DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

How to Draw the Line: A Note on Local Market Definition

by

Dieter PENNERSTORFER
Biliana YONTCHEVA

Working Paper No. 1917
October 2019

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

dieter.pennerstorfer@jku.at

How to Draw the Line:

A Note on Local Market Definition*

Dieter Pennerstorfer[†] Biliana Yontcheva[‡]

October 18, 2019

Abstract

This article presents a novel method of market delineation, which generates virtually isolated residential clusters using data on the spatial distribution of population. The performance of this approach is evaluated by contrasting it with traditional delineation techniques based on municipal boundaries. The estimation of simple entry models for five industries shows that markets defined using micro-level residence information perform better in terms of reducing cross-border spatial spillovers and predicting the equilibrium number of firms on the market more accurately. Additionally, the estimated entry threshold ratios using this method successfully reflect our expectations based on ex-ante knowledge about the investigated industries.

Keywords: market definition, entry models, spatial competition

JEL: Classification: L13, L11, R32

*We thank Céline Bonnet, Alexander McGlothlin, Frank Verboven, Christoph Weiss, the participants of the 2017 Winter Seminar of the Gesellschaft für Regionalforschnung, the ERSA Annual Congress 2017, the EARIE Conference 2018, the JEI Conference 2018, and the IIOC Conference 2019 for their helpful comments.

[†]University of Linz

[‡]Vienna University of Economics and Business

1 Introduction

The precise definition of the geographic market is usually an important first step in empirical competition analysis. Ideally, spatial markets are defined based on estimated demand and supply substitutability. As this approach is often infeasible due to demanding data requirements, researchers regularly use administratively defined regional units (e.g. municipalities) as a proxy for the relevant geographic market. Mapping firm locations as well as the spatial distribution of consumers—data that are often easily available at a disaggregated level—shows that the spatial distribution of economic activity is not necessarily congruent with administrative boundaries. This observation suggests that administratively defined regional units may be poor proxies for geographic markets that are expected to be independent of each other. In this article we utilize detailed, but easily available data on the distribution of consumers and offer a simple alternative approach to delineate local markets, characterized by much lower interdependence of neighboring units than geographic markets based on municipal boundaries.

The empirical literature investigating firm and market conduct suggests that approaches usually applied to define local markets can be classified in two groups:

In the first group, researchers rely on detailed price or revenue observations to model interaction between firms and evaluate whether they belong to the same market or are independent of each other (see Davis, 2006, and Ellickson et al., 2019, for applications of this approach for retail markets). Whenever price and revenue data are available, this method is effective in determining the scope of the relevant market from the perspectives of individual firms.

The second group comprises of applications in which location-specific price and sales data are not available, or the strategic firm behavior is discrete in character—such as market entry and exit—and does not permit a trivial calculation of elasticities. Researchers thus often focus on small and geographically isolated markets. This approach is particularly relevant when goods and services are provided at distinct locations that are difficult to substitute across space. Empirical investigations of this kind of services include competition analyses in the airline industry (Berry, 1992) or among motels located near highway intersections (Mazzeo, 2002). When the characteristics of the product do not rule out substitution, practitioners may try to satisfy the isolation assumption by focusing on rural markets with no nearby cities (Bresnahan and Reiss, 1991; Schaumans and Verboven, 2015). While this approach is warranted whenever the research is specifically concerned with market conduct in rural areas, it does not necessarily allow for the extrapolation of the results to urban markets (Aguirregabiria and Suzuki, 2016). Additionally, in densely populated areas the exclusion of all observations close to larger towns may reduce the sample size substantially. Seim (2006) resolves this problem by estimating the strength of interaction between distant spatial units within the same market, which allows for endogenous market definition. However, the model requires the implementation of a nested fixed-point algorithm over all locations within a given market in order to calculate entry probabilities. Since each additional location adds an equation to the algorithm, practitioners have an interest in limiting the total number of possible locations by applying an appropriate market definition strategy.

In this article we propose an alternative approach in dealing with the issue of market isolation. We aim to increase the level of independence of geographic markets in order to circumvent the need to model interactions across markets explicitly. We define local markets based on the actual spatial distribution of the residential population rather than relying on administratively defined spatial units.

We can draw on highly disaggregated spatial data for the entire population of Austria, based on grid cells of 250 $\times$ 250 meters (about 270 $\times$ 270 yards), and provide a simple and intuitive way to aggregate these small spatial units to local markets.[1] Note that population data at the grid level have become available in many countries in recent years (besides Austria similar data are available e.g. for France, Sweden or Japan) and will become even more widespread in the near future. Furthermore, the way to define markets

[1] This approach is conceptually similar to adding up counties to metropolitan statistical areas (MSAs) based on social and economic ties or to labor market areas (LMAs) based on commuting patterns. A parallel can also be made to aggregating zip code areas based on information on the distribution networks within firms. Empirical analyses of competitive conduct based on these approaches to delineate geographic markets include George and Waldfogel (2006), Cohen and Mazzeo (2007) and Ellickson (2006), respectively. These approaches, however, may result in geographical areas which are too large for many (in particular retail) industries, and are often coupled with higher data requirements.

