DIW BERLIN

2045

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2023

Agglomeration or Market Access? The Defining Factors of Firms' Location Choice

Dennis Gaus and Georg Hirte

IMPRESSUM

DIW Berlin, 2023

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
https://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
https://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
https://ideas.repec.org/s/diw/diwwpp.html
https://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Agglomeration or Market Access?
The Defining Factors of Firms' Location Choice

Dennis Gaus [a,*]
Georg Hirte [b]

June 2023

Abstract

As research indicates a gap between complex scientific measures of accessibility and simpler proxies used by firms, this paper analyses the impact of several market access indicators on the location decision of firms. It compares the role of inter- and intra-industry agglomeration as proxies of access with a newly developed gravity-based indicator incorporating transport distances and industry relations. The estimation results of a nested mixed multinomial logit model, based on a sample of 110,083 German firms, provide evidence that agglomeration effects play an essential role in firms' location choice, whereas the complex market access measure does not have a significant impact. This outcome holds true for large as well as small and medium sized enterprises and is confirmed in several robustness checks. Thus, the paper provides guidance to further research on companies' location decisions, highlighting that access indicators should be chosen specifically for the scientific context, as well as to firms to make more efficient location choices from the perspective of market access.

Keywords: Transportation; Accessibility; Location Choice; Agglomeration; Market Access

JEL classification: L14, O18, R12, R32

[a]) DIW Berlin e.V., Mohrenstraße 58, 10117 Berlin, DE.
[b]) Technische Universität Dresden, 01062 Dresden, Germany.
*) Corresponding author. E-mail address: dgaus@diw.de

1. Introduction

Recent Research published points out a discrepancy between increasingly complex methodologies measuring the attractivity of locations, on the one hand, and the reality of firms and individuals deciding on a location, on the other hand (Ahuja & Tiwari, 2021; Pot et al., 2021). This paper analyzes the impact of agglomeration and market access on firm's location decisions by comparing the role of various accessibility measures in a location choice model. Sophisticated scientific models explaining location decisions and why some regions are seemingly more attractive than others have been developed on the back of extensive datasets, increasing computational power, and refined theoretical and econometric approaches. An essential aspect in many studies is the accessibility of locations, assuming that access to relevant institutions is a crucial indicator of regions' attractivity and success (De Bok & Van Oort, 2011).

This holds true for many fields of economic research, with different decisions and decision makers at the focus of attention: From an urban planning perspective, individuals' access to public facilities, like hospitals and public transport, is of high importance (Schirmer et al., 2014); for labor researchers, access to jobs and personal development opportunities is of interest (Fingleton & Szumilo, 2019); and the field of industrial organization shows great interest in the relevance of firms' access to potential customers and suppliers, coined as *market access* (De Bok & Van Oort, 2011). While existing research emphasizes the importance of accessibility and provides evidence that access measures can at least partially explain location decisions, more recent results suggest that the progress in the understanding of the location decisions of firms has come to a halt (Balbontin & Hensher, 2019). Although more comprehensive methods of measuring accessibility are combined with increasing data availability, the additional explanatory power over simpler approaches is small, if existent at all. A potential explanation is that firms do not base their location decisions on complex measures of accessibility with high data and computational requirements but use more straightforward, easily observable proxies to identify attractive regions when (re-)locating their activities (ibd.). One of the most easily observable measures in this context is the existence of firms in the same industry: Firms tend to go where similar businesses already are. This extensively studied and empirically supported behavior leads to industry clusters, allowing firms to benefit from so called *agglomeration effects* (De Bok & Van Oort, 2011; Nielsen et al., 2017). Besides agglomeration, there is evidence that several other measures influence the location decisions of firms, including the existence of certain types of transport infrastructure and the availability of potential suppliers and customers (Balbontin & Hensher, 2019).

Based on these findings of existing literature, this paper takes a new perspective on the relevance of accessibility for the location decision of firms: Instead of defining market access in a specific way and analyzing whether this measure affects the decisions of firms, it compares the impact of four accessibility measures. On the one hand, it includes three easily observable proxies for market access: (1) the existence of firms in the same industry (intra-industrial agglomeration), (2) the existence of potential suppliers and customers (intra-supply-chain agglomeration), and (3) the relative economic importance of a firm within a region. On the other hand, it develops a novel indicator of market access by combining road transport distances with industrial relations and firm size to identify a firm- and region-specific market potential.

The impact of all four variables on the location decision of firms is analyzed and compared by estimating a nested multinomial logit (MNL) location choice model based on a sample of 110,083 German firms. In this model, firms choose their location from a set of 50 areas in Germany in two steps: first, they decide on a greater geographical region represented by the federal states, and second, they choose an area within the state based on rurality and socioeconomic status. A combination of location- and firm-specific characteristics is used to explain the observed decisions. This approach adds to the literature by determining which measures of access are most appropriate in which context, contributing to a better understanding of firms' locations decisions and improving economic methods building on them.

It is found that the proxies have a significant impact on firms' location decision, whereas the complex access measure is not considered in the decision-making process. Besides pointing out that the preferable market access indicator depends on the context and goal of the analysis, the findings reveal the efficiency of firms' location decisions: As market access is known to be a driver of economic development (Melo et al., 2017), using precise measures of accessibility when facing location decisions can be of crucial importance for companies' success.

The paper is structured as follows: Section 2 provides an overview of the existing literature. Section 3 describes the methodological approach used for the location choice analysis and the newly developed market access indicator, while Section 4 presents the dataset. Section 5 displays and discusses the estimation results and several robustness checks, and Section 6 closes with a summary, conclusions, and suggestions for further research.

2. Literature Review & Contribution

Since the advent of the new economic geography (NEG), extensive research seeks to understand those factors influencing firms' location decisions. NEG builds on the idea that economic activity is not distributed randomly in space but rather follows patterns of interaction between market participants (Fujita et al., 1999; Krugman, 1991; Venables, 2010). A direct consequence of this hypothesis is that firms decide actively on their location considering their market needs, the position of other market participants, external factors (e.g., politics), and other aspects. With constant discussions and developments of the underlying theories and approaches (Fingleton, 2007; Martin, 1999), a wide range of methods for empirical analyses has been developed (Commendatore et al., 2018). Two main subjects of research are observed in this context: the evaluation of those factors determining the location of economic activity and the assessment of the impact of location-specific characteristics on costs and prices.

The first aspect is commonly referred to as *location choice*. Including the location choice of individuals (representing consumer demand and labor supply) and firms (describing the supply of goods and services and labor demand), it aims to understand how these two groups choose their locations and interact in their location decisions. An extensive review of the literature on residential location choices is provided by Schirmer et al. (2014), identifying the built environment (e.g., number of houses), socioeconomic aspects (incl. house prices), points of interest (e.g., hospitals), and accessibility (e.g., public transport) as relevant factors determining residential choice. Zondag & Pieters (2005) provide a deeper analysis of the accessibility effects. O'Sullivan (2005) lays out the theory behind firms' location choices, while Balbontin & Hensher (2019) summarize the extensive empirical research on the topic. They identify *push factors* pushing firms away and *pull factors* pulling firms into locations, while differentiating between firm-specific characteristics and location-specific attributes influencing the attractiveness of a region for a firm.

Further research focuses on specific groups of establishments, such as relocating firms (Pellenbarg et al., 2002) or foreign direct investment (Kim & Aguilera, 2016; Nielsen et al., 2017). Analyses of German firms include Buch et al. (2005) and Krenz (2016). Methodologically, location choice models are commonly based on a MNL estimation (Carlton, 1983; McFadden, 1974; cf. Section 3.1), even though alternatives, like generalized extreme value models (Train, 2003) and probit models (Dahlberg & Eklöf, 2003; Kropko, 2008), exist.

The analysis of the effects of location aims to identify how the characteristics of a specific location influences an output variable. In the residential location choice, an example is the impact of a new means of transportation on housing prices (Gibbons & Machin, 2005). More common, however, is the evaluation of productivity effects: A wide range of literature exploring how the productivity of firms and regions depends on location and accessibility has evolved (Johansson, 1993; Melo et al., 2017). Following the approaches of productivity research in general, this strand relies on the estimation of production functions (Martín-Barroso et al., 2015), even though some authors use different methods such as propensity score matching (Petersen, 2011). Although accessibility is commonly found to have a significant effect on regional productivity, there is a methodological drawback to the production function approach: It is not entirely clear to what degree firms choose their location based on accessibility, predetermining their productivity level through their location choice. Endogeneity is a common issue in various production function applications (Felipe et al., 2008), with both Börjesson et al. (2019) and Graham & van Dender (2011) discussing the potential selectivity bias in the given context. Therefore, it is necessary to understand the location decision of firms when analyzing the productivity effects of locations and accessibility.

As the existing literature points out, accessibility plays a major role in researching locations and spatial interaction. However, there is no clear definition of accessibility, which is reflected by the wide range of indicators used across analyses. These range from measures describing the existence of certain infrastructure characteristics (Limao & Venables, 1999) to complex measures incorporating various modes of transport and their respective transport times and costs (Donaldson & Hornbeck, 2016; Graham & Gibbons, 2019). Ahuja & Tiwari (2021) provide an overview of some standard measures and Bröcker (2006) adds a theoretical discussion.

