

Schaffner, Florian

Conference Paper

Information transmission in high dimensional choice problems: The value of online ratings in the restaurant market

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Forecasting and Information Transmission, No. B01-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Schaffner, Florian (2016) : Information transmission in high dimensional choice problems: The value of online ratings in the restaurant market, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Forecasting and Information Transmission, No. B01-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/145585>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Information transmission in high dimensional choice problems: The value of online ratings in the restaurant market*

FLORIAN SCHAFFNER
Department of Economics, University of Zurich

July 8, 2016

Abstract

In choice problems with many alternatives and a priori uncertain outcomes, it has long been argued that individuals may use the observed choices of others as information to guide their own decisions. This paper analyses the role of these social interactions in the context of restaurant choice, while at the same time evaluating the value of a restaurant user rating in determining economic outcomes (as measured by restaurant checkins). The model follows a Polya urn logic, with choices being analyzed in a Dirichlet-multinomial framework. Data is provided by the online urban guide Yelp. Results show that correlation across choices of restaurant guests is present and can be modelled as a function of group-level variables related to information exchange. Ratings and other factors such as price categories do play a role, though traditional modelling approaches that ignore social interactions tend to overstate its importance both in terms of statistical and economic significance. The presented results from the restaurant market may well prove to be important in other choice contexts characterized by many alternatives and highly skewed outcome distributions.

Keywords: *Information transmission, social interactions, online reviews, Dirichlet multinomial, choice modelling*

JEL: *C46, C53, D71, D12*

*Address for correspondence: University of Zurich, Department of Economics, Zürichbergstrasse 14, CH-8032 Zurich, Switzerland, Tel. +41 44 634 22 97, florian.schaffner@econ.uzh.ch.

I would like to thank Johannes Kunz, Janina Nemitz, Rainer Winkelmann, Ali Yurukoglu, the participants of the 2014 Zurich Workshop on Economics and seminar participants at the University of Zurich for helpful comments and suggestions. Errors and omissions are my own.

1 Introduction

Choosing a place to eat is not easy. According to Beinhocker (2007), there exist over 50'000 restaurants in New York City alone. Restaurant choice is a demanding, highdimensional choice problem from the perspective of an individual - so much so that it has been argued (Simon, 1955; Ormerod et al., 2012; Sela and Berger, 2012) that with such an abundance of choices, individuals may lack the processing capacity to select the optimal choice, even when (or indeed especially when) complete information is available. Consequently, it has been suggested that individuals may therefore resort to other strategies to make their decisions, such as heuristics. In the case of restaurant choice, uncertainty about the payoffs is further complicating the problem, as not only alternatives are abundant, but also the quality of the food is only imperfectly observable until after the choice was made.¹

It is increasingly recognized that in the context of such problems, social interactions may play a crucial role, possibly affecting both efficiency and equity of the resulting allocation (Granovetter, 1985; Jackson, 2010; Vega-Redondo, 2007). Instead of processing the vast amount of imperfectly observed information and making independent choices, individuals may use the observed choices of others to help guide their decisions. In what they call social network markets, Potts et al. (2008) suggest that “the very act of consumer choice is governed not just by the set of incentives described by conventional consumer demand theory, but by the choices of others in which an individual’s payoff is an explicit function of the action of others.”² In other words, each individual’s action creates positive externalities, transmitting information about an alternative from which future decisionmakers can learn (De Vany and Walls, 1996). This in turn leads to correlated choices and individuals clustering around a limited set of the many choices available. Social interactions have been shown to play a significant role in a variety of contexts, including movie attendance (De Vany and Walls, 1996; McKenzie, 2008), music (Salganik et al., 2006), book sales (Beck, 2007), health care plan choice (Sorensen, 2006) and hospital choice (Guimaraes and Lindrooth, 2007; Pauly and Satterthwaite, 1981). Becker (1991) made a similar argument for restaurants in a theoretical framework.

In this paper, I analyze the role of information transmission via social interactions in the context of restaurant choice. More specifically, I put the focus on two aspects. First, I show that overdispersion in restaurant checkins is present in the data and that this overdispersion can be explained using a Polya Urn dynamic, incorporating previous guests choices in the decision process. I model the strength of social interactions as a function of both information

¹This is true even for repeated visits of an individual at the same restaurant, as food quality may vary from day to day or meal to meal.

²Note that I will use the terms “information transmission”, “observation of actions of others” and “social interactions” as a synonym for others’ choices.

exchange-related as well as socioeconomic group variables. To my knowledge, this paper is the first to embed social interactions in this manner and put the Polya Urn dynamic to use in the context of restaurant choice.

Second, I analyze the importance of user-provided ratings and other attributes of restaurants in the social interactions model. The importance of user-provided ratings has been hotly discussed in recent years both in popular media and research (for an example in the restaurant context, see Luca (2011)), with such ratings sometimes being described as a new form of currency or reputation for businesses. While debated, an assessment of the value of such ratings remains an open issue, and especially in a context with correlated behavior such ratings may provide little insights with respect to the economic outcomes of such businesses, as in these cases individuals follow the previous choices regardless of objective measures. Put simply, in a world where choices are perfectly correlated, a first individual chooses a restaurant and all successors adopt the same choice no matter what the rating says (whereas in the case of independent choices, individuals decide solely on attributes of a restaurant, disregarding the choices of others).

Researchers analyzing choice problems with aggregate data typically use McFadden's random utility framework and estimate a conditional logit model (McFadden, 1974). However, from an econometric perspective, social interactions lead to overdispersion and violate the multinomial assumption, resulting in a misspecified model. To account for information transmission from other individuals and the resulting overdispersion, I use a Dirichlet multinomial regression model, which treats choice probabilities as random variables rather than as constant parameters of the multinomial distribution. Under the Dirichlet multinomial distribution, conditional choice probabilities increase proportionally to the number of individuals who have previously chosen that alternative, introducing a sequential aspect by making today's choices dependent on past choices - even in the case when one only has cross-sectional data at hand. The multinomial distribution and thus independence of individual choices is nested in the Dirichlet-multinomial as a limit case, which allows directly testing the two models against each other.

While the Dirichlet-multinomial model has been used in the econometric literature before (e.g. Guimaraes and Lindrooth, 2007), researchers typically omit to back out the parameters of the Dirichlet distribution, thereby failing to draw final conclusions about the predictability of outcomes. As I am also interested in the predictability of success of a restaurant dependent on its rating, I will discuss these parameters in this paper.

I use a dataset provided by the online urban guide Yelp. The dataset contains information on checkins, ratings and other attributes of restaurants in the metropolitan area of Phoenix (AZ), covering 3'171 restaurants across 125 ZIP codes within a period from early 2010 to early 2015. While it may be true that platforms such as Yelp offer only selected samples,

I regard an ecosystem with an average 135 million monthly users and 67 million reviews as being interesting in itself. The fact that such platforms impact real-world decisions and their relevance is growing as the use of these platforms spreads across the population makes the study of such data important and meaningful beyond its own sake.

The model is estimated across different markets and allows the social interaction parameter to vary across these markets. By modelling social interactions within a market directly as a function of market level variables, I can broadly separate effects of within-market homogeneity of individuals from social interaction effects. For example, one would expect that in a market with low income inequality, interactions are higher than in markets with higher income inequality, as individuals are more alike (in terms of income) and may therefore learn more from their peers than in a high-inequality environment.

In the baseline specification, markets are defined on a ZIP code and price category level (each price category in each ZIP code constitutes a market). To check for the robustness of results, different market definitions are used, markets are evaluated across and within three time periods and choices are analysed on a higher aggregation level. All the results are robust vis-à-vis these alternating definitions and periods.

I find strong evidence for social interactions, justifying the use of the Dirichlet multinomial model. The interactions are driven by a combination of factors, including both information transmission variables (such as the total number of reviews in a market) and market-characteristic variables (such as the price level, the income level or the number of competitors). Meanwhile, within-group heterogeneity in income as measured by an inequality-proxy has no significant impact on correlation within a group. I also find that higher ratings have a positive effect on visit probabilities, but that the overall informational value of such ratings is limited.

Section 2 explains the role of internet data in information transmission, section 3 discusses data, followed by the outline of the model and estimation of the model in section 4. Section 5 presents main results, section 6 shows robustness checks. Section 7 concludes.

2 Information transmission and the role of internet data

In the classic perspective of choice theory, individuals observe a set of choices and their characteristics before choosing whatever option maximizes their utility. The restaurant market can be seen in such a choice context as well, but it has two features that differentiate it from more traditional choice applications. For one, there is uncertainty about the choices' arguably most important characteristic: the quality of a restaurant (including the food, the location and/or the atmosphere). A priori, individuals do not have certainty about the payoffs of eating at a particular restaurant. An individual may know that she has a preference for pizza,

but she does not know whether she will like the pizza served at a particular restaurant. All she can do is form an expectation about the quality based on restaurants' observables, which might be more or less accurate. At the same time, the a priori unobserved or only imperfectly observed quality component seems crucial in explaining the highly unequal economic outcomes of restaurants (as depicted by the checkins distribution on the left hand side in Figure 1). This is true even within a geographic area, price or food category.

The second particularity in the restaurant context is the abundance of choice. While in theory this does not change the choice problem, it has been argued that from a behavioral perspective, the abundance of alternatives and the corresponding information attached to these options is so vast that it makes it impractical for individuals to sift through all the options before making a choice (Simon, 1955). This is not just true in contexts of incomplete or imperfect information, but also in situations where all the relevant choice attributes are observed - indeed, complete information might worsen the problem as it raises the computational burden.

