Quality Differentiation and Spatial Clustering among Restaurants

Mossay, Pascal and Shin, Jong Kook and Smrkolj, Grega

Kyungpook National University, Universite catholique de Louvain, Newcastle University

18 February 2020

Quality Differentiation and Spatial Clustering among Restaurants*

Pascal Mossay[a] Jong Kook Shin[b] Grega Smrkolj[c]

August 19, 2020

Abstract

To explore the relationship between spatial location and quality differentiation, we build a dataset of over 30,000 restaurants rated by TripAdvisor, across large UK cities. Whereas top-rated restaurants tend to locate close to other top restaurants, bottom-rated restaurants tend to locate away from each other and closer to top ones. Our theoretical model can explain the main features of observed spatial patterns. We find that an increase in the population density in the city center reduces the spatial dispersion of both top and bottom restaurants but this reduction is larger in magnitude for top restaurants. Also, a larger quality difference between top and bottom restaurants increases both the absolute and relative dispersion of top restaurants.

Keywords: Spatial competition, Quality differentiation, Hotelling model, Restaurant industry
JEL: L13, L83, R12, R32

1 Introduction

The restaurant industry in the United States generated over USD 800 billion in sales in 2019. In fact, restaurants have become so widespread that the National Restaurant Association reports that nearly 6 in 10 adults have worked in the restaurant industry at some point during their lives.[1] The trend is similar in other western countries. In the United Kingdom, eating out is so popular that the revenue from restaurants and other food services constitutes the largest share of the leisure sector revenue.[2] Restaurants have now become something much more than simply a place to eat and are part of busy modern life, thus constituting an important component of urban

*This research made use of the Rocket High Performance Computing service at Newcastle University.

[a]Kyungpook National University, Korea, and CORE, Université catholique de Louvain, Belgium. Email: pmossay@outlook.com.

[b]Newcastle University Business School. Email: Jong.Shin@newcastle.ac.uk.

[c]Corresponding Author. Address: Newcastle University Business School, 5 Barrack Road, Newcastle upon Tyne, NE1 4SE, United Kingdom. Email: Grega.Smrkolj@newcastle.ac.uk.

[1]2019 Restaurant Industry Factbook, accessed January 25, 2020, `http://restaurant.org`.

[2]"Restaurant industry in the United Kingdom (UK) - Statistics & Facts", Statista Research Department, September 3, 2018, accessed January 25, 2020, `https://www.statista.com/topics/3131/restaurant-industry-in-the-united-kingdom-uk/`.

amenities. Nowadays, to stay competitive, restaurants need to do their best to provide a unique dining experience and many of them now offer a wide range of food and beverages, from coffee and cocktails to salads and healthy eating options. Taking care of their online reputation has become crucial for restaurant owners. Not only does a majority of consumers consult online reviews[3] but more than one in three consumers would also generally not eat in a restaurant rated below 4 stars on online review sites like TripAdvisor and Google.[4] Among various new sources of differentiation, location still plays an important role. According to the National Restaurant Association, 56% of consumers would choose a restaurant within a walking distance over another.

Given the intrinsic importance of location and product differentiation, the restaurant industry is a natural choice to study the relation between these two factors affecting competition. In this study, we explore whether there are any systematic differences in the location of top- and bottom-rated restaurants. In other words, we study the spatial competition among quality-differentiated firms, where quality is that perceived by consumers. To do this, we build a unique, hand-collected dataset, which maps over 30,000 restaurants listed and rated on TripAdvisor's online review site, across cities in England and Wales. We find that top-rated restaurants tend to be spatially more concentrated than bottom-rated ones and locate closer to the city center. Whereas top-rated restaurants tend to locate closer to other top-rated restaurants, bottom-rated ones tend to locate away from other bottom-rated restaurants and closer to top-rated ones. Moreover, bottom-rated restaurants tend to be less clustered than top-rated restaurants.

The spatial concentration of companies and firm clustering have attracted great attention both from a theoretical perspective (e.g., Krugman, 1991; Porter, 2000) and an empirical research perspective (Duranton and Overman, 2005; Marcon and Puech, 2003, and numerous others). Of particular interest is the question about which parameters determine the formation of clusters. Gordon and McCann (2000), for instance, compare the advantages and disadvantages of geographical proximity as perceived by business leaders in different sectors and conclude that agglomeration advantages usually far outweigh the disadvantages of increased competition. We contribute to this extensive literature with a novel study on the potential relation between quality differentiation and the spatial clustering of firms (restaurants) in cities.

The empirical literature on restaurants typically focuses on the impact of competition (e.g., an increase in the number of firms) on prices and on the relation between prices and quality. For instance, De Silva et al. (2016) recently show that competition does not decrease but rather increases restaurants' prices. Due to agglomeration economies, restaurants benefit from positive externalities in denser, well-served restaurant areas, which attract more consumers. They also show a positive relationship between the prices charged by restaurants and their quality.

In cities, population density affects the variety of available products. Schiff (2015) shows that a higher population density increases the diversity of cuisines and the range of restaurant quality levels in cities. In high-density areas, consumption benefits are large because consumers can

[3] "Over three-quarters of UK consumers check online reviews, with TripAdvisor 2.5 times more influential than Google," Uberall press release, June 26, 2019, accessed January 25, 2020, https://uberall.com/en-gb/company/press-releases/over-three-quarters-of-uk-consumers-check-online-reviews-with-tripadvisor-2-5-times-more-influential-than-google.

[4] "5 Industries Impacted by Online Reviews", accessed January 25, 2020, https://onlinegeniuses.com/5-industries-impacted-by-online-reviews-infographic/.

visit more restaurants they prefer. Couture (2016) estimates the average household's willingness to pay to move to a denser area. Both of these studies reemphasize the role of cities as centers of consumption (see Glaeser et al., 2001).

Our analysis focuses neither on the relationship between quality and price nor on quality competition. Rather, we contribute to the literature by exploring the relationship between restaurant quality and spatial clustering that we observe in the data. We also propose a competition model between restaurants that can explain the main properties of spatial patterns observed in our sample. Our model builds on models of imperfect competition among firms à la Hotelling. Four restaurants are competing in terms of price and location for consumers distributed along a line segment. As in the standard Hotelling duopoly model, the subgame perfect equilibrium in price and location is obtained by backward induction. Whereas the locations are chosen in the first stage of the game, the prices are set in the second one.

In the literature, few theoretical results exist about multiple-firm Hotelling models. Brenner (2005) analyzes a multi-firm Hotelling model under quadratic transport costs. He shows that in the second stage of the game, a price equilibrium exists and is unique for any number of firms. However, due to the analytical complexity related to the number of firms, Brenner relies on numerical computations to determine the location equilibria in the first stage of the game for up to nine firms. The multi-firm Hotelling model under linear transport costs was studied earlier by Economides (1993). In contrast to the model with quadratic costs by Brenner (2005), only part of the market is served in equilibrium.

In our model, unlike in the traditional Hotelling model, consumers also have idiosyncratic tastes about restaurants. This means that they visit and buy from all restaurants with a positive probability. The logit Hotelling model with multiple firms under linear transport costs was studied by de Palma et al. (1985). They show that if consumer taste heterogeneity is high enough, a single location equilibrium is obtained, in which all firms locate in the city center. In the case of three firms, de Palma et al. (1987) show that lower heterogeneity levels lead firms to disperse along the line segment and multiple location equilibria can emerge.

In our model, firms are differentiated by location but also by the quality of the good they serve to consumers. Like Tseng et al. (2010), we assume that the qualities of goods are exogenous (i.e., related to long-term decisions), whereas firms' locations are endogenous (i.e., a short-term choice). Tseng et al. (2010) show that spatial dispersion (resp. concentration at the center) occurs if the difference in quality is small enough (resp. large enough). However, our model is quite different from theirs as it involves two firms of each quality type, meaning that both inter- and intra-type competition is present. Moreover, unlike in Tseng et al. (2010), consumer heterogeneity in our model results from idiosyncratic tastes about restaurants rather than from differences in consumers' willingness to pay for quality.

As in Brenner (2005), our multi-firm model admits a unique price equilibrium in the second stage of the game. However, here, in addition to having multiple-firms as in the oligopoly model by Brenner (2005) or the logit oligopoly model by de Palma et al. (1985), the quality difference between top and bottom restaurants introduces additional complexity. Our model can only be solved using a numerical approach. We develop a numerical algorithm that relies on a state-of-the-art nonlinear solver to find both symmetric and asymmetric equilibria, of which there can be many. As in the triopoly model of de Palma et al. (1987), a lower level of consumer taste heterogeneity fosters spatial dispersion and induces the emergence of multiple location equilibria in the first stage of the game.

In contrast to the literature focusing on the effect of competition (e.g., the number of competitors in the market or the level of transport costs) on price and quality (see, e.g., Ma and Burgess, 1993), we focus on explaining properties of spatial patterns of top and bottom restaurants (e.g., their absolute and relative spatial dispersion) in relation to their quality.

When consumer taste heterogeneity is relatively low, transport costs play a prominent role in buying decisions. The city center attracts top restaurants, as it provides them with the best access to the customer base. Top restaurants are able to drive bottom restaurants away from the center through quality competition. The latter ones then disperse around top restaurants, primarily serving customers living outside the city center. However, when consumer taste heterogeneity is relatively high, buying decisions appear as rather random to restaurants. Consequently, the city center offers relatively little advantage in terms of access to the customer base. As competition is fiercer between top restaurants than bottom ones, top restaurants disperse more.

We also model non-uniform spatial densities of consumers. For a two-firm model, Anderson et al. (1997) have shown that tight density functions constitute an agglomeration force leading to lower prices. Considering quality-differentiated firms, we show that the magnitude of the quality difference between top and bottom restaurants qualifies this result. Although a larger concentration of consumers in the center attracts both types of restaurants, a larger quality difference increases the competition between top restaurants, forcing them to move apart. The relative magnitude of these two effects determines which restaurant type locates closer to the center.

We derive a number of testable predictions from our model regarding the spatial dispersion of top and bottom restaurants and the effect of quality. In turn, we run regressions to test these hypotheses and are able to validate most of them. We find that an increase in the population density in the city center reduces the spatial dispersion of both top and bottom restaurants but this reduction is larger in magnitude for top restaurants. A larger quality difference between top and bottom restaurants increases both the absolute and relative dispersion of top restaurants.

The rest of this paper is organized as follows: In Section 2, we describe the data and explore the spatial patterns of restaurants to establish some new stylized facts. In Section 3, we present a stylized model that is able to reproduce them and examine the relative spatial distribution of restaurants under different parameter configurations. In Section 4, we empirically test our model's predictions and outline those that leave room for further study. Section 5 concludes.

2 Spatial Analysis

2.1 Data

We built our dataset from four main sources.[5]

1. *Data on restaurants and their quality*

[5] All data were obtained and used in accordance with The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, UK.

First, we chose a set of over 100 of the largest cities in the UK by population in 2017.[6] We then listed for each city all the restaurants ranked on ©TripAdvisor's website from January 25 to February 14, 2018, using a restaurant search by city. We chose the restaurant rankings of TripAdvisor, as it is by far the most widely used review site for restaurants.[7]

We take a restaurant's ranking according to the TripAdvisor Popularity Index as an indicator of its overall quality perceived by customers. The index is calculated by a proprietary algorithm and is usually updated weekly. The index considers quantity, quality, and recency of reviews to determine reviewers' overall satisfaction with a restaurant. An online tracking system, as well as a "dedicated team of investigators," is employed by TripAdvisor to prevent and remove fake reviews.

