Reviews, Reputation, and Revenue: The Case of Yelp.com

Michael Luca

Working Paper 12-016

Reviews, Reputation, and Revenue: The Case of Yelp.com

Michael Luca
Harvard Business School

Working Paper 12-016

Reviews, Reputation, and Revenue: The Case of Yelp.com

Michael Luca[†]

Abstract

Do online consumer reviews affect restaurant demand? I investigate this question using a novel dataset combining reviews from the website Yelp.com and restaurant data from the Washington State Department of Revenue. Because Yelp prominently displays a restaurant's rounded average rating, I can identify the *causal* impact of Yelp ratings on demand with a regression discontinuity framework that exploits Yelp's rounding thresholds. I present three findings about the impact of consumer reviews on the restaurant industry: (1) a one-star increase in Yelp rating leads to a 5-9 percent increase in revenue, (2) this effect is driven by independent restaurants; ratings do not affect restaurants with chain affiliation, and (3) chain restaurants have declined in market share as Yelp penetration has increased. This suggests that online consumer reviews substitute for more traditional forms of reputation. I then test whether consumers use these reviews in a way that is consistent with standard learning models. I present two additional findings: (4) consumers do not use all available information and are more responsive to quality changes that are more visible and (5) consumers respond more strongly when a rating contains more information. Consumer response to a restaurant's average rating is affected by the number of reviews and whether the reviewers are certified as "elite" by Yelp, but is unaffected by the size of the reviewers' Yelp friends network.

[†] Harvard Business School, mluca@hbs.edu

1 Introduction

Technological advances over the past decade have led to the proliferation of consumer review websites such as Yelp.com, where consumers can share experiences about product quality. These reviews provide consumers with information about experience goods, which have quality that is observed only after consumption. With the click of a button, one can now acquire information from countless other consumers about products ranging from restaurants to movies to physicians. This paper provides empirical evidence on the impact of consumer reviews in the restaurant industry.

It is *a priori* unclear whether consumer reviews will significantly affect markets for experience goods. On the one hand, existing mechanisms aimed at solving information problems are imperfect: chain affiliation reduces product differentiation, advertising can be costly, and expert reviews tend to cover small segments of a market.[1] Consumer reviews may therefore complement or substitute for existing information sources. On the other hand, reviews can be noisy and difficult to interpret because they are based on subjective information reflecting the views of a non-representative sample of consumers. Further, consumers must actively seek out reviews, in contrast to mandatory disclosure and electronic commerce settings. [2]

How do online consumer reviews affect markets for experience goods? Using a novel data set consisting of reviews from the website Yelp.com and revenue data from the Washington State Department of Revenue, I present three key findings: (1) a one-star increase in Yelp rating leads to a 5-9 percent increase in revenue, (2) this effect is driven by independent restaurants; ratings do not affect restaurants with chain affiliation, and (3) chain restaurants have declined in

---

[1] For example, Zagat covers only about 5% of restaurants in Los Angeles, according to Jin and Leslie (2009).
[2] For an example of consumer reviews in electronic commerce, see Cabral and Hortacsu (2010). For an example of the impact of mandatory disclosure laws, see Mathios (2000), Jin and Leslie (2003), and Bollinger et al. (2010).

revenue share as Yelp penetration has increased. Consistent with standard learning models, consumer response is larger when ratings contain more information. However, consumers also react more strongly to information that is more visible, suggesting that the way information is presented matters.

To construct the data set for this analysis, I worked with the Washington State Department of Revenue to gather revenues for all restaurants in Seattle from 2003 through 2009. This allows me to observe an entire market both before and after the introduction of Yelp. I focus on Yelp because it has become the dominant source of consumer reviews in the restaurant industry. For Seattle alone, the website had over 60,000 restaurant reviews covering 70% of all operational restaurants as of 2009. By comparison, the Seattle Times has reviewed roughly 5% of operational Seattle restaurants.

To investigate the impact of Yelp, I first show that changes in a restaurant's rating are correlated with changes in revenue, controlling for restaurant and quarter fixed effects. However, there can be concerns about interpreting this as causal if changes in a restaurant's rating are correlated with other changes in a restaurant's reputation that would have occurred even in the absence of Yelp. This is a well-known challenge to identifying the causal impact of any type of reputation on demand, as described in Eliashberg and Shugan (1997).

To support the claim that Yelp has a *causal* impact on revenue, I exploit the institutional features of Yelp to isolate variation in a restaurant's rating that is exogenous with respect to unobserved determinants of revenue. In addition to specific reviews, Yelp presents the average rating for each restaurant, *rounded to the nearest half-star*. I implement a regression discontinuity (RD) design around the rounding thresholds, taking advantage of this feature. Essentially, I look for discontinuous jumps in revenue that follow discontinuous changes in

rating. One common challenge to the RD methodology is gaming: in this setting, restaurants may submit false reviews. I then implement the McCrary (2008) density test to rule out the possibility that gaming is biasing the results. If gaming were driving the result, then one would expect ratings to be clustered just above the discontinuities. However, this is not the case. More generally, the results are robust to many types of firm manipulation.

Using the RD framework, I find that a restaurant's average rating has a large impact on revenue - a one-star increase leads to a 5-9 percent increase in revenue for independent restaurants, depending on the specification. The identification strategy used in this paper shows that Yelp affects demand, but is also informative about the way that consumers use information. If information is costless to use, then consumers should not respond to rounding, since they also see the underlying reviews. However, a growing literature has shown that consumers do not use all available information (Dellavigna and Pollet 2007; 2010). Further, responsiveness to information can depend not only on the informational content, but also on the simplicity of calculating the information of interest (Chetty et al. 2009, Finkelstein 2009). Moreover, many restaurants on Yelp receive upward of two hundred reviews, making it time-consuming to read them all. Hence, the average rating may serve as a simplifying heuristic to help consumers learn about restaurant quality in the face of complex information.

Next, I examine the impact of Yelp on revenues for chain restaurants. As of 2007, roughly $125 billion per year is spent at chain restaurants, accounting for over 50% of all restaurant spending in the United States. Chains share a brand name (e.g., Applebee's or McDonald's), and often have common menu items, food sources, and advertising. In a market with more products than a consumer can possibly sample, chain affiliation provides consumers with information about the quality of a product. Because consumers have more information about chains than

about independent restaurants, one might expect Yelp to have a larger effect on independent restaurants. My results demonstrate that despite the large impact of Yelp on revenue for independent restaurants; the impact is statistically insignificant and close to zero for chains.

