

Effect of Agglomeration in the Restaurant Industry

Completed Research

Keertana V Chidambaram

IIT Madras

vc.keertana@gmail.com

Nargis Pervin

IIT Madras

nargisp@iitm.ac.in

Abstract

Although competition can dwindle the sales of a restaurant, it is found that restaurants tend to collocate with their competitors. Despite the associated costs, physical proximity to competitors can lead to positive externalities like increased demand and improved efficiency. We empirically examine the effect of agglomeration in the restaurant industry by understanding how the type and performance of neighboring competitors affect a restaurant. We build a Cox survival model on a dataset of 2,602 restaurants from the city of Phoenix, Arizona obtained from Yelp.

The study reveals three major insights: (1) improvement in neighbor performance (footfall and ratings) was detrimental to a restaurant's survival, (2) collocating with restaurants operating in the same price bracket had a positive influence while neighbors' food cuisine did not make a difference, and (3) franchises did better in neighborhoods crowded with other franchises and vice versa for non-affiliated restaurants.

Keywords

Agglomeration economics, survival model, restaurant failure.

Introduction

The restaurant industry is very important in the USA. The industry contributes about 4% of USA's GDP and restaurants sales are projected to be at \$798.7 billion (4.3% nominal growth) in 2017¹. However, despite the steady growth of the industry, restaurant businesses are very competitive because they are susceptible to high chances of failure. Previous studies have shown that approximately 25% of restaurants fail in their first year of operations and about 60% fail by year three (Parsa et al. 2005).

Despite operating in a very competitive industry, we find that restaurants tend to cluster in specific geographical locations, amid heavy pressure from neighboring restaurant units. This tendency can be largely explained by a powerful concept in economics called the agglomeration effect. Agglomeration refers to the phenomenon that business firms tend to cluster in specific geographical areas ignoring the concomitant competition. Agglomeration can be driven by demand and/or supply. Demand-driven agglomeration theory suggests that firms tend to agglomerate because they aspire to migrate towards a denser customer locality or to a place that is more accessible to their customers (Zhang and Enemark 2016). In case of supply-driven agglomeration, firms cluster broadly because of two strategic reasons (Marshall 1890): heightened demand and production enhancement. Firms can reduce consumer-search cost by spatially collocating, resulting in increased visitation. Knowledge spillover from neighboring competitors can keep a firm up-to-date with the latest technology and firms can also pool inputs like labor and supplies, both of which can improve a firm's productivity.

¹ Available at <https://www.restaurant.org/> (Last access: 20th Feb 2018)

But agglomeration can also have adverse effects due to increased competition. However, researchers and practitioners do not provide consistent evidence explaining the impact of agglomeration in the restaurant industry. Therefore, we posit the following research question

How does the type and performance of the neighboring restaurants impact a restaurant's success?

Herein, we seek to systematically analyze how various characteristics of neighbors contribute to the likelihood of survival or failure of a restaurant using a Cox survival model on a large restaurant dataset (2,602 restaurants) from Yelp. We consider three specific characteristics of neighboring restaurants in this study: (1) performance (defined in terms of its footfall, ratings, and failure history); (2) product offerings (similar price and food cuisine); and (3) affiliation to a franchise.

The study finds some interesting insights to comprehend the agglomeration phenomenon, especially in the restaurant industry. The three important horizons in the agglomerating behavior of the restaurants overall indicate that new establishments in the restaurant industry should be careful in choosing their neighbors and thereby their location. Findings demonstrate that locating in neighborhoods with higher failure rates results in higher odds of failure for a restaurant. A well-performing neighborhood, characterized by the neighboring restaurants' recent online performance in terms of ratings and customer footfall, can prove detrimental to a restaurant's success. Successful franchises like '*Pizza Hut*' gets benefitted if surrounded by other franchises like '*Baskin Robbins*'. However, individual restaurants should not be located near popular franchise restaurants. Restaurants that tried to blend in by serving the popular cuisine in the neighborhood and those that tried to differentiate by serving a distinctive cuisine found these efforts to be fruitless while locating in a cluster with similarly priced restaurants had a positive impact on the restaurant's survival.

The rest of the paper has been structured as follows: next, the theoretical background pertaining to this work has been presented followed by methodology description and model selection. Further, the results obtained from our analysis have been discussed followed by the conclusion and future work recommendations.