TripAdvisor's list of restaurants by city often includes restaurants located in nearby cities and towns. Moreover, some restaurants have no ranking and/or addresses. These observations were dropped from the sample.

2. *City boundaries*
   City boundaries in the form of geometric polygons with GPS coordinates were obtained from ©OpenStreetMap[8] and MapIt.[9] These maps did not always clearly distinguish between the city and its suburbs. We analyzed only cities that were clearly demarcated and matched closely TripAdvisor's definition of them.

3. *GPS locations of restaurants*
   Restaurants' mailing addresses, obtained from TripAdvisor, were converted to GPS coordinates via APIs provided by Google Maps™ and Bing Maps™. After geocoding, a few restaurants could not be located within relevant city boundaries. When addresses resulted in invalid GPS coordinates (e.g., towns in other countries), observations were dropped from the sample. We also excluded London as it is an outlier in terms of both the area and the number of restaurants.

4. *Population data*
   The UK does not have a central population registry. Unlike England and Wales, Scotland and Northern Ireland collect separate population data. For consistency, we decided to restrict our sample to the cities in the England and Wales as most of the largest UK cities are located

[6] 2017 World Population Review, "Population of Cities in United Kingdom (2017)," accessed January 25, 2018, http://worldpopulationreview.com/countries/united-kingdom-population/cities.

[7] A TripAdvisor-sponsored survey among its 9,500 registered consumers in the US and EU in 2018 claims that it is by far the most widely used and trustworthy restaurant review website. In the UK, 87% of all respondents agreed that online reviews influence their dining decisions. Moreover, 64% of respondents said they prefer to use TripAdvisor (Google 22%, Facebook 8%) while at home and 70% while traveling (Google 21%, Facebook 5%); 93% of UK respondents agreed that TripAdvisor's reviews matched their dining experiences (TripAdvisor, "Influences on Diner Decision-Making," accessed November 20, 2018, https://www.tripadvisor.com/ForRestaurants/r3227). Our own research confirms that TripAdvisor is the most comprehensive guide for the UK. For instance, the combined number of reviews for the five highest-rated restaurants in Newcastle upon Tyne was 125 on Yelp, 1,837 on Google, and 2,581 on TripAdvisor (data accurate as of July 4, 2020).

[8] Data is available under the Open Database Licence, http://www.openstreetmap.org/copyright.

[9] MapIt contains Ordnance Survey data ©Crown copyright and database rights 2010-17, which is licensed under the Open Government Licence v3.0.

Table 1: Descriptive statistics

	mean	sd	min	max
# top	82.44	70.23	18	374
# bottom	82.74	70.22	19	375
area (km^2)	207.91	340.75	10.64	2224.19
population (in 1000)	215.84	151.72	26.80	1073.05
Observations	96			

in those two countries. Our population data are sourced from the 2011 Census for England and Wales.[10] England and Wales are divided into 181,408 geographical units called Output Areas (OAs), the lowest geographical level for which census estimates are available. The average number of residents living in an OA was 309 in 2011. We allocated OAs to cities in the sample by using their population-weighted centroids with GPS coordinates.

Our final sample for our empirical analysis consists of 31,715 restaurants across 96 large cities in England and Wales (for the list of cities, see Appendix A).

2.2 Spatial Analysis

This analysis aims to explore the location differences between "good" and "bad" restaurants. To this end, we rank restaurants in each of the 96 cities in the sample from best to worst using the TripAdvisor Popularity Index. Next, all restaurants below the 25$^{\text{th}}$ percentile (the 1$^{\text{st}}$ quartile) of the ranking are labeled *Top* and all restaurants at or above the 75$^{\text{th}}$ percentile (the 4$^{\text{th}}$ quartile) are labeled *Bottom*.[11] Table 1 provides basic summary statistics of our dataset.

In the following, we investigate several differences in the location of top and bottom restaurants such as i) spatial dispersion, ii) spatial centricity, and iii) spatial dependence.

2.2.1 Spatial dispersion

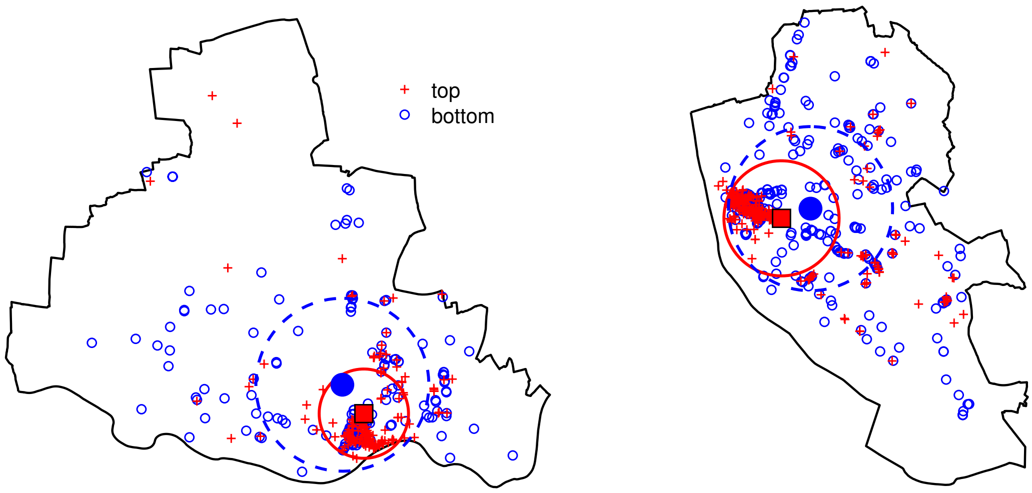
We first examine whether top (T) and bottom (B) restaurants disperse differently across the city. To this end, we calculate the average distance of restaurants in each group from that group's geometric centroid (respectively d_i^T and d_i^B) for each city i, $i \in [1, 96]$. All distances are calculated as great-circle distances.[12] Figure 1 shows the location of top and bottom restaurants in two sample cities—Newcastle and Liverpool. In both cities, bottom restaurants are more dispersed than top ones.[13]

[10] The Office for National Statistics licensed under the Open Government Licence v3.0.

[11] Our results are robust to restaurant classifications based on different percentiles. See Appendix B for various robustness checks.

[12] The "great-circle" distance is the shortest distance between two points on the spherical earth (i.e., as the crow flies).

[13] Maps of other cities in the sample are available on the corresponding author's website.



(a) Newcastle (b) Liverpool

Figure 1: Location of top and bottom restaurants in Newcastle and Liverpool. The dashed (blue) circle has a radius equal to the average distance of bottom restaurants from their geometric centroid (indicated by a blue dot). The solid (red) circle has a radius equal to the average distance of top restaurants from their geometric centroid (indicated by a red square).

We use the ratio $rd_i = d_i^B/d_i^T$ as a measure of relative dispersion of top versus bottom restaurants in city i. See Figure 2a for this measure of relative dispersion for all cities ranked by total area from the smallest to the largest. The correlation between the average distances of restaurants from their group's geometric centroid and the city area is positive and significant, although modest, for both top $(r = 0.239, p = 0.019)$ and bottom $(r = 0.391, p = 0.000)$ restaurants. However, the relative dispersion between top and bottom restaurants does not vary with city size $(r = 0.008, p = 0.936)$—the points in Figure 2a are uniformly scattered along the horizontal axis. In most cities, bottom restaurants seem relatively more dispersed: 64% of all points lie above the horizontal line $rd_i = 1$.

We formally test the null hypothesis of no difference in the spatial dispersions between the two groups of restaurants by using Anderson's (2006) distance-based multivariate generalization of Levene's test for homogeneity of variances.[14] In Figure 2a, the filled circles indicate the cities for which the dispersion difference between the two groups is significant at the .05 level. The dispersion difference is significant in 41 out of the 96 cities. In all but 5 significant cities, bottom restaurants are relatively more dispersed than top ones.

2.2.2 Spatial centricity

While the above test reveals that bottom restaurants tend to be more dispersed than top ones, it does not tell us whether they potentially locate in different places, in particular, whether one

[14] Specifically, we use a principal coordinate transformation of the dissimilarity matrix based on great-circle distances between restaurants to calculate the ANOVA F-statistic. We execute the more robust version of Levene's test and calculate deviations of group members from the group's spatial median based on principal coordinate axes without making any assumptions about the distribution of distances. Accordingly, we obtain the p-value by 9,999 permutations of the least-absolute-deviation residuals. The test was executed using the `R` package `vegan` 2.5-3 (Oksanen et al., 2018).

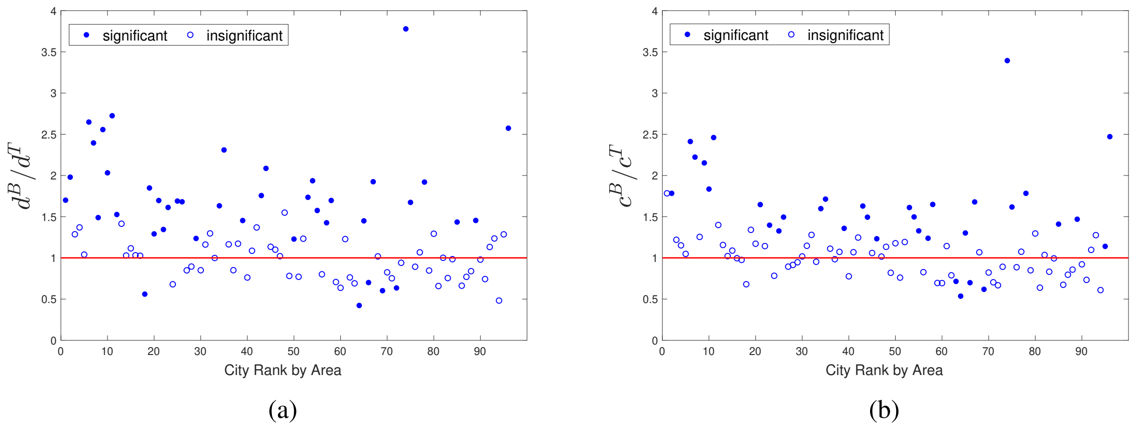


Figure 2: Relative dispersion (a) and relative centricity (b) between bottom and top restaurants across 96 cities in England and Wales. Cities are ranked by total area from the smallest to the largest.

group is more concentrated in the city center than the other. For this, we first need to define the city center. The literature documents many proxies that serve this purpose (e.g., a city map center point, main railway or bus station, main post office), of which none has a clear advantage. For our research, we define the city center as the geographic centroid (center of gravity) of all restaurants in the city, which can be intuitively interpreted as the city's dining center. As restaurants tend to concentrate in places where people spend time together, our center should indicate where the average individual shows up when going out for dinner.

For each group of restaurants, we calculate the average distance of restaurants to the city center (respectively c_i^T and c_i^B) and then use the Mann–Whitney–Wilcoxon two-sided test to decide if the difference between groups is significant. Figure 2b shows the relative average distances (c_i^B/c_i^T) of top and bottom restaurants across all cities ranked by area. The filled circles indicate cities for which the test is significant at the .05 level. In 61 out of the 96 cities, top restaurants are on average located closer to the center ($c_i^B/c_i^T > 1$), but the difference is significant for only 34 of them. However, out of the latter, in all but four cities, top restaurants are located closer to the center.