In the context of location choice, a common finding is that relatively simple proxies for accessibility, such as the existence of certain types of infrastructure, have a significant impact on the probability of firms to choose a particular location, whereas more complex measures have less of an effect (De Bok & Sanders, 2005; Graham & Gibbons, 2019). A measure that is consistently found to have a significant impact on firms' location decisions is the existence of firms in particular industries. Economic theory suggests, and empirical studies confirm, that firms tend to co-locate to similar firms, which is known as *agglomeration* (Rosenthal & Strange, 2004). Providing firms with economies of scale and scope as well as low transportation and transaction costs, agglomeration effects are the main reason for the formation of industrial clusters and an essential aspect of firms' location choice (De Bok & Van Oort, 2011).

In addition, agglomeration has a strong impact on firms' perception of accessibility in a certain region. Eickelpasch et al. (2016) and Pot et al. (2021) find that firms rely more on their perception of accessibility than on measured indicators, suggesting a bias in their location decisions. As empirical evidence for this specific behavior is lacking, this paper compares the impact of four measures of accessibility on the location choice of firms, identifying the differences between easily observable proxies and complex scientific measures. This is relevant for three reasons: First, understanding how firms define market access in their location decisions helps model these decisions more realistically in applications where accessibility is not the primary concern (e.g., general equilibrium models or transport network analyses; Hadas et al., 2017). Secondly, the analysis contributes to a quantification of the potential endogeneity bias in productivity analyses, providing evidence whether respective correction procedures are necessary (Bourguignon et al., 2007). Thirdly, the differentiation between perceived and actual accessibility can help promote the use of more comprehensive indicators and supporting firms in making more efficient location decisions.

3. Methodology

To explain the location decision of firms and analyze the impact of certain variables, a mixed MNL is estimated. The first part of this section explains this method in general (Carlton, 1983; McFadden, 1974; Train, 2003), while the second part provides details on the variables used in this application. Special attention is paid to the newly developed accessibility indicator.

3.1 Multinomial Logit Model

The explanation of the location decision of firms using MNL is based on the assumption of profit maximization: A firm settles where it can obtain the highest profits (Carlton, 1983). Assuming that firm *i* has full information and can obtain profit π_{il} in region *l*, its location is in the specific region k if $\pi_{ik} \geq \pi_{il} \ \forall \ l \neq k$. Following McFadden (1974), the underlying profit function consists of a deterministic part and a stochastic part: $\pi_{il} = \mu_{il} + \varepsilon_{il}$. In what some authors call a mixed MNL[1], the deterministic part μ_{il} explains profit by two sets of variables; a linear additive functional form is assumed for both: the first set contains firm-specific variables X_i with location-specific impacts β_l, and the second set consists of location-specific variables Z_{il} with coefficients α (Hoffman & Duncan, 1988). The MNL utilizes a probability distribution to deduce a probability function, assigning a probability P to each choice option for each firm: $P_i(k|X_i, Z_{il})$ is the probability that firm *i* locates itself in region *k* given the characteristics of the firm and all regions. Under the profit maximization condition and incorporating the two parts of the profit function, the probability term can be rewritten as follows (McFadden, 1974):

$$P_i(k|X_i, Z_{il}) = P(\pi_{ik} \geq \pi_{il}) = P(\varepsilon_{il} - \varepsilon_{ik} \leq \mu_{ik} - \mu_{il}) \ \forall \ l \neq k \tag{1}$$

[1] This model should not be confused with the Random Regressors MNL (Train, 2003), which is also commonly referred to as a Mixed Logit model.

This probability is converted into an exponential function using the fact that $\mu_{il} = X_i\beta_l + Z_{il}\alpha$ (Hoffman & Duncan, 1988; McFadden, 1974):

$$P_i(k|X_i, Z_{il}) = \frac{\exp(X_i\beta_k + Z_{ik}\alpha)}{\sum_{l=1}^{L}\exp(X_i\beta_l + Z_{il}\alpha)} \tag{2}$$

The observed behavior is expressed using a binary dependent variable for the estimation: It is 1 if a firm is located within region k and 0 otherwise. Deriving the log-likelihood of this expression, a maximum likelihood (ML) estimation is possible using a sample of observed location decisions and their respective circumstances.

There are two main assumptions inherent to this model. The first is intuitive in the context of probabilities: All estimated probabilities must be positive for all possible choices and characteristics (McFadden, 1974). The second assumption is known as Independence of Irrelevant Alternatives (IIA), stating that the ratio of the probabilities of two locations must be independent of any third location. Hoffman & Duncan (1988) point out that this is especially problematic if alternatives are similar and might exhibit unobserved common factors. Train (2003) explains that even a violation of the IIA assumption has only minor effects on the analysis of average preferences, which is supported by Dahlberg & Eklöf (2003). An alternative that does not depend on the IIA assumption is the nested logit model (NL), even though this approach requires a sufficiently large sample within the nests to obtain statistically robust results (Train, 2003).

Understanding profit as the difference between revenue and costs introduces accessibility to the model: Higher accessibility to relevant business partners is assumed to lower the costs of transactions and transportation, effectively increasing profits. Thus, *ceteris paribus*, a profit-oriented firm prefers a region with higher market access over a location with lower accessibility (Johansson & Forslund, 2008). There is, however, a drawback to the assumption of profit maximization in the given context: It does not account for the costs of establishing or relocating a business. Even though a firm might theoretically obtain the highest profits in a particular region, it might be costly or even infeasible for a small business (e.g., a craftsman's establishment) to relocate across the country (De Bok & Van Oort, 2011). Consequently, the assumption is expected to hold for relocating businesses and large firms, approaching their location decisions with a strategic perspective, but the degree to which it holds for established and small companies must be analyzed. To do so, separate models are estimated for large firms and small and medium sized enterprises, respectively, in addition to the baseline model analyzing the average effects of all firms.

A two-step decision-making progress is modelled to reproduce the location decision realistically. This is driven by the structure of the German economy with its high share of small firms: In many cases, these are flexible in their location choice only within a certain region or around a certain point (e.g., the owner's residency). While they commonly have a very strong preference for a particular part of the country, they decide on their location within the area under the profit maximization assumption. Consequently, firms are modelled to choose a greater region within Germany and the specific location within the region independently. Econometrically, this behavior corresponds to a nested MNL (McFadden, 1974): It separates the first-stage decision for a nest from the second-stage decision for a region within the nest. This allows to obtain valid results even if the choice of a greater region is determined by other factors than the selection of a specific location. For comparison, a non-nested version is also calculated.

The first part of the decision is modelled by choosing one of the 16 federal states, representing geographical areas, different political regimes, wage and taxation rates[2], and other aspects (BBSR, 2018; Krenz, 2016). In addition, topography and transportation accessibility differ between the states: While the northern states host the majority of German ports due to their adjacency to the sea, the southernmost regions have a significant share of mountains and several low mountain ranges are spread across the country. Although inland ports, air transport, and the highway network

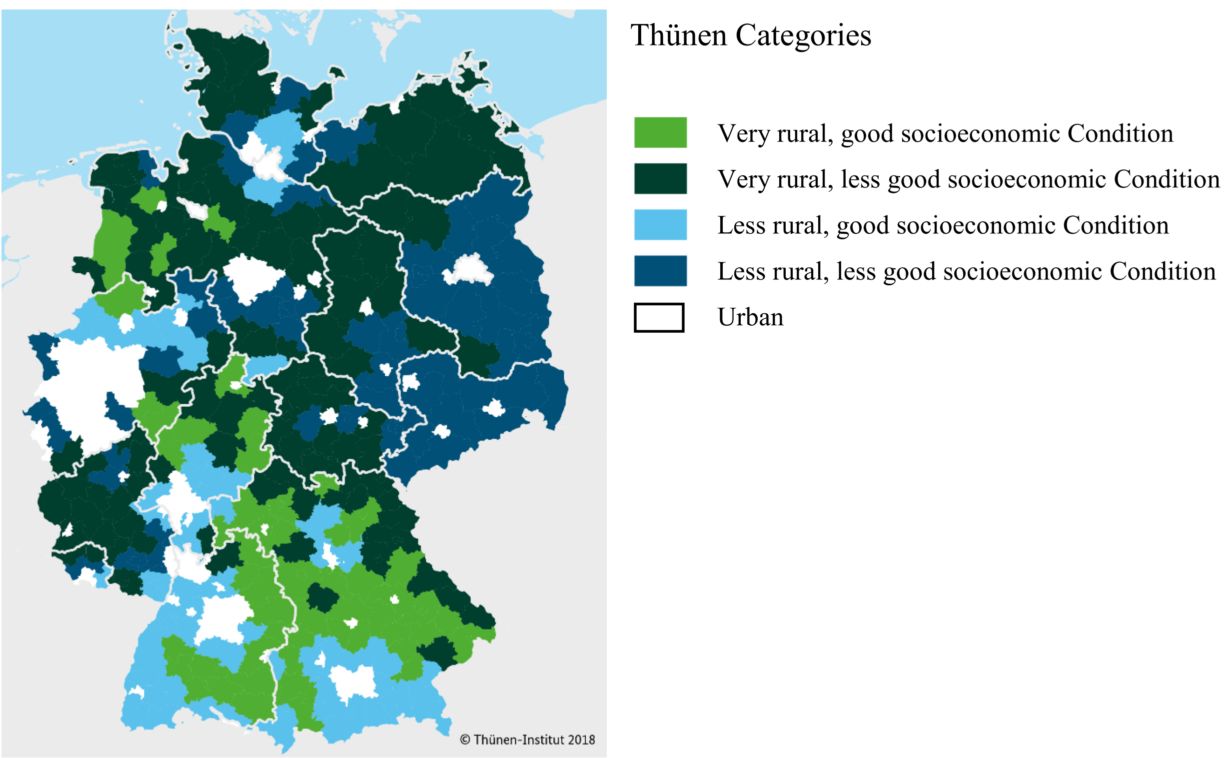

Figure 1 – Classification of German Counties based on States (White Borders) and Thünen Categories (Colors)
Sources: Küpper (2016); Thünen-Institut/BMEL (2016)

[2] E.g., wages are commonly lower in the former East German states (Kluge & Weber, 2018), and real estate transfer taxes are set by the states between a minimum of 3.5% and a maximum of 6.5% (Buettner & Krause, 2018).

are accessible from most locations, the range of accessibility reaches from states with minor ports and no international airports (e.g., Saxony-Anhalt) to states with several major ports and multiple international airports (e.g., North Rhine-Westphalia; DIW, 2019). Thus, it can be assumed that firms have a preference for a certain state independent of the specific location within the state.