Theoretical Background

A thorough literature review unfolded three distinct benefits of agglomeration phenomenon in the restaurant industry (Canina et al. 2005): (1) aggregation offers the customers access to a heterogeneous mix of products in the same geographic area which in turn increases the visits of those locations, (2) restaurants opening in an already established area can benefit from the spillover of the existing location attractiveness, (3) physical proximity to the competitors can help a restaurant to quickly respond to specific competitor moves by closely monitoring them. However, the effect of restaurant agglomeration on its success is not straight-forward. Parsa et al. (2011) showed that failure rate is higher in areas with higher restaurant concentration. On the contrary, Wang et al. (2015) found that failure rate decreases with increase in the number of competitors. Teller et al. (2016) argue that coopetition between retail firms in a cluster albeit complex in nature finally benefits every firm in the cluster. Given the consideration of the anomalies, we persuade to empirically understand the agglomeration phenomenon in the restaurant industry by analyzing types and performance of the neighboring restaurants in the cluster. Following characteristics of the neighborhood have been considered to conceptualize the survival of restaurants.

Effect of performance of neighbors

In the hospitality industry literature, it has been demonstrated that there is a positive spill-over of increased demand for medium and low rated hotels when they are located near high-rated hotels (Tsang and Yip 2009; Luo and Yang 2016). We envisage that a similar relationship could be realized in the restaurant industry and define the neighbor performance in terms of their recent ratings and footfall count. It is worthy to note that an increasing customer crowd for the neighboring competitors could give the restaurant more visibility and increase the attractiveness of the location but at the cost of losing the potential customers to its competitors. We also look at the performance history of the restaurants in the neighborhood to understand if new entrants can learn from previous failures or successes (Yang 2012) and examine whether the decision to avoid locations with higher failure rates has empirical evidence favoring it.

Effect of neighbors with similar product offerings

Literature in agglomeration economies argues that in order for firms to benefit better from agglomeration, they need to invest in product differentiation. In fact, the localized and direct competition between firms tends to push them apart while benefits of complementary differences attract them to dissimilar competitors (Baum and Haveman 1997; Picone et al. 2009). To study the dynamics of product differentiation for restaurants, we differentiate the firms along two dimensions: price and cuisine and investigate how the product offerings of neighbors affect a restaurant's success.

Effect of neighbors affiliated to a franchise

Studies in the past have found that restaurants affiliated to a franchise have multiple competitive advantages compared to individual restaurants: (1) they experience less failure (Michael and Combs 2008; Parsa et al. 2005, 2011) and obtain financial aids with ease; (2) they experience less income volatility during changes in the economic conditions (Koh et al. 2015); and (3) also enjoy better brand recognition with a ready customer portfolio (Salar and Salar 2014). Peiró-Signes et al. (2015) even give evidence for franchised hotels performing better in a cluster compared to individual units. From this analysis, we could infer that customers may have a preference for neighboring franchises which would be detrimental to a restaurant's survival. But prior studies suggest that small independent establishments can benefit from heightened demands caused by collocating with larger franchised establishments (Chung and Kalnins 2001; Tsang and Yip 2009). This increase in demand is caused primarily due to more extensive investment in marketing by the larger firms and hence the increased attractiveness of the location. Therefore, franchised restaurants as neighbors can be beneficial as well as detrimental to a restaurant's performance. We aim to find empirical evidence for the true nature of this relationship and study the effect of the composition of neighboring restaurants (in terms of the ratio of neighbors affiliated to a franchise) on the chance of survival for a restaurant.

Methodology

Data

In this study, we seek to identify the relationship between a restaurant's failure and its neighboring competitors' characteristics. Although there is no universal definition, literature reports a variety of definitions and proxies for business failure. The four major definitions for small businesses (as discussed by Watson and Everett (1996) are: (1) the discontinuance of the business for any reason, (2) formal bankruptcy, (3) termination to prevent losses, and (4) failure to achieve required returns or meet objectives. However, acquiring bankruptcy data which can explain failure corresponding to definitions (2) – (4) is difficult. To alleviate this problem, we use business data from Yelp which provides information about the discontinuance of a business.

Yelp, founded in 2004, hosts crowd-sourced reviews and ratings about local businesses and owns a monthly average of 29 million unique visitors, and a total of 148 million user reviews as on date². They provide very rich user and business-related data consisting of small, medium, and large businesses as listed on their website. Yelp users report restaurants that have ceased operations, which is verified by Yelp and then the restaurant is reported as 'closed' online. Unlike financial bankruptcy data, Yelp closure data is available for all listed businesses without much delay after closure. Hence, it is also easier for us to estimate the time period of closure from the timestamp of the latest review received by the business.