Empirically, changes in a restaurant's rating affect revenue for independent restaurants but not for chains. A standard information model would then predict that Yelp would cause more people to choose independent restaurants over chains. I test this hypothesis by estimating the impact of Yelp penetration on revenue for chains relative to independent restaurants. The data confirm this hypothesis. I find that there is a shift in revenue share toward independent restaurants and away from chains as Yelp penetrates a market.

Finally, I investigate whether the observed response to Yelp is consistent with Bayesian learning. Under the Bayesian hypothesis, reactions to signals are stronger when the signal is more precise (i.e., the rating contains a lot of information). I identify two such situations. First, a restaurant's average rating aggregates a varying number of reviews. If each review presents a noisy signal of quality, then ratings that contain more reviews contain more information. Further, the number of reviews is easily visible next to each restaurant. Consistent with a model of Bayesian learning, I show that market responses to changes in a restaurant's rating are largest when a restaurant has many reviews. Second, a restaurant's reviews could be written by high quality or low quality reviewers. Yelp designates prolific reviewers with "elite" status, which is visible to website readers. Reviews can be sorted by whether the reviewer is elite. Reviews written by elite members have nearly double the impact as other reviews.

This final point adds to the literature on consumer sophistication in responses to quality disclosure, which has shown mixed results. Scanlon et al. (2002), Pope (2009), and Luca and Smith (2010) all document situations where consumers rely on very coarse information, while

ignoring finer details. On the other hand, Bundorf et al (2009) show evidence of consumer sophistication. When given information about birth rates and patient age profiles at fertility clinics, consumers respond more to high birth rates when the average patient age is high. This suggests that consumers infer something about the patient mix. Similarly, Rockoff et al. (2010) provide evidence that school principals respond to noisy information about teacher quality in a way that is consistent with Bayesian learning. My results confirm that there is a non-trivial cost of using information, but consumers act in a way that is consistent with Bayesian learning, *conditional on easily accessible information.*

Overall, this paper presents evidence that consumers use Yelp to learn about independent restaurants but not those with chain affiliation. Consumer response is consistent with a model of Bayesian learning with information gathering costs. The introduction of Yelp then begins to shift revenue away from chains and toward independent restaurants.

The regression discontinuity design around rounding rules offered in this paper will also allow for identification of the causal impact of reviews in a wide variety of settings, helping to solve a classic endogeneity problem. For example, Amazon.com has consumer reviews that are aggregated and presented as a rounded average. RottenTomatoes.com presents movie critic reviews as either "rotten" or "fresh," even though the underlying reviews are assigned finer grades. Gap.com now allows consumers to review clothing; again, these reviews are rounded to the half-star. For each of these products and many more, there is a potential endogeneity problem where product reviews are correlated with underlying quality. With only the underlying reviews and an outcome variable of interest, my methodology shows how it is possible to identify the causal impact of reviews.

2 Data

I combine two datasets for this paper: restaurant reviews from Yelp.com and revenue data from the Washington State Department of Revenue.

2.1 Yelp.com

Yelp.com is a website where consumers can leave reviews for restaurants and other businesses. Yelp was founded in 2004, and is based in San Francisco. The company officially launched its website in all major west coast cities (and select other cities) in August of 2005, which includes Seattle. It currently contains over 10 million business reviews, and receives approximately 40 million unique visitors (identified by IP address) per month.

Yelp is part of a larger crowdsourcing movement that has developed over the past decade, where the production of product reviews, software, and encyclopedias, among others are outsourced to large groups of anonymous volunteers rather than paid employees. The appendix shows trends in search volumes for Yelp, Trip Advisor, and Angie's List, which underscores the growth of the consumer review phenomenon.

On Yelp, people can read restaurant reviews and people can write restaurant reviews. In order to write a review, a user must obtain a free account with Yelp, which requires registering a valid email address. The users can then rate any restaurant (from 1-5 stars), and enter a text review.

Once a review is written, anyone (with or without an account) can access the website for free and read the review. Readers will come across reviews within the context of a restaurant search, where the reader is trying to learn about the quality of different restaurants. Figure 1 provides a

snapshot of a restaurant search in Seattle. Key to this paper, readers can look for restaurants that exceed a specified average rating (say 3.5 stars). Readers can also search within a food category or location.

A reader can click on an individual restaurant, which will bring up more details about the restaurant. As shown in Figure 2, the reader will then be able to read individual reviews, as well as see qualitative information about the restaurants features (location, whether it takes credit cards, etc).

Users may choose to submit reviews for many reasons. Yelp provides direct incentives for reviewers, such as having occasional parties for people who have submitted a sufficiently large number of reviews. Wang (2010) looks across different reviewing systems (including Yelp) to analyze the social incentives for people who decide to submit a review.

2.2 Restaurant Data

I take the Department of Revenue data to be the full set of restaurants in the city of Seattle. The data contains every restaurant that reported earning revenue at any point between January 2003 and October 2009. The Department of Revenue assigns each restaurant a unique business identification code (UBI), which I use to identify restaurants. In total, there are 3,582 restaurants during the period of interest. On average, there are 1,587 restaurants open during a quarter. This difference between these two numbers is accounted for by the high exit and entry rates in the restaurant industry. Approximately 5% of restaurants go out of business each quarter.

Out of the sample, 143 restaurants are chain affiliated. However, chain restaurants tend to have a lower turnover rate. In any given quarter, roughly 5% of restaurants are chains. This can be compared to Jin and Leslie (2009), who investigate chains in Los Angeles. Roughly 11% of

restaurants in their sample are chains. Both of these cities have substantially smaller chain populations than the nation as a whole, largely because chains are more common in rural areas and along highways.

The Department of Revenue divides restaurants into three separate subcategories, in accordance with the North American Industry Classification System: Full Service Restaurants, Limited Service Restaurants, and Cafeterias, Grills, and Buffets. Roughly two-thirds of the restaurants are full service, with most of the others falling under the limited service restaurants category (only a handful are in the third group).

2.3 Aggregating Data

I manually merged the revenue data with the Yelp reviews, inspecting the two datasets for similar or matching names. When a match was unclear, I referred to the address from the Department of Revenue listing. Table 1 summarizes Yelp penetration over time. By October of 2009, 69% of restaurants were on Yelp. To see the potential for Yelp to change the way firms build reputation, consider the fact that only 5% of restaurants are on Zagat (Jin and Leslie, 2009).

The final dataset is at the restaurant quarter level. Table 2 summarizes the revenue and review data for each restaurant quarter. The mean rating is 3.6 stars out of 5. On average, a restaurant receives 3 reviews per quarter, with each of these reviewers having 245 friends on average. Of these reviews, 1.4 come from elite reviewers. "Elite" reviewers are labeled as such by Yelp based on the quantity of reviews as well as other criteria.