Researchers have long argued that both in contexts with uncertain payoffs as well as in contexts with high-dimensional choice sets, individuals resort to social interactions (Ormerod et al., 2012), either to learn new information from signals of others or to circumvent the burden of evaluating all the information themselves³. The interactions can broadly be thought of as externalities: every time an individual takes a decision, she gives away information to her peers that they in turn can take into account in their own decisions. Building on the previous choices, individuals deciding later in the sequence can filter out the best alternative. Repeating this process induces correlation across the decisions of individuals, which in turn can lead to highly unequal outcome distributions.

Such choices and interactions have been hard to measure and track in the past. But in recent years, internet data has been growing at an unprecedented pace. For example, by logging checkins and reviews of their users, Yelp measures and publishes information that was largely restricted to observations in limited geographic space (where people go) and word of mouth (what they tell about their experience) before the internet age. It is by using this data that I want to analyze how individuals choose restaurants, in light of the arguments made above.

The paper focuses on two measures in Yelp data: checkins and ratings. By including previous checkins of others in the decision process, one induces correlation across individuals' choices (or the same individual choosing multiple times), enabling the "rich-get-richer"-property that leads to the unequal outcome distribution seen in Figure 1. Ratings, on the other hand, are used as a proxy for the quality of a restaurant (assumed to be unbiased), where higher quality leads to a priori higher checkin probabilities, *ceteris paribus*. While these ratings are

³Social interactions have many labels such as information transmission, information diffusion or herding; I will use these terms interchangeably here.

a measure for quality, they can be noisy, especially in the beginning when the overall rating is based on only a few observations: Different individuals may rate the same restaurant very differently and may therefore not rely on these ratings all that much. In an extreme case, individuals may disregard others' opinions completely, deciding in complete uncertainty with respect to the quality of a restaurant. At the other end of the limit case, ratings are taken as an objective and valid measure of quality, in which case no person should go to a low-rated restaurant as long as there is a higher-rated one (conditional on cost). Data on the choices of individuals in combination with the model described below allow to assess the informational "value" of these ratings (where a rating is of no value in the former extreme case of complete uncertainty, but very valuable in the latter case). As individual ratings accumulate over time, the average rating should become more accurate, and one should move away from the first extreme, towards the second extreme. This dynamic should be detectable when analyzing the restaurant market at different points in time.

3 Data

Yelp is an online urban guide collecting visit and review data on businesses, most of which are active in the food and drink industry. The platform is visited by about 6 million people daily. The audience is characterized by an overweight of female visitors, an overweight in the 18 to 44 year old group and an above-average education and income level, relative to demographics of the average web user in the US.⁴

The data used in this paper is a collection of samples offered through the Yelp dataset challenge and includes information on businesses in the metropolitan area of Phoenix, Arizona, covering three snapshots over a period of five years between early 2010 and early 2015. Data on checkin counts (a feature that Yelp introduced in January 2010) as well as ratings and other restaurant attributes have been accumulated over the period from January 2010 to January 2015, with the snapshots taken in early 2013, early 2014 and early 2015. I use the checkins as a count measure for restaurant visits. Snapshots show cumulative checkins from 2010 onwards; having several snapshots across time allows me to calculate annual checkins for the years 2013 and 2014. The 3'171 restaurants in the final dataset span 42 cities, 125 ZIP code areas and count a total of 895'265 checkins as of early 2015. Roughly half of the restaurants are chain restaurants (defined as restaurants existing in more than 1 location). Additional data on ZIP-level economic and demographic characteristics are 5-year estimates from the 2013 US Census Survey.

Individuals checking in at a restaurant inform others about their choice. The probability of choosing a particular restaurant is calculated as the number of Yelp checkins divided by the total number of checkins in the same market. This checkin-based definition is different

⁴<https://www.quantcast.com/yelp.com>

from more traditional measurements such as revenue-based market share calculations, and as such it has its own characteristics. Rather than seeing checkins as a 1:1 reflection of visits, one should see it as a proxy for these visits. The number of checkins, averaging somewhere below 300 at the end of the sample period, is (presumably) much lower than the actual number of visits. Also, in my simple model, the unconditional probability of an individual checking in on Yelp is assumed to be constant across individuals, restaurants and time. This is clearly a strong simplification, and a violation of this assumption may introduce measurement error. If the measurement error is independent of the true probability p_j , it must be true that the variance of the observed probability is higher than the true variance. Alternatively, if the measurement error is like prediction error, that is, the observed choice probability is an unbiased prediction of the true probability, then the observed variance underestimates the true variance (Glaeser et al., 1996). Yelp users might differ with respect to the checkin probability among themselves as well as compared to non-Yelp users. Also, checkin probabilities might differ across restaurants, presumably overestimating true visit probabilities of trendy or especially good restaurants while underestimating true probabilities of less popular alternatives.

That being said, it is important to note that I am mainly interested in the behavior and the dynamics within the Yelp ecosystem; after all, Yelp users have access to that same information to make their decisions. In the end, these decisions happen in the “real” world, and actions based on information from online platforms become increasingly relevant as the use of such platforms spreads across the population. Alternatively, one can interpret a checkin of an individual as an implicit recommendation to others. While such a viewpoint is different, it serves equally well to a researcher interested in information transmission. Aside from that, more traditional proxies for market shares or choice probabilities do not come without their own drawbacks. Both revenue-based and profit-based calculations tend to overweigh expensive, high-margin restaurants, for example.

[Figure 1 about here.]

The restaurant population is restricted to restaurants that are recorded over the whole period of 2010 to 2015. A first look at the data in Figure 1 confirms the clustering around a few choice alternatives found in other social network markets. The distribution of cumulative checkins over a period of five years shows a large number of restaurants with very few checkins, and a small number of restaurants collecting a disproportionately high number of checkins: The top 10 percent of restaurants collectively combine as many checkins as the bottom 57 percent. That inequality is not reflected in the (bounded) ratings distribution shown in the right half of Figure 1, which is roughly bell-shaped and centered around a mean value of 3.5. Distributions on lower aggregation levels are characterized similarly. Note that, on an individual restaurant

basis, the variance of the ratings decreases the higher the average rating, an indication that individuals generally agree on what is a good restaurant, but have different opinions when it comes to bad ones (not shown here).

[Table 1 about here.]

There's a set of 8 variables used in estimation, with a number of additional interaction variables (price category and rating, price category and income). Summary statistics for all variables are shown in Table 1. The dependent variable CUMULCHECKINS measures the cumulative checkins at a restaurant. The checkin distribution is highly skewed, with few restaurants capturing a disproportional share of the market (see also Table 7 in the appendix). Note that checkins have no exact timestamp, but can be assigned to the years 2010-2012, 2013 or 2014. The main explanatory variable is the STARS rating, the (rounded) average rating given by users which is used as an initial quality guess (which supposedly gets better as the number of ratings grows). Ratings go from 1 to 5 in half-steps, are constant over time in the vast majority of cases and will serve as a measure of quality of a restaurant. As there are only few restaurants in the lower and upper parts of the STARS distribution, the variable has been recoded as a rating variable with four groups defined as $RATING1 = \{1, 1.5\}$, $RATING2 = \{2, 2.5\}$, $RATING3 = \{3, 3.5\}$, $RATING4 = \{4, 4.5, 5\}$. The variable PRICE indicates the price category of a restaurant, ranging from 1 to 4. Price category one is a restaurant serving food below 10 dollars, price category two ranges from 11 to 30 dollars, category three from 31 to 60 dollars and category four is for prices from 61 dollars. The categories three and four have been merged due to the small number of restaurants in these categories. The variable LNSUMREVIEWS is defined as the logarithmized total number of reviews within a group, serving as a measure for the overall Yelp activity within the group. The variable LNCOMP is the logarithmized count of the number of competitors in the same group. The variables LNINC and LNPOP are ZIP-level measures for log-income and log-population, respectively. Finally, the variable INEQ is an income inequality measure defined as the ratio of mean and median income (ZIP-level) used to approximate within-group homogeneity. An inequality measure above 1 implies a right-skewed income distribution. Within-group homogeneity is potentially important since in a more homogenous group, individuals are expected to respond more strongly to the information provided by others, while in more heterogeneous groups such signals may not be as important.

4 Model

The model of restaurant choice employed here is an adapted version from De Vany and Walls (1996), who used it to explain the unequal revenue distribution at the box office via word-of-mouth recommendations and correlated decisions of moviegoers. The model is a generalized

version of a Polya Urn scheme, where the first individual draws a ball of color j with some probability from an urn, replaces that ball and adds an additional ball of the same color to the urn, thereby enabling a “rich-get-richer” dynamic that is captured as social interactions in the present context. It is also closely related to the Chinese Restaurant Process (Aldous, 1985), but with a preset number of restaurants (blocks in the partition).

4.1 Restaurant choice with perfect information

I start with the benchmark of a simple, perfectly informed world. Suppose there is a sequence of $i = 1, \dots, N$ individuals in a market M who have to choose a place to eat from a choice set of R restaurants. The decision set of the i th consumer is denoted by $d_i = \{1, \dots, R\}$ and is identical for all individuals in the sequence. For simplicity, assume that each of these restaurants represents a distinct quality level q . Individuals maximize a utility with a quality-dependent payoff.

