Consumer Reviews and Regulation: Evidence from NYC Restaurants[†]

Preliminary and incomplete – please do not circulate.

Georgios Zervas
Questrom School of Business
Boston University

Chiara Farronato
Harvard Business School
and NBER

December 17, 2017

Abstract

We investigate complementarities and substitutabilities between two signals of restaurant quality: health inspections and online reviews. To protect consumers from unsafe dining, health inspections periodically evaluate restaurants on hygiene quality, and assign them health grades. Recently, consumers have increasingly been able to rate restaurant quality online, through platforms like Yelp. We first investigate whether online consumer reviews *detect* hygienic conditions that health inspectiors evaluate. To do this, we implement a text analysis machine learning algorithm to predict individual restaurant violations from the text of Yelp reviews. We find that consumer reviews are good predictors of food handling violations, but are poor predictors of facilities and maintenance violations. We then investigate how the hygienic information contained in online reviews affects consumer demand and supply incentives. On the demand side, we find that conditional on hygiene quality contained in online reviews, customers still use health grades to choose resturants. On the supply side, we find that relative to restaurants not on Yelp, restaurants reviewed on Yelp score better on hygiene dimensions detectable by customers than on dimensions not detectable by customers. The paper results have implications for the design of government regulation in a world where consumers rate their service experiences online.

1 Introduction

When consumers have limited information about service providers' underlying quality before making a purchase, a market for services can easily break down. Low quality providers can

[†]We thank Yi Fung for outstanding research assistance. We thank the New York City Department of Health and Mental Hygiene for sharing data and insights.

drive away high quality providers, and consumers might either stop purchasing or engage in risky transactions. This is the classical lemons market problem (Akerlof (1970)).

Regulation has historically been the solution to protect consumers from risky transactions and prevent a market break-down. Many services are subject to specific licensing requirements, and health and safety rules. Regulation relies on experts' evaluation of providers' quality. Restaurants, for example, are regularly inspected by public officials to evaluate their hygienic conditions.

In recent years, technology has facilitated the diffusion of online review systems. Tripadvisor reviews hotels and tourist attractions, Yelp reviews restaurants and other local services. Review systems typically count on customers' willingness to rate their prior experience with certain providers. In some situations all relevant dimensions of providers' quality are observable during a transaction. Examples might include hair braiders and travel agents. In other situations, such as with surgeons or lawyers, quality is not as easily observable by non-experts.

When providers' quality is easily observable by customers, aggregating their reviews as *wisdom of the crowds* might be an effective way to guarantee service quality and screen out bad providers. In these cases, and when an online system is capable of providing unbiased signals of providers' quality, online reviews can in principle partially substitute for regulation. Put differently, there can be situations in which the wisdom of the crowds, aggregated via online reviews, can be as good as the expert evaluation of providers' quality.

In this paper we investigate whether online reviews can partially substitute for health and safety inspections of New York City restaurants. We combine data on inspection records from the New York City Department of Health, and online reviews from Yelp.com. First, we predict which health violations are *observable* by the crowd, as measured in the text of Yelp reviews. We find that food handling violations are accurately predicted by the text of recent Yelp reviews. On the other hand, we find that violations about facilities and maintenance are not detectable through online reviews.

We use our prediction of hygienic conditions from Yelp reviews to evaluate the efficacy of health inspections along two dimensions: informing restaurant-goers about restaurant hygienic quality, and nudging restaurants to improve their hygiene. The two objectives are directly stated by the NYC department of health as the main goals of the health and safety inspection program. We will denote the first goal the *customer-facing informational objective*, and the second goal the *restaurant-facing incentive objective*.

We find that the customer-facing informational objective of health and safety inspections is completely absorbed by the information obtained on Yelp. First, we show that when we summarize the information from Yelp with the simple average rating, letter grading seems to

affect demand: restaurants with A-grade experience an increase in demand and a reduction in the probability of going out of business. However, Yelp contains much more information than just the average rating. This information is contained in the text of online reviews. Once we incorporate the restaurant hygiene prediction from Yelp text, the effect of letter grades on customer demand is indistinguishable from zero.

Related to the restaurant-facing incentive goal, we find that restaurants' response to inspections is (short/long)-lived. Restaurants improve their hygienic quality following a first inspection, but there is sizeable heterogeneity in their response. We present a diff-in-diff approach to estimate whether online reviews provide restaurants with incentives to be more hygienic. This approach exploits the fact that some violations are visible through online reviews, as predicted by our machine learning algorithm on the text of Yelp reviews, and some food establishments are not on Yelp.

Our empirical evidence suggests that online reviews can partially substitute for health inspections. The crowd is able to detect food handling violations. For other violations, the expert is still required. Our paper provides suggestive evidence that health inspections can be re-focused on violations that are not visible through the eyes of the crowd.

Of course the viability of online reviews as a partial substitute to regulation depends on the review platform's incentives to provide unbiased ratings, and on its ability to collect truthful reviews while policing providers' attempts to manipulate the system. We turn to this issue in our concluding remarks.

The debate over online reviews and regulation has resurged in the last years, with the diffusion of online peer-to-peer markets such as Uber and Airbnb. The debate focuses on whether these new businesses should be subject to existing regulation, just like taxis and hotels. On one side, new online marketplaces argue that review systems are effective at ensuring service quality and monitoring service providers. On the other side, policymakers are skeptical about the efficacy of online reviews, and instead favor stricter regulation.[1]

Regulation has proved valuable in ensuring providers' quality. There is extensive literature on the value of regulation in ensuring health and safety standards. Within the restaurant industry, existing papers have shown that inspections are an effective way to provide high-quality service, and that consumers are sensitive to them (Jin and Leslie (2003)).

But regulation can also be costly by limiting competition. Research has shown that occupational licensing (Federman et al. (2006)) can be harmful to manicurists who do not meet the licensing standards but whose services are nonetheless valuable to consumers.

[1]Last year both Uber and Lyft pulled out of Austin to protest fingerprint background checks (New York Times, 5/10/2016). http://www.nytimes.com/2016/05/10/technology/uber-and-lyft-stop-rides-in-austin-to-protest-fingerprint-background-checks.html

The need for regulation is often based on the premise that other ways to ensure consumer protection are impractical or too costly. However, recent technology developments have made it cheaper to collect and aggregate information about service providers. In particular, review systems now facilitate the collection, aggregation, and disclosure of consumers' past experiences, and can offer a cheaper and equally effective alternative to regulatory standards.

If accurate, online reviews offer some practical advantages over regulation. They are cheaper and more frequent than regulatory inspections. Major US cities are currently financially constrained, and conducting restaurant health inspections can be costly. In addition, inspectors have been shown to have discretionary power over their evaluations (Ibanez and Toffel (2017)). If a large number of customers contribute reviews about restaurants' hygiene, they have the potential to be more accurate than inspections.

The paper is organized as follows. In Section 2 we present the relevant literature. Section 3 presents a simple theoretical framework of a market for services with asymmetric information, where online reviews and regulation serve as signals of service quality. Section 4 describes the empirical context and the data used. Section 7 presents our results and Section 8 concludes.

2 Related work

Our work is most closely related to two strands of literature. The first set of related papers study the role of online reviews in ensuring trust and affecting consumer choice. Lewis and Zervas (2017) look at the effect of hotel ratings on travelers' demand. They find that an additional star on TripAdvisor, Expedia, and Hotels.com increases hotel demand by 25%, and the hotel is able to charge 9% higher prices. Luca (2011) studies the effect of Yelp star ratings on restaurant revenues, finding that an additional Yelp stars increases restaurant revenues by 5-9%. Other papers, including Resnick and Zeckhauser (2002) and Cabral and Hortacsu (2010), have looked at reputation mechanisms in e-commerce.

There is also a set of papers on the biases of online reviews, which is important to keep in mind when evaluating whether online reviews are an accurate measure of providers' quality. Mayzlin et al. (2014) show evidence that hotels have incentives to fake reviews on TripAdvisor, by submitting positive reviews to themselves and negative reviews to their competitors. Luca and Zervas (2016) confirm that incentives to manipulate reviews are also present on Yelp.com.

The second set of related work focuses on the effects of regulation on service quality. Jin and Leslie (2003) focus on a policy change in Los Angeles county that required restaurants to start displaying their health grades at their door. They find that consumer demand became

sensitive to changes in restaurant's hygiene quality, restaurants improved their hygienic conditions, and the number of hospitalizations due to foodborne illnesses decreased. Separate work by Simon et al. (2005) confirms their results. The improvements in restaurant hygiene occur despite the fact that restaurants have different reputational incentives (Jin and Leslie (2009)). Online, Hui et al. (2016) investigate the addition of a buyer protection program on eBay. They find a reduction in moral hazard after eBay started mandating that sellers refund buyers when items are not received as described. This reduction in moral hazard increases surplus, above and beyond the benefits from the reputational system already present on eBay.

In our work, we combine the study of the effect of online reviews and health inspection on customer demand. A joint analysis is important at this point because as Kang et al. (2013) and Harrison et al. (2014) have shown, online ratings and health inspection scores are correlated: when online reviews are positive, a restaurant is also more likely to be found clean by inspectors. By separately studying the effect of one measure of restaurant quality (e.g. health inspection scores) on customer demand, the estimates can actually confound the true effect of health grade with the separate effect of online reviews. In our empircal strategy we provide evidence that indeed the effect of health grades on demand is close to zero when the text of recent Yelp reviews is taken into account.

By incorporating Yelp review text as signals of restaurant quality, our work also follows a more recent research trend that uses text as data in a broad set of applications (Taddy (2013), Taddy (2015), Gentzkow et al. (2017), and Greenstein et al. (2016)).

Our work can inform the debate over the regulation of peer-to-peer markets. Recent papers have focused on the welfare benefits of flexible labor for Uber drivers (Chen et al. (2017)) and passengers (Cohen et al. (2016)). Farronato and Fradkin (2016) study the welfare implications from Airbnb for travelers, Airbnb hosts, and hotels jointly. But still little is known about the role that online reviews have in providing adequate information about providers' quality as an alternative to regulatory screening (Einav et al. (2016)). We are among the first to shed light on potential substitutabilities between online reviews and regulation.

3 Theoretical Framework

In this section we present a simple theoretical framework of a market with asymmetric information. Health inspections and online reviews constitute two distinct signals of restaurant quality, which affect demand and in turn restaurant's incentives to invest in quality.

Let us consider a single restaurant, whose demand is a function of its underlying quality θ.

Quality is a multi-dimensional vector, and can include food taste, hygiene, and maintenance. Consumers have some prior beliefs about restaurant quality θ_0.

In the absence of regulation or reputation mechanisms, restaurant demand is $D(\theta_0)$.

Consider the restaurant's quality choice. The quality investment cost function is denoted by $c(\theta)$. If restaurant's quality investments do not affect customer demand – $D = D(E(\theta|\theta_0))$ for any θ –, restaurants have no incentives to increase their quality from the minimum level, and at equilibrium θ_0 will coincide with this minimum quality level.

Now consider the role of regulation. Inspections guarantee some minimum quality level denoted $\underline{\theta}$. If that is their only objective, in equibrium customer beliefs about restaurant quality will be $\theta_0 = \underline{\theta}$, and this also coincides with the true quality level. Demand in this case would be $D(E(\theta|\theta_0, \underline{\theta}))$.

Inspections can also provide a signal of restaurant quality to inform consumers, $s_i(\theta)$. Consumers update their beliefs according to this signal and customer demand becomes $D(E(\theta|\theta_0, \underline{\theta}, s_i))$.

Online reviews provide an additional signal of quality $s_y(\theta)$, which is incorporated into demand $D(E(\theta|\theta_0, \underline{\theta}, s_i, s_y))$.

Customers solve an information aggregationg problem. In the empirical section we want to test whether $E(\theta|\theta_0, \underline{\theta}, s_i, s_y) = E(\theta|\theta_0, \underline{\theta}, s_y)$, i.e. demand is not affected by health grades.

Now recall that inspections also guarantee some minimum standards of quality, so eliminating inspections would eliminate $\underline{\theta}$. This implies that we cannot completely rule out that in the absence of inspections hygiene quality would not drop. In the empirical section, however, we show suggestive evidence that information from Yelp reviews incentivizes restaurants to be cleaner.

4 Data

This section decribes the data on health scores, online reviews, and demand that we use in our empirical analysis.

First, data on health inspections come from the New York City Department of Health and Mental Hygiene (DoH). As the DoH describes on its website,[2] *"[it] conducts unannounced inspections of restaurants at least once a year. Inspectors check for compliance in food handling, food temperature, personal hygiene and vermin control. Each violation of a regulation gets a certain number of points. At the end of the inspection, the inspector totals the points, and this number is the restaurant's inspection score"*. More points imply more violations, so

[2] `http://www1.nyc.gov/site/doh/services/restaurant-grades.page`

a higher *health score* implies worse restaurant health quality. We obtained inspection level data directly from the DoH between July 2007 and September 2016.

The DoH performs inspections of restaurants, coffee shops, bars, nightclubs, most cafeterias, and fixed food stands in the five boroughs of New York. The three stated objectives of the inspection program are: 1) give consumers better information about sanitary conditions of restaurants 2) improve restaurants' food preparation practices 3) reduce foodborne illness attributable to restaurants.

We have data at the level of each violation code. There are about 50 violations codes that check restaurants on multiple dimensions: vermin (e.g. evidence of mice), food temperature (hot food item not held at or above 140F), facilities (improper sewage disposal system), food handling (raw food not properly washed), overall hygiene (inadequate personal cleanliness of staff), contamination (worker does not wash hands throughly), and regulatory (wash-hand sign not posted).[3]

On July 27th 2010 the DoH started assigning a health grade, A through C, following each inspection. Since then, restaurants must display their grade card conspicuously enough for customers to see. The program uses dual inspections to help restaurants improve before being graded. Every year a restaurant undergoes an inspection cycle. An inspection cycle is a series of inspections consisting of an initial inspection, possibly followed by a re-inspection and compliance inspections that lead to a letter grade update. An initial inspection is followed by a re-inspection several weeks later for restaurants that do not receive an A grade on initial inspection. A score of less than 14 points on either initial or re-inspection results in an A grade; with a score of more than 13 points on initial and 14-27 points on re-inspection a restaurant receives both a B grade and a "Grade Pending" card; with a score of more than 13 points on initial and 28 or more points on re-inspection a restaurant receives both a C grade and a "Grade Pending" card. The restaurant may post the grade card, the pending card or both. A restaurant may attempt to adjudicate its pending grade to an A or B grade at an administrative tribunal.[4]

The DoH inspection data also provide us with restaurant level characteristics, such as the type of cuisine, the type of restaurant, and date of entry and exit.