One challenge with the revenue data is that it is quarterly. For the OLS regressions, I simply use the average rating for the duration of the quarter. For the regression discontinuity, the process is slightly more complicated. For these observations, I do the following. If the rating

does not change during a given quarter, then I leave it as is. If the rating does cross a threshold during a quarter, then I assign the treatment variable based on how many days the restaurant spent on each side of the discontinuity. If more than half of the days were above the discontinuity, then I identify the restaurant as above the discontinuity.[3]

3 Empirical Strategy

I use two identification strategies. I implement a regression discontinuity approach to support the hypothesis that Yelp has a causal impact. I then apply fixed effects regressions to estimate the heterogeneous effects of Yelp ratings.

3.1 Impact of Yelp on Revenue

The first part of the analysis establishes a relationship between a restaurant's Yelp rating and revenue. I use a fixed effects regression to identify this effect. The regression framework is as follows:

$$\ln(Revenue_{jt}) = \beta\, rating_{jt} + \alpha1_j + \alpha2_t + \epsilon_{jt}$$

where rating is $\ln(Revenue_{jt})$ is the log of revenue for restaurant j in quarter t, $rating_{jt}$ is the rating for restaurant j in quarter t. The regression also allows for year and restaurant specific unobservables. β is the coefficient of interest, which tells us the impact of a 1 star improvement in rating on a restaurant's revenue. While a positive coefficient on rating suggests that Yelp has a causal impact, there could be concern that Yelp ratings are correlated with other factors that

---

[3] An alternative way to run the regression discontinuity would be to assign treatment based on the restaurant's rating at the beginning of the quarter.

affect revenue. To support the causal interpretation, I turn to a regression discontinuity framework.

3.2 Regression Discontinuity

Recall that Yelp displays the average rating for each restaurant. Users are able to limit searches to restaurants with a given average rating. These average ratings are rounded to the nearest half a star. Therefore, a restaurant with a 3.24 rating will be rounded to 3 stars, while a restaurant with a 3.25 rating will be rounded to 3.5 stars, as in Figure 5. This provides variation in the rating that is displayed to consumers that is exogenous to restaurant quality.

I can look at restaurants with very similar underlying ratings, but which have a half-star gap in what is shown to consumers. To estimate this, I restrict the sample to all observations in which a restaurant is less than 0.1 stars from a discontinuity. This estimate measures the average treatment effect for restaurants that benefit from receiving an extra half star due to rounding. I also present estimates for alternative choices of bandwidth.

3.2.1   Potential Outcomes Framework

The estimation is as follows. First, define the binary variable T:

$$T = \begin{cases} 0 \ if \ rating \ falls \ just \ \textbf{below} \ a \ rounding \ threshold \ (so \ is \ rounded \ down) \\ 1 \ if \ rating \ falls \ just \ \textbf{above} \ a \ rounding \ threshold \ \ (so \ is \ rounded \ up) \end{cases}$$

For example, T = 0 if the rating is 3.24, since a Yelp reader would see 3 stars as the average rating. Similarly, T=1 if the rating is 3.25, since a Yelp reader would see 3.5 stars as the average rating.

The outcome variable of interest is ln (Revenue$_{jt}$). The regression equation is then simply:

$$\ln(Revenue_{jt}) = \beta \ T_{jt} + \gamma \ q_{ojt} + \alpha 1_j + \alpha 2_t + \epsilon_{jt}$$

where β is the coefficient of interest. It tells us the impact of an exogenous one-half star increase in a restaurant's rating on revenue. The variable q_{ojt} is the unrounded average rating. The coefficient of interest then tells us the impact of moving from just below a discontinuity to just above a discontinuity, *controlling for the continuous change in rating*.

In the main specification, I include only the restricted sample of restaurants that are less than 0.1 stars away from a discontinuity. To show that the result is not being driven by choice of bandwidth, I allow for alternative bandwidths. To show that the result is not being driven by non-linear responses to continuous changes in rating, I allow for a break in response to the continuous measure around the discontinuity. I also allow for non-linear responses to rating. I then perform tests of identifying assumptions.

3.3 Heterogeneous impact of Yelp

After providing evidence that Yelp has a causal impact on restaurant revenue, I investigate two questions regarding heterogeneous impacts of Yelp. First, I test the hypothesis that Yelp has a smaller impact on chains. The estimating equation is as follows:

$$\ln(Revenue_{jt}) = \beta \, rating_{jt} + \delta \, rating \, X \, chain_{jt} + \alpha 1_j + \alpha 2_t + \epsilon_{jt}$$

The coefficient of interest is then δ. A negative coefficient implies that ratings have a smaller impact on revenue for chain restaurants.

I then test whether consumer response is consistent with a model of Bayesian learning. The estimating equation is as follows:

$$\ln(Revenue_{jt}) = \beta \, rating_{jt} + \gamma \, rating \, X \, noise_{jt} + \alpha 1_j + \alpha 2_t + \epsilon_{jt}$$

The variable $rating\ X\ noise_{jt}$ interacts a rating with the amount of noise in the rating. A Bayesian model predicts that if the signal is less noisy, then the reaction should be stronger. The variable $rating\ X\ prior\ beliefs_{jt}$ interacts a rating with the precision of prior beliefs about restaurant quality. Bayesian learning would imply that the market reacts less strongly to new signals when prior beliefs are more precise. All specifications will include restaurant and year fixed effects.

Empirically, I will identify situations were ratings contain more and less information and where prior information is more and less precise. I will then construct the interaction terms between these variables and a restaurants rating.

There are two ways in which I measure noise. First, I consider the number of reviews that have been left for a restaurant. If each review provides a noisy signal of quality, then the average rating presents a more precise signal as there are more reviews left for each restaurant. Bayesian learners would then react more strongly to a change in rating when there are more total reviews. Second, I consider reviews left by elite reviewers, who have been certified by Yelp. If reviews by elite reviewers contain more information, then Bayesian learners should react more strongly to them.

4 Impact of Yelp on Revenue

Table 3 establishes a relationship between a restaurant's rating and revenue. A one-star increase is associated with a 5.4% increase in revenue, controlling for restaurant and quarter specific unobservables. The concern with this specification is that changes in a restaurant's rating may be correlated with other changes in a restaurant's reputation. In this case, the coefficient on Yelp rating might be biased by factors unrelated to Yelp.

