

ABSTRACT

Title of Document: *VOX POPULI: THREE ESSAYS ON THE USE OF SOCIAL MEDIA FOR VALUE CREATION IN SERVICES*

Jorge Mejia
Ph.D. Candidate of Information Systems, 2016
Department of Operations & Information Systems
Robert H. Smith School of Business
University of Maryland, College Park

Directed By: Anandasivam Gopal
Associate Professor of Information Systems
Department of Operations & Information Systems
Robert H. Smith School of Business
University of Maryland, College Park

Prior research shows that electronic word of mouth (eWOM) wields considerable influence over consumer behavior. However, as the volume and variety of eWOM grows, firms are faced with challenges in analyzing and responding to this information. In this dissertation, I argue that to meet the new challenges and opportunities posed by the expansion of eWOM and to more accurately measure its impacts on firms and consumers, we need to revisit our methodologies for extracting insights from eWOM. This dissertation consists of three essays that further our understanding of the value of social media analytics, especially with respect to eWOM. In the first essay, I use machine learning techniques to extract semantic structure from online reviews. These semantic dimensions describe the experiences of consumers in the service industry more accurately than traditional numerical variables. To demonstrate the value of these dimensions, I show that they can be used

to substantially improve the accuracy of econometric models of firm survival. In the second essay, I explore the effects on eWOM of online deals, such as those offered by Groupon, the value of which to both consumers and merchants is controversial. Through a combination of Bayesian econometric models and controlled lab experiments, I examine the conditions under which online deals affect online reviews and provide strategies to mitigate the potential negative eWOM effects resulting from online deals. In the third essay, I focus on how eWOM can be incorporated into efforts to reduce foodborne illness, a major public health concern. I demonstrate how machine learning techniques can be used to monitor hygiene in restaurants through crowd-sourced online reviews. I am able to identify instances of moral hazard within the hygiene inspection scheme used in New York City by leveraging a dictionary specifically crafted for this purpose. To the extent that online reviews provide some visibility into the hygiene practices of restaurants, I show how losses from information asymmetry may be partially mitigated in this context. Taken together, this dissertation contributes by revisiting and refining the use of eWOM in the service sector through a combination of machine learning and econometric methodologies.

VOX POPULI: THREE ESSAYS ON THE USE OF SOCIAL MEDIA FOR VALUE
CREATION IN SERVICES

By

Jorge Mejia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:

Professor Anandasivam Gopal, Chair

Professor Ritu Agarwal

Professor Shawn Mankad

Professor Michael Trusov

Professor Jennifer Golbeck, Dean's representative

© Copyright by
Jorge Mejia
2016

Dedication

To her, who stayed by my side since that summer we met. Back then, when we were young, I was impressed. Now that we are bit older, I am more impressed and fortunate you are still by my side.

Acknowledgements

I often think of buying presents for Anand, but the debt is so large that I never find a gift big enough, impressive enough, refined enough, to match my gratitude. I resolved long ago to instead enjoy his company and listen, yes listen.

I've learned so much from Shawn that I often feel like I've sat in his classes. I am thankful for his patience not only showing me how to do stuff, but in helping me organize my thoughts and ideas. He's been a great research partner and friend.

In my second semester in the program, I took Ritu's seminar class, and she asked me one day if I wanted to be a "star". After seeing her work, dedication, and brilliant feedback, I now know what it takes. She's been an inspiration to me, and I hope she continues to be my mentor and friend.

To Michael and Jen, I want to say thank you for being part of this and supporting my work. Your feedback and mentoring has made my work much better.

To the other members of the faculty, thank you for your support and constructive feedback. I have never knocked on a door that has not opened at Maryland.

To my family and friends, thank you for your support and patience. To my PhD program buddies, especially Isaac, Brad, Keongtae, Bryan, Alena, Shannon, Seoungwoo and JPP, thank you for keeping it fun.

To Justina and the Ph.D. Program Office, thank you for your encouragement and guidance. The road was not always clear, but you always seemed to have a flashlight.

I thought this was an end, and I know realize it is a beginning.

Table of Content

Dedication.....	ii
Acknowledgements.....	iii
Table of Content	iv
List of Tables	vi
List of Figures.....	vii
Chapter 1 Introduction	1
Chapter 2 More Than Just Words: Service Quality Dimensions in Online Reviews and Firm Survival	8
Introduction.....	8
Background and Theory.....	14
Measuring Service Quality in Service Operations.....	14
Online Reviews.....	18
Service Quality and Survival	20
Methodology.....	24
Research Context.....	24
Coarsened Exact Matching (CEM).....	25
Extracting Service Themes from Text.....	29
Econometric Models and Estimation.....	32
Longitudinal Generalized Linear Models.....	34
Robustness Check with the Cox Proportional Hazards Model.....	39
Discussion and Implications	41
Appendix A: Text Analysis Methodology.....	46
A.1. Latent Semantic Analysis (LSA).....	47
A.2. LSA and Other Topic Modeling Methods	52
Appendix B: Post Hoc Analysis: Predicting Closure	54
Chapter 3 Deal or No Deal? The Quality Implications of Online Daily Deals	56
Introduction.....	56
Background and Theory.....	62
Research Methodology	70
Data.....	70
Empirical Model	73
Model Estimation and Results	75
Discussion.....	77
Lab Experiments: Testing the Pre-Consumption Effects of Daily Deals	79
Experiment 1: Deals and Merchant Price	79
Experiment 2: Deals and Merchant Age.....	82
Experiment 3: Deals and Deal Competition	84
Discussion and Managerial Implications.....	86
Chapter 4 Watch where you eat: Restaurant Hygiene Inspections in New York City and Moral Hazard	94
Introduction.....	94
Context: The NYC Restaurant Inspection Program.....	99
The Data: Inspections and Online Reviews for New York City.....	102
Analysis and Results.....	109

Creating a Social Media Sourced Hygiene Dictionary	109
Building the List of Seed Words.....	111
Obtaining Training Data and Implementation	113
Validating SMASH.....	116
Using SMASH to Show the Effect of Moral Hazard on Restaurant Hygiene	118
Conclusion	120
Tables	123
Figures.....	135
Appendix.....	144
Bibliography	146

List of Tables

Table 2.1. Matching Variables.....	123
Table 2.2. Imbalance Comparison Table	123
Table 2.3. Comparing Case and Control Samples	123
Table 2.4. Top Words Associated with Each Semantic Component	123
Table 2.5. Reviews Associated with the Identified Five Semantic Components	124
Table 2.6. Variable Descriptions	125
Table 2.7. Summary Statistics	126
Table 2.8. Correlation Matrix	126
Table 2.9. GLMER Coefficient Estimates.....	127
Table 2.10. Cox Proportional Hazards Results (Positive Values Contribute to Closing)	128
Table 2.11. Proportionality Test To Verify Proportionality Assumption.....	128
Table 3.1. Variable Descriptions	129
Table 3.2. Summary Statistics	129
Table 3.3. Correlation Table	130
Table 3.4. Multi-level Hierarchical Bayesian Results	130
Table 4.1 Variable Descriptions	131
Table 4.2. Summary Statistics for the Difference-in-Differences Model.....	132
Table 4.3. Difference-in-Differences Results	132
Table 4.4. Linear Mixed Effects Model Results. DV=Inspection Scores.....	133
Table 4.5. Linear Mixed Effects Model Results. DV=SMASH WC Scores 90 Days After Inspection	133

List of Figures

Figure 2.1. These plots depict the Schoenfeld residuals by the mean rating, the number of reviews, and the semantic variables Quality_Overall and Food_Efficiency. A non-zero slope is evidence against proportionality.....	135
Figure 2.2. Cross-Validation and Singular Values for LSA	135
Figure 2.3. ROC Curves for GLMER models	136
Figure 3.1. Daily Deals Effects on eWOM: Timeline	136
Figure 3.2. Mean Yelp Rating by Offset from Daily Deal Offer Date	137
Figure 3.3. Lab Studies Stimuli Examples.....	138
Figure 3.4. Brand Evaluation as a Function of Deal Offered and the Price of Merchants.....	139
Figure 4.1. NYC Restaurant Hygiene Cards.....	140
Figure 4.2. Placement of Grade Card in a NYC Restaurant	140
Figure 4.3. Grades over time (Before and After the NYC Grading Program).....	141
Figure 4.4. Restaurant with A Grades by Borough in 2011, 2013, and 2015 in NYC	141
Figure 4.5. Restaurant with A Grades by Borough from 2011 to 2015.....	142
Figure 4.6. Score Trajectories from Restaurants that Consistently Score A in their Initial Inspections Versus those that Post a Grade P in their Initial Inspection	142
Figure 4.7. Average Scores of the Two Main Inspection Trajectories	143
Figure 4.8. Trends in SMASH Score After Different Types of Inspections for PAPA	143

Chapter 1 Introduction

In recent years, there has been a rapid increase in the volume and reach of electronic word of mouth (eWOM) available to consumers. With the rise in connectivity through the Internet and web 2.0 tools focused on sharing, platforms aggregating eWOM, such as online reviews sites and review dashboard providers, have become mainstream. For example, Yelp receives over 142 unique million visitors every quarter. Similarly, Trip Advisor has aggregated more than 225 million traveler reviews, while Amazon sells over 150 million products with customer reviews. Aggregators like ReviewTrackers and Podium.co provide dashboards that allow firms to quickly respond to their online reviews in one location, rather than tracking individual eWOM sites. The popularity of these sites and business models suggests that consumers have evolved into content creators and even critics. While most online reviews platforms focus on traditional categories like restaurants and hotels, there has also been an expansion to include other categories, such as the critical choice of a physician or dentist, for which 35% of consumers say they have searched reviews (Keckley and Coughlin 2012).

Online reviews also wield significant influence on consumer choice. For example, a 2014 consumer report estimates that almost 90% of consumers have looked at reviews in the last 12 months to help them make a decision on a local business. Further, 40% of consumers read reviews on a regular basis (Bright Local 2014). Apart from being an important tool in consumers' decision-making choices, these reviews also have significant influence in shaping opinions about businesses. Recent consumer surveys estimate that 72% of consumers say that positive reviews

increase trust in a business or product, while over 80% of consumers trust online reviews as much as personal recommendations (Bright Local 2014; Deloitte 2014). Moreover, the popularity and influence of these reviews to consumers and firms has given rise to a new breed of business models. For example, we now have companies dedicated solely to managing what consumers are saying online about products and services, such as Reputation.com, CzarMetrics, and Luminoso.

The rise of this consolidated consumer knowledge, through improved access to information and expanded opportunities to share experiences, poses a significant challenge to businesses as they attempt to fulfill the expectations of better-informed consumers. Two concurrent movements in the technology space make this a compelling question for firms. First, the volume and variety of online reviews available to the consumer from multiple platforms has expanded significantly in the last decade, often leading firms and managers to feel overwhelmed by the quantity of the discourse available (cite). Needless to say, manually understanding and responding to individual queries and feedback on such sites is simply not feasible any more, leading to a significant gap in the literature on how existing research in online reviews may be extended to a world where volume has exploded. Second, a lot of the earlier work focused on numerical data such as review valence and review volume – both of which captured economic significance in terms of quality and demand (cite). However, much of the current content in eWOM is unstructured and available in the form of free-form text, images, and sound. Analyses of review volume and valence may be overlooking insight that may be generated from these other sources of data.

In this dissertation, I focus attention on extending the considerable prior literature in online reviews and eWOM to contexts where the volume of discourse as well as the text available in the discourse is appropriately utilized. As businesses learn to react to and harness this growing power of consumers through eWOM platforms, concurrent research is needed to develop newer models, based on more modern text-analytic and experimental methodologies, that inform managers as well as researchers on how to generate actionable learning from eWOM. I argue that to meet the new challenges and opportunities posed by the expansion in reach, relevance and volume of eWOM, we need to revisit our methodologies and models. In this doctoral dissertation, I include three essays on social media analytics, each of which build on seminal work on online reviews and firm performance by combining econometric models with lab experiments and text analysis to generate insights from large corpuses of online reviews.

In the first essay of this proposal (Chapter 2), I describe how to extract the semantic structure behind the text in online reviews to predict firm performance outcomes. Given the importance and relevance of eWOM, it is not surprising that management scholars, especially in information systems and marketing, are interested in understanding the relationship between eWOM and firm performance outcomes. However, a salient gap in this literature is being able to extract meaning from the collective corpus of opinions captured in the text of online reviews (Archak et al. 2011; Cao et al. 2011). The first essay aims to fill this gap through the application of existing machine learning techniques to analyze a large corpus of text in online reviews. Our aim is to discover the *themes* behind online reviews in a single service

context using an automated, data-driven approach that can provide benefits to platforms, merchants, and consumers. Furthermore, and as proof of concept to the value of our proposed semantic themes, I attempt to predict a critical firm performance outcome, firm survival, using longitudinal econometric models. The models provided in the essay show that prediction based on the semantic themes provides a significant improvement beyond survival models that are only predicated on numerical scores that have been used thus far in the literature, speaking to the value of the semantic themes.

In the second essay (Chapter 3), I explore the effect on eWOM of online deals, one of the most popular and controversial tools in the current marketing mix. These deals are one of the most popular online marketing tools. Groupon, for example, has over 200 million active subscribers and receives over 160 million unique monthly visitors. Despite the apparent popularity behind these daily deals platforms, however, there are many diverging opinions about their value to both consumers and merchants. In fact, multiple surveys find that only approximately half of merchants make a profit from running such deals (Dholakia 2010; BusinessInsider 2011).

Moreover, the long-term effect of deals on the reputations of merchants has been questioned. Early findings in the computer science literature suggest that the effect of Groupon promotions on online reviews is strictly negative (Byers et al. 2012a; 2012b). However, I question the generalizability of these results and build on existing marketing theories to understand the mechanisms behind the effect of online deals on online reviews. Indeed, previous theories suggest that the direction of the

effect of marketing efforts (i.e. negative vs. positive) on consumer perceptions can depend on a number of factors. For example, in some cases consumers may respond positively to such promotions because they perceive the deal as a sign of high confidence from the merchant, while in other contexts such promotions may smack of desperation and hence elicit a negative response from consumers (Kirmani and Wright 1989). I build on this theoretical work to understand the conditions under which online deals affect a merchant's eWOM. Moreover, I take into account the merchant's competitive landscape, which typically includes other merchants and their competitive actions. Thus, I am able to identify the effect on eWOM for a merchant when its competitors offer online deals. I provide managerially relevant implications and possible strategies for mitigating the potential negative eWOM effects resulting from online deals.

In the third essay (Chapter 4), I expand the application of social media analytics beyond the typical contexts of marketing and sales. I focus on how social media can be used in the efforts to decrease foodborne illness, which is a significant public health concerns. There has been a recent public policy push in many cities in the U.S., such as Seattle and San Francisco, to establish a framework to publicize the results of restaurant health inspections to consumers. The motivations for implementing such programs are straightforward. For example, a study of Seattle restaurants showed that restaurants with improper food hygiene practices had 16 times greater risk of a foodborne outbreak (Irvin et al. 1989). In this essay, I follow the implementation of a restaurant letter-grading program by the New York City Department of Health and Mental Hygiene, one of the most ambitious public health

initiatives of Mayor Bloomberg's administration. One of the main goals of this program is to provide the public with information about inspection results and, in turn, provide restaurants with an incentive to follow best food safety practices. Beginning in July 2010, restaurants were required to prominently display a letter grade (A, B, or C) in their establishments. Using recent advances in machine learning to process the text in online reviews for restaurants in New York, I create and validate and social media hygiene dictionary to measure the hygiene of restaurants over time. I then show, using this proposed methodology, moral hazard in the manner in which restaurants achieve high scores in the program. I find that many restaurants use the design of the restaurant grading program in New York to their advantage and in fact perform at lower levels of hygiene than they are able to. I provide strategies and insights for policy makers to better design incentives, penalize underperforming restaurants, and better protect consumers from foodborne illness.

In summary, the three essays contribute to the literature in information systems in two specific ways. First, I revisit the question of how crowd-sourced content, such as eWOM and online reviews, are linked through the firm's demand side to questions of firm performance. While prior work has focused on performance outcomes associated with consumer choice and willingness to pay, I respond to calls for broadening the footprint of social media research beyond choice by considering outcomes that are of considerable importance, such as firm survival and hygiene ratings. Second, where much of the literature has focused on the numerical scores that are typically attached to eWOM, developing systematic models for combining the text-based information in addition to the numerical data has remained understudied in

the IS literature. Through the use of lab experiments, semantic analysis and the creation of a context-specific dictionary, I propose new models of performance that incorporate text as well as numerical information. It is my belief that given contemporaneous progress in research in machine learning as well as eWOM in information systems, such hybrid approaches to understanding the impact of eWOM on businesses will become mainstream. The essays in my dissertation contribute to this growing trend. In the next chapter, I describe the first essay in detail.

Chapter 2 More Than Just Words: Service Quality Dimensions in Online Reviews and Firm Survival

Introduction

A recurring theme within service operations research and practice has been the need to establish the relationship between elements of service quality with overall performance of the service provider (Chase and Apte 2007). Elements of service quality that are attuned to providing customers with the ideal customer experience should ideally lead to customer loyalty, increased revenues, positive word of mouth, and even stock prices (Zeithaml et al. 1996, Ramdas et al. 2013). Within the research domain, this relationship has been viewed through the lens of the service profit-chain (Heskett et al. 1994) or the capabilities-service quality-performance (C-SQ-P) triad (Roth and Jackson 1995), frameworks that provide a holistic view of the service provider model. For service operations professionals, this research has implications for service design (Goldstein et al. 2002), i.e. the specific combination of processes, people skills, technologies, and materials that are integrated to provide the “planned” service. If services are designed well and executed to meet customer expectations, customer loyalty and profitability should follow (Sasser et al. 1997, Roth et al. 1997). Thus, a vital part of this research stream has been to specify models that associate specific attributes of service design and delivery to measures of operational and financial performance for the service provider (Chase and Apte 2007, Soteriou and Zenios 1999).

While intellectually appealing, empirically establishing the relationship between service quality or customer feedback and financial outcomes is not without challenges, a significant one being finding appropriate measures of service quality and customer feedback that are scalable and easily available to the service provider (Voss et al. 2008, Rosenzweig et al. 2011). To capture service quality, scholars have used surveys such as the ServQual instrument (Roth and Jackson 1995, Parasuraman et al. 1988) as well as indirect indicators of quality, such as wait times and service errors (Soteriou and Zenios 1999). However, these approaches are potentially limited as they are time-consuming for the customer (Cronin and Taylor 1992), hard to generalize across providers (Roth and Jackson 1995), and not fully scalable (Roth and Menor 2003). Furthermore, as the service sector becomes “experience-oriented” (Pine and Gilmore 1999), wherein customer engagement and enthusiasm are more important than simply providing good quality service (Voss et al. 2008, Pullman and Gross 2004), alternative sources of information on customer feedback and service quality are needed so that better information on the quality of service design can be extracted. Indeed, (Pullman and Gross 2004, p.553) write that “successful experiences are those that the customer finds unique, memorable and sustainable over time, would want to repeat and build upon, and enthusiastically promote via word of mouth.” Evidence of such engaging service encounters is unlikely to appear in large-scale surveys aimed at the average consumer. This information is potentially available through one source that allows individual consumers to provide detailed feedback on their experiences, while also becoming increasingly ubiquitous and influential in the context of services - online reviews.

Online reviews have long been a topic of considerable interest and promise, starting with the early days of electronic commerce (Dellarocas 2003). Moving beyond simply influencing online sales (Chevalier and Mayzlin 2006), positive online reviews, representing word of mouth, have also been associated with a host of service contexts such as offline business outcomes such as offline retail sales (Duan et al. 2008), hotel services (Ye et al. 2009), and even medical care (Gao et al. 2012). When viewed as a data source, online reviews typically contain multiple pieces of information that can be used as indicative of service quality and customer engagement. At an aggregate level, overall service quality can be gauged by the star rating, typically on a 1-5 scale. Researchers have also used the number of reviews provided in a given time period to measure traffic associated with the service provider (Dellarocas and Narayan 2006). However, if experience-based services require a deeper measure of customer engagement, i.e. “emotionally engaged customers” (Voss et al. 2008, p.247), these numerical measures, though representing an aggregate positive or negative sentiment towards the service provider through the star rating and length of the review, are not enough. The text of the online reviews represent a vital and viable source of such information (Cao et al. 2011) that convey the benefits of ubiquity, scale, and relevance for the service provider. In this paper, we thus focus on extracting value from the text present in online reviews that go beyond the overall sentiment already captured in the various numerical data available.

While early work extracting elements of text from online reviews has focused on positive or negative sentiment, word counts and readability scores (Ghose and

Ipeirotis 2011), less research has considered extracting semantic meaning from the actual text of the reviews themselves, even though they represent a collective corpus of opinions, judgments, evaluations and suggestions of customers (Archak et al. 2011, Cao et al. 2011). As a consequence, business outcomes that are part of standard service operations models (such as the service-profit-chain or the service quality-performance relationship) are likely to be better explained by augmenting the oft-used ratings and review volume variables with information contained within the text. We use text-mining techniques to extract semantic information representing service design themes (Goldstein et al. 2002) from the text of online reviews on service providers. Subsequently we use these semantic factors, in addition to numerical review data, to explain an economic outcome fundamentally related to the health of the service provider – business survival (Rosenzweig et al. 2011). If profitability is linked inextricably to customer engagement and excitement, the lack of this engagement reflected in online reviews should indicate impending business failure over time. We thus show that the semantic themes extracted from the review text provide greater explanatory power of the survival of service providers, relative to baseline models that only include numerical review data in addition to other contextual variables of importance (Parsa et al. 2011). This approach builds on prior work linking service quality dimensions to performance, with two important differences. First, we allow the review text to generate themes rather than specify them, a priori, and second, we provide a much deeper and systematic measure of consumer feedback than would be possible through survey research or individual customer queries.

The specific context we model pertains to online reviews for restaurants in a large metropolitan area in the United States. The restaurant industry is one of the definitive experience-based service industries in the U.S. (Zeithaml et al. 1993). We collect a comprehensive set of online reviews, including text and numerical data, of restaurants within the Washington D.C. metropolitan area over a period of nine years from one popular review platform. The dataset includes reviews on all restaurants that were operating in the area at any point during this nine-year time period, including those that closed during the observation window. In total, we have access to over 130,000 reviews with 50,000 pages of review text associated with over 2,400 restaurants. We first identify restaurants that closed during specific time periods and then match these restaurants to those that are identical or similar but have remained open, thereby creating a case-control set. Second, we use the full corpus of restaurant reviews to identify the semantic dimensions that characterize consumer experience. Third, we estimate a series of econometric specifications to explain restaurant closure through both numerical and semantic measures.

Our semantic analysis identifies five specific components within review text that allow a deeper reflection of a restaurant's service themes (Chase and Apte 2007). These components reflect distinct aspects of a restaurant's service offering, and the viability of the business, such as overall quality, food quality, wait times and ambience. Econometric models that incorporate the semantic themes provide significantly greater explanatory power than models where the themes are omitted. In addition, we show that each semantic theme likely provides a different measure of the quality of the restaurant's service design (Goldstein et al. 2002) beyond the

quantitative information provided by ratings and number of reviews. The semantic themes we identify are also better able to capture the experience aspect of services, providing a fuller reckoning of restaurant quality not captured by ratings and review volume. They thus provide a better way to capture unobservable quality and are likely to be better predictors of financial or business outcomes.

Our work contributes to existing work in service operations in several ways. Though considerable work in service operations has related aspects of service design to firm performance, little work has explicitly looked at the text provided by customers on their experiences, through online reviews. If the next generation of service innovations involve “experiences” (Pine and Gilmore 1999, Heineke and Davis 2007), unstructured and non-traditional sources of feedback, like those provided through online reviews and social media, become increasingly relevant. However, these sources provide large volumes of text, beyond the ability of individuals (such as a restaurateur) to absorb. Our work thus extends models of the value of service design (Voss et al. 2008, Rosenzweig et al. 2011, Ramdas et al. 2013) by using text analysis techniques to identify service themes. We validate these themes by estimating models of survival, thereby adding to the the critical service quality-performance link in service operations (Heineke and Davis 2007).

Second, we add to the literature on online reviews by augmenting models of electronic word of mouth (eWOM) that primarily use rating and review volumes as independent variables. In the spirit of Archak et al. (2011) and Cao et al. (2011), we provide a more comprehensive model of the effects of eWOM in the service sector. As eWOM becomes increasingly influential in retail and online settings, the large

amounts of text available can be tapped to streamline service design; our work here provides one such methodology for doing so. Finally, our work provides an analytical tool that can be used by service providers, potential investors, as well as platforms like Yelp strategically. Restaurants can analyze their own reviews using our model and identify processes that need improvement. In addition, firms like Yelp.com and OpenTable.com can use these dimensions to provide specific guidance to both consumers and businesses. Our approach here can also be adapted to other service contexts, such as hotels and retailers, easily as long as reasonably large datasets of text-based reviews are available. While the semantic themes in other contexts will no doubt change, the underlying approach is generalizable to other service contexts and economic outcomes as well.

Background and Theory

Measuring Service Quality in Service Operations

Understanding the strategic importance of service quality, from consumer feedback, has been a major part of the research agenda in service operations (Heineke and Davis 2007). Beginning with numerous case and correlational studies highlighting the role of service or product quality in organizations (Schoeffler et al. 1974, Juran and Gryna 1980, Hart et al. 1989, Schlesinger and Heskett 1991, Schneider and Bowen 2010), scholars have sought to link quality improvements within service design to firm performance. (Buzzell and Gale 1987, pg.7), for example, assert: “in the long run, the most important single factor affecting a business unit’s performance is the quality of its products and services relative to those of

competitors.”

Beyond anecdotal cases, scholars have also proposed several theoretical frameworks through which to analyze the service quality-performance relationship. Rust et al. (1994, 1995), for example, propose treating service or product quality as an investment on the part of the service provider and appropriately define “return on quality” (ROQ) to postulate the mechanisms by which quality improvements yield positive financial returns. They argue that service quality investments should increase financial success by providing cost reductions, increased customer retention and attracting new customers. Alternatively, the service-profit chain model (Heskett et al. 1994, Sasser et al. 1997) argues that service quality and customer satisfaction lead to customer loyalty, which in turn leads to improved financial returns for the firm, thereby providing the service provider with a strong incentive to focus on service design. The service-profit chain has also received some empirical validation through empirically showing the relationship between service quality, customer satisfaction and financial outcomes (Sasser et al. 1997, Roth et al. 1997, Soteriou and Zenios 1999). A third approach to modeling the performance implications of service quality comes from (Roth and Jackson 1995) who propose the operational capabilities, service quality, and performance triad (C-SQ-P), arguing that service firms capable of delivering superior quality do so because they employ a generic set of capabilities better than competitors. They validate the model empirically using data from the banking sector.

While this research stream has empirically established the link between service quality and firm outcomes within multiple service settings, a significant

challenge has remained creating operational methods within the firm to measure quality and customer satisfaction on an ongoing basis (Rust et al. 1995). To this end, scholars have used a variety of methods to measure service quality or customer satisfaction over the years. Early research applied manufacturing-based metrics of quality to the service context, using numerical indicators of quality that reflected a production process approach (Wyckoff 1984). For example, Reichheld and Sasser (1990) proposed service firms count service ‘defects’, Hart (1988) counted dissatisfied customers communication, and Soteriou and Zenios (1999) used customer wait times and “service errors”.

However, as services became more intangible, such manufacturing-based approaches became less directly relevant. The definition of service quality changed from defects to the extent to which customer expectations are met through the service encounter, thereby providing a superior consumer experience (Parasuraman et al. 1985, Gronroos 1990, Reeves and Bednar 1994). Zeithaml et al. (1990), for example, note “only customers can judge quality; all other judgments are essentially irrelevant.” To measure whether a customer’s needs and expectations have been met, scholars proposed a variety of survey instruments administered to various customers to gather specific feedback. A notable methodology used here was SERVQUAL (Parasuraman et al. 1985, 1993), containing 22 questionnaire items, attempting to measure the gap between expectations and perceptions at the time of service. Variations on SERVQUAL include SERVPERF (Cronin and Taylor 1994) and the more recent HEDPERF (Abdullah 2006).

Despite finding wide adoption by scholars and practitioners (Metters and

Marucheck 2007), researchers have questioned the applicability of such survey instruments (and other similar approaches to collecting consumer feedback) to different service contexts and industries (Roth and Jackson 1995, Cronin and Taylor 1992, Carman 1990). Creating such surveys, with their requisite statistical properties, is particularly difficult in the service industry, where individual customers in varying service types (banks versus restaurants, for instance) place different weights on various service attributes (Carman 1990, Garvin 1988). Moreover, surveys are time-consuming for customers (Cronin and Taylor 1992); provide limited validity across similar providers, making benchmarking difficult (Roth and Jackson 1995); do not capture individual heterogeneity (Garvin 1988); and are hard to scale (Roth and Menor 2003). Thus, defining operational measures of service quality remains a challenge (Metters and Marucheck 2007).