Second, we scrape online reviews from Yelp.com. Yelp contains business information, such as zip code and phone number, and a historical record of reviews. From this record we are able to construct the average rating score of a business at any point in time. We will call this the *Yelp rating*. We also record the text of each review, and we describe in Section

[3]For more details, see `https://fivethirtyeight.com/features/how-data-made-me-a-believer-in-new-york-citys-restaurant-grades/`.

[4]More details on inspection regulation and grading can be found at `http://www1.nyc.gov/assets/doh/downloads/pdf/rii/inspection-cycle-overview.pdf`.

?? how to summarize the health information contained within.

Third, we obtain demand data from OpenTable. We have scraped OpenTable data since April 2013. For every day and restaurant on OpenTable, we have information on whether the restaurant had a table available for 2 people around 7PM.

We match businesses from the DoH with businesses on Yelp and OpenTable using Yelp search algorithm, and matching on restaurant name, address, and phone number.

4.1 Descriptives

Overall, of the 57,062 individual businesses present in the DoH data, 58 percent were matched to a business on Yelp, and 4.4 percent were matched on OpenTable. Table 1 presents descriptive statistics at the restaurant level for the three samples: all DoH restaurants, restaurants with Yelp reviews, and restaurants available for booking on OpenTable. Relative to all restaurants inspected by the DoH, restaurants with Yelp reviews tend to be more concentrated in Manhattan, are less likely to be fast food restaurants, and are less likely to go out of business within our sample period. That is even more true for restaurants on OpenTable.

Initial inspections occur throughout the year, with a lower number of inspections during vacation periods (summer and Christmas season). The interval between inspection cycles depends on the sanitary condition of the restaurant during the previous inspection. Namely, if a restaurant has an A-grade, it will be inspected again after approximately one year since the prior initial inspection. If a restaurant has a B-grade, it will be inspected after 5-7 months since the most recent re-inspection. If it has a C-grade, it will be inspected 3-5 months since the most recent compliance inspection. In practice, as pictured in Fig. 8 there is substantial amount of variability in the time interval between inspections. Such variability is intentional, given that inspectors show up unannounced to evaluate restaurants' sanitary conditions.

The distribution of violation scores at initial inspection is depicted in the top panel of Fig. 9. Only 25 percent of restaurants obtain an A grade during the initial inspection. The other restaurants obtain a violation score corresponding to a B (43 percent) or a C (32 percent). Those restaurants whose score would imply a B or a C-grade are re-inspected within a few weeks. After re-inspection, well over half of the restaurants get to display an A-card. As the bottom panel of Fig. 9 shows, 73 percent of restaurants end an inspection cycle with an A grade, 20 percent end it with a B grade, and only 7 percent end it with a C grade.

To study how Yelp ratings and health grades affect customer demand, Fig. 10 displays a restaurant's average probability of being sold out at 7PM on any given night on OpenTable,

as a function of a restaurant letter grade and average Yelp rating. At least at a descriptive level, there seems to be a small positive correlation between hygiene and customer demand, and a larger positive correlation between Yelp ratings and customer demand. In Section 7 we show that this result holds true once we control for other observable characteristics.

The empirical strategy relies on variation in posted hygiene grades and Yelp ratings. Table. **??** summarizes their joint distribution, which displays some variation across both signals of restaurant quality. We now turn to describe how we plan to exploit this variation to understand the effect of health grades and online reviews on demand, and the effect of online reviews on restaurants' hygiene incentives.

5 Incorporating Text Data

A Yelp rating typically reflects a consumer's aggregate utility, and it is a function of several components such as food quality, service quality, price, and – possibly – hygiene conditions. The utility function may also contain interactions between these components: for example, Lehman et al. (2014) show that ratings are less susceptible to unhygienic conditions for restaurants that are perceived as being more authentic. Therefore, the extent to which Yelp ratings can capture restaurant hygiene will depend on the weight consumers place to hygiene compared to other preferences as well as the interactions between such preferences and hygiene.

By contrast, review text often provides a breakdown of a consumer's overall assessment of a restaurant. For example, a consumer might reveal that she gave 4-stars to a restaurant with excellent food because the food arrived lukewarm. Specific to our context, review text might also reveal different types of violations: for instance, a review mentioning lukewarm food might be an indicator of violation code 02A ("Food not cooked to required minimum temperature.") Therefore, unlike ratings, review text potentially contains information that we directly relates to presence specific hygiene violations.

To incorporate review text in our analysis we need to reduce the dimensionality of our text data. Yelp reviews contain hundreds of thousands of unique words, and using each one of these as a covariate is impossible as we would end up with more covariates than observations in our data. We solve this problem by estimating low dimensional representations of reviews that capture variation with respect to various components of consumer utility that we want to measure. Specifically, we posit that the text of a consumer review is function of restaurant characteristics, hygiene conditions, and consumers' idiosyncratic experience at the restaurant.

Indexing reviews by j, and words by k, let $c_j = \{c_{jk}\}$ be a vector of word counts in review

j. We model the probability of c_j using a multinomial logit:

$$p(c_j|x_{i(j)}, m_j) = \text{MN}(c_j; q_j, m_j) \tag{1}$$

where c_j a vector of word counts in review j, $x_{i(j)}$ is a vector of covariates for restaurant i, $m_j = \sum_k c_{jk}$ is the length of review j, and q_j is vector of word probabilities. The probability of word k is given by

$$q_{jk} = \frac{e^{w_k}}{\sum_l e^{w_l}}$$

where $w_k = \alpha_k + x'_{i(j)}\phi_j$.

We are particularly interested in uncovering words related to specific hygiene violatons. For this reason, we associate each restaurant inspection with reviews that were submitted up to 3 months prior to the inspection. The word counts in these reviews constitute our response variable c_j. The independent variables include a vector of dummies, one for each violation code, indicating the presence (or absence) of each violation in that inspection. We also include restaurant fixed effects as controls, as well as the numeric score associated with each rating.

Estimating the multinomial logit model is prohibitely expensive because the coefficients for each word in c_j depend on the coefficents of each other word in c_j. Instead, we use the distributed multinomial regression framework of Taddy (2015), which allows efficient estimation this model under simplifying assumptions that work well in practice. The estimated model yields a set of coefficients for each word-covariate combination.

Finally, we use the estimated model to reduce the dimensionality of each review. To do so, we compute *sufficient reductions*, which are projections of the text that capture variation with respect to the covariates $x_i(j)$. In doing so, we reduce the dimensionality of the text data from hundreds of thousands to the number of covariates – *i.e.*, the number of violation codes – in our model. The reductions are simply the product of the coefficients and the word counts, $\nu_j = \phi_j c_j$. These reductions, as Taddy (2015) shows, have the property that $x_{i(j)} \perp c_j|\phi_j c_j$. In other words, the reductions are sufficient statitics for the covariates $x_{i(j)}$: conditional on the reductions, review text is orthogonal to violation codes. Based on this property, we work with sufficient reductions and ignore review text.

Figure 11 displays simple correlations between the sufficient reductions and the occurence of some common violations. We observe that text appears to predict the occurence of certain violations better than others. Moreover, the violations that appear to be more predictable are the ones that relate to directly to food – intuitively these are also the violations we would expect consumers to be most likely to notice.

6 How are inspectors assigned to restaurants?

We want to test whether inspectors are assigned to restaurants randomly, or at least randomnly conditional on observables. To do so, we will compute two probabilities. First, given a restaurant characteristic X we will compute $P(X = x)$, which is simply the fraction of x restaurants in the population. If inspectors are assigned to restaurants randomly conditional on X then we should expect that $P(X|Z) = P(X)$, where Z denotes inspector. The conditional probability $P(X|Z)$ is the second quantity we estimate from the data.

Figures

7 Empirical Strategy and Results

We want to estimate the causal effect of online reviews and regulation on demand – where do consumers choose to eat? – and supply – how do restaurants adjust their hygiene quality as a function of online reviews. We tackle the two questions separately.

7.1 Effect of Online Reviews and Hygiene Inspection on Consumer Demand

In order to analyze the effect of information contained in online reviews and health inspections on demand, we estimate the following equation:

$$Y_{it} = \alpha_1 X_i + \alpha_2 X_{it} + \beta doh_{it} + \gamma yelp_{it} + \epsilon_{it}$$

where i denotes a restaurant, and t denotes a day. Here, doh_{it} is a vector of dummies corresponding to the letter grade assigned by the DOH (and displayed conspicuously within the restaurant premises), $yelp_{it}$ is a vector of dummies corresponding to the Yelp average rating rounded to the nearest half star. So for example, a 2.73 average rating would be rounded to 2.5 stars. Controls include time-invariant restaurant characteristics such as cuisine and zipcode, and time controls such as day of the week and quarter fixed effects.

We estimate the effect on two outcome variables, the probability of being sold out at 7PM on OpenTable, and the probability of going out of business within 3 months of a health inspection. We consider both measures to be correlated with overall demand, with the sold out probability being a finer measure. At the end of this section we provide some evidence supporting the hypothesis that sold out probability and restaurant exits are correlated with restaurant demand.

We are worried about the possibility that some characteristics, unobservable to the econometrician, but observable to both the inspector and restaurant guests, might both affect the inspection outcome and guests' decision of where to eat. These characteristics could include, for example, the presence of mice, or adulterated food, that are both visible by the inspector, and by the guests eating at the restaurant. To separate the causal effect of health grade on demand, we use an instrumental variable strategy that relies on the random assignment of inspectors to restaurants. We use two instruments. The first is a dummy for whether the current inspector evaluated the same restaurant in the previous inspection cycle. This instrument has been shown to be a good predictor of inspection outcomes, since repeated interactions lead to more generous evaluations (Ibanez and Toffel (2017)). The second instrument is the average violation score that an inspector assigns to restaurants other than the restaurant under evaluation. Effectively, this is a measure of inspector severity, which is uncorrelated with the underlying quality of a restaurant under random assignment (Bartik (1991)).

Table 3 shows the estimates of linear probability models where the outcome is a dummy for whether a restaurant is sold out on OpenTable around 7PM on day t.[5] The results show that restaurants displaying an A-card are 5.4 percentage points more likely to be sold out on a given night relative to restaurants with a C-card. This represents a 28% increase in sold out probability relative to the .19 baseline. Yelp ratings seem to have a larger effect. Having a 4.5-star rating is associated with doubling of the probability of being sold out relative to having no Yelp reviews. Having a 2-star reviews is associated with a 95% redution in the sold out probability. When we control for restaurant fixed effects, the effect of letter grade disappears. Moreover, having Yelp reviews is always associated with a higher probability of being sold out relative to having no reviews, and higher rated restaurants also sell out more often.

When we instrument for the letter grade with inspector characteristics, the effect of letter grade on sold-out probability disapears (Table 4). This result suggests that our OLS estimates were simply picking up hygiene characteristics that were visible by customers in ways other than the hygiene score. Intuitively, given that OpenTable allows customers to book a restaurant online before seeing the letter grade displayed at the door, we would expect that the effect of hygiene inspections would be small at best. On the other hand, given that OpenTable users are likely to make their decisions based on online reviews, it is reassuring to see that Yelp stars have a large and significant effect on restaurant sold-out probability.

[5]In the current draft, we only include the tables of the linear probability models because the coefficients are easily interpretable as marginal effects. Qualitatively the results do not change when we estimate logit models.

The results found using the sold-out probability are qualitatively similar to the results using restaurant exit as an outcome, although this time the effects of hygiene grades and Yelp reviews are comparable in size. Table 6 shows the estimates of linear probability models where the outcome is a dummy variable for whether a restaurant goes out of business within 90 days since a health inspection by the DoH. Improvements in letter grades and Yelp star ratings are associated with a reduction in the probability that a restaurant goes out of business, and the results hold true after adding a large set of restaurant and time controls, and even after instrumenting for hygiene grades (Table 7).

Given our instrumental variable approach, and the fact that we have two separate instruments for hygiene grade, we can both test whether our instruments are weak, and test the assumption that hygiene grade is exogenous in the second stage regression. Tables 5 and **??** present our first-stage regression results. The regressions are qualitatively similar, but we report both sets of estimates because the samples on which they are estimated differ. Table 5 focuses on restaurants on OpenTable, while table **??** considers all restaurants in NYC. The F-statistics are always above 166, and inspector's strictness, as measured by inspection scored assigned by the same inspector to other restaurants, has a negative and significant effect on the probability that a restaurant obtains an A-grade. When we consider all the restaurants in Table **??**, the dummy variable for repeated interactions between the inspector and the restaurant during subsequent inspections, has a positive and significant effect on the probability of obtaining an A-grade. Finally, the Durbin-Wu-Hausman test rejects the null hypothesis that letter grade is exogenous to the second-stage in most of our specifications.

One might wonder whether the two outcome variables – sold out on OpenTable and going out of business – are good measures of demand. To the extent that restaurant success is the unique common underlying factor influencing both whether a restaurant is sold out on OpenTable and whether a restaurant goes out of business, we should expect these two measures to be negatively correlated. To verify this, we estimate three-month survival curves of two groups of restaurants: those with a low probability of being sold out on OpenTable, and those with a high probability of being sold out on OpenTable. For every quarter, we compute the cumulative probability of being sold out on OpenTable up to the beginning of that quarter. We then estimate Kaplan-Meier survival curves with two strata, above and below the median probability of being sold out. Fig. 12 plots these curves, together with confidence intervals. The blue curve at the top shows the survival probability for restaurants that start a quarter with above median probability of being sold out. The red curve at the bottom shows the survival probability for restaurants that start a quarter with below median sold-out probability. The fact that the red curve is always below the blue curve demonstrates that booking rates on OpenTable are correlated with exit. The higher the booking rate, the

lower the probability that a restaurant will go out of business within 3 months.