We focus on 2,602 restaurant businesses in the city of Phoenix, Arizona, USA in this study (of which 2051 are still in operation and 551 have failed). All the businesses provided by Yelp have at least 3 unfiltered reviews older than 14 days. Our sample consists of restaurants that have opened (had its first review) after 1st Jan 2008 and contains data points dated till 26th July 2017. Prior to 2008, only a small percentage of operational businesses were listed in Yelp which constitutes to about only 2% of the total number of reviews

² Available on <https://yelp.com/about>

available in the dataset. The data used comprises of two components: (1) static data, such as the restaurant's name, location, pricing, etc., that do not change over time and (2) dynamic data, such as reviews, review star ratings, etc., that change over time. We aggregated the static and dynamic data to a weekly level to form a panel dataset for this study.

Variable Description

A brief description of all the variables used in the model are given below:

Dependent Variable

Failure: We have defined a 0/1 binary variable to indicate the failure of restaurants. The restaurant is marked as operational from when its first review appears. For the restaurants that are being reported as closed, the closure week is assumed to be the week immediately succeeding the week in which the restaurant received its final review. We use virtual death as a proxy for real-world death.

Independent Variables

Neighborhood-level Variables

Six variables describing the characteristics of a neighborhood of the restaurant are used in this model. Neighbors are the restaurants in a neighborhood which is defined as the area geographically within r km radius from the restaurant in question and. We take r to be 1000m in the main model and later test the robustness of the result by implementing survival analysis for 2 different values for $r = 750\text{m}$ and $r = 1250\text{m}$. The following are the variables calculated for the neighborhood of each of the restaurants:

Neighboring Restaurant's Performance Variables:

1. *Neighbor Footfall*: The total number of reviews received by all neighbors is used as a proxy for the total number of customers visiting the neighbors during that particular week.
2. *Neighbors' Latest Star Rating*: For each of the neighboring restaurants, the mean of the latest three reviews is calculated as *Latest Star Rating*. The average value of all the *Latest Star Rating* of neighbors corresponds to the recent performance of the neighbors in terms of their Yelp star ratings.
3. *Neighbor Failure Ratio*: This variable captures how well restaurants have fared in the neighborhood. It is calculated as the number of restaurants in the neighborhood who have failed as on date, given as a ratio of the total number of restaurants that ever opened in the neighborhood in the dataset.

Neighborhood Composition Variables:

4. *Franchise Neighbor Ratio*: This variable is used to perceive the effects of having franchised restaurants in the nearby area. It is calculated as the number of live neighbors who have at least one another restaurant belonging to its franchise in the dataset and given as a ratio of the total number of neighbors. Two restaurants are considered to belong to the same franchise if they share the same restaurant name.
5. *Same Price Neighbor Ratio*: This variable is included in the model to incorporate the effects of neighboring restaurants that are offering similar products in the same price bracket. It is computed as the number of live neighbors who operate at the same price range as the restaurant in question and given as a ratio of the total number of neighbors.
6. *Same Category Neighbor Ratio*: This variable also captures the effect of neighboring restaurants making similar product offerings but in terms of the type of cuisine that is served. It is calculated as the number of live neighbors serving the same category of food as the restaurant in question given as a ratio of the total number of neighbors.

Control Variables

We define the following eight restaurant-level variables as control variables for the survival analysis:

Restaurant Attribute Variables:

1. *Pricing*: This variable incorporates the difference in failure rates that can be attributed to the price structure of the restaurant. On Yelp, restaurants are labeled as \$, \$\$, \$\$\$ or \$\$\$\$ for inexpensive, moderate, pricey, and ultra-high-end according to their price range. Pricing is an ordinal categorical variable that takes values 0, for low price (\$ and \$\$), and 1, for high price (\$\$\$ and \$\$\$\$). This is a static variable (i.e. varies for each restaurant but retains the same value over time).
2. *Age*: The variable measures how old the restaurant is. We include this in the model to incorporate the possible reduction in hazard rate that can be attributed to the reputation that a restaurant holds because of how old it is. Age at a particular point in time is calculated as the number of weeks from the appearance of its first review. This is a time-varying variable.
3. *Category Popularity*: This is a binary 0/1 variable which is a measure of the popularity of the type of cuisine the restaurant serves. We have classified the restaurants into 9 broad categories based on the type of restaurant and the genre of cuisines that the restaurant serves (e.g. Asian food, dessert parlors and cafés, Mediterranean food, etc.). It takes the value 1 if the category is popular (i.e. the number of restaurants in the dataset serving the same category is > 400, top 80 percentile) and 0 otherwise.
4. *Attribute Count*: Yelp lists various attributes about a restaurant (for example, parking space presence, Bring-Your-Own-Bottle facility, Wi-Fi availability, delivery service, etc.) for the customers' information. This variable is calculated as the total count of all the restaurant's attributes as listed on Yelp. This is included to incorporate the potential positive effect attributes may hold in reducing the failure risk of a restaurant.
5. *Franchise Restaurant Count*: This variable is used to measure the size of the franchise of the restaurant. It is calculated as the total number of restaurants in the dataset having the same name as the restaurant in question.

Restaurant Review Variables:

6. *Monthly Review Count*: This variable acts as a proxy for the total footfall that a restaurant receives and is also a measure of the online reputation of the restaurant. It is calculated as the total count of all the reviews that a restaurant receives during that month (last 4 weeks).
7. *Latest Star Rating*: This variable is also a measure of the online reputation and the quality of the restaurant. The variable is calculated as the average value of the latest three reviews received by the restaurant. We consider only the star ratings of the latest reviews because it is a good indication of the real-time quality of the restaurant's food and services. Moreover, in general, customers look out for the latest reviews on Yelp.
8. *Groupon Offering*: This is a 0/1 binary variable used to indicate whether a restaurant is offering Groupon deal during the time period. We consider the presence of the word 'Groupon' in the user reviews as an indication that the restaurant was offering Groupon deals during the time period.

The descriptive statistics and Pearson correlation values among the continuous variables are presented in Table 1 and Table 2, respectively. We find that no two variables are strongly correlated and hence, all the variables are retained for analysis.

Model

Since we are dealing with survival data we use the Cox proportional hazards model for this study. The probability of occurrence of an event at a given set of points in a time interval can be evaluated using survival analysis. In a survival model, the survival function $S(t)$ captures the probability that a restaurant lives longer than time t (Cox and Oakes 1984). It is given by $S(t) = P[T > t]$ where T is the time to failure for the restaurant. We can calculate the hazard, $h(t)$, defined as the probability of failure in the next instant given that the restaurant already survived till time t (Kleinbaum and Klein 2010):

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t < T < t + \Delta t | T > t]}{\Delta t} = \frac{-S'(t)}{S(t)} \quad (1)$$

| Variable | | Mean | S.D. | Min | Max | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|--|-----------|--------|--------|-----|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|----|
| <i>Neighbor Weekly Footfall_{i,t}</i> | 1 | 10.45 | 19.34 | 0 | 204 | 1 | | | | | | | | | | |
| <i>Neighbors' Latest Star Rating_{i,t}</i> | 2 | 3.39 | 0.43 | 1 | 5 | 0.10* | 1 | | | | | | | | | |
| <i>Neighbor Failure Ratio_{i,t}</i> | 3 | 13.75 | 13.62 | 0 | 100 | 0.36* | 0.04* | 1 | | | | | | | | |
| <i>Franchise Neighbor Ratio_{i,t}</i> | 4 | 40.14 | 22.86 | 0 | 100 | -0.21* | -0.37* | 0.00 | 1 | | | | | | | |
| <i>Same Price Neighbor Ratio_{i,t}</i> | 5 | 56.43 | 24.19 | 0 | 100 | -0.11* | -0.05* | -0.04* | 0.06* | 1 | | | | | | |
| <i>Same Category Neighbor Ratio_{i,t}</i> | 6 | 52.13 | 28.37 | 0 | 100 | 0.03* | -0.03* | 0.06* | 0.12* | -0.02* | 1 | | | | | |
| <i>Age_{i,t}</i> | 7 | 162.55 | 117.88 | 0 | 499 | 0.07* | -0.05* | 0.28* | 0.11* | 0.02* | 0.03* | 1 | | | | |
| <i>Attributes Count_i</i> | 8 | 8.58 | 6.16 | 1 | 22 | 0.04* | 0.01* | 0.02* | -0.01* | -0.01* | -0.01* | -0.01* | 1 | | | |
| <i>Franchise Restaurants Count_{i,t}</i> | 9 | 5.09 | 11.71 | 0 | 67 | -0.03* | -0.05* | 0.06* | 0.10* | 0.09* | 0.18* | 0.09* | -0.01* | 1 | | |
| <i>Monthly New Review Count_{i,t}</i> | 10 | 1.29 | 2.78 | 0 | 112 | 0.25* | 0.09* | 0.19* | -0.08* | -0.08* | 0.03* | -0.04* | 0.02* | -0.16* | 1 | |
| <i>Latest Star Rating_{i,t}</i> | 11 | 3.39 | 1.07 | 1 | 5 | 0.05* | 0.11* | 0.01* | -0.10* | -0.02* | -0.05* | -0.10* | -0.01* | -0.25* | 0.18* | 1 |

