_{s}^{17}/emp_{s}^{13}} \tag{1}$$

[4]Geodata provided by OW networks GmbH (2018) is used to merge the official municipality key with the Orbis database. To calculate distances based on centroids, shapefiles provided by the German Federal Agency for Cartography and Geodesy (GeoBasis-DE / BKG, 2017) and Schwochow Softwareentwicklung (2019) are used.

[5]There are several ways to analyze the impact of agglomeration on productivity. Common strategies used include looking at growth, new business creation and their employment, and analyzing wages and rents (Rosenthal and Strange, 2004, p.

as a proxy for economic growth. emp_{rs}^{17} and emp_{rs}^{13} represent the (total) employment (number of employees) in region r belonging to sector s in 2017 and 2013, while emp_{s}^{17} and emp_{s}^{13} indicate the total employment in sector s in 2017 and 2013.

Relative Regional Sectoral Specialization Beaudry and Schiffauerova (2009) find that the location quotient is the most frequently used specialization indicator.[6] The location quotient of region r belonging to sector s is given by

$$LQ_{rs} = \frac{emp_{rs}^{13}/emp_{r}^{13}}{emp_{s}^{13}/emp^{13}}. \tag{2}$$

The numerator ($emp_{rs}^{13}/emp_{r}^{13}$) is defined as the employment share of sector s in region r. The denominator (emp_{s}^{13}/emp^{13}) is the employment share of sector s in (overall) Germany.

To allow for non-linearities, we also include LQ_{rs}^{2} to examine whether specialization has a positive but decreasing influence on the emergence of agglomeration externalities due to congestion (e.g., road or other infrastructure congestion).

Relative Regional Sectoral Diversity Sectoral diversity can promote the process of cross-fertilization of experiences and ideas, and therefore fosters regional growth (Combes, 2000, p. 333).

The relative inverse Herfindahl-Hirschman index (HHI) (see also Beaudry and Schiffauerova, 2009, pp. 15–16) of region r and sector s is computed as

$$HHI_{rs} = \frac{1/\sum_{j=1,j\neq s}^{J}\left(emp_{rj}^{13}/(emp_{r}^{13} - emp_{rs}^{13})\right)^{2}}{1/\sum_{j=1,j\neq s}^{J}\left(emp_{j}^{13}/(emp^{13} - emp_{s}^{13})\right)^{2}}. \tag{3}$$

j is an index for the sector that is different from sector s, and J is the total number of sectors. Therefore, we consider all sectors except sector s. Higher HHI values indicate a relatively high level of diversification.

Relative Regional Sectoral Competition A greater extent of Porter externalities may arise in competitive, concentrated regions with firms belonging to the same sector, which is why we assume a positive relationship between competition and employment growth. Following Combes (2000, pp. 337-338), we calculate the relative inverse local HHI of productive concentration

$$comp_{rs} = \frac{1/\sum_{i\in s}(emp_{rsi}^{13}/emp_{rs}^{13})^{2}}{1/\sum_{i=1}^{I}(emp_{si}^{13}/emp_{s}^{13})^{2}}, \tag{4}$$

where emp_{rsi}^{13} is the employment of firm i belonging to sector s in region r and I is the total number of firms.

Relative Regional Sectoral Average Firm Size To measure relative firm size, we use the average size of firms in sector s belonging to region r normalized by the average size of firms in the same sector s in Germany. The normalized average firm size of sector s in region r is denoted as

$$size_{rs} = \frac{emp_{rs}^{13}/firms_{rs}^{13}}{emp_{s}^{13}/firms_{s}^{13}}, \tag{5}$$

where $firms_{rs}^{13}$ is the number of firms in sector s and located in region r and $firms_{s}^{13}$ is the total number of firms belonging to sector s (Combes, 2000).

2130–2132). Each of these strategies has some advantages and disadvantages. In this paper, regional employment growth is used. The link between agglomeration forces and regional employment growth rest on the assumption that agglomeration economies increase productivity, causing regions to grow faster (Rosenthal and Strange, 2004, p. 2130). The approach has several advantages, such as that the data are usually more readily available, or that the policy implications can be derived more directly. However, this operationalization also has a drawback in that firms are constrained by their prior decisions to invest in fixed capital, which affects the valuation of marginal labor. (Rosenthal and Strange, 2004, p. 2130).

[6] See Glaeser et al. (1992) and van Soest et al. (2002) for applications of this indicator.

4 Methodology

Because, by construction, the relative variables exhibit a right-skewed distribution, they are transformed using the natural logarithm:

$$\begin{aligned}&lGR_{rs} = \ln{(GR_{rs})},\ lLQ_{rs} = \ln{(LQ_{rs})},\ lLQ_{rs}^2 = (lLQ_{rs})^2,\ lHHI_{rs} = \ln{(HHI_{rs})},\\ &lcomp_{rs} = \ln{(comp_{rs})} \text{ and } lsize_{rs} = \ln{(size_{rs})}.\end{aligned} \tag{6}$$

As a baseline model, we estimate OLS regressions using all available region-sector combinations (rs)

$$lGR_{rs} = \beta_0 + \beta_1 lLQ_{rs} + \beta_2 lLQ_{rs}^2 + \beta_3 lHHI_{rs} + \beta_4 lcomp_{rs} + \beta_5 lsize_{rs} + \sum_{j=6}^{11} \beta_j S_{rs,j-5} + \epsilon_{rs}. \tag{7}$$

$S_{rs,1}, \ldots, S_{rs,6}$ take the value 1 if region-sector rs belongs to sector S_j $(j = 1, ..., 6)$ and 0 otherwise, using sector seven (social) as the reference sector. Note that a region r must not include all seven sectors. We consider seven different sectors (see Table 1), and the model for a specific sector s' is given as

$$lGR_{rs'} = \beta_0 + \beta_1 lLQ_{rs'} + \beta_2 lLQ_{rs'}^2 + \beta_3 lHHI_{rs'} + \beta_4 lcomp_{rs'} + \beta_5 lsize_{rs'} + \epsilon_{rs'}. \tag{8}$$

Both models impose the assumption of independent and identically distributed (iid) observations, implying that $\mathrm{E}(\epsilon_i \epsilon_j) = \mathrm{E}(\epsilon_i) \cdot \mathrm{E}(\epsilon_j) = 0$. However, the assumption of statistically independent observations is violated if there is a spatial dependence between the observations (LeSage and Pace, 2009, p. 2).

We test for spatial autocorrelation in the OLS residuals (null hypothesis: absence of spatial autocorrelation) using Moran's I, and provide results from a permutation bootstrap test. We model the spatial proximity of regional centroids (coordinates in degrees) as inverse distances, where the distances are calculated as great-circle distances, assuming the Earth to be a sphere with a radius of 6371.009 km (mean radius in the ellipsoidal model (Deza and Deza, 2012, p. 466)).[7] Combining the row-standardized inverse distances in a matrix, we gain the spatial weights matrix $\boldsymbol{W}$. We test the OLS regression residuals $\hat{\boldsymbol{\epsilon}}$ for spatial autocorrelation, taking into account that the residuals are correlated, and adjust the mean and variance of I accordingly (Cliff and Ord, 1981, p. 200).

The empirical p-value is calculated on the basis of a random permutation test (Bivand et al., 2013, p. 278). If we find spatial autocorrelation based on Moran's I, we use a spatial regression model.[8] Basically, three types of regional spillovers can occur: (1) values of the dependent variable and (2) values of the explanatory variables from other regions may have an impact, and (3) the errors might be correlated. All types are included in the Manski model

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \alpha \boldsymbol{\iota}_n + \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X} \boldsymbol{\theta} + \boldsymbol{u}, \quad \text{with} \quad \boldsymbol{u} = \lambda \boldsymbol{W} \boldsymbol{u} + \boldsymbol{\epsilon}, \tag{9}$$

where $\boldsymbol{y}$ is a $(n \times 1)$ vector with observations of the dependent variable, ρ is a spatial autoregressive parameter, and $\boldsymbol{W}$ is a row-standardized $(n \times n)$ spatial weight matrix. Therefore, $\rho \boldsymbol{W} \boldsymbol{y}$ represents the endogenous interaction effects between the dependent variables. $\boldsymbol{\iota}_n$ indicates a $(n \times 1)$ vector with ones. $\boldsymbol{\iota}_n$ is associated with the constant term parameter α, whereas $\boldsymbol{X}$ represents a $(n \times k)$ matrix of explanatory variables with $\boldsymbol{\beta}$ as the associated $(k \times 1)$ parameter vector. Correspondingly, $\boldsymbol{W} \boldsymbol{X} \boldsymbol{\theta}$ accounts for exogenous interaction effects with $\boldsymbol{\theta}$ as a $(k \times 1)$ parameter vector. $\boldsymbol{\epsilon}$ is a $(n \times 1)$ vector of error terms where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, and λ is an autoregressive parameter for the error lags (Elhorst, 2010, pp. 11–13).

As the parameters in the Manski model are unidentified, restrictions must be imposed. We follow Elhorst (2010), who suggests an approach presented in Figure 1, starting with the bottom-up approach but following with the top-down approach.[9] If the (robust) LM error and lag tests point to the SAR and SEM model, the spatial Durbin model should be estimated (Elhorst, 2010, p. 14).

[7] When analyzing all available region-sector combinations (Equation (7)), the inverse distance for sectors belonging to the same region is set to one, allowing spillovers between sectors in the same region.

[8] These models are, e.g., described in detail in LeSage and Pace (2009).

[9] While Florax et al. (2003) advocate a bottom-up approach expanding the linear regression with spatially lagged variables based on the result of LM (Lagrange multiplier) tests, LeSage and Pace (2009) support the use of a top-down approach starting from the spatial Durbin model (SDM) and utilizing LR (Likelihood ratio) tests.

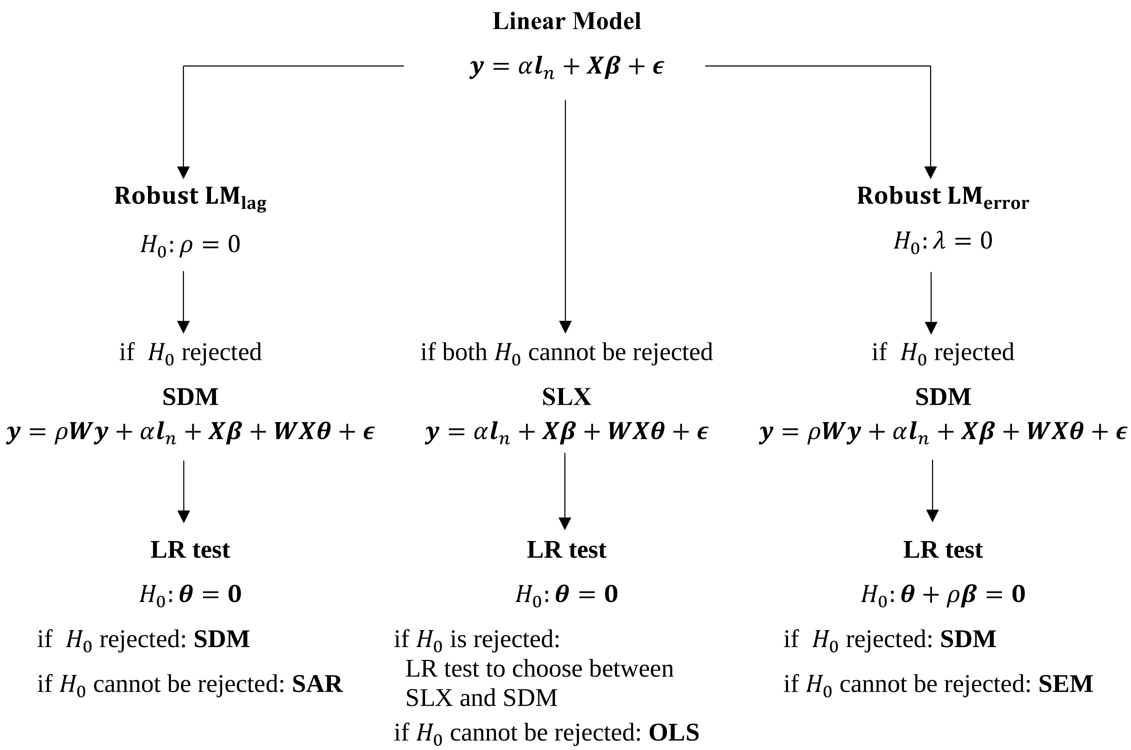


Figure 1: Model selection approach according to Elhorst (2010).