When we summarize the information contained in Yelp reviews with only the average star rating, we conclude that both the letter grades and Yelp reviews predict where consumers choose to eat. But the information contained in Yelp reviews is richer than the average star rating. In particular, users can read the text of the reviews, which often include the reasons for a particular star rating. Some of these reasons might involve hygienic conditions of the restaurant, that are currently captured by the positive coefficient of the A-card on the probability of being sold out on any given night, and on the probability of staying in business.

7.2 Effect of Online Reviews on Restaurants' Hygiene

In order to assess whether restaurants change their hygiene quality due to online reputation, we take advantage of the fact that the DoH inspects restaurants that have different degrees of visibility on Yelp. We also take advantage of the fact that the DoH evaluates restaurants on different dimensions of quality – different violation codes. The degree to which each violation code can be detected through online reviews differs across violation codes, as we have shown in Section 3. We define a restaurant *visibility* on Yelp to be the number of Yelp reviews received in the 90 days prior to a DoH inspection. We define a violation code *detectability* via Yelp to be the improvement in predicting the occurrence of that violation code from the text of Yelp reviews. We then run an OLS regression of the following type:

$$violation_{vit} = \nu_v + \nu_i + \alpha X_{vit} + beta_v * on_yelp_i + \gamma_v * has_reviews_{it} + \delta_v * log(nr_reviews_{it}+1) + \epsilon_{vit} \quad (2)$$

where $violation_{vit}$ is equal to one if restaurant i was found violating code v during inspection t. on_yelp_i is a dummy variable equal to one if the restaurant has a Yelp page, with at least one review. $has_reviews_{it}$ is equal to one if the restaurant has received some reviews in the 90 days prior to inspection t. $log(nr_reviews_{it} + 1)$ is the logarithm of the total number of reviews received in the 90 days prior to inspection t.

The coefficients of interest are γ_v and δ_v, one per violation. These coefficients measure how likely restaurants on Yelp are to violate on different measures of hygiene as their short-tem visibility on Yelp increases. We test two hypothesis. The first hypothesis is that γ_v is smaller for detectable violations, where detectability is defined as the improvement in predicting violation v using the text of Yelp reviews. The measure of detectability of each violation code is the result of our text analysis. The two hypotheses together will confirm whether restaurants on Yelp will tend to violate less on violation codes that are detectable by consumers via Yelp reviews. This test would provide suggestive evidence that even

if inspectors did not check for detectable violation codes, online reviews could be a good alternative to ensure high quality on those dimensions.

To clarify our hypotheses, take two types of violations, e.g. XXXX

Our causal statements take advantage of the data at hand. We have information on restaurants' violation performance on multiple dimensions of hygiene quality. If we only had the total inspection score, we could not separate hygiene in dimensions that are visible to the consumers and dimensions that are not visible to the consumers. In addition, simply running a regression of aggregate violation score on restaurants' Yelp visibility would suffer from endogeneity. Presumably, restaurants that receive a higher number of reviews on Yelp in the 90 days prior to a health inspection might be better in unobservable (to the econometrician) dimensions of quality that both affect their success with costumers and their hygiene performance. Our specification at the violation code level does not suffer from this endogeneity problem. X_{vit} includes, among other controls, the total inspection score during inspection t, so γ_v and δ_v measure the restaurant propensity to violate on specific hygiene dimensions, conditional on their overall hygiene level.

In general, restaurants on Yelp tend to violate less during health inspections (Figure 13). For their propensity to violate on specific codes, see Figures 14 and 15.

Until now we have simply compared restaurants on Yelp and restaurants not on Yelp, but Table 1 demonstrates that the two groups are significantly different along many observable dimensions. In order to minimize bias we match Yelp and non-Yelp restaurants based on their propensity score. We use the covariates in Table 1 to estimate the probability of treatment assignment for all our restaurants.

8 Conclusion

In a world where consumers cannot distinguish between high and low quality service providers before engaging in a transaction, the market can break down or consumers may engage in transactions that are too risky from a public policy perspective. In these contexts, regulation that guarantees minimum quality standards and informs consumers about different quality levels can help solve the market failure. However with the diffusion of online reviews – a new cheap way to collect provider quality information from consumers' past transactions –, online reputation provides an additional force to reduce information asymmetries.

In the context of restaurants in New York City, we have shown that Yelp reviews contain the type of information on restaurants' hygiene that consumers care about when choosing when to eat. In other words, the additional signaling value of health inspections in affecting where consumers eat is limited at best.

We have also shown that there are differences in the degree to which Yelp reviews can predict health violations. Specifically, Yelp reviews contain relevant information to predict food handling violations, but are poor predictors of facilities and maintenance violations.

In the next version of this paper, we plan to test whether the visibility of some dimensions of hygiene (food handling) drive restaurants to comply with minimum hygiene standards exactly in those more visible dimensions. Such a result would imply that regulators can better focus their resources to regulate quality dimensions that consumers cannot see (facilities and maintenance), while leaving the more visible dimensions to the *eyes of the crowd*.

Of course, relying on online reviews to guarantee hygiene quality raises a new set of important questions. In particular, the online platform collecting reviews has two important roles to play: it must convince customers to share their past experiences, and it must be able to aggregate and summarize this information so that it is useful for future transactions, incentivizes providers to continuously engage in high quality transactions, and does not create excessive barriers to entry for new providers.

At the information gathering stage, at least two issues are worth considering. The first is that reviews are not necessarily representative of all transaction experiences. Research has shown that buyers are more likely to leave a review after a positive experience. They have also shown that if consumers personally interact with service providers – on Airbnb for example, when guests and hosts sometimes meet in person –, they are less likely to leave negative reviews. And if they fear retaliation after a negative review, the bias is even worse. The second issue is that providers have incentives to manipulate the reviews, by offering discounts on future purchases, or by directly writing fake reviews about their business or their competitors'.

At the information aggregation stage, platforms need to consider how to display reviews, and how to rank service providers on the basis of those reviews. Aggregate ratings can be too coarse a measure of quality, and individual reviews can be too many to read, so inevitably consumers focus on just a (potentially non-representative) subset. In addition, if quality changes over time or if certain reviews, or non-reviews, are more useful than others, aggregation needs to incorporate those features. We leave the analysis of platform incentives to provide unbiased reviews to future research.

References

G. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, 1970. URL `http://www.jstor.org/stable/1879431`.

T. J. Bartik. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, 1991. ISBN 9780880991148. URL http://ideas.repec.org/b/upj/ubooks/wbsle.html.

L. Cabral and A. Hortacsu. The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics*, 58(1):54–78, 2010. URL https://scholar.googleusercontent.com/scholar.bib?q=info:yCFzWkIArGcJ:scholar.google.com/{&}output=citation{&}scisig=AAGBfm0AAAAAWQuVlWmNPQ4IG33SbsRAwmUj2CsH3Sv9{&}scisf=4{&}ct=citation{&}cd=-1{&}hl=en.

M. K. Chen, J. Chevalier, P. Rossi, and E. Oehlsen. The Value of Flexible Work: Evidence from Uber Drivers. Technical report, National Bureau of Economic Research, Cambridge, MA, mar 2017. URL http://www.nber.org/papers/w23296.pdf.

P. Cohen, R. Hahn, J. Hall, S. Levitt, and R. Metcalfe. Using Big Data to Estimate Consumer Surplus: The Case of Uber. Technical report, National Bureau of Economic Research, Cambridge, MA, sep 2016. URL http://www.nber.org/papers/w22627.pdf.

L. Einav, C. Farronato, and J. Levin. Peer-to-Peer Markets. *Annual Review of Economics*, 8(1):615–635, oct 2016. ISSN 1941-1383. doi: 10.1146/annurev-economics-080315-015334. URL http://www.annualreviews.org/doi/10.1146/annurev-economics-080315-015334.

C. Farronato and A. Fradkin. Market Structure with the Entry of Peer-to-Peer Platforms: the Case of Hotels and Airbnb. 2016.

M. Federman, D. Harrington, and K. Krynski. The impact of state licensing regulations on low-skilled immigrants: The case of Vietnamese manicurists. *The American Economic Review*, 2006. URL http://www.jstor.org/stable/30034649.

M. Gentzkow, B. T. Kelly, and M. Taddy. Text as Data. *NBER Working Paper*, 2017. URL http://www.nber.org/papers/w23276.pdf.

S. Greenstein, Y. Gu, and F. Zhu. Ideological Segregation among Online Collaborators: Evidence from Wikipedians. *NBER Working Paper*, 2016. URL http://www.nber.org/papers/w22744.

C. Harrison, M. Jorder, H. Stern, F. Stavinsky, and V. Reddy. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness—New York City, 2012–2013. *MMWR*, 2014. URL http://www.cdc.gov/MMWr/preview/mmwrhtml/mm6320a1.htm.

X. Hui, M. Saeedi, Z. Shen, and N. Sundaresan. Reputation and regulations: evidence from ebay. *Management Science*, 62(12):3604–3616, 2016.

M. Ibanez and M. W. Toffel. Assessing the Quality of Quality Assessment: The Role of Scheduling. *SSRN Electronic Journal*, 2017. ISSN 1556-5068. doi: 10.2139/ssrn.2953142. URL http://www.ssrn.com/abstract=2953142.

G. Jin and P. Leslie. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 2009. URL http://www.ingentaconnect.com/content/aea/aejmi/2009/00000001/00000001/art00011.

G. Z. Jin and P. Leslie. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451, 2003.

J. S. Kang, P. Kuznetsova, M. Luca, and Y. Choi. Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *EMNLP*, pages 1443–1448. Citeseer, 2013.

D. W. Lehman, B. Kovács, and G. R. Carroll. Conflicting social codes and organizations: Hygiene and authenticity in consumer evaluations of restaurants. *Management Science*, 60(10):2602–2617, 2014.

G. Lewis and G. Zervas. The Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry. 2017.

M. Luca. Reviews, Reputation, and Revenue: The Case of Yelp.Com. *SSRN Electronic Journal*, 2011. ISSN 1556-5068. doi: 10.2139/ssrn.1928601. URL http://www.ssrn.com/abstract=1928601.

M. Luca and G. Zervas. Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science*, 62(12):3412–3427, dec 2016. ISSN 0025-1909. doi: 10.1287/mnsc.2015.2304. URL http://pubsonline.informs.org/doi/10.1287/mnsc.2015.2304.

D. Mayzlin, Y. Dover, and J. Chevalier. Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, 104(8):2421–2455, aug 2014. ISSN 0002-8282. doi: 10.1257/aer.104.8.2421. URL http://pubs.aeaweb.org/doi/10.1257/aer.104.8.2421.

P. Resnick and R. Zeckhauser. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *Advances in applied microeconomics*, 11:127–157, 2002.

P. A. Simon, P. Leslie, G. Run, G. Z. Jin, et al. Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in los angeles county. *Journal of Environmental Health*, 67(7):32, 2005.

M. Taddy. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108(503):755–770, 2013. URL https://scholar.googleusercontent.com/scholar.bib?q=info:THyJGbWP2nEJ:scholar.google.com/{&}output=citation{&}scisig=AAGBfm0AAAAAWQuYdzhTFRyG0fIMAlJFwMw60FwvwHtd{&}scisf=4{&}ct=citation{&}cd=-1{&}hl=en.

M. Taddy. Distributed Multinomial Regression. *The Annals of Applied Statistics*, 9(3):1394–1414, 2015. URL https://scholar.googleusercontent.com/scholar.bib?q=info:9CFVSMWrqbEJ:scholar.google.com/{&}output=citation{&}scisig=AAGBfm0AAAAAWQuYzwMR5EUZjgIaHlo62ql205MMzC9L{&}scisf=4{&}ct=citation{&}cd=-1{&}hl=en.

Figure 1: Independence of inspectors and boro's.

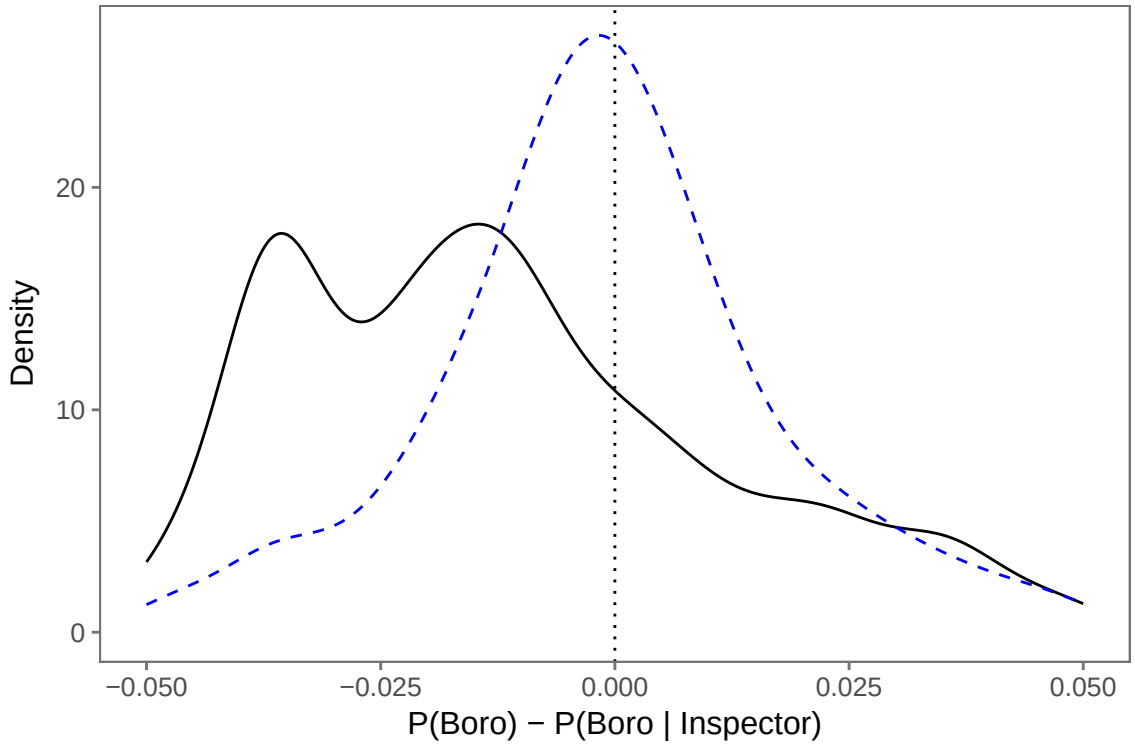


Figure 2: Independence of inspectors and venue.