In the case of perfect information, all individuals make choices observing quality perfectly, leading to a quality-dependent vector of choice probabilities $p(q) = \{p_1, \dots, p_R\}$ that is equal across individuals. The N individuals make their choices independently from one another and allocate themselves across restaurants $1, \dots, R$, resulting in the allocation vector $A = \{A_1, \dots, A_R\}$ whose realization can be defined as $a = \{a_1, \dots, a_R\}$ where

$$\sum_{k=1}^R a_k = N.$$

Any particular outcome is multinomial distributed with probability function

$$\begin{aligned} Pr\{A_1 = a_1, \dots, A_R = a_R | p_1, \dots, p_R\} = \\ Pr\{A = a | p\} = \frac{N!}{a_1! \dots a_R!} p_1^{a_1} \dots p_R^{a_R} \end{aligned} \tag{1}$$

4.2 Restaurant choice with imperfect information

Now suppose that the quality of a restaurant is unknown or only imperfectly known to individuals at the moment they make their decisions, introducing uncertainty about the payoffs of a particular choice. Instead of maximizing utilities, individuals now maximize expected utilities that, as I will show later, depend on their position in the choice sequence.

While in this setting, individuals cannot observe the quality of a restaurant directly, they now have two alternative measures at their disposal. For one, they can form an expectation about the quality of a restaurant using restaurant ratings $s = E(q)$, which on average is

assumed to be an unbiased quality estimate. On the other hand, individuals can observe the choices of the $(i - 1)$ individuals preceding them in the sequence and use these observations as additional signals for quality.

In statistical terms, p is now itself a random vector dependent on ratings, rather than a constant parameter reflecting quality as in the perfect information case. Whereas before, a was the random realization of an allocation vector generated by the parameters in p , there is now an additional random layer as the entries in p are generated from a random process as well.

The probability density of choosing a restaurant given p , $P(A = a|p)$, is still a known function: It's the multinomial density shown in the previous section. To account for the fact that p is a random vector, I need a prior probability distribution for p supposed to reflect individuals' beliefs about the ratings as a quality indicator. In other words, as p is a function of ratings, one can interpret this prior probability distribution as the informational value of a rating, absent any information about the choices of others.

Now consider what happens when individuals later in the sequence start taking into account the choices that were made previously. They can use the previous choices to update their beliefs and form a posterior distribution, where the posterior distribution of p is proportional to the product of the known density of choosing a restaurant given p and the prior distribution of p . One can now see what role previous choices play: They change the distribution of the number of guests visiting a restaurant. For example, it can be shown that on average, the distributions become more concentrated, reflecting an average gain in knowledge (Lindley, 1961). So while the first individual in a sequence can base his decision only on a prior distribution, the individual making the N th decision has access to much more accurate information to make his decision.

$P(A = a|p)$ is still the multinomial density defined previously. What I need to embed quality uncertainty is a prior distribution of p given the ratings, $P(p|\delta^{-1}\alpha)$, where the value of a rating is captured in the parameters δ and α . In my case, I assume p to be Dirichlet distributed. The Dirichlet distribution (also known as the multivariate Beta distribution) is a conjugate prior for the probability parameter p of the multinomial distribution with density:

$$f(p_1, \dots, p_R; \alpha_1, \dots, \alpha_R, \delta) = \frac{1}{B(\delta, \alpha)} \prod_{k=1}^R p_k^{\delta^{-1}\alpha_k - 1} \quad (2)$$

where $\sum_{k=1}^M p_k = 1$, $\alpha_k > 0$, $\delta > 0$. $B(\delta^{-1}\alpha)$ is the β function that can be expressed in terms of the Gamma function:

$$B(\delta, \alpha) = \frac{\prod_{k=1}^R \Gamma(\delta^{-1} \alpha_k)}{\Gamma(\sum_{k=1}^R \delta^{-1} \alpha_k)} \quad (3)$$

The marginal Beta distribution of the Dirichlet distribution ($\text{Beta}(\alpha_k, \sum_{k=1}^R \alpha_k - \alpha_k)$) has mean

$$E(p_j) = \frac{\delta^{-1} \alpha_j}{\sum_{k=1}^R \delta^{-1} \alpha_k} = \frac{\alpha_j}{\sum_{k=1}^R \alpha_k}. \quad (4)$$

The covariance between the choice probabilities for restaurants j and m is given by

$$\text{Cov}(p_j, p_m) = \frac{-\alpha_j \alpha_m}{(\sum_{k=1}^R \alpha_k)^2 (\sum_{k=1}^R \delta^{-1} \alpha_k + 1)}, \quad j \neq m \quad (5)$$

When $R = 2$, the Dirichlet reduces to the Beta distribution.

Figure 2 shows distributions over p_j for different values of δ (for simplicity, $\alpha_k = \alpha = 1$, $E(p_j) = 0.5$). When all components of $\delta^{-1} \alpha$ are equal to 1, the Dirichlet distribution reduces to the uniform distribution over the probability simplex. When the components of $\delta^{-1} \alpha$ are all greater than 1, the density is unimodal, and when the components of $\delta^{-1} \alpha$ are all less than 1, the density has sharp peaks at the boundaries. As $\delta \rightarrow 0$, the distribution over p degenerates to a constant, i.e. there is no uncertainty about the quality left and we're back in the perfect information case with the multinomial distribution.

[Figure 2 about here.]

Up to now I only considered a single market. To uncover parameters of the distribution, I will need to estimate across multiple markets, where restaurants with identical ratings are assumed to be identical in terms of expected quality (both within as well as across markets). This may sound simplistic and reductionist, but actually reproduces the choice problem faced by individuals: Before actually having eaten at a restaurant, a rating and some rudimentary information such as a price range is all an individual can base his decision on.

Consider a set of markets $m = 1, 2, \dots, M$. For the m th market, there is an associated vector $p^{(m)}$ of length $k = R$ for the probabilities of going to each restaurant. Suppose that one can model these M probability vectors as coming from a $\text{Dir}(\delta^{-1} \alpha)$ distribution and that we have N_m samples from the m th probability vector. This prior distribution is then compounded with the multinomial distribution. The resulting $\{a_m\}$ are realizations of a Dirichlet-multinomial distribution.

The $m = 1, 2, \dots, M$ sets of samples $\{a_m\}$ drawn from the M probability mass functions drawn from the $\text{Dir}(\delta^{-1}\alpha)$ distribution are conditionally independent given $\delta^{-1}\alpha$, so the likelihood of $\delta^{-1}\alpha$ can be written as the product

$$\Pr\{A = a\} = \prod_{m=1}^M \Pr\{A_m = a_m | \delta^{-1}\alpha\} \quad (6)$$

where the compounded, unconditional distribution of A_m is obtained by integrating over p :

$$\begin{aligned} \Pr\{A_m = a_m | \alpha\} &= \int \Pr\{A_m = a_m | p\} f(p | \alpha) dp \\ &= \frac{N_m!}{\prod_{k=1}^R a_{mk}!} \frac{\Gamma(\sum_{k=1}^R \delta_m^{-1} \alpha_k)}{\Gamma(N_m + \sum_{k=1}^R \delta_m^{-1} \alpha_k)} \prod_{k=1}^R \frac{\Gamma(a_{mk} + \delta_m^{-1} \alpha_k)}{\Gamma(\delta_m^{-1} \alpha_k) a_{mk}!} \\ &= \frac{NB(\delta^{-1}A_m, N_m)}{\prod_{k: a_{mk} > 0} a_k B(\delta_m^{-1} \alpha_k, a_{mk})} \end{aligned} \quad (7)$$

where $NB()$ denotes the negative binomial distribution. By substituting Equation (7) into Equation (6), one obtains the likelihood of the observed data (see Equation 12 in the estimation section).

Using the law of iterated expectations and the law of total variance, the Dirichlet multinomial can be shown to have expected value, variance and covariance given by (Ng et al., 2011)

$$\mathbb{E}(a_{jm}) = \mathbb{E}[\mathbb{E}(a_{jm} | p_{jm})] = N_m \mathbb{E}(p_{jm}) = \frac{N_m \alpha_j}{\sum_{k=1}^R \alpha_k}, \quad (8)$$

$$\begin{aligned} \text{Var}(a_{jm}) &= \text{Var}(\mathbb{E}[a_{jm} | p_{jm}]) + \mathbb{E}[\text{Var}(a_{jm} | p_{jm})] \\ &= \mathbb{E}(N_m p_{jm} (1 - p_{jm})) + \text{Var}(N_m p_{jm}) \\ &= N_m \frac{\alpha_{jm}}{\sum_{k=1}^R \alpha_k} \left(1 - \frac{\alpha_j}{\sum_{k=1}^R \alpha_k} \right) (1 + (N_m - 1) \rho_m) \end{aligned} \quad (9)$$

where $\rho = 1/(\delta_m^{-1} \sum_{k=1}^R \alpha_k + 1)$ is an overdispersion parameter, inflating the variance by a factor $(1 + (N_m - 1)\rho_m)$ vis-à-vis the multinomial distribution. Note that α_j is restaurant-specific, while δ_m is market specific. The expected value of the Dirichlet-multinomial distribution is independent of δ ; in the case of the variance, δ enters through ρ . In the case of perfect information and no uncertainty about payoffs, $\delta \rightarrow 0$, $\rho \rightarrow 1$ and the overdispersion disappears.

It is important to point out that the social interactions as modeled here could be interpreted both as a form of information transmission or individuals simply having a preference to be

surrounded by other individuals (or a combination of the two). In the former case individuals indirectly find out about the quality of a restaurant, while in the latter case the presence of others itself becomes an attractive feature of a restaurant. When only looking at choices on the level of individual restaurants, both explanations are observationally equivalent in the sense of Ellison and Glaeser (1997).

