Agglomeration or Market Access? The defining Factors for Firms' Location Choice.

Dennis Gaus, DIW Berlin e.V.

61st ERSA Congress 2022, Pécs, Hungary

1. Introduction

The role of accessibility for the development of firms, cities, and regions has seen great research attention ever since the seminal work of Hansen (1959). Firms´ access to customers and suppliers, coined as *market access*, and its role for the location of economic activity has been studied using a wide variety of access indicators – from measures based on the existence of certain infrastructure characteristics (Limao & Venables, 1999) to complex measures using various modes of transport and their respective transport times and costs (Donaldson & Hornbeck, 2016; Graham & Gibbons, 2019). An overview over some of the previously used measures is provided in Ahuja & Tiwari (2021). One indicator that is very commonly found to have a significant impact on the location decision of firms is the existence of firms of the same industry, which is known as agglomeration (cf. Rosenthal & Strange, 2004). Furthermore, literature using relatively simple proxies for accessibility, such as the existence of certain types of infrastructure, commonly finds a significant impact on the probability of firms choosing a certain location, whereas research based on more complex measures identifies less of an effect (cf. Graham & Gibbons, 2019; e.g., De Bok & Sanders, 2005). This pattern leads to the hypothesis that firms proxy accessibility using easily observable indicators when facing location decisions, but do not consider actual measures of market access.

This paper adds to the existing literature in two ways: First, it combines several existing concepts into a novel indicator of firms' market access based on road transport distance between firms, the relation between industries, and the size of enterprises. Second, it uses an extensive German firm-level dataset to estimate a location choice model incorporating the developed market access measure as well as several simpler indicators in order to determine the role of each individual measure for the location decision of firms.

2. Methodology

The access indicator aims at defining the term *market access* in a very intuitive way: it combines the market, identifying potential customers and suppliers by describing how likely

companies are to do business with each other based on the relation between their industries, with the access, using transport distances between individual firms. To identify industry relations, the German input-output-matrix (Statistisches Bundesamt, 2020) is adapted to describe the relative importance of each industry for each industry. The distance part is based on road transport distances obtained using the OpenRouteService (HeiGIT, 2008) with the "HGV" profile accounting for speed and weight restrictions for heavy goods vehicles, thus representing truck transport between companies. In addition, the importance of individual business partners is measured through revenue. These three aspects define the partnership potential between any pair of companies, following a gravity model approach:

$$p_{i,j} = \frac{ir_{i,j}^{\alpha_1} * m_j^{\alpha_2}}{d_{i,j}^{\alpha_3}} \quad (1)$$

The potential of firm *i* to do business with firm *j*, $p_{i,j}$, is defined through the industry relation between the industries of the firms, $ir_{i,j}$, the size of firm *j*, m_j, and the transport distance between the firms, $d_{i,j}$. The weighting parameters α_1 and α_2 get fixed to 1 following the literature, while the distance decay parameter α_3, reflecting that the importance of partners decreases with distance, gets fixed to 2. In this application, a relatively high decay parameter is necessary to ensure separability between locations. To identify the total market potential of a firm, the partnership potential between the firm and all other firms is added up:

$$M_i = \sum_{j \neq i} p_{i,j} \quad (2)$$

To identify the role of such a complex market access measure in the location decision of firms, a two-step multinomial logit location choice model is estimated. In the first step, a firm chooses a federal state, and in the second step, it determines the structural characteristics of its location. The federal states represent regions in this context, but also different political regimes and taxation rates. The structural characteristics are derived from the Thuenen classification, describing counties as very rural, less rural, or urban, and differentiating rural areas further into those with favourable socio-economic status and those in a less favourable situation (Küpper, 2016). Combining these two measures leads to a total of 50 possible choices clustered into 16 nests.