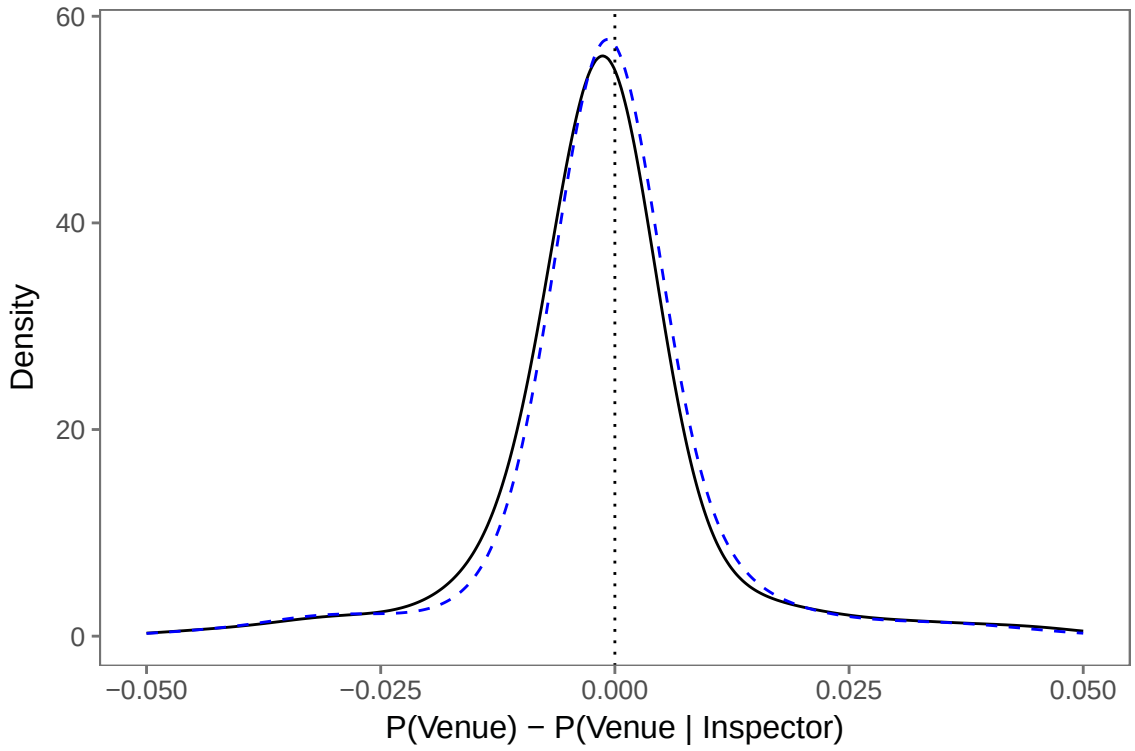


Figure 3: Independence of inspectors and service.

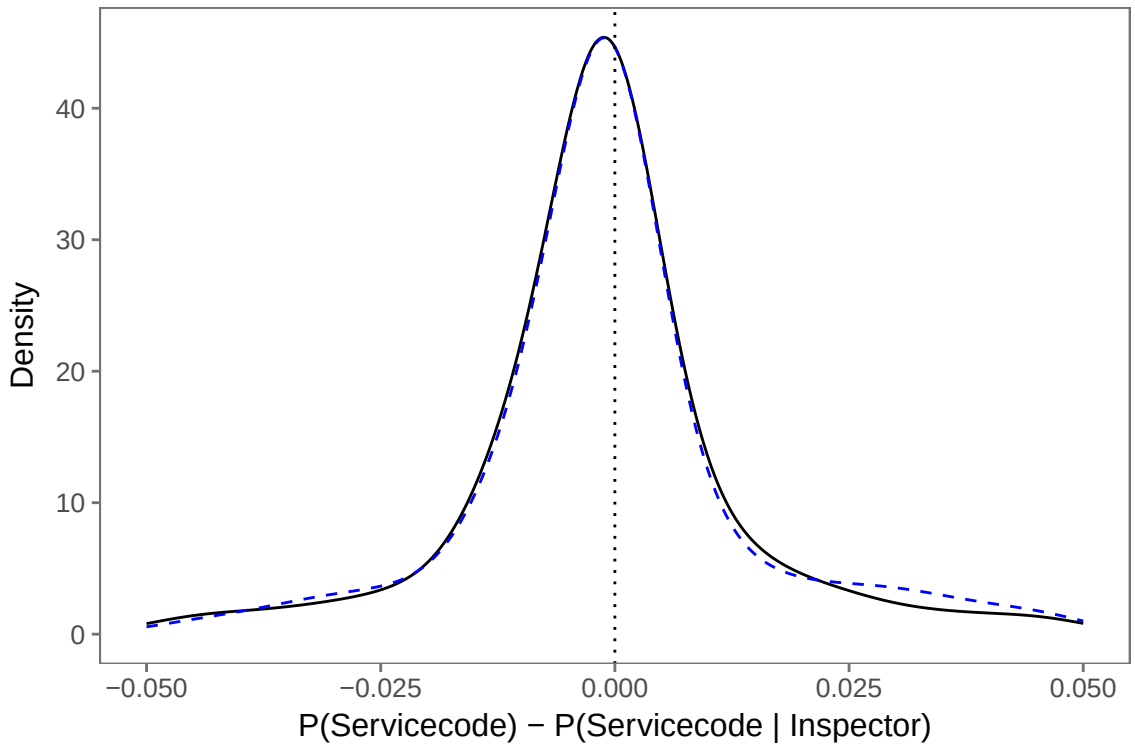


Figure 4: Independence of inspectors and cuisine.

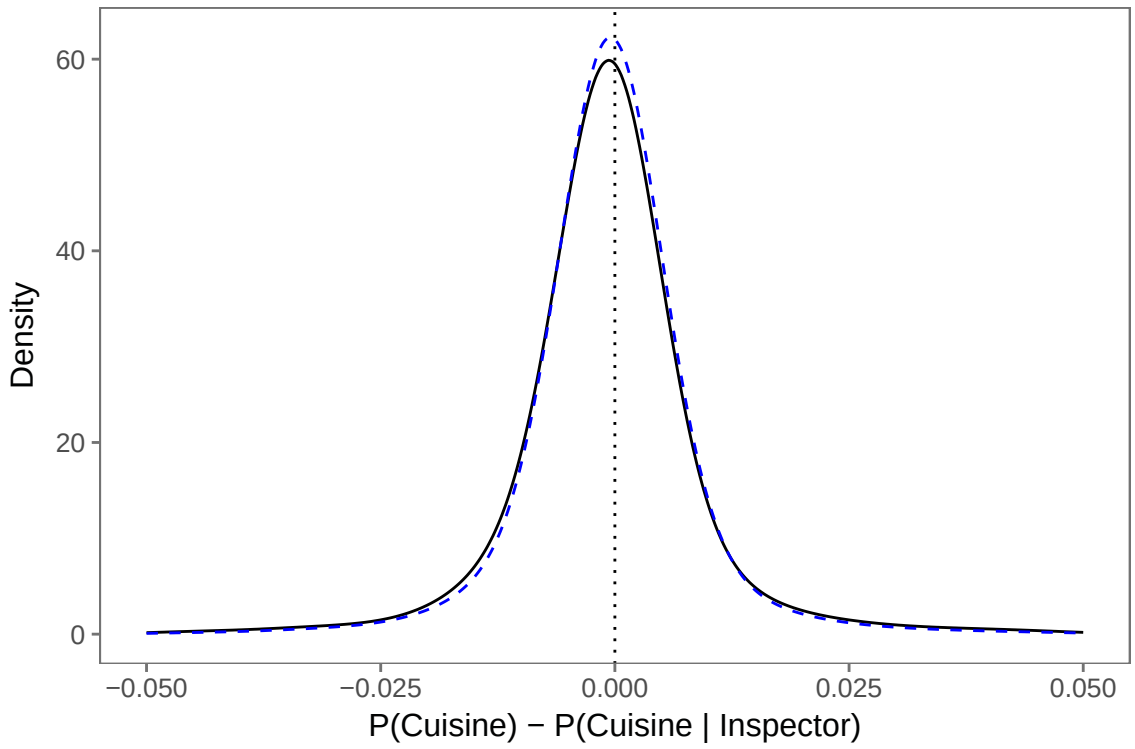


Figure 5: Independence of inspectors and zipcodes.

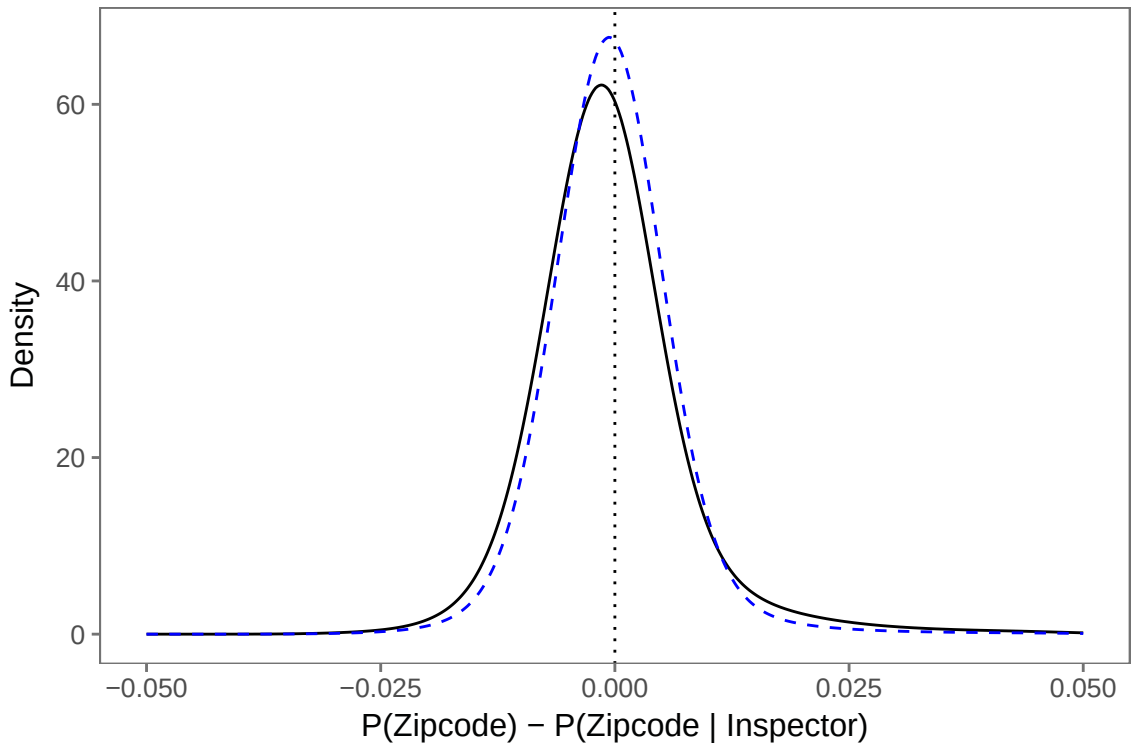


Figure 6: Independence of inspectors and zipcode-cuisines.

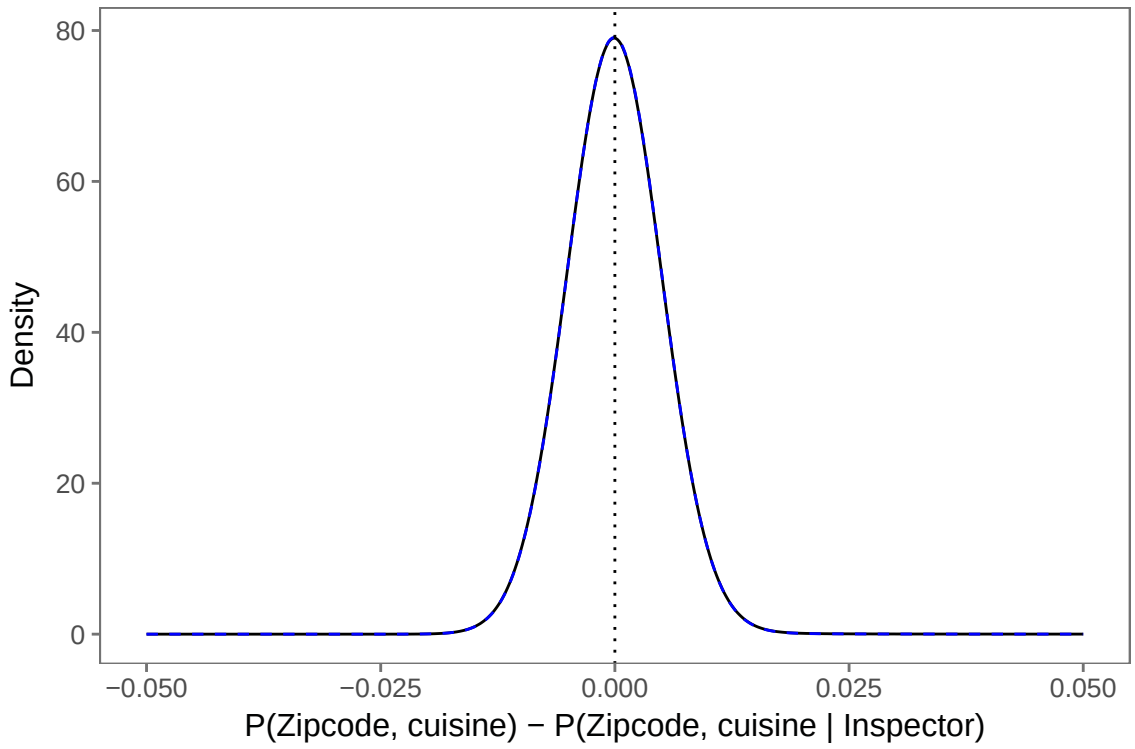


Figure 7: Independence of inspectors and camis.