proposed in this article is not restricted to grid data, and is also applicable whenever researchers wish to aggregate small geographic units of irregular shape, like blocks, block groups or census tracts, to local markets.[2]

To evaluate the level of market isolation generated by our approach, we estimate entry threshold models for five retail industries and compare the results to entry models using municipalities as local markets. We find, first, that our approach significantly reduces the number of residents and the number of firms located close to the border of neighboring markets, which decreases the likelihood of interactions across markets. Second, the regression results suggest that the estimated spillover effects between local markets are indeed much smaller and that the number of correctly predicted markets is higher. Third, industry knowledge allows us to make predictions regarding the competitive effect of firm entry for two of the retail markets in our sample. The estimated entry threshold ratios based on our approach to define local markets closely correspond to this prior knowledge. This is not the case for threshold ratios based on municipality data.

The remainder of the article is organized as follows: The market definition based on the spatial distribution of population is outlined in Section 2. The empirical results from the entry threshold models are presented and discussed in Section 3. Section 4 concludes.

2 Market Definition and Data

2.1 Market Definition

The way to define local markets proposed in this article is based on the notion that local markets can be interpreted as spatial clusters of the residential population. We can draw on detailed information of the spatial distribution of the population, collected in 2015 by the Austrian Statistical Office (Statistics Austria). Statistics Austria places regional statistical grid units over the entire territory of Austria. The grids are independent of administrative boundaries and the size of one grid cell is 250 $\times$ 250 meters. Each person is assigned to exactly one cell based on their postal address. This provides very detailed

[2] An R program for aggregation over administrative units is available from the authors upon request.

information about the spatial distribution of the population, as one square kilometer (square mile) is represented by 16 (41) cells.

The aim of the procedure outlined below is to define local markets that overlap less than conventionally used administrative units (e.g. municipalities). To do so we, first, discard all grid cells with a population size smaller than the average number of inhabitants (i.e. < 7 residents). Second, using the remaining cells with above-average population, two grid cells are considered to be connected if they share a common border (i.e. a common edge). All cells are in the same "spatial cluster" if they are either directly connected to each other, or if they are connected via pairs of connected cells. In this step all cells with more than 7 inhabitants are assigned to one specific cluster, precluding pairs of different clusters to share a common border.

Third, we restrict the number of local markets to be the same as the number of municipalities, to be able to compare the results of entry threshold models based on different ways to delineate local markets. We thus take the 2,089 largest clusters (i.e. the clusters representing the largest population) and assign the remaining small clusters to the closest large cluster.[3] Finally, all remaining grid cells discarded in the first step (i.e. all cells with below average population) are assigned to the closest cluster. Each grid cell is therefore assigned to exactly one compact spatial cluster, which we will call "cluster market" henceforth.

This procedure is illustrated in Figures 1 to 4 for a small region in our sample. The visualization suggests that the delineation of municipalities ("municipality markets") is not necessarily congruent to the spatial distribution of the population. This suspicion is supported by descriptive statistics reported in Table 1, showing that the population density close to the border of a different municipality is quite high, whereas the border regions of cluster markets are more thinly populated.

[3]The distance between a small and a large cluster is defined as the shortest Euclidean distance between (the centroids of) all pairs of cells, with one cell in each of the two clusters. Each small cluster is assigned to the closest large cluster. In case of ties small clusters are randomly assigned to one of the tied large clusters.