Table 1. The descriptive statistics ($N = 617,649$) and the Pearson correlation (with significance value) for all the continuous variables used in the model. (*, $p < 0.01$)

| Variable Name | Distribution | |
|--|--------------|----------|
| | 0 | 1 |
| <i>Pricing_i</i> | 610,908 | 6,741 |
| | (98.91%) | (1.09%) |
| <i>Category Popularity_{i,t}</i> | 489,116 | 128,533 |
| | (79.19%) | (20.81%) |
| <i>Groupon Offerings_{i,t}</i> | 616,723 | 926 |
| | (99.85%) | (0.15%) |

Table 2. Summary statistics of categorical variables ($N = 617,649$)

If we assume that this hazard function for a restaurant is given by the proportional hazards model, then for the set of explanatory variables X , the hazard function can be re-written as:

$$h(t, X) = h_0(t)\psi(X) \quad (2)$$

where $\psi(X)$ is some function of X such that $\psi(0) = 1$ and $h_0(t)$ is the baseline hazard function. A special form of the proportional hazards model also called the Cox Proportional Hazards Model as proposed by Cox (Cox 1972) (hereby also referred to as the Cox model) is used in our analysis. This model assumes that $\psi(X) = e^{(\beta X)}$, where β is a vector of regression coefficients. Now, the hazard function can be re-written as:

$$h(t, X) = h_0(t)e^{(\beta X)} = h_0(t)e^{(\sum_{i=1}^n \beta_i X_i)} \quad (3)$$

where β_i 's are the coefficients corresponding to X_i 's and n represents the total number of variables (X 's). This survival model makes the key assumption of proportional hazard or constant relative hazards, i.e. the survival curves for two strata, as determined by the choice of covariates, must have hazard functions that are proportional over time (Kleinbaum and Klein 2011). Hence, the proportional hazard assumption must be tested first before implementing the model. We test this assumption using Schoenfeld residuals (Schoenfeld 1982). Testing for a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time is equivalent to testing for time dependency of covariates. To perform the test and to run the models, we have used STATA 12 and the computation of all variables has been done on R.

The probability values obtained from the test is the result of the hypothesis test for a non-zero slope of residuals for each of the variables. The graphs of the residuals for the variables that pass the tests are also plotted to ensure that we do not miss out any non-linear relationships between the residual and the function of time which cannot be detected by a zero-slope test. Due to space constraints, the tables and graphs are not presented in this paper. At 90% confidence interval, four variables are time-dependent, namely, *Neighbors' Latest Star Rating_{i,t}*, *Franchise Restaurants Count_{i,t}*, *Monthly New Review Count_{i,t}* and *Latest Star Rating_{i,t}*. Time dependency of covariates result in the overestimation of the effects of covariates whose associated hazards are increasing while the coefficients are biased towards zero in the case of converging hazards, and hence, incorrect results and faulty inferences will be observed (Bellera et al. 2010; Kalbfleisch and Prentice 2010). Thus, we use an extended Cox model to incorporate the effects of the time-dependent variables (Kleinbaum and Klein 2010):

$$h(t, X(t)) = h_0(t)\exp\{\sum_{i=1}^n (\beta_i X_i + \gamma_i X_i f(t))\} \quad (4)$$

where X_i 's are a total of n number of independent variables, β_i 's and γ_i 's represent the coefficients for the main effects and the interaction effects with time respectively for each of the X_i 's. For time-independent variables, corresponding γ_i 's are set to 0. $f(t)$ is a function of time, here we take $f(t)$ to be $\ln(t)$. Including the effects of interactions between the independent variables, the final model can be written as:

$$h(t, X(t)) = h_0(t)\exp\{\sum_{i=1}^n (\beta_i X_i + \gamma_i X_i f(t)) + \sum_{i=1}^n \sum_{j=i+1}^n \delta_{i,j} X_i X_j\} \quad (5)$$

where $\delta_{i,j}$'s represents the coefficient corresponding to these interaction terms and $\delta_{i,j}$ is set to 0 if the interaction term is not included in the model.