After choosing the state, a firm chooses its location within the state based on aspects like population density, urbanization, and economic development. These structural characteristics are summarized in the Thünen classification of the 401 German counties, differentiating counties in the two dimensions of rurality and socioeconomic status (Küpper, 2016). With respect to the rurality, the Thünen Institute classifies counties as very rural, less rural, or urban. Besides population density, this aspect considers the share of agricultural and forestry area, the share of one- and two-family houses, the regional population potential, and the adjacency of larger cities. In terms of the socioeconomic characteristics, the classification uses a distinction between well and less well socioeconomically situated counties based on nine indicators (unemployment rate, average wage, median wage, municipal tax income, apartment vacancy rate, life expectancy of newborn boys and girls, respectively, balance of migration, school dropout rate). The final Thünen classification consists of five categories: urban (type 5), less rural and socioeconomically well situated (3), less rural and socioeconomically less well situated (4), very rural and socioeconomically well situated (2), and very rural and socioeconomically less well situated (1; Thünen-Institut/BMEL, 2016).

In the nested baseline specification, firms thus choose a state (defining the nest) in the first step and a Thünen category within the state in the second step. Combining the 16 states with the five Thünen categories provides a total of 50 possible choices, which are visualized in Figure 1. The choice option describing the counties of category 1 in Baden-Württemberg is used as the reference category in the MNL.

The two previously mentioned sets of explanatory variables are used to explain into which of the 50 choices (*l*) a company sorts itself. The first set consists of six firm-specific characteristics that are independent of the location. These are four dummy variables describing whether a firm has less than 500 employees (*SME*), is registered as a stock corporation (*SC*), was founded before 1990 (*OLD*)[3], and whether it has export revenues reported (*EXP*), a numerical variable describing the fixed assets ratio (fixed assets over total assets, *FIR*), and a categorical variable describing its

[3] This variable captures the division of Germany with its restrictions of location choice by adjusting the probability of firms founded before 1990 settling in a location.

industry group (*W*) based on the WZ2008 classification (Destatis, 2008) with industries being combined into a total of 13 industry groups. For the dummy and numerical variables, the method obtains one coefficient for each choice option except the first; for the categorical variable, a coefficient for each choice-industry-combination except the reference categories is estimated. Thus, the six variables of the first set provide a total of 833 estimated coefficients.

The second set of explanatory variables consists of four variables defining firms' accessibility at every possible location in different ways, allowing us to compare their impact on the location decision. The first three variables represent proxies that are easily observable for the firm without data collection or computation requirements. These include the number of firms in the same industry (*SI*) as well as the number of firms in the three most closely related industries[4] (*RI*) in the area, providing information on intra-industry and intra-supply-chain market access. Both numbers are calculated specifically for the foundation year, meaning they describe the number of respective firms that already existed in the area when the firm was established.[5] The impact of these data on the location decision of firms corresponds to the agglomeration effects described in the literature review: Positive coefficients are expected, suggesting that firms tend to locate themselves where firms in the same or related industries already are (De Bok & Van Oort, 2011). The third variable indicates the economic importance of a firm within a region, described as the ratio between the revenue of the firm and the total gross regional product (GRP) of the choice area (*EI*; Krenz, 2016). While this variable is measured in 2018 values due to data availability and computational constraints, the stability of firms over time supports the reliability of the variable (cf. Section 4). A positive value of this coefficient implies that firms prefer playing an important role, meaning they locate themselves in economically weak regions, whereas negative values signify a preference for economically strong regions even though firms have less of an economic role in these areas. As both options seem realistic, no expectation on this coefficient is formulated.

The final explanatory variable is the comprehensive market access indicator (*M*) explained in the next section. The arithmetic mean of all firms in the same industry[6] within the region is used to aggregate this firm-specific measure onto the regional level. It captures firms' potential market access in a detailed way, providing information on the theoretically optimal location from the perspective of market access. However, it is important to note that the measure is computed for

[4] The three most closely related industries are derived from the input-output matrix as described in Section 3.2, i.e., these are three out of 72 industries.
[5] Empirically, this is limited to the firms included in the dataset, i.e., the variables are calculated as firms in the dataset with an earlier founding year than the respective firm.
[6] The industry categorization of the WZ2008 with 72 industries as described in Section 3.2 is used.

2018 only, meaning that the accessibility is not assessed at the time of a firm's foundation. Computing the access measure for every company foundation is computationally infeasible, but the measure is expected to be stable over time due to the data structure (cf. Section 4).

Incorporating the described variables in the nested MNL framework gives a model with a total of 902 coefficients to be estimated (49 intercepts, 833 from the first set of explanatories, four from the second set of explanatories, 16 nest intercepts). As it is infeasible to present and discuss all these coefficients (and many of them are found to be insignificant), the focus in the result section lies on the coefficients of major interest, namely those for the access measures.

3.2 Access Indicator

The access indicator is newly developed based on the existing literature and aims at defining the term *market access* in an intuitive way: it combines the market, identifying potential customers and suppliers through the relations between industries, with the access, using transport distances between individual firms. To do so, it modifies the existing gravity model approach and combines it with a double-weighting scheme (Nijkamp & Ratajczak, 2021).

The starting point of the market access measure is, following the standard of gravity models, a size variable (Donaldson & Hornbeck, 2016). Based on the theory that large firms can serve larger parts of the market and offer lower prices due to economies of scale, it is assumed that a firm's relevance (or "economic mass") for other companies increases with its size (Graham & Gibbons, 2019). In this application, size is measured as operating revenue. In contrast to standard gravity models, however, the measure assumes that the attractivity of another company as a business partner depends only on the size of the other company, but not on the own size.

The second aspect of the accessibility measure describes how likely companies are to do business with each other based on the relationship between their industries. Even though it is evident that specific industries interact a lot, whereas other sectors have hardly any relation, this aspect is neglected in scientific analyses, ignoring an essential aspect of market potential. In the indicator, the industry interaction is reflected in the first weighting factor based on the German input-output matrix (Destatis, 2021a). This matrix describes the flows of money within and between 72 industries in Germany based on the WZ2008 classification (Destatis, 2008). The one-directional flows of the original matrix are transformed to capture demand and supply between industries, and a row-normalization provides a measure of the relative importance of each industry for each industry.[7] To use the industry relation on firm level, it is assumed that companies within

[7] To minimize computation efforts, relations below 0.0025 are manually set to zero.

the industries are homogenous and face similar intra- and inter-industry relations, which is reasonable considering the high level of detail of the industry classification. Thus, the first weighting factor describes the average relation between their industries for any pair of companies.

The third part of the access indicator corresponds to the distance part of the gravity model, assuming that close firms have a higher potential for business relations as transportation and transaction costs increase with distance (Ahuja & Tiwari, 2021; Graham & Gibbons, 2019). While the existing literature commonly uses geometric distance due to its relatively low computational requirements (Melo et al., 2017), this paper applies firm-to-firm road transport distances.[8] These are calculated using a local installation of the OpenRouteService (ORS; HeiGIT, 2008) and map data from Openstreetmap (OSM; Geofabrik, 2020) for Germany in 2018. The ORS allows the use of specific settings for heavy goods vehicles (HGVs), accounting for speed and weight restrictions as well as HGV preferences such as the avoidance of residential areas and preferred routing via highways. However, distances between companies are replaced by transport distances between the geometric center points of the companies' postal code areas for longer distances.[9] Due to the small size of most of the 8,169 German postal code areas, the difference between the firm-to-firm distance and the replacement is negligible, while computation time is significantly reduced.

Combining the size with the two weighting schemes of industry relations and transport distance in a gravity model approach, the business potential $p_{i,j}$ between any pair of companies can be defined as follows (cf. Nijkamp & Ratajczak, 2021):

$$p_{i,j} = \frac{s_j^{\alpha_1} * ir_{i,j}^{\alpha_2}}{d_{i,j}^{\alpha_3}} \tag{3}$$

In this expression, the potential for firm i to do business with firm j, $p_{i,j}$, is defined through the size of firm j, s_j, the industry relation between the industries of the firms, $ir_{i,j}$, and the transport distance between the firms, $d_{i,j}$. α_1 and α_2 represent the weighs of the size and business relation, respectively, and α_3 is a distance decay parameter reflecting that the importance of partners decreases with distance due to the increase in transportation and transaction costs. While α_1 and α_2 are fixed to 1 following the standard in the literature, the size of the distance decay parameter is difficult to determine and may significantly impact the results (cf. Rosenthal & Strange, 2020). While some applications (e.g., some productivity analysis settings) allow estimating the parameter

[8] Transport times are also calculated but lead to very similar results due to strong correlation (Pearson coefficient = 0.49). The results are not presented but are available from the author upon request.
[9] This is done if the distance between the postal code areas is further than 264km, representing the first quantile of the road transport distances between the geometric center points of all postal code areas.