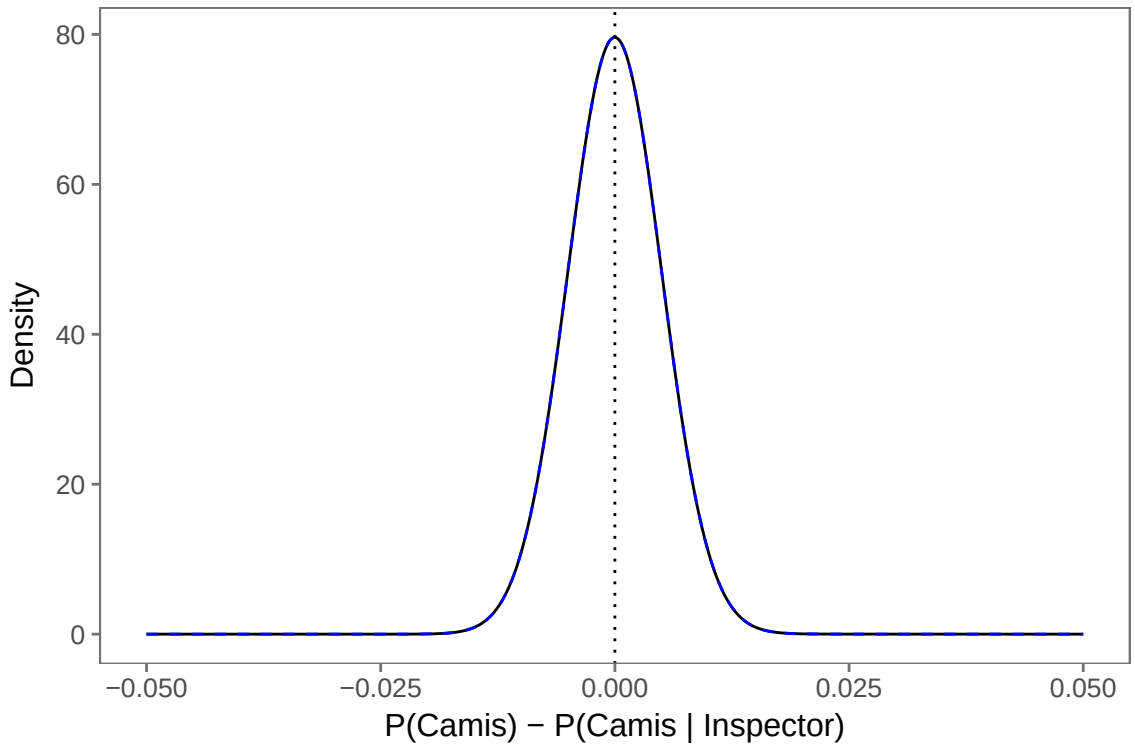
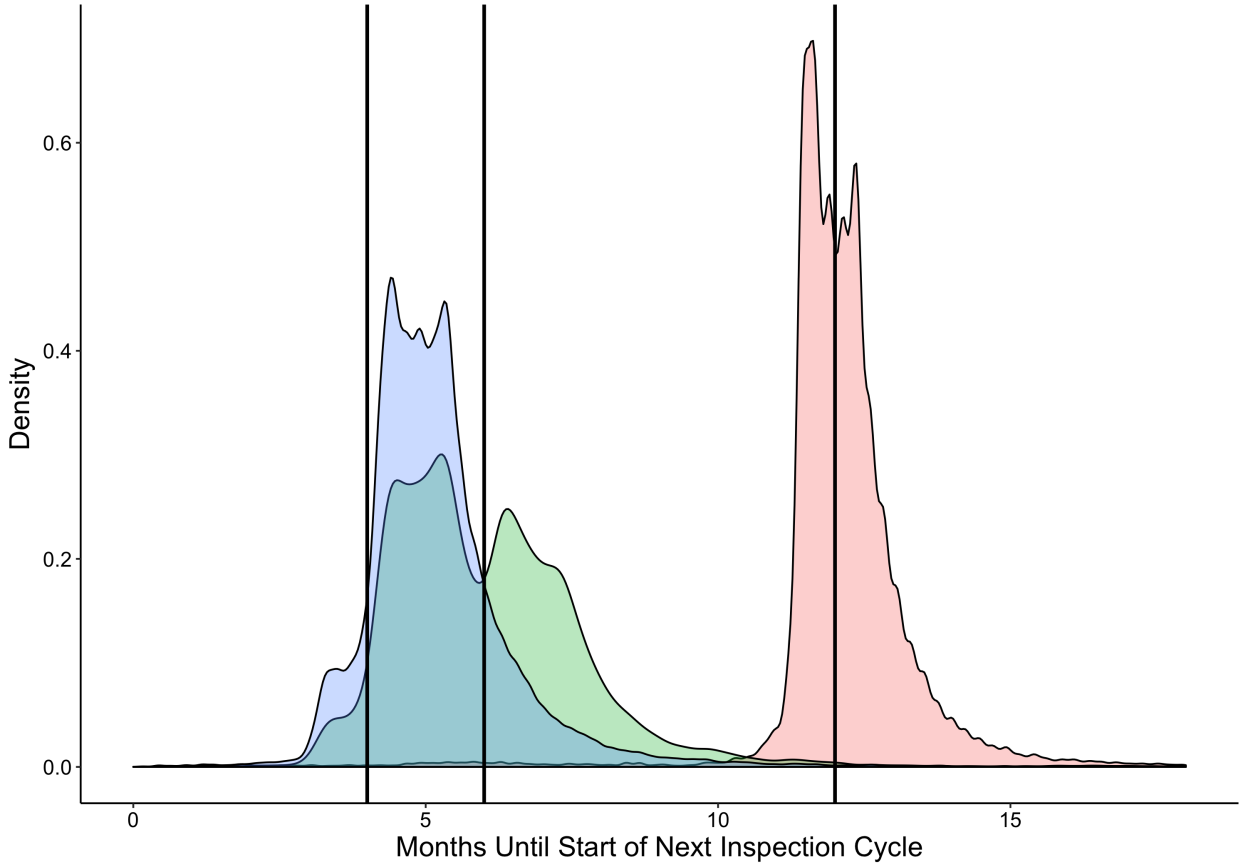
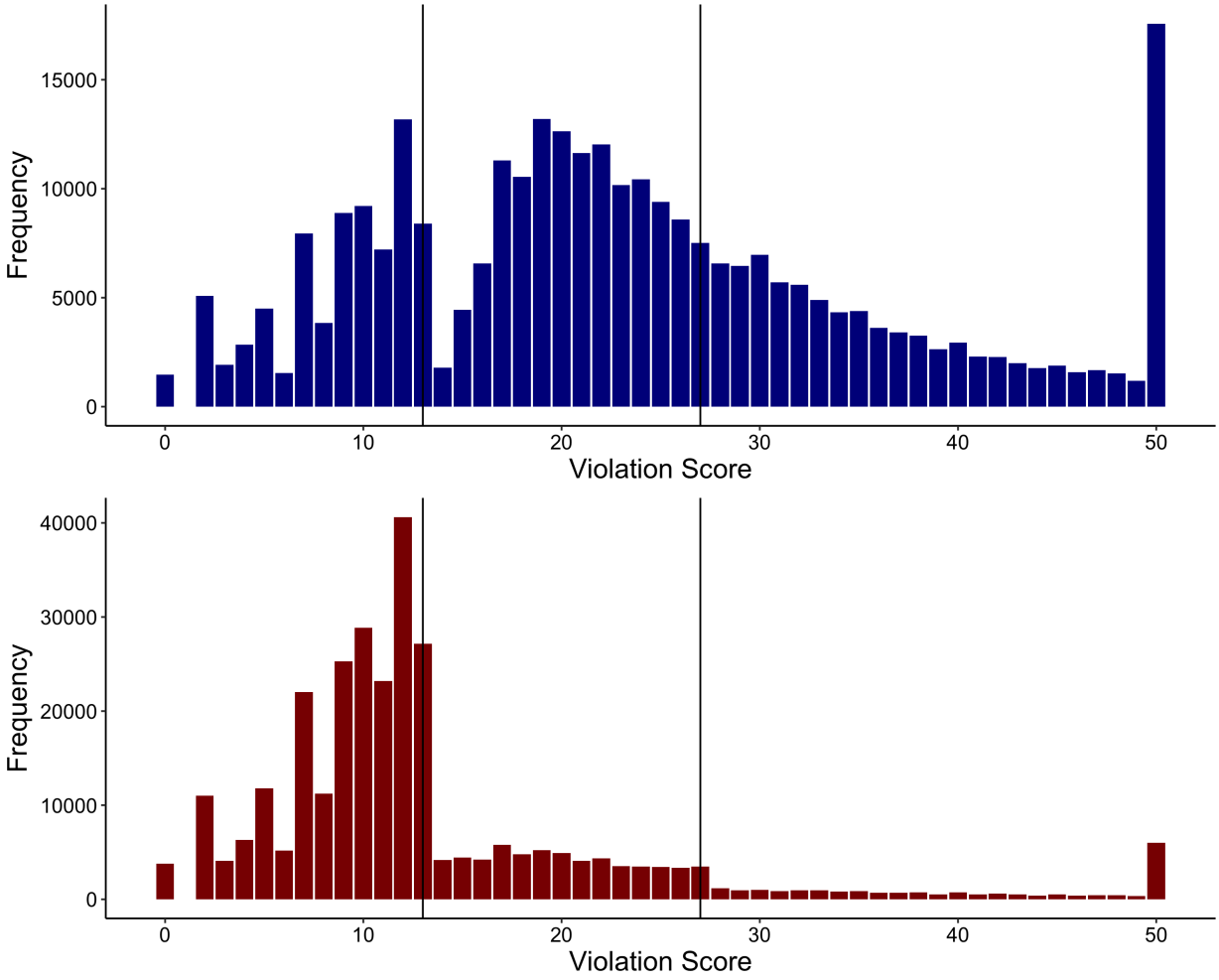


Figure 8: Time Between Inspections



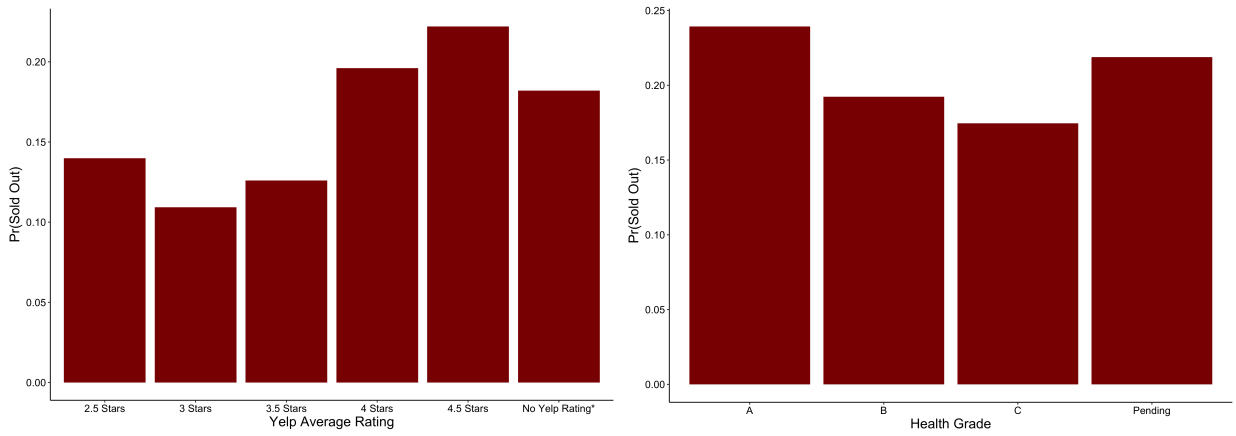
The plot shows the distribution of time between the last inspection within a cycle, and the initial inspection of the following cycle. For restaurants displaying letter grade A (pink), the expected time is 12 months since the initial inspection of the last inspection cycle. For restaurants displaying letter grade B, the expected time if 5-7 months since the most recent re-inspection. Finally, for restaurants diplsaying letter grade C, the expected time is 3-5 months since the last compliance inspection. the plots shows substantial variation in the time between inspections.

Figure 9: Violation Scores after Initial Inspection and after Re-inspection



For each inspection cycle, the top panel shows the distribution of violation scores that restaurants obtain after the initial inspection. The vertical lines correspond to the score thresholds that would assign A-B-C letter grades. Low score automatically give an A-card, while higher scores imply that a restaurant will be re-inspected within a few weeks. The bottom panel shows the distribution of violation scores that restaurants obtain after re-inspection, and it includes those restaurant that obtained an A-grade violation score at initial inspection.

Figure 10: Violation Scores after Initial Inspection and after Re-inspection



The two bar charts show the average probability of being sold out at 7PM on any given night as a function of Yelp reviews (left panel) and hygiene letter grades (right panel). 86% of restaurant-days fall into the A-card group, 9% fall into the B-card group, 2% fall into the C-card group, and 3% fall into the pending-card group. There are virtually no restaurants with average star rating below 2 on Yelp. Most restaurant-days (69%) are in the 3-4 star group, followed by the 4-5 star group (18%), and the 2-3 star or no-rating group (both at 6.5%).