To reinforce the causal interpretation, I turn to the regression discontinuity approach. In this specification, I look at restaurants that switch from being just below a discontinuity to just above a discontinuity. I allow for a restaurant fixed effect because of a large restaurant-specific component to revenue that is fixed across time. Figure 4 provides a graphical analysis of demeaned revenues for restaurants just above and just below a rounding threshold. One can see a discontinuous jump in revenue. Table 4 reports the main result, with varying controls. Table 4 considers only restaurants that are within a 0.1-star radius of a discontinuity. Table 5 varies the bandwidth.

I find that an exogenous one-star improvement leads to a roughly 9% increase in revenue. (Note that the shock is one-half star, but I renormalize for ease of interpretation). The result provides support to the claim that Yelp has a causal effect on demand. In particular, whether a particular restaurant is rounded up or rounded down should be uncorrelated with other changes in reputation outside of Yelp.

The magnitude of this effect can be compared to the existing literature on the impact of information. Gin and Leslie (2003) show that when restaurants are forced to post hygiene report cards, a grade of A leads to a 5% increase in revenue relative to other grades. In the online auction setting, Cabral and Hortacsu (2010) show that a seller experiences a 13% drop in sales after the first bad review. In contrast to the electronic commerce setting, Yelp is active in a market where (1) other types of reputation exist since the market is not anonymous (and many restaurants are chain-affiliated), (2) there may be a high cost to starting a new firm or changing names, leaving a higher degree of variation in rating, and (3) consumers must actively seek information, rather than being presented with it at the point of purchase.

In addition to identifying the causal impact of Yelp, the regression discontinuity estimate is information about the way that consumers use Yelp. First, it tells us that Yelp as a new source of information is becoming an important determinant of restaurant demand. The popularization of the internet has provided a forum where consumers can share experiences, which is becoming an important source of reputation. Second, the mean rating is a salient feature in the way that consumers use Yelp. Consumers respond to discontinuous jumps in the average rating. Intuitively, the average rating provides a simple feature that is easy to use. Third, this implies that consumers do not use all available information, but instead use the rounded rating as a simplifying heuristic. Specifically, if attention was unlimited, then consumers would be able to observe changes to the mean rating based on the underlying reviews. Then the rounded average would be pay-off irrelevant. Instead, consumers use the discontinuous rating, which is less informative than the underlying rating but also less costly (in terms of time and effort) to use.

4.1 Identifying Assumptions

This regression discontinuity approach heavily relies on random assignment of restaurants to either side of the rounding thresholds. Specifically, the key identifying assumption is that as we get closer and closer to a rounding threshold, all revenue-affecting predetermined characteristics of restaurants become increasingly similar. Restricting the sample to restaurants with very similar ratings, we can simply compare the revenues of restaurants that are rounded up to the revenues of restaurants that are rounded down.

This helps to avoid many of the potential endogeneity issues that occur when looking at the sample as a whole. In particular, restaurants with high and low Yelp scores may be very

different. Even within a restaurant, reputational changes outside of Yelp may be correlated with changes in Yelp rating over time. However, the differences should shrink as the average rating becomes more similar.

For restaurants with very similar ratings, it seems reasonable to assume that restaurants changes that are unrelated to Yelp would be uncorrelated with whether a restaurant's Yelp rating is rounded up or rounded down. The following section addresses potential challenges to identification.

4.1.1   Potential Manipulation of Ratings

One challenge for identification in a regression discontinuity design is that any threshold that is seen by the econometrician might also be known to the decision makers of interest. This can cause concerns about gaming, as discussed in McCrary (2008). In the Yelp setting, the concern would be that certain types of restaurants submit their own reviews in order to increase their revenue. This type of behavior could bias the OLS estimates in this paper if there is a correlation between a firm's revenue and decision to game the system. The bias could go in either direction, depending on whether high revenue or low revenue firms are more likely to game the system. In this section, I address the situation that could lead to spurious results. I then argue that selective gaming is not causing a spurious correlation between ratings and revenues in the regression discontinuity framework.

In order for the regression discontinuity estimates to be biased, it would have to be the case that restaurants with especially high (or alternatively with especially low) revenue are more likely to game the system. This is certainly plausible. However, it would also have to be the case that these restaurants stop submitting fake reviews once they get above a certain

discontinuity. In other words, if some restaurants decided to submit fake reviews while others did not, the identification would still be valid.

In order to invalidate the regression discontinuity identification, a restaurant would have to submit inflated reviews to go from a rating of 2.2 stars, only to stop when it gets to 2.4 stars. However, if a restaurant stopped gaming as soon as it jumped above a discontinuity, the next review could just drop it back down. While the extent of gaming is hard to say, it is a very restrictive type of gaming that would lead to spurious estimates.

I offer two further arguments against the gaming hypothesis: one economic and the other statistical. First, suppose the concern is that restaurants are gaming in this sophisticated manner, leading to a spurious impact of rating where none exists. This argument becomes circular because if no effect exists, then restaurants should not have the incentive to invest in gaming. Therefore even the existence of gaming would require that Yelp has a causal effect on revenue.

The second piece of evidence against the gaming hypothesis is based on a test offered by McCrary (2008). The intuition of the test is as follows. Suppose that restaurants were gaming Yelp in a way that would bias the results. Then, one would expect to see a disproportionately large number of restaurants just above the rounding thresholds.

I construct the test in the following way. I begin with a dataset at the restaurant / review level. For example, a restaurant that has five reviews would have five observations. The variable of interest would be the average rating after each review. If there was gaming, there should be "too many" observations with ratings just above rounding thresholds.

To formally test for this, I sum the number of observations for each 0.05-star interval, and compute the probability mass for each interval. I create a binary variable to indicate bins that fall just above a rounding threshold (e.g., 3.25-3.3 stars, 3.75-3.8 stars). The dependent variable is

the probability mass, and the independent variable is the indicator for bins that fall just above the discontinuity.

Table 6 presents the results of this test. The test shows that there is not any clustering of restaurants just above the discontinuity, suggesting that manipulation is not an issue with the regression discontinuity design.

5 The Impact of Yelp on Chains

How does the introduction of a new technology that increases information flow affect restaurants with chain affiliation? Historically, chain affiliation is valuable precisely because it reduces uncertainty about restaurant quality. Consumer reviews are coming to serve a similar purpose.

There are two ways in which Yelp ratings might affect chains. First, a chain's rating on Yelp may have an effect on revenue. Second, Yelp may cause an overall shift in demand between chains and independent restaurants if Yelp is providing more information about independent restaurants than about chains. In this section, I investigate both effects.