4.3 Parameters of interest

To see the role previous choices play in this process, it is instructive to look at the conditional probability of the $(N+1)$ st individual to eat at restaurant j given the first N individuals have led to the allocation vector $A_N = a_N$. For notational ease, I drop the m -subscript and go back to focussing only on a single market. If $(A|p) \sim \text{Multinomial}_k(N, p)$ and $P \sim \text{Dir}(\delta^{-1}\alpha)$, then $(P|A = a) \sim \text{Dir}(\delta^{-1}\alpha + a)$. Based on this, the conditional probability given the first N individuals can be shown to be equal to

$$\begin{aligned} Pr\{d_{N+1} = j|A = a\} &= \frac{\delta^{-1}\alpha_j + a_j}{N + \sum_{k=1}^R \delta^{-1}\alpha_k} \\ &= \frac{\delta^{-1}\alpha_j}{\sum_{k=1}^R \delta^{-1}\alpha_k} \frac{\sum_{k=1}^R \delta^{-1}\alpha_k}{N + \sum_{k=1}^R \delta^{-1}\alpha_k} + \frac{a_j}{N} \frac{N}{N + \sum_{k=1}^R \delta^{-1}\alpha_k} \quad (10) \\ &= wE(p_j) + (1-w)\frac{a_j}{N} \end{aligned}$$

where $w = \frac{\sum_{k=1}^R \delta^{-1}\alpha_k}{N + \sum_{k=1}^R \delta^{-1}\alpha_k}$. The last step in Equation (10) decomposes the probability into a weighted average of a prior probability and a likelihood component. The first line in Equation (10) nicely illustrates how information transmission depends on α and δ . The parameter vector α serves as a (conditional) quality measure (higher quality restaurants have higher α_j), which is scaled by δ on a market level. The conditional probability also illustrates an important property of the Dirichlet-multinomial distribution: While the individuals' decisions are (clearly) not independent, they are exchangeable, meaning that the order in which individuals choose is irrelevant.

A higher α_j increases the expected probability of an individual visiting the restaurant relative to restaurants with a lower rating, *ceteris paribus*. If social interactions are strong enough though, these relative differences may become irrelevant. This is reflected in the δ parameter. δ does not influence the marginal expectation of any alternative, but does influence the marginal variance, capturing social interactions. As $\delta \rightarrow 0$ ($\delta^{-1} \sum_{k=1}^R \alpha_k \rightarrow \infty$), the Dirichlet multinomial converges to the multinomial distribution with a constant p vector and independent choices. Differences in choice probabilities are entirely determined by differences in ratings reflected in α_j - reliance on social interactions becomes irrelevant.

$\delta \rightarrow \infty$, on the other hand, leads to a sparse distribution of p and a process heavily deter-

mined by social interactions (i.e. almost all people go or don't go to a particular restaurant, irrespective of quality or other attributes). In other words, the first individual flips a coin and chooses a restaurant at random, while all others following in the sequence adopt the first individual's choice.⁵ Outcomes become highly skewed, as they are heavily influenced by strong but a priori unpredictable social interactions. The α vector completely loses its predictive power. A special case is $\delta^{-1}\alpha_j = 1 \forall j$ with a uniform distribution over outcomes, i.e. all possible outcome allocations are a priori equally likely (also known as the Bose-Einstein distribution, discussed in De Vany and Walls (1996)).

4.4 Grouping restaurants by quality

Up to this point, the model assumed that each of the R restaurants represents a distinct quality level. When estimating the model, this will not be the case, as quality will be measured by a rating with a 9-point scale only. I will estimate the model on the individual restaurant level, assigning the same α_j parameter to restaurants with the same rating, making restaurants with the same rating are a priori indistinguishable from each other. This may seem like an oversimplification, but for the ratings to be of any value a high rating in one place should translate into an approximately equivalent signal for quality in another place (conditional on location and price category). If ratings and their informational value are idiosyncratic, they are of little help.

Alternatively, I also aggregate restaurants by their quality and estimate the model, taking these aggregated restaurant groups as the observational unit. With respect to the Dirichlet-multinomial distribution, partitioning $\{1, 2, \dots, R\}$ restaurants into $\{C_1, C_2, \dots, C_s\}$ with $s < R$ is straightforward, as

$$\left(\sum_{i \in C_1} P_i, \sum_{i \in C_2} P_i, \dots, \sum_{i \in C_s} P_i\right) \sim \text{Dir}\left(\sum_{i \in C_1} \alpha_i, \sum_{i \in C_2} \alpha_i, \dots, \sum_{i \in C_s} \alpha_i\right)$$

due to the Dirichlet's aggregation property (Frigyik et al., 2010). By definition then, the α_j in the aggregated model should be at least as high as the in the model working with individual restaurants.

4.5 Estimation

The model derived in the previous section can be estimated by maximum likelihood (Guimaraes and Lindrooth, 2007). Main results will focus on markets defined on a ZIP code and pricerange

⁵In such a case, the model would degenerate to a constant-only conditional logit model with restaurant fixed effects - which could be interesting from a marketing perspective (e.g. is the constant for restaurant A higher than the constant for restaurant B), but not interesting in the context of social interactions where individuals only have limited access to information. Also, overdispersion would be ignored.

level, thereby assuming information is exchanged within a given price range and limited geographic area (i.e. social interactions are limited within that group). Since I do not observe any individual-specific information, actual restaurant choices can be aggregated into a vector of counts $a_m = \{a_{1m}, \dots, a_{Rm}\}$ without any loss of information. Consequently, individuals in a given market face the same choice set with identical choice attributes. As noted previously, individual visits are not independent but exchangeable, making the counts independent of the ordering of individuals in the queue while maintaining the sequential nature of the choice problem.

As shown by Guimaraes and Lindrooth (2007), modelling p_{jm} as a random variable is equivalent to introducing unobservable market-specific effects that equally influence the decisions of all individuals belonging to the same market. These market-specific effects will induce correlation across individuals in the same market, which in turn leads to overdispersion in the a_{km} count. The choice probability of individual i in market m selecting choice j (conditional on the group random effects) is defined as

$$p_{ijm} = \frac{\exp(\beta' \mathbf{x}_j + \eta_{jm})}{\sum_{k=1}^R \exp(\beta' \mathbf{x}_k + \eta_{km})} = \frac{\alpha_j \exp(\eta_{jm})}{\sum_{k=1}^R \alpha_k \exp(\eta_{km})} \quad (11)$$

where $\alpha_j = \exp(\beta' \mathbf{x}_j)$, \mathbf{x}_j are observable characteristics of choice j , η_{jm} are the random effects that affect members in market m and ε_{ijm} are assumed to be independent conditional on the market random effects. Assume that the random market effects $\exp(\eta_{jm})$ are i.i.d. gamma distributed with parameters $\{\delta_m^{-1} \alpha_j, \delta_m^{-1} \alpha_j\}$ where $\delta_m > 0$ is a market-specific parameter. Then, $\exp(\eta_{jm})$ has unit expectation and a variance equal to $\delta_m \alpha_j^{-1}$. Moreover, the variables defined by the product $\alpha_j \exp(\eta_{jm})$ also follow independent gamma distributions with parameters $\{\delta_m^{-1} \alpha_j, \delta_m^{-1}\}$. Given that all these variables follow independent gamma distributions with the same scale parameter, the vector $p = \{p_{1m}, \dots, p_{Rm}\}$ follows a Dirichlet distribution with a density as defined in Equation 2 (Mosimann, 1962; Guimaraes and Lindrooth, 2007).

Compounding the Dirichlet with the multinomial and involving all markets M leads to the unconditional likelihood function

$$\begin{aligned} L_{DM} &= \prod_{m=1}^G \int \prod_{k=1}^R a_m! \frac{p_{km}^{a_{km}}}{a_{km}!} f_{DM}(p_{1m}, \dots, p_{R-1m}) dp_1 \dots dp_{R-1} \\ &= \prod_{m=1}^G \frac{a_m! \Gamma(\delta_m^{-1} \sum_{k=1}^{R_m} \alpha_k)}{\Gamma(\delta_m^{-1} \sum_{k=1}^{R_m} \alpha_k + a_m)} \prod_{k=1}^R \frac{\Gamma(\delta_m^{-1} \alpha_k + a_{km})}{\Gamma(\delta_m^{-1} \alpha_k) a_{km}!} \end{aligned} \quad (12)$$

where $p_R = 1 - \sum_{k=1}^{R-1} p_k$. If the market random effects η_{jm} have a variance of zero and the correlation coefficient tends to zero, the likelihood function of the Dirichlet multinomial collapses into the likelihood function of the multinomial logit model (or grouped conditional

logit model). Testing for the existence of social interactions (i.e. testing for $\delta_m > 0$) can therefore be implemented directly via a likelihood ratio test. Note that the null hypothesis for the test is in the boundary of the parameter space, and therefore the correct p -value is one-half that which is obtained from the χ_1^2 (Self and Liang, 1987; Gutierrez et al., 2001).