(Bröcker, 2006; Östh et al., 2014), this is not possible in the given context, since the model uses the access measure in an aggregated form. In the given context, a high decay is required to ensure separability between locations: A lower decay parameter means that the access to more distant firms becomes more relevant, leading to smaller differences across locations. Therefore, the main model uses a distance decay of $\alpha_3 = 2$, but the robustness of the results with respect to the decay parameter is analyzed by also estimating a model with $\alpha_3 = 0.5$.

Firm i's total market access is calculated as the sum of the business potential between the firm and all other firms:

$$M_i = \sum_{j \neq i} p_{i,j} \tag{4}$$

The measure can also be described in matrix form, where *IR* denotes the matrix of industrial relations, *D* marks the distance matrix, and *S* connotes the vector of operating revenues:

$$M = \left[IR^{\alpha_2} \odot \left(\frac{1}{D^{\alpha_3}} \right) \right] * \ S^{\alpha_1} \tag{5}$$

4. Data

The analysis is based on several datasets explained in this section. The main dataset consists of firm-level information on the firm- and the location-specific variables. To construct this dataset, supplementary calculations are necessary using additional databases.

The main source of data is the ORBIS firm database (Bureau van Dijk, 2022), which collects information on enterprises in Europe from company reports, financial statements, publications, and other sources. For 2018, the database holds the required information for 110,083 observations, corresponding to 3.2% of all German enterprises. Address-specific location data are used to identify the county, augmented by the Thünen category obtained from the Landatlas (Thünen-Institut/BMEL, 2016). The business information provided by Bureau van Dijk (2022) includes asset structure, year of foundation, number of employees, legal form, export revenue, and the industry of activity. These data are used to calculate the variables explained in the previous section, for which summary statistics are presented in Table 1.

A negative minimal value and a maximum value of 1.36 are found in the values of the fixed assets ratio. These are three observations in the ORBIS database, while all other firms have values between 0 and 1. For the dummy variables, only the mean is provided in Table 1, which can be interpreted as the percentage of firms with the value 1 in the respective variable. It indicates a very low share of large companies and stock corporations, which is typical for the German economy. It should still be noted that the database overrepresents large companies and holds a relatively low number of very small companies, which are subject to fewer reporting policies: While firms with a revenue below one million Euros constitute 88% of all firms in Germany and only 0.42% of all firms report revenues of more than 50 million Euros (Destatis, 2021b), the respective shares in the sample are 16% and 7%. However, the dataset is quite representative of the German economy in terms of industries and regional distribution. Table 4 in the appendix provides further insights into these distributions, quantifying the variation between regions. Concerning the distribution across industries, two classifications must be considered: The very detailed information in the ORBIS database is aggregated, on the one hand, into the 72 categories used in the input-output matrix (Destatis, 2021a) and, on the other hand, into the 13 industry groups used in the MNL estimation.

Concerning the age structure of the firms, Table 1 shows that only 27% of the sample were established before 1990 (50% were established after 2000). The differences between East and West Germany in the time before 1990 are captured by the dummy variable OLD, with a low share of firms founded in times significantly deviating from 1990: Only 6% of the observed firms were established before 1950. As the spatial distribution of the German economy and the country's transport infrastructure have only seen minor changes between 1990 and 2018, while the dummy controls for the different situation before 1990, the accessibility values of 2018 are thus assumed to match the values at the time of foundation well for a large part of the dataset.

	Description	Mean	Median	Var	Min	Max
SME	Dummy: <500 Employees	0.98				
SC	Dummy: Stock corporation	0.02				
OLD	Dummy: Founded before 1990	0.27				
EXP	Dummy: Export Revenue	0.03				
FIR	Fixed Assets Ratio	0.28	0.19	0.07	-1.12*	1.36*
SI	Number of pre-existing Firms in same Industry	51.39	13.00	$> 10^4$	0	2,593.00
RI	Number of pre-existing Firms in most closely Industries	144.55	54.00	$> 10^4$	0	5,504.00
EI	Ratio: Revenue over GRP (%)	0.14	0.01	15.49	0.00	2,494.11
OR	Operating Revenue (M €)	30.04	2.52	$> 10^{11}$	-3.42*	110,237.22
IR	Industry Relation (%)	1.37	0.32	0.14	0.00	63.22
D	Transport Distance between Postal Code Areas (km)	421.93	408.92	$> 10^4$	0.08	1,202.32
$M1_i$	Access (Firm): $\alpha_3 = 2$ (M)	484.60	15.41	$> 10^{14}$	0.00*	6,663,731.90
$M2_i$	Access (Firm): $\alpha_3 = 0.5$ (M)	5,670.41	5,279.88	$> 10^6$	0.00*	40,139.43
$M1_l$	Access (Location): $\alpha_3 = 2$ (M)	259.44	32.77	$> 10^7$	0.00*	891,376.40
$M2_l$	Access (Location): $\alpha_3 = 0.5$ (M)	5,096.28	4,842.81	$> 10^6$	0.00*	16,401.12

Notes: *) Further information on these outliers are provided in the text.

Table 1: Summary Statistics of Variables

Further data are necessary to calculate the location-specific variables explained in the previous section. The German input-output matrix (Destatis, 2021a), describing the relations between industries, is combined with the years of foundation that are included in the ORBIS data. This allows obtaining the number of firms in the same as well as in the three most closely related industries by counting the relevant firms in each location option with a founding year earlier than that of the respective firm. While a meaningful visualization of these 100 variables (two variables for each of the 50 locations) is impossible, a summary is provided in Table 2. As can be seen, an average (median) of 150 (51) firms in the same and 406 (189) firms in closely related industries were present in the area a firm located itself into, while in other regions it was 49 (13) and 139 (53) companies, respectively. This is in line with the finding that 17% of all firms are located in the area with the highest number of firms in the same as well as closely related industries, and another 9% in the region with the second-highest number. These numbers underline the expectation that these two variables, defining agglomeration of industries and supply chains, play an important role in the location decision of firms.

To determine the relevance of the individual firm for the regional economy, the firm-level operating revenue of 2018 is divided by the gross regional product (GRP) obtained from the System of National Accounts (Arbeitskreis VGRdL, 2018). The resulting 50 variables show that a firm accounts, on average, for 0.03% of the GRP in the region of choice, whereas it would account for an average 0.14% in the other locations (Table 2). While only 0.13% of all firms are located in the region where they have the highest possible economic impact, 16.63% choose the area in which they have the lowest possible share of GRP. This preference for economically well-situated regions with a high GRP is in line with the important role of agglomeration effects outlined by the descriptive analysis of the previous variables and previous literature (Johansson & Forslund, 2008).

	Own Location: Mean (Median)	Other Locations: Mean (Median)	Share of Firms (in %) located in Area with minimal (second-minimal) Value	Share of Firms (in %) located in Area with maximal (second-maximal) Value
SI	150.47 (51)	49.37 (13)	3.26 (3.33)	17.00 (8.51)
RI	406.20 (189)	139.21 (53)	0.94 (1.01)	16.86 (6.76)
EI	0.03 (0.00)	0.14 (0.01)	16.63 (5.75)	0.13 (0.17)
$M1_l$	484.60 M (101.18 M)	254.85 M (32.41 M)	0.24 (0.43)	4.23 (4.26)
$M2_l$	5670.41 M (5341.82 M)	5084.57 M (4820.68 M)	0.61 (0.27)	12.96 (7.09)

Table 2: Relation between Firms and Access Measures across Location Choices

The final variable necessary for the analysis is the indicator of market access consisting of three parts: size, industry relations, and transport distance. The size of firms is obtained from the ORBIS dataset (Bureau van Dijk, 2022). Using the operating revenue as the measure of size ensures that only revenue generated from activities that are relevant for other firms is accounted for. The descriptive statistics of this variable, included in Table 1, point out three negative observations, which are classified as data errors, as well as a commonly found pattern: The large gap between the mean and the median and the large variance can be traced back to a few very large companies increasing the mean by a large amount.

The industry relations are derived from the German input-output matrix (Destatis, 2021a) as described in Section 3.2. The resulting matrix points out that industries trade an average of 24.56% of their revenues with the single-most important partner industry and a total of 43.12% with their three most closely related sectors. For 26 of the 72 industries, the most important partner is the own industry, highlighting the vital role of intra-industry trade. However, while seven industries earn more than 30% of their revenues from within-industry trade, there are also twelve sectors with an intra-industry trade ratio of less than 1%. This underlines the heterogeneity in industrial relations and the importance of accounting for these structural differences when measuring accessibility.

An ORS installation with OSM-data for 2018 (Geofabrik, 2020) is used to calculate the transport distance between firms. Even though the data are obtained from the public open-source OSM project, their quality for Germany is extremely high, including the relevant information for the HGV profile. The matrix on the postal code area level gives an impression of the relevant measures (Table 1): It displays an average transport distance between any two of the 8,169 areas of 421.93 km, with the closest neighbor being on average 6.78 km away and the furthest other area distanced at an average of 902.38 km. A comparison of several example routes with the results of alternative route services confirm that these values are reasonable for Germany's size and road network, underlining the robustness of the obtained road transport distances.