Figure 1: Distribution of the population

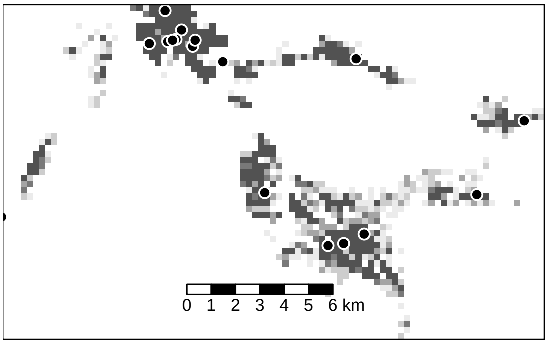


Figure 2: Large and small clusters

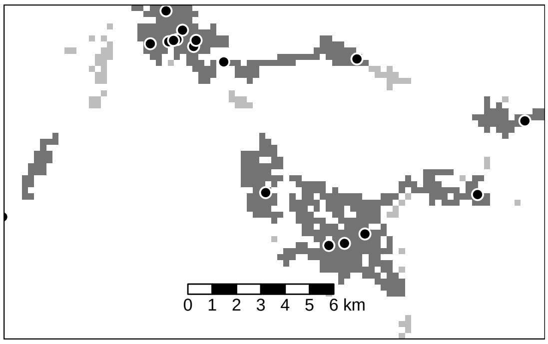


Figure 3: Cluster markets

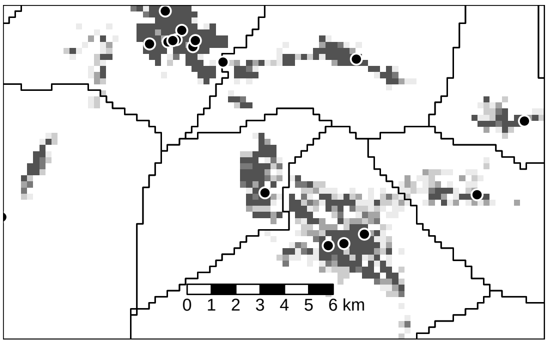
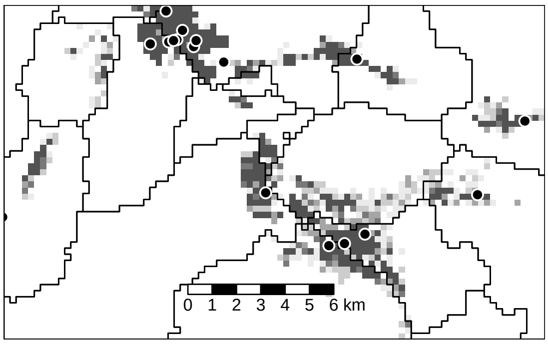


Figure 4: Municipality markets



Notes: Each pixel indicates one grid cell. Darker shades of gray (in Figures 1, 3 and 4) indicate larger population. Uninhabited cells are left blank. Large (small) clusters are colored in dark (light) gray in Figure 2. Supermarket locations are labeled with black circles with white frames. Black lines indicate borders of cluster markets and municipalities ("municipality markets") in Figures 3 and 4, respectively.

Table 1: Descriptive statistics on different types of markets

Delineation	Distance to next local market	Population density	Firm density: supermarkets	chimney sweeps	electricians	hairdressers	tourist agencies
cluster							
	$\leq$ 250m	10.9	1.0	0.1	0.5	0.8	0.1
	250m $< d_{ci} \leq$ 500m	41.5	2.0	0.2	1.3	1.6	0.4
	500m $< d_{ci} \leq$ 1km	87.7	3.9	0.7	2.2	4.1	0.7
	$>$ 1km	145.7	8.9	1.4	4.0	10.9	2.7
municipality							
	$\leq$ 250m	52.3	2.7	0.4	1.4	2.2	0.4
	250m $< d_{ci} \leq$ 500m	69.3	3.8	0.6	1.8	3.8	0.6
	500m $< d_{ci} \leq$ 1km	87.8	4.5	0.8	2.5	5.1	1.0
	$>$ 1km	137.3	8.4	1.3	3.8	10.4	2.6
difference between clusters and municipalities (in %)							
	$\leq$ 250m	−79	−61	−88	−62	−62	−69
	250m $< d_{ci} \leq$ 500m	−40	−48	−63	−29	−58	−41
	500m $< d_{ci} \leq$ 1km	0	−13	−13	−10	−21	−31
	$>$ 1km	+6	+6	+12	+5	+5	+1

Notes: Population density is measured in population per km^2 and firm density in number of firms per 100 km^2. d_{ci} denotes the Euclidean distance of grid cell c to the closest neighboring market i, i.e. $d_{ci} \leq d_{cj}\ \forall\ j \neq i$. Distances are calculated between the centroid of grid cell c and the centroid of the grid cell of market i closest to grid cell c.

2.2 Retail Industries and Market Demographics

We perform the empirical analysis for five retail industries in Austria. Firm level information for each industry was collected in 2015 by Herold Marketing CD 4/2015 (Herold Business Data Gmbh)—a telephone directory containing firms'[4] addresses and industry codes—and cross-validated by general research via internet.[5] The postal addresses of all firms are geocoded and linked to one specific local market. Following related work in this field, we include demographic market characteristics such as information on population, population density, age structure, educational attainment, income, employment and tourism, provided by Statistics Austria. A detailed description of these variables along with summary statistics are provided in Table A1 in the appendix.

We focus on supermarkets, chimney sweeps, electricians, hairdressers and tourist agencies. Three arguments were used when selecting these industries. The first and main criterion in our selection was to look at sectors where trade is likely to be highly localized and consumers seldom purchase from firms that are not in their immediate vicinity.[6] We expect that the market size for these industries is approximately equal to the size of an average municipality. Since competition is highly localized, the use of an appropriate geographic market definition plays a key role in estimating entry into these professions. Secondly, these purchases are likely to be made close to the consumers' place of residence and are usually private in character. This makes the use of residence data especially attractive. Lastly, entry and exit costs in these industries are limited, which allows for the use of cross-sectional data to estimate a model of equilibrium entry.

Assuming independent local markets implies that border regions between two markets accommodate neither consumers nor firms. Table 1 shows that the density of firms indeed decreases with proximity to the border to neighboring markets. While this spatial pattern is prevalent for both cluster and municipality markets, it is much more pronounced for the former market definition: In areas very close to the market boundaries (within 250m

[4]We use the term firm throughout the article to characterize a particular sales outlet.

[5]We are grateful to Dietmar Weinberger for preparing the data for us.