Guimaraes and Lindrooth (2007) show that the likelihood in Equation 12 can be reformulated as the fixed effects negative binomial model developed by Hausman et al. (1984), which makes estimation of the Dirichlet multinomial model readily implementable in standard statistical software packages (see also Guimaraes, 2005). As shown by Guimaraes and Lindrooth (2007), Guimaraes (2005) and - in the context of fixed effects negative binomial models - by Allison and Waterman (2002), the group parameter δ_m can be modelled as a function of group-level variables denoted by \mathbf{w}_m , i.e. $\delta_m = f(\gamma' \mathbf{w}_m)$, which allows to identify the driving forces of the social interactions and the overdispersion parameter. This is in contrast to the multinomial logit model, where market fixed effects cancel out. In my case, I assume $-\ln(\delta_m) = \gamma' \mathbf{w}_m$. If those variables are restricted to a single constant, it is implicitly assumed that all markets share a common δ ($\delta_m = \delta$).⁶

5 Main results

This section is structured as follows: First, I present coefficient estimates for the different rating categories in three different models. The first model is the traditional multinomial logit, where the probability vector p is treated as a constant. This model serves as a benchmark. In the second model, I estimate the Dirichlet multinomial model enforcing $\delta_g = \delta$ across all markets. The last model relaxes this assumption and models δ_g as a function of market-level variables, which allows insights into what drives the correlation across individuals. More specifically, I assume $-\ln(\delta_g) = \gamma' w_g$. Both Dirichlet-multinomial models can be tested directly using a likelihood ratio test comparing the likelihood of the Dirichlet multinomial to the likelihood of the multinomial logit which imposes $\delta = 0$ (the resulting p -value should be halved, as outlined previously). Robustness checks using only single-period checkins rather than cumulative checkins, specifications aggregating restaurants by rating category as well as results using alternative market definitions can be found in the appendix.

In a second step, I back out the parameters of the Dirichlet multinomial model using the estimates of the model with constant δ . The parameters $\delta^{-1}\alpha$ can be interpreted in the

⁶Note that I deliberately abstain from using restaurant fixed effects in estimation. Using fixed effects would imply effects that are only unobserved by the researcher but not the individual, whereas in my case I explicitly want to allow for individuals making decisions in a setting of imperfect information. Also, if the user-provided ratings are informative enough, unobserved restaurant fixed effects should be reflected in those ratings (whether they are is one of the questions this paper addresses). Other unobserved factors such as characteristics of the neighborhood do not affect results as restaurants are grouped by location and price category and all restaurants in a group would be equally affected by such factors.

context of the prior distribution, where the distribution over p converges to a constant as $\delta^{-1}\alpha \rightarrow \infty$, consequently indicating a high informational value for the Yelp ratings. Low parameter values, on the other hand, indicate a stronger role of social interactions.

[Table 2 about here.]

Main results are summarized in Table 2, where the multinomial logit model and two Dirichlet multinomial models are estimated using cumulative checkins at three different points in time. Main results focus on markets defined by both ZIP code and price category, i.e. each market is a unique ZIP x Price combination, resulting in 236 markets (robustness checks use different market definitions). Results for different time periods are estimated separately; note that since I use cumulative checkins for the main results, observations for later periods include checkins from all previous periods.

Coefficients on the rating variable are all positive and increasing, which is expected. Coefficients in the grouped conditional logit model are all highly significant at all points in time, whereas in the Dirichlet model the coefficient on the second rating category is not statistically significant at any conventional significance level. Put differently: Using the Dirichlet model, a higher rating does not correlate with higher choice probabilities if this rating is in the bottom half of the scale. Standard errors increase roughly by a factor of 4 or 5 compared to the multinomial logit model, depending on the period considered.

Focussing on the *Dir II* model, one can see what drives within-market social interactions (note that $\delta_g = \exp(-\gamma'w_g)$). Higher price categories, more competitors and higher income lead to a relatively stronger role for social interactions, while the total number of reviews in a market weakens them - which could be explained by ratings becoming more valuable as they are based on a larger number of opinions. Coefficients on the income inequality proxy and population are not significant. It is noteworthy that when including market level variables, the coefficient on the constant is close to zero and statistically insignificant. This opposed to the *Dir I* model where the coefficient on the constant is highly significant, suggesting that market level variables capture a large part of the social interaction parameter δ .

Comparison of the models using loglikelihood values, Pearson chi-square statistics and AIC statistics all clearly favor the Dirichlet models (note that Pearson statistics have been adjusted accordingly for the Dirichlet-multinomial; see Guimaraes and Lindrooth (2007); Mosimann (1962)). Interestingly, all these statistics change little over time in the case of the Dirichlet multinomial model, while they get progressively worse in the multinomial logit model as checkins accumulate over longer periods of time.

These statistics lead to the conclusion that social interactions (or previous checkins) do play a role in the decision process of individuals. The follow-up question then is: how much so?

In order to answer this question, I back out the parameter vectors α and δ of the *Dir I* model and report them in Table 3. I also report the overdispersion parameter, ρ . The top half of the table reports parameters for the cumulative checkins, the bottom half for single period checkins.

[Table 3 about here.]

The Dirichlet's $\alpha = \exp(x'\beta)$ parameters are a unidimensional function of a rating dummy, where higher β_j leads to higher α_j which in turn leads to a higher expected value of p_j , the probability of choosing a restaurant with rating j in a market. The α vector is then scaled by δ^{-1} , where $delta = \exp(-\gamma_0)$ and γ_0 refers to the constant, defining the Dirichlet parameters.⁷ The existence of social interactions can be tested directly by testing the null hypothesis $\delta = 0$ or performing a likelihood ratio test, where the multinomial logit model serves as the restricted model. As already apparent in the statistics reported in the footer of Table 2, the hypothesis of $\delta = 0$ is clearly rejected.

As outlined in the model section, a Dirichlet parameter of 1 results in a uniform distribution of p (the Bose-Einstein distribution). Parameters above 1 result in unimodal peaks at the expected value of p , while parameters below 1 lead to sparse distributions with peaks at the edges. Parameters are reported in Table 3. Figure 3 plots the densities of p for the different rating categories listed in Table 3.

As one can see from both the table and the figure, the Dirichlet parameter for highest rating category is the only one above 1 in the case of cumulative checkins. All other parameters are below one, indicating that little can be learned about the number of checkins from those ratings as checkins are mainly governed by a priori unpredictable social interactions. In the case of cumulative checkins, the combinations $\delta^{-1}\alpha$ stay roughly constant, even though both components increase over periods.

Single period checkins for the years 2013 and 2014 are displayed in the bottom half of the table and are based on the result in Table 5 in the appendix. 2013 results use 1-year lags, while 2014 results use 2-year lags. Generally, all parameters are lower than in the cumulative case, indicating that focussing only on more recent checkins and discarding checkins in the more distant past, the importance of social interactions only rises relative to the importance of ratings.

[Figure 3 about here.]

What one can learn from the preceding table and figures is that while the quality reflected

⁷ $delta = \exp(-\gamma'w_g)$ in the *Dir II* model.

in the ratings do impact expected choice probabilities, the final allocation of checkins is far from certain. They give better-quality restaurants a headstart, but the choice probabilities within a rating category and their ultimate outcomes still vary wildly as individuals are responding strongly to signals of others. Intuitively, this social dynamic is best explained by the mechanism of Polya’s urn: Each time an individual chooses a restaurant from the pool of possible candidates, the choice probability of that restaurant increases for the individuals choosing later in the sequence. As time passes, conditional choice probabilities (as defined in Equation 10) are less governed by the $\delta^{-1}\alpha_j$ parameters, and more by the previous choices of others.

6 Robustness checks

I perform a number of robustness checks. First, by the aggregation property of the Dirichlet distribution, restaurants can be aggregated and assigned to partitions (as mentioned in the Model section). For the results in Table 4, restaurants within a given ZIP code and price category are grouped by their rating, resulting in one count per rating in a group. By the properties of the Dirichlet distribution, aggregating should increase $\delta_g^{-1}\alpha_j$ and - as the number of choices within a group is decreased - mechanically decrease correlation across individuals, which is proportional to the number of choice alternatives. This is shown in Table 4. Results stay qualitatively the same, and even though you observe an increase in the parameters, social interactions still play a large role.

[Table 4 about here.]

Second, to address potential concerns of endogeneity between ratings and visits, Table 5 shows results where only checkins within a given year are used in combination with lagged ratings.⁸ Single-period visits have been obtained by differencing the cumulative checkins (which is why the 2010-2012 is not covered). The 2013 specification uses $(t - 1)$ lags, while the 2014 specification uses both $(t - 1)$ and $(t - 2)$ lags. Results stay qualitatively the same as the previously reported results. The exception is the coefficient for the second rating category, which becomes even weaker as before. Again, correlation across choices is highly significant, strongly favoring the Dirichlet multinomial to the multinomial logit model. Also note that coefficients increase from 2013 to 2014 when using 1-period lags, while staying constant when comparing 2013 results to 2014 results using $(t - 2)$ lags. Again, loglikelihood, Pearson and AIC statistics for the Dirichlet models show little change across periods, while they worsen quite significantly in the case of the multinomial logit.

[Table 5 about here.]

⁸Note that ratings themselves change only little over time, and if they do mostly in the early phase when the rating is based only on few user feedbacks.

As a last robustness check, I estimate the model under a wider grouping structure. Grouping restaurants by ZIP code and price category assumes away wider spread social interactions that might exist, either across ZIP codes or across price categories. Table 6 in the appendix therefore presents results of the three models when the grouping is changed to ZIP code (i.e. across different price categories), to a citywide group and to a group defined on the city and price category level (instead of the ZIP and price category level). The latter leads to little changes compared to previous estimations as shown in the middle section of the table, whereas extending the groups to multiple price categories results in lower correlation within group, which is expected. Since price categories now change within group, price variables as well as interaction variables between price category and ratings and price category and group-level income are included. Interestingly, price categories have no significant effect in the Dirichlet models.

[Table 6 about here.]