The limitations of survey-based or production-based measures of customer engagement or service quality are exacerbated when viewed through the changes in the service industry, where customer experiences are increasingly becoming central (Pine and Gilmore 1999).¹ Service providers need to find sources of consumer engagement that accurately describe and capture the experiences of consumers, but without the effort and intrusive nature of surveys (Voss et al. 2008). To the extent that customers are “emotionally engaged” with the brand or the service, they are likely to display loyalty as well as act as ambassadors for the service provider (Voss et al. 2008). Thus, online reviews, i.e., electronic word of mouth, have emerged as a

¹ Such a shift is observable even in practice within the retail sector:
<http://www.nytimes.com/2015/08/14/business/economy/stores-suffer-from-a-shift-of-behavior-in-buyers.html>

viable source of consumer information that could be used to gather feedback on consumer engagement as well as measure emotional engagement with the services offered. We review the literature in online reviews next.

Online Reviews

Prior to the Internet and the extensive use of electronic networks, word of mouth has been influential in affecting the diffusion of innovations and shaping consumer attitudes and choices (Katz and Lazarsfeld 1955, Buttle 1998). With the rise of the Internet, and the extensive deployment of eWOM applications, consumers have become accustomed to using online reviews to make decisions, with over 60% of US consumers reporting high or medium level of influence from online reviews in their purchasing decisions (Openshaw et al. 2014). Correspondingly, the process by which online reviews are generated has also been of interest to service providers and researchers alike (Mudambi and Schuff 2010, Forman et al. 2008). Indeed, providing online reviews and interacting with other consumers online has been included in a set of critical behaviors representing customer engagement in services research (Van Doorn et al. 2010). Firms thus have a strong incentive to encourage their consumers to provide online reviews of their services and products, and even choose to interact with consumers through such forums in cases where service recovery is needed (Gu and Ye 2014).

If online reviews represent customer engagement, it would follow that the performance of service firms on such online platforms should be associated with firm performance. Considerable prior work has focused on establishing these relationships, such as the effect of online reviews on sales and consumer choice

(Chen and Xie 2008), loyalty (Yoo et al. 2013), trust (Awad and Ragowsky 2008), and brand image (Dellarocas 2003). These studies typically use econometric models to explain some form of business outcome by using one or a combination of these numerical measures: dispersion (variance of the ratings), valence (the numerical rating), and volume (the number of ratings) (Dellarocas and Narayan 2006). The studies focusing on the effect of eWOM on firm performance have also spanned a variety of research contexts: television ratings (Godes and Mayzlin 2004), movie box office sales (Liu 2006, Dellarocas et al. 2007, Duan et al. 2008), beer sales (Clemons et al. 2006), and medical care (Gao et al. 2012).

A common element in much of this work is the use only of the numerical variables associated with online reviews discussed earlier - valence, dispersion, and volume. As Archak et al. (2011) note, there are a number of potential issues with only using these numerical variables to represent the consumer experience. First, by compressing a complex customer engagement to a single number, product or service quality is assumed to be one-dimensional, which is most likely inaccurate. Second, individual preferences are highly heterogeneous, so a single number might not be sufficient to convey the same information to everyone (i.e. a 4-star review might be differently perceived). Third, ratings have been shown to have significant bias (Li and Hitt 2008, Chen and Lurie 2013). Given these reasons, recent work has argued that a more granular approach to capturing customer engagement and service quality is warranted from such sources (Van Doorn et al. 2010). A significant opportunity is available therefore to consider not only the numerical data, but also the text of the reviews, which provide a more detailed and personalized account of the service

experience.

Some early work has adopted this strategy of analyzing the economic impact of text information and thematic structures present in online reviews (Decker and Trusov 2010, Archak et al. 2011, Netzer et al. 2012). Building on this research, we adopt a robust and scalable approach based on latent semantic analysis (LSA) (described later) of review text. LSA allows the identification of semantic themes within a corpus of review text and scores each service provider's reviews on these themes, allowing deeper measurement of the provider's service quality in addition to numerical scores. Such a data-driven approach builds on the wide availability of large amounts of text within online platforms as well as provides adaptability to the specific services provided, thereby addressing the point made by Zeithaml (1988): "To bridge the gap between specific characteristics and the abstract concepts of quality, it is useful to consider service quality in terms of broader dimensions." Once we extract such dimensions of service quality, we need to relate them, in addition to the available numerical data, to a measure of firm performance. In the next section, we provide arguments for why they influence survival, a critical outcome in service operations.

Service Quality and Survival

We argue that the information provided through online reviews has significant implications for the health and wellbeing of the service provider, and therefore study the association between consumer feedback provided through such reviews and the survival of the firm. Prior work has attested to the importance of survival as an important outcome in the services context (Kalnins and Mayer 2004, Rosenzweig et

al. 2011). We thus build on earlier work in service operations showing the relationship of service quality to financial outcomes (Chase and Apte 2007), as well as the connection between online reviews and firm performance (Chen and Xie 2008), business survival specifically here.

The survival of restaurants has been examined in the hospitality research area in some detail through in-depth case studies (Parsa et al. 2005, 2011) and through large-scale analyses using BLS data (Luo and Stark 2015). The factors identified through such sector-specific analyses fall well within the purview of prior research studying organizational survival. Two broad sets of factors have been identified to be influential in explaining survival; ecological factors around the firm, and factors associated with human capital, i.e. capabilities of the firm itself (Brüderl et al. 1992). Ecological factors refer to the competitive environment around the firm, the presence of strong institutional support, and resource munificence (Castrogiovanni 1991, Carroll and Khessina 2005). On the other hand, firm-specific attributes, such as the experience and skills of the founding team, the extent to which the new firm can rely on external resources (such as networks through franchises), and funding relationships, can also influence survival (Bates 1990, Kalnins and Mayer 2004, Bayus and Agarwal 2007, Luo and Stark 2015).

Within this stream of work, the specific role played by consumer feedback from online reviews (representing consumer acceptance and approval) has not been addressed. The closest analog emerges from papers studying the impact of media coverage on firm performance (Pollock and Rindova 2003, Pollock et al. 2008, Petkova et al. 2013) wherein commentary on firms from traditional media outlets

influences the reputation and brands of firms positively. Similar effects on firm reputation have been observed from discourse in social media (Luo et al. 2013) but not specifically from reviews of the firm's products and services itself. In hypothesizing the effects of online reviews on survival, we argue for two causal mechanisms. First, online reviews help firms gain attention from potential customers and be included in the consumers' consideration set (Dellarocas 2003, Roberts and Lattin 1997). Second, details provided within the reviews help consumers form judgments of the relative merits of the service provider (Ludwig et al. 2013), thereby increasing "conversion", i.e. consumption of the service.

Gaining the attention of potential consumers is usually the first step in any service encounter, especially in experience service settings (Rosenzweig et al. 2011). Attention is typically linked to the salience and availability of a stimulus (Davenport and Beck 2013, Pollock et al. 2008). The more salient information is available for the decision, the greater the likelihood that the decision maker invokes the availability heuristic (Sunstein 2002), thereby actively considering the object receiving the salient discourse (Roberts and Lattin 1997). In the service sector, online reviews have become increasingly salient in helping garner attention for the focal provider, thus allowing the entry of the provider into the consumer's consideration set (Vermeulen and Seegers 2009). A recent survey reported that 92% of users consider using a local businesses with a 4-star online rating while only 13% do so for a business with 1 star², attesting to the salience of online reviews in general. Beyond the rating, review text outlining specific interactions with the service provider in persuasive and

² <https://www.brightlocal.com/2014/07/01/local-consumer-review-survey-2014/>

informative text is likely to also contribute to enhancing the salience of the service provider, thereby increasing the odds of the marginal consumer including the service provider in his or her consideration set (Cao et al. 2011).

Beyond entering the consideration set of consumers, service providers are evaluated by potential consumers before actual consumption. Evaluation requires judgment and depends on the presence and level of specific attributes (Fiske and Taylor 2013), which in the context of online reviews requires processing the information contained in the text and noting details of the service experiences provided by prior consumers. In fact, the BrightLocal survey cited above showed that 85% of consumers surveyed read up to 10 reviews before making consumption decisions. The text in reviews therefore needs to reflect the positive “experience” that consumers need to make positive judgments. High quality service providers are likely to generate more text in their reviews that are associated with and indicative of engaging consumer experiences, expressed with enthusiasm and conviction (Pullman and Gross 2004). By contrast, lower quality providers will garner review text that are bland and do not mention any specifics of the consumers’ interactions with the restaurant. An engaged restaurant review (“great good, wonderful ambience, low wait time, attentive wait staff and positive overall experience”) is likely to significantly influence judgments, leading to consumption and repeat business. We argue that these influences may convey more effect than the numerical ratings per se.

When viewed in aggregate, service providers with higher ratings as well as strongly persuasive content in their reviews are likely to signal strength and quality in their competitive environments, thereby influencing first-time consumers. In addition,

the presence of positive ratings and text will also enhance the likelihood of repeat business and loyalty, thereby increasing the financial viability of the service provider (Parsa et al. 2005). In the aggregate, signals within online reviews conveying positivity and engagement should be linked to overall sales and profitability, thereby enhancing the odds of survival. We test these expectations in the restaurant context within one metropolitan area and describe, next, the methodology and data we use in our empirical analysis.

Methodology

Research Context

For this study, we have chosen to focus on restaurants in a large U.S. metropolitan area, Washington, D.C. Prior literature (Mangold et al. 1999, Gu et al. 2012) suggests that restaurants and other high involvement services provide a more ideal context through which to study the effectiveness of eWOM. We focus on restaurants since these represent a common and ubiquitous service context where online reviews are used extensively (Lu et al. 2013). As a first step, we first identify all current and open restaurants in the D.C. area, as of December 2013, using the Washington D.C. municipal city database. This process provides us with a master list of more than 2000 restaurants that are “going concerns” during the time of analysis. Subsequently, we assemble a comprehensive list of restaurant closures in the D.C. area, drawing from two data sources.³ First, we collect search results of restaurants

³ While it is possible to consider restaurants in Northern Virginia or Maryland, there are regulatory and tax implications within D.C. that are significantly different. Therefore, to reduce unobservable heterogeneity in our analysis, we focus only on restaurants that are located in D.C. We also omit

reported as closed on Yelp.com and FourSquare.com. Second, we collect reports of restaurant closings from Eater.com and Gaylot.com, two sites dedicated to local culinary events including restaurant openings and closings. In total, this data collection effort results in a list of 575 restaurants that closed between 2005 and 2013. Out of these 575 restaurants from the D.C. metropolitan area, some were located close to D.C. but in Northern Virginia and Maryland, which were deleted, leaving us with 446 restaurants located inside the District. Finally, we manually confirmed that these 446 restaurants truly closed instead of changing locations or other non-closure events. While this set of closed restaurants may not be comprehensive, it constitutes a representative cross-section of cuisines, segments and geographical locations within D.C. Therefore, for the purposes of our analysis, i.e. understanding the effect of reviews and review text on restaurant closures, our sample of closed restaurants is adequate and appropriate. The sources that we use to gather data on closed restaurants also provide reasonable dates regarding the actual closure event. Therefore, we are able to approximate the closing period of the restaurant to the nearest quarter (each time unit represents a quarter), which is the highest level of granularity possible.

Coarsened Exact Matching (CEM)

While it is possible to estimate longitudinal models of restaurant closure using only this set of closed restaurants (reported later in the paper), in order to identify factors that are associated with business outcomes (i.e. restaurant closure), it is ideal to augment this dataset with restaurants that have remained open. This matching

franchises of fast food firms, since they are typically treated as extensions of the brand and are unlikely to be reviewed independently.

process ensures balance across covariates to be used in regression and also account for unobservable variables that may be particular to the set of closed restaurants. For example, some idiosyncratic factor that is unobservable to us may influence the closure of low-priced restaurants, but not high-priced restaurants. In such contexts, it is possible to mistakenly attribute the effect to other variables that are observable. Through matching, we are able to reduce, by a significant amount, the impact of such unobservable factors that may affect specific restaurant types. Matching also ensures that survival models of similar restaurants are estimated within the dataset, thereby minimizing potential bias.⁴ To this end, we create a set of similar (matched) restaurants that have remained open through our observation period, thereby comparing restaurants that have closed (treated cases) with restaurants that have not closed (controls) but that are otherwise as closely matched as possible on all other observable characteristics. We implement Coarsened Exact Matching (CEM), introduced by Iacus et al. (2011b) and shown to be a monotonic imbalance bounding (MIB) matching method with several beneficial statistical properties compared with prior Equal Percent Bias Reducing models (Rubin 1976), such as Propensity Score Matching (PSM) and Mahalanobis distance-based matching. In the words of Iacus et al. (2011b), CEM “generates matching solutions that are better balanced and estimates of the causal quantity of interest that have lower root mean square error than methods under the older existing class, such as based on propensity scores, Mahalanobis distance, nearest neighbors, and optimal matching”. Given the longitudinal structure of our data and the fact that closures are much less common

⁴ We also repeat our analysis on the set of closed restaurants only, with consistent results, as discussed later.

over time, a technique like CEM is better equipped to handle our context than other matching methodologies, such as PSM (Iacus et al. 2011b). Operationalizing CEM consists of three main steps, after identifying the set of variables to be used in matching treated cases (closed restaurants) to one or more controls (open restaurants):

1. Temporarily coarsen each matching variable, i.e. iteratively classify (recode) each variable into smaller strata or partitions so that the “substantively indistinguishable values of the variable are grouped together and assigned the same numerical value” (Iacus et al. 2011a, p.8), thus “coarsening” the variable
2. Sort all the observations (cases and controls) into strata that contain all the possible interactions of the stratified matching variables
3. Discard the observations in any stratum that does not include at least one treated and one control unit and revert to the “uncoarsened” measure, thereby allowing the final dataset to have optimal balance of treated and control units within each stratum.

Starting with an initial list of 2,021 potential control restaurants (representing all open restaurants in our dataset), we follow the progressive coarsening method outlined in Iacus et al. (2009) to select the matching variables. The matching variables we use are listed in Table 2.1 and, consistent with prior work (Lu et al. 2013), include all characteristics (excluding review text) of the restaurant observable from Yelp.com. Since we have a longitudinal dataset, we match each closed restaurant with open restaurants exactly in the time-period exactly before the focal restaurant closed. Thus, ensuring that each closed restaurant is matched to a similar

open restaurant during the most relevant time-period. To ensure completeness of the panel, we include such matched restaurants across the full period of our panel (i.e. 2005-2013). A key goal of the matching process is make sure that the treated and control groups are balanced (Iacus et al. 2011a), i.e. their covariates have similar distributional characteristics. Table 2.2 displays an imbalance table of our numerical matching variables, a preprocessing step recommended by Iacus et al. (2009) to evaluate the balance of variables in the case and control samples. The L1 statistic, introduced by Iacus et al. (2009), is a univariate measure of the imbalance between treated and control units, which is very low for our numerical variables. However, we do observe imbalance between treated and control units in the distribution of the number of reviews at the 25, 50, 75, and 100 percentiles, which also indicates the extent to which a matching procedure like PSM would have led to significant imbalance between the two samples. The results of our CEM procedure results in 446 closed restaurants (the entire closed sample) matched to 605 open restaurants (out of 2,021 possible). Thus, the final sample is based on a 1: k matching, shown to be desirable in previous work (Stuart 2010). A comparison of the matched and treated samples (Table 2.3) show near-identical values on covariates. We stress here that in comparison to propensity score matching, where the method tried to find a match for every treated unit, the focus in CEM is to only include appropriate matches in subsequent analyses. Therefore, the fact that the final sample does not include exactly k matches for all treated units does not signify a weakness of the matching process. Rather, it conveys the notion that the matched dataset achieves balance on observable covariates across treated and control groups, thereby allowing unbiased estimation of

treatment effects.

Finally, we collect online reviews for each restaurant from Yelp.com. This data collection resulted in over 35,000 online reviews for closed restaurants and over 67,000 reviews from open restaurants between 2005 and 2013. This set of over 130,000 online reviews contains roughly 15 million words collectively used in our text-mining methodology and econometric models, described in the next section. Before detailing these procedures, we briefly introduce text analysis and the general statistical tools used in this paper.

Extracting Service Themes from Text

With the review text we follow the standard preprocessing procedures of transforming text to lowercase, removing words composed of less than 3 characters and very common words called stopwords, and stemming words. The general objective underlying these pre-processing steps is to emphasize meaningful words by removing uninformative ones, and to keep the number of unique terms that appear in the corpus from becoming extremely large, one of the main computational challenges in text mining. Examples of stopwords include “the”, “and”, and “of”. Stemming refers to the process of removing suffixes, so that words like values, valued and valuing are all replaced with valu. We use the Porter stemming algorithm, which iteratively applies linguistic rules to identify and remove suffixes. Porter stemming is a standard algorithm implemented in most text mining software, including the “tm” (text mining) package (Feinerer and Hornik 2012) within R (R Core Team 2013), used for our analysis.

After preprocessing the text, we utilize a topic modeling technique called

Latent Semantic Analysis (LSA) to represent each review with a set of numerical covariates that capture service themes within the review text. LSA is a classical technique for extracting such themes (often called topics, factors or dimensions in related fields) from text data. It strongly resembles Principal Component Analysis for numerical data, as both techniques are based on the Singular Value Decomposition (SVD) (see the Appendix for a detailed discussion of SVD in the context of text).

Recently probabilistic approaches, such as Latent Dirichlet Allocation (commonly referred to as topic modeling Blei et al. 2003), for recovering themes in text have become popular. While these probabilistic approaches can recover more interpretable results from the perspective of data exploration and information retrieval, they also require several parameters to be defined for estimation and results can change each time the estimation is performed on the same dataset. These features create significant challenges for assessing statistical significance and stability of results when combining the extracted service themes with econometric models. One of LSA's comparative advantages is that it is essentially parameter-free and returns the same results each time that LSA is applied to the data, thus enabling more accurate econometric estimation that are used to derive managerial insights. Further discussion on how LSA compares to its probabilistic alternatives is also provided in the Appendix.

Similar to PCA, it is necessary to first evaluate how many semantic themes may emerge from applying LSA to the reviews dataset. The process of identifying the most suitable number of themes, referred to as cross-validation, suggested five service themes underlying the review text (described in detail in the Appendix). Once

the themes are identified, each review is represented with numerical scores for each of the extracted themes (akin to PCA loadings), and each theme has a list of rank ordered keywords (shown in Table 2.4) that we inspect to understand the semantic meaning behind the extracted themes. We note that in many such text-based settings, it is up to the researcher to examine the underlying text associated with a specific theme and evaluate their relevance. In our case, as shown in Table 2.5, phrases from top reviews that load heavily on each service theme are interpretable and consistent with an underlying service dimension. Therefore, we name the five themes: Quality_Overall, Food_Efficiency, Food_Quality, Responsiveness and Atmosphere. We note that one could follow other strategies, such as simply naming the theme with the most important keyword that emerges from the analysis. However, to ensure meaningful and accurate results, we prefer to utilize domain knowledge when assigning the variable names. In our analysis, we discover interpretable weights assigned to individual words that could be used to understand the concepts behind each of the themes, hence the variable names associated with each of the five themes.

To extract insights and combine the text analysis with an econometric framework, we then summarize this information into a panel dataset where each restaurant-time period is the unit of observation. Specifically, we aggregate the data for each restaurant (restaurant variables, reviews characteristics, and the calculated LSA themes) within quarterly (three-month) periods. Restaurants without reviews in a time period are excluded. We also exclude restaurants-time periods occurring before the first review of the restaurant or after restaurant closure, thus resulting in an unbalanced panel data set. For each of the variables in our panel, we only consider

the new reviews that were added on Yelp for that restaurant in the previous time period (i.e. one quarter). In the case of the semantic components calculated through LSA, we average the reviews' loadings (weights) on each of the five components observed in the previous period (quarter). As a robustness check, we also perform our analysis by aggregating data at the monthly level instead of the quarterly level, with strongly consistent results.⁵ We next describe the econometric models estimated for survival, using all the data made available through online reviews, including those created through LSA.

Econometric Models and Estimation

The outcome variable of interest in our dataset of reviews is the variable Closure_{it} , which equals 1 if restaurant i has closed at time t (denoting year-quarter) and 0 otherwise. The independent variables of interest are: fixed restaurant characteristics RestChars_i , such as price point of restaurant i ; time-varying review characteristics RestChars_{it} , such as number of reviews during time period t for restaurant i ; and LSA variables LSA_{it} , which includes the average value of each LSA variable over all reviews of restaurant i occurring in time period t . While it is possible to include the actual sentiment of the review as well (calculated on $(-1,1)$ where 0 represents neutral tone), we observe that this variable is highly correlated with the numerical rating for each review. This correlation is not surprising since positive sentiment in the review, suggesting overall satisfaction with the restaurant, should garner higher review rating (Cao et al. 2011). Therefore, we only include

⁵ These results are available upon request.

rating in our models (robustness tests replacing rating with sentiment provide identical results). We also note that the explanatory ability of our models rest on the reasoning that the latent service quality themes identified through LSA are relevant, beyond the effects of aggregate rating or sentiment.

Further, following Card and Krueger (2000) and Parsa et al. (2005, 2011), we control for the specific geographical location of the restaurant, Loc_i , which includes the restaurant's zip codes (12 in total), and we control for the competitive environment of each restaurant by including competition variables $Comp_{it}$, such as the number of reviews of the competitors of restaurant i at time t . For each restaurant i , we define competitors as restaurants with the same price point and location. All independent variables were scaled and centered. Building on Chevalier and Mayzlin (2006) and Archak et al. (2011), we model the impact of review characteristics and themes on the closure of restaurants while controlling for restaurant characteristics. Our baseline econometric model therefore is:

$$\begin{aligned} \text{Closure}_{it+1} = & b_{i0} + \beta_0 + \gamma \text{RestChars}_i + \delta \text{RestChars}_{it} + \alpha \text{LSA}_{it} + \rho \text{Loc}_i \\ & + \xi \text{Comp}_{it} + \epsilon_{it}, \end{aligned}$$

where γ and δ are vectors containing the sets of coefficients corresponding to restaurant and review characteristics respectively, α contains the coefficients corresponding to the LSA variables, and ρ and ξ are vectors containing the set coefficients corresponding to location and competition. As suggested by Ghose and Ipeirotis (2011), we also control for the review length and the readability scores of

each review.⁶ All dependent and independent variables used in the analysis are described in Table 2.6, and summary statistics of each variable are shown in Table 2.7. The correlation matrix for the coefficients of each variable is shown in Table 2.8.

Longitudinal Generalized Linear Models

We proceed to model our specification using a generalized linear mixed model (GLMER) as described by McCulloch and Neuhaus (2001) with a binomial family using a logit link and a random intercept (equivalent to a panel logit model). As part of the generalized linear model, we estimate different variations on the baseline model with and without the semantic themes. We start with a regression specification that assumes that in each period, a restaurant chooses whether or not to exit the market. Clearly, while many factors influence this decision, the information within online reviews received in the preceding period captures one aspect of restaurant quality that may be significantly associated with the decision to exit the market. Therefore, we estimate the following models.

Model 1 only includes restaurant and review characteristics:

$$\text{Closure}_{it+1} = b_{i0} + \beta_0 + \gamma \text{RestChars}_i + \delta \text{RestChars}_{it} + \rho \text{Loc}_i + \xi \text{Comp}_{it} + \epsilon_{it}.$$

Model 2 includes restaurant and review characteristics and the LSA variables:

$$\begin{aligned} \text{Closure}_{it+1} = & b_{i0} + \beta_0 + \gamma \text{RestChars}_i + \delta \text{RestChars}_{it} + \alpha \text{LSA}_{it} + \rho \text{Loc}_i \\ & + \xi \text{Comp}_{it} + \epsilon_{it}. \end{aligned}$$

⁶ While we report model results using word length and the SMOG readability index following [Ghose and Ipeirotis \(2011\)](#) as the measures for the review length and readability, respectively, we also run similar models with other review length and readability measures, such as character length, the number of sentences, the Automated Readability Index (ARI), the Fog-Gumming readability index and the Coleman-Lindau readability index. These results are consistent and are available upon request.

Model 3 is the same specification as Model 2, except that the sample of restaurants i are restricted to the set of closed restaurants Ω , i.e., we require that $i \in \Omega$. Model 3 allows us to examine the robustness of the results to the matching process. In addition to Models 1-3, we fit two additional models using a generalized linear model (GLM) with a binomial family using a logit link as well as fixed effects for each restaurant and year. The added fixed effects provide more conservative coefficient estimates and allows restaurant-specific heterogeneity to be accounted for. The first GLM model (Model 4) applies an exchangeable correlation structure, while the Model 5 applies an AR(1) correlation structure. The correlation structure pertains to how observations from the same restaurant, i.e. within a group, are assumed to be correlated, and thus affect the standard error calculations but not the coefficient estimates themselves. An exchangeable correlation structure assumes that every pair of observations in a group has the same correlation. Alternatively, an AR(1) correlation structure assumes that observations that are closer in time are more strongly correlated with each other, a pattern that has been observed in prior work in online reviews (Moe and Trusov 2011).⁷ In general, evaluating the correlation structure in a GLM model through the application of different structures is important in assessing model reliability and adding credence to the model results (Liang and Zeger 1986). We thus estimate the following two specifications: Model 4 adds fixed effects for restaurant and year, assuming an exchangeable correlation structure:

$$\text{Closure}_{it+1} = \beta_0 + \beta_i + \theta_{\text{Year}_t} + \delta \text{RestChars}_{it} + \alpha \text{LSA}_{it} + \epsilon_{it}^{\text{exc}},$$

⁷ In further robustness tests, we estimated models with lagged variables using more than 1-period lags for the semantic components. The current period variables remained significant, while the lags were not significant. Therefore, we only report the model specifications that do not include lagged variables.

where Year_t is the year portion of the year-quarter time period t and ranges from 2004 to 2013. Model 5 also has fixed effects for restaurant and year, assuming an autoregressive AR(1) correlation structure:

$$\text{Closure}_{it+1} = \beta_0 + \beta_i + \theta_{\text{Year}_t} + \delta \text{RestChars}_{it} + \alpha \text{LSA}_{it} + \epsilon_{it}^{AR(1)}.$$

The results of each model are displayed in Table 2.9. These models show that a number of variables are significant in their associations with restaurant closure. First, certain fixed restaurant characteristics have a significant effect on the probability of restaurant closure. Restaurants in the higher-priced segments and certain cuisines (e.g. American food) have a higher probability of closure. Second, as the literature has previously shown, numerical review characteristics are significant predictors of restaurant closure, with both the average rating and number of reviews being strongly negatively associated with the probability of closure ($p < 0.001$). Interestingly, the effect of the number of reviews, representing foot traffic into the restaurant, is much stronger than that of average ratings in each time period, showing the importance of review volume.

Regarding the semantic themes, we observe a range of p-values and coefficient magnitudes and signs. The coefficient of `Quality_Overall` is highly significant ($p < 0.001$ in 3 out of 4 models) and strongly negatively associated with closure. In contrast, the coefficient for `Food_Efficiency` is significant ($p < 0.01$ in 3 out of 4 models) but positively associated with closure. `Responsiveness` is marginally significant ($p < 0.1$ in 2 out of 4 models) and negatively associated with closure. Strikingly, the magnitude of the marginal effect of `Quality_Overall` is much larger than that of the average rating in all models, and the magnitude of the marginal effect

of Food_Efficiency is larger than that of the average rating in all models, suggesting that certain thematic characteristics of review text might be more important than the ratings associated with reviews. However, not all semantic characteristics are significant. Food_Quality and Atmosphere are not significant in any of our models. Therefore, we find evidence not only that the content of the review text provides semantic structure that explains business outcomes, but that after controlling for all the possible dimensions of the review text in this context, certain dimensions affect restaurant closure more than others.