[6]Nishida (2015) provides empirical evidence of this for the convenience store market in Okinawa, Japan. He shows that the influence of the population on sales in the 1km grid cell, where the store is located, is about 50 times larger compared to the impact of inhabitants in neighboring grid cells, suggesting very narrow catchment areas.

distance) the firm density for cluster markets is between 61% and 88% lower than for municipality markets. In areas between 250m and 500m to the nearest border the firm density in cluster markets is still substantially smaller (between 29% and 63%). Based on these statistics, we expect spatial interactions between cluster markets to be much smaller than across municipality markets.

3 Entry Model and Threshold Analysis

In order to evaluate the level of interdependence of local markets we estimate a standard entry model (following Bresnahan and Reiss, 1991), but control explicitly for potential spatial spillover effects (as done in Lábaj et al., 2018). Specifically, the number of firms in each market is assumed to depend on two components. First, we model the overall profitability of the market π as a function of market size S, characteristics of the representative consumer X and an unobserved, normally distributed profitability shock ε. Additionally, markets incur profitability spillovers from neighboring markets. In order to account for these spillovers, we construct a spatial weights matrix W, which records the contiguity relationships between the spatial units (i.e. between local markets). For a given set of J markets, we define $C_i \in J$ as the set units of which share a border with market i. The matrix W has an element $w_{ij} = 1/\sum_j^J I(j \in C_i)$ if market j shares a border with unit i, and zero otherwise. As such, $W\pi$ represents the mean profitability of neighboring markets. This yields the following function for the profitability net of competitive effects:

$$\pi = \rho W\pi + \alpha \ln S + X\beta - \varepsilon$$

The parameter ρ measures the correlation in profitability across markets. This correlation occurs for two reasons. Firstly, consumers may choose to make purchases in nearby towns. This type of correlation can be interpreted as a violation of the independence assumption and implies that a complex model of interaction between sellers' entry decisions should be modeled across markets. Unmeasured similarities in economic conditions within a given region (for instance, similar zoning laws) may be the second source of correlation across spatial units. Since these similarities are not necessarily due to direct interaction between agents from the two markets, they do not present a violation of the independence assumption. The goal of our approach is to limit the amount of correlation across units in order to allow researchers to forego modeling spillover effects. This would then allow for a structural interpretation of all of the parameters without strategic effects spreading across local market borders.

The parameter of α should be constrained to unity under the representative consumer assumption. In our specification, it can be interpreted as the inverse of the standard deviation of the profitability shock, ε. Analogously, the vector of parameters β can be interpreted as the effect of X on the latent profitability scaled by the inverse of the standard deviation of the profitability shock.

The second component influencing the entry decision is determined by the decrease in market profitability due to competition, measured by the parameter θ_n, which is a fixed effect calibrating the magnitude of competitive pressure on profits given the presence of n competitors. The interaction of these two effects (market profitability and market structure) can be represented as a spatial ordered probit model:[7]

$$N = n \text{ if } \theta_n < \pi \leq \theta_{n+1}$$

Due to the limited number of observations with more than 3 sellers, we do not allow for competitive effects beyond the entry of the third company. As such, N is a censored variable.

Unlike most previous studies, we retain large markets in the sample in order to be able to estimate spatial interactions between adjacent markets. The model is estimated for cluster and municipality markets separately to compare the performance of each delineation technique afterwards. The results of the estimation from the cluster and municipality specifications are reported in the appendix (Table A2 and Table A3, respectively).

Since our main aim is to decrease the interdependence across markets, we first evaluate

[7]See LeSage and Pace (2009) for a discussion of the estimation procedure.

the parameter estimates of the spatial autocorrelation parameter ρ. The point estimates and the statistical significance levels are reported in Table 2, while a detailed illustration of the distribution of the sampled values is provided in Figure A1 in the appendix. The parameter estimates are significantly different from zero at the 5% significance level for all five retail industries when using municipality markets, while this is the case for only one out of five models when using cluster markets. The spatial autocorrelation parameters ρ are equally precisely estimated in both cluster and municipality markets. The difference in statistical significance stems from the point estimates being substantially closer to zero for cluster markets: The absolute values of the point estimates are between 22% (for hairdressers) and 94% (for tourist agencies) smaller when using our approach to define local markets, suggesting that profitability levels between contiguous cluster markets are substantially less correlated compared to adjacent municipalities. The difference in parameters is significant at the 5% level for three out of the five industries (chimney sweeps, electricians and tourist agencies). We also compare the draws of ρ in each round of the Bayesian MCMC estimation scheme and find that the absolute value using the cluster specification is smaller than in the municipal specification in most rounds. The percentage of times when this is true ranges between 64.3% (for hairdressers) and 97.7% (for chimney sweeps).

To evaluate the fit of the model based on these two alternative market definitions we investigate the share of markets for which the empirical models predict the number of firms correctly.[8] The cluster delineation results in a higher share of correctly predicted markets for all five industries, as reported in Table 3. The share of correctly predicted markets is between two and four percentage points higher for cluster markets, suggesting a modest gain in accuracy. Table 3 shows that the entry models based on cluster markets predict in particular empty markets and markets with at least three firms more precisely.