To explain into which of the 50 choices (l) a company sorts itself, two sets of variables are used. The first set includes firm-specific characteristics and consists of four dummy variables describing whether a firm has more than 500 employees (*KMU*), is registered as a stock corporation (Aktiengesellschaft, *AG*), has been founded before 1990 (*old*), and whether it has export revenues reported (*ex*), as well as a numerical variable for the fixed asset ratio (fixed assets over total assets, *FIR*) and a categorical variable for its industry *W*. The second set comprises the location-specific market access variables. On the one hand, these are three easily observable proxies of accessibility, namely the number of firms of the same industry that have existed in the choice area at the founding time of the company *C*, the number of firms of the three most closely related industries that have existed in the choice area at the founding time of the company *R*, and the ratio between the revenue of the firm and the total gross regional product (GRP) of the choice area *P*. While the first two variables account for industry- and supply-chain-specific agglomeration effects, respectively, the third variable identifies the relative regional importance of the individual firm. On the other hand, the mean market access of companies of the same industry in the choice area *AM* is included, based on the market access measure explained above. This variable describes the market potential of a location for the individual firm in a very detailed way, but requires significant data processing and calculation efforts. Following the formulated hypothesis, it is expected that the easily observable variables have a significant impact on firms' location choice, whereas the complex measure is not considered by firms. Including all these variables, the following relation is estimated as a nested multinomial logit model:

$$\pi_{il} = P(L_i = l) = f(C_{il}, R_{il}, AM_{il}, P_{il}, KMU_i, AG_i, old_i, ex_i, FIR_i, W_i) \quad (3)$$

3. Data

The firm-level data used for the analysis stems from the ORBIS firm database. For the year 2018, the database holds information on roughly 1.5 million companies in Germany (about 43% of all German firms). As the data are incomplete for many firms, however, we use 110,083 observations with the required information for the analysis. While the data slightly overrepresent large companies – which is not surprising as small companies are subject to less reporting policies –, the distribution in terms of industries represents the structure of the German economy well.

	Description	Mean	Median	Var	Min	Max
M1	Access: Distance, $\alpha_3 = 0.5$ (M)	5,670.41	5,279.88	$> 10^{18}$	0.00	40,139.43
M2	Access: Distance, $\alpha_3 = 2$ (M)	484.60	15.41	$> 10^{14}$	0.00	6,663,731.90
FIR	Fixed Asset Ratio	0.28	0.19	0.07	-1.12	1.36
old	Dummy: Founded before 1990	0.27				
KMU	Dummy: <500 Employees	0.98				
AG	Dummy: Stock corporation	0.02				
EXP	Dummy: Export Revenue	0.03				

Table 1: Summary Statistics of Variables

Table 1 shows the descriptive statistics of the derived accessibility measures explained in the previous section and the firm-specific data. As they are dummy variables except for the fixed asset ratio, only the mean is provided, which can be interpreted as the percentage of firms with the value 1 in the respective variable. The firm distribution among the categorical variables and the location-specific variables can be obtained from the author, as these 200 variables cannot be displayed in a meaningful way.

4. Results & Discussion

Table 2 displays the coefficients of the location-specific variables. As can be seen, there are two variables with a strongly significant influence on the location choice: The previous existence of firms of the same industry, and the firm's share of the GRP. The importance of the existence of similar firms, and the corresponding positive coefficient, is a clear indicator of agglomeration effects, as also found in previous literature. The negatively significant coefficient found for the regional economic importance of a firm can be interpreted as a tendency of firms to locate themselves in economically strong regions, as a firm has a lower

Symbol	Variable	Baseline Model	Non-nested Model	Low Decay
μ_C	Firms of the same	0.000310***	0.000310***	-0.000317***
	Industry	(0.000043)	(0.000032)	(0.000037)
μ_R	Firms of related	0.000025*	0.000025*	-0.000011
	Industries	(0.000015)	(0.000014)	(0.000015)
μ_P	Share in GRP	-5.78***	-5.78***	-5.61***
		(1.68)	(1.34)	(1.19)
μ_{AM}	Average Market	-0.00	-0.00***	0.00***
	Access	(0.00)	(0.00)	(0.00)
R^2	R-squared	0.030	0.030	0.076
	(McFadden)			