Another modeling approach assumes that there are no spillovers, but allows for a regional error component. Even assuming that there are no regional spillover effects, different regional characteristics may still affect observations differently. Therefore, in this case, in the linear mixed model, all observations from the same region are assumed to be equally affected by these characteristics. The additional consideration of spatial correlation between them leads to the following spatial linear mixed model. We define a spatial linear mixed model (for an introduction to mixed models, see, e.g., Cameron and Trivedi (2005, pp. 774–777)), with a random regional intercept as

$$\boldsymbol{y} = \alpha \boldsymbol{\iota}_n + \boldsymbol{X\beta} + \boldsymbol{b} + \boldsymbol{\epsilon}, \tag{10}$$

where $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi})$, and $\boldsymbol{\Psi}$ is a covariance matrix, which models spatial correlation. As the covariance function, a class Matérn function of the form

$$C(d) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\rho d)^{\nu} K_{\nu}(\rho d), \tag{11}$$

for distances $d > 0$, again calculated as great-circle distance assuming the Earth to be a sphere with a radius of 6371.009 km, is used with positive parameters ρ (scale) and ν (smoothness) (Rousset and Ferdy, 2020, p. 67). $K_{\nu}(\cdot)$ is a modified Bessel function of order ν, and $\Gamma(\cdot)$ is the Gamma function. Correlation for zero distance is set to one (Rousset and Ferdy, 2020, p. 67).

For the class Matérn covariance function, correlation decreases as the distance increases (Diggle and Ribeiro, 2007, p. 51). To evaluate statistical significance in the mixed model, we use the t-as-z approach (Luke, 2017, p. 1495), which is based on the standard normal distribution to obtain p-values, as the number of observations for the global regression (Equation (7)) is quite high.[10]

[10] All analyses in this paper were done with R (R Core Team, 2020).

5 Descriptive results

Three different spatial scales are analyzed. At the 2-digit zip code level, there are 95 regions, 401 regions at the counties and independent cities level, and 671 regions at the 3-digit zip code level. We illustrate the number of employees in 2017 per region (Figure 2), (not normalized) growth rate of employees from 2013 to 2017 (emp_r^{17}/emp_r^{13}, Figure 3), and the absolute growth ($emp_r^{17} - emp_r^{13}$, Figure B1) for all spatial scales to demonstrate how changes in areal unit boundaries and aggregation level affect the descriptive results.

From Figure 2, we find that regions with high employment (in 2017) seem to cluster in the old federal states (West Germany). This is the more evident, the finer the scale. The regional growth rate seems to have no clear trend (Figure 3), whereas the absolute growth depicted in Figure B1 in the appendix shows a similar picture to that of regional employment.

Descriptive statistics of the untransformed and transformed variables using the natural logarithm for the three spatial scales are shown in Tables 2 and A6. We observe that the range between the lowest and highest values of the variables, and variation within variables, is higher at the finer spatial scales (county and independent city and 3-digit zip code) compared to the 2-digit zip code scale. Due to the construction of relative measures, we observe right-skewed distributions. For instance, the maximum value of LQ, particularly at the 3-digit zip code level, is more than 30 times higher than the mean. In the regression analysis, we therefore use all obtained regional measures in logarithms to reduce the skewness.

To demonstrate that the distribution of the variables changes depending on the sector in question, we also calculate the summary statistics for the three spatial scales at the sectoral level for the untransformed and transformed variables. The corresponding tables can be found in the appendix (Tables A7 and A8 for the 2-digit zip code scale, Tables A9 and A10 for the county and independent city scale, and Tables A11 and A12 for the 3-digit zip code scale).

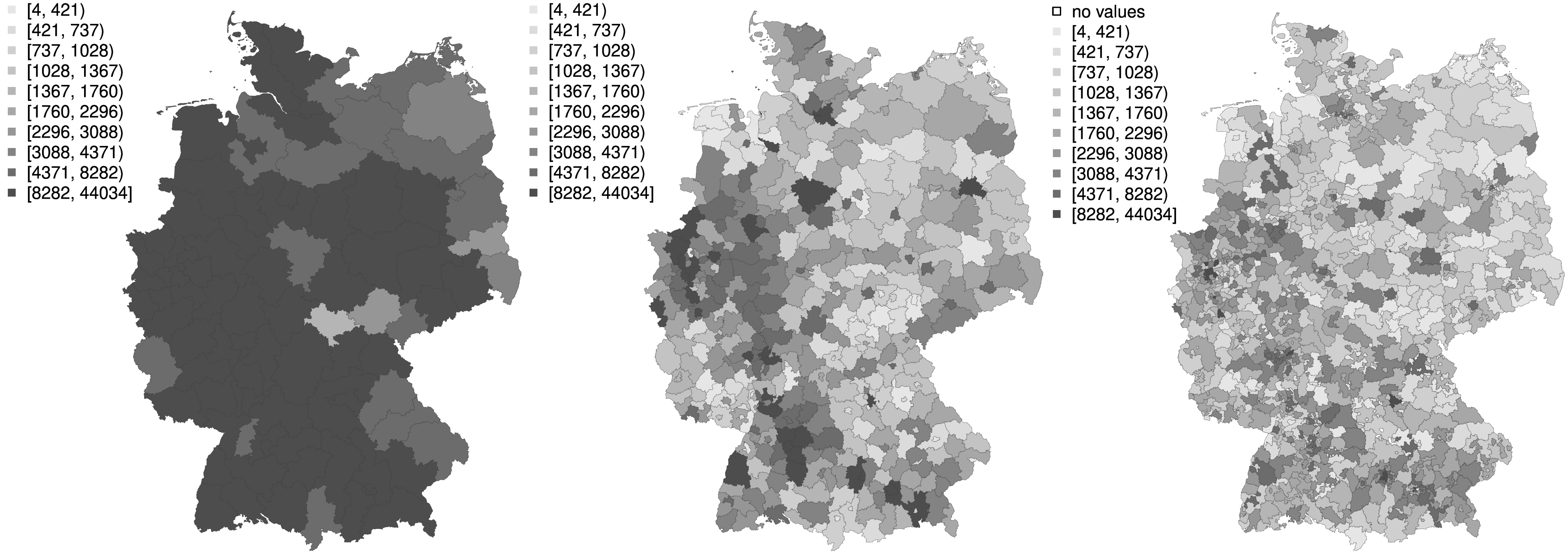
At the sectoral level, we again find the tendency that the range of variables increases with the number of regions. Characteristics within sectors are quite different, e.g., at the 3-digit zip code level (Table A11), the range of GR is smallest in the manufacturing sector at 8.0148 and largest in the sector finance and communication at 35.2514. For relative regional sectoral specialization LQ, median values range between 0.4634 for finance and communication and 1.7092 for social. We also find larger differences for competition ($comp$) and firm size ($size$), where the maximum sectoral median is about ten or, respectively, three times higher than the minimum median value. For diversity, we detect a rather small range in median values of 0.5682 (infrastructure) and 0.6582 (services). The 3-digit zip code level consists of 671 regions. However, only 125 regions inherit firms from the social sector (smallest number), whereas 645 regions inherit firms from the trade sector (largest number).

Since it can be assumed that the effects of specialization and diversification on growth oppose one another, we more closely analyzed the relationship between the two. The results from Table A13 show that the correlation between specialization (lLQ) and diversification ($lHHI$) is generally rather low.

6 Regression results

In this section, we first consider Moran's I (null hypothesis: absence of spatial autocorrelation; alternative: spatial autocorrelation) for the residuals from Equations (7) and (8) with adjusted mean and variance for regression residuals for all three spatial scales. We use a spatial weight matrix based on the great circle (geographic) distance between the regional centroids, e.g., d_{ij} for the distance between centroids from regions i and j, with $p_{ij} = d_{ij}^{-1}$. We present Moran's I and the observed p-value in Table 3. A random permutation test is performed (same null and alternative hypothesis as above) and the median and mean of the Monte-Carlo simulations (999 simulations plus observed statistic) of Moran's I and the corresponding pseudo p-value are presented.

If the absence of spatial autocorrelation is rejected (p-value < 0.1), we proceed with the model selection approach of Elhorst (2010). However, if Moran's I does not indicate the presence of spatial autocorrelation, we estimate an OLS model. To allow a comparison, we provide OLS estimations for all spatial scales at the global level, despite rejecting the null hypothesis of no spatial autocorrelation at the 5% significance level for the 3-digit zip code scale.

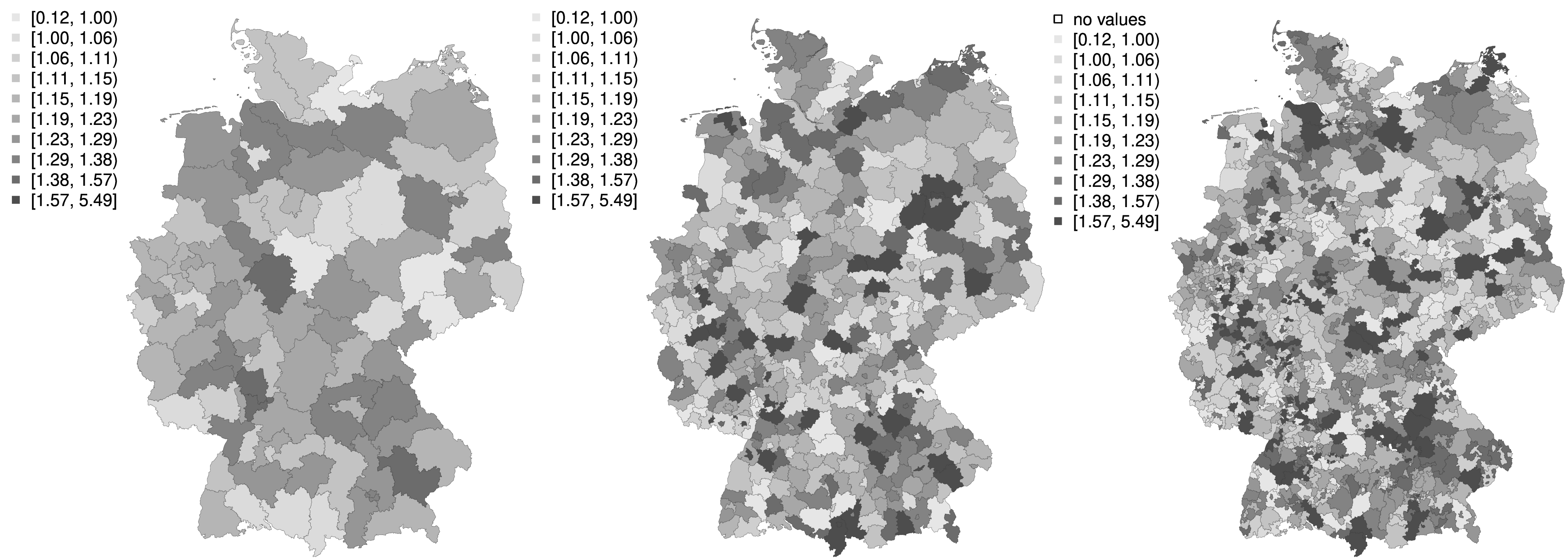


(a) 2-digit zip

(b) county and independent cities

(c) 3-digit zip

Figure 2: Number of employees in 2017 per region (emp_r^{17}) for all three spatial scales. Intervals are based on the deciles of emp_r^{17} for all spatial scales combined.