In our final approach to compare cluster with municipality markets we calculate the per-firm market size necessary for n firms to breakeven on an average market (i.e. with X set to the average value of the sample). We restrict our attention to food retailing and

[8] An overview of the frequency of specific market structures in our dataset is available in Table A5 in the appendix.

Table 2: Estimated spatial correlation coefficients ρ

Delineation	Super-markets	Chimney sweeps	Electricians	Hair-dressers	Tourist agencies
cluster	-0.064^{*} (0.033)	-0.270^{***} (0.055)	0.025 (0.037)	-0.054 (0.034)	-0.009 (0.058)
municipality	-0.120^{***} (0.033)	-0.424^{***} (0.053)	-0.078^{**} (0.039)	-0.070^{**} (0.033)	-0.162^{***} (0.055)
difference (in % in absolute values)	–46	–36	–67	–22	–94
p-value§	0.118	0.021	0.029	0.363	0.027
$\lvert\rho_{\text{cluster}}\rvert < \lvert\rho_{\text{municipality}}\rvert$ (% draws)	88.52	97.76	81.14	64.34	95.56

Notes: ***, ** and * indicate that parameters are significantly different from zero at the 1%, 5% and 10% levels, respectively. Standard errors are in parentheses.
§The p-value is calculated based on the following z-statistics using a one-sided test ($H_0 : \rho_{\text{cluster}} - \rho_{\text{municipality}} \leq 0$): $z = (\bar{\rho}_{\text{cluster}} - \bar{\rho}_{\text{municipality}})/\sqrt{\sigma_{\bar{\rho}_{\text{cluster}}} + \sigma_{\bar{\rho}_{\text{municipality}}}}$

Table 3: Goodness of fit (% of correctly predicted markets) by market structure

Model fit	Delineation	Super-markets	Chimney sweeps	Electricians	Hair-dressers	Tourist agencies
all markets	cluster	$\underline{62}$	$\underline{83}$	$\underline{61}$	$\underline{58}$	$\underline{86}$
	municipality	59	81	57	55	84
markets with 0 firms	cluster	$\underline{70}$	97	$\underline{84}$	$\underline{73}$	98
	municipality	52	97	77	63	98
markets with 1 firm	cluster	73	$\underline{35}$	42	60	17
	municipality	$\underline{80}$	33	$\underline{48}$	$\underline{65}$	$\underline{20}$
markets with 2 firms	cluster	27	$\underline{20}$	15	22	2
	municipality	$\underline{32}$	9	$\underline{17}$	$\underline{27}$	$\underline{3}$
markets with $\geq$ 3 firms	cluster	$\underline{53}$	$\underline{41}$	$\underline{33}$	$\underline{50}$	$\underline{34}$
	municipality	47	31	25	45	25

chimney sweeps, because we expect a relatively linear relationship between market size and the number of firms for these two industries: Food retailing is highly concentrated in Austria, with the companies REWE (comprising of the brands Billa, Merkur and Penny) and Spar controlling nearly 60% of all retail outlets (see Table A4 in the appendix for details). While about one fourth of all supermarkets are controlled by other chains, only about 15% are run by independent retailers. Except for these independent retailers, all brands price uniformly across markets (see Götz et al., 2008). As also the range of products and the size and design of the stores are harmonized within brands, the possibilities for retail stores to adapt to local market conditions are limited, and we thus expect to find an approximately linear relationship between the population and the number of firms in a local market.[9]

For chimney sweeps, price competition is restricted due to the presence of agreed tariffs at the federate state level.[10] Demand elasticity is likely to be low, since there is one sweep per building and individual tenants are thus often unable to switch suppliers. Furthermore, the Guild of Chimney Sweeps ("Landesinnung der Rauchfangkehrer") ensures homogeneity of the product across service providers, reducing the scope of firms' non-price reactions to a change in competitive pressure. Firms are thus unlikely to respond strongly to changes in competition, and we again expect a linear relationship between the number of firms and local market size, resulting in entry threshold ratios close to unity.[11]

The entry thresholds and entry threshold ratios, reported in Table 4, confirm our hypothesis when using cluster markets, but deviate from our expectation for municipality markets. For cluster markets, 1,166 inhabitants are necessary for the first supermarket to break even, as opposed to 916 residents using municipalities, reflecting that monopoly

[9]While individual brands do not adjust their pricing or product range to respond to local market characteristics, we acknowledge that the identities (brands) of the firms may depend on the market size. If this is the case, then retail stores in rural markets may be smaller because they don't belong to the major brands, which could result in small non-linearities in the estimated threshold ratios across markets of different size.

[10]Austria is divided into nine federal states, accommodating between 288 thousand and 1.7 million inhabitants.