The market access indicator, obtained at the firm level, is a combination of the three measures of size, industry relations, and transport distance. The descriptive statistics of the measure with a distance decay factor of $\alpha_3 = 2$ (main specification) and with $\alpha_3 = 0.5$ (for robustness checks) are presented in Table 1. These point out the high variation between firms with a few outliers pulling the average upwards. To be noted are nine cases with a calculated market access of 0, which can be traced back to issues in the definition of the address or the industry. To include this variable in the MNL estimation, the average of each location choice area and industry is calculated. The resulting values are used to determine the hypothetical market access for every firm in every choice region, providing another 50 variables.[10] It is found that firms enjoy an average market access of 484.6 M (median: 101.2 M) in the region they choose, whereas the access would be 254.8 M (32.4 M) in other regions (Table 2). Even though firms choose an above-average market access, only 4.23% of all firms are located in the region with the highest achievable market access, which is considerably lower than the numbers obtained for the agglomeration variables. This hints at agglomeration and the existence of firms in certain industries playing more important roles in the location decision of firms than the actual market access. At the same time, it supports the hypothesis that firms could locate themselves more efficiently and achieve higher market access values by considering more complex measures in their location decisions. The results of the MNL estimation, presented in the next section, provide a more detailed view on this observation.

5. Estimation Results & Discussion

This section presents and discusses the estimation results of the location choice model. The model is estimated as a nested MNL using a ML estimator with a total of 902 coefficients. This paper discusses all results, but it is impossible to report all coefficients. Therefore, the quantitative results are presented as follows: Summary statistics of the intercepts, the four firm-specific dummy variables, and the fixed assets ratio (FIR) are reported in Table 3. They include the number of coefficients that are statistically significant at the 90% level as well as the minimum and maximum of the 49 coefficients per variable, representing the range of the impact of these variables. The same numbers for the 49 coefficients of each of the 13 industry groups are included in Table 5 in the Appendix. The four most relevant coefficients, namely the impact of the location-specific variables, are reported in Table 3 with standard errors and significance. The coefficients of the 16 nests, describing the federal states, are included in Table 6 in the Appendix. The full results of all models can be obtained from the author upon request.

5.1 Baseline Model

As can be seen in the first column of Table 3, the main model specification supports the formulated expectations concerning the accessibility variables: On the one hand, the coefficient describing the impact of the number of firms in the same industry (β_{SI}) is positive and statistically significant at the 99% confidence level, confirming the importance of inter-industry agglomeration for the location decision of firms. This is supported by the significantly positive coefficient of the number

[10] For simplification, only the details of the measure with $\alpha_3 = 2$ are discussed; details on the robustness check measure with $\alpha_3 = 0.5$ are available upon request.

of firms in related industries (β_{RI}; at the 90% level). These two coefficients clearly indicate that firms tend to go where similar firms already are, leading to the agglomeration effects commonly found in the existing literature (De Bok & Van Oort, 2011; Rosenthal & Strange, 2003). The impact of the economic importance of a firm within a region (β_{EI}) is significantly negative at the 99% level, meaning that companies prefer economically well-situated regions over the opportunity to play an important role themselves. On the other hand, the calculated accessibility indicator does

Symbol	Note	Variable	Baseline Model	Large Companies	Non-nested	Low Decay
β_0	a)	Intercept	*: 11 [-0.06, 4.75]	*: 0 [-20.09, 3.57]	*: 30 [-0.06, 4.75]	*: 27 [0.19, 4.73]
β_{SME}	a)	Small Enterprises (Dummy)	*: 0 [-0.64, 1.05]		*: 0 [-0.64, 1.05]	*: 0 [-0.70, 1.09]
β_{SC}	a)	Stock Corporation (Dummy)	*: 0 [3.55, 6.09]	*: 0 [-1.47, 22.19]	*: 0 [14.69, 17.23]	*: 0 [3.88, 6.43]
β_{OLD}	a)	Founded before 1990 (Dummy)	*: 13 [-4.30, 0.04]	*: 0 [-23.95, 0.05]	*: 27 [-4.30, 0.04]	*: 27 [-4.31, 0.03]
β_{EXP}	a)	Export Revenue reported (Dummy)	*: 0 [-1.29, 0.35]	*: 0 [-19.50, 1.41]	*: 1 [-1.29, 0.35]	*: 0 [-1.25, 0.35]
β_{FIR}	a)	Fixed Assets Ratio	*: 7 [-1.05, 1.41]	*: 0 [-1.54, 3.34]	*: 14 [-1.05, 1.41]	*: 12 [-1.11, 1.33]
α_{SI}	b)	Firms in the same Industry	0.000310*** (0.000043)	0.000550 (0.000570)	0.000310*** (0.000032)	0.000317*** (0.000037)
α_{RI}	b)	Firms in related Industries	0.000025* (0.000015)	-0.000006 (0.000103)	0.000025* (0.000014)	-0.000011 (0.000015)
α_{EI}	b)	Share in GRP	-5.78*** (1.68)	-5.05 (4.82)	-5.78*** (1.34)	-5.61*** (1.19)
α_M	b)	Average Market Access	-0.00 (0.00)	0.00 (0.00)	-0.00*** (0.00)	0.00*** (0.00)
$Sign_{90}$	c)	Significant Coefficients at 90%	63	0	157	143
$Sign_{95}$	c)	Significant Coefficients at 95%	35	0	118	104
$Coef.$	c)	Total Coefficients	902	657	886	902
N		Observations	109,813	2,636	109,813	109,813
R^2		R-squared (McFadden)	0.030	0.076	0.030	0.030

Notes: a) "*:" marks number of significant coefficients at 90%-Level; "[β_1, β_2]" marks minimum and maximum coefficients. - b) *, **, & *** relate to significance on the 90, 95, and 99%-Level, respectively; standard errors in brackets. - c) Total number of estimated (significant) coefficients.

Table 3: Estimation Results of the Location Choice Models

not impact firms' location decisions with an insignificant value of 0. Even though this variable describes the supply-chain accessibility in a detailed firm- and location-specific way, it does not play a role for firms when choosing their location. This is in line with the formulated expectations as well as the hypothesis that perceived accessibility plays a larger role than actual market access. This provides evidence that firms base their location decision not on complex measures of real market potential but instead on easily observable proxies for accessibility.

In addition to the effects of these four variables, the model shows a reasonably good performance. For all firm-specific explanatory variables, the effects across locations are diverse, with many of the coefficients being statistically insignificant: The number of employees, the legal form of a firm, and its reported export revenues do not influence location decisions, with all 49 of their respective coefficients falling out of the 90% confidence interval. While it is reasonable that the size and legal form of a firm do not necessarily influence its preferred location, the lack of impact of the export orientation is suspected to be caused by unreliable data. Only 3,395 of the 110,083 observed firms have export revenues reported, which is significantly lower than the assumed range (Kaus & Leppert (2017) find 9.1% of German firms to export goods). The age variable describing whether firms were founded before 1990 has a significant impact in 13 of the 49 locations. Accounting for the structural differences of location decisions during the division of Germany, these effects are distributed across nine states and all five Thünen categories, underlining the importance of this variable for the model. The fixed assets ratio has an impact in seven cases, where companies with a higher input of fixed assets are less likely to be in urban locations – suggesting that capital intensive firms avoid city locations; for example, due to higher land prices (cf. Pellenbarg et al., 2002) –, and more likely to be in socioeconomically vulnerable locations commonly characterized by low land prices and wages. As expected, the coefficients associated with the industry groups show large variation, with service and hospitality industries likely located in urban areas and producing firms commonly found in rural regions (cf. Baraklianos, 2018).

Even though the conclusions drawn from the estimation results are as expected and in line with economic theory, it must be noted that only 63 of the 902 coefficients are significant at the 90% level and only 35 at the 95% level. In addition, the goodness of fit is low with an R^2 of 0.03, despite MNL estimations commonly having extremely low values as they explain probabilities for a binary variable (Bartlett, 2014). The low level of significance is most likely caused by the very high number of estimated coefficients in combination with several areas with relatively few firms. In addition, the categorical variables are unevenly distributed with few observations in certain categories, and the data on export revenue appears to be unreliable in the database. Despite these drawbacks, the results concerning the main variables can be interpreted clearly and provide evidence for the importance of industry- and supply-chain-specific agglomeration effects. While the model is inadequate for predicting the choices of firms, it provides a reliable explanation of the location decision and the role of the relevant variables (cf. Train, 2003).

5.2 Company Size Models

As the main model describes the behavior of all firms in the sample, it can be assumed that the location decision plays a different role for companies of different sizes: For an entrepreneur, it is undoubtedly easier to open his small business at his residency than to move across the country for the theoretically optimal location (Arauzo Carod & Manjón Antolín, 2004). With increasing size of the firm, however, location decisions are of higher strategic importance and complex variables like accessibility might be considered. To understand whether such differences exist between small and large firms, separate models are estimated with the sample being split up based on the variable *SME*. With only 2,636 companies having more than 500 employees, however, this requires several industry groups (*W*) to be combined in the model of large enterprises, leaving a total of nine groups with a minimum of 149 firms. Furthermore, the high number of coefficients estimated from a small sample leads to high standard errors and low rates of significance (second column of Table 3).

The main observation is that the estimates of the four accessibility-related variables are close to those obtained in the main specification, even though no statistical significance is found. While the number of firms in the same industry has a slightly higher impact than in the baseline model, the importance of firms in closely related industries is slightly lower, hinting at large enterprises making even more use of agglomeration effects than the average of all firms. It is worth noting that Krenz (2016) finds an opposite effect, with large firms not being influenced by agglomeration, even though using different data and methodological approaches. The firm's economic role within the region has a slightly lower impact than in the baseline model, while the market access indicator remains with an insignificant coefficient of 0.

The coefficients of the firm-specific variables contribute to the goodness of fit, which is higher than in the baseline model with an R^2 of 0.08, but show higher variation than in the reference model. Thus, the model does not provide evidence that large enterprises approach their location decisions differently than the overall average of firms. This is also found in the estimation using firms with fewer than 500 employees: The results are almost identical to those of the main model (and, therefore, not reported), underlining that firms in all sizes use easily observable proxies for market access to make their location decisions but do not consider actual accessibility indicators.