Figure 11: Correlation between health violation and sufficient reductions of Yelp review text

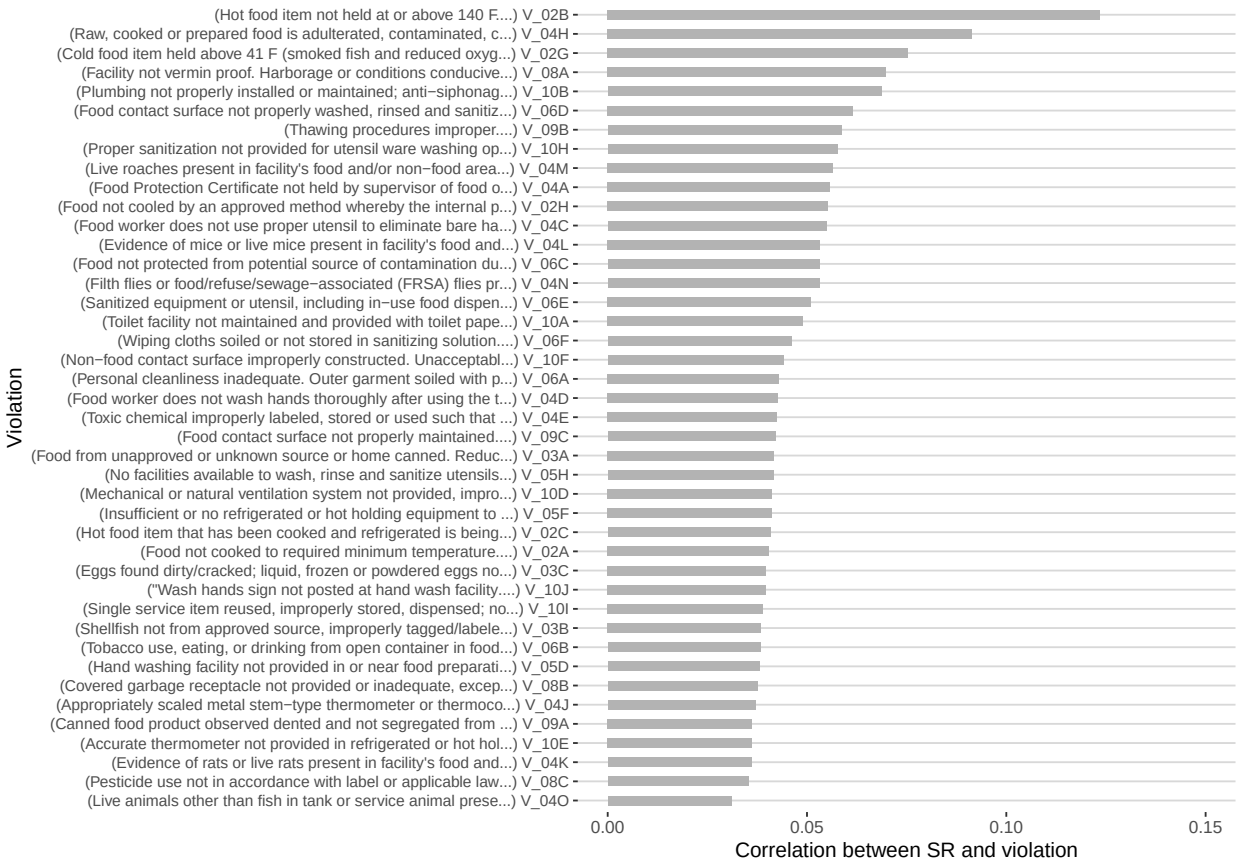
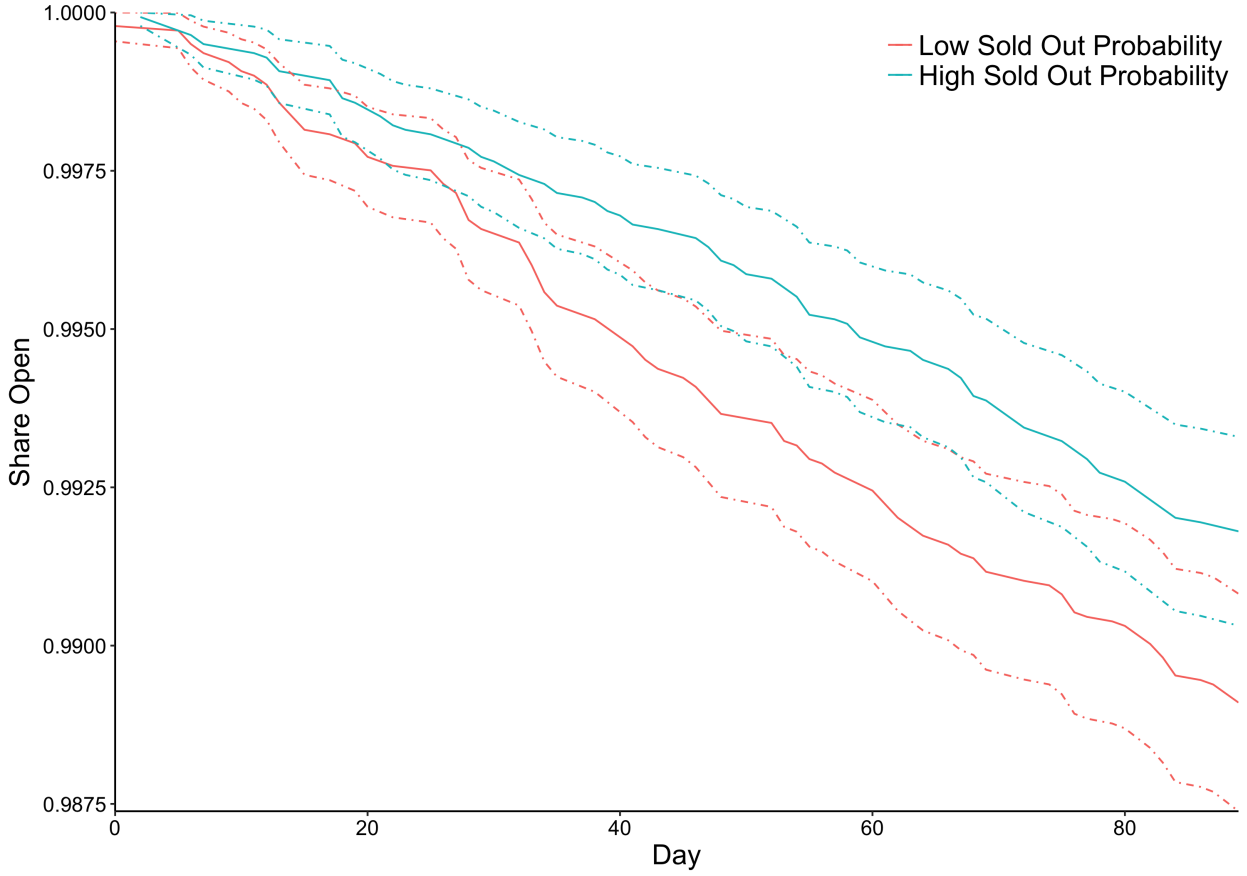


Figure 12: Survival Curves of Restaurants on OpenTable



The figure shows three-month survival curves for restaurants. For every quarter, we compute the cumulative probability of being sold out on OpenTable up to the beginning of that quarter. We then estimate Kaplan-Meier survival curves with two strata, above and below the median probability of being sold out (median across all restaurants and quarters for which we have observations). The blue curve shows the survival probability for restaurants that start a quarter with above median probability of being sold out. The red curve shows the survival probability for restaurants that start a quarter with below median sold-out probability. The fact that the red curve is always below the blue curve demonstrates that booking rates on OpenTable are correlated with exit. The higher the booking rate, the lower the probability that a restaurant will go out of business within 3 months.

Figure 13: Inspection Score and Restaurant Visibility on Yelp

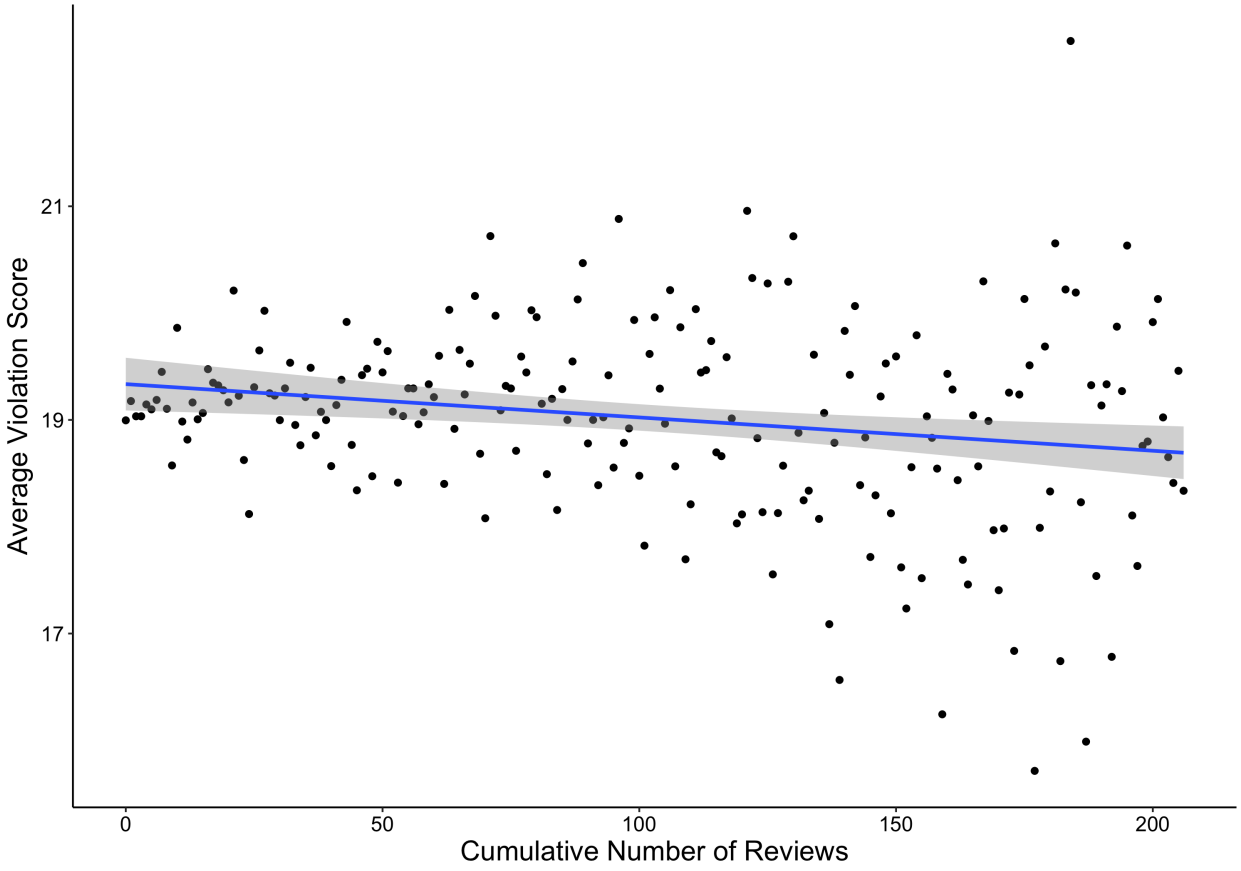
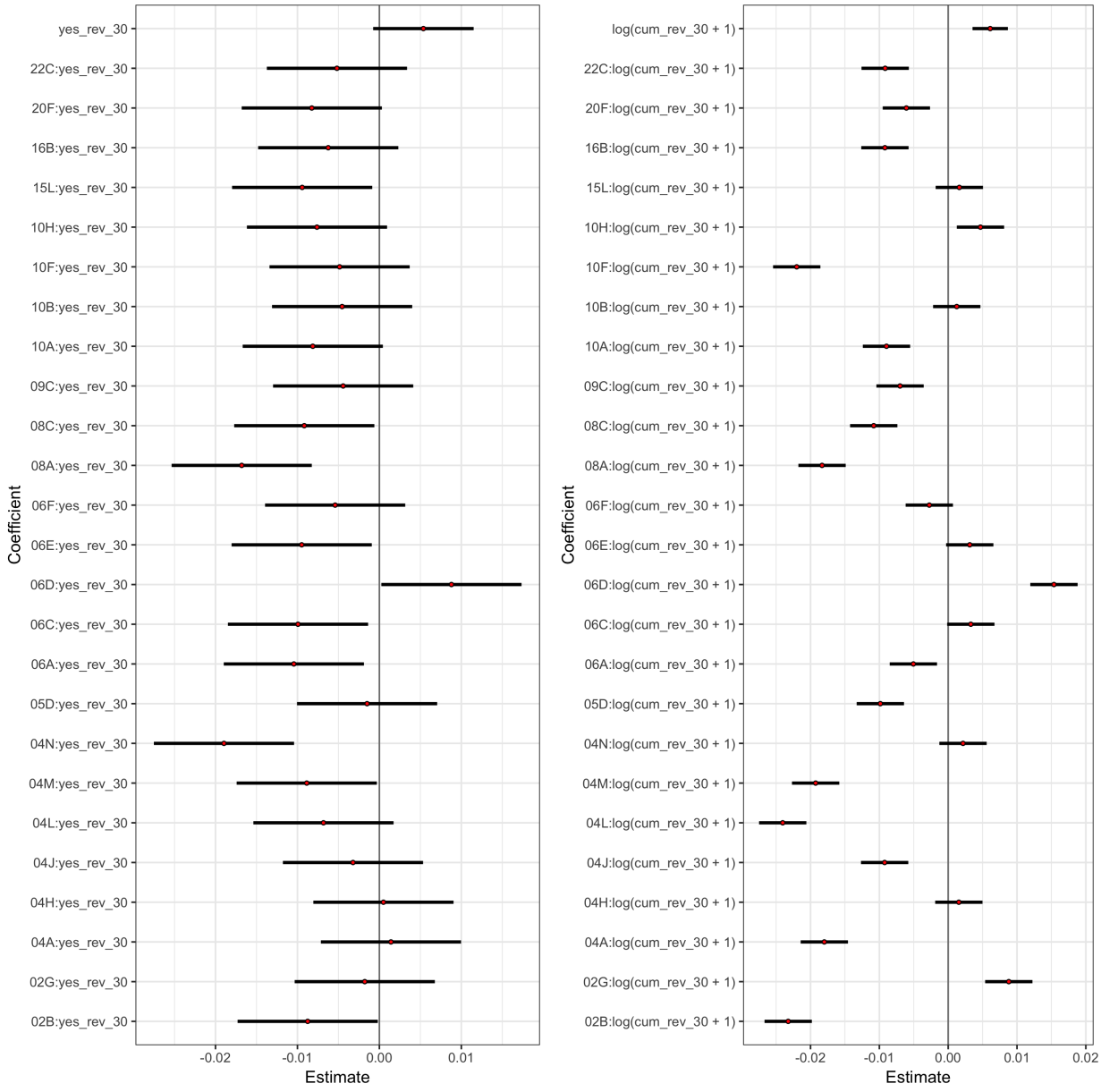
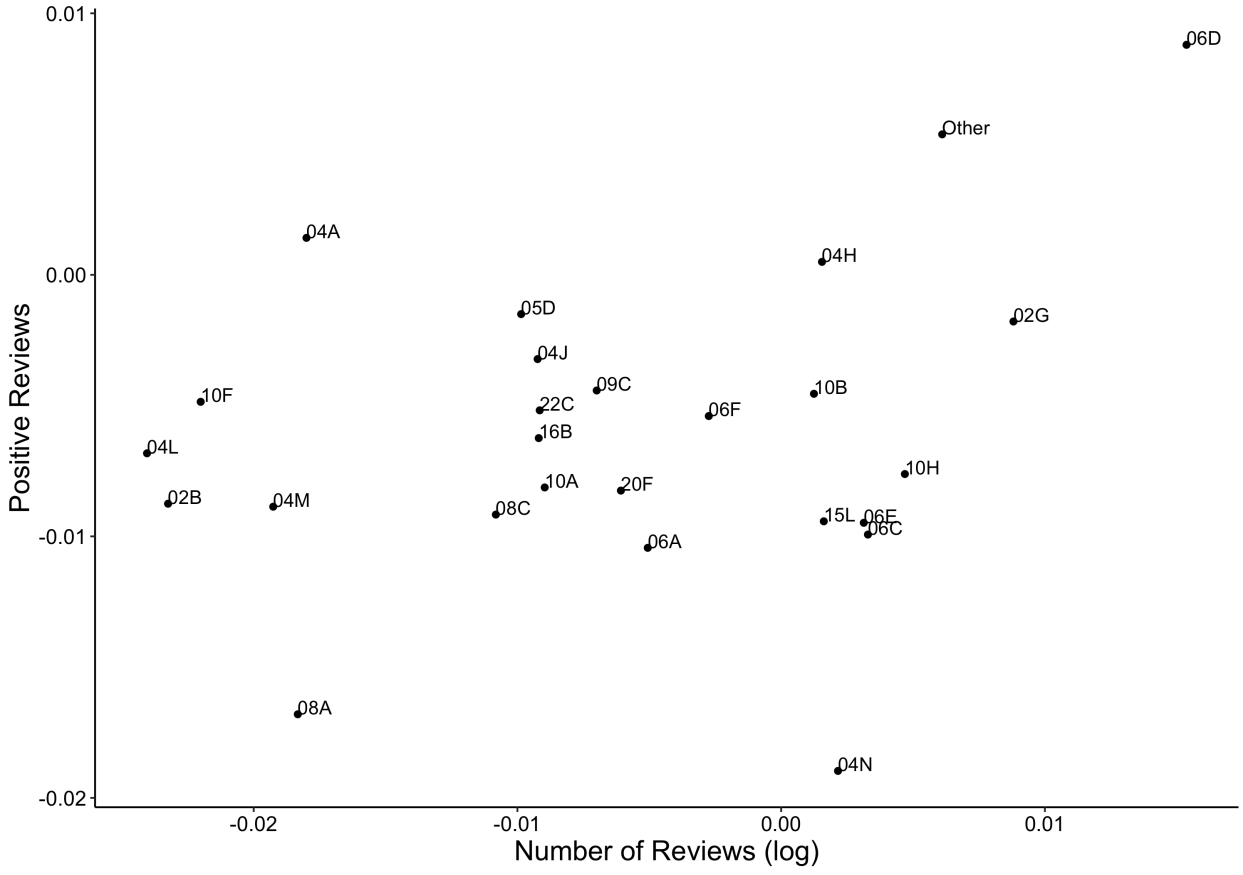