(a) 2-digit zip

(b) county and independent cities

(c) 3-digit zip

Figure 3: Not normalized growth rate of employees from 2013 to 2017 (emp_r^{17}/emp_r^{13}) for all three spatial scales. Intervals are based on the deciles of emp_r^{17}/emp_r^{13} for all spatial scales combined.

Table 2: Descriptive statistics for the variables used in the global regressions. For better interpretability given in levels, for the regressions the variables are transformed using the natural logarithm (see Equation (6)). The first row for each spatial scale denotes the number of available region-sector combinations.

Scale	Statistic	GR	LQ	LQ^2	HHI	$comp$	$size$
2-digit zip	n	634					
	min	0.2656	0.0055	0.0000	0.2784	0.0022	0.0108
	median	0.9940	0.8763	0.7680	0.8475	0.0191	0.8516
	mean	1.1754	1.0343	1.7848	0.8381	0.0263	1.0248
	max	20.5658	8.3625	69.9310	1.2568	0.1794	7.8700
	sd	1.1755	0.8457	4.3593	0.1629	0.0233	0.8917
county and indep. cities	n	2154					
	min	0.0005	0.0019	0.0000	0.2528	0.0014	0.0093
	median	0.9768	0.8768	0.7688	0.6627	0.0072	0.6486
	mean	1.2707	1.2838	4.9127	0.6666	0.0109	1.0084
	max	86.7423	37.5587	1410.6595	1.3857	0.1653	20.7296
	sd	2.3951	1.8068	39.2912	0.2062	0.0128	1.4792
3-digit zip	n	3361					
	min	0.0072	0.0019	0.0000	0.2350	0.0014	0.0093
	median	0.9786	0.8457	0.7152	0.6134	0.0057	0.5914
	mean	1.2258	1.3461	5.8129	0.6301	0.0084	1.0068
	max	35.2586	41.5243	1724.2655	1.3857	0.1219	36.9052
	sd	1.4011	2.0002	52.2496	0.2023	0.0091	1.6490

At the sectoral level, we reject the null hypothesis at the 10% significance level for sector manufacturing at the 2-digit zip code scale, for the sectors infrastructure, trade, and social at county and independent city levels, and for the sectors manufacturing, infrastructure, trade, and social at the 3-digit zip code scale.

Global regression results The global regression results are given in Table 4. We estimate OLS regressions at each spatial scale for comparison. The model selection approach of Elhorst (2010) indicates a SDM at the 3-digit zip code level. As the coefficients of the SDM cannot be directly interpreted as marginal effects, we also calculate the direct and indirect effects.

Consistent with our expectation, the OLS regression results of counties and independent cities and the 3-digit zip codes are similar. However, they differ from the 2-digit zip code level, which might be due to the high aggregation level. Concerning statistically significant parameters, both regressions indicate a highly significant negative influence of the specialization measure lLQ at the 1% level, which is in line with the regression results of Combes (2000) for France. Van Oort (2007) finds similar results for the Netherlands. The concentration of distribution activities hampers employment growth, whereas the spatial lag of the concentration variable affects growth positively.

Since specialization effects are two-sided, e.g., MAR externalities are fostered by labor market pooling, but specialized regions may also amplify the effect of negative shocks, it is unclear whether specialization per se is conducive to growth. Both regressions display a statistically significant negative effect of the variable $lsize$ at the 1% significance level, indicating that larger firms have lower growth rates. This finding supports the result of Combes (2000), who stresses that a negative relationship between relative average firm size and employment growth may portray a so-called ‘firm’s life-cycle effect’, indicating that the growth of small firms is initially faster, but will slow down if an ‘optimal’ size is reached. Belonging to either the energy (baseline sector) or the finance and communication sector shows lower growth at the county and independent city and 3-digit zip code scale than the other sectors.

The regression results of the spatial Durbin model at the 3-digit zip code scale are equal to the OLS model in the sign of coefficients. We see a shift in statistical significance of the direct impacts from specialization to diversity (at the 10% significance level). We further find that spatial lags, except diversity at the 10% significance level, and indirect effects of the variables measuring specialization, diversity, and competition, are not statistically significant, demonstrating that spatial effects of agglomeration externalities on employment growth may be limited in extent.

Table 3: Moran's I for the residuals for all three spatial scales.

Dataset	Group	Moran's I	obs. p-value	MC_{median}	MC_{mean}	ps. p-value
2-digit zip						
	All	0.0058	0.3047	-0.0026	-0.0016	0.2900
	Manufac.	0.0156	0.0399	-0.0121	-0.0111	0.0540
	Energy	-0.0176	0.6279	-0.0133	-0.0115	0.6300
	Infr.	-0.0040	0.3213	-0.0117	-0.0103	0.3390
	Trade	-0.0143	0.5586	-0.0117	-0.0108	0.5700
	F. & C.	-0.0220	0.7315	-0.0111	-0.0111	0.7560
	Services	-0.0118	0.4853	-0.0130	-0.0109	0.4680
	Social	-0.0311	0.7627	-0.0157	-0.0141	0.7660
county and indep. cities	All	0.0035	0.1715	-0.0002	-0.0002	0.1930
	Manufac.	0.0025	0.1410	-0.0031	-0.0027	0.1280
	Energy	-0.0087	0.7686	-0.0049	-0.0043	0.7860
	Infr.	0.0113	0.0031	-0.0029	-0.0025	0.0170
	Trade	0.0112	0.0032	-0.0031	-0.0025	0.0100
	F. & C.	-0.0105	0.8618	-0.0037	-0.0036	0.8980
	Services	-0.0044	0.5962	-0.0031	-0.0029	0.6190
	Social	0.0247	0.0081	-0.0123	-0.0104	0.0240
3-digit zip						
	All	0.0056	0.0064	-0.0004	-0.0004	0.0130
	Manufac.	0.0038	0.0637	-0.0019	-0.0018	0.0590
	Energy	-0.0079	0.7580	-0.0034	-0.0030	0.7970
	Infr.	0.0100	0.0011	-0.0018	-0.0016	0.0060
	Trade	0.0041	0.0630	-0.0018	-0.0015	0.0810
	F. & C.	0.0029	0.1681	-0.0025	-0.0023	0.1480
	Services	-0.0014	0.4495	-0.0022	-0.0020	0.4090
	Social	0.0341	0.0092	-0.0082	-0.0072	0.0260

Similar to earlier findings, such as from Burger et al. (2010), the direct and indirect effects of the agglomeration variables (lLQ, $lHHI$, $lcomp$) are mostly diametrically opposed. For example, the measure for diversity directly affects employment growth positively, but has a negative indirect effect on employment growth. This means that the emergence of agglomeration externalities may depend on the spatial scale.

Comparing these results to those from the spatial linear mixed model (Table 5), we find similar results for the 2-digit zip code and counties and independent cities level. While Moran's I indicated spatial autocorrelation in the OLS residuals for the 3-digit zip code level (pseudo p-value 0.0130), the residuals of the SMM model indicate that the random effect with covariance function of class Matérn was sufficient to take care of spatial autocorrelation (pseudo p-value 0.9490). As in the SDM (direct impact), we see a positive statistically significant effect of diversity at the 10% level.

Sectoral regression results In light of the finding that the intensity of agglomeration forces depends on the observed sector (Combes, 2000; Burger et al., 2010), the three global regression models are separately re-estimated for the seven sectors. For the 2-digit zip code scale, Moran's I pointed towards the OLS model for all sectors, except manufacturing, for which the model selection according to Elhorst (2010) indicated an OLS model (Table A14). At the county and independent city level, Moran's I, and the model selection pointed to OLS for most sectors (Table A15), but for the sectors trade and social, the selection process indicated a SLX model (Table A16). At the 3-digit zip code scale, OLS models could be estimated (Table 6) for most sectors, but for the manufacturing sector, the selection process pointed to a SDM (Table 7). Similar to the global regression results and consistent with the results of Combes (2000) for France and Illy et al. (2009) for Germany, the measure of specialization lLQ affects growth negatively in most sectors at all spatial scales. This effect is statistically significant for sector trade at the 2-digit zip code scale, infrastructure, finance and communication, and social at the county and independent city scale, as well as social and manufacturing at the 3-digit zip code scale. For the squared specialization variable lLQ^2, we find statistically significant effects for sector social (positive) at the 2-digit zip code scale, for the sectors infrastructure (positive), services (negative), and social (positive) at the county and independent city scale, and for the sectors trade (positive), social (positive) and manufacturing (negative) at the 3-digit zip code scale.

Table 4: Global regression results for all three spatial scales.

Coef./Fit	2-digit zip	county and indep. cities	3-digit zip	Coef./Fit	3-digit zip			direct	indirect	total
constant	−0.0626	−0.2000***	−0.2056***	constant	−0.8046	ρ	0.1845			
	(0.0946)	(0.0772)	(0.0696)		(0.6852)					
lLQ	−0.0436	−0.0467***	−0.0279***	lLQ	−0.0204	Lag lLQ	0.2286	−0.0200	0.2753	0.2553
	(0.0352)	(0.0143)	(0.0104)		(0.0163)		(0.1630)	(0.0144)	(0.1912)	(0.1905)
lLQ^2	0.0419**	−0.0058	0.0000	lLQ^2	0.0003	Lag lLQ^2	0.0125	0.0003	0.0154	0.0158
	(0.0208)	(0.0069)	(0.0049)		(0.0049)		(0.0339)	(0.0034)	(0.0430)	(0.0437)
$lHHI$	−0.0531	0.0217	0.0340	$lHHI$	0.0671**	Lag $lHHI$	−0.4218*	0.0663*	−0.5013	−0.4350
	(0.0763)	(0.0370)	(0.0237)		(0.0338)		(0.2489)	(0.0362)	(0.3174)	(0.3034)
$lcomp$	0.0082	−0.0159	−0.0199	$lcomp$	−0.0096	Lag $lcomp$	0.0063	−0.0096	0.0055	−0.0041
	(0.0204)	(0.0161)	(0.0136)		(0.0169)		(0.1150)	(0.0174)	(0.1553)	(0.1524)
$lsize$	−0.0653*	−0.1091***	−0.0969***	$lsize$	−0.0975***	Lag $lsize$	0.0852	−0.0973***	0.0823	−0.0150
	(0.0361)	(0.0205)	(0.0145)		(0.0205)		(0.1060)	(0.0172)	(0.1254)	(0.1160)
F. & C.	0.0411	0.0157	0.0291	F. & C.	0.0396	Lag F. & C.	0.6379	0.0408	0.7900	0.8308
	(0.0648)	(0.0501)	(0.0391)		(0.0420)		(0.5663)	(0.0337)	(0.6927)	(0.7072)
Infr.	0.0927*	0.1619***	0.1346***	Infr.	0.1363***	Lag Infr.	0.1005	0.1366***	0.1539	0.2905
	(0.0549)	(0.0387)	(0.0320)		(0.0442)		(0.7735)	(0.0504)	(1.0197)	(1.0537)
Manufac.	0.0729	0.1558***	0.1371***	Manufac.	0.1406***	Lag Manufac.	0.2895	0.1412***	0.3862	0.5273
	(0.0571)	(0.0390)	(0.0325)		(0.0492)		(0.9022)	(0.0498)	(1.0327)	(1.0664)
Services	0.0411	0.0907**	0.0879**	Services	0.1292***	Lag Services	1.3861**	0.1319***	1.7262**	1.8581**
	(0.0603)	(0.0453)	(0.0348)		(0.0376)		(0.5810)	(0.0386)	(0.7912)	(0.8092)
Social	0.1473*	0.2019***	0.1323**	Social	0.1317**	Lag Social	0.7812	0.1332**	0.9862	1.1194
	(0.0767)	(0.0713)	(0.0564)		(0.0596)		(0.7211)	(0.0527)	(0.8240)	(0.8411)
Trade	0.0930	0.1564***	0.1572***	Trade	0.1819***	Lag Trade	0.7915	0.1835***	1.0101	1.1936
	(0.0587)	(0.0374)	(0.0324)		(0.0536)		(0.9961)	(0.0546)	(1.2257)	(1.2682)
Model	OLS	OLS	OLS	Model	SDM					
Obs.	634	2154	3361	Obs.	3361					
BP test	82.1204***	48.9188***	118.2506***	SP BP test	121.5968***					
R^2	0.1992	0.0872	0.0752	Log likelihood	−2108.0323					
Adj. R^2	0.1850	0.0826	0.0722							
Wald (F)	4.6602***	8.6545***	13.8712***							

Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis 'homoscedasticity' of the studentized Breusch-Pagan (BP) test had to be rejected (5% level), adjusted for spatial regressions (SP) if needed. Standard errors are given in parentheses. For the impacts, inference based on their empirical distribution (100 simulation runs due to computational burden) is presented.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

Table 5: Global regression results for all three spatial scales using spatial mixed models (SMM).