Overall, based on the two tests in Figure 2, we conclude that top restaurants tend to be spatially more concentrated than bottom ones and locate closer to the city center.

2.2.3 Spatial dependence

Our next question is whether restaurants of a certain type tend to be surrounded by restaurants of the same or another type. To answer this, we use Ripley's bivariate K function (see, e.g., Baddeley et al., 2015; Dixon, 2002), which corresponds to the expected number of type j restaurants within a distance r of a typical type i restaurant, standardized by dividing by the density (number per unit area) of type-j restaurants (λ_j):

$$K_{i,j}(r) = \frac{1}{\lambda_j}\mathbb{E}[\text{number of type } j \text{ restaurants within a distance } r \\ \text{of a randomly chosen type-}i \text{ restaurant}]. \tag{1}$$

For a city map of area A with n_i type-i restaurants and n_j type-j restaurants, an estimator of the above function is

$$\hat{K}_{i,j}(r) = \left(\hat{\lambda}_i\hat{\lambda}_j A\right)^{-1} \sum_{k=1}^{n_i}\sum_{l=1}^{n_j} I(d_{i_k,j_l} \leq r)e(i_k, j_l), \tag{2}$$

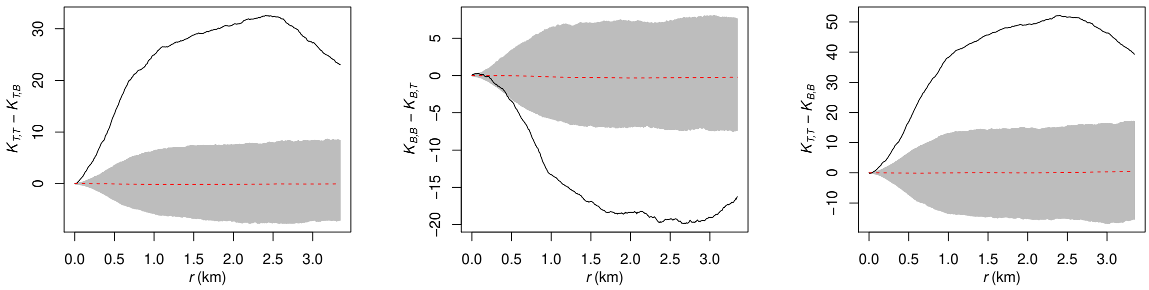
where $\hat{\lambda}_i = {}^{n_i}/_{A}$ and $\hat{\lambda}_j = {}^{n_j}/_{A}$ are the observed densities of type-i and type-j restaurants, respectively; d_{i_k,j_l} is the distance between the k-th restaurant of type i and the l-th restaurant of type j; $I(z)$ is an indicator function that takes 1 if z is true and 0 otherwise; and $e(i_k, j_l)$ is an edge correction weight. The latter is employed to reduce the estimation bias due to the fact that restaurants outside the city boundary are not counted even though they are within a distance r. We therefore use the isotropic correction method, in which $e(i_k, j_l)$ is the reciprocal of the fraction of the circumference of a circle centered at the k-th restaurant of type i with radius d_{i_k,j_l} that lies inside the city boundary. As the variance of the estimator increases with distance r, we apply Ripley's rule of thumb and take as its maximum value one quarter of the shortest side of the rectangle enclosing the city polygon (for details, see Baddeley et al., 2015).

Under the null hypothesis of random labeling, the type of each restaurant (top or bottom) is determined randomly, independently of other restaurants, so that $K_{T,B}(r) = K_{B,T}(r) = K_{T,T}(r) = K_{B,B}(r) = K(r)$. We examine departures from random labeling using three pairwise differences between bivariate K functions: $K_{T,T}(r) - K_{T,B}(r)$ and $K_{B,B}(r) - K_{B,T}(r)$ to evaluate whether restaurants tend to be surrounded by restaurants of the same or the other type, and $K_{T,T}(r) - K_{B,B}(r)$ to evaluate whether restaurants of one type cluster more or less than those of the other type.

We make inferences based on a permutation test, fix the number of observed restaurants of each type and their spatial locations in a city accordingly, and then randomly permute the types (labels) of restaurants. For each of 999 randomly permuted datasets, we then evaluate three pairwise differences between the K functions and compute their pointwise envelopes by finding for each distance r their 25-th largest and 25-th smallest values among all simulated values (upper and lower 2.5% quantiles). For a given distance r, the test rejects the null hypothesis of random labeling if the empirical estimate of the relevant pairwise difference lies outside the envelope limits at the $(25 + 25)/(999 + 1) = 0.05$ significance level.

All tests and simulations in this section were executed using the well-known `spatstat` (Baddeley et al., 2015) package for spatial point pattern analysis, written in `R`. As the pertinent K functions are based on Euclidean distances, all GPS locations of restaurants were projected using the British National Grid projection (OSGB36).

Figure 3 shows the results for a city in our sample, Newcastle upon Tyne, which is the most populous city in North Eastern England. In Figure 3a, the null hypothesis would be rejected at the 5% significance level for any choice of distance r. Positive deviations of $\hat{K}_{T,T}(r) - \hat{K}_{T,B}(r)$



(a) (b) (c)

Figure 3: Spatial dependence between restaurants in Newcastle. Solid line: empirical estimate of pairwise differences. Dashed line: sample mean of simulations from the null hypothesis of random labeling. Gray shading: pointwise 5% critical envelopes based on 999 simulations.

suggest that top restaurants are more likely to be found close to other top restaurants than if the type of restaurant was randomly allocated.[15] However, In Figure 3b, the deviations of $\hat{K}_{B,B}(r) - \hat{K}_{B,T}(r)$ are negative and below the 2.5% quantile for all distances above 0.5 km, which suggests that bottom restaurants are more likely to be found close to a top restaurant. That is, while top restaurants concentrate, bottom restaurants tend to disperse.[16] Finally, in Figure 3c, the deviations of $\hat{K}_{T,T}(r) - \hat{K}_{B,B}(r)$ are positive and significant for all distances. This suggests that top restaurants cluster more than bottom ones, which is in line with our findings in the previous section.

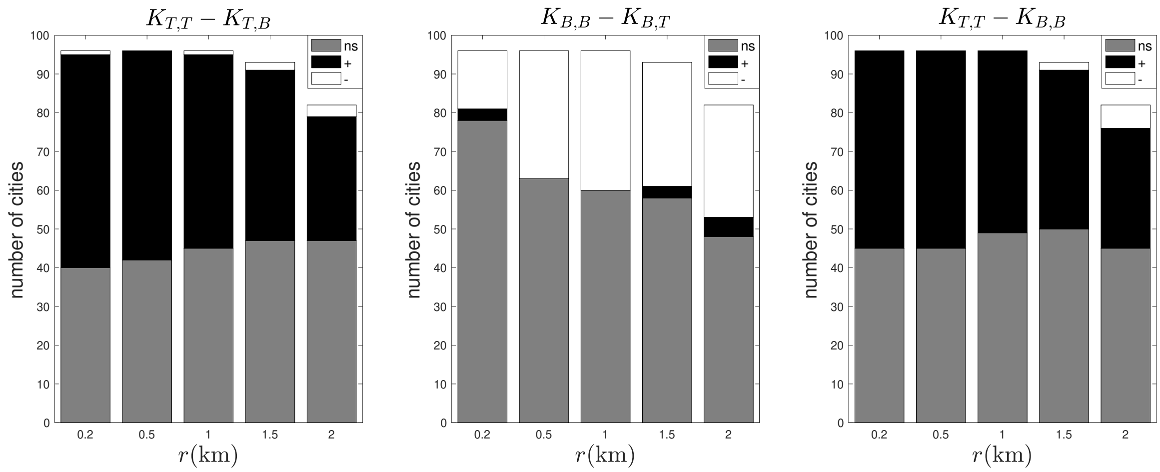
Figure 4 summarizes the results for the entire sample of cities for a radius of up to 2 km.[17] We observe that, for most cities with significant differences, top restaurants tend to locate closer to other top restaurants (Figure 4a), while bottom restaurants tend to locate away from other bottom restaurants and closer to top restaurants (Figure 4b). Also, bottom restaurants as a group tend to be less clustered than the top ones (Figure 4c). For instance, at a distance of 1 km, the difference $\hat{K}_{T,T}(r) - \hat{K}_{T,B}$ is significant (and positive for all but one city) in 51 of the 96 cities. Further, the difference $\hat{K}_{B,B}(r) - \hat{K}_{B,T}$ is significant and negative in 36 cities. Finally, the difference $\hat{K}_{T,T}(r) - \hat{K}_{B,B}$ is significant and positive in 47 cities. Results for other distances are similar.

Acknowledging the limitations of aggregating results for cities of different sizes, we present, in Figure 5, the estimates of pairwise differences between pooled bivariate K functions, together

[15] In Figures 3 and 4, we replace $\hat{K}_{T,B}(r)$ and $\hat{K}_{B,T}(r)$ with a more efficient estimator, $K^*_{T,B}(r) = \frac{\hat{\lambda}_B \hat{K}_{T,B}(r) + \hat{\lambda}_T \hat{K}_{B,T}(r)}{\hat{\lambda}_T + \hat{\lambda}_B}$ (Lotwick and Silverman, 1982).

[16] Note that our test is conditional on observed locations and thus valid despite the potential inhomogeneity of λ (density of restaurants) across a city. However, as it is not conditional on spatial covariates, we are unable to differentiate between co-location due to true dependence among restaurants and that due to exogenous features of city territory. We can therefore discuss only unconditional or observed dependence. In our experience, analyses based on spatial covariates are very sensitive to assumptions about the nature of interactions and the set of covariates included; therefore, this paper does not pursue this line of research.

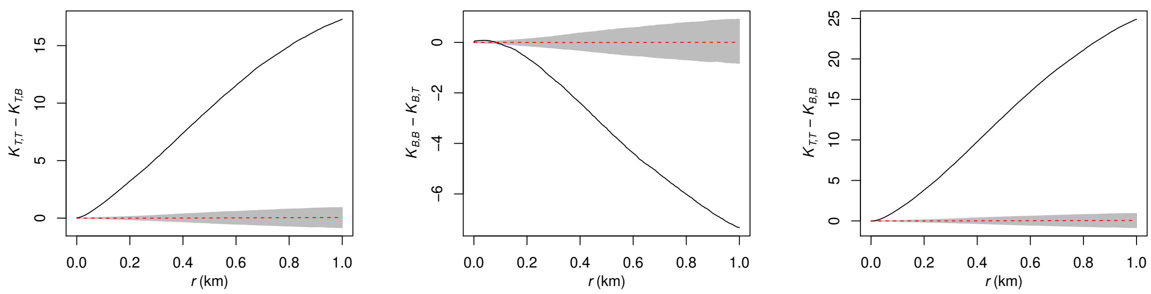
[17] Recall that, for result reliability, we impose an upper limit on the value of r for which K is calculated, so that for distances above 1 km, smaller cities start to fall out of the sample.