The results from Quality_Overall and Responsiveness show negative coefficients, which is consistent with the notion that restaurants with reviews that clearly articulate quality and responsiveness are less likely to close (more likely to stay open, i.e. $Closure_{it} = 0$). However, the positive coefficient for the Food_Efficiency component is puzzling, since it appears that the articulation of efficiency-related words in the review is associated with restaurant closure. We investigated this in some detail in our text by reading and analyzing several dozen reviews with high topic weights on this theme. We see that typical reviews with high weights on Food_Efficiency tend to use words like “wait”, “time” and “hour” in describing their experiences; the reviews tended to describe the waiting time involved, the use of the bar for waiting and the use of time measures in hourly intervals (half-hour, quarter of an hour).⁸ Even though there may not be clear negative sentiment attached to these reviews, the use of terms like “ask”, “wait” and “hour” are typically associated with negative outcomes in service contexts, given the

⁸ It may be more appropriate to rename this theme “Food_Inefficiency” given the meaning conveyed in the reviews.

same overall ratings and review volume. Moreover, reviews that explicitly mention wait times associated with service are more likely to be negatively tinged. This could be one explanation for why this particular component is associated with higher odds of restaurant closure. Though physically scanning all 130,000 reviews for this information is daunting, the several dozen random reviews that we manually evaluate for this component suggested enough ambiguity around wait times and efficiencies articulated in the reviews for the focal restaurant. A restaurant that tends to show reviews with high topic weights on this theme would benefit from evaluating the efficiency of the restaurant's operations from a timeliness perspective.

The results from Model 3, which only includes closed restaurants and does not include the case-control matched data, are very similar to Model 2, which includes the matched sample. The similarity in the direction and pattern of results suggests that the matching process is not directly influencing any inferences we draw about the effects of the semantic themes. Finally, we note that including the LSA variables in the regression significantly improves the overall fit of the model, as suggested by a 25% reduction in the AIC of Model 3 compared with Models 1 and 2. The AIC reduction demonstrates noticeably better fitting models and highlights the significant value of the semantic information captured from the review text in the LSA variables, in the presence of traditional numerical variables. In the Appendix, we explore survival prediction using these models, and show that the enhanced model fits when using the semantic themes actually lead to more accurate predictions.

Robustness Check with the Cox Proportional Hazards Model

We have thus far viewed restaurant closure as a binary variable, i.e., we assume that the restaurant decides in each period whether to remain open or close. However, the length of time a restaurant survived before closure varies greatly and provides information about the overall success of a restaurant before closure. Furthermore, the length of time a restaurant has survived allows us to differentiate between a restaurant that has been running for years and one that only opened in recent weeks or months (Luo and Stark 2015). Survival or duration models are designed to incorporate information on both cases for which an event of interest has occurred and those for which an event has not yet occurred, correcting for the effects of censoring (Lin and Wei 1989). We examine the effect of restaurant characteristics, review characteristics, and our themes on the failure rate of restaurants by using the semi-parametric Cox proportional hazards regression method (Cox 1972). The Cox model specifies the failure rate of restaurants $h_i(t)$, as the product of a vector of covariates $X_i'(t)\beta$, and a time dependent baseline rate $h_0(t)$, so that the form of the hazard function is given by:

$$h_i(t, X_i'(t), \beta) = h_0(t)e^{X_i'(t)\beta}.$$

To quantify the effects of restaurant and review characteristics on the timing of restaurant closure, we employ a Cox proportional hazards survival model including all the same variables as in Model 2 above. Because we make no claims about the functional form of time dependence, the Cox proportional hazards model offers a plausible approach to modeling survival in restaurants.

The results of this model, shown in Table 2.10, are qualitatively similar to our

GLMER models in sign and significance. To allow comparisons to the previous GLM models, a positive (negative) coefficient in this model corresponds to a higher (lower) probability of closing. As in the GLMER model, higher rating and number of reviews are highly significant, with the number of reviews having a much stronger effect than average rating. Also consistent with GLMER models, *Quality_Overall* is highly significant ($p < 0.001$) and strongly negative, *Food_Efficiency* is significant ($p < 0.01$) but positive, and *Responsiveness* is marginally significant ($p < 0.1$) and negative. Again, we observe that marginal effects of the themes on closure (*Quality_Overall*, *Food_Efficiency*, and *Responsiveness*) are stronger than the marginal effect of the average rating of reviews for a time period, thus highlighting the importance of the LSA variables that capture the underlying content in review text. Also, we see the positive coefficient of *Food_Efficiency*, which suggests that the use of words associated with wait times and efficiency tend to provoke negative implications for quality. The remaining two themes, *Food_Quality* and *Atmosphere*, are not significant in this model.

Cox proportional survival models assume the proportional hazards condition, which states that covariates are multiplicatively related to the hazard function. We test whether this assumption holds here. We follow the diagnostics tests suggested (Grambsch and Therneau 1994, Maindonald and Braun 2006) to implement both a regression and a graphical approach to test non-proportionality. In the regression approach, we measure the correlation of scaled Schoenfeld residuals and time (ρ) to test the proportionality of the main predictors in the model. None of the variables are statistically significant ($p < 0.05$), as shown in Table 2.11, indicating no violation of

the proportionality assumption. In a graphical approach, we plot the Schoenfeld residuals (Schoenfeld 1982), which are based on the individual contributions of a covariate to the derivative of the log partial likelihood (Hosmer et al. 2013). If the Schoenfeld residual shows a random pattern at each failure time, suggesting that the covariate effect does not change with respect to time, the proportionality assumption holds. Figure 2.1 does not show significant deviations from a horizontal line, indicating no systematic violations of the proportionality assumption, and thus consistency of the estimates and robustness of results.

Discussion and Implications

The question of how firms can leverage service quality and customer satisfaction metrics to improve firm performance has been on the agenda of service scholars since the early 1970s (Chase and Apte 2007). This question has inspired a large body of research on the strategic importance and measurement of service quality for organizations. Although scholars have identified a wide variety of measures for service quality relying mostly on questionnaires (Parasuraman et al. 1985) and have successfully established a relationship between service quality and performance through multiple empirical frameworks (Heskett et al. 1994, Roth and Jackson 1995), this literature continues to bemoan the difficulty in defining standardized measures of quality for the service industry (Metters and Marucheck 2007). This issue is particularly troublesome at a time when customers are increasingly focused on experiences (Pine and Gilmore 1999) and engagement (Voss et al. 2008), which are not easily captured by survey-based measures (Roth and Menor 2003). In the words of Soteriou and Zenios (1999), “In order to answer the

'how' questions we need to address 'what' questions first. What are the operational characteristics of a service that translate to customers' high levels of quality?"

In this study, we argue that online reviews, along with the large corpuses of text written by customers about their experiences with service operators, can be of significant value to measure customer satisfaction and to further understand the link between service quality and firm performance. One of our contributions is thus in applying text mining techniques to extract, from online reviews text, distinct service quality dimensions that bridge the gap between specific operational characteristics and service quality, often the focus of the extant work (Goldstein et al. 2002). We apply this approach to a comprehensive set of online reviews of restaurants in the Washington D.C. area from 2005 to 2013. The context of restaurants is particularly appropriate for a study of online reviews where a recent report by National Restaurant Association (2012) stated "Simply put: online reviews can help or break your business" and found that more than half of diners report that information a peer review site is likely to affect their decision to choose a restaurant. Through a series of econometric models, we provide strong evidence that the semantic components extracted from the review text firstly capture relevant information above and beyond numerical attributes, and second, are significant predictors of business outcomes such as business closure. Of the five components that were extracted from the text, three are significant predictors of restaurant closure. More importantly, we show that the marginal effects of these variables are significant even when accounting for the mean rating for the reviews in the same time period, suggesting that there is considerable information within the text that is likely used by human consumers of the reviews but

not accounted for in large-scale econometric models of performance. Our approach, founded on classical text analysis methods, allows us to extract relevant service quality dimensions associated with latent “restaurant quality”, thereby allowing for a fuller view of restaurant performance.

Our work has limitations that are worth discussing. First, we focus on one specific segment of online reviews to analyze, raising questions about the applicability of this method to other contexts. As outlined above, we believe the methods can be extended but future work is needed to ensure generalizability. Second, we do not provide a comprehensive model of restaurant closure, and we do not have access to other factors that may be influential, such as revenue data and personnel issues within the merchant. Similarly, we do not consider information about the specific reviewer and their background that may color their reviews (Dai et al. 2012). These refinements are out of the scope of this paper and are also not the true objective of our work, which is to validate the process of service dimension identification through text analysis. Finally, we model survival, a discrete and final outcome variable. Other interim service operations outcomes may also be modeled with greater ease, such as health violations and responses to specific promotions; we are working on extending our models to these other contexts.

Beyond the service operations context, our work contributes to the research on the strategic legitimation efforts carried out by organizations. A recent but growing literature has examined how various forms of communications contribute to the development of legitimacy and access to resources (Martens et al. 2007, Porac et al. 2002). However, much of this work has focused on content created and distributed

by traditional media outlets (Rindova et al. 2007, Kennedy 2008, Petkova et al. 2013). In contrast, our study examines information created by customers in a social media platform where everyone can express their opinion. Our results show that legitimacy as well as quality signals can also be provided by the presence and textual content of these reviews as they influence the choices of future customers and ultimately firm survival. In line with Kennedy (2008), a study of the effect of press releases to the survival of 74 firms, we also are among the first (to our knowledge) to investigate the influence of social media on the survival of over 1000 merchants, an essential outcome in industrial organization.

Our work also provides important insights to practitioners. Prior work argues that service providers learn from their interactions with customers, especially in the service context (Clark et al. 2013), but these benefits come with significant costs. In contrast, online reviews provide, with little cost, many essential pieces of information, beyond the numerical ratings, that are of considerable value to such merchants (Cao et al. 2011). Beyond Yelp, other sources of review text, such as OpenTable and TripAdvisor, are also available to the service provider. Clearly, assimilating all this data manually is infeasible for either the restaurateur or the consumer. However, this information does have economic implications for both the restaurateur (as we show here) and the consumer. To the extent that more granular firm-level data is available (such as sales, promotions or restaurant check sizes), further refinements of the underlying method are possible and the individual effects of semantic themes on these outcomes can be estimated. Moreover, using our approach, restaurant managers and potential investors can evaluate a restaurant's

performance relative to that of its competitors, by considering the weights on the semantic components as well as the numerical ratings. Prior work in marketing has argued for the value in benchmarking for service organizations, primarily as a learning tool (Soteriou and Zenios 1999, Vorhies and Morgan 2005); our model allows for the development of quality-based benchmarking services for service firms.

Alternatively, platforms owners, such as Yelp.com, could use our approach to create multi-dimensional scores of quality using review text that might more accurately translate the review text and describe the true quality of a restaurant. These methodologies, if implemented on large corpuses of text, could provide value-added services through benchmarking for merchants and quality-based trends for consumers. While firms like Yelp do provide consumers with aggregate statistics on the ratings of reviews received over time, they are limited when it comes to processing the text within these reviews. Our work would help address this gap in their offerings, and even raise the possibility of additional merchant-specific services for additional revenues. More broadly, our work also paves the way for more work that combines text analytic or data mining applications with econometric models, thereby capitalizing on the large and varied forms of crowd-sourced and social media that are increasingly becoming of relevance in the services context today. We believe that more such work is needed to allow managers to fully understand consumer engagement and to optimize their service design appropriately.

Appendix A: Text Analysis Methodology

In this appendix, we discuss the procedures involved in analyzing text from the reviews in some detail. Note that, as mentioned in the main document, the text corpus of all online reviews for the restaurants in the final dataset undergoes preprocessing. After preprocessing the corpus, we construct a document-term matrix, which decomposes each document by the set of words contained in it. The matrix contains a column for each word that appears anywhere in the (preprocessed) corpus and a row for every document. Each matrix entry counts how often each term appears in each document. Let X be the document-term matrix for a corpus with n reviews and p unique terms. To understand the relationship between n documents and a response variable Y_i (e.g. restaurant closure) one could simply use the document-term matrix X as an input for a statistical model. One could try to estimate the following model

$$Y = F(X\beta + Z\gamma),$$

where Z are control or other variables of interest, and β and γ are the regression coefficients to be estimated. The coefficients β would quantify how different words or phrases affect Y . In practice, however, such a model is not identifiable, because the number of terms p is typically much larger than the number of observations n .

One possible remedy to this problem is to perform dimension reduction. In fact, a common strategy in the eWOM literature is to replace the term document-matrix with word count and valence (tone or sentiment). These variables can be computed by taking the sum over terms, $\sum^j X_{ij}$ as the word count for document i , and

valence can be viewed as a weighted sum $\sum^j w_j X_{ij}$, where w_j is positive or negative depending on the tone of the term. A drawback of using only word count and valence is that the underlying content within each document is not well captured. To address this shortcoming, we follow an alternative methodology that starts with the idea that many unique words can be summarized succinctly with a fewer number of keywords. For instance, one may try to combine the columns corresponding to *beer*, *whisky*, and *gin* with *alcohol*, if contextually appropriate. The underlying idea in the example is similar to that of Principal Component Analysis (PCA), where one replaces variables (words in the text setting) by their linear combination,

$$V = XH^T,$$

where columns of the $K \times p$ matrix H are factor loadings that serve to reduce the column space of X sufficiently for the model to be estimable. Thus, instead of the model in Equation 1, one tries to estimate:

$$Y = F(V\beta + Z\gamma),$$

where β quantifies how themes in the text captured in the columns of V affect Y . In the next section, we discuss how a set representing V is computed.

A.1. Latent Semantic Analysis (LSA)

Latent Semantic Analysis (Dumais 2004) is a matrix factorization technique that computes the factor loadings H in Equation 2. The classical LSA model relies on Singular Value Decomposition (SVD), where the document-term matrix X is decomposed as

$$X = U_K \Sigma_K H_K^T,$$

where U_K and H_K are orthogonal matrices, and Σ_K is a $K \times K$ diagonal matrix with

positive entries. The interpretation of each column of H_K is similar to that of PCA, where each column has length p and contains weights that are used to take linear combinations of the columns (terms) of X . The weights also have the same function of projecting the data onto axes of greatest variation. The first axis, defined by the weights in the first column of H_K , explains the most variation; the second column of H_K explains the second most variation and is orthogonal to previous columns, and so on.

The relation in Equation 4 is exact when the inner rank $K = \min(n, p)$, i.e., the inner rank equals the rank of X . If $K < \min(n, p)$ then the equality no longer holds and Equation 4 is an approximation. However, in this case the Eckart-Young theorem (Eckart and Young 1936) establishes that the SVD-based approximation is the most accurate in the sense of minimizing the Frobenious norm:

$$U_K \Sigma_K H_K^T = \operatorname{argmin}_D \|X - D\|_F^2.$$

Since the main goal of LSA is dimension reduction, K is always chosen empirically to be relatively small in order to summarize the data effectively while achieving parsimony. Often a technique similar to examining a scree-plot in PCA is used to choose the number of components to retain, which can be constructed by plotting the singular values in Σ_K . Another rigorous procedure to choose the number of components to retain is cross-validation, where the objective is to use random subsets of the data to fit LSA and another subset to assess its accuracy. The number of retained components is then cycled over and the one that corresponds to the lowest test error is chosen. Since we are primarily interested in factoring a matrix, we employ the two-dimensional cross validation of Owen and Perry (2009), which

selects sub-matrices for training and test data. The procedure for choosing the number of components to retain is statistically rigorous with consistency under appropriate regularity conditions (Owen and Perry 2009). As shown in the left panel of Figure 2.2, this cross validation procedure indicates that we retain 3 components.

LSA requires performing SVD on typically very large and sparse document-term matrices, which can easily create extraordinary computational and memory demands when using the well-known power algorithm for computing singular vectors. To overcome these limitations, extensive work in numerical linear algebra has led to fast and memory-efficient algorithms. We use an augmented implicitly restarted Lanczos bidiagonalization algorithm (implemented in the R package “irlba”), which is an iterative approach for calculating singular vectors in large, sparse matrices that is numerically stable and efficient from both a computational and memory perspective (Baglama and Reichel 2014). Once a rank K approximation has been estimated, then the i th document can be projected onto the space of H_K with the following:

$$\begin{aligned} V &= XH_K^T\Sigma_K^{-1}, \\ V &= [V_1, V_2, \dots, V_K], \end{aligned}$$

where V_1, V_2, \dots, V_K are new variables that capture the essential meaning of each document along different dimensions in the text. The purpose of right multiplying by Σ_K^{-1} is to rescale each dimension by the percentage of variation that the component explains. As mentioned above, we select 3 components for our study, i.e. $K = 3$. Accordingly, V_1, V_2 , and V_3 are identified to be used in subsequent econometric models of restaurant closure.

To understand the semantic meaning behind V_1, V_2 and V_3 , we inspect the

words with largest positive and negative weights and assign theme names based on these keywords. We note that one could follow other heuristics that assign variable names with, for example, the word with the largest magnitude weight. However, to ensure meaningful and accurate results, we prefer to utilize domain knowledge when assigning the variable names. In our analysis, we discover interpretable weights assigned to individual words that could be used to understand the concepts behind V_1 , V_2 and V_3 . With other datasets, it may be useful to also include phrases (called “n-grams” in text mining) when constructing the term document-matrix to recover more unique representations.

The top positively and negatively weighted words for each of the 3 retained components are shown in Table 2.4. Since X only contains non-negative values, by the Perron-Frobenius theorem about eigenvectors of non-negative matrices (Meyer 2000), the first component (column or set of factor loadings) of H_K will contain exclusively non-negative weights. Thus, the first theme V_1 takes a strictly additive combination of words, emphasizing ones that are most “interesting” in the sense that they maximize variation. Accordingly, we find, on examining the words associated with V_1 , that it emphasizes words related to broad aspects of service quality at the restaurant. For instance, words like order, time and service may relate to reliability and responsiveness; words like menu and food may relate to the business context of variety and taste. We refer to V_1 as Quality_Overall for convenience throughout the rest of the document; it represents an overall sense of the quality and value provided by the restaurant and by virtue of being the first extracted dimension, captures the most “variance” in the review text (comparable to the first principal component in

PCA). Subsequent components are likely to be associated with more specific issues pertaining to the restaurant's offerings. The second theme V_2 is a contrast between waiting time and food quality. If a review's score is strongly positive, then words like order, time, wait, hour and minute are utilized, so the specific review with high positive weights on this component reflects an emphasis on efficiency of service. On the other end of this component, reviews that have high negative weights emphasize food quality and taste, with keywords such as dish, flavor, chicken, sauce and food.

While V_2 can be entered directly into an econometric model, we note the underlying weights can be non-intuitive since the sign of the weights is arbitrary. For instance, words like "good" and "great" that have positive sentiment have negative weights, which can occur because the weights are *not* related to sentiment. They reflect the observed joint association of these words together in a review. Therefore, to avoid confusing caused by the positive / negative weights and to allow the variables to be more directly interpretable within a regression context, we split $V_2 = V_2^+ - V_2^-$ into its positive part ($V_2^+ = 0.5(|V_2| + V_2)$) and negative part ($V_2^- = 0.5(|V_2| - V_2)$) so that each part has weights that are non-negative. If a document loads heavily on V_2^+ , it likely focuses on efficiency-related issues – we call this theme Food_Efficiency. Similarly, if a document loads heavily on V_2^- , then it focuses on food quality and we call this theme Food_Quality.

Along the same lines of logic, we observe that the third theme V_3 is a contrast between responsiveness (keywords are server, waiter and ask) and restaurant tangibles, like atmosphere and entertainment. If a review's score is strongly positive, then it puts emphasis on the restaurant's responsiveness, while if a review's score is

strongly negative, then it puts emphasis on atmosphere and entertainment. V_3 is also split according to its positive and negative parts, which are referred to as Responsiveness and Atmosphere, respectively.

Returning briefly to the GLMER and survival model results presented in the main text, we note that the first theme (Quality_Overall) by itself allows considerable semantic structure to be extracted from the text. Every additional component adds to the model of restaurant closure but likely with lower power, given the sample of the dataset (as shown in the right panel of Figure 2.2). As larger datasets become easily available, the ability to identify smaller non-zero marginal effects in the econometric models improves significantly.

A.2. LSA and Other Topic Modeling Methods

We note that there are alternative techniques, such as Latent Dirichlet Allocation (Blei et al. 2003) or Probabilistic LSA (Hofmann 1999) that could instead be used to construct the matrix of factor loadings H that are based on more appropriate distributional assumptions for text data. Specifically, these alternative approaches are founded on a Poisson distributional assumption, instead of joint normality, which is assumed by LSA. These alternative techniques have the same goal of representing a document by its conceptual content through the factor loadings. In fact, as shown in Arora et al. (2012), Latent Dirichlet Allocation and Probabilistic LSA can be written in the algebraic form of Equation 4, but with different constraints. Instead of orthogonality constraints, UK and HK are subject to probability constraints, which can improve interpretability of the learned topics. However, there are several properties of probabilistic models that lead us to prefer the classical SVD-based LSA

approach. First, probabilistic models often provide results that change each time the data is analyzed. As written in Blei (2012), “As for many modern probabilistic models of interest-and for much of modern Bayesian statistics-we cannot compute the posterior because of the denominator, which is known as the evidence. A central research goal of modern probabilistic modeling is to develop efficient methods for approximating it.” Typically the approximation requires some degree of randomization, which causes path-dependent solution, whereas the SVD-based LSA can be computed exactly. Given that our ultimate goal is to understand the economic impact of online reviews, the SVD-based LSA avoids the additional and open methodological challenges associated with probabilistic models of accurately computing standard errors, performing hypothesis tests, and so on, within an ensuing regression analysis. The second drawback of probabilistic models is that the components of H_K are always highly positively correlated, since the potential improvement in interpretability is due to the relaxation of orthogonality constraints. In our experiments, Latent Dirichlet Allocation resulted in variables that were highly collinear, which would again create estimation issues in any ensuing analysis. Thus, uniqueness and orthogonality are properties of LSA that are advantageous for estimation of statistical and economic models in subsequent analysis. Moreover, as shown in Table 2.5, phrases from top reviews in each component are interpretable and consistent with the component keywords.⁹

⁹ In the interest of robustness, we apply probabilistic LSA and Latent Dirichlet Allocation to our corpus of text. While the topics identified in such analyses closely resemble the themes identified through LSA, they do not have the benefits of orthogonality, which aid in regression analyses subsequently. We estimated GLM models using these themes as well and found broadly similar results in terms of two or three themes being significantly associated with survival. However, we found the

The use of LSA and our overall results raise interesting methodological questions for future research. The extant text mining literature assesses performance along primarily two dimensions: algorithmic properties and accuracy in information retrieval tasks. Further study of how different text mining algorithms compliment econometric regression modeling to inform business decision-making would be a valuable contribution to multiple fields within the business community.

Appendix B: Post Hoc Analysis: Predicting Closure

The performance of GLMs as predictive models has been extensively studied using a receiver operating characteristic (ROC) curve (Pepe 2000, Bensoussan et al. 2009). One important feature of ROC curves is that they display the trade-offs possible between increasing the detection of true positives and increasing false positive rates as the positivity criterion varies. ROC curves are particularly useful for comparing models since tests are put on the same scale and the scale relates directly to the notion of accuracy (Pepe 2000). To generate out-sample prediction probabilities of restaurant closure as a function of time and generate the ROC curves for the GLMER models (Table 2.9), we perform longitudinal ten-fold cross-validation (Heagerty et al. 2000, Heagerty and Zheng 2005), which ensures that future observations do not predict past observations. Specifically, we divide the data into ten sequential groups of equal size. For each group $k = 2, \dots, 10$, the model is trained on the previous $k - 1$ groups and that model is used to generate predictions for group k . The associated ROC curves are shown in Figure 2.3, with their respective “Area

best model improvement with LSA, as expected given the properties of LSA described here. These extended results are available upon request.

Under the Curve” statistics (AUC). We observe a large increase in the ability to predict when and if a restaurant will close from our baseline model (Model 1) without semantic variables to our models containing semantic variables (Model 2 and Model 3). Moreover, our model with only closed restaurants (Model 3) has very similar performance to our model using the case-control data set (Model 2), showing that our predictive accuracy is not an artifice of the matching procedure. However, including both sets of restaurants is a more accurate reflection of the task facing restaurateurs and platform owners, since it is not known ex ante which restaurants are likely to close. The similarity of results from the closed set (Model 3) and those that include the matched sample (Model 2) is a sign of the robustness of our CEM-based matching method.

Chapter 3 Deal or No Deal? The Quality Implications of Online Daily Deals

Introduction

Daily deals offered by firms, such as Groupon and LivingSocial have become increasingly popular as consumers have flocked to their use; four out of ten New Yorkers reported using online deals for redemption at assorted retail outlets in 2014. Groupon, the market leader in daily deals, has over 53 million active customers, 200 million subscribers and reports selling over 400 million individual deals thus far globally (Groupon 2015a). While these numbers suggest that such platforms are an unqualified success, the effects on merchants offering them are equivocal. On one hand, research shows that only 55% of merchants offering such deals actually made a profit, while 26% reported losses (Dholakia 2010). On the other hand, daily deals have also led to increased foot traffic, revenues and visibility for merchants at a lower customer acquisition cost (Dholakia 2011a, 2011b). Reports in the practitioner press reflect this equivocality; articles describing their effects on business outcomes for merchants outline both positive as well as negative effects, adding to the ambiguity surrounding deals (Clifford and Miller 2012, Agrawal 2013, Cohan 2012).

In this paper, we focus on one specific business outcome that is associated with the offering of daily deals and also has implications for the long-term performance of the merchant offering deals – electronic word of mouth (eWOM). eWOM has emerged as an important factor, particularly in the services sector;

numerous studies have shown that eWOM associated with a firm and its products has a direct impact on firm performance metrics, such as sales, customer satisfaction, and brand evaluations (e.g., Chevalier and Mayzlin 2006, Clemons et al. 2006, Duan et al. 2008, Chen and Xie 2008, Zhu and Zhang 2010). 88% of consumers read online reviews (representing the firm's eWOM) to determine the quality of a local business, which then informs their purchase decisions (BrightLocal 2014). Thus, understanding the factors that affect eWOM is critical for merchants. Our work here address an important question in this domain: how does the offering of a daily deal affect the firm's resulting eWOM?

Studies on the question of whether daily deals influence brand evaluations, and the resulting eWOM discourse, have provided contradictory evidence thus far in the literature; on the one hand, Kimes and Dholakia (2011) find that a merchant's brand equity suffers no loss as a result of offering daily deals. On the other hand, in a conference presentation, Byers et al. (2012a) report that the offering a Groupon has a strictly negative impact on resulting Yelp ratings of merchants, implying that those considering daily deals as part of their marketing mix should exercise caution. Interestingly, both perspectives are significantly at odds with extant research on price promotions and advertising (that bear distinct commonalities with online daily deals), which propose various mechanisms by which price promotions and advertising efforts may affect (even positively affect) brand attitudes and consumer perceptions of quality (Biswas et al. 2013, Chen and Kirmani 2015). This divergence in the literature regarding the effect of daily deals suggests not only a significant gap in our understanding of how daily deals may influence eWOM (inasmuch as eWOM reflects

brand attitude and quality perceptions) but also a need for further research evaluating this question, potentially using multiple methods to isolate the specific effect.

In this paper, we address this gap; we extend prior work in marketing addressing price promotions and advertising to the online daily deal context. Eschewing the strong positions taken by Kimes and Dholakia (2011) and Byers et al. (2012a), we argue for a middle ground, contending that the effect deals may have on brand evaluations and eWOM will depend on the particular conditions of the merchant. Merchants who opt for daily deals are heterogeneous in terms of their characteristics as well as the environments in which they operate; we contend that this heterogeneity is likely to influence the manner in which daily deals affects their specific eWOM. More specifically, we focus on two sets of factors that moderate the impact of daily deal on eWOM. First, there is likely to be heterogeneity driven by specific merchant characteristics, such as the age of the merchant and the price segment in which the merchant operates. Second, the competitive environments in which merchants operate, especially the actions of competitors, may influence the focal merchants. The presence of daily deals in the competitive landscape thus may influence the response to a focal merchant's daily deal.

Prior work in marketing and IS provides theoretical arguments for why these moderators may indeed provide a more nuanced effect of daily deals on the resulting eWOM. Daily deals are effectively a combination of price promotions and opt-in advertisements (Edelman et al. 2011, Shivendu and Zhang 2013), since subscribers sign up to receive information about deals, which include elements of advertising copy and deep discounts. Research shows that the magnitude of a price promotion, in

dollar terms, tends to influence consumers' response to the promotion; the larger the perceived difference between the consumers' price expectation and the discounted price, the more positive the response (Grewal et al. 1998, Wu et al. 2004, Biswas et al. 2013). Thus, merchants in the premium price segment offering deep discounts should see a more positive response to deals than merchants in the lower price segment, even if the discount rate is the same. Alternatively, the persuasion knowledge model (Friestad and Wright 1994) asserts that promotional efforts perceived by consumers as appropriate and confident are likely to have a positive effect on brand evaluations, and potentially on eWOM. However, merchants perceived as defensive or desperate are likely to see a negative response to their promotional efforts (Kirmani 1990, Kirmani 1997, Chen and Kirmani 2105). Therefore, new merchants or merchants with strongly positive eWOM ex ante may garner more positive eWOM from offering daily deals in their attempt to gain new customers, in contrast to the negative effect shown by Byers et al. (2012a) or the null effect from Kimes and Dholakia (2011).