Coef./Fit	2-digit zip	county and indep. cities	3-digit zip
constant	-0.0650	-0.2011^{**}	-0.2053^{***}
	(0.1004)	(0.0853)	(0.0760)
lLQ	-0.0431	-0.0449^{***}	-0.0253^{**}
	(0.0289)	(0.0156)	(0.0115)
lLQ^2	0.0416^{***}	-0.0057	0.0003
	(0.0093)	(0.0045)	(0.0032)
$lHHI$	-0.0544	0.0317	0.0408^{*}
	(0.0631)	(0.0337)	(0.0247)
$lcomp$	0.0075	-0.0168	-0.0198
	(0.0241)	(0.0174)	(0.0149)
$lsize$	-0.0666^{**}	-0.1107^{***}	-0.1004^{***}
	(0.0327)	(0.0163)	(0.0122)
F. & C.	0.0406	0.0129	0.0267
	(0.0483)	(0.0420)	(0.0329)
Infr.	0.0922^{*}	0.1603^{***}	0.1342^{***}
	(0.0489)	(0.0418)	(0.0336)
Manufac.	0.0722	0.1528^{***}	0.1367^{***}
	(0.0495)	(0.0421)	(0.0337)
Services	0.0401	0.0878^{**}	0.0850^{**}
	(0.0504)	(0.0422)	(0.0334)
Social	0.1475^{***}	0.2007^{***}	0.1288^{**}
	(0.0569)	(0.0616)	(0.0510)
Trade	0.0923^{*}	0.1544^{***}	0.1568^{***}
	(0.0493)	(0.0415)	(0.0329)
Model	SMM	SMM	SMM
Obs.	634	2154	3361
$\hat{\nu}$	0.5000	0.5000	0.5000
$\hat{\rho}$	0.4052	0.0054	0.0187
$\hat{\sigma}_b^2$	0.0010	0.0014	0.0042
$\hat{\sigma}_\epsilon^2$	0.1067	0.2276	0.2025
Log likelihood	-193.1154	-1466.8825	-2111.0548

Notes: Conditional standard errors are given in parentheses. p-values are obtained based on the t-as-z approach. As covariance function, the Matérn function is used (ρ scale, ν smoothness parameter). Distances are calculated as great-circle distance, assuming the Earth to be sphere with a radius of 6371.009 km.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

Although most tendencies are similar across the three spatial scales, we also discover some diametrically opposed effects. For instance, for the trade sector, the variables lLQ and lLQ^2 negatively affect growth at the 2-digit zip code and county and independent city scales, but have a positive effect on growth at the 3-digit zip code scale. This finding indicates that our employment growth models are not robust across the spatial scales, hinting at the existence of MAUP.

We choose the 3-digit zip code scale to discuss sectoral results, as it is our finest spatial scale. The goodness of fit varies strongly across sectors, for instance, the finance and communication sector regression displays an adjusted R^2 of 0.0606 (smallest value), whereas the social sector has a value of 0.2507 (largest value). This result suggests that the extent to which agglomerations contribute to growth differs strongly between sectors, and may be explained by the fact that sectoral input mixes are heterogenous.

Regarding the effect of specialization (lLQ and lLQ^2) at the 3-digit zip code scale, we determine only for the sectors trade and social, a positive significant relation between growth and high specialization (lLQ^2), with a significant negative effect for just specialization for the social sector. For the manufacturing sector, specialization (both lLQ and lLQ^2) has a significant negative impact on growth.

The coefficients for sectoral diversity and competition are not statistically significant for most sectors. We discover that diversity contributes significantly positively to growth for the finance and communication sector, implying that Jacobs externalities have a positive effect only in certain sectors. Competition affects growth significantly negatively only for the infrastructure sector, which is in line with Combes' (2000) result

Table 6: Regression results for sectors at 3-digit zip code level, which are estimated by OLS.

Coef./Fit	Energy	Infr.	Trade	F. & C.	Services	Social
constant	−0.0268	−0.4405***	−0.0682	0.3049	0.1187	0.0801
	(0.3632)	(0.1495)	(0.1321)	(0.2588)	(0.2353)	(0.2853)
lLQ	−0.0395	−0.0226	0.0121	−0.0120	−0.0437	−0.1682***
	(0.0313)	(0.0178)	(0.0248)	(0.0372)	(0.0465)	(0.0554)
lLQ^2	−0.0114	0.0241	0.0631***	−0.0092	0.0030	0.0819***
	(0.0076)	(0.0164)	(0.0172)	(0.0135)	(0.0121)	(0.0242)
$lHHI$	0.1014	0.0178	−0.0033	0.1837**	0.0367	−0.0342
	(0.0786)	(0.0514)	(0.0428)	(0.0880)	(0.0714)	(0.1381)
$lcomp$	−0.0105	−0.0791**	−0.0099	0.0638	0.0264	0.0571
	(0.0800)	(0.0307)	(0.0269)	(0.0568)	(0.0447)	(0.0964)
$lsize$	−0.0247	−0.1818***	−0.1291***	−0.1340**	−0.0861**	−0.0318
	(0.0309)	(0.0481)	(0.0250)	(0.0554)	(0.0343)	(0.0506)
Model	OLS	OLS	OLS	OLS	OLS	OLS
Obs.	332	607	645	476	550	125
BP test	5.9741	121.8203***	42.4687***	19.3280***	5.7181***	31.8934***
R^2	0.0332	0.1319	0.1823	0.0705	0.0989	0.2809
Adj. R^2	0.0183	0.1247	0.1759	0.0606	0.0906	0.2507
Wald (F)	2.2368*	6.0294***	14.9765***	3.6522***	11.9441***	2.7080***

Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis 'homoscedasticity' of the studentized Breusch-Pagan (BP) test had to be rejected (5% level). Standard errors are given in parentheses.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

for the industry and services sectors, but contradicts the finding of van Oort (2007), who use an inverse of the variable *size* as a proxy for competition. Caragliu et al. (2016) also find that Porter externalities are not a driver of employment growth in their sample of European NUTS2 regions, which includes twelve service and manufacturing sectors.

In line with our global regression results, the effect of average firm size shows a negative contribution to growth and is consistent (and most often significant) throughout all sectors. This result is also in line with findings of Combes (2000) for France and de Araújo et al. (2019) for Brazil.

7 Conclusion

Consistent with earlier MAUP-related research, such as Fotheringham and Wong (1991) and Burger et al. (2010), our descriptive and regression results show that agglomeration effects vary strongly, depending on the spatial scale, implying that the MAUP is present and should not be neglected. As the regression results of the counties and independent cities and 3-digit zip codes exhibit similar tendencies concerning the statistical significance and direction of the parameter estimates, it is reasonable to assume that a finer scale aggregation leads to more robust and accurate results. For some sectors, the observed Moran's I and associated significances differ greatly across the spatial scale. Therefore, in addition to the MAUP, the presence of spatial autocorrelation seems to depend on the spatial scale.

At the global level, we find a positive effect of sectoral diversity on employment growth only for the spatial regression models at the 3-digit zip code level. Concerning the effects of competition, our sectoral regressions do not provide direct evidence of a relationship between local competition and employment growth, as we find very heterogeneous and statistically insignificant competition effects at the global level.

According to our results, clustering, in the sense of regional sectoral specialization, has a rather negative effect on employment growth. Analyzing all available region-sector combinations, we found a negative effect, and only for the 2-digit zip code level, a significant positive effect of high specialization. Consistent with the literature, our results imply that the effect of specialization and competition are ambiguous, for which there may be several reasons. For instance, although spatial concentration may lead to labor market pooling, improve job matching (Kim, 1989), or attract suppliers and consumers,

Table 7: Regression results for the sector manufacturing at 3-digit zip code level, which are estimated by SDM.

Coef./Fit	Manufac.			direct	indirect	total
constant	1.0766	ρ	−0.1084			
	(1.3012)					
lLQ	−0.0443*	Lag *lLQ*	0.5737*	−0.0446*	0.5223	0.4777
	(0.0241)		(0.3170)	(0.0254)	(4.2858)	(4.2917)
lLQ^2	−0.0148**	Lag lLQ^2	0.2310**	−0.0150*	0.2100	0.1950
	(0.0071)		(0.1141)	(0.0078)	(1.8714)	(1.8743)
lHHI	0.0274	Lag *lHHI*	0.5059	0.0271	0.4540	0.4811
	(0.0440)		(0.7736)	(0.0459)	(8.0791)	(8.0923)
lcomp	−0.0327	Lag *lcomp*	0.1852	−0.0328	0.1704	0.1376
	(0.0260)		(0.2224)	(0.0264)	(0.3951)	(0.3932)
lsize	−0.1335***	Lag *lsize*	0.3713	−0.1338***	0.3483	0.2145
	(0.0239)		(0.2462)	(0.0264)	(7.2692)	(7.2816)
Model	SDM					
Obs.	626					
SP BP test	17.3919***					
Log likelihood	−245.1618					

Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis 'homoscedasticity' of the studentized Breusch-Pagan (BP) test had to be rejected (5% level), adjusted for spatial regressions (SP). Standard errors are given in parentheses. For the impacts, inference based on their empirical distribution (2000 simulation runs) is presented.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

a specialized region may be less resilient to shocks, as negative sectoral shocks have a large effect on the region (Duranton and Puga, 2004). Therefore, we cannot conclude directly that regional policies will not positively affect regional growth. Furthermore, it is neccessary to analyze whether a specialized regional structure is beneficial to a region in the long run (Illy et al., 2009; Combes, 2000).

References

Acs, Z. and Sanders, M. (2014). Endogenous growth theory and regional extensions. In *Handbook of Regional Science*, page 193. Springer Berlin.

Arbia, G. and Petrarca, F. (2011). Effects of MAUP on spatial econometric models. *Letters in Spatial and Resource Sciences*, 4(3):173–185.

Arrow, K. J. (1971). The economic implications of learning by doing. In *Readings in the Theory of Growth*, pages 131–149. Springer.

Beaudry, C. and Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? the localization versus urbanization debate. *Research Policy*, 38(2):318–337.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. Springer, 2 edition.

Blien, U., Suedekum, J., and Wolf, K. (2006). Local employment growth in west germany: A dynamic panel approach. *Labour Economics*, 13(4):445–458.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, 51(3):351–364.

Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3):287–302.

Burger, M. J., van Oort, F. G., and van der Knaap, B. (2010). A treatise on the geographical scale of agglomeration externalities and the modifiable areal unit problem. *Scienze Regionali*, 9(1):19–40.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Caragliu, A., de Dominicis, L., and de Groot, H. L. (2016). Both marshall and jacobs were right! *Economic Geography*, 92(1):87–111.