(a) (b) (c)

Figure 4: Spatial dependence between restaurants in England and Wales. Gray: observed deviation from random labeling is not significant (ns); black: observed deviation is significant and positive (+); white: observed deviation is significant and negative (-).

with pointwise 5% critical envelopes corresponding to the null hypothesis of random labeling. We first use the ratio-of-sums estimator and calculate each pooled $K_{i,j}$ as a weighted mean of individual $K_{i,j}$ estimates for the 96 cities with weights proportional to $n_i n_j$.[18] In the next step, we calculate the relevant pairwise difference using the pooled $K_{i,j}$ estimates from the previous step. To obtain the critical envelopes, we randomly relabel the restaurants in each city 999 times and calculate a relevant pairwise difference between the pooled $K_{i,j}$ functions each time. Out of the 999 simulated pairwise differences, we select the upper and lower 2.5% quantiles. Our conclusion is that, on average, top restaurants in England and Wales concentrate, while bottom restaurants disperse (except for distances below 0.1 km). Top restaurants also cluster more.



(a) (b) (c)

Figure 5: Pooled estimates for 96 cities in England and Wales. Solid line: pooled estimate of the pairwise difference; dashed line: sample mean of simulations from the null hypothesis of random labeling; gray shading: pointwise 5% critical envelopes based on 999 simulations.

[18]Observe that the estimator in (2) can be expressed as the ratio $\hat{K}(r) = \hat{Y}(r)/\hat{X}(r)$, where the denominator $\hat{X}(r)$ is the number of pairs $n_i n_j$.

3 Theoretical Analysis

In this section, we present a minimal model that can explain our stylized facts. To address the relative dispersion of top and bottom restaurants, we need at least 2 firms of each quality type.

3.1 Model

The model is based on Hotelling's paradigm of a linear city. Four firms (restaurants) sell goods that are vertically differentiated by quality in a city that is defined as a unit interval $[0, 1]$. Consumers are distributed on this interval according to the log-concave density $f(y) \geq 0$, $y \in [0, 1]$, with their mass normalized to one. Each consumer buys a unit of the good provided by the restaurants. The indirect utility of a consumer located at y and purchasing from restaurant $j = 1, ..., 4$ is

$$u_j(y) = q_j - p_j - (y - x_j)^2 + \sigma\varepsilon_{j,y}, \tag{3}$$

where $q_j > 0$ is the quality of the good,[19] $p_j > 0$ is its price, $x_j \in [0, 1]$ is the location of restaurant j, and $\varepsilon_{j,y}$ is an idiosyncratic preference shock. The latter describes the fact that consumers differ in some dimensions that the restaurants cannot observe and, therefore, they also cannot predict with certainty the decision of any particular consumer. The scale parameter σ controls the degree of consumer taste heterogeneity. For $\sigma > 0$, there is always a positive probability that any consumer will buy from any restaurant. The larger σ is, the less consumers discriminate between restaurants and the more random their buying decisions appear. The quadratic term captures the disutility incurred by consumers due to transport costs.[20]

Assuming that idiosyncratic shocks are independently and identically distributed according to a type I extreme value distribution (the logit model), the probability that a consumer located at y buys from restaurant j is

$$P_j(y; \mathbf{p}, \mathbf{x}) = \frac{\exp\{(q_j - p_j - (y - x_j)^2)/\sigma\}}{\sum_{k=1}^{4} \exp\{(q_k - p_k - (y - x_k)^2)/\sigma\}}, \tag{4}$$

where $\mathbf{p}$ denotes the price vector (p_1, p_2, p_3, p_4) and $\mathbf{x}$ the vector of restaurant location choices (x_1, x_2, x_3, x_4).

Restaurants play a two-stage game. In the first stage, they select their location, whereas, in the second one, they select the price they charge consumers. We assume that the quality of

[19] In our model, the quality of goods determines the utility consumers get from purchasing them. Such a general concept of quality is better captured by restaurant rankings based on consumers' ratings (e.g., TripAdvisor) than rankings based on experts' opinions (e.g., Michelin, Zagat), whose exquisite taste is not often perceptible to the common eye (see O'Loughlin, Marina, "Is Zagat's guide to London restaurants on another planet?", *The Guardian*, accessed July 4, 2020, https://www.theguardian.com/lifeandstyle/2012/sep/10/zagat-guide-london-restaurants.)

[20] These costs may include the value of the time spent in travel.

restaurants is given and fixed.[21,22] The marginal costs of production are normalized to 0. The profit of restaurant j, which is a continuous function of prices and location choices, is given by

$$\Pi_j(\mathbf{p}, \mathbf{x}) = p_j \int_0^1 P_j(y; \mathbf{p}, \mathbf{x}) f(y)\, \mathrm{d}y \tag{5}$$

and is bounded below by zero.

3.2 Computation of equilibria

We search for subgame perfect Nash equilibria in pure strategies. To determine them, we proceed with backward induction. We rely on a numerical algorithm consisting of two loops: the inner one calculates the price equilibrium for any given set of firms' locations, whereas the outer one calculates all the location equilibria.

3.2.1 Price equilibrium

The inner loop takes a vector of given restaurant locations $\bar{\mathbf{x}}$ and calculates the price equilibrium as a solution to the following system of four nonlinear equations (firms' first-order conditions for profit maximization):

$$\frac{\partial \Pi_j(\mathbf{p}, \bar{\mathbf{x}})}{\partial p_j} = 0, \quad j = 1, ..., 4. \tag{6}$$

The equations above involve derivatives of integral functions that must be evaluated numerically. We approximate these derivatives using centered finite difference schemes and compute the integrals using MATLAB's global adaptive quadrature method. The solution to (6) is then obtained using the Artelys Knitro 12.0 solver.

Based on theoretical results from Caplin and Nalebuff (1991), the price equilibrium can be shown to exist and be unique. The key argument for the existence of the equilibrium consists of showing that a firm's profit $\Pi_j(\mathbf{p}, \mathbf{x})$ is quasi-concave in its own price, which is the case as $P_j^{-1}(y; \mathbf{p}, \mathbf{x})$ can be shown to be convex in the firm's own price. The uniqueness of the price equilibrium follows from the log-concavity of the density of consumers $f(y)$. Note that the uniqueness result was previously obtained by Anderson and de Palma (1988) for the duopoly model but the general argument of Caplin and Nalebuff (1991) extends it to the case of more firms. The uniqueness result is particularly important to us because it ensures that all firms have a unique best response and that no other price equilibrium exists. Caplin and Nalebuff (1991)

[21] The assumption of exogenous quality can be interpreted as assuming that high quality restaurants do not offer low quality services, whereas low quality restaurants do not have a capacity to offer high quality services. In this sense, quality is related to a long-run decision. The owner's business expertise can take a lifetime to build and is often specialized and directed to a certain type of business and consumers. We are also not primarily interested in quality choices, but rather in location differences between restaurants of different quality levels. As the multiple-firm model with the multiplicity of equilibria is computationally demanding even with the assumption of exogenous quality, we necessarily leave the possibility of endogenous quality, interesting in its own right, for future work.

[22] Irmen and Thisse (1998) consider a model with strategic location and quality. Their model, in which two firms compete in a multi-characteristic space, leads to maximum differentiation in one dimension and minimum differentiation in all others.

also show that the price equilibrium game is log-supermodular, which implies that the price equilibrium vector is globally stable under many learning and adjustment processes. This makes our algorithm suitable for finding the numerical solution in the inner loop.

3.2.2 Location equilibria

The outer loop solves the system of first-order necessary conditions for a location equilibrium

$$\frac{\partial \Pi_j(\mathbf{p}^*(\mathbf{x}), \mathbf{x})}{\partial x_j} = 0, \quad j = 1, ..., 4, \tag{7}$$

where $\mathbf{p}^*$ is a vector of equilibrium prices for a given set of locations $\mathbf{x}$. The Knitro solver (*knitromatlab_lsqnonlin*) starts at some initial point $\mathbf{x_0}$ and, using the interior-point algorithm, attempts to find the minimum of the sum of squares of the functions that appear on the left-hand side of equation (7), subject to the bounds $0 \leq x_j \leq 1, \forall j = 1, ..., 4$. During the search for the solution $\mathbf{x}^*$ to system (7), whenever any profit function needs to be evaluated, the outer loop calls the inner loop to first calculate equilibrium prices for a given set of locations.

Usually, there are many candidates for an equilibrium satisfying the first-order optimality conditions (7). To find all of them, we restart the algorithm from $1,000$ initial locations, obtained by randomly drawing components of $\mathbf{x_0}$ from the interval $[0, 1]$. This process gives us a set of solutions to the firms' first-order conditions, which are only necessary but not sufficient conditions for an interior equilibrium. In the second step, we then eliminate candidates that fail to satisfy second-order conditions for a local maximum. Finally, in the third step, we check that each remaining candidate actually satisfies global optimality. For this, we check that firm j's location choice is a global maximum given the location of all the other firms. In other words, for each equilibrium candidate, we look for the location $x_j \in [0, 1]$ of firm j that maximizes its profit function while the locations of other firms are kept fixed to their equilibrium value. To do this, we restart the Knitro optimizer from 100 initial points to obtain the best feasible solution for firm j on the interval. We repeat the exercise for all $j = 1, ..., 4$ and keep an equilibrium candidate only if it is a global maximum for all firms. In general, we obtain several location equilibria.

3.2.3 Stability

When dealing with multiple location equilibria, it is useful to distinguish stable from unstable equilibria. In the literature, the latter are usually disregarded as they are considered not to be sustainable in practice. This is because a small perturbation around an unstable equilibrium will grow in size rather than diminish over time. When defining local stability, we follow Dixit (1986) and assume that, for a given location vector $\mathbf{x}$, each firm adjusts its location over time at a rate proportional to the marginal profitability of the adjustment,

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = s_j \frac{\partial \Pi_j(\mathbf{p}^*(\mathbf{x}), \mathbf{x})}{\partial x_j}, \tag{8}$$

where $s_j > 0$ is the adjustment speed of restaurant j, $j = 1, ..., 4$. By performing a linear approximation around the equilibrium $\mathbf{x}^*$, a system of perturbation equations is obtained. Following the stability conditions in Dixit (1986), we require that all eigenvalues of the corresponding 4×4 matrix $\mathbf{J}_{ij} = \partial^2 \Pi_i(\mathbf{p}^*, \mathbf{x}^*)/\partial x_i \partial x_j$ for $i, j = 1, ..., 4$, have negative real parts.