7 Discussion

The aim of this paper is to explore the role of information transmission through social interactions in the choice process of individuals in a highly competitive market with a wide range of hard to distinguish alternatives, exemplified by the restaurant market. The presence of social interactions leads to correlation across individuals, highly skewed allocations and generally more uncertainty in the prediction of economic outcomes. Here, social interactions are embedded in a simple but insightful Dirichlet multinomial framework. The choice of the model turns out to be superior to the traditional multinomial logit approach based on a number of different measures as well as simulations.

Social interactions are found to be present in Yelp’s restaurant data, and can be modelled largely as a function of variables relating to aggregate information exchange variables, while socioeconomic variables such as income or income heterogeneity within a market play a negligible role as a driver of correlation. Results are robust against both restricting choices to a single period and lagged ratings as well as to widening the market definition and aggregating restaurants into their rating levels. I also find that user-provided ratings only have a limited impact on individuals’ decisions, as interpreted from the magnitude of the parameters of the Dirichlet-multinomial distribution. Generally, ratings at the top end of the scale are more informative than ratings at the bottom of the scale.

This paper only deals with counts aggregated on a restaurant level and with no information on the level of the individuals themselves. It is therefore important to note that I cannot provide details on the exact definition of the social interactions. Specifically, the social interactions found here could both be a form of information transmission or individuals simply having

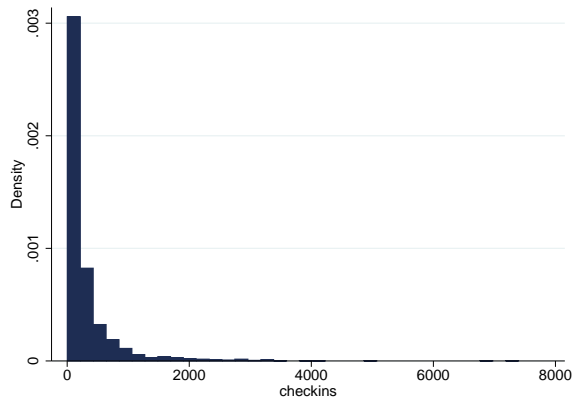
a preference to be surrounded by other individuals (or a combination of the two). In the former case individuals indirectly find out about the quality of a restaurant, while in the latter case the presence of others itself becomes an attractive feature of a restaurant. When only looking at choices on the level of individual restaurants, both explanations are observationally equivalent in the sense of Ellison and Glaeser (1997).

References

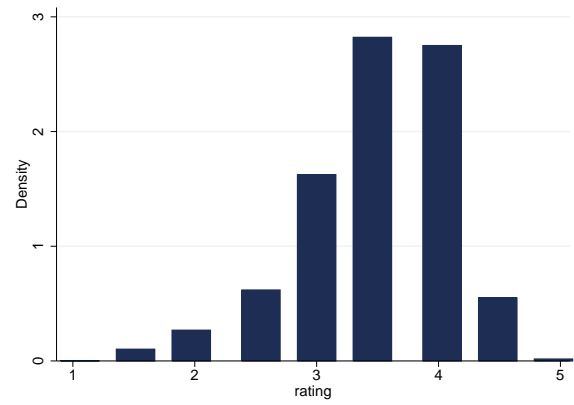
- Aldous, D. J. (1985). *Exchangeability and related topics*. Springer.
- Allison, P. D. and Waterman, R. P. (2002). Fixed effects negative binomial regression models. *Sociological Methodology*, 32(1):247–265.
- Beck, J. (2007). The sales effect of word of mouth: a model for creative goods and estimates for novels. *Journal of Cultural Economics*, 31(1):5–23.
- Becker, G. S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy*, 99(5):1109.
- Beinhocker, E. D. (2007). *The origin of wealth: Evolution, complexity, and the radical re-making of economics*. Random House.
- De Vany, A. and Walls, W. D. (1996). Bose-einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal*, pages 1493–1514.
- Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *The Journal of Political Economy*, 105(5):889–927.
- Frigyik, B. A., Kapila, A., and Gupta, M. R. (2010). Introduction to the dirichlet distribution and related processes. *Department of Electrical Engineering, University of Washington, UWEETR-2010-0006*.
- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions*. *The Quarterly journal of economics*, 111(2):507–548.
- Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness. *American journal of sociology*, pages 481–510.
- Guimaraes, P. (2005). A simple approach to fit the beta-binomial model. *Stata Journal*, 5(3):385–394.
- Guimaraes, P. and Lindrooth, R. C. (2007). Controlling for overdispersion in grouped conditional logit models: A computationally simple application of dirichlet-multinomial regression. *The Econometrics Journal*, 10(2):439–452.
- Gutierrez, R. G., Carter, S., and Drukker, D. M. (2001). On boundary-value likelihood-ratio tests. *Stata Technical Bulletin*, 10(60).
- Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r & d relationship. *Econometrica: Journal of the Econometric Society*, pages 909–938.

- Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.
- Lindley, D. V. (1961). Dynamic programming and decision theory. *Applied Statistics*, pages 39–51.
- Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp. com. *Com (September 16, 2011)*. *Harvard Business School NOM Unit Working Paper*, (12-016).
- McFadden, D. (1974). Conditional logit analysis of qualitative choices. *Zarembka. P.(eds): Frontiers*.
- McKenzie, J. (2008). Bayesian information transmission and stable distributions: Motion picture revenues at the australian box office*. *Economic Record*, 84(266):338–353.
- Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika*, pages 65–82.
- Ng, K. W., Tian, G.-L., and Tang, M.-L. (2011). *Dirichlet and related distributions: Theory, methods and applications*, volume 888. John Wiley & Sons.
- Ormerod, P., Tarbush, B., and Bentley, R. A. (2012). Social network markets: the influence of network structure when consumers face decisions over many similar choices. *arXiv preprint arXiv:1210.1646*.
- Pauly, M. V. and Satterthwaite, M. A. (1981). The pricing of primary care physicians services: a test of the role of consumer information. *The Bell Journal of Economics*, pages 488–506.
- Potts, J., Cunningham, S., Hartley, J., and Ormerod, P. (2008). Social network markets: a new definition of the creative industries. *Journal of cultural economics*, 32(3):167–185.
- Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856.
- Sela, A. and Berger, J. (2012). Decision quicksand: how trivial choices suck us in. *Journal of Consumer Research*, 39(2):360–370.
- Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.
- Sorensen, A. T. (2006). Social learning and health plan choice. *The Rand Journal of Economics*, 37(4):929–945.
- Vega-Redondo, F. (2007). *Complex social networks*. Number 44. Cambridge University Press.

Figures



(a) Checkins distribution as of 2015



(b) Ratings distribution as of 2015

Figure 1: Restaurant visits and ratings as of 2015

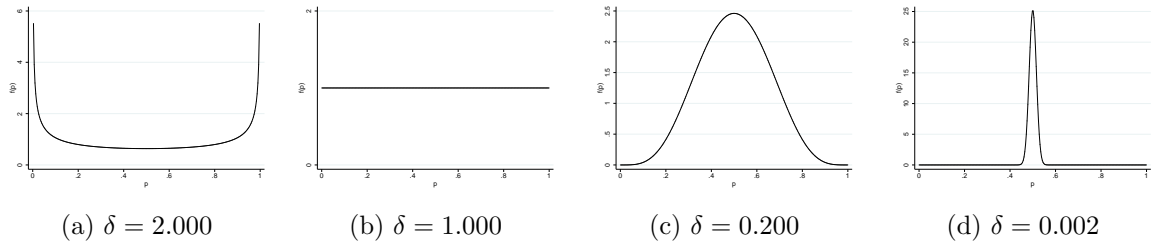


Figure 2: Dirichlet distributions for a set of different δ

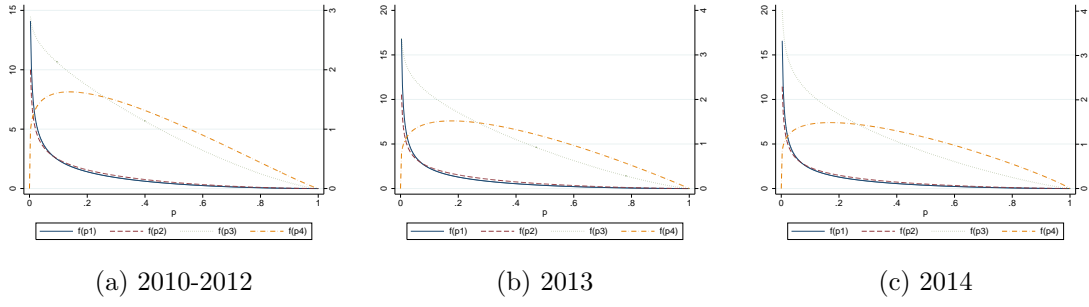


Figure 3: Dirichlet parameters for different rating categories, over time

Tables

Table 1: Summary Statistics

| | <i>2010-2012</i> | | <i>2013</i> | | <i>2014</i> | |
|---------------------|------------------|----------|-------------|----------|-------------|----------|
| | Mean | SD | Mean | SD | Mean | SD |
| CUMULCHECKINS | 125.06 | (218.77) | 210.05 | (363.23) | 282.42 | (480.47) |
| CHECKINS (1 period) | 125.06 | (218.77) | 84.87 | (154.08) | 72.41 | (128.49) |
| RATING1 | 0.01 | (0.11) | 0.01 | (0.11) | 0.01 | (0.11) |
| RATING2 | 0.10 | (0.30) | 0.10 | (0.29) | 0.10 | (0.30) |
| RATING3 | 0.47 | (0.50) | 0.50 | (0.50) | 0.51 | (0.50) |
| RATING4 | 0.41 | (0.49) | 0.39 | (0.49) | 0.38 | (0.49) |
| PRICE | 1.56 | (0.59) | 1.56 | (0.59) | 1.56 | (0.59) |
| LNSUMREVIEWS | 6.51 | (1.51) | 6.84 | (1.47) | 7.11 | (1.44) |
| LNSUMREVIEWS, adj. | 3.55 | (0.78) | 3.88 | (0.74) | 4.15 | (0.72) |
| LNCOMP | 2.96 | (0.95) | 2.96 | (0.95) | 2.96 | (0.95) |
| INEQ | 1.36 | (0.21) | 1.36 | (0.21) | 1.36 | (0.21) |
| LNINC | 11.14 | (0.37) | 11.14 | (0.37) | 11.14 | (0.37) |
| LNPOP | 10.34 | (0.68) | 10.34 | (0.68) | 10.34 | (0.68) |
| Observations | 3'159 | | 3'169 | | 3'170 | |

Standard errors in parentheses. *Price*, *lncompetitors*, *lninc* and *lnpop* are constant over time and within ZIP.