Results & Discussion

Survival Analysis

Table 3 depicts the hazard estimates obtained from the Cox proportional hazards regression. Two models have been presented: (I) with only the restaurant attribute variables and (II) with both neighborhood and restaurant attributes variables, respectively. Hazard ratios of the covariates have been tabulated in Table 3.

Effects of the Control Variables:

In both Model I and II, as expected, higher age (leading to a better tested market strategy and higher reputation), more number of listed attributes (providing more facilities for customers), larger size of franchises the restaurant is affiliated to (leading to all benefits gained by a franchise as discussed in the theory section), and high star ratings (indicative of better quality of food and enhanced online reputation)

| Variables | Hazard Ratio | |
|--|--------------|------------|
| | I | II |
| Main Effects | | |
| <i>Neighbor Weekly Footfall_{i,t}</i> | | 1.0030* |
| <i>Neighbors' Latest Star Rating_{i,t}</i> | | 0.0022** |
| <i>Neighbor Failure Ratio_{i,t}</i> | | 3.8223*** |
| <i>Franchise Neighbor Ratio_{i,t}</i> | | 1.4696* |
| <i>Same Price Neighbor Ratio_{i,t}</i> | | 0.7240* |
| <i>Same Category Neighbor Ratio_{i,t}</i> | | 0.943 |
| 1. Pricing _i | 1.3594 | 1.2149 |
| Age _{i,t} | 0.9985*** | 0.9986*** |
| 1. Category Popularity _{i,t} | 0.8285** | 0.8326 |
| Attribute Count _i | 0.9884* | 0.9876* |
| Franchise Restaurants Count _{i,t} | 0.1141*** | 0.1217*** |
| Monthly New Review Count _{i,t} | 5.4519*** | 6.2802*** |
| Latest Star Rating _{i,t} | 0.0387*** | 0.0623** |
| 1. Groupon Offerings _{i,t} | 10.6009*** | 10.4747*** |
| Interaction Effects | | |
| Monthly New Review Count _{i,t} x Latest Star Rating _{i,t} | 0.9209*** | 0.9187*** |
| Franchise Restaurants Count _{i,t} x Franchise Neighbor Ratio _{i,t} | | 0.8224** |
| Time-Varying Covariates | | |
| <i>Neighbors' Latest Star Rating_{i,t}</i> | | 2.7456** |
| Franchise Restaurants Count _{i,t} | 1.3691** | 1.3731** |
| Monthly New Review Count _{i,t} | 0.7922*** | 0.7739*** |
| Latest Star Rating _{i,t} | 1.6321** | 1.5117** |
| Number of Subjects | 2602 | 2602 |
| Number of Failure | 551 | 551 |
| $P > \chi^2$ | 0.000 | 0.000 |

***, **, * coefficients significant at the $p < 0.01, 0.05$, and 0.10 levels respectively.

Table 3. Cox survival model regression estimation for failure of restaurants for (I) only control variables and (II) both control and neighbor variables

are significantly beneficial to reduce the risk of failure of the restaurant. Of note, all the variables have a hazard ratio less than 1. Groupon deals increase the chances of failure ($H > 1$) significantly which is congruent with the previous studies dictating that customers availing Groupon coupons on an average give lower ratings to the restaurant (Byers et al. 2012). It is observed that reviews with higher ratings are more helpful while price range does not have any significant effect on restaurant survival.

Effects of the Neighborhood Variables:

Neighbors' performance: In Model II, we notice that the hazard ratio of neighbors' footfall is 1.0030, indicating that one unit increase in *Neighbor Weekly Footfall* increases the chance of failure of the restaurant by 0.3%. While in the case of neighbors' star ratings, the hazard value is less than 1 ($H=0.0022$), and its interaction term with time has a hazard ratio greater than 1 ($H_t=2.7456$). This indicates that at $t=0$ (Jan 2008), an increase in neighbors' star improved the chance of success for a restaurant. But over time this benefit deteriorates steadily. Using Equation 6, we notice that Neighbors' star starts adversely affecting the restaurant's survival chance since $t=241$ (Aug 2012). We also observe that restaurants located in a neighborhood with higher failure ratio are more prone to failure ($H=3.8233$).