3.3 Simulation results

In our baseline specification, we consider a uniform density of consumers ($f(y) = 1$ for $y \in [0, 1]$, and 0 otherwise) and assume that firms provide goods of the same quality ($q_j = 1,\ \forall j = 1, ..., 4$). For a relatively high level of consumer taste heterogeneity (i.e., $\sigma = 1$), we observe that, in equilibrium, all firms locate in the center (see Figure 6a). Lower values of taste heterogeneity σ induce multiple equilibria[23]; however, only a single equilibrium satisfies the stability condition. For $\sigma = 0.1$, we find 51 equilibria, most of which differ only in the permutation of firm names. Among them, we identify 5 distinct types of equilibria, shown in Figure 6b, of which only the last one—type 5, with two firms in the center and the other two around it—is stable.[24]

When consumer taste heterogeneity σ decreases, the consumer's choice of a restaurant is guided more by the restaurants' objective characteristics and less by his or her idiosyncratic taste. This increases the competition among the restaurants. For $\sigma = 0.08$, the central agglomeration equilibrium (type 1) disappears and types 2–5 remain. In turn, for $\sigma = 0.06$, the equilibrium with 3 firms at the same place (type 2) also disappears and only types 3–5 remain. Then, for $\sigma = 0.05$, the equilibria with two firms at the same place but outside the center (types 3 and 4) disappear, so only equilibrium type 5 remains. At the same time, a new type of symmetric equilibrium (type 6) appears, where firms non-uniformly spread out throughout the city. Of types 5 and 6, only equilibrium type 6 is stable (see Figure 6c). This stable equilibrium, with firms spread out throughout the city, remains the only equilibrium when sigma approaches zero (i.e., $\sigma = 0.01$ in Figure 6d).[25] As in de Palma et al. (1987), a lower level of consumer taste heterogeneity encourages the spatial dispersion of restaurants.

3.3.1 Equilibria with quality differentiation

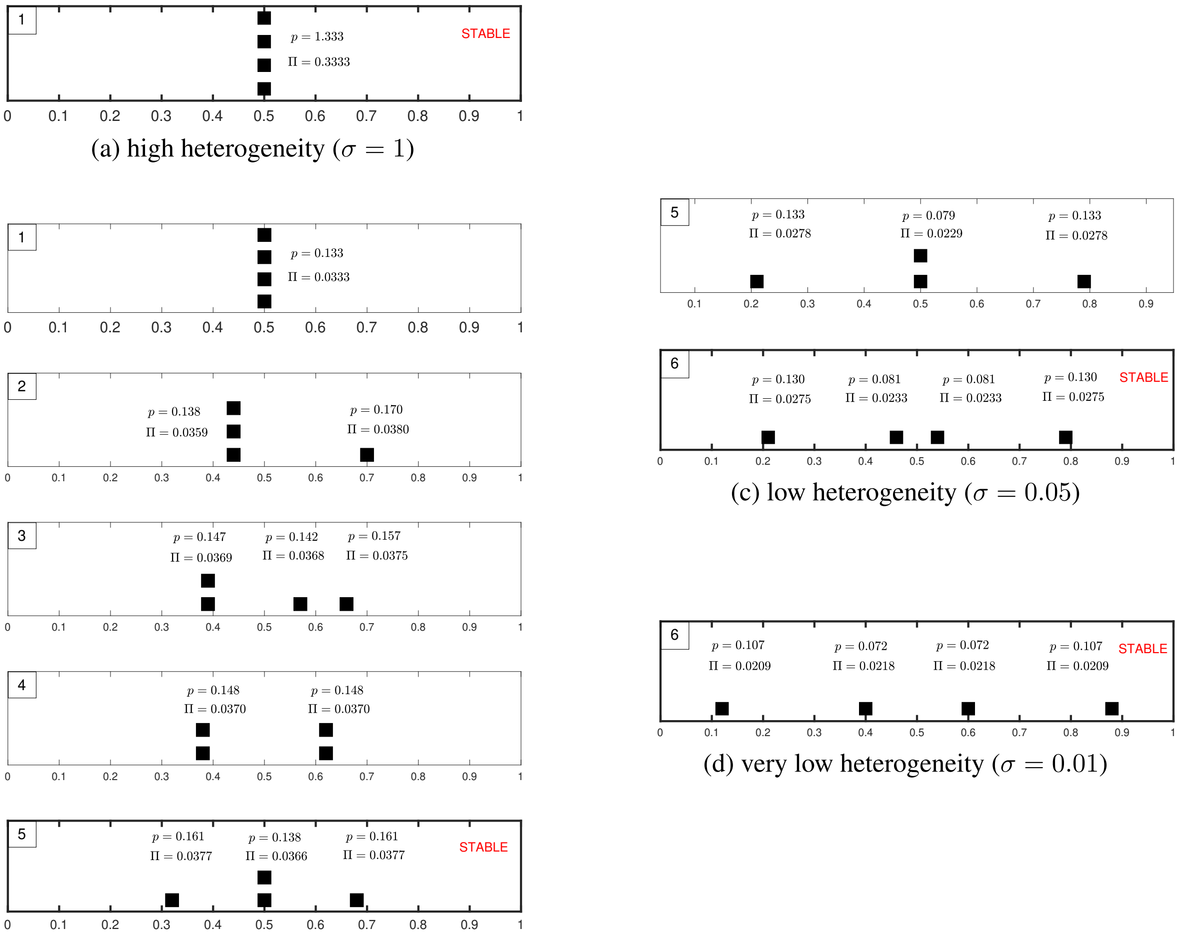
To explore the effects of quality differentiation on the location of the restaurants, we consider a model where 2 of the restaurants provide a high-quality good and the other 2 provide a low-quality good. For convenience, we term the high-quality (top) restaurants firms 1 and 2, and the low-quality (bottom) restaurants firms 3 and 4.

Solving the model over a wide range of parameter values, we observe that when the value of consumer taste heterogeneity σ is high, a sufficiently large quality difference between top and bottom firms is needed to relocate some firms away from the center. Moreover, intermediate values of the consumer taste heterogeneity σ are needed to obtain (stable) equilibria in which top firms are more spatially concentrated than bottom ones.

[23] This observation is similar to a result in de Palma et al. (1987), who consider symmetric equilibria in a 3-firm logit model.

[24] Equilibrium types 2 and 3 in Figure 6b also have their mirror counterparts $1 - \mathbf{x}$. Throughout the text, we classify any such mirror equilibria as one single type.

[25] It is interesting to observe that equilibrium locations $(0.124, 0.396, 0.604, 0.876)$, which we obtain for $\sigma = 0.01$, are identical, to three decimal places, to those in Brenner (2005), who considers a n-firm Hotelling model with homogenous consumers.



(a) high heterogeneity ($\sigma = 1$)

(b) intermediate heterogeneity ($\sigma = 0.1$)

(c) low heterogeneity ($\sigma = 0.05$)

(d) very low heterogeneity ($\sigma = 0.01$)

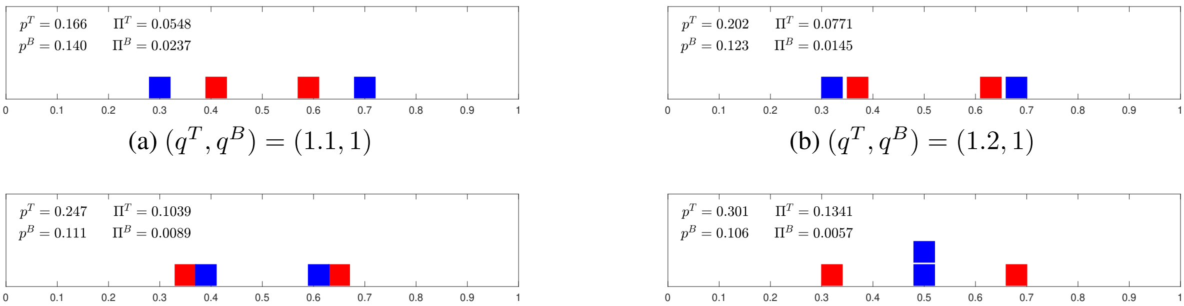
Figure 6: Location equilibria for different levels of consumer taste heterogeneity (σ), when firms provide goods of the same quality ($q_1 = q_2 = q_3 = q_4 = 1$).

When the value of consumer taste heterogeneity is high (for instance, $\sigma = 0.2$) and the quality difference is small—e.g., $(q_1, q_2, q_3, q_4) = (1.1, 1.1, 1, 1)$, which we shorten to $(q^T, q^B) = (1.1, 1)$—in equilibrium, all firms locate in the center. For a larger quality difference, we obtain a different equilibrium in which bottom firms locate in the center, whereas top firms locate around them—for $(q^T, q^B) = (2, 1)$, the equilibrium locations are $(0.214, 0.786, 0.5, 0.5)$, with prices $(0.584, 0.584, 0.205, 0.205)$ and profits $(0.2776, 0.2776, 0.0051, 0.0051)$, respectively. For an even larger quality difference, the top firms move even farther away from the center (and from each other).

For intermediate values of the consumer taste heterogeneity σ, we are able to find top firms relatively closer to the center, as long as the quality difference between top and bottom firms is not too large. For instance, consider $\sigma = 0.1$ with a small quality difference, namely, $(q^T, q^B) = (1.1, 1)$. There are now four types of equilibria (compared to Figure 6b): i) top firms together at one place and bottom firms together at another place, $(0.546, 0.546, 0.312, 0.312)$; ii) top firms in the center, $(0.5, 0.5, 0.256, 0.744)$, iii) bottom firms in the center, $(0.370, 0.630, 0.5, 0.5)$; and, iv) all firms spread out, with top firms closer to the center, $(0.406, 0.594, 0.301, 0.699)$. However, only the last equilibrium is stable. Note that, for a larger quality difference, namely,

$(q^T, q^B) = (1.2, 1)$, the first equilibrium above disappears. There are three types of equilibria left, of which only the last one is stable: $(0.5, 0.5, 0.208, 0.792)$, $(0.369, 0.631, 0.5, 0.5)$, and $(0.37, 0.63, 0.317, 0.683)$. For an even larger quality difference, namely, $(q^T, q^B) = (1.3, 1)$, the equilibrium with the top firms in the center disappears and the following two equilibria remain: the unstable one with the bottom firms in the center $(0.347, 0.653, 0.5, 0.5)$ and the stable one with all the firms spread out $(0.346, 0.654, 0.386, 0.614)$. Finally, for a very large quality difference, namely, $(q^T, q^B) = (1.4, 1)$, the only equilibrium left, which is stable, shows that bottom firms locate in the center, $(0.320, 0.680, 0.5, 0.5)$.

Top restaurants' higher (relative) quality increases their market power but also the competition among them. This drives top restaurants to locate farther from one another. This is especially visible when we compare the stable equilibria at different levels of quality differences in Figure 7. When the quality difference is small, $(q^T, q^B) = (1.1, 1)$, the top firms locate close to the center as well as to each other, whereas the bottom firms locate around them (Figure 7a). As the relative quality of the top firms increases (Figure 7b–7c), the top firms start moving apart, whereas the bottom ones are increasingly attracted to the center. When the quality difference is large, namely, $(q^T, q^B) = (1.4, 1)$, in Figure 7d, the initial situation is completely reversed—it is now the top firms that locate around the bottom firms, both of which are in the center. In all these equilibria, the top firms charge higher prices and realize larger profits than the bottom firms (compare p^T and Π^T with p^B and Π^B, respectively, in Figure 7).



(a) $(q^T, q^B) = (1.1, 1)$

(b) $(q^T, q^B) = (1.2, 1)$

(c) $(q^T, q^B) = (1.3, 1)$

(d) $(q^T, q^B) = (1.4, 1)$

Figure 7: Stable equilibria for different levels of quality difference between top (red) and bottom (blue) restaurants; $\sigma = 0.1$.