Figure 14: Diff-in-Diff Coefficient Estimates



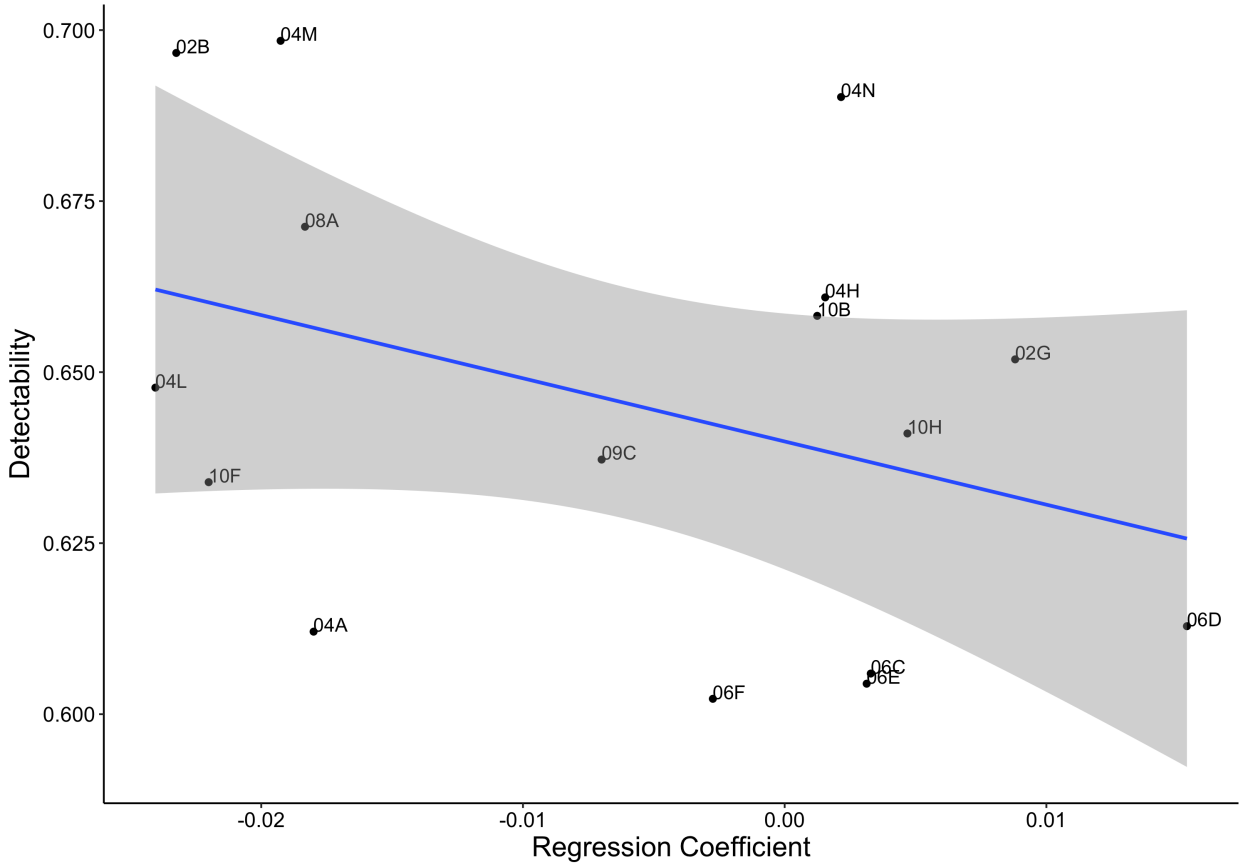
Coefficient estimates from equation 2. The left panel shows the propensity of Yelp restaurants to violate relative to restaurants without Yelp reviews in the last 90 days, γ_v. The right panel shows the propensity of Yelp restaurants to violate as the number of recent Yelp reviews increases, δ_v.

Figure 15: Diff-in-Diff Coefficient Estimates



Coefficient estimates from equation 2. The figure displays the coefficients estimates from Figure 14 as a scatterplot. On the x-axis we plot δ_v, on the y-axis we plot γ_v.

Figure 16: Propensity to violate and detectability on Yelp



The variable on the x-axis is the estimated δ_v coefficient from equation 2, one per violation. The variable on the y-axis is the R-squared from the prediction of violation v from restaurant characteristics and the text of Yelp reviews.

Tables

Table 1: Restaurant Characteristics

Characteristics	All	Yelp	OpenTable
Cuisine - American	20.6%	21.8%*	30.1%*
Cuisine - Chinese	10.2%	10.7%*	1.0%*
Cuisine - Cafe/Bakery	6.4%	7.1%*	0.5%*
Cuisine - Latin/Mexican	6.0%	5.9%	6.0%
Cuisine - Pizza	6.0%	7.3%*	2.0%*
Cuisine - Italian	3.1%	4.5%*	18.5%*
Borough - Manhattan	36.6%	44.0%*	83.5%*
Borough - Brooklyn	25.5%	25.3%	11.8%*
Borough - Queens	23.7%	20.3%*	2.9%*
Borough - Bronx	10.2%	6.5%*	1.1%*
Borough - Staten Island	3.9%	3.8%	0.8%*
Venue - Restaurant	47.5%	60.8%*	89.6%*
Venue - Fast Food	7.9%	8.8%*	0.2%*
Venue - Bar/Pub	4.7%	5.2%*	2.9%*
Share Chain Restaurants	10.2%	11.4%*	1.2%*
Share Closed	57.1%	42.8%*	22.2%*
Share Newly Opened	67.0%	67.2%	66.9%
N	57,045	31,218	2,534

$^{*}p < 0.01$

The table presents a summary of restaurant characteristics for the three samples: all restaurants inspected by the New York City Department of Health, restaurants with Yelp reviews, and restaurants on OpenTable.

Table 2: Effect of Yelp Reviews and Health Grades on Sold Out Probability.

	Sold Out at 7PM (OLS)			
	(1)	(2)	(3)	(4)
1.5 Yelp Stars	−0.194***	0.207	0.221	0.086*
	(0.013)	(0.143)	(0.146)	(0.052)
2 Yelp Stars	−0.178***	−0.228***	−0.247***	0.034
	(0.046)	(0.058)	(0.061)	(0.052)
2.5 Yelp Stars	−0.052	−0.087**	−0.096**	−0.023
	(0.036)	(0.039)	(0.041)	(0.017)
3 Yelp Stars	−0.088***	−0.077***	−0.087***	−0.018
	(0.019)	(0.019)	(0.021)	(0.011)
3.5 Yelp Stars	−0.031*	−0.031**	−0.042**	−0.006
	(0.016)	(0.015)	(0.017)	(0.008)
4 Yelp Stars	0.088***	0.094***	0.093***	0.023***
	(0.018)	(0.017)	(0.018)	(0.008)
4.5 Yelp Stars	0.176***	0.177***	0.185***	0.036***
	(0.039)	(0.035)	(0.034)	(0.011)
5 Yelp Stars	0.212***	0.298**	0.322**	0.097**
	(0.046)	(0.132)	(0.127)	(0.039)
Grade A	0.054***	0.043***	0.045***	0.008
	(0.015)	(0.015)	(0.015)	(0.006)
Grade B	0.010	0.015	0.016	0.007
	(0.015)	(0.015)	(0.015)	(0.007)
Grade Pending	0.028*	0.035**	0.037**	0.005
	(0.016)	(0.015)	(0.015)	(0.007)
Constant	0.188***			
	(0.019)			
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	1,765,291	1,765,291	1,765,291	1,765,291
R^2	0.025	0.088	0.102	0.439
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear probability model of the probability of being sold out on health outcomes and Yelp reviews. One observation is a restaurant-day. The outcome variable is equal to 1 if the restaurant has no availability on OpenTable between 6:30PM and 7:30PM. Controls include time-invariant restaurant characteristics such as cuisine and zipcode, and time controls such as day of the week and quarter fixed effects. Standard errors are clustered at the restaurant level.

Table 3: DMR topics.

Violation	Words
02B	vile, mice, driver, rush, minimum, front_desk, health, demand, incorrect, sick tourist_trap, roti, grade, counter, fusion, eel_avocado, hilton, absurd, overcook, ice
02G	rancid, applebe, rip_off, tourist_trap, mother, puke, horrend, redeem, diarrhea, wave straw, club, fall_apart, fishi, omelett, front_desk, hamburg, ticket, frozen, limp
04A	erica, promot, seafood_tower, lair, buzzer, sport, stair, samosa, smoke, threaten carrot_halwa, risotto, lentil, drunk, shoe, pot_pie, groupon, avocado, dough, smoke
04H	automat_gratuiti, trashi, douch_bag, aisl, lame, interrupt, atroci, egg_benedict, piss, attend pool_player, cafe_henri, bro, pitcher, pour, empti, suck, she, shell, stair
04J	porchetta, bouncer, turkey, honey_mustard, bland, bu_boy, overpr, garlic, raw, ined jaeger_schnitzel, rosti, chicharon_bulaklak, vindaloo, hash_brown, hudson, reuben, cod, bone, basket
04L	certif, 3d, ipad, voucher, medium_rare, ketchup, oil, music, spit, review lahmacun, flavorless, penn, frozen, chef, oliv_oil, nasti, hype, serv, entre
04M	abita, marti, fork, roach, mice, host, shove, turkey, tap, kitchen kunjip, omlett, bu_boy, onion_ring, bubbl_tea, yellowtail, eggplant, deep, frozen_margarita, smoke_salmon
04N	ghetto, toilet, voucher, straw, drunk, strike, hard_earn, disinterest, owner, tourist croissant, smoke, concept, raw, cart, pork_chop, orang_juic, pour, eventu, ice
05D	ticket_holder, brie, seafood_tower, doormen, coconut, bed_bug, deliveri, ticket, glass, appar il_bambino, chicharon_bulaklak, ticket_holder, crawfish, squid, shisha, stair, sesam, asian, skewer
06A	cl, paul, nan, bother, uncal, hill, baguett, kitchen, omelett, hung paella, bamboo_steamer, kimche, omelet, frost, nacho, soho, dough, salti, pour
06C	wife, grade, black_bean, silverwar, year, third, done, absurd, chef, threw indiffer, blech, stew, fat, tourist_trap, accept, mash_potato, either, potato, meh
06D	nightmar, inappropri, desk, tab, sarcast, shirt, remov, busboy, cash, shoe tourist_trap, bro, imposs, bummer, chipotl, smoke, chili_oil, bought, push, bone
06E	dandan_noodl, spa_castl, ticket, caraf, nacho, discount, atroci, bouncer, sit, upstair undersp, beef_randang, seva, pot_pie, cevich, oili, snapper, danc, ketchup, medium_rare
06F	pongsri, tourist_trap, guy, hair, black, argu, show, him, loung, waitress matsuhisa, bouncer, mac_n, raw, teriyaki, szechuan, mash_potato, michelin_star, music, sirloin
08A	mice, redeem, discount, downhil, voucher, social, gratuiti, grubhub, groupon, crappi reheat, toilet, gratuiti, runni, groupon, blah, forgett, hollandais_sauc, badli, anytim_soon
08C	dp, beard, dieci, coupon, groupon, cop, samosa, spinach, spanish, risotto croqueta_de, floyd, gra, do_camino, ihop, ghetto, rabbit, pancit, pork_chop, guac
09C	gil, snow_crab, maitr_d, iranian, forgotten, float, needless, entre, fee, parti pigal, matsuhisa, hostest, macaroni, schnitzel, mani_pedi, mash_potato, yucca, groupon, loud
10B	risotto, rush, ipad, bu_boy, remov, grade, oxtail, dick, scream, chef ketchup, tourist_trap, bleh, rigatoni, poach_egg, magnolia, hash_brown, ined, eggplant, dri
10F	certif, jacket, feet, coupon, fight, stolen, credit_card, tourist_trap, ticket, minimum ticket_holder, greas, asian, passabl, bonchon, ehh, veal, glass, bottl, medium
10H	insult, doorman, acknowledg, wear, credit_card, internet, honor, later, attend, brown porch_swing, sum, shabu, scallop, veal, sea_bass, fish, potenti, barista, plate
rating1	refund, unprofession, disgust, poison, worst, refus, zero, disrespect, insult, appal
rating2	meh, tasteless, below_averag, flavorless, downhil, unprofession, underwhelm, worst, disgust, unaccept
rating3	write_home, meh, alright, averag, eh, wow_factor, unmemor, fell_short, underwhelm, solid
rating4	highli_recommend, gem, hidden_gem, perfect, yum, solid, delici, fantast, excel, divin
rating5	highli_recommend, gem, hidden_gem, heaven, impecc, amaz, everi_penni, phenomen, perfect, fantast

Words with the highest loadings in 1- and 2-star reviews.