Neighbors' product offerings: We find that restaurants benefitted from agglomerating along the lines of the same price bracket, as the hazard ratio of *Same Price Neighbor Ratio* is less than 1 ($H=0.7240$). This implies that a restaurant situating in an area with 100% neighbors belonging to the same price bracket experiences 27.6% less chances of failure as compared to 0% neighbors in the same price bracket. *Same Category Neighbor Ratio* is not statistically significant. The findings can be attributed to the fact that price similarity with the neighbors acted in a complementary manner.

Neighbors' affiliation: The hazard rates of *Chain Neighbor Ratio* is 1.4696) and that of its interaction with *Chain Restaurants Count* is 0.8224. This suggests that increasing *Chain Neighbor Ratio* for a restaurant increases its risk of failure. However, it has a reverse effect on a restaurant affiliated to a franchise showing increase in the chance of survival.

Robustness Test

We examine the model outputs for three different radius values for the neighborhood, $r = 750m, 1000m$, and $1250m$. We find that the results are consistent for all the three cases. However, the details have been omitted in this report due to space constraints.

Limitations & Conclusions

A systematic execution of the Cox proportional hazard model on a rich restaurant dataset from Yelp unraveled the nature and the magnitude of the effect of various characteristics of the agglomerating neighborhood on the survival of a restaurant. The outcomes of this paper showcase some interesting insights that can be explained by the theory of agglomeration and are of importance to restaurant owners, financial institutions, and governments. The findings suggest that restaurants those are surrounded by competitors with poorer performance and less failure history in the locality tend to have a lower chance of failure. Prior to the burst of the social media bubble around 2010 in the USA, locating near high-rated neighbors resulted in spill-over of positive effects, however, the trend has reversed with high-rated neighbors negatively influencing restaurants in the recent years. This could possibly be because earlier customers lacked information about every single establishment and were more reliant on the overall quality of the location. But with the digital penetration, customers have more information on individual restaurant's quality of products and services which might lead them to be more selective about their preferred restaurant. Hence, the findings indicate that new restaurants should try to avoid locating near well-performing neighbors.

We also observed that a change in the neighbor composition based on the neighboring restaurants' cuisine did not help or hurt a restaurant. Hence, it is insignificant for restaurants to try to differentiate themselves through product offerings alone. But it is seen that locating near a neighborhood with more similar priced restaurants brought a positive externality. This could also be because people agglomerate based on their income, (Waldfogel 2010) and thus, demand for restaurants belonging to a certain price range also agglomerates.

Our study is also not without limitations. In this paper, we have only made use of data from Phoenix, Arizona. It would be interesting to explore the relationship with a dataset covering a more diverse geographic area. Due to unavailability of data, we have omitted few control variables pertaining the demographic, psychographic and internal business data (like profits, turnover, etc.) that could make the findings more robust. The outcomes of this study can be used as complementary with other preference models to determine effective locations for new entrants in the restaurant industry.

References

- Baum, J. A. C., and Haveman, H. A. 1997. "Love Thy Neighbor? Differentiation and Agglomeration in the Manhattan Hotel Industry, 1898-1990," *Administrative Science Quarterly* (42:2), p. 304.
- Bellera, C.A., MacGrogan, G., Debled, M., de Lara, C.T., Brouste, V. and Mathoulin-Pélissier, S., 2010. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC medical research methodology*, 10(1), p.20.
- Byers, J. W., Mitzenmacher, M., and Zervas, G. 2012. "The Groupon Effect on Yelp Ratings," in *Proceedings*