Chen, J. (2019). Geographical scale, industrial diversity, and regional economic stability. *Growth and Change*, 50(2):609–633.

Cieślik, A., Gauger, I., and Michałek, J. J. (2018). Agglomeration externalities, competition and productivity: empirical evidence from firms located in ukraine. *The Annals of Regional Science*, 60(1):213–233.

Cliff, A. and Ord, J. (1981). *Spatial Processes: Models & Applications*. Pion.

Combes, P.-P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of urban economics*, 47(3):329–355.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, volume 5, pages 247–348. Elsevier.

Dauth, W. (2013). Agglomeration and regional employment dynamics. *Papers in Regional Science*, 92(2):419–435.

de Araújo, I. F., Gonçalves, E., and Almeida, E. (2019). Effects of dynamic and spatial externalities on local growth: Evidence from brazil. *Papers in Regional Science*, 98(2):1239–1259.

Delgado, M., Porter, M. E., and Stern, S. (2014). Clusters, convergence, and economic performance. *Research policy*, 43(10):1785–1799.

Deza, M. and Deza, E. (2012). *Encyclopedia of Distances*. Springer. Springer Heidelberg.

Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer New York.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier.

Duschl, M., Scholl, T., Brenner, T., Luxen, D., and Raschke, F. (2015). Industry-specific firm growth and agglomeration. *Regional Studies*, 49(11):1822–1839.

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial economic analysis*, 5(1):9–28.

Eurostat (2008). Nace rev. 2—statistical classification of economic activities in the european community. *Methodologies and working papers*. Theme: General and regional statistics.

Fischer, M. and Nijkamp, P. (2013). Regional science at full gallop: Editorial introduction. In *Handbook of Regional Science*, pages xxi – xxxvii. Springer.

Florax, R., Folmer, H., and Rey, S. J. (2003). Specification searches in spatial econometrics: the relevance of hendry's methodology. *Regional Science and Urban Economics*, 33(5):557–579.

Flowerdew, R. (2011). How serious is the modifiable areal unit problem for analysis of English census data? *Population trends*, 145(1):106–118.

Fotheringham, A. S. and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7):1025–1044.

Frenken, K., van Oort, F., and Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies*, 41(5):685–697.

Garcia-López, M.-À. and Muñiz, I. (2013). Urban spatial structure, agglomeration economies, and economic growth in barcelona: An intra-metropolitan perspective. *Papers in Regional Science*, 92(3):515–534.

GeoBasis-DE / BKG (2017). *VG1000*. German Federal Agency for Cartography and Geodesy. 2017–01–01. http://www.bkg.bund.de.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in cities. *Journal of political economy*, 100(6):1126–1152.

Griffith, D. A., Wong, D. W. S., and Whitfield, T. (2003). Exploring relationships between the global and regional measures of spatial autocorrelation. *Journal of Regional Science*, 43(4):683–710.

Helsley, R. W. and Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*, 20(2):189–212.

Henderson, V., Kuncoro, A., and Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5):1126–1090.

Hoogstra, G. J. and van Dijk, J. (2004). Explaining firm employment growth: does location matter? *Small business economics*, 22(3–4):179–192.

Illy, A., Hornych, C., Schwartz, M., and Rosenfeld, M. T. W. (2009). Urban growth in germany – the impact of localization and urbanization economies. IWH Discussion Papers 19/2009, Halle (Saale).

Jacobs, J. (1969). *The economy of cities*. Vintage Book Edition.

Kim, S. (1989). Labor specialization and the extent of the market. *Journal of Political Economy*, 97(3):692–705.

Krugman, P. (1991a). History and industry location: the case of the manufacturing belt. *The American Economic Review*, 81(2):80–83.

Krugman, P. (1991b). Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499.

Lesage, J. P. and Fischer, M. M. (2008). Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis*, 3(3):275–304.

LeSage, J. P. and Pace, R. K. (2006). Interpreting spatial econometric models. In Fischer, M. M. and Nijkamp, P., editors, *Handbook of Regional Science*, page 1535. Springer.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Li, Z., Ding, C., and Niu, Y. (2019). Industrial structure and urban agglomeration: evidence from chinese cities. *The Annals of Regional Science*, 63(1):191–218.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in r. *Behavior research methods*, 49(4):1494–1502.

Marshall, A. (1890). *Principles of Economics*. 1. Great Mind Series.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38:1–42.

Openshaw, S. and Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In *Statistical methods in the spatial sciences*, pages 127–144. N. Wrigley.

OW networks GmbH (2018). *Postleitzahlen Liste von Deutschland 2018*. https://www.geodaten-deutschland.de.

Pessoa, A. (2014). Agglomeration and regional growth policy: externalities versus comparative advantages. *The Annals of regional science*, 53(1):1–27.

Porter, M. E. (1990). The competitive advantage of nations. *Harvard Business Review*, 68(2):73–93.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 3.6.3.

Regionaldatenbank Deutschland (2020). *Erwerbstätigenrechnung d.Bundes u.d. Länder*. https://www.regionalstatistik.de. Code: 13312.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–1037.

Rosenthal, S. S. and Strange, W. C. (2001). The determinants of agglomeration. *Journal of urban economics*, 50(2):191–229.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2119–2171. Elsevier.

Rothgang, M., Cantner, U., Dehio, J., Engel, D., Fertig, M., Graf, H., Hinzmann, S., Linshalm, E., Ploder, M., Scholz, A.-M., et al. (2017). Cluster policy: insights from the german leading edge cluster competition. *Journal of Open Innovation: Technology, Market, and Complexity*, 3(18):1–20.

Rousset, F. and Ferdy, J.-B. (2020). *spaMM (reference manual): Mixed-Effect Models, Particularly Spatial Models*. R package version 3.3.0; Date: 2020–05–30.

Schwochow Softwareentwicklung (2019). *Polygons of German zip code levels*. 2019–08–07. https://www.suche-postleitzahl.org.

Statistisches Bundesamt (2008). *Klassifikation der Wirtschaftszweige*. Wiesbaden.

Stavropoulos, S., van Oort, F., and Burger, M. (2020). Heterogeneous relatedness and firm productivity. *The Annals of Regional Science*, pages 1–35.

Sternberg, R. and Litzenberger, T. (2004). Regional clusters in germany–their geography and their relevance for entrepreneurial activities. *European Planning Studies*, 12(6):767–791.

StMWi (2020). *CLUSTER OFFENSIVE BAYERN – Im Netzwerk zum Erfolg*. Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie, München. Version: May 2020.

Tanaka, K. and Hashiguchi, Y. (2020). Agglomeration economies in the formal and informal sectors: a bayesian spatial approach. *Journal of Economic Geography*, 20(1):37–66.

van Oort, F. G. (2007). Spatial and sectoral composition effects of agglomeration economies in the Netherlands. *Papers in Regional Science*, 86(1):5–30.

van Soest, D. P., Gerking, S., and van Oort, F. G. (2006). Spatial impacts of agglomeration externalities. *Journal of Regional Science*, 46(5):881–899.

van Soest, D. P., Gerking, S. D., and van Oort, F. G. (2002). Knowledge externalities, agglomeration economies, and employment growth in Dutch cities. Technical report, CENTER Discussion Paper No. 2002-41 .

Wong, D. (2009). The modifiable areal unit problem (MAUP). In Fotheringham, A. S. and Rogerson, P. A., editors, *The SAGE handbook of spatial analysis*, pages 105–124. SAGE publications London.

Yule, G. U. and Kendall, M. G. (1950). *Introduction to the Theory of Statistics.* Harper Publishing Company, 3 edition.

Appendix

A Tables

Table A1: Validity of the dataset.

Variable	Criteria
period per fiscal year	12 months
accounting practice	local GAAP
total assets in 2013 and 2017	$\geq$ €10000
sales in 2013 and 2017	$\geq$ €10000
number of employees in 2013 and 2017	≥ 1 & ≤ 2000
GR	≤ 100

Table A2: Comparison of employee shares per sector for Germany in 2017. We compare the Orbis data with census data from the regional database (Regionaldatenbank Deutschland, 2020).

NACE	WZ08	Orbis	Regional DB
01–03	A	0.0050	0.0139
05–39	B–E	0.2448	0.1848
41–43	F	0.0307	0.0560
45–98	G–T	0.7195	0.7453

Table A3: Comparison of employee shares per federal state for Germany in 2017. We compare the Orbis data with census data from the regional database (Regionaldatenbank Deutschland, 2020).

Federal State	Orbis	Regional DB
Baden-Wuerttemberg	0.1744	0.1413
Bavaria	0.1497	0.1701
Berlin	0.0442	0.0443
Brandenburg	0.0123	0.0251
Bremen	0.0059	0.0096
Hamburg	0.0301	0.0284
Hesse	0.1288	0.0779
Mecklenburg-Vorpommern	0.0065	0.0169
Lower Saxony	0.0589	0.0917
North Rhine-Westphalia	0.2672	0.2130
Rhineland-Palatinate	0.0318	0.0456
Saarland	0.0215	0.0120
Saxony	0.0214	0.0462
Saxony-Anhalt	0.0125	0.0227
Schleswig-Holstein	0.0243	0.0316
Thuringia	0.0105	0.0236

Table A4: Literature overview of MAUP-related studies.

Author(s)	Method	Data	Scale/Zoning	Results
(1) Literature Review				
Openshaw (1984)	Overview of prior MAUP literature	No own data	Different scaling and zoning systems in reviewed studies	The zoning problem is greater than the scaling problem because there is more freedom to choose the boundaries than the number of zones.
(2) Univariate and bivariate analysis				
Openshaw and Taylor (1979)	Correlation and spatial autocorrelation analysis	Percentage vote and population over 60 from the 1970 Census data for 99 counties in Iowa, United States	Different arrangements of counties in Iowa	Correlation coefficients vary from a moderate level for relatively disaggregated data to a high level for highly aggregated data.
Flowerdew (2011)	Correlation analysis	English 2001 Census data on ethnicity, housing tenure, marital status, car ownership, illness, and employment	Output areas, wards, districts	In most cases, correlations across different scales do not vary a lot. But in some cases, there are substantial effects of the MAUP. Correlation coefficients tend to increase with increasing scale aggregation.
Griffith et al. (2003)	Spatial autocorrelation analysis	Population density data of the 1990 Census data of 48 coterminous states and the District of Columbia, USA	Census block groups, counties, states	A weak-to-moderate level of positive spatial autocorrelation is found at all aggregation levels. The spatial autocorrelation increases with finer scale aggregations.
(3) Multivariate analysis				
Fotheringham and Wong (1991)	OLS regression, logit model, and spatial autocorrelation analysis	1980 American Census data for the Buffalo metropolitan area	Different aggregations of the original 871 block groups	Different spatial scales lead to changes in parameter estimates in both directions. An increase in the aggregation level leads to a rise in the standard error.
Rosenthal and Strange (2001)	OLS regression	Dun & Badstreet Marketplace database, 1996–1997	Zip codes, counties, and states	The discovered variability in results at different scales may be attributable to idiosyncratic characteristics of the explanatory variables.
Briant et al. (2010)	OLS regression and gravity model	Wage, employment, and market data from the French National Institute of Statistics and Economics, 1976-1996	Administrative regions, administrative departments, employment areas, grid zoning systems	Modeling issues have a greater influence on the results than the MAUP and therefore, are more important.
Burger et al. (2010)	OLS regression and SDM	Sectorial employment growth and agglomeration externalities using the Dutch LISA 2006 dataset	Municipalities, economic geographic areas, labor market regions	The spatial level of analysis is a scaling and gerrymandering problem. MAUP is highly relevant for agglomeration economies.
Arbia and Petrarca (2011)	Spatial simulation studies using the SARAR(1,1) model	Artificial data	Different scale levels on a square lattice grid	In general, aggregation leads to a loss in efficiency. However, the extent of efficiency loss depends on the extent of the spatial autocorrelation.
Chen (2019)	OLS regression, SAR model, SEM, SDM	Regional economic instability, industrial diversity, control variables from the Bureau of Economic Analysis, County Business Patterns, Census	Counties, states, economic areas, MSA	The relation between diversity and stability is strongly influenced by the scale.