5.1 Do Ratings Affect Chains?

Table 7 presents the differential impact of Yelp ratings on chain restaurants. While ratings have a large impact on revenue overall, the effect is being driven entirely from independent restaurants. Because chains already have relatively little uncertainty about quality, their demand does not respond to consumer reviews

5.2 Do Consumer Reviews Crowd Out Demand for Chains?

Given the differential impact of Yelp on chains and independent restaurants, one might expect chains to become less popular after the introduction of Yelp. This is because the increased information about independent restaurants leads to a higher expected utility conditional on going to an independent, restaurant. Hence Yelp should not only shift demand between independent restaurants, it should also increase the value of going to an independent restaurant relative to a chain.

Consistent with this, table 8 shows that chains experienced a decline in revenue relative to independent restaurants in the post-Yelp period. Higher Yelp penetration leads to an increase in revenue for independent restaurants, but a decrease in revenue for chain restaurants.

One may be concerned with this specification if chain restaurants had been trending downward in the period before Yelp was introduced. To address this concern, I show that the result is robust to the inclusion of chain-specific time trends.

6   Evidence of Bayesian Learning

How do consumers update beliefs based on information obtained from consumer reviews? On the one hand, a standard model of Bayesian learning predicts that the market would react more strongly when ratings contain more precise information and when prior beliefs are less precise. On the other hand, we have already seen that consumers use the average rating as a simplifying heuristic. This may cast doubt on the sophistication of consumer response.

It is possible to test for Bayesian learning, taking a restaurant's rating as a public signal of quality. The market response to the signal depends on two things: consumers' prior beliefs about product quality, and the precision of the signal. The precision of information contained in user reviews depends on the number of reviews and the credibility of the reviewers.

In this section, I identify situations where the signal is more and less precise in order to test for Bayesian learning.

6.1 Number of Reviews

If each consumer review presents a noisy signal of quality, then having many reviews should cause the overall rating to contain more information and hence have a larger impact. Table 9 shows that this is in fact the case.

The first column looks at all restaurants, and shows that a change in a restaurant's rating has 50% more impact when the restaurant has at least 50 reviews (compared to a restaurant with fewer than 10 reviews). However, this interpretation could be concerning if restaurants that are more rating-sensitive receive more reviews. To allay this concern, I restrict the sample to restaurants that have at least 50 reviews as of October 2009 (column 2). I then consider how responsive these restaurants are to changes in rating as they receive more reviews.

Under this specification, a restaurant with at least 50 reviews is roughly 20% more rating sensitive than when it had fewer than 10 reviews.

6.2 Certified Reviewers

Consumer reviews are written by a non-representative sample of voluntary reviewers who often have little or no connection to the reader. In order to find a review useful, a consumer must find it relevant, accurate, and credible. One way to achieve this is to certify the quality of a reviewer.

Yelp has a reviewer credentialing program, where they formally certify certain reviewers who have written a lot of reviews that Yelp has deemed helpful. These reviewers are marked as

"elite," and in addition to knowing whether a reviewer was elite, readers can filter to only look at reviews by elite reviewers.

If elite certification gives reviewers a reputation for leaving informative reviews, then reviews by elite members should have a larger impact. Consistent with the Bayesian hypothesis, Table 10 shows that elite reviewers have roughly double the impact of other reviewers. Despite the fact that the econometrician cannot observe the criteria for certifying a reviewer as elite, this suggests a strong role for reviewer reputation.

An alternative explanation of this result is that Yelp simply certifies reviewers who are better at predicting average consumer preferences. There are several difficulties with this interpretation. First, Yelp does not have access to revenue data at the restaurant level, so this would require Yelp to know consumer preferences. Second, if Yelp knew the distribution of preferences over restaurants, they could simply announce them. Third, the regression includes restaurant fixed effect. In order for the result to be spurious, elite reviewers would then have to be more likely to review restaurants whose reputation is about to improve. To some extent, this seems plausible. However, if it is, then rational consumers should be responding more heavily to elite reviews, which are then more indicative of a restaurant's reputation. This argument would therefore not nullify the result. Further, I find that elite reviewers only have an effect after becoming certified as elite.

A second way to think of certifying reviewer quality is through the number of "friends" the reviewer has. Yelp reviewers are able to form online connections, called "friends" with other reviewers. Having many friends might plausibly signal that a reviewer writes precise reviews, or has tastes that reflect popular opinions. Empirically, I estimate this by weighting the overall

rating by the number of friends each reviewer has. I find that the number of friends each reviewer has does not affect the impact of a review.

7   Discussion

The overall message of this paper is simple. Online consumer review websites improve the information available about product quality. The impact of this information is larger for products of relatively unknown quality. As this information flow improves, other forms of reputation such as chain affiliation should continue to become less influential. On the consumer side, simplifying heuristics and signals of reviewer quality seem to increase the impact of quality information. In this section, I put some of the results into broader context and discuss possible areas of future work.

7.1 Comparing Consumer Reviews with Mandatory Disclosure

This paper shows that a one-star increase leads roughly to a 9% increase in revenue. One relevant comparison is between consumer reviews and mandatory disclosure laws. Jin and Leslie (2003) find that a restaurant whose hygiene report card grade moves from a B to an A experiences a 5% increase in revenue relative to other grades. Bollinger, Leslie, and Sorensen (2010) find that calorie posting laws cause consumers to consumer 6% fewer calories at Starbucks.

Ultimately, the policy goal of quality disclosure laws is to (1) provide information to consumers, so that they can make better decisions and (2) hold firms accountable. This paper suggests that consumer review websites can be equally as effective at altering demand, although there is no hard evidence on the correlation between Yelp ratings and more objective quality

measures. In ongoing work, I am estimating the correlation between Yelp ratings and other objective quality measures.

7.2 Comparing Yelp and Other Reviews

Clearly Yelp is not the only way in which consumers learn about restaurant quality. However, Yelp is striking in the sheer number of restaurants that contain non-trivial numbers of reviews. Appendix 4 shows the percent of restaurants covered by different review systems in urban areas. As discussed, Yelp currently contains reviews of 70% of restaurants in Seattle. In contrast, Zagat is only a 5% sample (Gin and Leslie 2009) in Los Angeles. My own data shows that the Seattle Times - a local paper that also reviews restaurants - contains even fewer restaurants, as does the magazine Food & Wine.