[11]As electricians, hairdressers and tourist agencies are subject to less coordination, we expect that prices in these industries are likely to respond to changes in the local competitive climate but cannot make specific predictions about the level of the change. Given that industry knowledge does not provide a clear guidance on deriving threshold ratios which we could compare our empirical results with, the estimated entry thresholds for these industries are reported in the appendix (Table A6).

markets are in fact larger than they appear when relying on administrative boundaries. The cluster method seems better suited to take into account the fact that in certain areas, where municipalities are small, the relevant market may correspond to several adjacent administrative units. The per-firm entry threshold ratio between duopoly and monopoly markets is equal to 1.527 for municipalities. This suggests sizable competitive effects—which is unlikely due to uniform pricing of most brands. The per-firm entry threshold ratio between the second and the first firm based on cluster markets is 1.227, which is much closer to unity and hence to the expected ratio. In the case of chimney sweeps, on the other hand, the necessary population for three firms to break even is much smaller for cluster markets. This may be due to the ability of the cluster method to distinguish sub-markets in large municipalities. Again, entry threshold ratios are much closer to unity using cluster markets rather than municipality markets, further increasing our confidence in the approach to delineate local markets proposed in this article.

Table 4: Breakeven population and entry threshold ratios

Delineation	Supermarkets				Chimney sweeps			
	breakeven population		sequential entry threshold ratios		breakeven population		sequential entry threshold ratios	
	s_1	1,166			s_1	5,625		
cluster	s_2	1,431	s_2/s_1	1.227	s_2	9,534	s_2/s_1	1.695
	s_3	1,648	s_3/s_2	1.152	s_3	12,633	s_3/s_2	1.325
	s_1	916			s_1	7,664		
municipality	s_2	1,399	s_2/s_1	1.527	s_2	16,862	s_2/s_1	2.200
	s_3	1,631	s_3/s_2	1.166	s_3	23,533	s_3/s_2	1.396

Note: s_n denotes the per firm entry threshold population for the n^{th} firm.

4 Discussion

The increase in the availability of micro-level data over the last decade has led to substantial improvements in the accuracy of economic analysis. However, the high level of disaggregation also poses certain challenges when determining how individual information should be aggregated to the market level.[12] Additionally, as towns grow or infrastructure projects are carried out, original administrative boundaries diverge from the actual distribution of economic activity, manifested in the locations of both firms and potential consumers. Census data regarding residency, often available at a very detailed spatial level, reflect such changes timely and can be used to redefine local markets prior to the redrawing of administrative boundaries by public authorities.

In the present analysis we utilize such detailed population data and propose a methodology allowing researchers to generate so-called "cluster markets". This approach aims at defining local markets that are virtually isolated, which facilitates the analysis of firm and market conduct considerably. These cluster markets are proved to be characterized by border regions with much less economic activity as compared to local markets based on municipal boundaries, a market definition commonly applied in empirical competition analysis. Estimation results from entry threshold models based on cluster markets predict the observed market structure more accurately and suggest that cluster markets are indeed characterized by substantially less interdependence across neighboring markets. Furthermore, estimated entry threshold ratios meet our expectations—based on ex-ante industry knowledge—more closely than the results based on municipality markets.

The present application relies on grid data, but this approach can easily be applied for population data at the block, block group, census tract or municipal level as well. Data at the lowest level of regional aggregation available can serve as starting units to be added up into bigger, economically meaningful clusters. Such an endeavor would suffer from the shortcomings of administratively defined units, especially the lack of comparability across regional entities due to differences in size. This drawback is particularly distinct when the areas of the starting units are large relative to the catchment area of the good or service analyzed. We thus suggest to start with population data at the lowest level of regional aggregation available, so that these small units can be aggregated in a very

[12] An overview of Eurostat Census data from 2001 shows over 30% of the countries providing demographic data at the LAU2 (approx. municipal) level report at least one municipality with zero inhabitants. Furthermore, of the 90,062 observations in the Census dataset, 7,813 have fewer than 100 inhabitants. While this level of disaggregation allows researchers to analyze economic processes in detail, it also means that municipalities are not likely to act as independent economic units.

flexible way, reflecting the market size of the industry under scrutiny.

In the present analysis, we constrained ourselves to generating markets which are comparable to municipal markets. This influenced both the number of markets we generate and the type of industries used. This approach restricts the power of the proposed method. We would advice practitioners to impose other restrictions which are motivated more by market characteristics than by comparability to existing administrative units. One of the strengths of this method is its ability to generate markets of varying sizes, depending on the definition of "small" and "large" clusters.

The size of the generated markets tends to be bigger in urban areas, where clusters with high population density tend to cover a larger area. This allows us to identify suburbs which are in fact part of an urban agglomeration, despite being in a separate administrative unit. The cluster method is also useful in identifying natural boundaries within urban areas, such as rivers or large parks. If practitioners wish to derive multiple sub-markets from a single urban area, we recommend that they classify spatial units as having above-average population using mean values derived at the regional level.

Utilizing data on infrastructure or information on the mobility between regional units could improve the accuracy of the approach to define local markets outlined in this article. This, however, would considerably increase the data requirements for the implementation of this methodology. The approach proposed in this article thus negotiates the trade-off between precision and applicability in order to provide an easily implementable solution to the isolated markets problem inherent to the literature on localized competition analysis in general, and on entry threshold models in particular.

References

Aguirregabiria, V. and Suzuki, J. (2016). *Empirical Games of Market Entry and Spatial Competition in Retail Industries*, chapter 9, pages 201–233. Edward Elgar Publishers.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60:889-917.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009.