- of the 13th ACM Conference on Electronic Commerce - EC '12, p. 248.
- Canina, L., Enz, C.A. and Harrison, J.S., 2005. Agglomeration effects and strategic orientations: Evidence from the US lodging industry. *Academy of management journal*, 48(4), pp.565-581.
- Canina Ph D, L. and Harrison Ph D, J., 2005. The Agglomeration Conundrum: How Co-location Helps Some Hotels and Hurts Others.
- Chung, W., and Kalnins, A. 2001. "Agglomeration Effects and Performance: A Test of the Texas Lodging Industry," *Strategic Management Journal* (22:10), pp. 969-988.
- Cox, D.R., 1992. Regression models and life-tables. In *Breakthroughs in statistics* (pp. 527-541). Springer, New York, NY.
- Cox, D.R. and Oakes, D., 1984. *Analysis of survival data* (Vol. 21). CRC Press.
- Kalbfleisch, J.D. and Prentice, R.L., 2011. *The statistical analysis of failure time data* (Vol. 360). John Wiley & Sons.
- Kalnins, A., and Chung, W. 2004. "Resource-Seeking Agglomeration: A Study of Market Entry in the Lodging Industry," *Strategic Management Journal* (25:7), pp. 689-699.
- Kleinbaum, D.G. and Klein, M., 2010. *Survival analysis* (Vol. 3). New York: Springer.
- Koh, Y., Rhou, Y., Lee, S. and Singal, M., 2015. Does Franchising Alleviate Restaurants' Vulnerability to Economic Conditions?. *Journal of Hospitality & Tourism Research*, p.1096348015619411.
- Lee, S. K. 2015. *Conditional Agglomeration Externalities in Lodging Markets*, (39:4), pp. 540-559.
- Luo, H., and Yang, Y. 2016. "Intra-Metropolitan Location Choice of Star-Rated and Non-Rated Budget Hotels: The Role of Agglomeration Economies," *International Journal of Hospitality Management* (59), Elsevier Ltd, pp. 72-83.
- Marshall, A. 1890. "1920. Principles of Economics," *London: Mac-Millan*, pp. 1-627.
- Michael, S. C., and Combs, J. G. 2008. "Entrepreneurial Failure: The Case of Franchisees," *Journal of Small Business Management* (46:1), pp. 73-90.
- Parsa, H. G., Self, J., Sydnor-Busso, S., and Yoon, H. J. 2011. "Why Restaurants Fail? Part II - The Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis," *Journal of Foodservice Business Research* (14:4), pp. 360-379.
- Parsa, H. G., Self, J. T., Njite, D., and King, T. 2005. "Why Restaurants Fail," *Cornell Hotel and Restaurant Administration Quarterly* (46:3), pp. 304-322.
- Picone, G. A., Ridley, D. B., and Zandbergen, P. A. 2009. "Distance Decreases with Differentiation: Strategic Agglomeration by Retailers," *International Journal of Industrial Organization* (27:3), Elsevier B.V., pp. 463-473.
- Peiró-Signes, A., Segarra-Oña, M. del V., Miret-Pastor, L., and Verma, R. 2015. "The Effect of Tourism Clusters on U.S. Hotel Performance," *Cornell Hospitality Quarterly* (56:2), pp. 155-167.
- Salar, M., and Salar, O. 2014. "Determining Pros and Cons of Franchising by Using Swot Analysis," *Procedia - Social and Behavioral Sciences* (122), Elsevier B.V., pp. 515-519.
- Schoenfeld, D. 1982. "Partial Residuals for the Proportional Hazards Regression Model," *Biometrika* (69:1), pp. 239-241.
- Schemper, M., 1992. Cox analysis of survival data with non-proportional hazard functions. *The Statistician*, pp.455-465.
- Teller, C., Alexander, A., and Floh, A. 2016. "The Impact of Competition and Cooperation on the Performance of a Retail Agglomeration and Its Stores," *Industrial Marketing Management* (52:0), pp. 6-17.
- Tsang, E. W. K., and Yip, P. S. L. 2009. "Competition, Agglomeration, and Performance of Beijing Hotels," *Service Industries Journal* (29:2), pp. 155-171.
- Waldfoegel, J. 2010. "Who Benefits Whom in the Neighborhood? Demographics and Retail Product Geography," [Book] - *Agglomeration Economics* (February), pp. 181-209.
- Wang, L., Gopal, R., Shankar, R., and Pancras, J. 2015. "On the Brink: Predicting Business Failure with Mobile Location-Based Checkins," *Decision Support Systems* (76), pp. 3-13.
- Watson, J., and Everett, J. 1996. "Small Business Failure Rates: Choice of Definition and the Size Effect," *Journal of Entrepreneurial & Small Business Finance* (5:3), p. 271.
- Yang, N., 2012. *March of the chains: Herding in restaurant locations*. Yale SOM working paper.
- Zhang, J., and Enemark, A. E. 2016. "Factors Influencing Business Performance in Hotels and Restaurants," *Asia-Pacific Journal of Innovation in Hospitality and Tourism* (5:1), pp. 1-20.