Discussion. We can interpret the results of stable equilibria for different levels of consumer taste heterogeneity as follows. When consumer taste heterogeneity is relatively high, buying decisions appear rather random to restaurants. Consequently, the city center offers relatively little advantage in terms of access to the customer base. As competition is fiercer between top restaurants than bottom ones, top restaurants disperse more. However, when consumer taste heterogeneity is relatively low, transport costs play a prominent role in buying decisions. The city center attracts top restaurants, as it provides them the best access to the customer base. Top restaurants are able to drive away bottom restaurants in the center through quality competition. The latter then disperse around top restaurants, primarily serving customers living outside the city center. When the quality difference between restaurant groups increases, it results in fiercer competition between top restaurants, which forces them to move apart.

Here, we see that our model is able to produce stable equilibria in which top restaurants are concentrated, whereas bottom ones are dispersed, thereby establishing the link with our stylized facts. Further, our model predicts that top restaurants become relatively more dispersed when the quality difference between top and bottom restaurants increases.

The analysis above assumes a uniform distribution of consumers. In reality, city centers attract businesses due to the high concentration of consumers and, therefore, the high demand available at these locations. Higher consumer densities in city centers are the rule rather than the exception.[26] Therefore, in the next section, we introduce a non-uniform density of consumers.

3.3.2 Equilibria with a non-uniform density of consumers

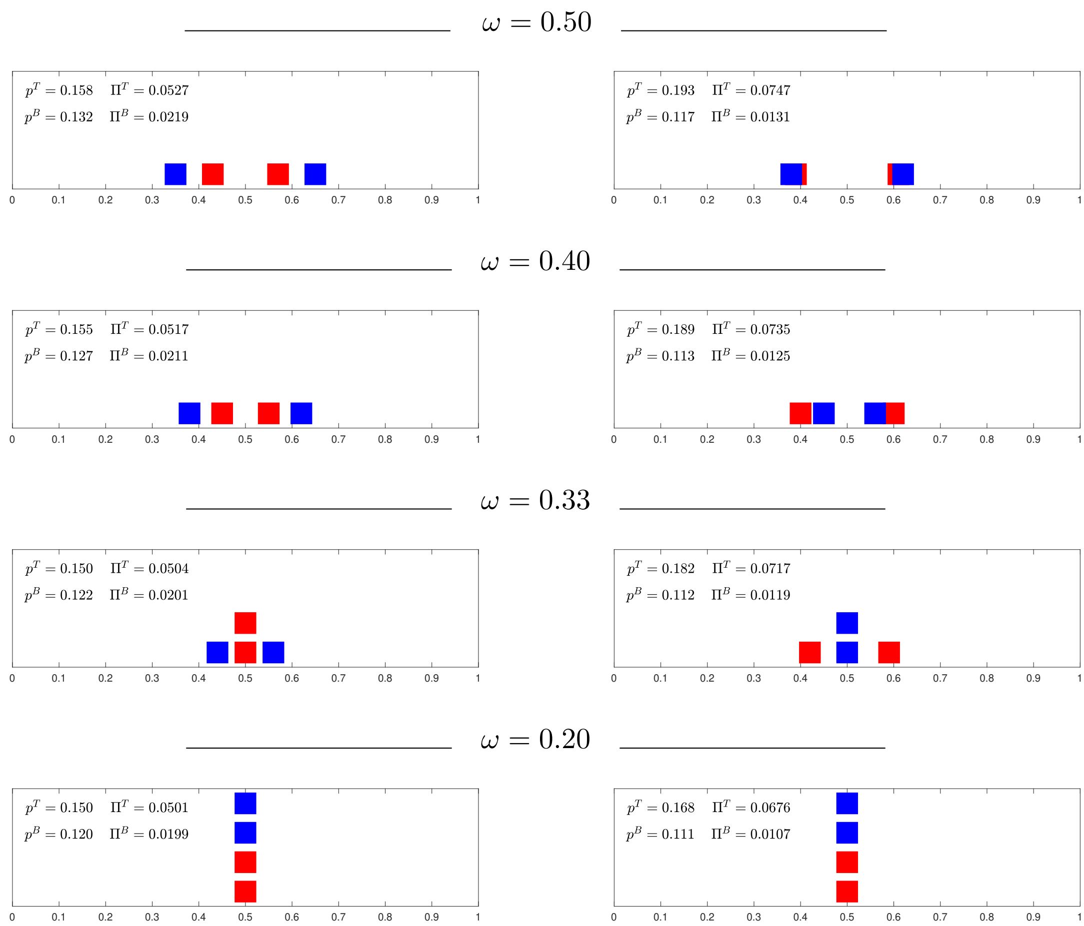
We assume that the density of consumers $f(y)$, $y \in [0, 1]$, is given by a normal distribution truncated over the city interval $[0, 1]$. For its mean, we take the city center ($\mu = 0.5$). We then vary the standard deviation of the parent normal distribution (ω) to see how the degree of consumer concentration in the city center affects the (relative) location of restaurants.

When the values of the other parameters are kept fixed, a relatively larger mass of consumers in the center (lower ω) calls for a larger quality difference to create equilibria in which any of the firms locate away from the city center. It also induces firms in any non-agglomeration equilibrium to locate closer to the center than they would if ω was higher. These results were expected. More interesting is the observation that the magnitude of the quality difference between the top and bottom restaurants plays a crucial role in how consumer density influences their relative locations.

In Figure 8, we illustrate this for $\sigma = 0.1$. We focus on stable equilibria. In the first column, the quality difference is small, $(q^T, q^B) = (1.1, 1)$. When the relative mass of consumers in the center increases (ω falls), the top firms move closer to the center. Although this leads to a lower price due to the increased competition, the market size effect dominates. Although bottom firms are also attracted to the center, they remain separated and farther from the center than the top ones, except in the case of a very densely populated center ($\omega = 0.2$), where all the firms locate in the center.

The situation is very different when the quality difference is larger. In the second column of Figure 8, $(q^T, q^B) = (1.2, 1)$. A lower value of ω attracts both types of restaurants to the center (compared to the case of the uniform distribution of consumers in the first row of Figure 7) but the effect is stronger for bottom restaurants. When $\omega = 0.4$, the latter are already closer to the center than top restaurants. When the quality difference is large, the bottom restaurants are inferior rivals and competition exists mostly between the top restaurants. This increased competition between the top restaurants makes them relatively less willing to locate close to each other. Specifically, the closer top restaurants are to each other, the more consumers' choices are guided by prices, which drives both prices and profits down.

[26] Our data confirm this for population densities. See footnote 27.



(a) $(q^T, q^B) = (1.1, 1)$ (b) $(q^T, q^B) = (1.2, 1)$

Figure 8: Stable equilibria for varying values of consumer dispersion (ω). The quality difference is small in the first column and large in the second one. In all the plots, $\sigma = 0.1$

4 Regression Analysis

The analysis of our theoretical model in the previous section allows us to develop empirically testable behavioral predictions as indicated below.

Behavioral Predictions:

1. An increase in the population density in the city center reduces the spatial dispersion of both top and bottom restaurants.

2. When the quality difference between top and bottom restaurants increases,

   (a) the absolute spatial dispersion among top restaurants increases,

   (b) the relative spatial dispersion between top and bottom restaurants increases, and

   (c) the absolute spatial dispersion of bottom restaurants decreases.

To test these predictions, we construct a dataset by combining the city-level spatial statistics reported in Section 2 with the data on different city characteristics obtained from the Office for National Statistics. Table 2 presents the summary statistics for the key variables.

4.1 Description of variables

Geographical dispersion of restaurants. Measures of restaurants' geographical dispersion are our dependent variables. Here, we use $TdistC$ and $BdistC$, which are the average distances of "top" and 'bottom" restaurants, respectively, from the city center (see Section 2.2.2). Based on Box-Cox tests, we take the logarithmic transformation of these variables to form $\ln TdistC$ and $\ln BdistC$, respectively. To test Behavioral Prediction 2(b) concerning the relative dispersion of the restaurants, we use $\ln TBdistC = \ln TdistC - \ln BdistC$.

Population dispersion. The variable PD corresponds to the population density in the city center relative to that in the surrounding area. Specifically, we measure this non-parametrically using $PD_{X_1-X_2} = den(X_1) - den(X_2)$, for $X_1 < X_2$, where $den(X)$ is the population density within an X-kilometer radius of the city center, expressed as thousands of people per km^2. To obtain an estimate of the total population living within the circle of the X-kilometer radius, we count the number of people living within the OAs whose centroids lie inside the circle. We report the results for $PD_{2-4} = den(2) - den(4)$.[27,28]

Restaurant quality dispersion. We use the *overall rating* of each restaurant to construct quality dispersion measures that are comparable across different cities.[29] We compute two alternative measures. The first one is σ_Q, which is the standard deviation of the overall rating for *all* the restaurants in the city. Our theory in Section 3 relies on two types of restaurants. However, in reality, when deciding where to locate, a new restaurant would take into account the location of all types of restaurants. As a robustness check, we also use dQ, which is defined as the difference between the average overall rating of the top and bottom quartiles of restaurants in the city. This measure is more closely related to our theoretical model, but potentially ignores the effect of restaurants of intermediate quality.

Control variables. Our theoretical model is based on a city whose geographical size is normalized. Since two of our dependent variables, *TdistC* and *BdistC*, represent physical distances in kilometers, it is possible that these measures reflect, to some extent, the effect of the geographical size of cities. Similarly, it is possible that the average distance of restaurants is influenced by tourist activities and the total number of restaurants. To address such concerns, in a robustness check, we use the logarithm of city area ($\ln Area$), the logarithm of the number of restaurants ($\ln N_R$), and the dummy variable $Tourist$ as our control variables. The variable $Tourist$ takes the value 1 if, between 1999 and 2018, the city was mentioned at least once on

[27] We observe that PD_{2-4} is positive for 90 out of 96 cities in our sample.

[28] For robustness, we also tried alternative sets of distances, such as $PD_{1-3} = den(1) - den(3)$. The results are qualitatively similar and available upon request.

[29] *Overall rating* is a raw score based on reviewers' overall assessment of the restaurant, which ranges from 1 (terrible) to 5 (excellent). As such, it captures the cardinal dimension of restaurant quality, whereas TripAdvisor's popularity ranking is an ordinal measure that takes into account overall ratings as well as other information, such as the recency and number of reviews. TripAdvisor recommends the use of the popularity ranking for its recency and consistency of information instead of, for example, the overall rating. However, the dispersion of ordinal measures, such as popularity rankings, is not informative.

the list of 20 most visited towns and cities by overseas visitors according to the "inbound town data" released by the Office for National Statistics. Otherwise, it takes the value 0.

4.2 Empirical Results

To empirically test our behavioral predictions, we run cross-sectional regressions of the restaurants' geographical dispersion measures on the dispersion of the population and restaurant quality. Table 3 reports the results based on ordinary least squares.

The estimated coefficients for PD_{2-4} in columns (1)–(4) and (9)–(12) are all negative and significant. This is in line with Prediction 1—the geographical dispersion of both types of restaurants is smaller when the population density in the city center is relatively larger. Note that our theoretical model does not give clear predictions about the relative geographical dispersion of top and bottom restaurants. In columns (5)–(8), the coefficients for PD_{2-4} are also negative, although not always significant. Hence, the effect of the population density in the city center appears stronger for top restaurants than for bottom ones.