Table A5: Literature overview of studies on agglomeration externalities.

Author(s)	Economic growth	Agglomeration indicators/externalities	Method	Results
Glaeser et al. (1992)	Change in the natural log of employment and wage	• Specialization/MAR: location quotient • Diversity/Jacobs: fraction of the city's employment of the largest 5 industries other than the regarded industry • Urbanization: tested by regressing growth outside the four largest industries on the employment growth in these industries • Competition: numbers of firms per worker in city-industry divided by the number of firms per worker in country-industry	OLS regressions	Specialization affects employment growth negatively while competition and city diversity enhances employment growth.
Combes (2000)	Difference between the (natural log) employment growth of sector s in zone i and the national employment growth of s	• Specialization/localization: location quotient • Urbanization: total employment density • Diversity: relative inverse HHI of sectoral employment concentration • Competition: inverse HHI of productive concentration	Sectoral OLS regressions	Agglomeration externalities significantly affect local growth. Employment density and diversity have different effects across the industrial and service sectors.
Rosenthal and Strange (2003)	Total new establishments' employment and number of firm births	• Specialization/localization: own industry employment (in different concentric-circles) • Urbanization: other industry employment (in different concentric-circles) • Lack of diversity: HHI of sectoral employment concentration (in different concentric-circles)	Tobit and probit fixed effects models	Localization economies attenuate rapidly with the distance.
Blien et al. (2006)	Local employment growth of sector s in zone i and time t	• Specialization: local employment share • Diversity: Krugman-diversification index	Dynamic panel model	Diversity positively influences employment growth of local industries. There is no clear evidence that positive localization externalities exist.
Van Soest et al. (2006)	Change in the natural log of employment	• Specialization: location quotient • Lack of diversity: employment in five other biggest industries in location i divided by the total employment in i • Competition: share of firms in industry s and location i of the total employment in i divided by the share of firms in s in the region of the total employment in the region • All variables are weighted with the distance	Regression using distance weighted variables (calcuated in distance between centroids)	Spatially lagged dependent variable are not statistically significant, suggesting that effects of agglomeration externalities decline quickly with distance.
Frenken et al. (2007)	Employment, productivity, unemployment, and inactivity growth	• Localization: LOS-index (degree of technological clustering based on input-output relations) • Urbanization: population density • Unrelated variety: entropy at the two-digit level • Related variety: weighted sum of entropy within each sector	Spatial lag model and first-order contiguity matrix (but also tested different matrices)	Related variety, i.e., Jacobs externalities, increases employment growth.

Continued on next page

Author(s)	Economic growth	Agglomeration indicators/externalities	Method	Results
van Oort (2007)	Change in the natural log of employment	• Localization: location quotient • Lack of diversity/Jacobs: area-based Gini-coefficient • Competition: share of establishments per worker in a municipal industry s of the same industry's establishments per worker in Netherland	OLS and spatial lag model with first-order contiguity matrix (but also tested different matrices)	Proximate externality measured by the spatially lagged explained and explanatory variables have limited effects on growth.
Illy et al. (2009)	Local employment growth rate of sector s in zone i normalized by the industry growth rate	• Specialization: location quotient • Diversity/Jacobs: relative inverse HHI of sectoral employment concentration • Competition: inverse HHI of productive concentration, firms are classified for different size ranges concerning the employees	OLS regression and Moran's I with row-standardized (binary) contiguity matrix	High levels of specialization affect urban growth positively. Jacobs externalities are not statistically significant.
Burger et al. (2010)	Mean-corrected increase in number of employees of sector s in region i	• Localization: local employment share • Urbanization: population density • Lack of diversity/Jacobs: area-based Gini-coefficient	OLS regression and SDM with row-standardized inverse distance matrix (but also tested different matrices)	It it difficult to distinguish whether different effects of agglomeration externalities at different spatial scales are due to the MAUP or diverse functioning of the externalities.
Dauth (2013)	Natural log of employment in region r and industry i at time t and difference in the natural log of employment	• Agglomeration: Ellison-Glaeser index • Agglomeration: Cluster index of Sternberg and Litzenberger (2004) • Diversity: Krugman diversification index	OLS regression and dynamic panel data model, including spatially lagged exogenous variables to account for spatial dependence	Employment growth has a higher degree of persistence in industrial agglomerations in Germany.
Garcia-López and Muñiz (2013)	Local (natural log of) employment concentration growth of sector s in zone i	• Specialization/MAR: location quotient • Diversity/Jacobs: relative inverse HHI of sectoral employment concentration • Localization: employment in sector s and municipality i • Urbanization: total employment density outside s in i • Distance-weighted specialization and diversity index	OLS and SEM with municipal and sectoral fixed effects using first-order contiguity matrix (but also tested different matrices)	The urban spatial structure is essential to explain intra-metropolitan growth.
Caragliu et al. (2016)	Gross percentage employment change in region industries	• Specialization/MAR: location quotient • Unrelated variety: entropy at the two-digit level • Related variety: weighted sum of entropy within each sector	Sectoral OLS regression and SARAR estimation to control for spatial heterogeneity of economic activities	Positive impact of MAR and Jacobs externalities on employment growth.
Cieślik et al. (2018)	Total factor productivity (TFP) of Ukrainian firms	• Specialization: 1) level of concentration of the industry within the regarded region and 2) level of concentration of the region within the regarded industry in the whole sample • Diversification: 1) competition between industries within the same region, 2) local competition between firms of the same industry	Panel model	Agglomeration externalities and competition affect the TFP. Other factors, such as firm size or capital intensity, also have an impact on the TFP.

Continued on next page

Author(s)	Economic growth	Agglomeration indicators/externalities	Method	Results
de Araújo et al. (2019)	Difference in the natural log of average wages in region r and sector s	• Specialization/MAR: location quotient • Diversity: relative inverse HHI of sectoral employment concentration	Dynamic panel model	In Brazil, high levels of regional specialization have a positive impact on regional growth in the long-run while diversity positively affects local growth.
Li et al. (2019)	Firm's value added	• Localization: total employment in own-industry activities • Urbanization: total employment in other-industry activities • Regional industrial dominance: Herfindahl index of sales	Panel model (within estimator to control the firm fixed effect)	Localization economies are less likely to arise in regions in which sectors are more dominated by a few large firms.
Stavropoulos et al. (2020)	Labor productivity	• Related variety: weighted sum of entropy within each two-digit sector • Unrelated variety: two-digit level entropy • Localization: location quotient • Density: population per km^2	Multilevel model	Related variety and localization have a positive effect on productivity. It has a larger effect in high-tech and high-service sectors.
Tanaka and Hashiguchi (2020)	Natural log of total factor productivity for sector s and region i	• Specialization/Localization: industry employment density • Diversity/Jacobs: natural log of inverse HHI of sectoral employment concentration • Competition: natural log of inverse HHI of sales concentration	Bayesian spatial method using a weight matrix based on travel times for the shortest paths	Employment density affects the productivity in informal sectors positively. Spatial dependence is present among informal firms.

Table A6: Descriptive statistics for the variables used in the global regressions. The first row for every spatial scale denotes the number of available region-sector combinations.

Scale	Statistic	lGR	lLQ	lLQ^2	$lHHI$	$lcomp$	$lsize$
2-digit zip	n	634					
	min	-1.3258	-5.2030	0.0000	-1.2787	-6.1211	-4.5272
	median	-0.0061	-0.1320	0.1977	-0.1654	-3.9595	-0.1607
	mean	0.0497	-0.2926	0.9202	-0.1977	-3.9375	-0.2408
	max	3.0236	2.1238	27.0709	0.2285	-1.7181	2.0631
	sd	0.3667	0.9136	2.0633	0.2128	0.7762	0.7569
county and indep. cities	n	2154					
	min	-7.6178	-6.2639	0.0000	-1.3753	-6.5862	-4.6805
	median	-0.0235	-0.1315	0.4774	-0.4115	-4.9278	-0.4329
	mean	0.0429	-0.3348	1.5883	-0.4566	-4.9064	-0.5268
	max	4.4629	3.6259	39.2365	0.3262	-1.7999	3.0316
	sd	0.5008	1.2150	3.1927	0.3274	0.8333	1.0642
3-digit zip	n	3361					
	min	-4.9359	-6.2639	0.0000	-1.4482	-6.5862	-4.6805
	median	-0.0216	-0.1676	0.5578	-0.4887	-5.1686	-0.5253
	mean	0.0469	-0.3395	1.7174	-0.5160	-5.1197	-0.5694
	max	3.5627	3.7263	39.2365	0.3262	-2.1050	3.6084
	sd	0.4726	1.2657	3.2718	0.3358	0.7763	1.0906

Table A7: Descriptive statistics for the variables used in the sectoral regressions at the 2-digit zip code level. For better interpretability given in levels, for the regressions the variables are transformed using the natural logarithm (see Equation (6)). The first row for every sector denotes the number of available regions.

Sector	Statistic	GR	LQ	LQ^2	HHI	$comp$	$size$
Manufac.	n	95					
	min	0.5159	0.0641	0.0041	0.4081	0.0031	0.2724
	median	1.0022	1.0840	1.1750	0.8537	0.0158	0.9514
	mean	1.0262	1.0755	1.4377	0.8583	0.0172	1.0981
	max	1.9208	3.1006	9.6139	1.1911	0.0724	6.0231
Energy	n	92					
	min	0.3030	0.0154	0.0002	0.4752	0.0087	0.0711
	median	0.9168	0.7067	0.5003	0.8535	0.0211	0.6986
	mean	1.2179	1.1063	2.4604	0.8330	0.0237	0.9423
	max	8.5117	6.9012	47.6260	1.0971	0.0729	5.5306
Infr.	n	95					
	min	0.7425	0.1887	0.0356	0.3013	0.0037	0.5819
	median	0.9909	1.0086	1.0173	0.8215	0.0256	0.9243
	mean	1.0604	1.1518	1.8731	0.8145	0.0281	1.0189
	max	3.4389	4.8495	23.5172	1.2070	0.0766	3.1601
Trade	n	95					
	min	0.6865	0.0784	0.0061	0.3307	0.0022	0.1929
	median	1.0009	0.9274	0.8600	0.8130	0.0169	0.8425
	mean	1.0590	0.9708	1.0877	0.8261	0.0225	1.0807
	max	3.5721	2.1584	4.6587	1.2568	0.0789	5.7918
F. & C.	n	94					
	min	0.2656	0.0433	0.0019	0.2936	0.0046	0.1305
	median	1.0057	0.6329	0.4009	0.8701	0.0195	0.6391
	mean	1.1791	0.9104	1.4603	0.8553	0.0224	0.9836
	max	5.9864	3.3784	11.4137	1.1853	0.0669	7.0887
Services	n	94					
	min	0.5363	0.0933	0.0087	0.3344	0.0031	0.1971
	median	1.0102	0.8866	0.7860	0.8484	0.0112	0.9405
	mean	1.0165	0.9418	1.0999	0.8451	0.0149	1.0594
	max	1.7924	2.2121	4.8936	1.2145	0.0544	4.1865
Social	n	69					
	min	0.6855	0.0055	0.0000	0.2784	0.0392	0.0108
	median	1.0466	0.4849	0.2351	0.8482	0.0480	0.4108
	mean	1.8542	1.1016	3.5753	0.8331	0.0661	0.9744
	max	20.5658	8.3625	69.9310	1.1265	0.1794	7.8700

Table A8: Descriptive statistics for the variables used in the sectoral regressions at the 2-digit zip code level. The first row for every sector denotes the number of available regions.