5.3 Robustness Checks

Two further aspects are controlled for in additional robustness checks: the role of the nests and the impact of the distance decay parameter. For the first robustness check, the baseline model is estimated without grouping the location options into the 16 nests of the states (column 3 in Table 3). This model assumes that firms settle in one of the 50 options without a preference for, or limitation to, a greater region. Methodologically, this means that 16 coefficients less are estimated and that the standard errors are not corrected for the nest structure (Train, 2003). Consequently, the standard errors are smaller than in the baseline model, giving a higher significance level: 157 of the 886 coefficients are significant at the 90% level and 118 at the 95% level. Most of these coefficients belong to the variables found to have a significant impact in the nested specification: the intercepts, 14 coefficients of the fixed assets ratio, 27 coefficients of the age variable, and many of the industry group-specific coefficients. Although all four access variables significantly affect the location decision if the model does not account for the nests, the effect of the complex market access indicator remains extremely small (-1.70×10^{-12}). In addition, a reverse-nested model is estimated for comparison, supporting the findings of the other specifications.[11]

The results underline the importance of the correct model specification: The assumption that the average firm prefers to be located within a specific state – either due to characteristics of the state and the firm (e.g., seaport availability for exporting firms) or due to a limited choice set (e.g., around the residency of the owner) – and chooses merely its location within the region to maximize profits implies that the higher significance level of the non-nested specification is caused by inappropriately specified standard errors rather than by a better model fit. While this leads to the conclusion that the nested model is the more reliable and, therefore, preferred one, the differences between the two specifications are small, emphasizing the robustness of the results with respect to the standard error specification.

The final model analyzes the robustness of the results regarding the distance decay parameter used in the market access indicator (column 4 in Table 3; Melo et al., 2017). It uses a low decay of $\alpha_3 = 0.5$, meaning that further-away firms have a stronger impact on accessibility than in the baseline model. The results are found to be very similar to those obtained in the standard specification, even though the significance level is higher: 143 of the 902 coefficients are significant at the 90% level (104 at the 95% level); these are, as in the other specifications, mostly

[11] In this model, the nests are defined by the Thünen categories instead of by the states, i.e., it is assumed that firms choose the socioeconomic structure of their location before choosing their region. The results are very similar to those of the standard specification and available from the author upon request.

related to the fixed assets ratio, the age dummy, and the industry groups. In addition, strong agglomeration effects are found within the industry, whereas the closely related industries do not have a significant impact in this estimation.

Using the lower decay parameter, the access indicator has a significant effect on the location decision of firms, even though the coefficient is still very small. With the distance between firms playing less of a role in this model, however, the focus of this variable shifts toward the existence of firms in related industries. Consequently, the variable has a similar interpretation as the two access proxies describing the numbers of pre-existing firms, which explains the significance of the access indicator and the insignificant effect of the number of firms in closely related industries. To ensure the differentiation between locations and the impact of the transport distance on the access indicator, it is thus necessary to use a high decay parameter as done in the baseline specification.

6. Conclusions & Further Research

This paper analyzes the role different measures of accessibility play in the location decision of firms. A nested MNL is estimated based on a sample of 110,083 firms in Germany. Firms choose their location out of 50 options defined by federal states (in the first step) and Thünen categories (in the second step) based on two sets of explanatory variables: The first set includes firm-specific characteristics, such as size and industry, while the second set consists of four location-specific access measures. Three of these variables are proxies for market access that the firm can observe without much information collection or data processing; the fourth variable is a comprehensive measure of market potential with significant computational requirements. It combines the size of companies with industrial relations and the road transport distance between firms. The results provide evidence that firms do not account for such complex access measures but choose their location using easily observable proxies such as the number of firms in certain industries – most commonly, firms in their own industry. This outcome is in line with the expectations and the findings of existing research (De Bok & Van Oort, 2011; Johansson & Forslund, 2008; Melo et al., 2017).

On the one hand, agglomeration effects play a significant role in analyzing the distribution of economic activity. On the other hand, existing research finds strong effects of market access on the location decision when measuring accessibility with simple measures, whereas the evidence for a significant effect of more complex indicators is less pronounced. Several robustness checks confirm the reliability of these results, allowing us to draw two main conclusions.

First, the results clarify that the reliability of research findings depends on a realistic depiction of accessibility. If the goal is to model the location decision of firms and understand the influence of market access, it is essential to use variables that are available to the average firm. These include, for instance, the number of firms in specific industries or the economic importance of an individual company within a region, but certainly not the multidimensional indicators developed and used in scientific contexts. These complex measures have their own realm: capturing market access very accurately, they enable the analysis of the economic effects of accessibility. If the goal is to understand the impact of market access on firm (or regional) productivity, costs, prices, and similar variables, a comprehensive accounting of accessibility, as facilitated by the more comprehensive indicators, provides more reliable results. Thus, the findings of this paper show that the preferable market access indicator depends on the context: When it comes to decision-making, easily observable proxies are more realistic impact factors; when the focus is on the impact of accessibility, complex measures provide detailed and reliable insights.

Secondly, the results indicate that the observed location decision process leads to inefficient choices in terms of market access. Proxying accessibility with simpler measures, more than 95% of the observed firms are not located in the area offering the greatest actual market access, meaning they could benefit from better market access and lower transportation costs by relocating. Clearly, accessibility is one of many factors influencing location decisions, meaning that firms might consciously choose to forgo market access in exchange for other benefits. Nevertheless, the strong agglomeration effects found in the results indicate the important role of accessibility in firms' location decisions. From the perspective of firms, it might be advisable to put a stronger focus on more complex measures of market potential to make more efficient location decisions and benefit from the full advantages of the best possible access to suppliers and customers.

Even though the paper comes to meaningful and reliable results, some aspects remain open for further research. First, this applies to the dataset. Collecting further data and improving the reliability of the export variable could be important for obtaining even more robust findings. In addition, other variables, like foreign direct investment (Nielsen et al., 2017) or R&D intensiveness (Fritsch & Falck, 2007), may also play a role in firms' location decisions. Secondly, the realism of the market access indicator would improve by incorporating two aspects: imports and internet relations. The indicator in its current form does not account for cross-border trade or firms in other countries, thus underestimating the market potential for firms close to borders. Considering the open borders of the European Union and including firms in surrounding countries through additional data would depict the market access of many firms more realistically and contribute to more robust results. In addition, many firms and even entire industries depend stronger on an internet connection to their business partners than on road transport (Duso et al., 2021). Thus, including an internet access measure in the MNL and conducting an analysis focusing on the requirements of individual industries could provide further insights.

References

Ahuja, R., & Tiwari, G. (2021). Evolving term “accessibility” in spatial systems: Contextual evaluation of indicators. *Transport Policy*, *113*(January), 4–11. https://doi.org/10.1016/j.tranpol.2021.03.006

Arauzo Carod, J. M., & Manjón Antolín, M. C. (2004). Firm size and geographical aggregation: An empirical appraisal in industrial location. *Small Business Economics*, *22*(3–4), 299–312. https://doi.org/10.1023/b:sbej.0000022216.09083.76

Arbeitskreis VGRdL. (2018). *Database: Volkswirtschaftliche Gesamtrechnungen der Länder*. https://www.statistik-bw.de/VGRdL/

Balbontin, C., & Hensher, D. A. (2019). Firm-specific and location-specific drivers of business location and relocation decisions. *Transport Reviews*, *39*(5), 569–588. https://doi.org/10.1080/01441647.2018.1559254

Baraklianos, I. (2018). The impact of accessibility on the location choices of the business services. Evidence from Lyon urban area. *Region et Developpement*, *48*, 85–104.

Bartlett, J. (2014, February 8). *R squared in logistic regression*. The Stats Geek. https://thestatsgeek.com/2014/02/08/r-squared-in-logistic-regression/

BBSR. (2018). *Database: INKAR online*. https://www.inkar.de

Börjesson, M., Isacsson, G., Andersson, M., & Anderstig, C. (2019). Agglomeration, productivity and the role of transport system improvements. *Economics of Transportation*, *18*(June 2018), 27–39. https://doi.org/10.1016/j.ecotra.2018.12.002

Bourguignon, F., Fournier, M., & Gurgand, M. (2007). Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. *Journal of Economic Surveys*, *21*(1), 174–205. https://doi.org/10.1111/j.1467-6419.2007.00503.x

Bröcker, J. (2006). Parametric and Non-parametric Accessibility Measurement. *46th ERSA Congress*.

Buch, C. M., Kleinert, J., Lipponer, A., & Toubal, F. (2005). Determinants and effects of foreign direct investment: Evidence from German firm-level data. *Economic Policy*, *20*(41), 52–110. https://doi.org/10.1111/j.1468-0327.2005.00133.x

Buettner, T., & Krause, M. (2018). Föderalismus im Wunderland: Zur Steuerautonomie bei der Grunderwerbsteuer. *Perspektiven Der Wirtschaftspolitik*, *19*(1), 32–41. https://doi.org/10.1515/pwp-2017-0028

Bureau van Dijk. (2022). *Bureau van Dijk | Private company information – Orbis*. https://www.bvdinfo.com/en-gb/

Carlton, D. W. (1983). The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables. *The Review of Economics and Statistics*, *65*(3), 440–449.

Commendatore, P., Kubin, I., Bougheas, S., Kirman, A., Kopel, M., & Bischi, G. (2018). *The Economy as a Complex Spatial System*.