7.3 Comparative Incentive Problems of Consumer Reviews and Chains

Chain affiliation helps to increase the amount of information available about restaurant quality. However, chain affiliation can also lead to free-riding (Jin and Leslie 2009) and high monitoring costs (Kaufmann and Lafontaine 1994). Consumer reviews may reduce these incentive problems. This is one reason why consumer demand is shifting from chain to independent restaurants in the period following the introduction of Yelp. On the other hand, consumer reviews create separate incentive issues, such as an underprovision problem (Avery 1999) and selection of reviewer.

7.4 Welfare Gains from Yelp

It seems uncontroversial to assert that providing this information to consumers might improve welfare in various ways. As evidence, I discuss two results.

First, Yelp causes demand to shift from chains to independent restaurants. By revealed preference, consumers' expected utility from going to independent restaurants must then be higher. This can be viewed as a welfare gain resulting from either better restaurants or better sorting between consumers and restaurants.

Second, revenue is a key determinant of a restaurant's decision to exit. Hence, Yelp may have a long-run effect on exit behavior of firms. Assuming Yelp measures are a reasonable measure of true quality, then Yelp may help to drive worse restaurants out of business, which would be a second source of welfare gain. In ongoing work, I am estimating the relationship between Yelp and exit decisions.

References

Avery, Christopher, Paul Resnick, and Richard Zeckhauser, 1999. "The Market for Evaluations," *American Economic Review*, Vol. 89, No. 3, 564-584.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, 2010. "Calorie Posting in Chain Restaurants," *American Economic Journal*, Economic Policy.

Bundorf, Kate, Natalie Chun, Gopi Shah Goda, and Daniel Kessler, 2009. "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics*, Vol. 28, No. 3, 718-727.

Cabral, Luis and Ali Hortacsu, 2010. "The Dynamics of Seller Reputation: Theory and Evidence from eBay," *Journal of Industrial Economics*, Vol. 58, 54-78.

Chetty, Raj, Adam Looney, and Kory Kroft, 2009. "Salience and Taxation: Theory and Evidence," *American Economic Review*, Vol. 99, No. 4, 1145-1177.

Chevalier, Judith and Dina Mayzlin, 2006. "The Effect of Word-of-Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, Vol. 43, No 3.

Dellarocas, Chrysanthos, 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, Vol. 49, No. 10, 1407-1424.

Dellavigna, Stefano and Joshua Pollet, 2007. "Demographics and Industry Returns," *American Economic Review*, Vol. 97, No. 5, 1667-1702.

Dellavigna, Stefano and Joshua Pollet, 2010. "Investor Inattention and Friday Earnings Announcements," *Journal of Finance*, Vol. 64, No. 2, 709-749.

Eliashberg, Jehoshua and Steven Shugan, 1997. "Film Critics: Influencers or Predictors?," *Journal of Marketing*, Vol. 61, 68-78.

Ellison, Glenn and Drew Fudenberg, 1995. "Word-of-Mouth Communication and Social Learning," *Quarterly Journal of Economics*, Vol. 110, No. 1, 93-125.

Finkelstein, Amy, 2009. "E-ZTAX: Tax Salience and Tax Rates," *Quarterly Journal of Economics*, Vol. 124, No. 3, 969-1010.

Imbens, Guido and Thomas Lemieux, 2008. "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, Vol. 142, No. 2, 615-635.

Jin, Ginger and Phillip Leslie, 2009: "Reputational Incentives for Restaurant Hygiene," *American Economic Journal: Microeconomics*, Vol. 1, No. 1, 236-267.

Jin, Ginger and Phillip Leslie, 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*, Vol. 118, No. 2, 409-451.

Kaufmann, Patrick and Francine Lafontaine, 1994. "Costs of Control: The Source of Economic Rents for McDonald's Franchises," *Journal of Law and Economics*, Vol. 37, No. 2, 417-453.

Lewis, Gregory, 2009. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," *working paper*.

Luca, Michael and Jonathan Smith, 2010. "Salience in Quality Disclosure: Evidence from The US News College Rankings," *working paper*.

Mathios, Alan, 2000. "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market," *Journal of Law and Economics*, Vol. 43, No. 2, 651-677.

McCrary, Justin, 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, Vol. 142, No. 2, 698-714.

Pope, Devin, 2009. "Reacting to Rankings: Evidence from 'America's Best Hospitals'," Vol. 28, No. 6, 1154-1165.

Reinstein, David and Christopher Snyder, 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," *The Journal of Industrial Economics*, Vol. 53, No. 1, 27-51.

Resnick, P. and R. Zeckhauser, 2002. "Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *The Economics of the Internet and E-Commerce,* M. Bayes, ed., Amsterdam: Elsevier Science**,** 11: 127-157.

Rockoff, Jonah, Thomas Kane, Douglas Staiger, and Eric Taylor, 2010. "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools," *working paper.*

Scanlon, Dennis, Michael Chernew, Catherine McLaughlin, and Gary Solon, 2002. "The Impact of Health Plan Report Cards on Managed Care Enrollment," *Journal of Health Economics*, Vol. 21, 19-41.

Tadelis, Steven, 1999. "What's in a Name? Reputation as a Tradeable Asset," *American Economic Review*, Vol. 89, No. 3, 548-563.

Wang, Zhongmin, 2010. "Anonymity, Social Image, and the Competition for Volunteers: A Case Study of the Online Market for Reviews," *The B.E. Journal of Economic Analysis and Policy*, Vol. 10, No. 1, Article 44.

Table 1: Penetration of Yelp Over Time

| | Year | Quarter | # Restaurants | | # Reviews | | |
			Total	In Yelp	Quarterly	Total	Per Restaurant
Pre Yelp	2003	1Q	1,434	0%	0	0	0.0
	2003	2Q	1,485	0%	0	0	0.0
	2003	3Q	1,497	0%	0	0	0.0
	2003	4Q	1,497	0%	0	0	0.0
	2004	1Q	1,509	0%	0	0	0.0
	2004	2Q	1,515	0%	0	0	0.0
	2004	3Q	1,512	0%	0	0	0.0
	2004	4Q	1,545	0%	10	10	1.4
	2005	1Q	1,530	1%	4	14	1.6
	2005	2Q	1,569	1%	21	35	1.6
Post Yelp	2005	3Q	1,575	15%	537	572	2.4
	2005	4Q	1,566	22%	529	1,101	3.3
	2006	1Q	1,554	28%	931	2,032	4.6
	2006	2Q	1,563	33%	882	2,914	5.6
	2006	3Q	1,572	37%	1,070	3,984	6.8
	2006	4Q	1,596	40%	1,591	5,575	8.8
	2007	1Q	1,578	44%	2,372	7,947	11.4
	2007	2Q	1,590	47%	2,973	10,920	14.5
	2007	3Q	1,581	51%	3,519	14,439	18.0
	2007	4Q	1,578	54%	3,450	17,889	21.1
	2008	1Q	1,548	57%	4,766	22,655	25.7
	2008	2Q	1,548	60%	5,083	27,738	30.0
	2008	3Q	1,560	61%	5,905	33,643	35.1
	2008	4Q	1,575	62%	5,682	39,325	40.0
	2009	1Q	1,545	66%	7,640	46,965	46.4
	2009	2Q	1,557	67%	7,481	54,446	52.4
	2009	3Q	1,587	69%	8,263	62,709	57.4