Notes: Standard Errors in Brackets

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

Table 2: Estimation Results of the Location Choice Model for the Location-specific Variables

GRP share in an economically successful region than in an economically weak area. In addition, the existence of firms from the most important related industries plays a minor role: The coefficient of this variable is positive, but significant only at the 90% level. The mean accessibility of firms of the same industry in the region, however, does not have a significant influence on the location decision of firms. The baseline model thus clearly supports the hypothesis that firms consider proxies for market access, such as numbers of firms or existence of transport infrastructure, when deciding on their location, but do not take more complex measures for accessibility into account. For the other variables, one coefficient for each choice is obtained, giving a total of 646 coefficients that can be obtained from the author upon request. Among these coefficients, 63 are significant at least on the 90% level and contribute accordingly to the explanatory power of the model.

Assuming that the strategic importance of location decisions depends on the size of a firm, I also estimate the model using only the available large companies (minimum 500 employees), but find that the sample size (2,636 companies) is insufficient to estimate this complex model in a statistically reliable way. However, the same model estimated only with small- and medium-sized enterprises (i.e., firms with less than 500 employees) leads to almost identical results as the estimation with the full sample.

To analyse the effect of the nests on the standard errors, a non-nested model (second column of Table 2) and a reverse-nested model (with Thuenen classes as nests, not displayed) are estimated. As the standard errors in the non-nested specification are smaller, the mean accessibility variable has a significant, but still extremely small impact. Another robustness check is conducted with respect to the distance decay parameter: If the access measure is calculated using $\alpha_3 = 0.5$, a very low, but significant positive impact of the market access is found (third column of Table 2). In this model, however, the differentiation between locations is vastly lower than in the baseline model, making the interpretation of this coefficient less reliable. This points out the importance of the model specification: Assuming that the location decision is indeed based on the region and the structure within the region independently, the baseline model with $\alpha_3 = 2$ is the most reliable and therefore preferred one.

Even though the details depend on the model specification, the findings are overall robust and reliable, providing strong support for the hypothesis that firms make their location decisions based on easily observable variables and do not consider more complex

accessibility indicators. In addition, a large portion of the firms' choices can be explained by company characteristics such as their industry.

5. Conclusion

This paper examines the accessibility-related factors determining the location choice of firms. A novel indicator of the market access of firms is developed, combining the size of companies with industrial relations and address-specific transport distances. Consequently, its impact is compared with the effect of several easily observable proxies for accessibility in a multinomial logit location choice model. The results point out that firms do indeed consider the proxies such as the number of firms of certain industries – most commonly, the own industry –, but do not account for more complex access measures. When studying the location decisions of firms, it might therefore be more suitable to use simple measures than to develop and calculate complex indicators. To identify optimal locations, however, the use of elaborate accessibility measurements might be adequate, as companies tend to locate themselves suboptimally from an accessibility perspective using simple proxies: While 17% of the companies in the sample are located in the region with the highest number of firms of the own industry and another 17% positioned themselves in the area with the highest number of firms from closely related industries, 96% of all observed firms are not located in the area that would provide them with the highest market access. This explains the agglomeration effects that are commonly found in the literature, and it leads to a growing gap between prospering regions and less favourably situated areas when it comes to attracting further firms.

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

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

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

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

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76. https://doi.org/10.1080/01944365908978307

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

Jiang, Y., Liao, F., & Jin, L. (2021). Effects of locational accessibility on firm diffusion characteristics: The case of Sino-Europe Economic Corridor. *Transport Policy*, *105*(August 2020), 80–93. https://doi.org/10.1016/j.tranpol.2021.02.013

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

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.

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

Statistisches Bundesamt. (2020). *Volkswirtschaftliche Gesamtrechnung (VGR) Deutschland*. https://www.destatis.de/DE/Themen/Wirtschaft/Volkswirtschaftliche-Gesamtrechnungen-Inlandsprodukt/_inhalt.html#sprg227266