Dahlberg, M., & Eklöf, M. (2003). *Relaxing the IIA Assumption in Locational Choice Models: A Comparison Between Conditional Logit, Mixed Logit, and Multinomial Probit Models* (2003:9).

De Bok, M., & Sanders, F. (2005). Firm relocation and accessibility of locations: Empirical results from the Netherlands. *Transportation Research Record*, *1902*, 35–43. https://doi.org/10.3141/1902-05

De Bok, M., & Van Oort, F. (2011). Agglomeration economies, accessibility and the spatial choice behavior of relocating firms. *Journal of Transport and Land Use*, *4*(1), 5–24. https://doi.org/10.5198/jtlu.v4i1.144

Destatis. (2008). *Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)*. https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html

Destatis. (2021a). *Input-Output-Tabelle (Revision 2019) - Inländische Produktion (Herstellungspreise): Deutschland, Jahre*.

Destatis. (2021b). *Rechtliche Einheiten nach Umsatzgrößenklassen*. https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Unternehmen/Unternehmensregister/Tabellen/unternehmen-umsatzgroessenklassen-wz08.html

DIW. (2019). *Database: Verkehr in Zahlen 2019/2020* (S. Radke (ed.)). BMVI. https://doi.org/10.1007/978-3-642-59063-4_10

Donaldson, D., & Hornbeck, R. (2016). Railroads and American Economic Growth: A “Market Access” Approach. *Quarterly Journal of Economics*, *131*(2), 799–858. https://doi.org/10.1093/qje/qjw002

Duso, T., Nardotto, M., & Schiersch, A. (2021). Broadband and Productivity: Structural Estimates for Germany. In *DIW Discussion Papers* (No. 1988). https://doi.org/10.2139/ssrn.3998913

Eickelpasch, A., Hirte, G., & Stephan, A. (2016). Firms’ Evaluation of Location Quality: Evidence from East Germany. *Journal of Economics and Statistics*, *236*(2), 241–273. https://doi.org/10.1515/jbnst-2015-1014

Felipe, J., Hasan, R., & McCombie, J. S. L. (2008). Correcting for Biases when estimating Production Functions: An Illusion of the Laws of Algebra? *Cambridge Journal of Economics*, *32*(3), 441–459. https://doi.org/10.1093/cje/bem043

Fingleton, B. (2007). *New Directions in Economic Geography* (B. Fingleton (ed.)). Edward Elgar Publishing. https://doi.org/10.1111/j.1757-7802.2008.00008.x

Fingleton, B., & Szumilo, N. (2019). Simulating the impact of transport infrastructure investment on wages: A dynamic spatial panel model approach. *Regional Science and Urban Economics*, *75*(January), 148–164. https://doi.org/10.1016/j.regsciurbeco.2018.12.004

Fritsch, M., & Falck, O. (2007). New business formation by industry over space and time: A multidimensional analysis. *Regional Studies*, *41*(2), 157–172. https://doi.org/10.1080/00343400600928301

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The Spatial Economy: Cities, Regions and International Trade*. The MIT Press.

Geofabrik. (2020). *OpenStreetMap Data Extracts*. https://download.geofabrik.de/

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, *57*(1), 148–169. https://doi.org/10.1016/j.jue.2004.10.002

Graham, D. J., & Gibbons, S. (2019). Quantifying Wider Economic Impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation*, *19*(May), 100121. https://doi.org/10.1016/j.ecotra.2019.100121

Graham, D. J., & van Dender, K. (2011). Estimating the agglomeration benefits of transport investments: some tests for stability. *Transportation*, *38*(3), 409–426. https://doi.org/10.1007/S11116-010-9310-0

Hadas, Y., Gnecco, G., & Sanguineti, M. (2017). An approach to transportation network analysis via transferable utility games. *Transportation Research Part B: Methodological*, *105*, 120–143. https://doi.org/10.1016/j.trb.2017.08.029

HeiGIT. (2008). *Openrouteservice: Homepage*. https://openrouteservice.org/

Hoffman, S. D., & Duncan, G. J. (1988). Multinomial and Conditional Logit Discrete-Choice Models in Demography. *Demography*, *25*(3), 415–427.

Johansson, B. (1993). Infrastructure, Accessibility and Economic Growth. *International Journal of Transport Economics*, *XX*(2), 131–156.

Johansson, B., & Forslund, U. (2008). The analysis of location, colocation and urbanization economies. In C. Karlsson (Ed.), *Handbook of Research on Cluster Theory* (Issue 67, pp. 39–66). Edward Elgar Publishing. https://doi.org/10.4337/9781848442849.00009

Kaus, W., & Leppert, P. (2017). Aussenhandelsaktive Unternehmen in Deutschland: Neue Perspektiven durch Micro Data Linking. *WISTA*, *3*, 22–38.

Kim, J. U., & Aguilera, R. V. (2016). Foreign Location Choice: Review and Extensions. *International Journal of Management Reviews*, *18*(2), 133–159. https://doi.org/10.1111/ijmr.12064

Kluge, J., & Weber, M. (2018). Decomposing the German East–West wage gap. *Economics of Transition*, *26*(1), 91–125. https://doi.org/10.1111/ecot.12137

Krenz, A. (2016). *Firm structure and the location decision of German manufacturing firms: Evidence from official firm-level data* (No. 298; Cege Discussion Papers).

Kropko, J. (2008). *Choosing Between Multinomial Logit and Multinomial Probit Models for Analysis of Unordered Choice Data*. University of North Carolina at Chapel Hill.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, *99*(3), 483–499. https://doi.org/10.1086/261763

Küpper, P. (2016). *Abgrenzung und Typisierung ländlicher Räume* (No. 68; Thünen Working Paper).

Limao, N., & Venables, A. J. (1999). Infrastructure, geographical disadvantage, and transport. *World Bank Policy Research Working Paper*, *WPS 2257*(December), 1–37.

Martín-Barroso, D., Núñez-Serrano, J. A., & Velázquez, F. J. (2015). The effect of accessibility on productivity in spanish manufacturing firms. *Journal of Regional Science*, *55*(5), 708–735. https://doi.org/10.1111/jors.12200

Martin, R. (1999). The new "Geographical Turn" in Economics: Some Critical Reflections. *Cambridge Journal of Economics*, *23*(1), 65–91.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. Academic Press.

Melo, P. C., Graham, D. J., Levinson, D., & Aarabi, S. (2017). Agglomeration, accessibility and productivity: Evidence for large metropolitan areas in the US. *Urban Studies*, *54*(1), 179–195. https://doi.org/10.1177/0042098015624850

Nielsen, B. B., Asmussen, C. G., & Weatherall, C. D. (2017). The location choice of foreign direct investments: Empirical evidence and methodological challenges. *Journal of World Business*, *52*(1), 62–82. https://doi.org/10.1016/j.jwb.2016.10.006

Nijkamp, P., & Ratajczak, W. (2021). Gravitational Analysis in Regional Science and Spatial Economics: A Vector Gradient Approach to Trade. *International Regional Science Review*, *44*(3–4), 400–431. https://doi.org/10.1177/0160017620980519

O'Sullivan, A. (2005). Where Do Firms Locate? In A. O'Sullivan (Ed.), *Urban Economics*. McGraw-Hill.

Östh, J., Reggiani, A., & Galiazzo, G. (2014). Novel methods for the estimation of cost-distance decay in potential accessibility models. In A. Condeço-Melhorado, A. Reggiani, & J. Gutiérrez (Eds.), *Accessibility and Spatial Interaction* (pp. 15–37). Edward Elgar Publishing. https://doi.org/10.4337/9781782540731.00008

Pellenbarg, P. H., van Wissen, L. J. G., & van Dijk, J. (2002). Firm relocation: state of the art and research prospects. In *SOM-theme D: Regional Science*. http://www.rug.nl/staff/p.h.pellenbarg/artikelen/publicaties/35._firm_relocation-_state_of_the_art_and_research_prospects.pdf

Petersen, T. (2011). *Estimating the link between accessibility and productivity with propensity score matching*.

Pot, F. J., van Wee, B., & Tillema, T. (2021). Perceived accessibility: What it is and why it differs from calculated accessibility measures based on spatial data. *Journal of Transport Geography*, *94*(March), 103090. https://doi.org/10.1016/j.jtrangeo.2021.103090

Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, *85*(2), 377–393. https://doi.org/10.1162/003465303765299882

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2119–2171). Elsevier. https://doi.org/10.1016/S1574-0080(04)80006-3

Rosenthal, S. S., & Strange, W. C. (2020). How close is close? The spatial reach of agglomeration economies. *Journal of Economic Perspectives*, *34*(3), 27–49. https://doi.org/10.1257/jep.34.3.27

Schirmer, P. M., van Eggermond, M. A. B., & Axhausen, K. W. (2014). The role of location in residential location choice models: A review of literature. *Journal of Transport and Land Use*, *7*(2), 3–21. https://doi.org/10.5198/jtlu.v7i2.740

Thünen-Institut/BMEL. (2016). *Landatlas*. https://www.landatlas.de/laendlich/typologie.html

Train, K. E. (2003). *Discrete choice methods with simulation* (Vol. 9780521816). Cambridge University Press. https://doi.org/10.1017/CBO9780511753930

Venables, A. J. (2010). New Economic Geography. In *Palgrave Dictionary of Economics*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230280823_26

Zondag, B., & Pieters, M. (2005). Influence of accessibility on residential location choice. *Transportation Research Record*, *1902*, 63–70. https://doi.org/10.3141/1902-08