Table 2: Summary Statistics

Variable	Obs	Mean	Std Dev	Min	Max
Revenue ($)	41,766	176,105	440,723	33	8,774,281
Rating	14,593	3.6	0.9	1	5
Reviews	14,593	3.0	4.8	0	82
Elite Reviews	14,593	1.4	2.0	0	28
Friends of Reviewers	14,593	244.5	506.3	0	15,751

Note: All statistics are per quarter per restaurant.

Table 3: Impact of Yelp on Revenue

Dependent Variable = ln (Revenue)	
On Yelp	0.089***
	(0.008)
Rating	0.054***
	(0.007)
Quarter FE	x
Restaurant FE	x
Observations	41766
Restaurants	3582

Notes: Rating is measured in deviations from the mean. "On Yelp" indicates whether the restaurant was in Yelp at the time of each observation. Robust standard errors are reported. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 4: Regression Discontinuity Estimate

	Dependent Variable = ln (Revenue)		
Discontinuity	0.094**	0.092**	0.093**
	(0.041)	(0.042)	(0.042)
Rating	x	x	x
Rating Quadratic		x	
Rating X Above			x
Observations	2169	2169	2169
Restaurants	854	854	854

Note: All specifications also control for restaurant and quarter FE.
Regressions include all observations within 0.1 stars of a
discontinuity. Robust standard errors are reported. *, **, ***
denote significance at the 10%, 5%, and 1% level.

Table 5: RD Estimates for Different Bandwidths

	Dependent Variable = ln (Revenue)	
Discontinuity	0.054**	0.094**
	(0.028)	(0.041)
Rating	x	x
Observations	3569	2169
Restaurants	1001	854
Bandwidth (Stars)	0.3	0.2

Note: All specifications also control for restaurant and quarter FE.
Robust standard errors are reported. *, **, *** denote significance
at the 10%, 5%, and 1% level.

Table 6: McCrary Test for Quasi-Random Assignment

Dependent Variable = Prob Mass of 0.05 Star Bin	
Treatment (0.05 star interval above rounding threshold)	0.001
	(0.005)
Observations	78

Note: Dependent variable is the probability mass of observations in each 0.05 star interval. The treatment variable indicates intervals that are just above a rounding threshold.

Table 7: Differential Response for Chains

	Dependent Variable = ln (Revenue)		
	All Restaurants	Only Independents	Only Chains
On Yelp	0.097 ***	0.097 ***	0.000
	(0.008)	(0.008)	(0.038)
On Yelp X Chain	-0.086 ***		
	(0.029)		
Rating	0.065***	0.065***	0.005
	(0.009)	(0.009)	(0.025)
Rating X Chain	- 0.055**		
	(0.027)		
Observations	41766	39283	2483
Restaurants	3582	3439	143

Notes: All specifications include restaurant and quarter fixed effects. Rating is measured in deviations from the mean. "On Yelp" indicates whether the restaurant was in Yelp at the time of each observation. Robust standard errors are reported. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 8: Test for Crowding out of Chains

	Dependent Variable = ln (Revenue)		
Yelp Penetration	0.027 **	0.059 *	0.070 *
	(0.011)	(0.035)	(0.036)
Yelp Penetration X Chain	-0.078 **	-0.078 **	-0.283 **
	(0.037)	(0.037)	(0.131)
Chain	x	x	x
Restaurant FE	x	x	x
Time Trend		x	x
Chain Specific Time Trend			x
Observations	41766	41766	41766
Restaurants	3582	3582	3582

Notes: Yelp penetration measures the percent of restaurants on Yelp in a given quarter. Robust standard errors are reported. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 9: Response to Rating by Number of Reviews

	Dependent Variable = ln (Revenue)	
	(1)	(2)
On Yelp	0.082***	0.153***
	(0.022)	(0.047)
Rating	0.053***	0.115***
	(0.009)	(0.018)
Rating ×	0.010*	0.006
(11-20 reviews)	(0.005)	(0.008)
Rating ×	0.015***	0.015*
(21-30 reviews)	(0.005)	(0.008)
Rating ×	0.017***	0.019**
(31-40 reviews)	(0.006)	(0.008)
Rating ×	0.018***	0.021**
(41-50 reviews)	(0.007)	(0.009)
Rating ×	0.027***	0.026***
(50+ reviews)	(0.007)	(0.002)
Observations	41766	6080
Restaurants	3582	369

Notes: Rating is measured in deviations from the mean.
Regressions control for number of reviews, and include
restaurant and quarter fixed effects. Specification (1)
includes all restaurants. Specification (2) includes only
restaurants that have at least 50 reviews as of December
2009. Robust standard errors are reported. *, **, ***
denote significance at the 10%, 5%, and 1% level.

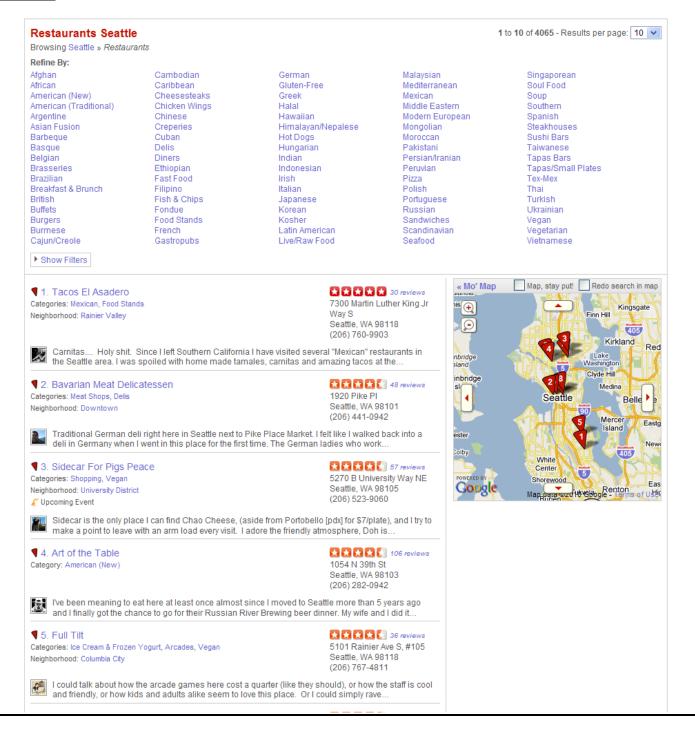**Table 10: Response to Rating Weighted by Reviewer Type**