Table 4: Effect of Yelp Reviews and Health Grades on Sold Out Probability.

	Sold Out at 7PM (IV Second Stage)			
	(1)	(2)	(3)	(4)
Grade A	0.029	0.008	0.020	-0.014
	(0.025)	(0.024)	(0.025)	(0.009)
1.5 Yelp Stars	-0.048*	-0.050	-0.068	0.014
	(0.027)	(0.043)	(0.043)	(0.017)
2 Yelp Stars	0.164	0.143	0.137	0.007
	(0.228)	(0.216)	(0.216)	(0.018)
2.5 Yelp Stars	-0.081**	-0.079**	-0.076**	-0.009
	(0.035)	(0.033)	(0.033)	(0.016)
3 Yelp Stars	-0.136***	-0.105***	-0.102***	-0.007
	(0.027)	(0.024)	(0.024)	(0.012)
3.5 Yelp Stars	-0.108***	-0.084***	-0.082***	-0.001
	(0.026)	(0.023)	(0.023)	(0.011)
4 Yelp Stars	-0.021	0.008	0.015	0.011
	(0.025)	(0.022)	(0.022)	(0.011)
4.5 Yelp Stars	0.078**	0.102***	0.112***	0.015
	(0.036)	(0.033)	(0.034)	(0.012)
5 Yelp Stars	0.077	0.093	0.100	0.047
	(0.100)	(0.097)	(0.094)	(0.041)
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	1,635,033	1,635,033	1,635,033	1,635,033
R^2	0.027	0.079	0.091	0.412
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear probability model of the probability of being sold out on health outcomes and Yelp reviews. One observation is a restaurant-day. The outcome variable is equal to 1 if the restaurant has no availability on OpenTable between 6:30PM and 7:30PM. Same regression as in Table 3, except that the letter grade posted on the restaurant door – dummy for grade A – is instrumented with two variables that shift the inspection violation score, but are unlikely to be correlated with other quality characteristics observable by restaurant guests. The first instrument is whether the inspector of the current inspection was the same as the inspector of the previous restaurant inspection. The second instrument is an average of the inspector's violation scores assigned to restaurants other than the one being inspected. Controls include time-invariant restaurant characteristics such as cuisine and zipcode, and time controls such as day of the week and quarter fixed effects. Standard errors are clustered at the restaurant level. Depending on the specifications, we sometimes reject the null hypothesis of exogeneity of hygiene letter grades.

Table 5: Effect of Yelp Reviews and Health Grades on Sold Out Probability.

	Grade A (IV First Stage)			
	(1)	(2)	(3)	(4)
1.5 Yelp Stars	0.177***	0.210***	0.227***	0.672***
	(0.010)	(0.038)	(0.038)	(0.080)
2 Yelp Stars	0.039**	0.025	0.016	0.293**
	(0.020)	(0.018)	(0.021)	(0.117)
2.5 Yelp Stars	−0.081**	−0.086**	−0.087**	0.089*
	(0.036)	(0.038)	(0.038)	(0.048)
3 Yelp Stars	−0.017	−0.019	−0.025	0.076***
	(0.015)	(0.016)	(0.016)	(0.030)
3.5 Yelp Stars	0.002	0.007	−0.003	0.105***
	(0.012)	(0.012)	(0.012)	(0.022)
4 Yelp Stars	0.041***	0.051***	0.043***	0.116***
	(0.011)	(0.012)	(0.012)	(0.020)
4.5 Yelp Stars	0.072***	0.075***	0.072***	0.087***
	(0.015)	(0.016)	(0.016)	(0.023)
5 Yelp Stars	0.074	0.092	0.103	0.045
	(0.053)	(0.062)	(0.063)	(0.055)
Same Inspector	−0.006	−0.001	0.001	0.041*
	(0.022)	(0.021)	(0.021)	(0.025)
Avg. Inspector's Score	−0.017***	−0.017***	−0.016***	−0.016***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	1.218***			
	(0.020)			
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	1,635,033	1,635,033	1,635,033	1,635,033
R^2	0.056	0.090	0.097	0.340
Note:				*p<0.1; **p<0.05; ***p<0.01

First stage of Table 4. The smallest F-statistic of the first stage regressions is 166.

Table 6: Effect of Yelp Reviews and Health Grades on Restaurant Exit

	Out of Business within 90 Days (OLS)			
	(1)	(2)	(3)	(4)
1 Yelp Star	−0.008***	−0.003	−0.002	−0.005***
	(0.0005)	(0.002)	(0.002)	(0.001)
1.5 Yelp Stars	−0.003	0.003	0.003	0.005
	(0.006)	(0.006)	(0.006)	(0.011)
2 Yelp Stars	−0.0001	0.005	0.005	0.002
	(0.003)	(0.003)	(0.003)	(0.004)
2.5 Yelp Stars	0.0001	0.004**	0.003**	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)
3 Yelp Stars	−0.003***	0.001	0.001	−0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
3.5 Yelp Stars	−0.004***	−0.001	−0.001	−0.003***
	(0.0005)	(0.001)	(0.001)	(0.001)
4 Yelp Stars	−0.005***	−0.003***	−0.003***	−0.005***
	(0.0004)	(0.0005)	(0.001)	(0.001)
4.5 Yelp Stars	−0.006***	−0.005***	−0.005***	−0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
5 Yelp Stars	0.006	0.003	0.004	0.010
	(0.014)	(0.014)	(0.014)	(0.017)
Grade A	−0.010***	−0.009***	−0.010***	−0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Grade B	−0.007***	−0.007***	−0.007***	−0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Grade Pending	0.001	0.001	0.002	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Closed by DOH	−0.001	−0.002	−0.002	−0.004***
	(0.002)	(0.002)	(0.002)	(0.002)
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	330,447	330,447	330,447	330,447
R^2	0.004	0.011	0.012	0.408
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear probability model of the probability of being out of business. For all inspection types A and B, the outcome variable measures whether a restaurant closes within 90 days since the inspection. Yelp ratings are computed as cumulative average of all the reviews received up to the date of the inspection. In column (1) we only include inspection type fixed effects as controls. Column (2) includes inspection type fixed effects, as well as fixed effects for zipcode, cuisine type, service type, venue, and whether it's a chain restaurant. Column (3) includes all the controls in column (2) plus time controls, i.e. day of the week FE, year-quarter FE, restaurant age group FE. Column (4) has inspection type FE, restaurant FE, and time controls from column (3). Standard errors are clusterd at the restaurant level. Because the DoH can actually decide to close a restaurant during an inspection if hygiene violations are deemed too severe, we run the regressions both with and without those restaurants that went out of business immediately following a closure by the DoH. The table presents the results exluding those inspections, but the results are virtually identical.

Table 7: Effect of Yelp Reviews and Health Grades on Restaurant Exit

	Out of Business within 90 Days (IV)			
	(1)	(2)	(3)	(4)
Grade A	-0.003**	-0.001	-0.005***	0.001
	(0.001)	(0.001)	(0.002)	(0.002)
1.5 Yelp Stars	-0.009***	-0.003	-0.004	-0.005***
	(0.001)	(0.002)	(0.003)	(0.001)
2 Yelp Stars	-0.001	0.005	0.005	0.007
	(0.008)	(0.008)	(0.008)	(0.012)
2.5 Yelp Stars	-0.002	0.003	0.003	0.005
	(0.004)	(0.004)	(0.004)	(0.006)
3 Yelp Stars	0.001	0.004**	0.004**	0.003
	(0.002)	(0.002)	(0.002)	(0.003)
3.5 Yelp Stars	-0.004***	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
4 Yelp Stars	-0.004***	-0.001**	-0.001*	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
4.5 Yelp Stars	-0.006***	-0.004***	-0.003***	-0.004***
	(0.000)	(0.001)	(0.001)	(0.001)
5 Yelp Stars	-0.007***	-0.006***	-0.006***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.011***			
	(0.001)			
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	218,764	218,764	218,764	218,764
R^2	0.001	0.009	0.011	0.408
Note:				*p<0.1; **p<0.05; ***p<0.01

Linear probability model of the probability of being out of business. The specifications are the same as in Table 6, but the health grade is instrumented with inspector-specific characteristics: same inspector as in the previous inspection, and inspector's average violation points assigned to other restaurants. In all specifications we reject the null hypothesis of exogeneity of hygiene letter grades.

Table 8: Effect of Yelp Reviews and Health Grades on Restaurant Exit

	Grade A (IV First Stage)			
	(1)	(2)	(3)	(4)
1 Yelp Star	0.161*	0.023	−0.090	−0.034
	(0.091)	(0.097)	(0.103)	(0.152)
1.5 Yelp Stars	0.078	0.016	−0.067	−0.072*
	(0.060)	(0.049)	(0.049)	(0.043)
2 Yelp Stars	0.071***	0.070***	−0.019	−0.010
	(0.021)	(0.020)	(0.020)	(0.025)
2.5 Yelp Stars	0.031***	0.046***	−0.041***	0.006
	(0.012)	(0.011)	(0.011)	(0.012)
3 Yelp Stars	0.066***	0.091***	0.009	0.012*
	(0.006)	(0.006)	(0.006)	(0.007)
3.5 Yelp Stars	0.069***	0.101***	0.020***	0.021***
	(0.004)	(0.004)	(0.004)	(0.006)
4 Yelp Stars	0.108***	0.132***	0.052***	0.024***
	(0.004)	(0.004)	(0.004)	(0.006)
4.5 Yelp Stars	0.180***	0.185***	0.108***	0.041***
	(0.007)	(0.008)	(0.007)	(0.010)
5 Yelp Stars	0.283***	0.261***	0.192***	0.108**
	(0.027)	(0.029)	(0.035)	(0.043)
Same Inspector	0.028***	0.032***	−0.010*	0.011*
	(0.006)	(0.006)	(0.006)	(0.006)
Avg. Inspector's Score	−0.017***	−0.017***	−0.014***	−0.014***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Geo and Restaurant Controls	No	Yes	Yes	No
Time Controls	No	No	Yes	Yes
Camis Fixed Effects	No	No	No	Yes
Observations	330,447	330,447	330,447	330,447
R^2	0.094	0.132	0.196	0.402
Note:				*p<0.1; **p<0.05; ***p<0.01

First stage of the IV regressions in Table 7. The smallest F-statistic of these first stages is 2,321.

Table 9: Restaurant visibility on Yelp and hygiene performance

	(log) Inspection Score		
	(1)	(2)	(3)
On Yelp	0.069***	0.001	−0.044***
	(0.007)	(0.008)	(0.013)
Has Recent Yelp Reviews	0.019**	0.022**	−0.020
	(0.010)	(0.009)	(0.012)
(log) Number of Recent Yelp Reviews	−0.007*	−0.017***	0.025***
	(0.004)	(0.004)	(0.008)
(log) Restaurant Age		0.023***	0.001
		(0.001)	(0.002)
Constant	2.431***		
	(0.004)		
Geo and Restaurant Controls	No	Yes	No
Restaurant FE	No	No	Yes
Observations	244,121	244,117	244,117
R^2	0.001	0.028	0.180
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 10: Effect of restaurant visibility on Yelp on restaurant incentives to improve hygiene performance

	Violation Found
Inspection Score (log)	0.109^{***}
	(0.0004)
Restaurant Age (log)	0.001^{***}
	(0.0001)
On Yelp	-0.010^{***}
	(0.001)
Has Recent Yelp Reviews	0.0003
	(0.001)
(log) Number of Recent Yelp Reviews	0.004^{***}
	(0.001)
On Yelp * Detectable Violation	0.013^{***}
	(0.002)
Has Recent Yelp Reviews * Detectable	-0.004^{*}
	(0.002)
(log) Number of Recent Yelp Reviews * Detectable	-0.006^{***}
	(0.001)
Violation Code FE	Yes
Restaurant FE	Yes
Observations	3,326,192
R^2	0.140
Note:	*p<0.1; **p<0.05; ***p<0.01

Coefficient estimates of equation 2. Here, instead of computing one set of (β, γ, δ) per violation, we compute $(\beta_d, \gamma_d, \delta_d)$ for the 8 most detectable violations and $(\beta_{nd}, \gamma_{nd}, \delta_{nd})$ for the 8 least detectable violations, as measured by our text analysis.