Beyond these sources of merchant heterogeneity, the competitive environment around the merchant contributes by setting reference prices for the focal merchant (Mazumdar et al. 2005, Bell and Lattin 2000). If all merchants in the competitive market offer daily deals, effectively the consumer's price expectation for similar products or services may shift lower towards the discounted price. In such contexts, merchants offering similar deals, partly as a response to competitive pressures, are unlikely to experience the noted negative effect. An interesting corollary to this

reasoning is the potential effect of deals offered by competitors on the focal merchant not offering a deal. These empirical questions remain unaddressed in the literature.

Finally, it is possible that merchants underestimate the operational complexities from short-term demand spikes that arise from offering daily deals (Blattberg et al. 1995, Pauwels et al. 2002). Resource flexibility, and in particular workforce flexibility, is critical in being able to handle demand fluctuations (Ebben and Johnson 2005). Firms that are better at these capabilities will be able to better accommodate the demand increase resulting from offering daily deals and more likely to generate positive eWOM. However, these effects will manifest only after consumption, i.e. when the deal is redeemed, raising the open empirical question of pre-consumption versus post-consumption effects of the daily deal.

To address these open questions empirically, we use data collected on restaurants as the focal industry; prior research shows that eWOM is of particular importance in this industry, as are daily deals, making it a suitable context for our study (Lu et al. 2013). Additionally, restaurant online reviews have been shown to reflect elements of quality, customer satisfaction, and referral intention (Chen and Laurie 2013), thereby allowing us to test the contingent effects of daily deals in this context. We conduct our analysis in two stages. In the first stage, we use data on daily deals and online reviews of restaurants in a major metropolitan area in the United States over a 13-month period. Online reviews were collected from Yelp while deals information was acquired from Yipit, a deal aggregator. In addition to restaurants offering deals, we also collected eWOM data on a census of restaurants in the area, thereby providing us with a control set of restaurants that did not offer deals.

We test for heterogeneity in the effect of the deal on the resulting review valence for deal merchants (those who offer deals); in addition, we also test the effect of nearby competing deals on merchants who do not offer deals. Our results do indeed verify an average negative “Groupon effect” reported in the literature (Byers et al 2012a) but also that this effect is significantly moderated by merchant age and price segment. Interestingly, we also see strong evidence of a competitive effect; the presence of competing deals in the neighborhood affects eWOM even for those restaurants that do not offer deals.

The econometric model based on secondary data does not allow us to differentiate whether the resulting moderation is based on redemption of the coupon, or if the deal changed the intrinsic evaluation of the restaurant, i.e. pre-consumption versus post-consumption effects of the deal. In the second stage of our analysis, we address this question through experiments. Building on Raghurir and Corfman (1999) who study the effect of past promotional activity on pre-trial evaluations, we study pre-consumption responses to daily deals using three lab experiments. The results show the presence of a deal effect on brand evaluation even when there is no possibility of consumption; thus, while operational flexibility is critical, we show that the hypothesized “deals” effect on eWOM exists even before the deal is redeemed.

Our study contributes to the literature in multiple ways. First, we go beyond simply documenting a negative link between daily deals and online reviews; we show that these effects can be mitigated for merchants with specific characteristics. Second, we document, for the first time, the unique effects of competition in how daily deals affect brand evaluation and eWOM for both merchants offering deals and those that

choose not to. While competitive effects have been observed in the promotions literature in marketing (Mazumdar et al. 2005), we extend this to the newer context of online daily deals. Third, using a series of experiments, we are able to show how deals, combining elements of advertising and price promotions, can influence brand valuation even before consumption. Specifically, we note the high level of consistency between the results obtained from archival data and those obtained from experiments. We briefly discuss the theory regarding online deals and eWOM next before delving into the research methodology subsequently.

Background and Theory

Online daily deals, also referred to as group-buying deals, social coupons or group discount vouchers (Luo et al. 2014, Kumar and Rajan 2012), are discount coupons posted online through a platform, such as Groupon or LivingSocial. These sites operate double-sided platforms (Parker and Van Alstyne 2005) on which merchants offer deals (on one side of the platform) while individuals buy the deals (on the other side of the platform). The platform appropriately extracts revenues from the merchant side of the platform, while subsidizing the consumer side. Platform owners typically work with merchants to offer deals for a specific period of time (usually 2 weeks). Once purchased, the deal's discounted price can be redeemed over a longer time-period (typically three months after the deal is posted) at the merchant, with some conditions applied on the bundling of coupons or on the specific form of services offered (Dholakia 2010). Once the coupon expires, the consumer can still typically redeem the original dollar value of the coupon without the discount.

Since the launch of online deals in 2008, they have experienced strong growth, with IBIS reporting that between 2009 and 2014, revenues have grown at an annual rate of 147.1%, reaching \$3.4 billion in 2014 (2015). Groupon and LivingSocial are market leaders, totally accounting for roughly 60% of the deals market. In recent years alone, the number of global active deals offered daily has grown from 180,000 in the first quarter of 2014 to 425,000 in the first quarter of 2015 (IBIS 2015). Because of the continued popularity of daily deals with merchants, practitioner news outlets have argued that daily deals are a new and integral part of the online and mobile marketing mix for merchants (Tuten and Ashley 2011, Integreon 2012, Krasnova et al. 2013, Bharadwaj et al. 2013). However, and despite their popularity with customers as well, online deals have remained a contentious subject in practice (Agarwal 2013, Cohan 2012).

Within the academic literature, a small but growing body of work in marketing has focused on the extent to which deals affect firm performance. Specifically, scholars have provided insight into how offering a daily deal affects firm revenues (Dholakia 2010, Dholakia 2011a, Edelman et al. 2011, Dholakia and Kimes 2011, Dholakia 2012, Reiner and Skiera 2013, Shivendu and Zhang 2013). In addition to revenue, Dholakia (2010, 2011a, 2011b) surveyed merchants to investigate the profitability of daily deals. The results, largely consistent with the practitioner press, indicate that only 50% of merchants offering deals actually made any surplus. One of the primary reasons for this is based on the observation that most consumers often do not spend more than the deal value (Dholakia 2011b). However, on the positive side, daily deals have been found to provide many of the immediate

benefits of price promotions observed in the literature (Guadagni and Little 1983, Neslin, Henderson, and Quelch 1985, Blattberg and Neslin 1990) by increasing foot traffic, revenues, and visibility for merchants, at a lower customer acquisition cost (Dholakia 2011a, 2012c). Beyond revenues and margins, how do daily deals influence the focal merchants' electronic word of mouth? We address this specific question next.

Recent work has attempted to quantify the extent to which daily deals may influence a merchant's brand attitudes and eWOM. In a survey of 931 U.S. consumers, Kimes and Dholakia (2011) examine consumer responses to daily deals and find there is no loss in brand equity for merchants offering deals. In fact, the authors comment: "To the contrary, respondents offered favorable comments about the restaurant and their dining experience" (p. 18). These authors also find that consumers are aware of the deals offered by multiple daily deal sites, with 50% or more of consumers reporting awareness of deals from Groupon, restaurant.com, LivingSocial, BuyWithMe, and TravelZoo. Alternatively, Byers et al. (2012a, 2012b) focus on a single platform, Groupon, and report that offering a Groupon leads to an average 0.2 star rating decrease in Yelp¹⁰. Their study follows approximately 5,000 Groupons from many categories in the U.S, which were then linked to the online reviews for that merchant on Yelp. While Kimes and Dholakia (2011), by virtue of their design, do not account for the heterogeneity among merchants who offer deals, Byers et al. (2012a) include all retailers offering Groupon deals without accounting again for the heterogeneity of deal merchants. We argue that there is likely a middle

¹⁰ The authors refer to this as the "Groupon effect" in their analysis (Byers et al. 2012a).

ground here driven by heterogeneity of the merchants offering deals, and their competitive environments. Accounting for this heterogeneity will deliver, arguably, a more nuanced view of the effect of daily deals, suggesting that the negative or null effect from deals may not extend to all merchants. We consider the influence of these sources of heterogeneity below.

As a first step, we note that daily deals represent a new online mixture of price promotions and opt-in advertisements (Edelman et al. 2011). Daily deal subscribers typically need to sign up with the platform (such as Groupon or LivingSocial) to access the deals and receive email messages with the equivalent of advertising copy for the merchants, which typically contain positively framed images and endorsements; these are akin to advertisements for the focal merchant (Shivendu and Zhang 2013). Additionally, daily deals represent price promotions, since they offer deep price discounts. We use these two aspects of daily deals to theoretically motivate the moderating effects of heterogeneity and competition in their effects on eWOM.

Prior work has established that consumer expectations and attitudes are shaped by pricing strategies, more specifically, the promotional activities used by merchants (Mazumdar et al. 2005, Lynch and Ariely 2000, Yoon et al. 2014, Lee and Tsai 2015). Research has established that higher prices are associated with higher quality evaluations and more positive eWOM, all else being equal (McGregor et al. 2007, Li and Hitt 2010). We are, however, interested in how the *marginal* effect of the daily deal on the resulting eWOM may change depending on the merchant's characteristics, or more specifically, its price segment. Research studying the effects of price promotions on consumer attitudes and brand evaluations shows that the effect

is driven largely by the discounted dollar amount (Blattberg et al. 1995). Winer (1986) first proposed that the larger the difference between the initial price and the purchase price, the greater the resulting consumer utility and the more positive the consumer attitude towards the merchant or product. He argued that this effect is predicated on the discounted dollar value rather than the discount rate per se; this effect has been empirically observed in various other product contexts as well (Della Bitta et al 1981, Grewal et al. 1998, Wu et al. 2004, Biswas et al. 2013).

Following this logic in the daily deals context, we argue that deep discounts offered by merchants in the low-price segment are likely to be viewed more negatively, given that the resulting dollar value of the discount is small. Additionally, if lower prices are already associated with lower quality, further price discounts in this context are likely to be viewed even more negatively by consumers. Alternatively, deals offered by high-price segment merchants start from an expectation of high quality (Li and Hitt 2010) and also offer consumers greater discounts in absolute value. Thus, merchants in the high-price segment may not experience the same negative effect on their eWOM as merchants operating in a lower-price segment. In fact, strong associations made between price and qualities, which benefit high-price merchants, may even lead to a positive response on eWOM for such merchants. These arguments suggest that the marginal effects of daily deals on ex post eWOM are likely to be significantly negative for low-price merchants (consistent with Byers et al. (2012a)) but not so for merchants in the premium segment.

To the extent that daily deals share features of opt-in advertisements, the persuasion knowledge model proposed by Friestad and Wright (1994) suggests that under certain conditions, the motivations behind such promotional efforts are perceived as appropriate and reflect confidence in the product or service advertised. However, in other contexts, promotional or advertising efforts may smack of desperation and produce negative brand evaluations (Kirmani 1997, Chen and Kirmani 2015). Extending this argument to the daily deals context suggests that merchants providing daily deals under conditions that reflect confidence may not receive negative eWOM but may actually benefit from such daily deals. One such contingency in which it is viewed as legitimate to offer price promotions and invest in advertising effort is when the business is new (McDougall and Robinson 1990, Carter et al. 1994); anecdotal evidence also suggests that new businesses may find it advantageous to offer daily deals to recruit new customers (BizJournals 2013). In contrast, established merchants offering deals may lead to perceptions of weakness and desperation. Whether the merchant is truly in such a state may not matter; if offering a daily deal leads to some subset of consumers to draw such conclusions, brand perceptions are likely to suffer, leading to a higher probability of negative eWOM ex post. We thus expect that the age of the merchant to influence the extent to which daily deals affect eWOM, leading to a testable proposition.

Beyond sources of merchant heterogeneity, merchants do not operate in a vacuum, but within a competitive landscape that typically includes other merchants and their competitive actions, which cumulatively determines the options available to consumers. This competitive environment, in turn, contributes to setting reference

prices for the merchant segment (Mazumdar et al. 2005, Bell and Lattin 2000). Moreover, it is well accepted that price changes affect the demand for other products, through cross-price elasticity (Sethuraman et al. 1999). Frequent price promotions on a segment within a geographical area can also lead to lower reference prices for that category, which affects price perceptions of all merchants in that category (Mayew and Winer 1992). Thus, if nearby competitors offer frequent daily deals, the reference price for similar products or services should shift downward along with consumer quality expectations, and the resulting eWOM, on average (Li and Hitt 2010). However, it is likely that those merchants that offer deals, as a response of high deal competition, will not experience a negative deal effect as they are in fact matching prices with its competitors (Raghubir and Corfman 1995, 1999). The eWOM literature has also documented how the competitive environment affects eWOM (Forman et al. 2009, Li et al. 2011, Jabr and Zheng 2013, Kwark et al. 2014). This reasoning brings into sharp focus the possibility that nearby deals might also negatively affect merchants who *never* offered deals, leading to a perverse and unexplored effect of daily deals within a competitive market. We thus test for the effect on a focal merchant's eWOM when its competitors offer daily deals but it does not.

Beyond these theoretical mechanisms, and from an operations management perspective, some firms are better equipped to react to the demand fluctuations created by online deals. Previous work in the price-promotions and marketing literatures has shown that promotions can have a sudden and unpredictable effect on demand (Blattberg et al. 1995, Pauwels et al. 2002, Alvarez and Vázquez Casielles

2005). Moreover, it is well established that resource flexibility, and in particular workforce flexibility, are important firm attributes to handle demand changes (Paul and Jonathan 1991). For example, Ebben and Johnson (2005) show that flexible firms have very specific characteristics and that not all firms are able to achieve operational flexibility. Thus, it is plausible that certain merchants will be able to successfully accommodate the increased demand that results from online deals, while others will not. The practitioner press contains many such anecdotes, with a business owner recently stating: “We had thousands of orders pouring in that really we hadn't expected to have” (BBC 2011, 2012). However, these effects should only be observable after the deal is redeemed and not necessarily before consumption. This highlights that there may be pre-consumption as well as post-consumption effects of daily deals on eWOM. While these differences are not directly observable using archival data, we address this question in more detail later in the paper.

In summary, there is theoretical and anecdotal support for the notion that the effects of daily deals on eWOM are likely to be moderated by sources of merchant heterogeneity (price segment and merchant age) as well as competition. Given the multiple possible mechanisms for the effects of daily deals, rather than provide formal hypotheses, we allow the analysis to provide us with guidance. We next detail the data and methodology used to test for these effects.

Research Methodology

Data

We focus on online reviews and online deals for restaurants in a large U.S. metropolitan area, Washington, D.C. Prior work in online reviews (Mangold et al. 1999, Gu et al. 2012) suggests that restaurants provide a suitable context for studying eWOM, given the high-involvement nature of food. Existing research studying daily deals has also focused on services, particularly restaurants, to understand their appeal within this sector (Farahat et al. 2012). The online reviews for the restaurants in our sample were collected from Yelp.com, which has published over 14 million online reviews for restaurants, receives over 135 million monthly visitors, and is the market leader in North American online reviews (Yelp 2015). For the purposes of this study, we collected data on 2,012 restaurants operating in Washington, D.C., roughly comparable in volume to the 2,035 operating restaurants reported by the National Restaurant Association in D.C. for 2012, the focal year of our data collection¹¹. Each Yelp restaurant listing contains general information on restaurant characteristics, such as location, cuisine, price segment and ambience, and online review information, such as the average rating and number of reviews. Furthermore, we collected each individual review published for each restaurant, which resulted in 143,745 reviews between 2004 (Yelp's initial release) and 2012. Each review contains a numerical rating, text comments, and timestamp.

¹¹ In order to reduce unobservable heterogeneity in our sample, we do not include restaurants in Northern Virginia or Maryland in our sample. Restaurants outside Washington, D.C., operate under different licensing, tax, and regulatory regimes, adding further complexities to our model that are outside the purview of this paper.

We match this dataset with data provided by Yipit.com, a service provider aggregating deal data across multiple daily deal platforms, for a 13-month period between December 2011 and December 2012. We chose Yipit.com because it aggregates transaction data from over 97% of daily deal sites (Yipit.com 2015). Unlike prior research focused on measuring the effect a single daily deals vendor (e.g. Byers et al. 2012a), we observe deal offers from 31 vendors, including LivingSocial, Groupon, Google Offers, Yelp Deals, and other smaller vendors. In total we observe 2,425 deals corresponding to 935 restaurants in Washington, D.C. Each deal listing contains information about the merchant, such as phone number, name and geographical location; deal characteristics, such as price, discount and duration; and deal performance metrics, such as quantity sold and revenues generated from the deal. However, this dataset does not provide information on deal redemption. Figure 3.1 displays a typical daily deals timeline in our dataset, where the deal is sold for two weeks and can be redeemed for approximately 12 weeks after the sale period.

These two data sources—Yelp and Yipit—form the core of our empirical data collection strategy. To aggregate these into a single panel data set, we first summarize the review and deal data for each restaurant into two-week periods. Thus, our unit of analysis is restaurant-period¹². This results in an unbalanced panel data set, since not all restaurants have reviews in every period. As our primary interest is modeling the deal-eWOM relationship, we discard any periods beginning before the start date of our deals data set. Thus, our panel contains aggregated information for 28 two-week time periods covering 13 months (from December 1, 2011, to December

¹² In robustness tests, we also perform our analyses with three-week and four-week periods, with consistent results.

31, 2012). After aggregating and summarizing the dataset into two-week periods, we have 19,691 restaurant-period observations in the unbalanced panel. Figure 3.2 displays the mean rating of reviews published in Yelp 100 days before and after a daily deal is initially offered for those restaurants that offer deals. Consistent with the results of Byers et al. (2012a), we observe a discontinuity in Yelp ratings before and after offering a deal in our dataset.

Of the initial 2,012 restaurants identified on Yelp, 1,390 had Yelp reviews published during this time period. Of these, 922 offered at least one deal during this time period, while the remaining restaurants did not. To confirm that the latter is an appropriate control group, we tracked the text of the reviews available for these restaurants back to 2004 (the earliest available dates on Yelp) for keywords indicating possible deals, such as *deal*, *Groupon*, *online deal*, and *coupon*,¹³ which is consistent with the approach used by Byers et al. (2012b). We found that fewer than 10% of these restaurants had one or more deal keywords from 2004 to 2012, and none had deal keywords during our study period or the preceding six months. This suggests that most of the restaurants in this group rarely or never offer deals, thereby forming an ideal “control” group. Conversely, for restaurants that do offer deals, we find deal keywords in 98% of the reviews during the corresponding deal periods.

Our dependent variable of interest, $Rating_{it}$, is the average numerical rating of the reviews for restaurant i published in time period t . The primary independent variable of interest is $Deals_{it}$, which is equal to 1 if restaurant i is engaging in one or more online deals during time period t and 0 otherwise. A deal is considered offered

¹³ A full set of keywords used to search for daily deals is available upon request from the authors.

during time period t if that period overlaps with the deal sales period. We model contemporaneous effects on eWOM because brand evaluations are likely formed at the point of deal offers. Additionally, Yipit reports that over 40% of the deals bought are redeemed in the first two weeks; beyond this two-week period, additional confounding may influence eWOM, making identification harder.

Since we are interested in evaluating the deal effect contingent on restaurant heterogeneity and competition, we define these variables as follows. Sources of restaurant heterogeneity, such as price segment and restaurant age, are collected from the Yelp restaurant listing. Deal intensity within the competitive environment is captured by $DealsInZip_{it}$, equal to the number of deals being offered by competitors of restaurant i during time period t , where a competitor is defined as any other restaurant in the same geographical area that has the same cuisine type and price segment. Finally, we control for all other observable characteristics of the restaurant in Yelp, such as location, cuisine, ambience, and noise level. A full description of the variables used in our models can be found in Table 3.1. Summary statistics and a correlation table for the resulting panel data set can be seen in Table 3.2 and 3.3, respectively. We now describe our econometric model.

Empirical Model

We model the effect of deals offered by a merchant and its nearby competitors on the ratings of that merchant over time, accounting for sources of heterogeneity and competition. We employ a hierarchical model to capture these effects on the merchant's eWOM as follows. First, to capture heterogeneity of longitudinal

dynamics across restaurants, we allow the effect of our variables of interest to vary across restaurant. As such, the first-level model in the hierarchy is:

$$Rating_{it} = b_{i_0} + \beta_0 + \beta_{1i}Deal_{it} + \beta_{2i}DealsInZip_{it} + \beta_3RestInZip_i + \beta_4BaseRating_i + \beta_5BaseNumReviews_i + \epsilon_{it}$$

where i indexes restaurants and t indexes time periods. Second, we argue that deal and competition effects might systematically vary based on restaurant characteristics. Hence, in the second-level model in our hierarchy, we regress each subject-level coefficient in the first level on all the observable characteristics of the restaurant captured in Yelp. Thus the second-level model is:

$$\beta_{ji} = \gamma_0 + \gamma_{j1}Price_i + \gamma_{j2}Age_i + \sum_{p=1}^{12} \gamma_{j3}^{(p)} Loc_i^{(p)} + \sum_{q=1}^{16} \gamma_{j4}^{(q)} Cuisine_i^{(q)} + \sum_{r=1}^{15} \gamma_{j5}^{(r)} Char_i^{(r)} + \epsilon_{ji},$$

where $j \in \{1, 2\}$ indexes the predictors in the first level (i.e. $Deal_{it}$, $DealsInZip_{it}$), p indexes locations (zipcodes), q indexes cuisine types, and r indexes other restaurant characteristics (e.g. ambience, noise level, parking options). $Loc_i^{(p)}$, $Cuisine_i^{(q)}$ and $Char_i^{(r)}$ are binary variables.

We model this specification using a Hierarchical Bayes model (Rossi and Allenby 2003, Gelman et al. 2014) to account for the observable and unobservable heterogeneity of the merchants. Hierarchical Bayes (HB) models have been highly popular as a tool to model multi-faceted, non-linear phenomena (Rossi et al. 2012). Bayesian methods are particularly appropriate to the decisions modeled in marketing problems, where there are many units of analysis (e.g. customers or sites), each with multiple observations, and there is a desire to account for individual differences (Rossi and Allenby 2003) compared to OLS regression models. Furthermore,

previous work both in marketing and IS have used Bayesian methods to study the dynamics and effects of online reviews (e.g. Zhao et al. 2013, Trusov et al. 2010, Moe and Trusov 2011). For example, both Dellarocas et al. (2007) and Dickinger and Mazanec (2008) analyze how online reviews affect firm performance using hierarchical models. Similarly, Zhou and Duan (2010) model the impact of user reviews and professional reviews in the context of software downloads using a Bayesian framework. Thus, there is a significant body of work in the extant literature supporting the use of Bayesian methods to model the effects of online reviews and we base our Bayesian analyses on these established methods.

Model Estimation and Results

We estimate the specified model using a Bayesian Markov Chain Monte Carlo (MCMC) sampling methodology with standard conjugate diffuse priors. Starting values were taken from the maximum likelihood parameter estimates from independent linear models estimated on the same dataset. The MCMC chain was run for 10,000 iterations including an initial burn-in period of 1,200 iterations, and the chain achieved convergence quickly. The posterior distributions of the coefficients of 8,800 draws were extracted and analyzed.

Table 3.4 summarizes the posterior distributions of the model coefficients for the proposed HB specification. Below, we report the posterior mean for each coefficient of interest, followed by the 95% highest posterior density (HPD) interval. In agreement with previous work (Byers et al. 2012a), we find a negative effect of offering an online deal. In particular, we find that offering a deal results in a decrease in mean rating of 0.902 [0.672, 1.436] during the same period. However, we also

find two strong moderators to this effect: price segment and restaurant age. Restaurants with a one-level higher price segment experience a reduction of the “deal effect” by 0.539 [0.211, 0.567]. Furthermore, younger restaurants are less negatively affected by deals, with a one standard deviation reduction in age implying a reduction in the deal effect of 0.665 [0.333, 0.748]. Thus, premium restaurants, as well as new ventures, experience less negative fallout from the offering of a deal.

However, how does the presence of daily deals within the competition affect the restaurant? The results from the HB model show a significant negative effect of deal competition on the average rating for all restaurants. For every deal offered by a proximal competitor, we find a decrease in mean rating of 0.235 [0.152, 0.368] during the same period. Surprisingly, this result extends even to those restaurants that rarely or never offer deals; the presence of deals in their neighborhood negatively affects their ratings as well. This effect is striking, since the offering of a deal by a neighboring restaurant should have no correlation with the focal restaurant’s eWOM. Yet, the results show evidence of a negative externality imposed by deal intensity in a market segment even on non-deal restaurants. Furthermore, we do not observe any significant moderation of this “deal competition effect” by characteristics of the focal restaurant – this effect is entirely based on the presence of deals in proximal competition, defined as restaurants in the same price segment, cuisine and geographical area. The fact that most restaurants in the control set have little or no deal activity (either through the Yipit dataset or through our keyword analysis of their existing reviews) suggests that these effects are based on spillovers from proximal “deal” restaurants. While this spillover effect of offering deals has been hinted at in

the practitioner press¹⁴, here we present empirical evidence of this effect for the first time to the best of our knowledge.

Discussion

Our empirical analysis of restaurant reviews and online deals demonstrates that offering a deal has an overall negative effect on the reviews arriving within the two-week period during which the deal was offered, consistent with Byers et al. (2012a). However, we add further nuance to this broad result by arguing and providing evidence for the influence of moderators of this effect. Specifically, we identify the price segment and age of the restaurant as moderators, suggesting systematically weaker negative effects for premium and new restaurants. Most notably, we also find evidence for a deal competition effect. That is, all merchants (even those who never offer deals) are negatively affected by nearby competing merchants offering deals, showing that the effects of daily deals are not limited to participating merchants but also create spillover effects that impact other merchants.

These conclusions are based on a large dataset and an unconstrained modeling framework that allows coefficient estimates to vary by merchant and controls for a range of factors that may affect the rating of the restaurant. However, there may be a possibility that our findings are somewhat influenced by unobserved variables, selection issues, and/or reverse causality. Reverse causality may be at play if merchants exhibiting poor performance (beyond what we observe in their online reviews) due to some unobservable factors, seek to influence their short-term online ratings by offering a deal (Farahat et al. 2012). In the presence of such an omitted

¹⁴ <http://www.cnbc.com/id/49092709>

variable, our estimates of the deal effect would be biased. However, we do not believe this to be likely here, since online deals in large cities typically go live several months after the terms of the deal are finalized (Groupon 2015b, LivingSocial 2015, GrouponWorks 2014, Zabranova 2012). In other words, the gap in time between the restaurant opting for a deal and the eventual offering of the deal is often several months, which is significantly longer than the two-week periods (or even a four-week periods, which provided similar results in robustness tests) we model. Therefore, the effects of reverse causality, in terms of previously unobserved lower ratings driving the decision to offer a daily deal, are unlikely here given the design of the analysis.

One limitation of the negative “deal effect” identified by our empirical model is that we are unable to distinguish between negative eWOM due to a decrease in performance of the merchant during the deal period versus negative eWOM arising primarily from reduced quality expectations from a restaurant that is observed to offer a daily deal. That is, the negative response could be due to the impression that the focal restaurant offering a deal is “in distress”, thereby leading to lower reviews even amongst existing customers or those without coupons. Effectively, it is not clear if the actual consumption and redemption of a deal is needed for lowered brand evaluations and online ratings, or if reviewers respond to the fact that a restaurant is offering a daily deal and hence reduce their quality expectations.

To address these issues, we complement our econometric analyses with three lab experiments, specifically designed to tease out these confounding effects and provide cleaner identification. In the first two experiments, we test for the effect of offering an online deal on consumer quality expectations without the possibility of

actual deal redemption to test for the moderating influence of price segment (Experiment 1) and restaurant age (Experiment 2). We follow these up with Lab Experiment 3, in which we test for the deal competition effect. We describe these experiments in detail next.

Lab Experiments: Testing the Pre-Consumption Effects of Daily Deals

Experiment 1: Deals and Merchant Price

Procedure, Data and Measures

Experiment 1 tests whether consumers' online evaluations of a merchant's services are affected by online deals and whether the price segment of the merchant moderates this effect. Evidence of this effect was observed in the Bayesian analyses reported earlier. For the purposes of the experiment, 400 U.S. respondents (191 women) between the ages of 18 and 63 ($\text{Mean}_{\text{age}} = 35$) from Amazon's Mechanical Turk (MTurk) were recruited for pay (\$0.35) in this study. This pool of participants has been shown to be reliable for experimental research (Goodman, Cryder, and Cheema 2013), to represent the broader population (Buhrmester, Kwang, and Gosling 2011), and to generate high quality results (Ipeirotis et al. 2010). Further, we selected only workers with a Hit Approval Rate (HIT), or the rate of completed jobs that are approved by a worker, of 95% or above. Respondents were randomly assigned to one of four 2 (deal offered: yes vs. no) x 2 (price segment: high vs. low) between-subjects conditions.