Sector	Statistic	lGR	lLQ	lLQ^2	$lHHI$	$lcomp$	$lsize$
Manufac.	n	95					
	min	-0.6618	-2.7478	0.0001	-0.8962	-5.7804	-1.3003
	median	0.0022	0.0806	0.1034	-0.1582	-4.1471	-0.0498
	mean	0.0107	-0.0954	0.4615	-0.1710	-4.2608	-0.0411
	max	0.6527	1.1316	7.5505	0.1749	-2.6255	1.7956
Energy	n	92					
	min	-1.1940	-4.1713	0.0031	-0.7439	-4.7416	-2.6436
	median	-0.0868	-0.3480	0.5457	-0.1584	-3.8603	-0.3599
	mean	0.0387	-0.4715	1.7629	-0.1953	-3.8721	-0.4320
	max	2.1414	1.9317	17.3997	0.0927	-2.6185	1.7103
Infr.	n	95					
	min	-0.2978	-1.6675	0.0001	-1.1996	-5.6046	-0.5414
	median	-0.0091	0.0086	0.1498	-0.1966	-3.6645	-0.0788
	mean	0.0316	-0.0494	0.4078	-0.2255	-3.8151	-0.0238
	max	1.2352	1.5789	2.7807	0.1881	-2.5698	1.1506
Trade	n	95					
	min	-0.3762	-2.5461	0.0000	-1.1065	-6.1211	-1.6454
	median	0.0009	-0.0754	0.0831	-0.2071	-4.0820	-0.1714
	mean	0.0293	-0.1260	0.2598	-0.2203	-4.0783	-0.0641
	max	1.2732	0.7694	6.4827	0.2285	-2.5392	1.7564
F. & C.	n	94					
	min	-1.3258	-3.1395	0.0000	-1.2256	-5.3715	-2.0366
	median	0.0057	-0.4579	0.6227	-0.1392	-3.9368	-0.4478
	mean	0.0520	-0.5058	1.2095	-0.1773	-3.9974	-0.3679
	max	1.7895	1.2174	9.8567	0.1700	-2.7042	1.9585
Services	n	94					
	min	-0.6231	-2.3716	0.0000	-1.0954	-5.7655	-1.6239
	median	0.0101	-0.1204	0.1101	-0.1644	-4.4935	-0.0613
	mean	-0.0099	-0.2213	0.4613	-0.1920	-4.4030	-0.1288
	max	0.5835	0.7940	5.6245	0.1943	-2.9111	1.4319
Social	n	69					
	min	-0.3776	-5.2030	0.0002	-1.2787	-3.2402	-4.5272
	median	0.0456	-0.7238	1.2687	-0.1646	-3.0362	-0.8897
	mean	0.2490	-0.6963	2.2739	-0.2033	-2.8386	-0.7827
	max	3.0236	2.1238	27.0709	0.1191	-1.7181	2.0631

Table A9: Descriptive statistics for the variables used in the sectoral regressions at the county and independent city level. For better interpretability given in levels, for the regressions the variables are transformed using the natural logarithm (see Equation (6)). The first row for every sector denotes the number of available regions.

Sector	Statistic	GR	LQ	LQ^2	HHI	$comp$	$size$
Manufac.	n	393					
	min	0.2019	0.0277	0.0008	0.2576	0.0014	0.0454
	median	0.9966	1.1450	1.3111	0.6665	0.0050	0.6990
	mean	1.1451	1.2365	2.1449	0.6614	0.0068	1.1626
	max	6.0973	3.3727	11.3750	1.2273	0.0376	15.2419
Energy	n	252					
	min	0.0486	0.0019	0.0000	0.2672	0.0087	0.0093
	median	0.8901	0.7168	0.5139	0.6910	0.0101	0.3590
	mean	1.2514	1.7857	10.0706	0.6790	0.0136	0.8596
	max	25.4999	16.3378	266.9233	1.1647	0.0412	14.4216
Infr.	n	377					
	min	0.5084	0.0178	0.0003	0.2621	0.0015	0.0440
	median	0.9566	1.1908	1.4181	0.6040	0.0063	0.9022
	mean	1.1333	1.6144	5.1242	0.6156	0.0085	1.1022
	max	8.2616	12.0086	144.2075	1.2357	0.0817	20.7296
Trade	n	391					
	min	0.3455	0.0461	0.0021	0.2677	0.0014	0.0926
	median	1.0090	0.8760	0.7673	0.6160	0.0067	0.6390
	mean	1.1726	1.0238	1.4842	0.6410	0.0089	0.9876
	max	14.6027	3.8039	14.4695	1.3857	0.0882	18.0706
F. & C.	n	296					
	min	0.0346	0.0033	0.0000	0.2528	0.0046	0.0135
	median	0.9772	0.4362	0.1903	0.6927	0.0078	0.3219
	mean	1.6049	0.8344	1.8161	0.6942	0.0106	0.7854
	max	86.7423	7.6410	58.3841	1.2449	0.0751	16.0323
Services	n	351					
	min	0.0005	0.0046	0.0000	0.2684	0.0027	0.0098
	median	0.9918	0.6131	0.3759	0.7024	0.0052	0.4391
	mean	1.3152	0.8744	1.4453	0.7057	0.0074	1.0825
	max	19.2772	3.8359	14.7141	1.2722	0.0649	14.4918
Social	n	94					
	min	0.3428	0.0169	0.0003	0.3565	0.0392	0.0108
	median	0.9314	1.2305	1.5143	0.7540	0.0392	0.3081
	mean	1.5897	2.8354	38.7691	0.7334	0.0523	0.8988
	max	20.5658	37.5587	1410.6595	1.1290	0.1653	14.3346

Table A10: Descriptive statistics for the variables used in the sectoral regressions at the county and independent city level. The first row for every sector denotes the number of available regions.

Sector	Statistic	lGR	lLQ	lLQ^2	$lHHI$	$lcomp$	$lsize$
Manufac.	n	393					
	min	-1.5998	-3.5848	0.0000	-1.3562	-6.5382	-3.0921
	median	-0.0034	0.1354	0.3225	-0.4057	-5.3083	-0.3581
	mean	0.0544	-0.0820	0.8221	-0.4675	-5.2478	-0.2402
	max	1.8078	1.2157	12.8510	0.2049	-3.2813	2.7240
Energy	n	252					
	min	-3.0231	-6.2639	0.0000	-1.3196	-4.7416	-4.6805
	median	-0.1164	-0.3329	1.2491	-0.3696	-4.5942	-1.0245
	mean	-0.0219	-0.4389	2.8916	-0.4290	-4.3946	-1.0441
	max	3.2387	2.7935	39.2365	0.1525	-3.1889	2.6687
Infr.	n	377					
	min	-0.6765	-4.0305	0.0000	-1.3392	-6.4986	-3.1233
	median	-0.0444	0.1747	0.3822	-0.5042	-5.0640	-0.1029
	mean	0.0314	0.0787	0.9238	-0.5341	-5.0979	-0.0830
	max	2.1116	2.4856	16.2448	0.2116	-2.5041	3.0316
Trade	n	391					
	min	-1.0626	-3.0779	0.0000	-1.3179	-6.5862	-2.3794
	median	0.0090	-0.1324	0.2288	-0.4846	-5.0083	-0.4478
	mean	0.0652	-0.2029	0.5613	-0.5091	-5.0255	-0.3081
	max	2.6812	1.3360	9.4735	0.3262	-2.4282	2.8943
F. & C.	n	296					
	min	-3.3629	-5.7204	0.0000	-1.3753	-5.3715	-4.3079
	median	-0.0230	-0.8297	1.1407	-0.3671	-4.8481	-1.1334
	mean	0.0143	-0.9307	2.7867	-0.4106	-4.7638	-0.9579
	max	4.4629	2.0335	32.7225	0.2191	-2.5890	2.7746
Services	n	351					
	min	-7.6178	-5.3893	0.0000	-1.3153	-5.9139	-4.6296
	median	-0.0082	-0.4893	0.6297	-0.3533	-5.2632	-0.8229
	mean	0.0642	-0.7426	2.2390	-0.3908	-5.1779	-0.7147
	max	2.9589	1.3444	29.0443	0.2407	-2.7351	2.6736
Social	n	94					
	min	-1.0707	-4.0779	0.0019	-1.0314	-3.2402	-4.5272
	median	-0.0711	0.2073	0.8142	-0.2823	-3.2402	-1.1775
	mean	0.1326	0.0789	2.0319	-0.3458	-3.0228	-0.9683
	max	3.0236	3.6259	16.6291	0.1213	-1.7999	2.6627

Table A11: Descriptive statistics for the variables used in the sectoral regressions at the 3-digit zip code level. For better interpretability given in levels, for the regressions the variables are transformed using the natural logarithm (see Equation (6)). The first row for every sector denotes the number of available regions.

Sector	Statistic	GR	LQ	LQ^2	HHI	$comp$	$size$
Manufac.	n	626					
	min	0.1169	0.0024	0.0000	0.2576	0.0014	0.0757
	median	1.0060	1.0794	1.1652	0.6221	0.0041	0.6211
	mean	1.1574	1.2147	2.1949	0.6325	0.0055	1.1226
	max	8.1317	3.3292	11.0837	1.2299	0.0483	12.8277
Energy	n	332					
	min	0.1254	0.0019	0.0000	0.2350	0.0087	0.0093
	median	0.8909	0.8833	0.7804	0.6331	0.0087	0.3339
	mean	1.1757	1.7625	8.4546	0.6403	0.0116	0.8228
	max	14.8041	12.1844	148.4593	1.1647	0.0336	15.4232
Infr.	n	607					
	min	0.3044	0.0301	0.0009	0.2370	0.0015	0.0440
	median	0.9547	1.0965	1.2023	0.5682	0.0046	0.8362
	mean	1.1458	1.6775	6.1307	0.5811	0.0057	1.0340
	max	10.5336	15.0569	226.7098	1.2239	0.0342	20.7296
Trade	n	645					
	min	0.1343	0.0041	0.0000	0.2677	0.0014	0.0926
	median	1.0041	0.8610	0.7414	0.5805	0.0052	0.6095
	mean	1.2089	1.0721	1.7530	0.6029	0.0067	1.1123
	max	17.5096	4.0836	16.6754	1.3857	0.0536	36.9052
F. & C.	n	476					
	min	0.0072	0.0038	0.0000	0.2528	0.0046	0.0135
	median	0.9840	0.4634	0.2147	0.6536	0.0070	0.3253
	mean	1.3906	0.9923	2.6747	0.6626	0.0088	0.8336
	max	35.2586	8.4743	71.8139	1.2221	0.0491	20.9726
Services	n	550					
	min	0.0090	0.0040	0.0000	0.2684	0.0027	0.0098
	median	0.9909	0.6400	0.4096	0.6582	0.0049	0.3806
	mean	1.2619	0.9088	1.5708	0.6744	0.0060	1.0134
	max	18.7484	4.0009	16.0070	1.2545	0.0293	11.9155
Social	n	125					
	min	0.3428	0.0433	0.0019	0.2564	0.0392	0.0108
	median	0.9264	1.7092	2.9213	0.6459	0.0392	0.2919
	mean	1.3910	3.9740	66.9355	0.6500	0.0444	0.8707
	max	20.5658	41.5243	1724.2655	1.1290	0.1219	14.3346

Table A12: Descriptive statistics for the variables used in the sectoral regressions at the 3-digit zip code level. The first row for every sector denotes the number of available regions.