Cohen, A. M. and Mazzeo, M. J. (2007). Market structure and competition among retail depository institutions. *The Review of Economics and Statistics*, 89(1):60–74.

Davis, P. (2006). Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics*, 37(4):964-982.

Ellickson, P. B. (2006). Quality competition in retailing: A structural analysis. *International Journal of Industrial Organization*, 24(3):521–540.

Ellickson, P. B., Grieco, P. L., and Khvastunov, O. (2019). Measuring competition in spatial retail. mimeo.

George, L. M. and Waldfogel, J. (2006). The New York Times and the market for local newspapers. *American Economic Review*, 96(1):435–447.

Götz, G., Krauskopf, T., and Winker, P. (2008). Die Bestimmung regionaler Preisindizes—Das Beispiel Österreich. https://www.uni-giessen.de/fbz/fb02/fb/professuren/vwl/goetz/forschung/publikationenordner/arbeitspapiere/die-bestimmung-regionaler-preisindizes-das-beispiel-oesterreich.

Lábaj, M., Morvay, K., Silanič, P., Weiss, C., and Yontcheva, B. (2018). Market structure and competition in transition: Results from a spatial analysis. *Applied Economics*, 50(15):1694–1715.

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

Mazzeo, M. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics*, 33(2):221-242.

Nishida, M. (2015). Estimating a model of strategic network choice: The convenience-store industry in Okinawa. *Marketing Science*, 34(1).

Schaumans, C. and Verboven, F. (2015). Entry and competition in differentiated products markets. *The Review of Economics and Statistics*, 97(1):195–209.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, pages 619–640.

Appendix: Supplementary Tables and Figures

Table A1: Descriptive statistics on market demographics

Variable name	Description	Market definition	Mean	Std. Dev.	Min	Max
population	population size (2015)	cluster	4,107.4	37,056.8	381.0	1,580,885.0
		municipality	4,085.4	40,274.3	46.0	1,796,670.0
density	population per square kilometer (in 1,000 inhabitants)	cluster	0.113	0.202	0.002	4.283
		municipality	0.135	0.247	0.001	4.335
elderly	number of population aged 65 or older (in 2013) over population	cluster	0.184	0.030	0.118	0.374
		municipality	0.184	0.035	0.100	0.423
education	number of population with tertiary education (in 2013) over population	cluster	0.070	0.029	0.016	0.231
		municipality	0.068	0.031	0.000	0.270
income	total income (in 2012) over population (in 1,000 Euro)	cluster	17.929	4.044	0.688	35.200
		municipality	17.931	3.938	0.567	35.200
employment	number of employed individuals (in 2013) over population	cluster	0.490	0.034	0.344	0.581
		municipality	0.493	0.038	0.344	0.621
tourism	average daily overnight stays (between Nov. 1st 2014 and Oct. 31st 2015) over population	cluster	0.073	0.216	0.000	2.786
		municipality	0.077	0.230	0.000	2.786

Notes: All data have been collected by Statistics Austria. Information on population is provided at the grid cell level, while all other variables are available at the municipality level. Total income includes wages, business income, pensions, transfers and other income. The number of observations equals 2,089 for cluster markets and 2,099 for municipality markets. Data at the municipality level is translated into cluster markets by taking the weighted average of all municipalities comprised by a particular cluster. Thus, the weight of municipality j for calculating the variable for cluster i, ω_{ij}, is defined as the population of municipality j in cluster i, pop_{ij}, over the total population in cluster i, pop_i, i.e. $\omega_{ij} = \frac{pop_{ij}}{pop_i}$.

Table A2: Estimation results (cluster delineation)

	Super-markets	Chimney sweeps	Electricians	Hair-dressers	Tourist agencies
intercept	-9.431^{***}	-8.994^{***}	-8.576^{***}	-9.858^{***}	-9.733^{***}
	(0.795)	(1.207)	(0.855)	(0.813)	(1.250)
population (log)	1.176^{***}	1.175^{***}	1.043^{***}	1.167^{***}	1.095^{***}
	(0.035)	(0.054)	(0.037)	(0.036)	(0.055)
density	-0.550^{***}	-1.155^{***}	-0.739^{***}	-0.579^{***}	-0.266
	(0.140)	(0.236)	(0.162)	(0.146)	(0.178)
elderly	5.471^{***}	4.244^{**}	0.254	3.997^{***}	1.259
	(1.219)	(1.789)	(1.330)	(1.235)	(1.890)
education	-6.355^{***}	2.046	1.075	-2.473^{**}	1.761
	(1.068)	(1.635)	(1.093)	(1.086)	(1.621)
income	0.006	-0.014	0.014^{*}	0.008	-0.002
	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
employment	0.926	-3.397^{**}	0.638	1.504	-0.309
	(1.117)	(1.707)	(1.189)	(1.133)	(1.715)
tourism	0.871^{***}	-0.383	0.460^{***}	0.675^{***}	0.737^{***}
	(0.121)	(0.274)	(0.128)	(0.122)	(0.162)
ρ	-0.064^{*}	-0.270^{***}	0.025	-0.054	-0.009
	(0.033)	(0.055)	(0.037)	(0.034)	(0.058)
θ_2	1.055^{***}	1.434^{***}	0.979^{***}	0.956^{***}	0.751^{***}
	(0.024)	(0.057)	(0.034)	(0.054)	(0.055)
θ_3	1.698^{***}	2.241^{***}	1.657^{***}	1.570^{***}	1.208^{***}
	(0.045)	(0.090)	(0.054)	(0.054)	(0.081)
# of observations	2,089				

Notes: ***, ** and * indicate that parameters are significantly different from zero at the 1%, 5% and 10% levels, respectively. Standard errors are in parentheses.