Participants were first asked to review information about a restaurant and then assess the quality of the restaurant. We presented the aggregate information for the restaurant as displayed in Yelp. To avoid providing any biasing cues, we withheld any review text in the description of the restaurant. In all four (deal X price) conditions, the restaurant was named “Italian Kitchen” and given a fictional address to control for possible familiarity with an actual restaurant. We were also careful to use the user interface of Yelp.com by using the exact same fonts and colors (shown in Figure 3.3). The stimuli were developed by selecting a restaurant from our Yelp data set with the most popular cuisine (*Italian*), an average number of reviews (65), and an average rating (4.0). Deal offered was manipulated by showing that the restaurant was offering a 50% off discount. This offer was not presented in the no-deal condition. Further, to create the high and low price conditions, we set the price segment of the restaurant to low (\$) and high (\$\$\$\$).

Brand evaluation was measured as a composite of purchase intention and perceived quality ($r=0.87$). We adapted purchase intention from Jamieson and Bass (1989): “If you were thinking about going to an Italian restaurant, how likely would you be to visit this restaurant?” (1 = “very unlikely” to 7 = “very likely”). Measures for perceived quality were adapted from Kirmani and Wright (1989): “Given the information provided about this restaurant, please rate the likely overall quality of this restaurant” (1 = “very low” to 7 = “very high”). Finally, as a manipulation checks of pricing we asked respondents to rate the restaurant along two dimensions: price expectations (1 = “low-priced” to 7 = “high-priced”) and cost (1 = “cheap” to 7 =

“expensive”). We expect both the deal and price manipulations to affect these measures in this experiment.

Results and Discussion

The manipulation checks confirmed that participants in the no-deal condition (Mean = 4.56) find the restaurant higher priced than participants in the deal condition (Mean = 2.67; $F(1, 396) = 4.69, p < 0.05$). Likewise, participants in the high price-segment condition (Mean = 6.18) find the restaurants higher priced than participants in the low price-segment condition (Mean = 1.79; $F(1, 396) = 4.2, p < 0.05$). Moreover, we find similar results with the cost manipulation check.

More importantly, and following the results of HB model where price acts as a moderator of the deal effect on online ratings, we proceed to test for a significant deal X price interaction using ANOVA. We find a significant deal X price segment interaction effect ($F(1,396) = 6.31, p < 0.01$; see Figure 3.4). A first set of planned contrasts show that for non-deal restaurants, a higher price segment had no significant effect on brand evaluation (Mean_{no deals-low} = 5.14 vs. Mean_{no deals-high} = 5.28; $F(1,396) = 2.12, p = 0.33$). For deal-offering restaurants, however, having a higher price segment significantly increased brand evaluations (Mean_{deal-low} = 3.94 vs. Mean_{deal-high} = 5.6; $F(1,396) = 3.59, p < 0.05$). A different set of planned contrasts show that for restaurants associated with a lower price segment, there is a significant decrease in brand evaluation when a deal is offered (Mean_{low-non deal} = 5.14 vs. Mean_{low-deal} = 3.94; $F(1,396) = 4.11, p < 0.05$). However, this effect becomes marginally significant and positive for restaurants with a high price (Mean_{high-non deal} = 5.28 vs. Mean_{high-deal} = 5.6; $F(1,396) = 3.71, p = 0.07$).

Therefore, in a controlled setting without the possibility of consumption, unlike in our empirical model, we observe that the price segment of the restaurant moderates the negative effect of the deal on brand evaluations. These results suggest that even before visiting the restaurant and experiencing the service provided, there is a decrease in brand evaluations for certain merchants who offer online deals. This finding adds further evidence that certain merchants are more likely to be perceived as potentially “distressed”, regardless of their true financial status (in our case the low-priced merchants, representing the non-premium segment) whereas other merchants will be perceived as “confident” (in our case the high-priced or premium merchants), as first suggested by Kirmani (1990). Beyond the price segment of the restaurant, is it possible that daily deals offered by new restaurants are viewed less negatively? We explore this contrast in the next experiment.

Experiment 2: Deals and Merchant Age

Procedure, Data and Measures

Experiment 2 tests whether consumers’ brand evaluations are affected by the offering of a daily deal and whether the newness (age) of the restaurant moderates this effect. 398 respondents (175 women; ages 18-72, $Mean_{age} = 38$) from MTurk participated for pay (\$0.35) in this study. Respondents were randomly assigned to one of four 2 (deal offered: yes vs. no) X 2 (new restaurant: yes vs. no) between subjects conditions. The procedure and stimuli were identical to Experiment 1, except that for the new condition, we added a banner showing that the restaurant opened recently using existing Yelp’s user interface, as seen in Figure 3.3. As in the previous study, participants were first asked to review the information in the

simulated webpage for the restaurant and then assess the quality of the restaurant. We again performed manipulation checks on the pricing and cost of the restaurant. As in the first study, we measure brand evaluation, which is a composite of purchase intention and perceived quality ($r = 0.88$).

Results and Discussion

The manipulation checks confirmed that participants in the no-deal condition (Mean = 4.50) find the restaurant higher priced than participants in the deal condition (Mean = 2.56; $F(1, 394) = 4.61, p < 0.05$) with similar results with the cost manipulation check. More importantly, and reflecting the moderating role of age on the deals effect in our HB model, we find a significant deal X age interaction ($F(1, 394) = 6.44, p < 0.02$; see Figure 3.5). A first set of planned contrasts shows that for non-deal restaurants, being new had no significant effect on brand evaluation (Mean_{no deal-established} = 5.84 vs. Mean_{no deal-new} = 5.17; $F(1, 398) = 1.15, p = 0.28$). For deal-offering restaurants, however, being new significantly increased brand evaluations (Mean_{deal-established} = 4.03 vs. Mean_{deal-new} = 6.13; $F(1, 394) = 4.59, p < 0.05$). A different set of planned contrasts show that for already established restaurants, there is a significant decrease in behavioral intentions when a deal is offered (Mean_{established-no deal} = 5.84 vs. Mean_{established-deal} = 4.03; $F(1, 394) = 5.32, p < 0.05$). However, this effect becomes marginally significant and positive for newly established restaurants (Mean_{new-no-deal} = 5.17 vs. Mean_{new-deal} = 6.13; $F(1, 394) = 3.40, p = 0.06$).

Therefore we observe that the newness of the restaurant does indeed moderate the negative effect of the deal in a controlled experimental setting, in agreement with our empirical results. These results also add credence to the notion

that brand evaluations (perceived quality and purchase intentions) of established restaurants decrease even before consumption simply by offering an online deal. By contrast, new restaurants are expected to offer daily deals as a way to incentivize new consumers to take a chance on the merchant (Dholakia 2012); thus, offering a deal does not reduce the perceptions of quality in such cases. We test for the effect of deal competition next.

Experiment 3: Deals and Deal Competition

Procedure, Data and Measures

Experiment 3 tests whether consumers' brand evaluations are affected if the focal merchant offers an online deal and if nearby competitors also offer online deals. 404 respondents (187 women; ages 21-70, Mean_{age} = 41) from MTurk participated for pay (\$0.35) in this study. Respondents were randomly assigned to one of four 2 (deal offered: yes vs. no) X 2 (deal competition: high vs. none) between subjects conditions. The procedure and stimuli were identical to study 1 and 2, except that for the deal competition condition, we added a measure of the degree of deal competition for similar restaurants nearby. More specifically, a restaurant with high nearby competition a graphic that showed, consistent with Yelp's user interface: "There are 20 deals for similar restaurants in this area" as seen in Figure 3.3. As before, participants were asked to first read information about the restaurant and then assess the quality of the restaurant followed by a manipulation check on the pricing and cost of the restaurant. The measure are the same as in Experiment 1 and 2, brand evaluations, which is a composite of purchase intention and perceived quality ($r = 0.91$).

Results and Discussion

The manipulation checks confirmed that participants in the no-deal condition (Mean = 4.46) find the restaurant higher priced than participants in the deal condition (Mean = 2.11; $F(1, 400) = 4.18, p < 0.05$) with similar results with the cost manipulation check. More importantly, and reflecting the role of competition in our HB model, we find a significant deal x deal competition interaction ($F(1,400) = 6.74, p < 0.01$; see Figure 3.6). A first set of planned contrast show that for non-deal restaurants, having a high deal competition had a significant negative effect on brand evaluation (Mean_{no deal-no deal competition} = 5.78 Vs. Mean_{no deal-deal competition} = 4.19; $F(1, 400) = 5.92, p < 0.02$). For deal offering restaurants, however, high deal competition marginally increased brand evaluations (Mean_{deal-no deal competition} = 3.78 vs. Mean_{deal-deal competition} = 4.33; $F(1,400) = 4.59, p < 0.07$). A different set of planned contrasts show that for restaurants with no nearby deal competition, there is a significant decrease in behavioral intentions when a deal is offered (Mean_{no deal competition-no deal} = 5.78 vs. Mean_{no deal competition-deal} = 3.78; $F(1,400) = 4.92, p < 0.05$). However, this effect is not significant for restaurants with high nearby deal competition (Mean_{deal competition-no deal} = 4.19 vs. Mean_{deal competition-deal} = 4.33; $F(1,400) = 1.46, p = 0.22$).

Therefore we observe evidence of a deal competition effect, reflecting our empirical model results. That is, even merchants who do not offer deals are affected by nearby competitors offering online deals. In this study, however, we go beyond this finding and show that for merchants without nearby deal competition, offering a deal would lead to a significant decrease in brand evaluations. However, for merchants with high deal competition, we do not find evidence of any change in

brand evaluations as a result of offering a deal. This is in line with previous work suggesting that when the causes of promotions are attributable to external factors (i.e. high deal competition), there is no negative effect on the brand from offering a promotion (Raghubir and Corfman 1995, 1999). Our results show that in environments with high deal intensity, the negative effects associated with offering a deal are attenuated, as offering a deal here is viewed as standard practice with little or no penalty imposed on the merchant. In summary, we observe that the results from the econometric model using archival data, regarding the contingent effects of deals on eWOM, were fully supported by the three lab experiments presented here.

Discussion and Managerial Implications

The revenues from online deals are expected to climb to \$5.5 billion in 2016 according to industry analysts (BIA Kelsey 2014). However, these figures notwithstanding, offering daily deals in the service industry raises several questions about their effects on merchants to both practitioners and researchers. One article from CNBC proclaims “Groupon isn’t a Good Deal for Businesses”, while another one from the same website states, without irony or explanation, “Groupon is Good for Business”¹⁵. While directly linking daily deals to revenues or sales is challenging, we aim to shed light on how online deals and the competitive deal landscape affect consumer perceptions. We focus our work on the restaurant sector, where daily deals are highly popular, and evaluate the effect of deals on the resulting eWOM, captured through Yelp review valence. In a longitudinal analysis of restaurants reviews from

¹⁵ www.cnbc.com/id/49092709, www.cnbc.com/id/49092710

Yelp.com and a data set covering the population of online deals offered by the same restaurants in a major metropolitan area in the U.S., we find evidence that restaurants' short-term ex post ratings are generally negatively affected by offering a deal in line with previous work (Byers et al. 2012a).

However, while the average effect of a deal is indeed negative, the individual effect for a restaurant depends on its characteristics and the extent to which deals are common in the competitive environment. Indeed, in specific contingencies, we show that the effect of the daily deal on eWOM may actually be positive. Restaurants operating in a higher priced segment experience responses to daily deals that appear considerably less negative and even positive, suggesting that these restaurants are viewed as offering deals from a position of confidence. Similarly, new restaurants are not penalized as much for offering daily deals as a way to recruit new customers, while established restaurants are more likely to be viewed as distressed when they offer deals (Friestad and Wright 1994). Finally, we show that deal intensity in the competition has a significant impact on eWOM; if deals appear to be part of the marketing mix within the competition, the negative effect on eWOM is muted for the focal restaurant. Surprisingly, we also observe that restaurants that do not offer deals are also affected by deal intensity in the competitive landscape, providing evidence of a clear spillover effect.

Beyond our econometric results, we also attempted to replicate these findings in a controlled setting. In Lab Studies 1 and 2, we find consistent support for the overall average negative effect of daily deals on brand evaluation, as well as a significant moderation effect of price segment and age. Indeed, our results reflect

Kirmani and Wright's (1989) assertion that not all advertising and promotions are viewed positively. They write: "many people spontaneously assume high advertising expense implies managerial confidence and high quality unless [...] desperation undermine is salient to them." Further, we are able to establish that the negative effects on eWOM from daily deals are not solely attributable to poor performance of the restaurant during redemption, as has been discussed in the press. Indeed, we find that even before there is any product or service consumption (as is the case with our lab subjects), the restaurant's brand evaluation suffers. Furthermore, in Experiment 3, we replicate our empirical model's finding that nearby competitors offering online deals are also affecting the rating of merchants.

Taken together, our results present a more complete view of how the recent and highly popular phenomenon of online daily deals affects consumer brand evaluations, leading to several theoretical contributions to the literature. First, building on early work on daily deals by Kimes and Dholakia (2011) suggesting that daily deals do not affect brand equity and Byers et al. (2012a) suggesting that online deals decrease online ratings, we provide specific conditions under which online deals affect online reviews and brand evaluations. The nuances in how deals affect consumer perceptions are important in being able to assess the true value of such programs. Second, our work extends the work of Friestad and Wright (1984) and Kirmani and Wright (1989) in understanding how consumers interpret a marketer's efforts, motives, and tactics. We show that in the current information-rich environment from online sources, consumers do form quality attributions based on the cues found in online reviews, online deals, and the particular conditions

characterizing each merchant. In fact, since consumers are keenly aware of the deep discounts found in online deals, they may be using information about the marketing campaigns of the merchant over time as indicators of the risks taken by the merchant to acquire customers (i.e. customer acquisition cost).

We also report an intriguing spillover effect from daily deals to proximal restaurants that do not offer any deals but see their brand evaluations decrease. These results suggest that consumers do not interpret marketing efforts in a vacuum, especially deep price discounts of the sort offered by daily deal firms. In fact, by examining the promotional actions of nearby competitors, our work suggests that the consumers “schemer schema” for interpreting, evaluating, and responding to influence tactics from merchants, which was first proposed by Wright (1986), is more complex than previously thought of and includes not only the tactics of the focal merchant but the tactics used in relation with the competitive environment. Moreover, as Bharadwaj et al. (2013) argues, many firms have tried to use social media and other marketing tools in isolation. Our work suggests that an effective *digital business strategy* should take into account how online deals affect the social media standing of merchants.

To our knowledge, we are among the first to outline the conditions under which market-specific quality determinants, such as the presence of proximal daily deals within the competition, affect an individual’s quality attributions. One possible explanation for the deal competitions effect is that a higher number of nearby competitors offering deals leads to a reduction in reference prices for the restaurant’s services. That is, a drastic reduction in the price of some services might affect quality

attributions for all other merchant in that market. This is important since it suggests that online review or electronic word of mouth for firms or merchants are likely to be affected by the actions of others, a form of cross-market elasticity that has not been addressed in the literature. Finally, we also contribute to the price promotions literature by suggesting that there are actually two layers in the effect promotions have on brand evaluations –a pre-consumption effect and a post-consumption effect. We are among the first to show that such pre-consumption effects are not only possible but also actually salient and likely reflect a combined influence of media, anecdotal evidence and offline word of mouth.

Apart from these contributions, there are significant implications from our work for practice, especially for merchants and platform owners. While there has been significant recent media attention to the failures of daily deals merchants, we find conditions under which daily deals can be beneficial in terms of online reputation to merchants, which is a key question for both merchants and platform owners. A direct implication of our work for merchants offering daily deals or considering them is to more fully examine how their customers may perceive their motivations for offering such deep discounts and to understand their competitive environment as well. Established merchants offering daily deals may consider investing in alternative signals of quality or munificence that could help offset the perceptions of distress that may be conveyed involuntarily. Alternatively, these merchants could choose lower price discounts so as to allow modest demand increases without the commensurate negative brand perceptions. Similarly, restaurants operating in markets where deals are common should consider the effects of not offering deals; if their online

reputations suffer through the spillover effect, there might indeed be value in simply offering deals themselves. In either case, our results provide some guidelines for when and why merchants may opt to offer daily deals.

For platform owners, such as Google and LivingSocial our work also raises many pertinent issues. First, since consumers appear to be interpreting online deals either as a signal of high confidence or desperation, the daily deals platform might want to highlight the cues signaling high confidence, through the selection of the merchant or the structuring of the promotion itself. Additionally, we believe that deal platforms could actually benefit from using and reporting the information available on the merchant in online reviews platforms, such as Yelp.com and Foursquare.com. For example, if some reviewers mention the words “deals, Groupon, or Living Social” in the text and seem to be pleased with the service and deal provided by the merchant, then the customers making quality inferences could take this information into account. Another interesting perspective raised by our work is the possibility of offering deals only for consumers on demand, i.e. offering deals to consumers around a particular area and the possibility of buying deals only when consumers are actually at the merchant’s place. Using the geographical location from mobile devices, for example, might allow daily deal vendors to showcase offers at the right time (i.e. just before consumption). Furthermore, our findings in regard to how deal competition affect the reviews of focal merchants, highlights the importance for daily deal platforms to appropriately select merchants at a given time in a given location to avoid potentially driving reference prices down for all merchants in a given segment. Our discussions with managers at a leading deal platform firm suggested that the firm was

indeed aware of these broader reference price effects and had considered strategically targeting selected retailers or merchants so as to avoid these externalities. We believe this is also an area where more research is warranted.

Our study also has some important limitations. In regards to our empirical model, we are limited in our ability to generalize our findings since our data is for a single major city in the United States. However, we observe the population of online deals for over a year, which allows us to account for seasonality effects during the year. Further, while we are able identify a deal competition effect in our empirical model and confirm this effect in a controlled setting, we cannot rule out or specifically test individual mechanisms that might drive a decrease in ratings for focal merchants under high deal competition. Future work should address the specific mechanisms behind the deal competition effect to understand consumer choice given a range of different competitive environments. This would be a more accurate representation of the current state of daily deal vendors and restaurants in many cities in the U.S, and is likely to require a series of studies using archival as well as experimental methods to truly establish the key causal mechanisms. Finally, there are many underlying drivers of restaurant eWOM and daily deals represent only one such driver. However, we have focused our attention here only on the marginal effect of deals and heterogeneity therein, where we believe we have cleaner identification.

In summary, we provide the first empirical analyses of the heterogeneity in the response to daily deals in merchant eWOM, and show that restaurant heterogeneity and competition have a significant role to play in moderating the so-called “Groupon effect” (Byers et al. 2012a). Our work also provides significant

managerial implications for platform firms like Groupon, as well as merchants who are customers of such platform firms. Our work helps to open up the black box of the daily deals effect to an extent, using both econometric as well as experimental methodologies. However, we believe that daily deals operate across product and service categories in many different ways, and much more work is necessary to fully understand their implications. While current online marketing and digital business strategies have accepted daily deals as a legitimate part of the retailers' marketing mix, their interactions with consumer perceptions of quality as well as interactions with other elements of the marketing mix represent many open questions that provide many fruitful avenues of future research.

Chapter 4 Watch where you eat: Restaurant Hygiene

Inspections in New York City and Moral Hazard

Introduction

There is significant public awareness of the importance of the nutritional content of food, or “watching what we eat”. However, another important public health issue concerns the hygiene practices of food establishments, or where we eat (Jones and Angulo 2006). The CDC estimates that, as a result of foodborne illness, one in six Americans gets sick and approximately 3,000 Americans die each year (CDC 2016a). Many of these outbreaks are restaurant-related, with approximately 60% of cases estimated to be a result of food prepared at restaurants (Hedberg et al. 2006; Gould et al. 2013). Public policy makers have reacted by launching programs to inspect and certify the hygiene of restaurants. The potential benefits of such programs are clear and have been empirically shown to decrease foodborne illness outbreaks, (Bucholz et al. 2002; Irwin et al. 1989), particularly in cities where the hygiene scores are publicized (Jin and Leslie 2009; Jin and Lee 2014). However, inspecting restaurants is a costly and time-consuming process, and real-time changes in hygiene quality are difficult to observe through infrequent inspections. Thus, as with any certification scheme with imperfect information (Shapiro 1986), there is a possibility for moral hazard.

In this study, we provide evidence about moral hazard in restaurant hygiene by analyzing the New York City (NYC) restaurant inspection program. NYC has one

of the highest number of restaurants per-capita in the world (Bloomberg 2015), and over 55% of lunches and dinners take place in restaurants (Zagat 2015). The NYC restaurant inspection program was started in 2010 in an effort to curb the city's large number of food poisoning outbreaks. The program assigns hygiene scores and corresponding grades (A, B or C), which are then displayed on a sticker on the front of the restaurant. Consistent with previous findings in the literature, the program led to a 14% decrease in foodborne hospitalizations between 2010 and 2012 (Wong and Matis 2013) and observed an increase in the number of "A" grade restaurants from 27% in 2010 to 41% in 2012 (Farley 2012).

However, one particular aspect of the NYC inspection program is how it treats restaurants that do not achieve an A grade. While restaurants that achieve an A grade receive an 'A' sticker following inspection, those that do not instead receive a 'P' (Pending) grade sticker and are scheduled for re-inspection several weeks later. In fact, the majority of restaurants in the program initially receive a P grade (with average scores that would yield a C grade) and then receive an A grade after the re-inspection. The 'P' sticker is then replaced with the grade received in the re-inspection. Clearly, these firms are able to meet high standards of hygiene. However, in future inspection cycles, over 70% of these restaurants display a similar trend, again receiving a P grade (with average scores that would again yield a C grade), being scheduled for re-inspection, and finally achieving an A grade upon re-inspection.

Moral hazard is a framework that can help explain the difference between low, largely C-grade initial inspection scores, and high, largely A-grade re-inspection

scores. Under moral hazard, firms use the guarantee of a P grade in the initial inspection as a type of temporary insurance from the real inspection grade, then make quick and significant investments in hygiene quality that yield an A grade upon re-inspection, when it counts in terms of public perception. Once a firm attains an A grade, the next inspection cycle can take several quarters, during which time firms lose their incentives to invest in proper hygiene. At the center of this issue is the fact that it is not economically feasible to continuously monitor restaurant hygiene quality. NYC, for example, has over 20,000 restaurants, and inspections often take hours and require significant costs for the city in terms of organization and deployment and for restaurants in terms of lost revenue. NYC and some other cities rely on random inspections to choose restaurants for inspections at irregular intervals. However, given the large scope of the NYC restaurant population, the NYC program is often criticized by the popular press for being backlogged and behind schedule (NY Times 2012, NY Daily News 2015).

But are restaurants with Grade Pending, Grade A, Grade Pending, Grade A (PAPA) trajectories inherently different from those with straight As (AA)? Or is their difference in scores during initial inspections and re-inspections explained by moral hazard? To identify the effects of moral hazard from restaurant heterogeneity, we propose a novel methodology to measure hygiene from consumer-generated content in social media. We apply recent advances in machine learning and computer science to develop, test, and externally validate a social-media hygiene dictionary to measure the hygiene of restaurants continuously. While inspection results are only measured at discrete time points, this provides a continuous measure of “user-reported hygiene”

for each restaurant over time. One of the advantages of this methodology is that consumers act as inspectors inadvertently by sharing their content online, which also reduces concerns about endogenous hygiene quality choices.

To fully identify the effects of moral hazard, we take four different empirical strategies. First, we compare the hygiene inspection results of the group of restaurants that change their performance depending on the type of inspections with restaurants that get consistent scores independent of the type of inspection. We control for a number of observable characteristics of restaurants (e.g. location, segment, cuisine) and inspectors (e.g. experience, previous scores). Second, we use a difference-in-difference model to study the change in user-reported hygiene following an inspection that results in a P grade. With this model, we show that such inspections significantly improve hygiene practices as measured through our user-reported measure. Moreover, we show a differential effect for different types of establishments: for example, lower-priced restaurants display a greater change in user-reported hygiene, highlighting the role status may play in moral hazard. Third, we compare user-reported hygiene in the months before and after initial inspections and re-inspections, and we find that user-reported hygiene is strikingly different for restaurants with inconsistent inspection scores versus those with consistent inspection scores. Finally, we conduct two robustness checks. First, we show that user-generated hygiene is unlikely to be influenced by scores by comparing the user-generated hygiene of restaurants before and after the introduction of the NYC inspections program and find no significant differences before and after the program in the consumer experiences shared in social media, thereby further allaying concerns

of endogeneity. Second, we show that, as expected, users can accurately predict hygiene problems that are observable by consumers while unable to predict hygiene problems that are unobservable (e.g. problems in the kitchen or paperwork).

After controlling for restaurant characteristics, economic indicators, and inspector characteristics, we estimate that moral hazard explains approximately 30% of the A-grade restaurants in the NYC program and appear to showcase high quality in hygiene while in fact earn C-level inspection scores. The study also contributes to the emerging literature leveraging machine learning methodologies in economics (Athey 2015, Athey and Imbens 2015) by providing a methodology to measure hygiene quality based on user-generated content in social media. Moreover, we build on prior work on public health surveillance showing that social media content can accurately identify episodes of foodborne illness (Harrison, et al. 2014) and is predictive of hygiene inspection scores (Kang et al. 2013) to provide, to the best of our knowledge, the first publicly available hygiene dictionary, which can be reused by other researchers or municipalities to identify worse offenders. Finally, imperfect information about quality can have a large impact on individual and institutional behavior. As the 2015 Chipotle hygiene scandal suggests (Wall Street Journal 2015), even highly successful and reputable firms can fail to follow proper hygiene practices, which may lead to serious public health concerns and significant brand damage. The current study aims to more fully inform the debate on how to design and structure incentives for efficient restaurant hygiene quality investments, leveraging the power of big data and social media analytics to tackle the organizational and economic barriers to continuous monitoring.

Context: The NYC Restaurant Inspection Program

To operate a food service establishment in New York City, owners must have their restaurant inspected and graded by the NYC Department of Health and Mental Hygiene (Health Department from now on). The program was started in 2010 during Mayor's Bloomberg administration with three goals: first, give consumers easy access to information about the quality of hygiene of restaurants; second, improve restaurants' hygiene practices; and third, reduce the amount of restaurant-related foodborne illness.

Food service establishments are defined as fixed-site food vendors, a category that includes restaurants, coffee shops, bars, nightclubs, and most cafeterias (DOHM 2014). The program excludes mobile food vending units or temporary food service establishments, such as food trucks, correctional facilities, and charitable organizations. As per the inspection program, every food service establishment receives an unannounced random inspection at least once per year from the Health Department. The visit may take place anytime the establishment is open to the public or preparing food. The Health Department inspects approximately over 20,000 food service establishments each year to monitor their compliance with food safety regulations (DOHM 2012b).

For each inspection, the inspector follows an established procedure to record in a mobile device the observed violations to the health code. Lower inspection scores show better adherence to the Health Code. Each violation is associated with a range of points, which depends on the type of violation and the risk it presents to the potential consumer. At the end of the inspection, the points are summed, and the total

becomes the final inspection score, which is made publicly available. Scores with 13 or fewer points, 14 to 27 points, and 28 or more points result in A, B, and C grades, respectively. Figure 4.1 displays the grade cards that are displayed at the entrance of restaurants, which to comply with the health code must be displayed within 5 feet of the entrance.

One element that makes the NYC program distinct from others in the country is their two-step inspection process (DOHM 2014; Ho 2012). Each restaurant is inspected through an “inspection cycle”, which begins with an initial inspection, but which can also include a re-inspection to ensure that any identified problems have been corrected. If the initial inspection yields an A grade, the establishment receives and posts an ‘A’ sticker and is not subject to a re-inspection until the next inspection cycle (roughly 12-18 months later). However, if a restaurant receives a B or C grade, the establishment is scheduled for re-inspection approximately within 3 months. Meanwhile, the restaurant posts a "grade pending" or ‘P’ sticker. The score generated from the re-inspection, which may be an A, B or C grade, must be posted immediately unless the restaurant requests a hearing at the NYC Health Administrative Tribunal. Grade cards must be displayed within 5 feet of the entrance within clear view of consumers. Figure 4.2 shows an example grade card in a restaurant.

There are several types of violations as specified in the in the NYC health code (DOHM 2016). First, violations are classified as “critical” or “general”. Critical violations are ones that contribute the most to foodborne illness and pose a significant risk to consumers. As such, critical violations receive more points than general

violations and represent more egregious hygiene risks. For example, failing to maintain certain ingredients at a safe temperature is a critical violation and receives between 7 and 28 points, while failing to maintain a toilet facility clean is a general violation and receives between 2 and 5 points. Furthermore, critical violations that pose an immediate “public health hazard” receive the highest number of points. If a restaurant does not correct such violations before the end of the inspection, it may result in the restaurant being closed.