In columns (1)–(4), σ_Q and dQ are all positive and significant. This finding supports Prediction 2(a) about the effect of quality difference on the absolute dispersion of top restaurants. A larger quality difference between top and bottom restaurants increases the distance between top restaurants.

The positive and significant coefficients for σ_Q and dQ in columns (5)–(8) present some evidence supporting Prediction 2(b) concerning the effect of quality difference on the relative dispersion of the two groups of restaurants. However, Prediction 2(c) about bottom restaurants is not adequately supported by our data. The only negative point estimate is nonsignificant, whereas others are positive and all, except one, are nonsignificant. Here, it emerges that our stylized logit Hotelling model does not work well.

The effects of the control variables can be found in the even-numbered columns. As anticipated, the geographical dispersion of both restaurant types is larger in cities with larger areas and more restaurants. A city's being an important tourist destination reduces the dispersion of both restaurant types in it. This is in line with the fact that city centers usually exhibit the highest concentration of visitor-related facilities and services. Importantly, the inclusion of various control variables does not qualitatively affect our key findings.

To provide a further robustness check, we also redo our analysis using spatial regression methods. In the first set of analyses, we model spatial dependence through the error terms and estimate the models using the generalized spatial two-stage least squares. Following the spatial regression literature, we assume that spatial effects decay with distance and use the inverse distance weighting matrix, where the distance between cities is measured as the distance between their city centers.[30] The results are reported in Table 4, which shows that our key results are unaffected. Using the same weighting matrix, we also perform a spatial regression analysis

[30] The estimated model is $\mathbf{y} = \sum_{k=1}^{K} \beta_k \mathbf{x}_k + \mathbf{u}$, $\mathbf{u} = \rho \mathbf{W} \mathbf{u} + \epsilon$, where $\mathbf{y}$ is a vector of observations of the dependent variable, $\mathbf{x}_k$ a vector of observations of the k-th covariate and β_k is the corresponding regression parameter. The errors $\mathbf{u}$ are spatially autoregressive. $\mathbf{W}$ is a spatial weighting matrix, ρ is the spatial autoregressive coefficient, and ϵ is a vector of innovations that are assumed to be independent but heteroskedastically distributed, with heteroskedasticity being of unknown form. This model addresses the possibility that unobserved variables affect nearby cities.

Table 2: Summary statistics for variables used in the empirical analysis ($N = 96$)

	mean	sd	min	max	Remarks
$TdistC$	1.760	0.845	0.348	4.347	Average distance of top restaurants from the city center (in km)
$BdistC$	1.890	0.798	0.620	4.970	Average distance of bottom restaurants from the city center (in km)
PD_{2-4}	1.315	0.924	-0.789	5.152	Dispersion of population density in the city (Population density within a two-km radius of the city center minus population density within a four-km radius)
σ_Q	0.770	0.083	0.601	1.032	Standard deviation of TripAdvisor's "overall rating" for all restaurants in the city
dQ	1.346	0.200	0.861	2.053	Difference in TripAdvisor's average "overall rating" between the top and bottom groups
$Area$	207.911	340.746	10.635	2224.189	Area of the city (in km^2)
N_R	330.365	280.912	74.000	1498.000	The number of restaurants in the city (according to TripAdvisor)
$Tourist$	0.219	0.416	0	1	The dummy variable takes the value 1 if the city appears at least once on the list of 20 most visited cities between 1999 and 2018, and 0 otherwise.

based on a model that allows for spatial dependence through both the dependent variable and the error terms.[31] We present the estimates in Table 5. Again, we find our results remain unchanged under this alternative estimation strategy.

5 Concluding remarks

This study explores the location of restaurants, grouped into top- and bottom-rated restaurants based on TripAdvisor's ratings, across large cities in England and Wales. Using a unique dataset, we find that in many cities in the sample, there exist significant differences in the way top-ranked and bottom-ranked restaurants locate. Top restaurants tend to be less spatially dispersed throughout the city and locate relatively closer to the city center than bottom ones. We also explore the spatial dependence between restaurant types and find that in the majority of cities, top restaurants locate closer to their own type and are, as a group, more clustered than bottom restaurants. For bottom restaurants, meaningful patterns are comparably less frequent but when they occur, they involve bottom restaurants usually locating away from their own type and closer to top restaurants.

To explore location incentives theoretically, we extend the standard logit Hotelling model to 4 firms by introducing both vertical differentiation of goods and a non-uniform density of consumers. We develop a numerical algorithm to find both symmetric and asymmetric equilibria and distinguish stable from unstable ones. Our theoretical model is able to produce stable equilibria in which top restaurants are concentrated, whereas bottom ones are dispersed. We find a relatively higher concentration of top restaurants when consumer taste heterogeneity is sufficiently low and the quality difference between both restaurant groups is not "too large." A larger quality difference, ceteris paribus, increases competition between top restaurants and induces them to move apart, whereas bottom restaurants are increasingly attracted to the void left in the center.

Our regression analysis shows that cities with a relatively larger population density in the center and those that are important tourist destinations experience a higher concentration of restaurants in the center. A large quality difference leads to the dispersion of top restaurants, as predicted by the model. However, the data does not confirm our prediction according to which a larger quality difference would pull bottom restaurants toward the center. The stylized logit model does not perform well in this case. This means that, in practice, some additional factors are likely to affect location decisions. We see this as an interesting direction for future research.

Several other possible directions for future work emerge following our analysis. In practice, restaurants are also differentiated by the type of food they offer (Italian, vegan, etc.). A more complex model could explore how product characteristics interact with quality differentiation for location decisions, both in theory and practice. Another interesting direction is about the inclusion of the time dimension, especially the entry and exit times of different restaurant types and the evolution of their location over time. Richer data might also enable one to perform a

[31] The estimated model is $\mathbf{y} = \sum_{k=1}^{K} \beta_k \mathbf{x}_k + \lambda \mathbf{W}\mathbf{y} + \mathbf{u}$, $\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \epsilon$, where λ is the spatial autocorrelation parameter corresponding to the spatial lag of the dependent variable $\mathbf{W}\mathbf{y}$ (denoted by "Spatial DV" in Table 5). In addition to the autocorrelation in the error term, the current model also allows for outcomes of nearby cities to be interdependent.

Table 3: Regression analysis of the spatial dispersion of restaurants (OLS)

	$\ln TdistC$				$\ln TBdistC$				$\ln BdistC$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PD_{2-4}	-0.180**	-0.168**	-0.191**	-0.178**	-0.068*	-0.021	-0.077*	-0.028	-0.112**	-0.148**	-0.114**	-0.150**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
σ_Q	1.823**	1.661**			1.216**	1.018**			0.606	0.642^{+}		
	(0.66)	(0.48)			(0.40)	(0.32)			(0.52)	(0.37)		
dQ			0.446^{+}	0.441*			0.457**	0.354**			-0.012	0.087
			(0.24)	(0.19)			(0.17)	(0.13)			(0.19)	(0.14)
$\ln Area$		0.218**		0.221**		0.083^{+}		0.086^{+}		0.135**		0.135**
		(0.05)		(0.06)		(0.05)		(0.05)		(0.04)		(0.04)
$\ln N_R$		0.320**		0.328**		-0.102^{+}		-0.098		0.422**		0.427**
		(0.10)		(0.11)		(0.06)		(0.06)		(0.06)		(0.07)
$Tourist$		-0.421*		-0.434*		-0.233*		-0.232*		-0.188^{+}		-0.202^{+}
		(0.18)		(0.19)		(0.10)		(0.10)		(0.10)		(0.11)
const.	-0.731	-3.339**	0.087	-2.699**	-0.967**	-0.651^{+}	-0.635*	-0.370	0.236	-2.688**	0.722**	-2.329**
	(0.51)	(0.61)	(0.31)	(0.55)	(0.32)	(0.37)	(0.24)	(0.32)	(0.40)	(0.40)	(0.26)	(0.36)
# obs.	96	96	96	96	96	96	96	96	96	96	96	96
Adj. R^2	0.161	0.466	0.110	0.426	0.097	0.295	0.083	0.276	0.065	0.563	0.049	0.547

Notes: Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4: Regression analysis of the spatial dispersion of restaurants (Spatial model I)

	$\ln TdistC$				$\ln TBdistC$				$\ln BdistC$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PD_{2-4}	-0.185**	-0.176**	-0.184**	-0.182**	-0.068*	-0.020	-0.075*	-0.024	-0.107**	-0.144**	-0.105**	-0.148**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
σ_Q	1.962**	1.724**			1.246**	1.012**			0.556	0.707+		
	(0.64)	(0.44)			(0.39)	(0.31)			(0.52)	(0.38)		
dQ			0.398+	0.456*			0.429*	0.324*			-0.038	0.097
			(0.24)	(0.19)			(0.17)	(0.13)			(0.19)	(0.14)
$\ln Area$		0.213**		0.220**		0.083+		0.086+		0.138**		0.136**
		(0.05)		(0.06)		(0.05)		(0.05)		(0.04)		(0.04)
$\ln N_R$		0.313**		0.328**		-0.102+		-0.096+		0.421**		0.425**
		(0.10)		(0.10)		(0.06)		(0.06)		(0.06)		(0.07)
$Tourist$		-0.423*		-0.437*		-0.234*		-0.232*		-0.186*		-0.199+
		(0.17)		(0.18)		(0.10)		(0.10)		(0.09)		(0.10)
const.	-0.829+	-3.301**	0.134	-2.704**	-0.989**	-0.649+	-0.605*	-0.360	0.270	-2.766**	0.744**	-2.348**
	(0.49)	(0.54)	(0.31)	(0.52)	(0.31)	(0.36)	(0.24)	(0.32)	(0.40)	(0.41)	(0.27)	(0.36)
Spatial Error	-0.309	-0.526	0.248	-0.217	0.723	0.174	0.816	0.421	0.501	0.887+	0.610	0.819
	(0.57)	(0.71)	(0.41)	(0.62)	(1.05)	(0.58)	(0.87)	(0.47)	(0.55)	(0.46)	(0.50)	(0.52)
N	96	96	96	96	96	96	96	96	96	96	96	96
Pseudo R^2	0.179	0.493	0.129	0.456	0.116	0.332	0.102	0.314	0.084	0.586	0.069	0.570

Notes: Spatial heteroskedasticity robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 5: Regression analysis of the spatial dispersion of restaurants (Spatial model II)