Table A3: Estimation results (municipality delineation)

	Super-markets	Chimney sweeps	Electricians	Hair-dressers	Tourist agencies
intercept	-7.557^{***}	-8.315^{***}	-8.308^{***}	-8.987^{***}	-11.226^{***}
	(0.747)	(1.160)	(0.846)	(0.781)	(1.314)
population (log)	1.077^{***}	0.998^{***}	0.993^{***}	1.111^{***}	1.165^{***}
	(0.037)	(0.056)	(0.040)	(0.040)	(0.062)
density	-0.257^{**}	-0.181	-0.173	-0.174	0.043
	(0.113)	(0.142)	(0.117)	(0.117)	(0.142)
elderly	5.441^{***}	6.336^{***}	1.173	4.247^{***}	3.081^{*}
	(1.025)	(1.575)	(1.186)	(1.093)	(1.791)
education	-3.390^{***}	1.627	0.351	-2.011^{*}	0.993
	(1.006)	(1.438)	(1.071)	(1.033)	(1.446)
income	0.001	-0.008	0.009	0.014^{**}	0.006
	(0.007)	(0.009)	(0.008)	(0.007)	(0.010)
employment	-1.231	-3.487^{**}	0.744	0.261	0.281
	(0.996)	(1.520)	(1.113)	(1.020)	(1.671)
tourism	0.778^{***}	-0.191	0.551^{***}	0.694^{***}	0.790^{***}
	(0.115)	(0.216)	(0.123)	(0.113)	(0.159)
ρ	-0.120^{***}	-0.424^{***}	-0.078^{**}	-0.070^{**}	-0.162^{***}
	(0.033)	(0.053)	(0.039)	(0.033)	(0.055)
θ_2	1.203^{***}	1.479^{***}	1.011^{***}	0.972^{***}	0.852^{***}
	(0.030)	(0.077)	(0.028)	(0.042)	(0.056)
θ_3	1.805^{***}	2.217^{***}	1.701^{***}	1.575^{***}	1.324^{***}
	(0.030)	(0.094)	(0.045)	(0.079)	(0.078)
# of observations	2,099				

Notes: ***, ** and * indicate that parameters are significantly different from zero at the 1%, 5% and 10% levels, respectively. Standard errors are in parentheses.

Table A4: Number of retail outlets

Brand	# of outlets	Share (in %)
Billa	1,045	21.1
Merkur	123	2.5
Penny	273	5.5
Spar	1,538	31.0
Hofer	455	9.2
Lidl	195	3.9
small chains:		
Mpreis, Unimarkt, Zielpunkt	559	11.3
independent retailers:		
Adeg, Nah & Frisch	772	15.6
total	4,960	100.0

Table A5: Market counts by profession and number of entrants

Delineation	Number of firms	Super-markets	Chimney sweeps	Electricians	Hair-dressers	Tourist agencies
cluster	0	702	1,642	1,143	843	1,757
	1	703	359	538	601	187
	2	297	59	217	268	58
	3+	387	29	191	377	87
municipality	0	549	1,615	1,066	730	1,733
	1	817	402	591	646	214
	2	313	56	242	306	60
	3+	420	26	200	417	92

Table A6: Breakeven population and entry threshold ratios for electricians, hairdressers and tourist agencies

Delineation	Electricians: breakeven population		Electricians: sequential entry threshold ratios		Hairdressers: breakeven population		Hairdressers: sequential entry threshold ratios		Tourist agencies: breakeven population		Tourist agencies: sequential entry threshold ratios	
cluster	s_1	2,026			s_1	1,364			s_1	6,063		
	s_2	2,588	s_2/s_1	1.277	s_2	1,547	s_2/s_1	1.134	s_2	6,017	s_2/s_1	0.992
	s_3	3,305	s_3/s_2	1.277	s_3	1,747	s_3/s_2	1.129	s_3	6,092	s_3/s_2	1.012
municipality	s_1	1,949			s_1	1,248			s_1	6,760		
	s_2	2,697	s_2/s_1	1.394	s_2	1,497	s_2/s_1	1.200	s_2	7,021	s_2/s_1	1.039
	s_3	3,600	s_3/s_2	1.335	s_3	1,717	s_3/s_2	1.147	s_3	7,017	s_3/s_2	0.999

Note: s_n denotes the per firm entry threshold population for the n^{th} firm.

Figure A1: Histogram of sampled values of ρ