The number of points received for a particular violation also depends on the condition level, which is the extent and frequency of the violation. Some violations have more condition levels and parameters than others. Conditions can vary from level 1 to level 5, with higher levels receiving higher numbers of points and signifying a more severe violation (DOHM 2011). For example, the presence of a single contaminated food item would constitute a lower condition level (Level 1 and 7 points) whereas the presence of 4 or more different contaminated food items would earn a higher condition level for the violation (Level 5 and 28 points). Figure A-3 in the appendix shows a partial list of violations, condition levels and corresponding points from DOHM (2010a).

At the end of the inspection, the inspector reviews the results of the inspection with the operator, explains the violation and condition scores and makes suggestions to improve food safety. The inspector then issues an inspection report, which contains a list of all violations and their corresponding points and severity, and the total inspection score (DOHM 2010a). Depending on the specific violations identified, restaurants are required to pay fines, which range from \$200 to \$2000 and may be

higher for repeated violations (DOHM 2012a). As a further penalty, restaurants are automatically closed if they score a grade of C in three consecutive inspection cycles (DOHM 2014).

In general, the program has enjoyed wide acceptance by consumers in NYC. A recent survey from Baruch College showed that 89% of New Yorkers consider grades when dining, 91% approve of publicizing grades, and 77% feel more confident dining in an A grade restaurant (CUNY 2012). In terms of changing restaurant hygiene practices, the number of restaurants that receive an A grade at the end of an inspection cycle has increased significantly over time, from less than 30% in 2010, the first year of the program, to 42% in 2012 (CUNY 2011), and according to internal sources at the Health Department, 80% in 2014 (DOHM 2014). Restaurants have also received less severe and critical violations over time. Perhaps more importantly, the rates of foodborne illness in NYC have declined significantly; according to the CDC, NYC had over 2.1 million foodborne illness episodes in 2009 and approximately 1 million in 2014 (CDC 2016b). Moreover, there has been a decrease a 14% decrease in the reported cases in NYC of Salmonella, one of the most dangerous foodborne bacteria, between 2010 and 2013 (DOHM 2014).

The Data: Inspections and Online Reviews for New York City

The data on restaurant inspections is available through the NYC Open Data program (NYCOD). NYC Open Data makes large amounts of public data generated by various New York City agencies and other City organizations available for public use (NYCOD 2016). The dataset contains information of all NYC restaurant inspections since the beginning of the restaurant grading program in 2010 and is

updated every month. It includes restaurant-specific information, such as the restaurant's address, phone, and cuisine; inspection-specific information, such as the inspection date, type, and resulting grade; and violation-specific information, such as the violation code, type, severity, and points. A full list of the inspection dataset variables is available in Table 4.1. We supplement the information on each restaurant with restaurant-specific data from Yelp.com, which is the leading online reviews platform for restaurants with over 95 Million reviews in the third quarter of 2015 (Yelp 2016). Restaurants in the two datasets are matched using the restaurant address and phone with over 95% overlap. The data in Yelp contains more detailed restaurant characteristics, such as price point, average consumer rating, hours, and parking options. A list of these variables is also available in Table 4.1.

We find that, as commonly reported, a majority of inspections result in A grades, and the proportion of A grades has been increasing year to year. Figure 4.3 shows the grade distribution for the three years before and after the grading program started in 2010 (DOHM 2014). Further, Figure 4.4 illustrates that the distribution of A grades displays little geographical variance and illustrates the overall increase in the proportion of A grades between 2011 and 2015. Likewise, Figure 4.5 illustrates a similar percentage of restaurants with an A grade in each of the five New York boroughs. For example, in 2015 almost 90% of all restaurants in NYC received A grades, ranging from 87.7% in Queens to 90.2% in Manhattan. Finally, the rate of A grades appears to be unrelated to poverty rates available also from NYC Open Data for matching boroughs and neighborhoods (2016), as illustrated in Figure A-1 in the

Appendix. The NYC hygiene inspection program therefore appears to be successful across the board and not only in a single borough, neighborhood or economic group.

While many restaurants eventually receive an A grade, these restaurants may have received that grade upon initial inspection or upon re-inspection. In fact, approximately 55% of restaurants that achieve an A grade actually receive a P Grade in their initial inspection, with an initial score that would have resulted in a B or C grade. The Health Department reports that there are 4,000 restaurants (or approximately 20% of all restaurants in NYC) with a P grade at any given time (Health Department 2015); Figure A-2 in the appendix illustrates this large proportion of restaurants with a P grade at a recent date in 2015 by plotting the grades over a map of NYC.

Restaurants that receive an A grade in their initial inspection and those that receive an A grade upon re-inspection have significantly different initial inspection scores ($\text{Mean}_A = 8.75$; $\text{Mean}_{PA} = 21.93$; $p < 0.001$ in a two-sided t-test). In fact, the median initial inspection score for restaurants that receive an A grade upon re-inspection is 26.5, which is near the cutoff for a C grade, 28 points. Yet upon re-inspection, these restaurants are able to perform at A-grade level and achieve scores that are not significantly different from those of restaurants that achieve an A grade in their initial inspection ($\text{Mean}_A = 8.75$; $\text{Mean}_{PA} = 9.45$; $p = 0.76$ in a two-sided t-test). These restaurants are therefore able to achieve high levels of hygiene despite committing significant, and often critical, hygiene violations in their initial inspections.

This behavior appears to be analogous to the one documented in classic contract theory, and in particular in signaling theory (Spence 1973; 2002), where firms (restaurants) aim to convey credible information, often regarding quality, to potential buyers (consumers) and where an impartial third party (Health Department) provides the certification or quality signal. This has also been demonstrated more recently in the importance of brands signaling quality in online marketplaces (Waldfogel and Chen 2006) and in many other contexts, such as online auctions (Lewis 2011; Dimoka, Hong, Pavlou 2012), and investment options (Goldlücke and Schmitz 2014). According to signaling theory, we would then expect firms that earn a P grade in their initial inspection, and thus signal a temporarily uncertain level of quality, to improve their inspection scores for the re-inspection and signal the highest possible hygiene level, an A grade.

We proceed to test whether the possibility of earning a P grade, which is only available for initial inspections, causes an improvement in hygiene as measured by inspection scores. Effectively, we want to confirm our expectations from signaling theory that the public grading program incentivizes restaurants to improve their performance. We employ a longitudinal difference-in-differences (DiD) model, which allows the same restaurants to serve as treatment and control groups at different points in time. Since we are interested in understanding the effect an event or treatment has on a quality outcome for a particular subject over time, utilizing a difference-in-differences approach is an appropriate methodology for causal inference (Bertrand et al. 2004). The essential idea of DiD is to examine a group of treated units before and after the treatment. In our case, restaurants are considered part of

the treatment group when an inspection occurs that has a possibility of a P grade (initial inspection) and are considered part of the control group when an inspection occurs that does not have the possibility of a P grade (re-inspection). The control group is an important part of the DiD framework since other factors that influence hygiene maybe have changed over time (see Lechner 2011 for a thorough review of this literature).

Moreover, a DiD framework is a particularly appropriate framework for our context for several reasons: First, the treatment (initial inspection) is randomly assigned and all restaurants receive the treatment at some point in the dataset. This is of particular importance to identify a treatment effect in longitudinal models (Athey and Imbens 2006). Second, the common trends assumption, which asserts that the differences in the expected control outcomes over time are not related to being part of the treated or control group in the post-treatment period. The implication is that if the treated group had not been subjected to the treatment, it would have experienced the same time trends. Since all groups in our data belong to a single category of merchants, restaurants, and are located in the same geographical region, the common trends assumption is highly logical for our context. Third, having a high number of time periods (particularly similar time periods) and groups (particularly similar restaurants) of control units is important as it has been shown to provide more precise estimation of treatment effects, provide more reliable testing of the common trends assumption, and more precise inference (Lechner 2011).

The unit of analysis is restaurant-time period, where the unit of time is quarters (we find consistent results using months as the unit of time). The outcome

variable is the numerical inspection score resulting, which is available from the inspection report. We include two types of controls as suggested by Imbens and Wooldridge (2007), group-level and treatment-level. Group-level controls, which have been shown to improve identification in DiD models, include all observable restaurant characteristics (e.g. price point, location). Treatment-level controls, which help account for within-group variation and reduce standard errors (Imbens and Woolridge 2008), include inspector characteristics and the type of violations. Summary statistics for the DiD model are shown in Table 4.2. To check the appropriateness of a DiD framework, we test the common trends assumption using the leads and lags methods of Author (2003) and Pischke (2014). We find no evidence of a violation of the common trends assumption.

$$INS_{it} = \beta_0 + \beta_1 TreatmentGroup_i + \beta_2 AfterTreatment_t + \beta_3 TreatmentGroup_i * AfterTreatment_t + \gamma Controls_{it} + \varepsilon_{it}$$

Here, INS_{it} is the numerical score resulting from an inspection of restaurant i during time period t . Table 4.3 displays the initial results from the difference-in-differences model capturing the effect of an initial inspection with the possibility of a grade P on inspection scores. The results show a highly negative and significant (-6.23; $p < 0.001$) average treatment effect (ATE) of an initial inspection on inspection scores, illustrating that the occurrence of initial inspections leads to a significant improvement in hygiene performance following the inspection. Moreover, we see a differential treatment effect for restaurants in different price segments. Specifically, the ATE for high-priced restaurants (3-4 Yelp dollar signs) is less negative, as seen by the positive and significant interaction coefficient (1.47; $p < 0.01$), while the ATE

for low-priced restaurants (1-2 Yelp dollar signs) is slightly more negative (-0.15; $p < 0.05$). This suggests that more high-end establishments are more stable in their hygiene performance before and after initial inspections.

From the DiD model, we conclude that the two-step inspection cycle policy of the Health Department indeed seems to be incentivizing restaurants to perform at higher levels of hygiene quality in the re-inspection. But how do restaurants perform in future inspections cycles? We find that approximately 85% of restaurants that achieve an A grade in their initial inspection also achieve an A grade in future initial inspections, while approximately 90% of restaurants that receive a P grade before achieving an A grade upon re-inspection do so again in future cycles. These two groups of A-grade restaurants, which we call “AA” and “PAPA” respectively, represent approximately 40% and 50% of all restaurants in NYC, illustrating that the majority of restaurants tend to repeat their behavior in inspection cycles over time. Figure 4.6 shows the inspection scores for all restaurants in these two groups for the first two inspection cycles; Figure 4.7 shows the corresponding average inspection scores. The dramatic difference between re-inspection scores in the first inspection cycle and the initial inspection scores in the second inspection cycle for the PAPA group suggest that many restaurants only improve their quality when a re-inspection is imminent (i.e., an initial inspection has recently occurred). We propose moral hazard as an explanation for this behavior.

More precisely, we argue that consistent with the information asymmetry literature (Stiglitz 2009), firms use regulatory frameworks, or government interventions, to their advantage, which may yield optimal outcomes for the firm. In

the words of Stiglitz “Even when markets are efficient, they may fail to produce socially desirable outcomes. The wealthy and powerful may “exploit” others in an “efficient” way” (2009). In our context, we propose that the behavior of PAPA firms is consistent with this argument as they take advantage of the P grade to produce suboptimal results in their hygiene quality. At the core of this issue is the high cost of perfect information or monitoring found in this literature (Akerlof 1970) and more recently the food safety literature (Starbird 2005). Specifically, because the Health Department cannot continuously monitor hygiene, limiting our ability to measure restaurants’ hygiene quality before and after inspections and thereby identify moral hazard. We therefore propose tracking hygiene quality through a secondary data source in order to identify how the behavior of different groups of restaurants (e.g. PAPA and AA) might differ. More exactly, we propose analyzing the text of online reviews using recent advances in machine learning to create a continuous measure of restaurant hygiene quality. We describe the machine learning procedure and its validation in the next section.

Analysis and Results

Creating a Social Media Sourced Hygiene Dictionary

We construct a Social Media Sourced Hygiene Dictionary (SMASH) to extract information about restaurant hygiene from the text of online reviews. There is a large body of work studying online reviews in management (Dellarocas 2003). Typically, these user-generated comments have been linked to sales in the marketing and information systems literatures. (Chevalier and Mayzlin 2006, Forman et al.

2008). Moreover, online reviews have been studied in many contexts, such as film industry (Duan et al. 2008), hotel services (Ye et al. 2009), and the medical care (Gao et al. 2012). However, most of these studies use numerical summaries, such as average ratings, in their analysis and largely ignore the text contained in the reviews. There is a small but growing literature, such as Decker and Trusov (2010), Archak et al. (2011) and Cao et al. (2011) that have attempted to summarize the corpus of text contained in the reviews. Closer to our context, and perhaps our closest analog is a conference poster from Kang et al. (2013), which uses the text in online reviews to predict inspection scores in the inspection program in Seattle. However, to the best of our knowledge, we are among the first to provide a scalable methodology for the analysis of social media text with the objective of grading hygiene in restaurants.

The general idea behind SMASH is to use a thesaurus to repeatedly augment an initial seed list of hygiene words based on synonym and antonym relationships. Specifically, an initial set of hygiene-related words with labeled polarity, or strength of relatedness to hygiene, are first defined. This initial list of seed words is augmented using synonyms in WordNet (Miller 1995; Feinerer and Hornik 2016), a popular online dictionary and thesaurus. After the newly identified words are added to the seed list, the process of augmenting the seed words using WordNet is repeated. This iterative procedure continues until no more new synonyms can be found. Once the growing process has finished, the dictionary is manually curated to remove clear errors that result from homonyms. For example, a word such as “roach” is short for “cockroach” (relevant) and also synonymous with “Mexican valium” and other drugs (not relevant). Instances like this are manually erased from the dictionary

We note that our overall procedure closely resembles the standard approach in the sentiment and opinion mining literature (Tsai et al. 2013; Feldman 2013; Liu 2015) with one important extension. In sentiment mining, the initial set of seed words are often manually specified (see Valitutti 2004; Hu and Liu 2004). However, manually specifying words for the context of hygiene is much less obvious compared to tonal sentiment and may be unduly influenced by the researcher’s own vocabulary or lack thereof. Indeed, poorly created seed lists can lead to less accurate dictionaries (Tang et al. 2009). Therefore, to mitigate this potential bias, we generate the initial word list through the Naïve Bayes classifier, a machine learning technique that identifies the initial word list in a data-driven manner following Liu (2015). We discuss this technique next.

Building the List of Seed Words

We first introduce some notation that will help facilitate our discussion of the Naïve Bayes classifier (Hand et al. 2001; Tang et al. 2009). Suppose we are given a training dataset with n documents that are labeled by their hygiene polarity, or d_j , which we define as a binary variable denoting whether a review (document) j is discussing hygiene negatively, where $j = 1, \dots, n$. Let p denote the total number of unique words that appear in all reviews, and let w_{jk} denote the number of times word k occurs in review j for $k = 1, \dots, p$.

The Naïve Bayes classifier estimates the probability that each document discusses hygiene negatively based on the word occurrences, i.e., $P(d_j = 1 | w_{j1}, w_{j2}, \dots, w_{jp})$. Through an application of Bayes Rule, one could in principle directly calculate the probability:

$$P(d_j = 1 | w_{j1}, w_{j2}, \dots, w_{jp}) = \frac{P(w_{j1}, w_{j2}, \dots, w_{jp} | d_j = 1) P(d_j = 1)}{P(w_{j1}, w_{j2}, \dots, w_{jp})}.$$

However, in practice calculating the joint distributions requires an unrealistically large amount of data (Hand et al. 2001). To overcome this issue, the joint distribution is simplified under the assumption of conditional independence:

$$P(w_{j1}, w_{j2}, \dots, w_{jp} | d_j = 1) = \prod_{k=1}^p P(w_{jk} | d_j = 1).$$

After applying the law of total probability (Ross 1996),

$$P(w_{j1}, w_{j2}, \dots, w_{jp}) = \prod_{k=1}^p P(w_{jk} | d_j = 1) P(d_j = 1) + \prod_{k=1}^p P(w_{jk} | d_j = 0) P(d_j = 0),$$

the probability that a document discusses hygiene negatively based on the word occurrences can be expressed as

$$P(d_j = 1 | w_{j1}, \dots, w_{jp}) = \frac{\prod_{k=1}^p P(w_{jk} | d_j = 1) P(d_j = 1)}{\prod_{k=1}^p P(w_{jk} | d_j = 1) P(d_j = 1) + \prod_{k=1}^p P(w_{jk} | d_j = 0) P(d_j = 0)},$$

where $P(w_{jk} | d_j = 1)$ and $P(w_{jk} | d_j = 0)$ can easily be calculated by inspecting how often the k th word appears in documents that are labeled as discussing hygiene negatively ($d_j = 1$) or those that are not ($d_j = 0$) (Hand et al. 2001).

Thus, the conditional independence assumption is a fundamental one that defines the Naïve Bayes classifier. Even though from a probabilistic perspective the conditional independence assumption is not realistic, its performance in many different machine learning contexts has been demonstrated (Hand et al. 2001). In fact, there is a long history of Naïve Bayes classification in text mining and sentiment analysis (see McCallum 1998; Sebastiani 2002; Go and Bhayani 2009; Feldman 2013 and references therein). After estimating the Naïve Bayes classifier as described in

the next section, we sort all words by their estimated $P(w_{jk}|d_j = 1)$ and keep the top 5% as the initial seed list. This ensures that words that are chosen are strongly associated with negative discussion of hygiene (see McCallum 1998; Sebastiani 2002; Go and Bhayani 2009; Feldman 2013 and references therein).

Obtaining Training Data and Implementation

To construct SMASH, we begin with a dataset containing the text of online reviews from Yelp.com for restaurants in our inspection dataset. This dataset contains approximately 1.3 million unique text reviews (documents), which are matched to 85% of the restaurants in the inspections dataset. Restaurants are matched based on name, address and phone number, as described in the previous section. We performed standard preprocessing of all review text, such as converting to lower case, removing stopwords, and stemming (Feinerer and Hornik 2012). Since the Naïve Bayes classifier is a supervised learning technique that requires training data to estimate the conditional probabilities and define the initial seed words, we begin by creating a training dataset.

In order to identify documents in which hygiene is discussed negatively (i.e. lower hygiene ratings from consumers), we first randomly sample 1,200 restaurants with high inspection scores, which are most likely to have hygiene problems, and then randomly sample one document for each selected restaurant. Since we have documents (reviews) from before the start of the NYC inspection program (prior to 2010), we sample documents from before the start of the program (2004-2009) to avoid any potential bias on the reviews from the program. The next task is to manually assign a label regarding the hygiene relation of each document. We

recruited 1,200 subjects from Amazon Mechanical Turk (MTurk) for pay. This pool of participants has been shown to be reliable for empirical research (Goodman, Cryder, and Cheema 2013), to represent the broader population (Buhrmester, Kwang, and Gosling 2011), and to generate high quality results (Ipeirotis et al. 2010). Each subject was asked, “Given a restaurant review, answer questions about whether a review indicates problems related to hygiene,” and was then presented a single document. After reading the document, the subject indicated whether the document was related to the hygiene of the restaurant using a 7-point scale, which was adopted from Egan et al. (2006). Subjects were also asked to select the type(s) of hygiene problem described (e.g. food preparation, cleanliness). As a manipulation test, we also asked subjects to rate whether the document was positive or negative and checked whether subjects’ answers matched the numerical rating posted for a given review. We discarded nine responses based on this test because of the incorrect responses in the manipulation check. Approximately 15% of the documents were labeled as regarding the hygiene of the restaurant; the labels for each review d_i were thus defined.

To further verify the quality of the MTurk responses, we tested for the psychometric properties of the hygiene scale (Egan et al. 2006). The composite reliability was calculated and varied from 0.88 to 0.94, thereby establishing their reliability. Finally, as a robustness check, we ran a separate batch of 120 MTurk jobs with the same document, which produced a Kappa statistic of 0.81, signifying substantial inter-rater reliability (McHugh 2012).

These 1,191 training documents, their MTurk-generated labels d_j , and the word occurrences w_{jk} allow us to calculate $P(w_{jk}|d_j = 1)$, which forms the basis of the seed word list for SMASH. By expanding the seed word list using the synonym approach described above, we create a “dictionary” of hygiene from social media. This entire process was performed using single words, two-word phrases, and three-word phrases, which are also known as n-grams of up to order three in natural language processing (Lodhi et al. 2002), so that the final dictionary contains single words along with two- and three-word phrases. For example, as a result of using n-grams, phrases like "barely edible" are kept in our dictionary even when the individual words "barely" and "edible" are not included. There is been a recent call in this literature to move beyond word-level analysis, since it is a simplification of language (see Cambria (2013) for a recent example). Therefore, we build on the majority of existing work that relies only on dictionaries based on individual words (see Liu 2015 and references therein). However, for simplicity we refer to all words and phrases as “words”.

Finally, for computational simplicity and because our final dictionary includes terms that may not have been observed in the training data (due to the augmentation step), in lieu of a formal probability we again build on classical approaches in sentiment analysis to define the word counts as a continuous measure of hygiene quality over time, which can be combined with inspections data to study the behavior restaurants before and after initial inspections. For a given document j , the SMASH score is defined as the total number of times words in the SMASH dictionary appear

in the document. Letting $\delta_k = 1$ if word k is included in the SMASH dictionary and 0 otherwise,

$$WC_j = \sum_k w_{jk} \delta_k.$$

where w_{jk} is the number of times word k appears in document j . For the purposes of our longitudinal econometric models, we can summarize the SMASH scores for a particular restaurant and time period by summing over the reviews published in that time period for that restaurant. The next section describes how we validate this measure and use it to investigate moral hazard in the context of the NYC restaurant inspections program.

Validating SMASH

Since we use reviews prior to the start of the inspection program to generate SMASH, we validate SMASH using a different, non-overlapping dataset after the program began in 2010. We fit the following longitudinal mixed effects model, where t indexes one-month¹⁶ time periods:

$$INS_{it} = \beta_{1i} + \beta_2 WC_{it} + \beta_3 Rating_{it} + \beta_4 Reviews_{it} + \beta_4 Price_i + \boldsymbol{\gamma} \mathbf{Chars}_i + \varepsilon_{it}.$$

INS_{it} is the numerical score resulting from an inspection of restaurant i during time period t . WC_{it} is the total SMASH score of Yelp reviews of restaurant i published in time period t , $Rating_{it}$ is the average rating, and $Reviews_{it}$ is the number of reviews. $Price_i$ is the Yelp price segment of the restaurant (1-4 dollar signs), and \mathbf{Chars}_i is a vector of other characteristics of restaurant i available from Yelp.com such as parking options, payment methods, and ambiance. We normalize each

¹⁶ We also fit the model using quarters as the unit of time and found consistent results.

variable to represent standard deviations from the mean. To account for restaurant heterogeneity, we also include fixed effects for each restaurant.

Table 4.4 displays the results from this model (model 1). We observe that the coefficient for $Rating_{it}$ is negative and significant ($p < 0.001$), illustrating that negative reviews are associated with higher (worse) inspection scores. The coefficient for our SMASH score is positive and also highly significant ($p < 0.001$) and in fact has slightly higher magnitude than that of average rating. To account for inspector heterogeneity as well as any potential time trends, we also fit a model with fixed effects for each inspector and time. The results for this model (model 2) are displayed in Table 4.4 and are consistent with the findings discussed above.

While these models clearly show that consumers are able to observe some hygiene-related issues (and report these observations through social media), they may not be able to do so for all types of violations. For example, consumers may be unable to accurately report on critical violations, such as the cleanliness of the kitchen, which are only observable by having full access to the restaurant premises. To test this potential limitation, we divide the inspections data set into two parts: part 1 contains inspection scores made of critical violations; part 2 contains only the non-critical violation scores. As expected, we find that consumer feedback, as measured by either Yelp rating or SMASH, is unable to significantly explain critical violations ($p = 0.2$ and $p = 0.41$ respectively). On the other hand, consumer hygiene-related feedback, as measured by SMASH, is strongly associated with non-critical violations ($p < 0.001$). Interestingly, average Yelp ratings are only weakly associated with non-critical violations ($p = 0.08$)

Furthermore, we include a general sentiment measure for each review to check whether sentiment is able to explain inspection results equally or better than SMASH scores. We find that sentiment is highly correlated with the average rating, as might be expected, but is uncorrelated with our hygiene dictionary. More importantly, our results remain consistent after including sentiment in the model.

As a final robustness check, we employ data from a recent competition sponsored by Harvard, Yelp.com, and DrivenData to use the text of online reviews to predict inspection scores for restaurants in Boston (DrivenData 2016). Submissions were open to the public and the submission with the lowest prediction Root Mean Squared Logarithmic Error (RMSLE) was declared the winner. Using the program's publicly available training and test datasets, we compare the performance of SMASH with the winning algorithm and find SMASH predictions to have approximately 3% large RMSLE, indicating similar predictive performance.

Using SMASH to Show the Effect of Moral Hazard on

Restaurant Hygiene

Equipped with a methodology to monitor the hygiene of restaurants over time, we now proceed to study the behavior of restaurants prior to and following inspections by the Health Department. In particular, as discussed in the Context and Data section, we focus on the behavior of A-grade restaurants that differ in their trajectories: those restaurants that consistently achieve an A grade in their initial inspections (AA) and those that consistently receive a P grade before achieving an A grade upon re-inspection (PAPA). We theorized that moral hazard may explain the

differences in initial inspections scores for these two groups, pictured in Figure 4.7. Using SMASH as a continuous measure of hygiene quality over time, we track the hygiene of restaurants in the PAPA group in the months after achieving an A grade, during which another inspection is unlikely.

Our strategy is twofold. We first graphically display the SMASH trends after the inspections. Figure 4.8 displays the daily SMASH score for 60 days following four different inspections of restaurants in the PAPA group: first initial inspection (plot 1), first re-inspection (plot 2), second initial inspection (plot 3), and second re-inspection (plot 4). Lower SMASH scores indicate improved hygiene performance. Plot 1 displays a downward trend following an initial inspection, consistent with the results of our DiD model. Conversely, plot 2 shows an upward trend following re-inspection, suggesting that the hygiene performance of restaurants in the PAPA group tends to immediately worsen after posting an A, with the safety of knowing that the next following inspection cycle will take place approximately a year later. Plots 3 and 4 show a similar pattern of improvement following an initial inspection and worsening after re-inspection.

Second, we employ a longitudinal model to compare the SMASH scores of the AA and PAPA restaurants. We consider daily SMASH scores in the 90 days following the inspection where an A was achieved (an initial inspection for the AA group and a re-inspection for the PAPA group). Letting $Offset_{it}$ be the number of days since such an inspection for restaurant i , we fit the following model:

$$WC_{it} = \beta_{1i} + \beta_2 Offset_{it} + \beta_3 Offset_{it} * AA_i + \beta_4 Offset_{it} * PAPA_i \\ + \beta_5 Rating_{it} + \beta_6 Reviews_{it} + \beta_7 Price_i + \gamma OtherChars + \varepsilon_{it},$$

where WC_{it} is again the average SMASH score of restaurant i at time t , and AA_i and $PAPA_i$ are indicators of restaurant i 's membership in group AA or PAPA, respectively. As before, we control for restaurant characteristics and include a fixed effect for each restaurant. Results are displayed in Table 4.5 (model 1). While the main effect for $Offset_{it}$ is approximately zero, the significant coefficient for the interaction term with $PAPA_i$ shows that the PAPA group displays a positive and significant trend in SMASH scores following re-inspection ($p < 0.001$). However, the coefficient for the interaction with AA_i is near zero, suggesting that the AA group displays consistent hygiene practices after conclusion of the inspection cycle. These results show that consumers are able to observe worsening levels of hygiene quality from restaurants in the PAPA group after posting an A grade in their re-inspections. For additional robustness, we add fixed effects for the reviewers generating the online review and find strongly consistent results (see Table 4.5, model 2).

Conclusion

This paper presents a detailed empirical analysis of the effects of moral hazard in a restaurant hygiene inspection program, which is to the best of our knowledge the first such analysis. To surmount the limitations of the inspections data and continuously monitor hygiene, we propose a novel methodology to measure hygiene from the text of social media using crowdsourcing and recent advances in machine

learning. We use this continuous measure to show evidence of moral hazard in the NYC hygiene inspection program.

These effects appear to be important in the New York City restaurant industry, particularly for those restaurants that post A grades, which are the large majority of restaurants in NYC. Our results provide support for information asymmetry and basic contract theory. We observe clearly that without the threat of a low public hygiene grade and with long periods between inspection cycles, many restaurants tend to display worsening hygiene quality after a successful inspection. The results presented illustrate the need for safeguards against moral hazard in restaurant hygiene inspections programs, such as employing social media as a hygiene tracking tool through our proposed methodology. Other safeguards might include having shorter inspection cycles windows or posting initial inspection grades of B or C prior to re-inspections.