	$\ln TdistC$				$\ln TBdistC$				$\ln BdistC$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PD_{2-4}	-0.153**	-0.145**	-0.144**	-0.138**	-0.070*	-0.029	-0.081**	-0.035	-0.102*	-0.127**	-0.092*	-0.124**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
σ_Q	1.770**	1.323**			1.127*	0.790*			0.524	0.572		
	(0.66)	(0.49)			(0.45)	(0.35)			(0.54)	(0.40)		
dQ			0.383	0.317^{+}			0.413*	0.258^{+}			-0.054	0.055
			(0.24)	(0.18)			(0.19)	(0.14)			(0.20)	(0.15)
$\ln Area$		0.204**		0.206**		0.067		0.069		0.139**		0.137**
		(0.05)		(0.05)		(0.05)		(0.05)		(0.04)		(0.04)
$\ln N_R$		0.303**		0.313**		-0.138*		-0.133*		0.431**		0.435**
		(0.10)		(0.10)		(0.06)		(0.06)		(0.06)		(0.06)
$Tourist$		-0.434*		-0.457*		-0.214*		-0.215^{+}		-0.199*		-0.215*
		(0.17)		(0.18)		(0.11)		(0.11)		(0.09)		(0.10)
const.	-0.923*	-3.084**	-0.185	-2.624**	-0.795*	-0.065	-0.441	0.184	0.252	-2.891**	0.632*	-2.572**
	(0.46)	(0.52)	(0.30)	(0.45)	(0.40)	(0.44)	(0.29)	(0.36)	(0.40)	(0.40)	(0.29)	(0.36)
Spatial DV	0.444*	0.334*	0.673**	0.492**	1.064^{+}	1.336*	1.335*	1.530**	0.074	0.318	0.239	0.377*
	(0.21)	(0.16)	(0.21)	(0.14)	(0.62)	(0.53)	(0.59)	(0.50)	(0.31)	(0.19)	(0.30)	(0.17)
Spatial Error	-0.802	-0.907	-0.564	-1.044	0.080	-0.872	-0.037	-0.917	0.464	0.656	0.464	0.519
	(0.60)	(0.78)	(0.62)	(0.86)	(1.24)	(0.86)	(1.23)	(0.88)	(0.59)	(0.42)	(0.57)	(0.45)
# obs.	96	96	96	96	96	96	96	96	96	96	96	96
Pseudo R^2	0.193	0.506	0.151	0.483	0.126	0.327	0.109	0.312	0.084	0.581	0.065	0.574

Notes: Spatial heteroskedasticity robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

firm-level analysis and control for spatial covariates, such as the location of shopping centers and tourist attractions, so as to reach conclusions about the conditional spatial dependence between restaurants. The generalization of our results to cities in other countries and a comparison with results for quality-differentiated firms in other industries (e.g., hotels and hairdressing salons) are also yet to be explored.

Appendices

A List of Cities

List of 96 cities in the sample, ordered alphabetically:

Basildon, Basingstoke, Bath, Bedford, Birkenhead, Birmingham, Blackburn, Blackpool, Bolton, Bournemouth, Bradford, Brighton, Bristol, Burnley, Cambridge, Canterbury, Cardiff, Carlisle, Chelmsford, Cheltenham, Chester, Chesterfield, Chichester, Colchester, Coventry, Crawley, Croydon, Darlington, Derby, Doncaster, Dudley, Durham, Eastbourne, Enfield, Exeter, Gloucester, Hereford, Lancaster, Leamington Spa, Leeds, Leicester, Lichfield, Lincoln, Liverpool, Luton, Maidstone, Manchester, Mansfield, Middlesbrough, Milton Keynes, Newcastle upon Tyne, Newport, Northampton, Norwich, Nottingham, Nuneaton, Oldham, Oxford, Peterborough, Plymouth, Poole, Portsmouth, Preston, Reading, Rochdale, Rotherham, Salford, Salisbury, Sheffield, Solihull, Southampton, Southend on Sea, Southport, St Albans, Stevenage, St Helens, Stockport, Stoke-on-Trent, Sunderland, Sutton, Sutton Coldfield, Swansea, Swindon, Telford, Wakefield, Warrington, Watford, Wembley, West Bromwich, Wigan, Winchester, Woking, Wolverhampton, Worcester, Worthing, York.

B Robustness Checks

B.1 Different percentile groups

Table 6 shows that conclusions in Figure 2 are robust to changes in the percentiles used to split restaurants into the top and bottom groups. For instance, even when the median is used to split the restaurants (the last row), the dispersion difference is significant for 42 cities out of the 96 cities. In all but 8 significant cities, bottom restaurants are relatively more dispersed than top ones. Likewise, the difference in the average distance from the city center is significant for 35 cities. Out of the latter, in all but 5 cities, top restaurants are located closer to the center.

B.2 Alternative ranking

As an alternative to TripAdvisor's default Popularity Index, we calculate a simple overall rating, which is based on reviewers' overall assessment of restaurants. The latter has five possible qualitative levels—terrible, poor, average, very good, and excellent—to which we assign the scores from 1 (terrible) to 5 (excellent). We then obtain the overall rating for each restaurant as

Table 6: Spatial dispersion and centricity for different percentiles used to split restaurants into the top and bottom groups

	d^B/d^T		c^B/c^T	
percentiles	> 1	< 1	> 1	< 1
10^{th} – 90^{th}	64 (32)	32 (3)	62 (25)	34 (4)
20^{th} – 80^{th}	63 (35)	33 (4)	60 (32)	36 (3)
30^{th} – 70^{th}	65 (36)	31 (7)	63 (30)	33 (5)
40^{th} – 60^{th}	61 (35)	35 (7)	60 (30)	36 (6)
50^{th} – 50^{th}	62 (34)	34 (8)	62 (30)	34 (5)

Notes: The number of significant cities in parentheses ($n = 96$). All restaurants below the first stated percentile of the ranking are labeled Top and all restaurants at or above the second stated percentile are labeled Bottom.

a weighted average of the scores it received, where each score is weighted by the number of reviewers that submitted it. For accuracy, we only consider restaurants with at least 5 reviews. Unlike the Popularity Index, the overall rating does not favour restaurants with more recent reviews or restaurants with a higher number of them.

When we rank the restaurants using the overall rating, our conclusions are fundamentally the same. For instance, when we use the 25^{th} and the 75^{th} percentile to split the restaurants, like in Figure 2, the ratio d^B/d^T is larger than 1 in 65 cities out of the 96 cities, but the difference in dispersion is significant for only 25 cities. Out of the latter, in all but 4 cities, bottom restaurants are relatively more dispersed than top ones. In 63 cities out of the 96 cities, top restaurants are on average located closer to the city center, but the difference is significant for only 19 cities. Out of the latter, in all but 2 cities, top restaurants are located closer to the center.

B.3 Controling for horizontal differentiation

When we only compare top and bottom restaurants of the same cuisine type (e.g. British, Italian), we obtain similar conclusions about their relative dispersion and centricity as in Figure 2. Likewise, our conclusions about spatial dependence appear quite robust. For instance, Figure 9 differs from Figure 3 in that it considers only restaurants serving British food and still leads to qualitatively identical conclusions.

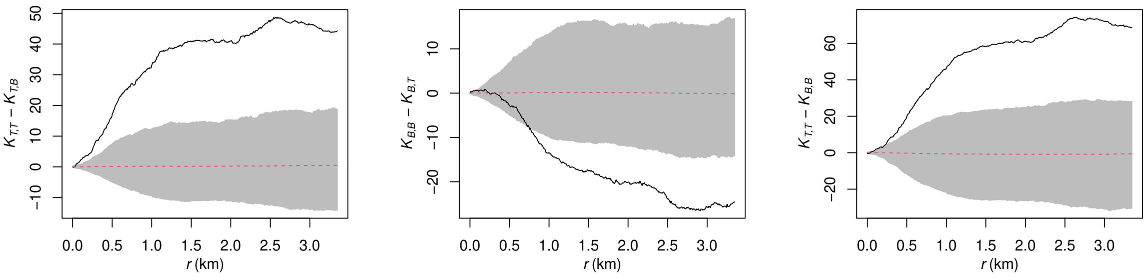


Figure 9: Spatial dependence between British restaurants in Newcastle.

References

ANDERSON, M. J. (2006): "Distance-Based Tests for Homogeneity of Multivariate Dispersions," *Biometrics*, 62, 245–253.

ANDERSON, S. P. AND A. DE PALMA (1988): "Spatial Price Discrimination with Heterogeneous Product," *Review of Economic Studies*, 55, 573–592.

ANDERSON, S. P., J. K. GOEREE, AND R. RAMER (1997): "Location, Location, Location," *Journal of Economic Theory*, 77, 102–127.

BADDELEY, A., E. RUBAK, AND R. TURNER (2015): *Spatial Point Patterns: Methodology and Applications with R*, Chapman and Hall/CRC.

BRENNER, S. (2005): "Hotelling games with three, four, and more players," *Journal of Regional Science*, 45, 851–864.

CAPLIN, A. AND B. NALEBUFF (1991): "Aggregation and imperfect competition: On the existence of equilibrium," *Econometrica*, 59, 25–59.

COUTURE, V. (2016): "Valuing the Consumption Benefits of Urban Density," *Working Paper*.

DE PALMA, A., V. GINSBURGH, Y. PAPAGEORGIOU, AND J.-F. THISSE (1985): "The Principle of Minimum Differentiation Holds under Sufficient Heterogeneity," *Econometrica*, 53, 767–781.

DE PALMA, A., V. GINSBURGH, AND J.-F. THISSE (1987): "On existence of location equilibria in the 3-firm Hotelling problem," *The Journal of Industrial Economics*, 36, 245–252.

DE SILVA, D., C. ELLIOT, AND R. SIMMONS (2016): "Entry, exit and price competition in UK restaurants," *Working Paper, Lancaster University*.

DIXIT, A. (1986): "Comparative statics for oligopoly," *International Economic Review*, 27, 107–122.

DIXON, P. M. (2002): "Ripley's K function," *Encyclopedia of Environmetrics*, 3, 1796–1803.

DURANTON, G. AND H. G. OVERMAN (2005): "Testing for localization using micro-geographic data," *The Review of Economic Studies*, 72, 1077–1106.

ECONOMIDES, N. (1993): "Hotelling's 'Main Street' With More Than Two Competitors," *Journal of Regional Science*, 33, 303–319.

GLAESER, E., J. KOLKO, AND A. SAIZ (2001): "Consumer city," *Journal of Economic

Geography, 1, 27–50.

GORDON, I. R. AND P. MCCANN (2000): "Industrial clusters: complexes, agglomeration and/or social networks?" *Urban studies*, 37, 513–532.

IRMEN, A. AND J.-F. THISSE (1998): "Competition in Multi-characteristics Spaces: Hotelling Was Almost Right," *Journal of Economic Theory*, 78, 76–102.

KRUGMAN, P. R. (1991): *Geography and trade*, MIT press.

LOTWICK, H. AND B. SILVERMAN (1982): "Methods for analysing spatial processes of several types of points," *Journal of the Royal Statistical Society. Series B (Methodological)*, 406–413.

MA, C.-T. A. AND J. BURGESS (1993): "Quality competition, welfare, and regulation," *Journal of Economics*, 58, 153–173.

MARCON, E. AND F. PUECH (2003): "Evaluating the geographic concentration of industries using distance-based methods," *Journal of Economic Geography*, 3, 409–428.

OKSANEN, J., F. G. BLANCHET, M. FRIENDLY, R. KINDT, P. LEGENDRE, D. MCGLINN, P. R. MINCHIN, R. B. O'HARA, G. L. SIMPSON, P. SOLYMOS, M. H. H. STEVENS, E. SZOECS, AND H. WAGNER (2018): *vegan: Community Ecology Package*, R package version 2.5-3.

PORTER, M. E. (2000): "Location, competition, and economic development: Local clusters in a global economy," *Economic development quarterly*, 14, 15–34.

SCHIFF, N. (2015): "Cities and product variety: evidence from restaurants," *Journal of Economic Geography*, 15, 1085–1123.

TSENG, C.-C., W.-J. LIANG, AND K.-C. A. WANG (2010): "Spatial agglomeration with vertical differentiation," *Papers in Regional Science*, 89, 841–858.