Lnsunreviews and *lnsunreviewsadj* are calculated on the Price x ZIP level.

Lnsunreviewsadj is the sum of reviews in the group divided by the number of restaurants in the group.

Source : Yelp, US Census. Own calculations.

Table 2: Cumulative checkins over time

| <i>Variable</i> | 2010-2012 | | | 2013 | | | 2014 | | |
|--------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|
| | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> |
| RATING2 | 0.537 (0.041) | 0.178 (0.160) | 0.248 (0.673) | 0.866 (0.037) | 0.277 (0.163) | 0.333 (0.162) | 0.634 (0.033) | 0.229 (0.162) | 0.309 (0.161) |
| RATING3 | 1.400 (0.040) | 0.576 (0.154) | 0.673 (0.153) | 1.787 (0.036) | 0.663 (0.157) | 0.739 (0.156) | 1.673 (0.032) | 0.611 (0.156) | 0.707 (0.155) |
| RATING4 | 1.846 (0.040) | 0.795 (0.154) | 0.881 (0.153) | 2.299 (0.036) | 0.956 (0.156) | 1.023 (0.156) | 2.240 (0.032) | 0.939 (0.155) | 1.026 (0.156) |
| <i>Group level</i> | | | | | | | | | |
| PRICE2 | | | 0.301 (0.068) | | | 0.268 (0.068) | | | 0.217 (0.068) |
| PRICE3 | | | 0.727 (0.160) | | | 0.603 (0.155) | | | 0.454 (0.152) |
| REVIEWS | | | -0.408 (0.059) | | | -0.394 (0.060) | | | -0.344 (0.060) |
| LNCOMP | | | 0.479 (0.089) | | | 0.458 (0.089) | | | 0.375 (0.088) |
| INEQ | | | -0.171 (0.170) | | | -0.118 (0.168) | | | -0.184 (0.166) |
| LNINC | | | 0.115 (0.067) | | | 0.010 (0.066) | | | 0.055 (0.066) |
| LNPOP | | | -0.064 (0.050) | | | -0.053 (0.050) | | | -0.067 (0.049) |
| CONST | | -0.631 (0.153) | 0.003 (1.004) | | -0.765 (0.156) | -0.012 (0.995) | | -0.771 (0.155) | 0.713 (0.987) |
| No. of markets | 236 | 236 | 236 | 239 | 239 | 239 | 238 | 238 | 238 |
| LogL | -188'783 | -15'740 | -15'696 | -310'566 | -17'289 | -17'253 | -416'735 | -18'144 | -18'113 |
| Pearson | 443'831 | 3'559 | 3'522 | 742'451 | 3'450 | 3'426 | 983'624 | 3'352 | 3'338 |
| AIC | 377'573 | 31'487 | 31'414 | 621'137 | 34'587 | 34'528 | 833'475 | 36'296 | 36'315 |

Source: Yelp. Dependent variable: Cumulative checkins of individual restaurants. Standard errors in parentheses. Markets are defined on a ZIPxPRICE level. Dirmul I refers to $\delta_g = \delta$, Dirmul II to $\delta_g = f(x_g) = \exp(-\gamma'w_g)$. *RATING* dummy variables reflect a 4-step scale. *PRICE* dummy variables reflect three price categories. *REVIEWS* is the logarithm of the sum of reviews in a market. *LNCOMP* is the logarithm of the number of competitors in a market. *INEQ* is the ratio of average income and median income in a ZIP code area, proxying income inequality. *LNINC* is the logarithm of average income in the ZIP code area. *LNPOP* is the logarithm of the ZIP population.

Table 3: Expectations and correlations across time

| <i>Variable</i> | 2010-2012 | | 2013 | | 2014 | |
|-------------------------------|------------|--------------|------------|--------------|------------|--------------|
| | <i>MNL</i> | <i>Dir I</i> | <i>MNL</i> | <i>Dir I</i> | <i>MNL</i> | <i>Dir I</i> |
| <i>Cumulative checkins</i> | | | | | | |
| δ | | 1.879 | | 2.149 | | 2.162 |
| ρ | | 0.233 | | 0.239 | | 0.245 |
| $\delta^{-1}\alpha_1$ | | 0.532 | | 0.465 | | 0.462 |
| $\delta^{-1}\alpha_2$ | | 0.636 | | 0.614 | | 0.582 |
| $\delta^{-1}\alpha_3$ | | 0.947 | | 0.903 | | 0.852 |
| $\delta^{-1}\alpha_4$ | | 1.179 | | 1.210 | | 1.183 |
| <i>Single period checkins</i> | | | | | | |
| δ | | | | 2.868 | | 3.090 |
| ρ | | | | 0.336 | | 0.351 |
| $\delta^{-1}\alpha_1$ | | | | 0.349 | | 0.324 |
| $\delta^{-1}\alpha_2$ | | | | 0.355 | | 0.331 |
| $\delta^{-1}\alpha_3$ | | | | 0.512 | | 0.477 |
| $\delta^{-1}\alpha_4$ | | | | 0.758 | | 0.719 |

Source: Yelp. The upper half of the table shows results for cumulative checkins, the lower half shows results for single-period checkins in years 2013 and 2014. Dirmul I refers to $\delta_g = \delta = \exp(-constant)$. Calculations assume that exactly one restaurant of each rating category exist within a group and are based on the results in Table 2 and 5. Bottom half uses L1-lags for 2013 and L2-lags for 2014. δ measures the strength of interactions, with $\delta \rightarrow 0$ indicating no social interactions, while $\delta \rightarrow \infty$ presents a the limit case where visits depend exclusively on other individuals' previous choices. $\rho = 1/(1 + \sum_{j=1}^R \delta^{-1}\alpha_j)$ is the overdispersion parameter converging to 0 as $\delta \rightarrow 0$.

Table 4: Cumulative checkins over time, aggregated by rating

| <i>Variable</i> | 2010-2012 | | | 2013 | | | 2014 | | |
|---------------------|------------------|--------------------|-------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|
| | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> |
| RATING2 | 0.946 (0.041) | 0.354 (0.189) | 0.505 (0.188) | 1.502 (0.037) | 0.444 (0.186) | 0.566 (0.186) | 1.306 (0.033) | 0.433 (0.184) | 0.588 (0.185) |
| RATING3 | 3.522 (0.040) | 1.640 (0.184) | 1.908 (0.188) | 4.216 (0.036) | 1.787 (0.181) | 2.022 (0.188) | 4.058 (0.032) | 1.742 (0.179) | 1.992 (0.185) |
| RATING4 | 3.723 (0.040) | 1.895 (0.183) | 2.127 (0.188) | 4.403 (0.036) | 1.990 (0.181) | 2.199 (0.188) | 4.295 (0.032) | 1.948 (0.179) | 2.127 (0.184) |
| <i>Group level</i> | | | | | | | | | |
| PRICE2 | | | 0.003 (0.177) | | | -0.104 (0.176) | | | 0.057 (0.168) |
| PRICE3 | | | 0.193 (0.342) | | | 0.137 (0.349) | | | 0.087 (0.320) |
| REVIEWS | | | -0.409 (0.128) | | | -0.323 (0.132) | | | -0.447 (0.129) |
| LNCOMP | | | 0.911 (0.214) | | | 0.760 (0.216) | | | 0.811 (0.209) |
| INEQ | | | -0.446 (0.514) | | | -0.292 (0.529) | | | 0.016 (0.511) |
| LNINC | | | 0.002 (0.188) | | | 0.043 (0.196) | | | 0.158 (0.183) |
| LNPOP | | | -0.057 (0.127) | | | -0.088 (0.136) | | | 0.094 (0.126) |
| CONST | | - 0.948 (0.176) | 0.160 (2.698) | | -1.091 (0.172) | -0.260 (2.958) | | -1.132 (0.169) | -3.191 (2.796) |
| No. of observations | 592 | 592 | 592 | 598 | 598 | 598 | 593 | 593 | 593 |
| No. of markets | 220 | 220 | 220 | 223 | 223 | 223 | 221 | 221 | 221 |
| LogL | -37'723 | -2'172 | -2'157 | -63'836 | -2'377 | -2'365 | -89'520 | -2'480 | -2'468 |
| Pearson | 74'659 | 354 | 372 | 132'279 | 370 | 382 | 183'405 | 342 | 355 |
| AIC | 75'451 | 4'352 | 4'335 | 127'677 | 4'762 | 4'751 | 179'046 | 4'969 | 4'957 |

Source: Yelp. Dependent variable: Cumulative checkins, aggregated by rating within a market. Standard errors in parentheses. Markets are defined on a ZIPxPRICE level. Dirmul I refers to $\delta_g = \delta$, Dirmul II to $\delta_g = f(x_g) = \exp(-\gamma'w_g)$. *RATING* dummy variables reflect a 4-step scale. *PRICE* dummy variables reflect three price categories. *REVIEWS* is the logarithm of the sum of reviews in a market. *LNCOMP* is the logarithm of the number of competitors in a market. *INEQ* is the ratio of average income and median income in a ZIP code area, proxying income inequality. *LNINC* is the logarithm of average income in the ZIP code area. *LNPOP* is the logarithm of the ZIP population.