There are several limitations of the current study that should be noted. First, we consider the restaurant grades program in NYC only, which limits the generalizability of our results. Second, while we use Yelp reviews to create a secondary measure of hygiene, other sources, such as Twitter, Facebook or even Instagram, could be incorporated to create a more comprehensive social media sourced hygiene detection tool. Finally, while we uncover moral hazard in this context, we do not propose a methodology for optimizing the scheduling of inspections. We aim to take this step in future work.

In the words of the CDC and the New York City Health Department, "foodborne illness remains one of the top public health challenges for the city, the

state, and the entire country" (CDC 2016a). We believe that our approach to apply recent advances in machine learning to supplement inspections program with a continuous, social media-based measure of hygiene can be used to assist policy makers in the design and implementation of hygiene inspection programs.

Tables

Table 2.1. Matching Variables

Matching Variable	Description
Mean rating	The average rating of the restaurant
Number of reviews	The number of reviews of the restaurant
Pricepoint	The price point of the restaurant
Cuisine	Dummies for the cuisine of the restaurant (e.g. American)
Other attributes	Dummies for other restaurant attributes (e.g. attire, ambiance, noise level)

Table 2.2. Imbalance Comparison Table

Matching Variable	Type	L1	Min	25%	50%	75%	Max
Mean rating	(diff)	< 0.001	0	0	0	0	0
Number of reviews	(diff)	< 0.001	0	-5	-19	-36	7
Pricepoint	(diff)	< 0.001	0	0	0	0	0

Table 2.3. Comparing Case and Control Samples

Sample	n	Mean Rating	Number of Reviews	Pricepoint
Case	446	3.26	54.80	1.76
Control	605	3.31	59.50	1.65

Table 2.4. Top Words Associated with Each Semantic Component

Quality_Overall	Food_Efficiency	Responsiveness	Food_Quality	Atmosphere
food	place	order	good	place
good	bar	server	veri	crowd
place	order	came	great	happi
like	time	ask	dish	bar
order	one	dish	nice	beer
friend	drink	menu	flavor	neighborhood
time	wait	waiter	chicken	like
great	minut	minut	restaur	food
nice	hour	tabl	sauc	drink
service	ask	meal	food	music

Table 2.5. Reviews Associated with the Identified Five Semantic Components

Reviews that Load onto Quality Overall
<p>“Food was okay. My shrimp tempura roll was good, but the donburi wasn't. The tempura ice cream was my favorite part of the meal. The service was pretty good. Our server was a genuine sweetheart so I might go again for the rolls and the service. Pretty place too.”</p> <p>“The atmosphere of the place was kind of weird. The food wasn't all that impressive for Thai food in the DC area... If you get \"Beef Red Curry\", you kind of expect more than 4 small pieces of beef... The service though was outstanding. The server was always around for water/drink refills and was very nice.”</p> <p>“We weren't seated in the main dining area (that's for the highrollers)... Palena was a good meal overall, not stressful as some good restaurants tend to get around the busiest dinner hours.”</p>
Reviews that Load onto Food Efficiency
<p>“It took them 30 minutes to make 2 burgers.”</p> <p>“It's just as good as toki underground, and better yet there is not a ridiculous two hour wait...you order at the counter and they prepare it right away. you get your food within 5 min... it kept me coming back again and again.”</p> <p>“We went on a weekday night.. and we were told there was a 15 minute wait. The host scooped us out of the bar not even a minute later with an available table.”</p> <p>“We got there earlier than our reservation time (7PM on a Thursday) and were able to be seated right away.”</p>
Reviews that Load onto Responsiveness
<p>“my medium well burger came out medium rare, and they put bacon on my wife's veggie burger (she did not ask for this obviously).”</p> <p>“I really, really liked the vending machine that was dispensing beer and cigarettes -- you had me at hello. I also thought the bartender was super friendly and accommodating”</p> <p>“Not only did the server have great difficulty comprehending the neatly written break-down, he came back and told us that we were \$1 short of the \"suggested gratuity\" shown on the receipt.”</p> <p>“Service was prompt and pleasant”</p> <p>“My fiancé and I were gracious to be sat quickly at a one of the last tables in the full dining room set for four.... Notably, different members of the staff delivered plates and shared a little bit about what we were about to enjoy (without aimlessly listing off ingredients)”</p> <p>“Our waitress Rosalin(?) was nice enough, but very very confident and almost seemed like she was acting out a scene as she told us a bit about the restaurant and theme.”</p>
Reviews that Load onto Food Quality
<p>“The shrimp were perfect, really perfect, I had to resist stealing more from my friend. I enjoyed the fried oysters in cornmeal too. The shrimps were just really flavorful, just really good.”</p> <p>“The delicate rings of squid -- so exquisitely supple save for one or two pieces -- were served on a velvety polenta, fire-roasted tomato fondue, and fresh pesto. A perfectly portioned and divine way to begin what was about to be our oceanic adventure... The lobster was excruciatingly tender and sweet, paired with a clarified lemon-herb butter. The oysters were enormous and juicy, paired with a mouth-puckering mignonette. The jumbo shrimp were bigger than jumbo, delicate meat executed perfectly, paired with a spicy cocktail sauce. The mussels were wonderfully plump and meaty.”</p> <p>“We consider it one of the best meals we've ever had so far.... Course after course titillated and awed, I loved every minute of it.”</p> <p>“everyone was really pleased with their food....The tempura was quite good. The soup was a little bit too salty and the dumpling really greasy, but overall, everything was good.”</p>

<p>“Corned beef sliders - good. Heavy, yummy.... Shepherd's pie ...was just ground lamb and potatoes, ...Lemon something for dessert - looked very yummy. Eh. Even the shortbread was eh.”</p>
<p>Reviews that Load onto Atmosphere</p>
<p>“The atmosphere of the place was kind of weird. There are rocks on the ceiling and it was a little bit too dark. I guess they were going for a cave theme, but the important question is WHY???”</p> <p>“It smells greasy. Atmosphere is lacking, but since this is more a take out place for lunch time, it doesn't have to have atmosphere. Then again, why not work a bit at it. Easiest first step - turn on some music, not much, not loud, just a little basic background. I understand New Orleans is famous for its music...and the guys working here might also like that.”</p> <p>“what i like about this place is you get great food for a very reasonable price in a laid back, low key, and personable environment.”</p>

Table 2.6. Variable Descriptions

Variable	Description
$Closure_{it}$	1 if restaurant i is closed in time period t ; 0 otherwise
$meanrating_{it}$	Average rating of reviews for restaurant i in time period t
$numreviews_{it}$	Number of reviews for restaurant i in time period t
$Quality_Overall_{it}$	LSA measure for the overall experience in restaurant i in time period t
$Food_Efficiency_{it}$	LSA measure for the reliability and wait times in restaurant i in time period t
$Food_Quality_{it}$	LSA measure for the food quality in restaurant i in time period t
$Responsiveness_{it}$	LSA measure for the service responsiveness in restaurant i in time period t
$Atmosphere_{it}$	LSA measure for the atmosphere in restaurant i in time period t
$Pricepoint_i$	Price point for restaurant i
WL_{it}	Average word count of reviews for restaurant i in time period t
$readability_{it}$	Average SMOG readability index of reviews for restaurant i in time period t
$comp_meanrating_{it}$	Average mean rating for restaurant i 's competitors in time period t
$comp_numreviews_{it}$	Number of reviews for restaurant i 's competitors in time period t
$numcompetitors_{it}$	Number of competitors for restaurant i in time period t
$Cuisine_i$	Set of binary variables indicating whether each of 16 cuisines is listed in cuisine type for restaurant i (restaurants can have multiple cuisines)
Loc_i	Set of binary variables indicating the zip code of restaurant i
$OtherChars_i$	Set of binary variables describing 15 other characteristics for restaurant i , such as payment method, parking, attire, group-friendly, kid-friendly, waiter, Wi-Fi, alcohol, etc.

Table 2.7. Summary Statistics

VarID	Variable	Mean	Std. Dev.
1	<i>meanrating_{it}</i>	3.316	0.921
2	<i>numreviews_{it}</i>	5.173	5.085
3	<i>Quality_Overall_{it}</i>	0.002	0.001
4	<i>Food_Efficiency_{it}</i>	0.001	0.001
5	<i>Food_Quality_{it}</i>	-0.001	0.001
6	<i>Responsiveness_{it}</i>	0.001	0.001
7	<i>Atmosphere_{it}</i>	-0.001	0.001
8	<i>Pricepoint_i</i>	1.774	0.638
9	<i>WL_{it}</i>	120.627	68.394
10	<i>readability_{it}</i>	9.387	2.037
11	<i>comp_meanrating_{it}</i>	3.313	0.362
12	<i>comp_numreviews_{it}</i>	2.898	2.060
13	<i>numcompetitors_{it}</i>	39.203	25.245

Table 2.8. Correlation Matrix

VarID	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1												
2	0.06	1											
3	-0.06	0.1	1										
4	-0.24	-0.02	0.26	1									
5	-0.13	-0.09	-0.26	0.16	1								
6	-0.18	0.09	0.23	0.21	-0.13	1							
7	-0.11	0.04	-0.21	-0.2	-0.04	0.14	1						
8	-0.05	0.24	0.22	-0.05	-0.23	0.2	0.05	1					
9	-0.08	0.09	0.14	0.12	-0.16	0.26	-0.1	0.21	1				
10	0.04	0.01	0.19	0.06	-0.04	0.06	-0.03	0.07	0.23	1			
11	0.09	-0.06	-0.02	-0.03	0	-0.02	0	-0.13	-0.01	0.01	1		
12	-0.07	0.21	0.04	-0.01	-0.07	0.1	0.07	0.15	0.02	-0.03	-0.13	1	
13	-0.01	0.01	-0.02	0.05	0.06	-0.06	-0.09	-0.11	-0.03	-0.01	-0.02	0.09	1

Table 2.9. GLMER Coefficient Estimates

Variable	Model 1: Base model	Model 2: Semantic Variables	Model 3: Closed Sample	Model 4: Fixed Effects (EX)	Model 5: Fixed Effects (AR-1)
<i>(Intercept)</i>	-5.29 (0.23) ***	-5.33 (0.23) ***	-2.94 (0.16) ***	-2.55 (0.17) ***	-2.59 (0.18) ***
<i>meanrating_{it}</i>	-0.12 (0.05) **	-0.1 (0.05) *	-0.1 (0.04) *	-0.09 (0.01) *	0.1 (0.02) *
<i>numreviews_{it}</i>	-2.07 (0.18) ***	-2 (0.18) ***	-1.58 (0.16) ***	-2.04 (0.39) ***	-2.06 (0.4) ***
<i>Quality_Overall_{it}</i>		-0.57 (0.17) ***	-0.4 (0.17) *	-0.97 (0.13) ***	-0.99 (0.14) ***
<i>Food_Efficiency_{it}</i>		0.19 (0.07) **	0.15 (0.07) *	0.37 (0.09) ***	0.38 (0.09) ***
<i>Food_Quality_{it}</i>		-0.07 (0.06)	-0.06 (0.06)	-0.07 (0.08)	-0.07 (0.09)
<i>Responsiveness_{it}</i>		-0.12 (0.07) ψ	-0.12 (0.00)	0.01 (0.12)	0.02 (0.12)
<i>Atmosphere_{it}</i>		0.05 (0.06)	0.08 (0.06)	0.03 (0.09)	0.02 (0.09)
<i>Pricepoint_i</i>	-0.05 (0.09)	-0.03 (0.09)	-0.02 (0.07)		
<i>WL_{it}</i>	-0.34 (0.07) ***	0.18 (0.21)	0.01 (0.26)		
<i>readability_{it}</i>	-0.02 (0.05)	-0.02 (0.05)	0.01 (0.05)		
<i>comp_meanrating_{it}</i>	-0.05 (0.06)	-0.05 (0.06)	-0.07 (0.05)		
<i>comp_numreviews_{it}</i>	0.48 (0.06) ***	0.48 (0.06) ***	0.56 (0.06) ***		
<i>numcompetitors_{it}</i>	-0.02 (0.07)	-0.02 (0.07)	-0.08 (0.06)		
Cuisines	Included	Included	Included	Included	Included
Location	Included	Included	Included	Included	Included
OtherChars	Included	Included	Included	Included	Included
Restaurant and year fixed effects	Omitted	Omitted	Omitted	Included	Included
n (observations)	16515	16515	4787	16515	16515
groups (restaurants)	1035	1035	437	1035	1035
AIC	3817.7	3565.1	2566.8		
QIC				2863	2286

Table entries are Estimated Values (SE) with significance codes: '***' p < 0.001 '**' p < 0.01 '*' p < 0.05 ' ψ ' p < 0.1

Table 2.10. Cox Proportional Hazards Results (Positive Values Contribute to Closing)

Variable	Estimated Values (SE)
<i>meanrating_{it}</i>	-0.18 (0.04) *
<i>numreviews_{it}</i>	-1.66 (0.16) ***
<i>Quality_Overall_{it}</i>	-0.49 (0.16) **
<i>Food_Efficiency_{it}</i>	0.12 (0.01) *
<i>Food_Quality_{it}</i>	-0.07 (0.06)
<i>Responsiveness_{it}</i>	-0.16 (0.07) ψ
<i>Atmosphere_{it}</i>	-0.02 (0.06)
<i>Pricepoint_i</i>	0.21 (0.07) **
<i>WL_{it}</i>	0.25 (0.14)
<i>readability_{it}</i>	0 (0.02)
<i>comp_meanrating_{it}</i>	-0.01 (0.05)
<i>comp_numreviews_{it}</i>	0.31 (0.07) *
<i>numcompetitors_{it}</i>	0.08 (0.05)
Cuisines	Included
Location	Included
OtherChars	Included
Number of observations: 16515	
Number of events: 454	
Likelihood ratio test: 352.5 on 28 df, p=0	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ψ ' 0.1

Table 2.11. Proportionality Test To Verify Proportionality Assumption

Variable	ρ	χ^2	$Pr(> z)$
<i>meanrating_{it}</i>	-0.06	1.01	0.47
<i>numreviews_{it}</i>	0.03	0.05	0.56
<i>Quality_Overall_{it}</i>	0.01	0.12	0.73
<i>Food_Efficiency_{it}</i>	0.06	0.08	0.18
<i>Food_Quality_{it}</i>	-0.03	0.42	0.52
<i>Responsiveness_{it}</i>	-0.02	0.15	0.73
<i>Atmosphere_{it}</i>	-0.01	0.22	0.76

Table 3.1. Variable Descriptions

Variable	Description
<i>Rating_{it}</i>	Average rating of reviews published during time period <i>t</i> for restaurant <i>i</i>
<i>Deal_{it}</i>	Binary variable indicating whether deal offered restaurant <i>i</i> during time period <i>t</i>
<i>Price_i</i>	Price segment of restaurant <i>i</i> (number of Yelp dollar signs)
<i>Age_i</i>	The number of days from the first review published for restaurant <i>i</i> to Dec. 1, 2011
<i>BaseNumReviews_i</i>	Number of reviews for restaurant <i>i</i> prior to Dec. 1, 2011
<i>BaseRating_i</i>	Rating for restaurant <i>i</i> prior to Dec. 1, 2011
<i>RestInZip_i</i>	Number of competitor restaurants of restaurant <i>i</i> (same cuisine, price segment and zip code)
<i>DealsInZip_{it}</i>	Number of deals being offered by competitors of restaurant <i>i</i> during time period <i>t</i>
<i>Cuisine_i</i>	Binary variables indicating whether each cuisine (16 cuisines in total) is listed in the cuisine type for restaurant <i>i</i> (restaurants may have multiple cuisines)
<i>OtherChars_i</i>	Binary variables describing other restaurant characteristics, such as: payment methods, parking, attire, group-friendly, kid-friendly, waiter, Wi-Fi, alcohol, etc. (15 in total)
<i>Location_i</i>	Categorical variable for the zip code of the restaurant (12 in total)

Table 3.2. Summary Statistics

Variable	Mean (SD)	Range
<i>Rating_{it}</i>	3.43 (1.07)	1, 5
<i>Deal_{it}</i>	0.03 (0.16)	0, 1
<i>Price_i</i>	1.89 (0.72)	1, 4
<i>Age_i</i>	560 (14)	11, 2921
<i>BaseNumReviews_i</i>	66.81 (129.24)	0, 1963
<i>BaseRating_i</i>	3.47 (0.53)	1, 5
<i>RestInZip_i</i>	206.31 (88.27)	86, 330
<i>DealsInZip_{it}</i>	2.43 (2.13)	0, 12

Table 3.3. Correlation Table

	Variable	1	2	3	4	5	6	7	8
1	<i>Rating_{it}</i>	1.00							
2	<i>Deal_{it}</i>	-0.03	1.00						
3	<i>Price_i</i>	0.02	0.03	1.00					
4	<i>Age_i</i>	0.13	0.05	0.01	1.00				
5	<i>BaseNumReviews_i</i>	0.10	0.02	0.15	0.22	1.00			
6	<i>BaseRating_i</i>	0.35	-0.04	0.01	0.02	0.19	1.00		
7	<i>RestInZip_i</i>	0.01	-0.02	-0.05	0.07	0.03	0.01	1.00	
8	<i>DealsInZip_{it}</i>	-0.02	0.08	-0.03	0.05	0.01	-0.01	0.14	1.00

Table 3.4. Multi-level Hierarchical Bayesian Results

	Mean (SD)	HPD (Lower)	HPD (Upper)
First Level			
Intercept	0.845 (0.73)	0.785	0.922
<i>Deal_{it}</i>	-0.902 (0.07)	-1.436	-0.672
<i>BaseNumReviews_i</i>	0.000 (0.00)	-0.001	0.001
<i>BaseRating_i</i>	0.855 (0.02)	0.798	0.986
<i>RestInZip_i</i>	9.245e-5 (1.29e-4)	-1.141e-6	1.054e-4
<i>DealsInZip_{it}</i>	-0.235 (0.04)	-0.368	-0.152
Second Level: <i>Deal_{it}</i>			
Intercept	0.055 (0.08)	0.017	0.222
<i>Price_i</i>	0.539 (0.00)	0.211	0.567
<i>Age_i</i>	-0.665 (0.00)	-0.748	-0.333
Second Level: <i>DealsInZip_{it}</i>			
Intercept	0.111 (0.04)	0.051	0.287
<i>Price_i</i>	0.001 (0.41)	-0.112	0.241
<i>Age_i</i>	-0.034 (0.55)	-0.211	0.099
Cuisines (16 in total)	Included		
OtherChars (15 in total)	Included		
Location (12 in total)	Included		
Sample size (unique restaurants)	N=19,691 (1,390)		
BIC	4783.4		

Table 4.1 Variable Descriptions

Dataset	Variable (s)	Description
Open Data	Boro	NYC Borough
Open Data	Address	Physical address
Open Data	Phone	Phone number
Open Data	Cuisine	Main cuisine declared
Open Data	Inspection Date	The date the inspection took place
Open Data	Inspection Type	They type of inspection (i.e. initial, re-inspection)
Open Data	Inspection Grade	The grade received as a result of the inspection (i.e. A, B, C, or P)
Open Data	Violation Code	The specific health code violation
Open Data	Violation Description	Description of the health code violation
Open Data	Violation Critical	1 if the violation is a critical violation
Open Data	Violation Score	The points added for the violation
Yelp.com	Rating	The overall rating
Yelp.com	Number of Reviews	Total number of reviews
Yelp.com	Price-point	Price range of the restaurant
Yelp.com	OtherChars	Hours, Takes Reservations, Delivery, Take-out, Accepts Credit Cards, Accepts Apple Pay, Good For, Parking Options, Bike Parking, Good for Kids, Good for Groups, Attire, Ambience, Noise Level, Alcohol Beer & Wine, Outdoor Seating, Wi-Fi, Has TV, Waiter Service, Caters

Table 4.2. Summary Statistics for the Difference-in-Differences Model

Measure	Number
Number of restaurants	24,625
Number of inspections	493,804
Mean inspections per restaurant	6.33
Mean violation score per inspection	15.11
Fraction of critical violations per inspection	31.24
Restaurant characteristics:	
Mean Pricepoint	1.84
Mean Yelp Rating	3.54
Mean Yelp Number of Reviews	103

Table 4.3. Difference-in-Differences Results

Variable	Estimate (SE)
Treat (Initial Inspection)	-6.23 (1.44) ***
Rating	-5.22 (3.01) *
Reviews	0.0001 (0.00007) ψ
Price (low-1-2)	0.0003 (0.0002) ψ
Price (high 3-4)	0.004 (0.003) ψ
Treat*price (low=1-2)	-0.15 (0.08) *
Treat*price (high=3-4)	1.47 (0.61) **
OtherChars	Included
Groups	23,393
Observations	350,895
Other characteristics	Included
Auto Correlation Standard Errors	Included
Robust Standard Errors	Included
Heteroskedasticity-Consistent Standard Errors	Included
Fixed Effects	Restaurant, Inspector, time (yr-qt)

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ψ ' 0.1

Table 4.4. Linear Mixed Effects Model Results. DV=Inspection Scores

	Model 1	Model 2
Variable	Estimate (SE)	Estimate (SE)
Intercept	0.73 (0.51) ψ	0.81 (0.68)
WC	0.46 (0.16) **	0.51 (0.17) **
Rating	0.43 (0.15) **	0.37 (0.11) ***
Reviews	0.008 (0.0055) ψ	0.007 (0.005) ψ
Pricepoint	0.12 (0.11)	0.55 (0.45)
OtherChars	Included	Included
AIC	10,666	8,371
BIC	10,639	8.338
Groups	23,393	23,393
Observations	1,052,685	1,052,685
Fixed Effects	Restaurant	Restaurant, Inspector, Time (yr-month)

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ψ ' 0.1

Table 4.5. Linear Mixed Effects Model Results. DV=SMASH WC Scores 90 Days After Inspection

	Model 1	Model 2
Variable	Estimate (SE)	Estimate (SE)
Intercept	0.98 (0.69) ψ	0.66 (0.41) ψ
Offset	0.0005 (0.001)	0.01 (0.0065) ψ
Offset*AA	0.01 (0.0055) *	0.01 (0.004) **
Offset*PAPA	0.73 (0.21) ***	0.80 (0.33) **
Rating	0.65 (1.22)	0.72 (0.91)
Reviews	-0.16 (0.44)	-0.65 (0.83)
Pricepoint	0.10 (0.85)	0.24 (0.64)
OtherChars	Included	Included
AIC	13,291	11,590
BIC	13,349	11,611
Groups	23,393	23,393
Observations	1,579,028	1,579,028
Fixed Effects	Restaurant	Restaurant, Reviewer

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ψ ' 0.1

Figures

Figure 2.1. These plots depict the Schoenfeld residuals by the mean rating, the number of reviews, and the semantic variables **Quality_Overall** and **Food_Efficiency**. A non-zero slope is evidence against proportionality.

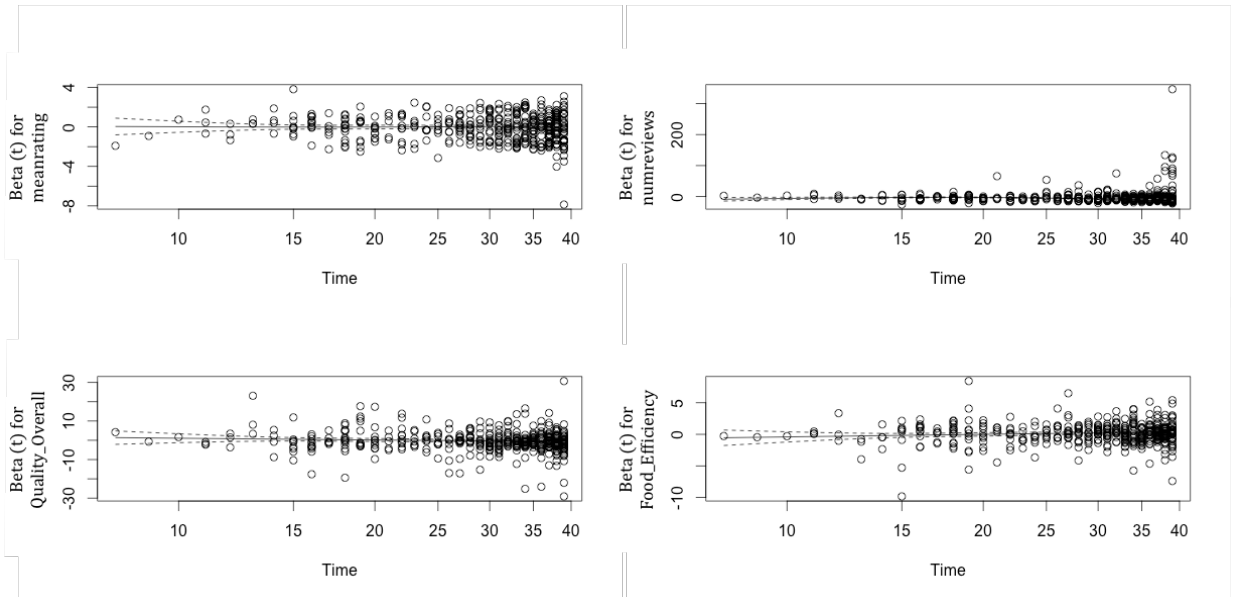


Figure 2.2. Cross-Validation and Singular Values for LSA

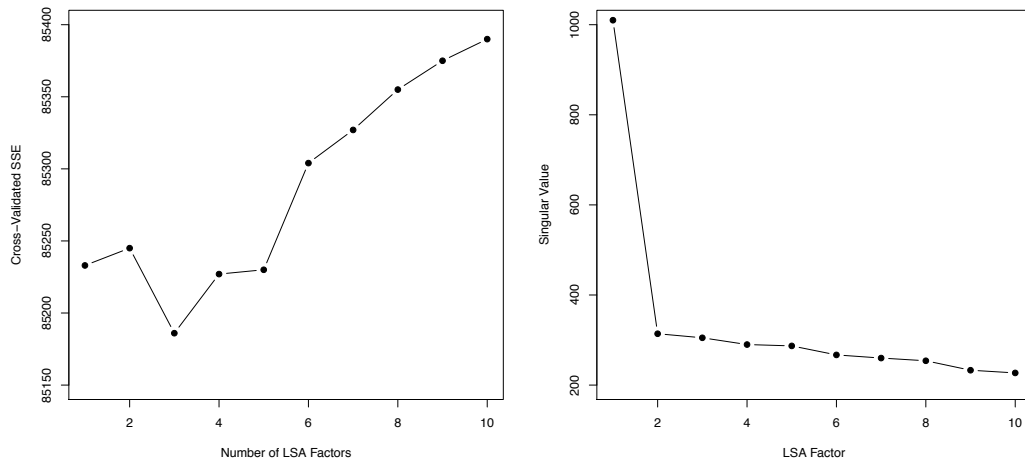


Figure 2.3. ROC Curves for GLMER models

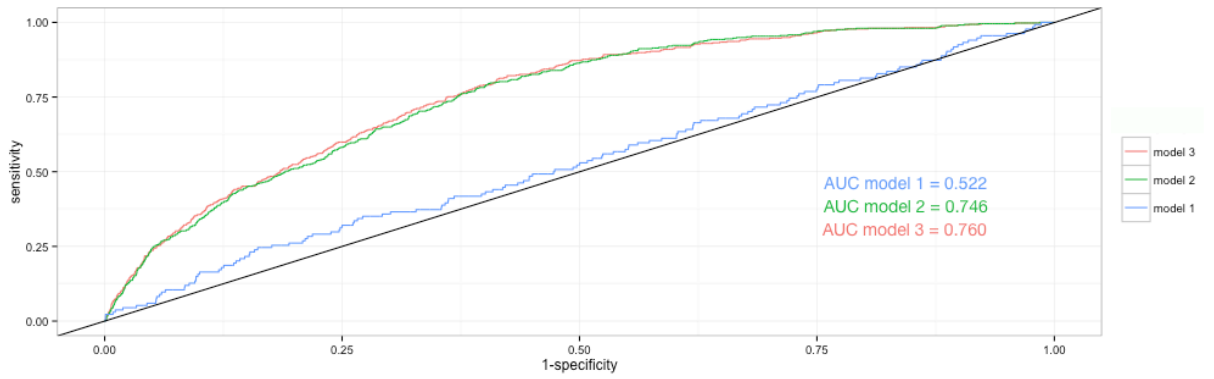


Figure 3.1. Daily Deals Effects on eWOM: Timeline

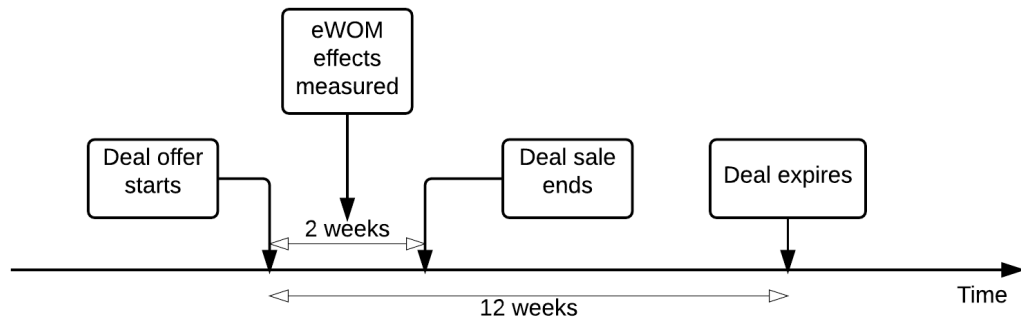


Figure 3.2. Mean Yelp Rating by Offset from Daily Deal Offer Date

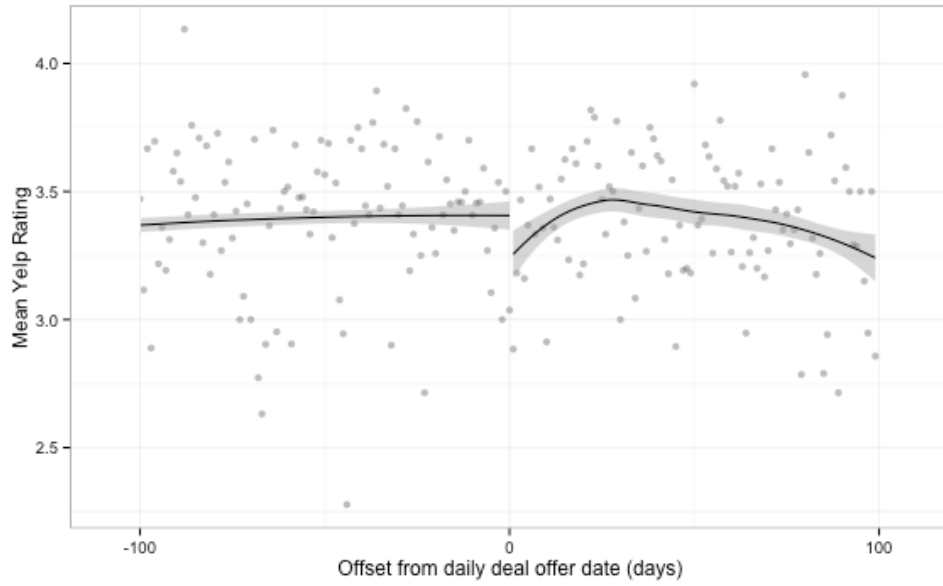


Figure 3.3. Lab Studies Stimuli Examples

Study 1 Deal-High Priced Cell

Study 2 Deal-New Cell

Study 1 Deal-High Priced Cell:

\$25 for \$50 Deal at Italian Kitchen

- Limit 1 voucher(s) per table.
- Not valid for delivery.