Sector	Statistic	lGR	lLQ	lLQ^2	$lHHI$	$lcomp$	$lsize$
Manufac.	n	626					
	min	-2.1468	-6.0163	0.0000	-1.3562	-6.5382	-2.5812
	median	0.0060	0.0764	0.3961	-0.4746	-5.5088	-0.4762
	mean	0.0592	-0.1990	1.2550	-0.5113	-5.4595	-0.3173
	max	2.0958	1.2027	36.1955	0.2069	-3.0309	2.5516
Energy	n	332					
	min	-2.0766	-6.2639	0.0001	-1.4482	-4.7416	-4.6805
	median	-0.1156	-0.1243	1.2005	-0.4571	-4.7416	-1.0970
	mean	-0.0176	-0.3389	2.4971	-0.4946	-4.5202	-1.0878
	max	2.6949	2.5002	39.2365	0.1525	-3.3931	2.7359
Infr.	n	607					
	min	-1.1895	-3.5023	0.0000	-1.4399	-6.4986	-3.1233
	median	-0.0464	0.0921	0.5240	-0.5652	-5.3837	-0.1789
	mean	0.0300	0.0165	1.1566	-0.5955	-5.3971	-0.1462
	max	2.3546	2.7118	12.2661	0.2020	-3.3743	3.0316
Trade	n	645					
	min	-2.0076	-5.4960	0.0000	-1.3179	-6.5862	-2.3794
	median	0.0041	-0.1496	0.2802	-0.5439	-5.2669	-0.4952
	mean	0.0755	-0.2248	0.7889	-0.5678	-5.2601	-0.3320
	max	2.8627	1.4070	30.2065	0.3262	-2.9260	3.6084
F. & C.	n	476					
	min	-4.9359	-5.5742	0.0000	-1.3753	-5.3715	-4.3079
	median	-0.0162	-0.7692	1.2411	-0.4252	-4.9583	-1.1230
	mean	0.0299	-0.8895	3.0620	-0.4607	-4.8767	-1.0332
	max	3.5627	2.1370	31.0721	0.2005	-3.0141	3.0432
Services	n	550					
	min	-4.7152	-5.5284	0.0000	-1.3153	-5.9139	-4.6296
	median	-0.0091	-0.4463	0.7097	-0.4183	-5.3239	-0.9660
	mean	0.0683	-0.7191	2.2071	-0.4425	-5.2819	-0.7822
	max	2.9311	1.3865	30.5635	0.2268	-3.5299	2.4778
Social	n	125					
	min	-1.0707	-3.1407	0.0045	-1.3611	-3.2402	-4.5272
	median	-0.0765	0.5360	0.9455	-0.4371	-3.2402	-1.2314
	mean	0.0621	0.4003	2.2004	-0.4763	-3.1514	-1.0330
	max	3.0236	3.7263	13.8851	0.1213	-2.1050	2.6627

Table A13: Correlation between the regression variables lLQ and $lHHI$, both on a global and a sectoral level.

Scale/Sector	2-digit zip	county and indep. cities	3-digit zip
Scale	0.0542	0.0085	0.0097
Manufac.	-0.0952	-0.0392	-0.0524
Energy	0.0907	-0.1264	-0.0302
Infr.	0.0875	-0.0256	-0.0299
Trade	0.2584	0.2515	0.1570
F. & C.	-0.0753	0.0696	0.0438
Services	0.0957	0.2023	0.1982
Social	0.1512	-0.0491	-0.0323

Table A14: Regression results for all sectors at 2-digit zip code level.

Coef./Fit	Manufac.	Energy	Infr.	Trade	F. & C.	Services	Social
constant	−0.1097	−0.3388	−0.2255	0.0732	0.0326	0.3956**	0.0731
	(0.1598)	(0.3754)	(0.1581)	(0.1345)	(0.3313)	(0.1963)	(0.3820)
lLQ	0.0240	−0.0630	−0.0569	−0.1571**	−0.1047	−0.0146	−0.1871
	(0.0595)	(0.0573)	(0.0402)	(0.0708)	(0.0912)	(0.1010)	(0.1215)
lLQ^2	−0.0018	−0.0273	0.0819	−0.0157	0.0207	0.0226	0.0954***
	(0.0229)	(0.0313)	(0.0516)	(0.0921)	(0.0421)	(0.0397)	(0.0176)
$lHHI$	−0.1160	0.6613*	−0.3008**	−0.2047*	0.0848	−0.1418	0.4608
	(0.0750)	(0.3921)	(0.1422)	(0.1074)	(0.2185)	(0.1088)	(0.3856)
$lcomp$	−0.0225	−0.1136	−0.0390	0.0256	0.0148	0.1022**	−0.0132
	(0.0373)	(0.1132)	(0.0414)	(0.0342)	(0.0810)	(0.0448)	(0.1338)
$lsize$	−0.1837***	−0.1977**	−0.1725	0.0047	−0.0428	−0.0287	0.1473
	(0.0568)	(0.0867)	(0.1142)	(0.0769)	(0.1106)	(0.0804)	(0.1283)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Obs.	95	92	95	95	94	94	69
BP test	12.4788***	12.7985***	39.2964***	60.5326***	12.9107***	10.0555***	15.5902***
R^2	0.2165	0.1921	0.3195	0.1974	0.1435	0.1158	0.4130
Adj. R^2	0.1724	0.1451	0.2813	0.1523	0.0948	0.0655	0.3665
Wald (F)	8.8267***	3.6805***	2.7012***	3.1648***	2.0732***	2.3047***	8.6115***

Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis ‘homoscedasticity’ of the studentized Breusch-Pagan (BP) test had to be rejected (5% level). Standard errors are given in parentheses.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

Table A15: Regression results for all sectors at county and independent city level, which are estimated by OLS.

Coef./Fit	Manufac.	Energy	Infr.	F. & C.	Services
constant	−0.1987	0.1581	−0.2345**	0.2695	−0.0482
	(0.1511)	(0.3190)	(0.1158)	(0.3289)	(0.2900)
lLQ	−0.0020	−0.0458	−0.0616***	−0.1478**	−0.0980
	(0.0328)	(0.0315)	(0.0237)	(0.0697)	(0.0733)
lLQ^2	−0.0047	−0.0009	0.0520***	−0.0135	−0.0465***
	(0.0219)	(0.0143)	(0.0164)	(0.0140)	(0.0177)
$lHHI$	0.0099	0.1757*	−0.0556	0.1206	−0.0539
	(0.0632)	(0.0968)	(0.0710)	(0.1263)	(0.1185)
$lcomp$	−0.0425	0.0416	−0.0353	0.0680	−0.0029
	(0.0308)	(0.0756)	(0.0271)	(0.0718)	(0.0574)
$lsize$	−0.1609***	−0.0583	−0.1575***	−0.0193	−0.1507***
	(0.0381)	(0.0422)	(0.0388)	(0.0647)	(0.0533)
Model	OLS	OLS	OLS	OLS	OLS
Obs.	393	252	377	296	351
BP test	18.4137***	19.4199***	45.8683***	10.2168***	7.3175
R^2	0.1105	0.0918	0.2040	0.0710	0.1087
Adj. R^2	0.0990	0.0733	0.1933	0.0549	0.0958
Wald (F)	6.0968***	2.4398***	7.4494***	4.4295***	8.4137***

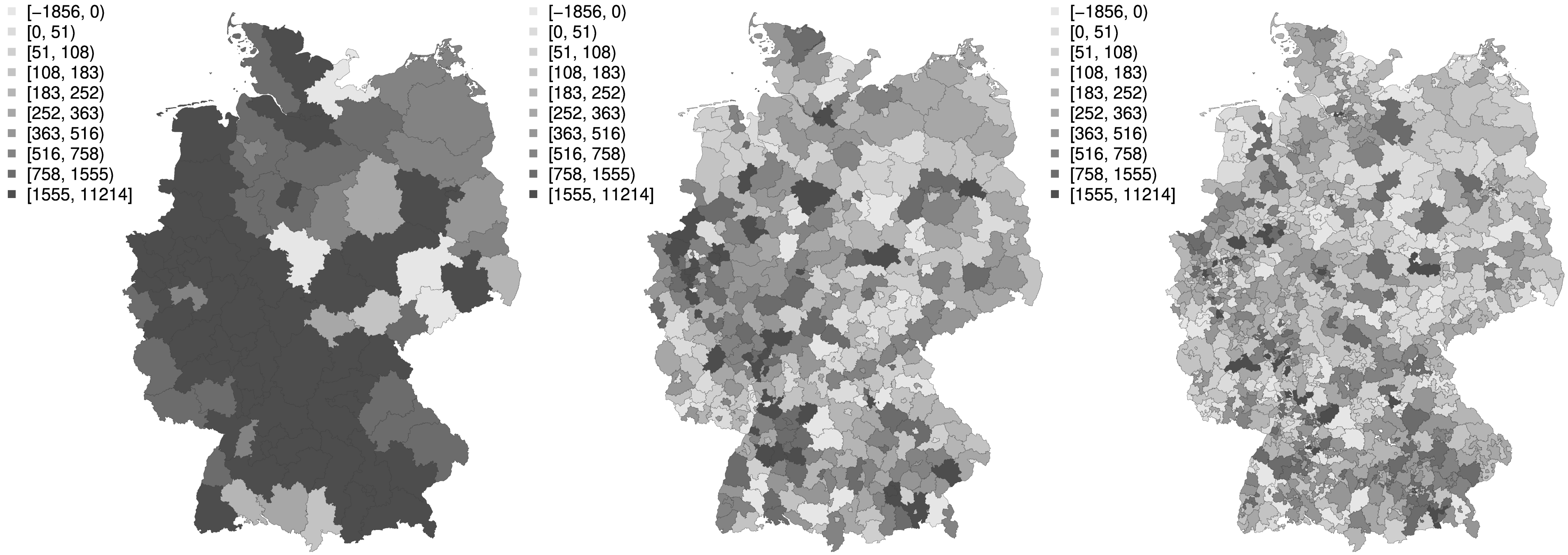
Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis 'homoscedasticity' of the studentized Breusch-Pagan (BP) test had to be rejected (5% level). Standard errors are given in parentheses.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

Table A16: Regression results for all sectors at county and independent city level, which are estimated by SLX.

Coef./Fit	Trade	Social
constant	1.3276	-10.1482^{*}
	(1.6974)	(5.4139)
lLQ	-0.0420	-0.1437^{*}
	(0.0280)	(0.0735)
lLQ^2	-0.0212	0.0717^{***}
	(0.0188)	(0.0222)
$lHHI$	0.0698	-0.0217
	(0.0737)	(0.2125)
$lcomp$	-0.0581	-0.0188
	(0.0442)	(0.1435)
$lsize$	-0.1501^{***}	-0.0425
	(0.0407)	(0.0868)
Lag lLQ	-0.4856	0.7711
	(0.3350)	(0.5455)
Lag lLQ^2	-0.3059	0.0123
	(0.3929)	(0.1654)
Lag $lHHI$	-0.9674^{*}	-0.9678
	(0.5686)	(2.9922)
Lag $lcomp$	0.4095	-2.7021
	(0.4034)	(1.8187)
Lag $lsize$	-0.2307	-1.5259^{*}
	(0.3745)	(0.8350)
Model	SLX	SLX
Obs.	391	94
BP test	41.9859^{***}	29.5128^{***}
R^2	0.1187	0.3572
Adj. R^2	0.0955	0.2798
Wald (F)	3.5014^{***}	2.3201^{***}

Notes: Robust standard errors (covariance matrix under HC0) are used if the null hypothesis 'homoscedasticity' of the studentized Breusch-Pagan (BP) test had to be rejected (5% level). Standard errors are given in parentheses.
* Significance at the 10% level. ** Significance at the 5% level. *** Significance at the 1% level.

B Figures



(a) 2-digit zip

(b) county and independent cities

(c) 3-digit zip

Figure B1: Not normalized growth of employees from 2013 to 2017 ($emp_r^{17} - emp_r^{13}$) for all three spatial scales. Intervals are based on the deciles of $emp_r^{17} - emp_r^{13}$ for all spatial scales combined.