Table 5: Single period checkins over time

| <i>Variable</i> | 2013 | | | 2014 | | | | | |
|--------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|
| | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir II</i> | <i>Dir II</i> |
| LRATING2 | 0.524 (0.051) | 0.017 (0.182) | 0.062 (0.182) | 0.912 (0.072) | 0.217 (0.208) | 0.241 (0.208) | | | |
| LRATING3 | 1.470 (0.050) | 0.384 (0.174) | 0.466 (0.174) | 1.939 (0.070) | 0.617 (0.200) | 0.687 (0.200) | | | |
| LRATING4 | 2.052 (0.050) | 0.776 (0.174) | 0.853 (0.174) | 2.666 (0.070) | 1.048 (0.200) | 1.110 (0.201) | | | |
| L2RATING2 | | | | | | | 0.603 (0.040) | 0.022 (0.183) | 0.076 (0.182) |
| L2RATING3 | | | | | | | 1.558 (0.039) | 0.387 (0.175) | 0.481 (0.175) |
| L2RATING4 | | | | | | | 2.174 (0.039) | 0.798 (0.175) | 0.886 (0.175) |
| <i>Group level</i> | | | | | | | | | |
| PRICE2 | | | 0.025 (0.070) | | | 0.042 (0.026) | | | 0.047 (0.068) |
| PRICE3 | | | 0.112 (0.163) | | | -0.015 (0.052) | | | 0.059 (0.158) |
| LREVIEWS | | | -0.178 (0.061) | | | -0.148 (0.064) | | | |
| LCOMP | | | 0.140 (0.092) | | | 0.025 (0.095) | | | |
| L2REVIEWS | | | | | | | | | -0.171 (0.059) |
| L2COMP | | | | | | | | | 0.108 (0.089) |
| INEQ | | | -0.312 (0.181) | | | -0.317 (0.190) | | | -0.257 (0.179) |
| LNINC | | | 0.027 (0.071) | | | -0.080 (0.072) | | | -0.041 (0.069) |
| LNPOP | | | -0.067 (0.053) | | | -0.029 (0.055) | | | -0.061 (0.052) |
| CONST | | -1.054 (0.173) | 0.440 (1.041) | | -1.521 (0.198) | 0.993 (1.078) | | -1.128 (0.173) | 1.027 (1.014) |
| No. of markets | 237 | 237 | 237 | 238 | 238 | 238 | 236 | 236 | 236 |
| LogL | -147'888 | -14'139 | -14'111 | -134'399 | -13'390 | -13'359 | -268'138 | -15'753 | -15'626 |
| Pearson | 327'815 | 2'427 | 2'448 | 294'048 | 2'151 | 2'173 | 605'410 | 2'311 | 2'326 |
| AIC | 295'781 | 28'270 | 28'245 | 268'805 | 26'788 | 26'740 | 536'283 | 31'514 | 31'341 |

Source: Yelp. Dependent variable: Single-period checkins of individual restaurants. Standard errors in parentheses. Markets are defined on a ZIPxPRICE level. Dirmul I refers to $\delta_g = \delta$, Dirmul II to $\delta_g = f(x_g) = \exp(-\gamma'w_g)$. *RATING* dummy variables reflect a 4-step scale. *PRICE* dummy variables reflect three price categories. *REVIEWS* is the logarithm of the sum of reviews in a market. *LNCOMP* is the logarithm of the number of competitors in a market. *INEQ* is the ratio of average income and median income in a ZIP code area, proxying income inequality. *LNINC* is the logarithm of average income in the ZIP code area. *LNPOP* is the logarithm of the ZIP population.

Table 6: Cumulative checkins over time, higher grouping levels

| <i>Variable</i> | ZIP, 2014 | | | City X Price, 2014 | | | City, 2014 | | |
|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir I</i> | <i>Dir II</i> | <i>MNL</i> | <i>Dir II</i> | <i>Dir II</i> |
| RATING2 | 0.689 (0.032) | 0.219 (0.160) | 0.280 (0.159) | 0.677 (0.032) | 0.239 (0.157) | 0.316 (0.156) | 0.675 (0.032) | 0.209 (0.158) | 0.281 (0.158) |
| RATING3 | 1.646 (0.031) | 0.543 (0.154) | 0.622 (0.154) | 1.752 (0.031) | 0.604 (0.150) | 0.701 (0.150) | 1.685 (0.031) | 0.526 (0.153) | 0.612 (0.153) |
| RATING4 | 2.139 (0.031) | 0.817 (0.159) | 0.889 (0.158) | 2.313 (0.031) | 0.921 (0.151) | 1.011 (0.150) | 2.174 (0.031) | 0.780 (0.158) | 0.856 (0.158) |
| PRICE2 | 1.010 (0.082) | 0.123 (0.798) | 1.024 (0.949) | | | | 4.554 (0.127) | 3.753 (1.294) | 1.001 (1.660) |
| PRICE3 | -0.519 (0.190) | 1.134 (2.032) | 2.394 (2.115) | | | | 5.877 (0.267) | 5.355 (3.116) | 3.387 (3.278) |
| PRICERAT2 | 0.152 (0.004) | 0.122 (0.047) | 0.133 (0.046) | | | | 0.148 (0.004) | 0.133 (0.047) | 0.129 (0.046) |
| PRICERAT3 | -0.086 (0.013) | 0.002 (0.124) | 0.027 (0.122) | | | | 0.075 (0.013) | 0.082 (0.126) | -0.316 (0.295) |
| PRICEINC2 | -0.074 (0.007) | -0.017 (0.070) | -0.101 (0.084) | | | | -0.384 (0.011) | -0.343 (0.116) | -0.096 (0.149) |
| PRICEINC3 | 0.084 (0.016) | -0.088 (0.177) | -0.206 (0.186) | | | | -0.534 (0.023) | -0.488 (0.279) | -0.316 (0.295) |
| <i>Group level</i> | | | | | | | | | |
| PRICE2 | | | | | | 0.240 (0.116) | | | |
| PRICE3 | | | | | | 0.414 (0.208) | | | |
| REVIEWS | | | -0.349 (0.064) | | | -0.418 (0.126) | | | -0.548 (0.224) |
| LNCOMP | | | 0.385 (0.097) | | | 0.353 (0.154) | | | 0.565 (0.294) |
| INEQ | | | -0.182 (0.164) | | | 0.049 (0.184) | | | 0.203 (0.249) |
| LNINC | | | 0.168 (0.078) | | | -0.164 (0.149) | | | 0.142 (0.292) |
| LNPOP | | | -0.066 (0.049) | | | 0.045 (0.058) | | | -0.008 (0.081) |
| CONST | | -0.909 (0.154) | -0.543 (1.087) | | -0.859 (0.149) | 2.092 (1.793) | | -0.992 (0.153) | -0.464 (2.989) |
| No. of markets | 118 | 118 | 118 | 71 | 71 | 71 | 37 | 37 | 37 |
| LogL | -439'340 | -19'310 | -19'277 | -514'537 | -20'046 | -20'016 | -520'212 | -20'488 | -20'462 |
| Pearson | 1'085'552 | 3'630 | 3'610 | 1'413'224 | 4'288 | 4'271 | 1'433'497 | 4'371 | 4'350 |
| AIC | 878'697 | 38'641 | 38'584 | 1'029'080 | 40'099 | 40'054 | 1'040'443 | 40'997 | 40'954 |

Source: Yelp. Dependent variable: Cumulative checkins of individual restaurants. Standard errors in parentheses. Markets are defined on a ZIP level. Dirmul I refers to $\delta_g = \delta$, Dirmul II to $\delta_g = f(x_g) = \exp(-\gamma'w_g)$. *RATING* dummy variables reflect a 4-step scale. *PRICE* dummy variables reflect three price categories. *REVIEWS* is the logarithm of the sum of reviews in a market. *LNCOMP* is the logarithm of the number of competitors in a market. *INEQ* is the ratio of average income and median income in a ZIP code area, proxying income inequality. *LNINC* is the logarithm of average income in the ZIP code area. *LNPOP* is the logarithm of the ZIP population.

Appendix

Table 7: Overdispersion in restaurant checkins at different levels of aggregation

| <i>Level of aggregation</i> | <i>Global</i> | <i>ZIP</i> | <i>Price category</i> | <i>ZIP x Price</i> |
|-----------------------------|---------------|------------|-----------------------|--------------------|
| Mean | 282.42 | 282.42 | 282.42 | 282.42 |
| SD | 480.47 | 178.25 | 119.53 | 228.72 |
| Skew | 1.90 | 1.44 | 0.07 | 1.78 |
| Kurtosis | 42.87 | 5.35 | 1.05 | 6.20 |

Source : Yelp. 2015 data used, covering cumulative checkins across a 5 year period.