	Dependent Variable = ln (Revenue)			
On Yelp	0.088***	0.089***	0.090***	0.091***
	(0.008)	(0.008)	(0.008)	(0.008)
Rating		0.054***	0.035***	0.029**
		(0.007)	(0.009)	(0.013)
Rating - elite wtd			0.036***	0.034***
			(0.008)	(0.009)
Rating - friend wtd				0.008
				(0.013)
Quarter FE	x	x	x	x
Restaurant FE	x	x	x	x
Observations	41766	41766	41766	41766
Restaurants	3582	3582	3582	3582

Notes: Rating is measured in deviations from the mean. "On Yelp" indicates whether the restaurant was in Yelp at the time of each observation. Rating - elite wtd is the average of the ratings left by elite reviewers. Rating - friend wtd weights the average rating by the number of friends a reviewer has as of December 2009. Robust standard errors are reported. *, **, *** denote significance at the 10%, 5%, and 1% level.
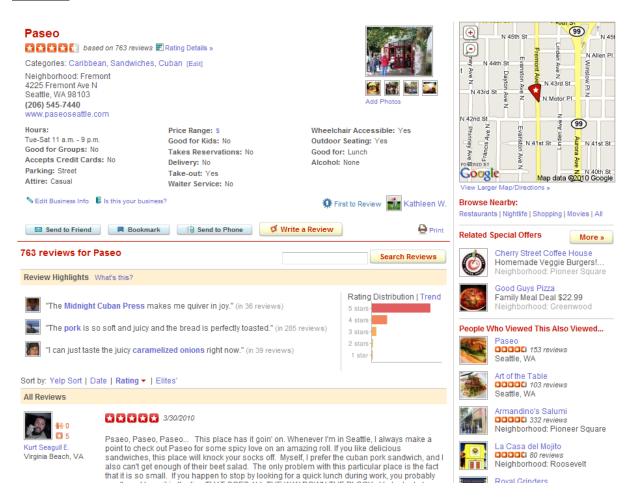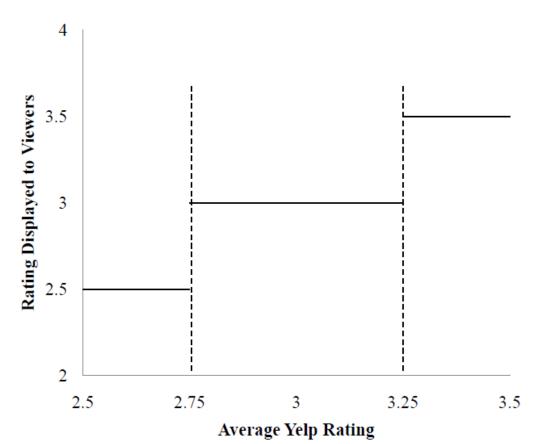
Figure 1

Figure 2

Figure 3: Yelp Displays Each Restaurant's Rounded Average Rating

Notes: Yelp prominently displays a restaurant's rounded average rating. Each time a restaurant's rating crosses a rounding threshold, the restaurant experiences a discontinuous increase in the displayed average rating.
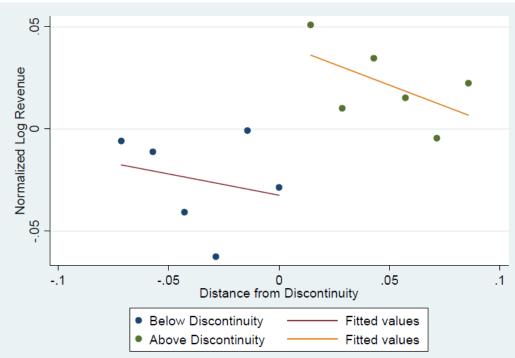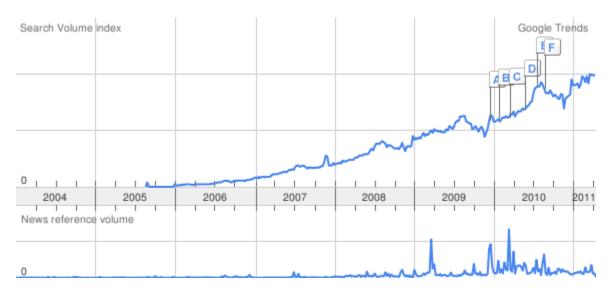
Figure 4: Average Revenue around Discontinuous Changes in Rating
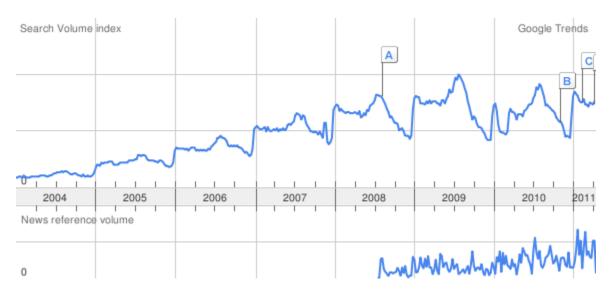

Revenues averaged within 0.01 star bins.

Notes: Each restaurant's log revenue is de-meaned to normalize a restaurant's average log revenue to zero. Normalized log revenues are then averaged within bins based on how far the restaurant's rating is from a rounding threshold in that quarter. The graph plots average log revenue as a function of how far the rating is from a rounding threshold. All points with a positive (negative) distance from a discontinuity are rounded up (down).
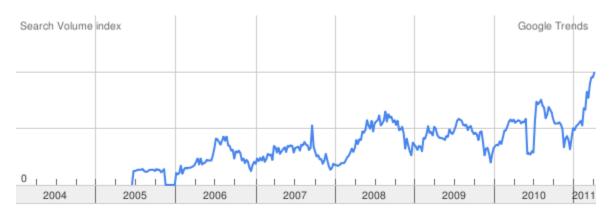
Appendix

Appendix 1: Yelp Search Volume

Appendix 2: Trip Advisor Search Volume

Appendix 3: Angie's List Search Volume

Appendix 4: Restaurants Covered by Different Information Sources