Appendix

Appendix A: Distribution of Firm Dataset across States and Industries

WZ2008 Classification	BW (DE1)	BY (DE2)	BE (DE3)	BB (DE4)	BR (DE5)	HH (DE6)	HE (DE7)	MV (DE8)	NI (DE9)
Mining and Quarrying (B)	33	92	2	8	1	6	25	5	37
Manufacturing (C)	4.204	3.844	370	465	97	259	1.604	217	1.461
Electricity, Gas, Steam, and Air Conditioning Supply (D)	273	252	27	95	18	37	131	81	183
Water Supply; Sewerage, Waste Management and Remediation Activities (E)	130	160	23	60	9	30	80	24	82
Construction (F)	2.401	3.201	643	610	84	257	1.389	345	1.379
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles (G)	3.386	4.553	641	485	166	624	2.249	207	1.708
Transportation, Storage (H)	629	921	117	147	98	255	451	86	432
Accommodation and Food Service Activities (I)	283	569	210	45	19	74	249	56	162
Information and Communication (J)	674	1.035	403	62	36	238	455	36	211
Financial and Insurance Activities (K)	197	308	107	35	14	145	169	8	111
Real Estate Activities (L)	422	713	282	176	39	192	373	85	278
Professional, Scientific, and Technical Activities (M)	1.355	2.034	676	201	67	381	1.077	105	519
Administrative and Support Service Activities (N)	789	1.149	349	191	56	238	613	97	388
Education (P)	62	93	62	20	4	28	58	5	29
Human Health and Social Work Activities (Q)	308	446	243	119	19	118	318	71	329
Arts, Entertainment, and Recreation (R)	139	191	65	19	11	35	103	12	80
Other Service Activities (S)	233	289	154	55	17	66	180	27	90
Others	62	85	6	181	2	6	31	88	83
Total	15.580	19.935	4.380	2.974	757	2.989	9.555	1.555	7.562
Share	14,15%	18,11%	3,98%	2,70%	0,69%	2,72%	8,68%	1,41%	6,87%

Notes: For a comparison with the entire German economy at state level for 2018, see Table 52111-04-01-4-B at https://www.regionalstatistik.de/

Table 4 - Part I: Distribution of Firm Dataset across States and Industries

Continuation of table on next page

WZ2008 Classification	NW (DEA)	RP (DEB)	SL (DEC)	SN (DED)	ST (DEE)	SH (DEF)	TH (DEG)	Total	Share
Mining and Quarrying (B)	52	23	4	18	9	12	5	332	0,30%
Manufacturing (C)	5.250	1.133	204	1.005	449	611	518	21.691	19,70%
Electricity, Gas, Steam, and Air Conditioning Supply (D)	349	68	25	54	72	110	59	1.834	1,67%
Water Supply; Sewerage, Waste Management and Remediation Activities (E)	304	64	22	70	46	39	36	1.179	1,07%
Construction (F)	3.617	981	198	765	418	649	345	17.282	15,70%
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles (G)	6.419	1.281	267	695	332	842	318	24.173	21,96%
Transportation, Storage (H)	1.230	284	53	171	119	181	75	5.249	4,77%
Accommodation and Food Service Activities (I)	538	120	23	84	44	75	27	2.578	2,34%
Information and Communication (J)	1.057	152	49	140	35	108	47	4.738	4,30%
Financial and Insurance Activities (K)	405	53	15	45	14	44	16	1.686	1,53%
Real Estate Activities (L)	1.084	172	30	224	111	118	100	4.399	4,00%
Professional, Scientific, and Technical Activities (M)	2.418	381	75	339	138	244	134	10.144	9,21%
Administrative and Support Service Activities (N)	1.674	272	58	218	128	191	111	6.522	5,92%
Education (P)	151	16	12	28	21	21	17	627	0,57%
Human Health and Social Work Activities (Q)	815	148	40	149	92	110	97	3.422	3,11%
Arts, Entertainment, and Recreation (R)	294	59	11	32	21	30	14	1.116	1,01%
Other Service Activities (S)	514	99	57	82	37	62	33	1.995	1,81%
Others	118	39	5	108	162	41	99	1.116	1,01%
Total	26.289	5.345	1.148	4.227	2.248	3.488	2.051	110.083	
Share	23,88%	4,86%	1,04%	3,84%	2,04%	3,17%	1,86%		

Notes: For a comparison with the entire German economy at state level for 2018, see Table 52111-04-01-4-B at https://www.regionalstatistik.de/

Table 4 - Part II: Distribution of Firm Dataset across States and Industries
Continuation from previous page

Appendix B: Estimation Results for Industry Groups

ID	Industry	WZ2008	Baseline Model	Large Companies	Non-nested	Low Decay
1	Agriculture, Natural Resources	A, B	#	(ID: 12)	#	#
2	Machine & Plant Manufacturing	C26-C30	*: 1 [-3,56, 0,52]	#	*: 9 [-3,56, 0,52]	*: 7 [-3,32, 0,42]
3	Construction	F	*: 1 [-2,78, 1,01]	(ID: 10)	*: 6 [-2,78, 1,01]	*: 2 [-2,65, 1,09]
4	Chemical	C19-C23	*: 1 [-3,19, 0,81]	*: 0 [-0,94, 47,18]	*: 5 [-3,19, 0,81]	*: 4 [-3,10, 0,60]
5	Retail	G45, G47	*: 1 [-2,94, 0,76]	*: 0 [-1,09, 42,83]	*: 8 [-2,94, 0,76]	*: 6 [-2,75, 0,87]
6	Hospitality	I	*: 2 [-2,24, 3,24]	(ID: 10)	*: 3 [-2,24, 3,24]	*: 3 [-2,07, 3,29]
7	Wholesale	G46	*: 1 [-3,30, 1,78]	*: 0 [-1,57, 42,29]	*: 8 [-3,30, 1,78]	*: 7 [-3,17, 1,56]
8	Metal	C24, C25	*: 1 [-3,09, 0,53]	(ID: 4)	*: 9 [-3,09, 0,53]	*: 5 [-2,89, 0,69]
9	Public Services	E, O-R, T	*: 1 [-2,58, 2,31]	*: 0 [-1,42, 40,28]	*: 4 [-2,58, 2,31]	*: 3 [-2,62, 2,17]
10	Decentralized Private Services	J59, J60, L, M71, M74, M75, N77, N79-N82, S95, S96	*: 2 [-2,57, 2,37]	*: 0 [-1,19, 41,44]	*: 6 [-2,57, 2,37]	*: 6 [-2,57, 2,29]
11	Centralized Private Services	D, J58, J61-J63, K, M69, M70, M72, M73, N78, S94	*: 4 [-3,07, 2,30]	*: 0 [-22,35, 19,76]	*: 11 [-3,07, 2,31]	*: 8 [-2,99, 2,25]
12	Other Production	C10-C18, C31-C33	*: 0 [-2,41, 1,63]	*: 0 [-1,07, 39,50]	*: 3 [-2,41, 1,63]	*: 2 [-2,46, 1,22]
13	Transport	H	*: 2 [-3,67, 2,45]	*: 0 [-1,99, 42,95]	*: 9 [-3,67, 2,45]	*: 5 [-3,38, 2,22]

Notes: "*:" marks number of significant coefficients at 90%-Level; "[β_1, β_2]" marks minimum and maximum coefficients; # marks the reference group.

Four groups are included in other groups in the model of large companies, marked with the respective ID of the group with which they are joined.

Table 5: Estimation Results of the Location Choice Models for Industry Groups

Appendix C: Estimation Results for Nests (States)

Nest	Baseline Model	Large Companies	Non-nested	Low Decay
DE1 (BW)	1.0003***	1.0000		1.0000***
	(0.2680)	(0.8349)		(0.1178)
DE2 (BY)	1.0002***	1.0000		1.0000***
	(0.1957)	(0.6108)		(0.1110)
DE3 (BE)	1.0006**	1.0000		1.0000***
	(0.4177)	(1.1517)		(0.1529)
DE4 (BB)	0.9996	1.0000		1.0000***
	(0.8249)	(3.9581)		(0.1928)
DE5 (BR)	0.9999	1.0000		1.0000***
	(0.7826)	(2.4305)		(0.2855)
DE6 (HH)	1.0001**	1.0000		1.0000***
	(0.3960)	(1.1022)		(0.1360)
DE7 (HE)	1.0001***	1.0000		1.0000***
	(0.2217)	(0.6090)		(0.1183)
DE8 (MV)	1.0002	1.0000		1.0001***
	(0.7648)	(4.2935)		(0.2417)
DE9 (NI)	1.0001**	1.0000		1.0000***
	(0.4319)	(0.9236)		(0.1373)
DEA (NW)	1.0001***	1.0000		1.0000***
	(0.1470)	(0.8853)		(0.1015)
DEB (RP)	1.0001***	1.0000		1.0000***
	(0.3271)	(0.8401)		(0.1348)
DEC (SL)	1.0000**	1.0000		1.0000***
	(0.4931)	(5.6019)		(0.2099)
DED (SN)	1.0001***	1.0000		1.0000***
	(0.3097)	(1.6536)		(0.1679)
DEE (ST)	1.0000**	1.0000		1.0000***
	(0.4610)	(1.5297)		(0.1924)
DEF (SH)	1.0001*	1.0000		1.0000***
	(0.5224)	(1.3340)		(0.1693)
DEG (TH)	1.0003	1.0000		1.0000***
	(0.7293)	(2.4087)		(0.2029)

Notes: Standard Errors in Brackets

*, **, & *** relate to significance on the 90, 95, and 99%-Level, respectively

Table 6: Estimation Results of the Location Choice Models for the Nests (States)