Reg. Price	Discount	Savings
\$50	50%	\$25

Study 2 Deal-New Cell:

\$10 for \$20 Deal at Italian Kitchen

- Limit 1 voucher(s) per table.
- Not valid for delivery.

Reg. Price	Discount	Savings
\$20	50%	\$10

Study 3 Deal-Competition

Study 3 Deal-Competition:

\$10 for \$20 Deal at Italian Kitchen

- Limit 1 voucher(s) per table.
- Not valid for delivery.

Reg. Price	Discount	Savings
\$20	50%	\$10

There are 20+ deals for similar restaurants in this area.

Figure 3.4. Brand Evaluation as a Function of Deal Offered and the Price of Merchants

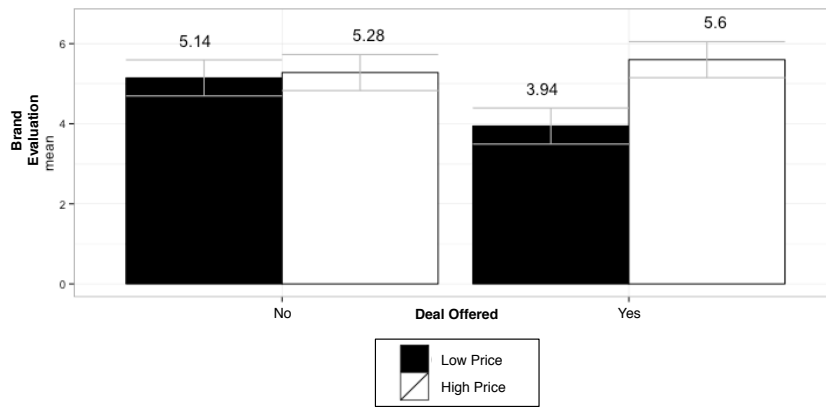


Figure 3.5. Brand Evaluation as a Function of Deals Offered and the Age of Merchants

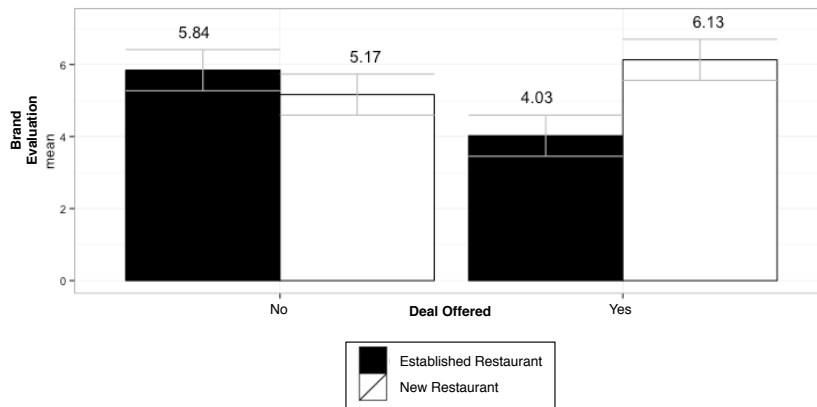


Figure 3.6. Brand Evaluation as a Function of Deal Offered and Deal Competition

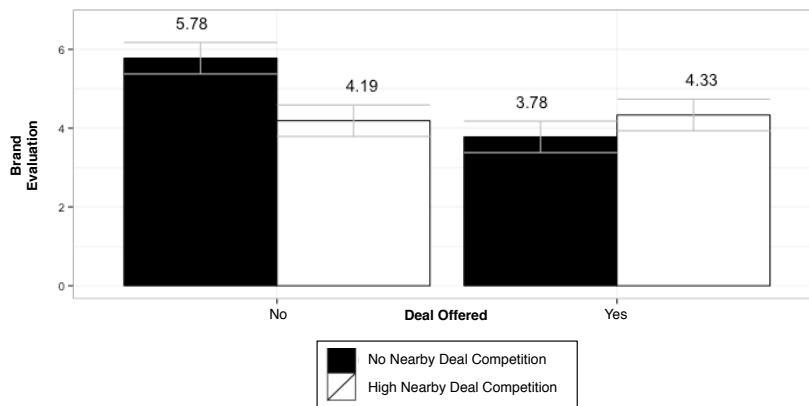


Figure 4.1. NYC Restaurant Hygiene Cards



Figure 4.2. Placement of Grade Card in a NYC Restaurant



Figure 4.3. Grades over time (Before and After the NYC Grading Program)

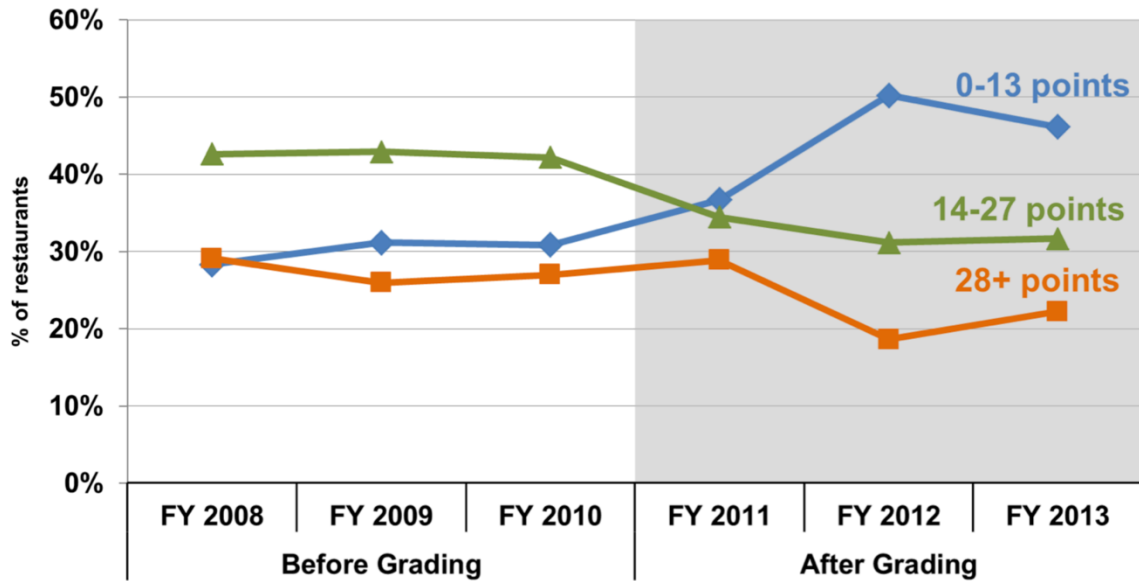


Figure 4.4. Restaurant with A Grades by Borough in 2011, 2013, and 2015 in NYC

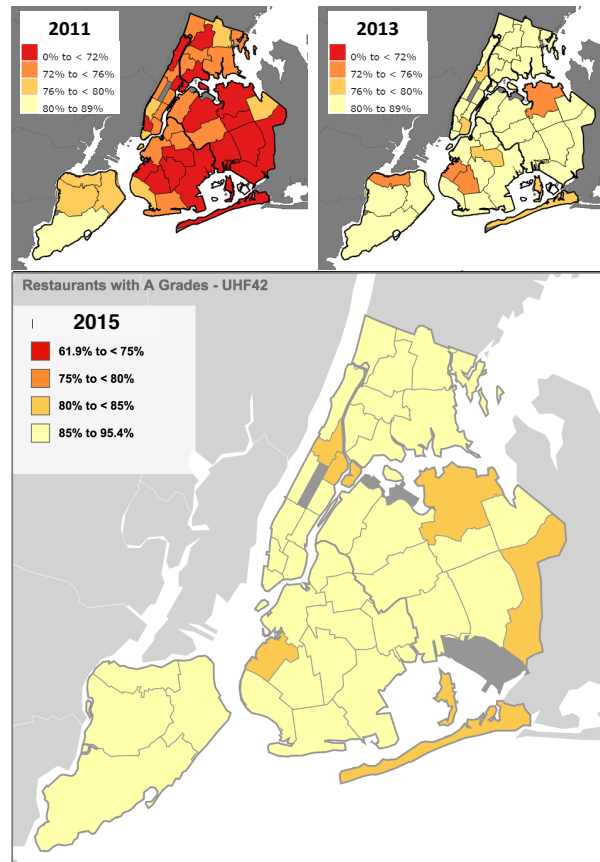


Figure 4.5. Restaurant with A Grades by Borough from 2011 to 2015

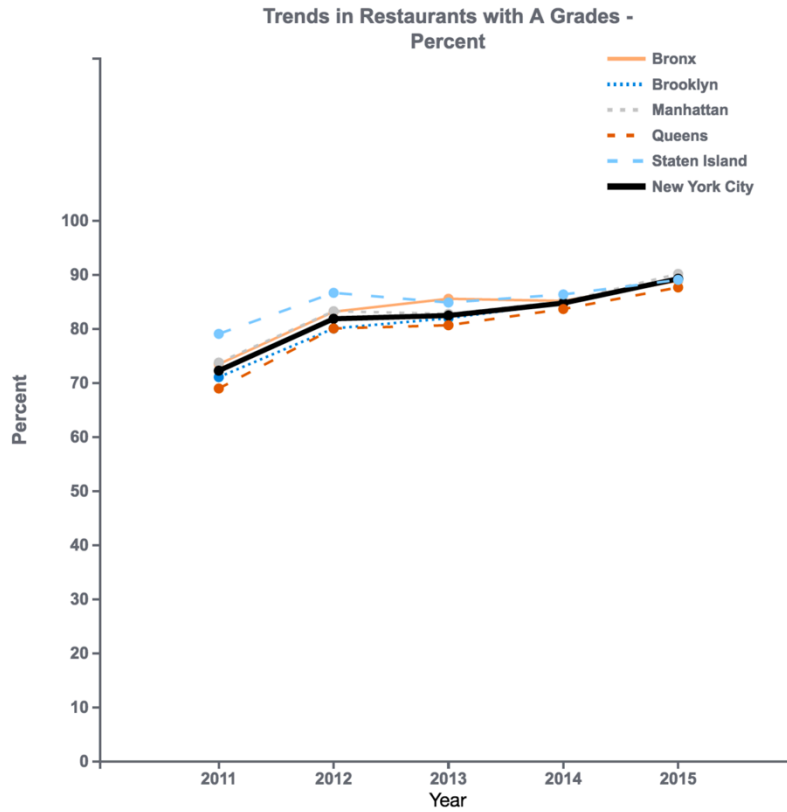


Figure 4.6. Score Trajectories from Restaurants that Consistently Score A in their Initial Inspections Versus those that Post a Grade P in their Initial Inspection

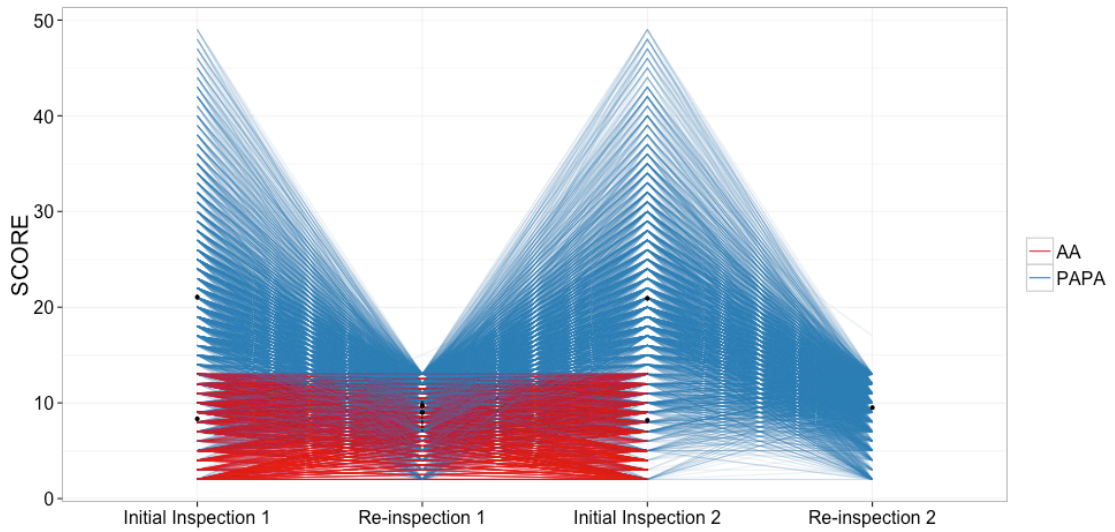


Figure 4.7. Average Scores of the Two Main Inspection Trajectories

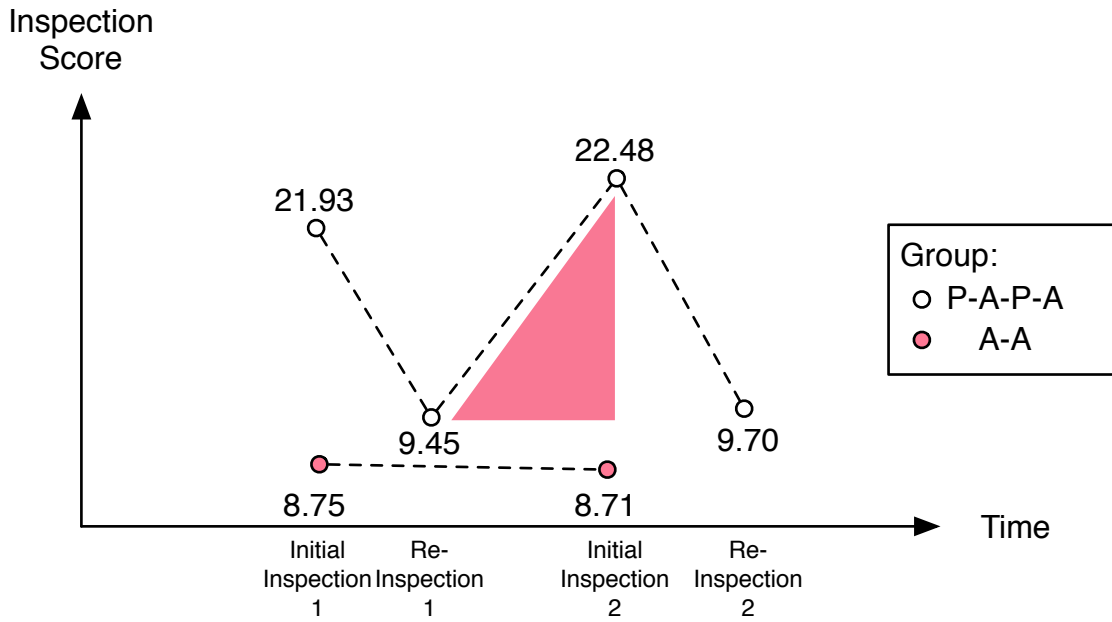
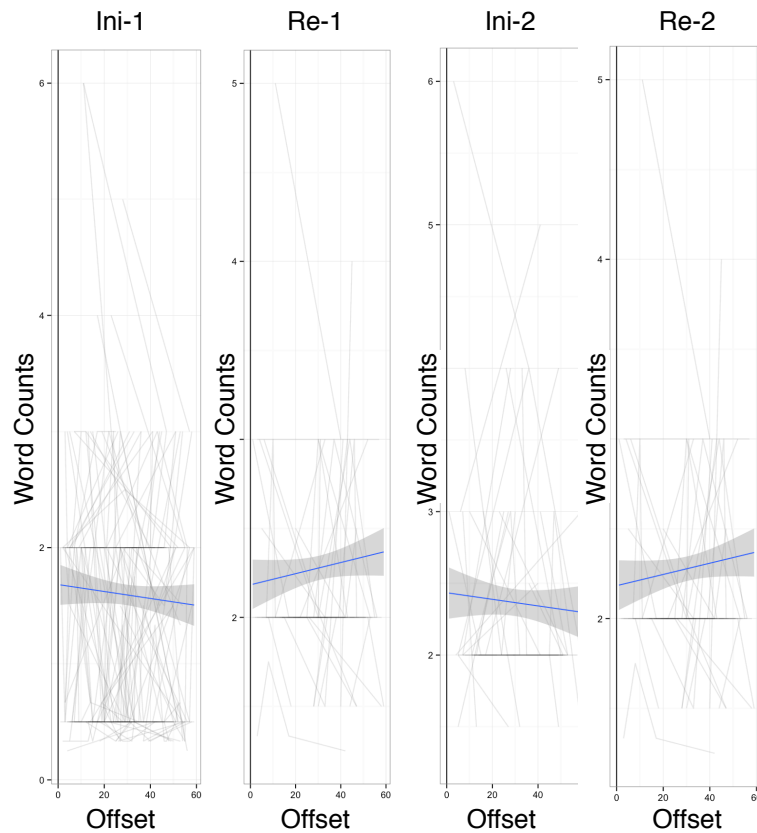


Figure 4.8. Trends in SMASH Score After Different Types of Inspections for PAPA



Appendix

Figure A-1. Restaurant with A Grades by Neighborhood Poverty Rates in New York City

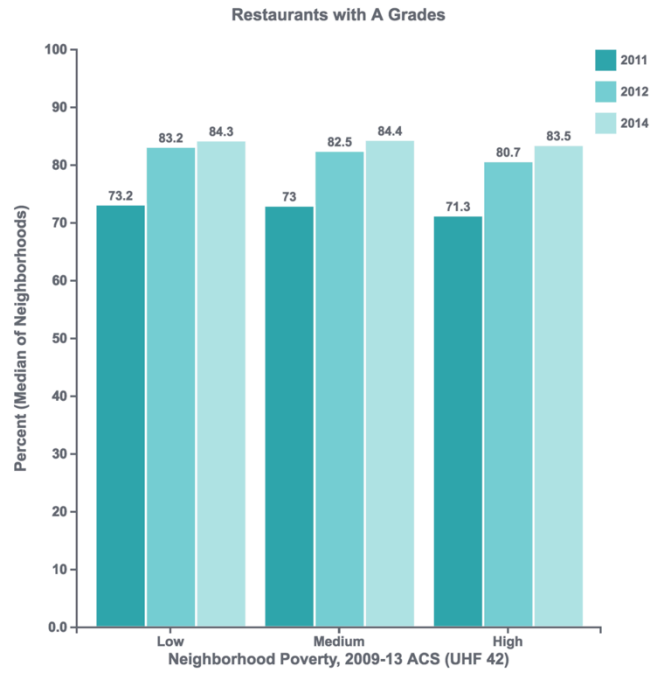


Figure A-2. Restaurant Grades (Including Grade P) Over NYC's Map

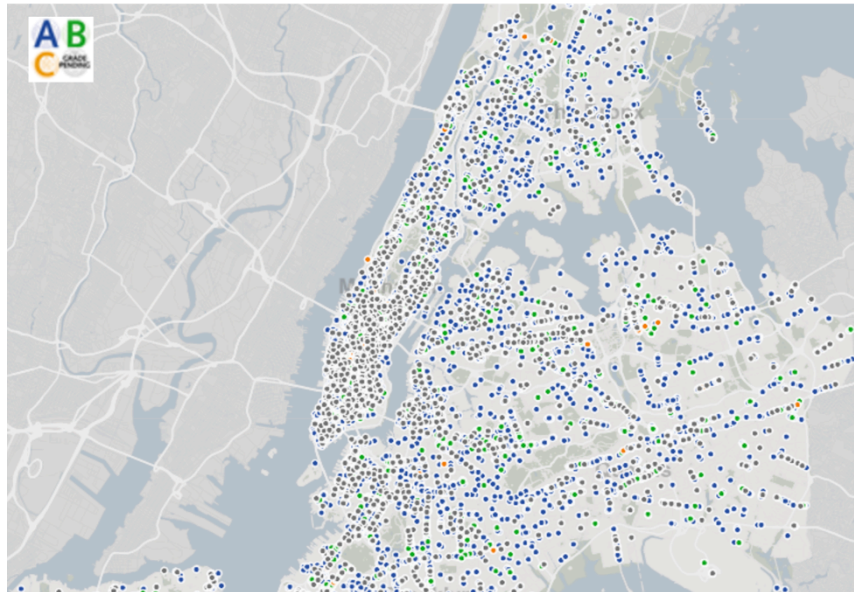


Figure A-3. Inspector Worksheet Showing Violations and Conditions (DOHM 2010b)

Rev Dec. 2010



Bureau of Food Safety and Community Sanitation

Form D

SELF-INSPECTION WORKSHEET FOR FOOD SERVICE ESTABLISHMENTS

CRITICAL VIOLATIONS		CONDITIONS					SCORE
		I	II	III	IV	V	
FOOD TEMPERATURE							
2A*	Food not cooked to required minimum temperature: • Poultry, meat stuffing, stuffed meats ≥ 165°F for 15 seconds • Ground meat and food containing ground meat ≥ 158°F for 15 seconds • Pork, any food containing pork ≥ 155°F for 15 seconds • Rare roast beef, rare beefsteak except per individual customer request ≥ required temperature and time • All other foods except shell eggs per individual customer request ≥ 145°F for 15 seconds	-	-	-	10	28	
2B*	Hot food item not held at or above 140°F.	7	8	9	10	28	
2C	Hot food item that has been cooked and refrigerated is being held for service without first being reheated to 165°F or above within 2 hours.	5	6	7	8	-	
2D	Precooked potentially hazardous food from commercial food processing establishment that is supposed to be heated, but is not heated to 140°F within 2 hours.	5	6	7	8	-	
2E	Whole frozen poultry or poultry breasts, other than a single portion, is being cooked frozen or partially thawed.	5	6	-	-	-	
2F	Meat, fish or molluscan shellfish served raw or undercooked without prior notification to customer.	-	-	-	8	-	
2G*	Cold food item held above 41°F (smoked fish and reduced oxygen packaged foods above 38°F) except during necessary preparation.	7	8	9	10	28	
2H*	Food not cooled by an approved method whereby the internal product temperature is reduced from 140° F to 70° F or less within 2 hours, and from 70°F to 41°F or less within 4 additional hours.	7	8	9	10	28	
2I	Food prepared from ingredients at ambient temperature not cooled to 41°F or below within 4 hours.	5	6	7	8	-	
2J*	Reduced oxygen packaged (ROP) foods not cooled by an approved method whereby the internal food temperature is reduced to 38°F within two hours of cooking and if necessary further cooled to a temperature of 34°F within six hours of reaching 38°F.	7	8	9	10	28	
FOOD SOURCE							
3A*	Food from unapproved or unknown source or home canned. Reduced oxygen packaged (ROP) fish not frozen before processing; or ROP foods prepared on premises transported to another site.	-	-	-	10	28	
3B*	Shellfish not from approved source, improperly tagged/labeled; tags not retained for 90 days.	-	-	-	10	28	
3C*	Eggs found dirty/cracked; liquid, frozen or powdered eggs not pasteurized.	7	8	9	10	28	
3D*	Canned food product observed swollen, leaking or rusted, and not segregated from other consumable food items.	7	8	9	10	28	
3E*	Potable water supply inadequate. Water or ice not potable or from unapproved source. Cross connection in potable water supply system observed.	-	-	-	10	28	
3F*	Unpasteurized milk or milk product present.	-	-	-	10	28	
3G	Raw food not properly washed prior to serving.	5	6	7	8	-	
FOOD PROTECTION							
4A	Food Protection Certificate not held by supervisor of food operations.	-	-	-	-	10	
4B*	Food worker prepares food or handles utensil when ill with a disease transmissible by food, or has exposed infected cut or burn on hand.	-	-	-	10	28	
4C*	Food worker does not use proper utensil to eliminate bare hand contact with food that will not receive adequate additional heat treatment.	7	8	9	10	28	
4D*	Food worker does not wash hands thoroughly after using the toilet, coughing, sneezing, smoking, eating, preparing raw foods or otherwise contaminating hands.	-	-	-	10	28	
4E*	Toxic chemical improperly labeled, stored or used such that food contamination may occur.	7	8	9	10	28	
4F*	Food, food preparation area, food storage area, area used by employees or patrons contaminated by sewage or liquid waste.	-	-	-	10	28	
4G*	Unprotected potentially hazardous food re-served.	-	-	-	10	28	
4H*	Raw, cooked or prepared food is adulterated, contaminated, cross-contaminated or not discarded in accordance with HACCP plan.	7	8	9	10	28	
4I	Unprotected food re-served.	5	6	7	8	-	

* Public Health Hazards (PHH) must be corrected immediately

+ Pre-permit Serious (PPS) Violations that must be corrected before permit is issued

Bibliography

- Abdullah, F. 2006. "Measuring Service Quality in Higher Education: Hedperf Versus Servperf," *Marketing Intelligence & Planning* (24:1), pp. 31-47.
- Ackelof, G. A. 1970. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* (84).
- Alvarez Alvarez, B., and Vázquez Casielles, R. 2005. "Consumer Evaluations of Sales Promotion: The Effect on Brand Choice," *European Journal of Marketing* (39:1/2), pp. 54-70.
- Angulo, F. J., and Jones, T. F. 2006. "Eating in Restaurants: A Risk Factor for Foodborne Disease?," *Clinical Infectious Diseases* (43:10), pp. 1324-1328.
- Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485-1509.
- Arora, S., Ge, R., and Moitra, A. 2012. "Learning Topic Models-Going Beyond Svd," in: *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*. IEEE, pp. 1-10.
- Athey, S. 2015. "Machine Learning and Causal Inference for Policy Evaluation," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: ACM*, pp. 5-6.
- Athey, S., and Imbens, G. W. 2006. "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica* (74:2), pp. 431-497.
- Athey, S., and Imbens, G. 2015. "A Measure of Robustness to Misspecification," *The American Economic Review* (105:5), pp. 476-480.

- Author, D. H. 2003. "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing," *Journal of Labor Economics*, January).
- Awad, N. F., and Ragowsky, A. 2008. "Establishing Trust in Electronic Commerce through Online Word of Mouth: An Examination across Genders," *Journal of Management Information Systems* (24:4), pp. 101-121.
- Baglama, J., and Reichel, L. 2014. "Irlba: Fast Partial Svd by Implicitly-Restarted Lanczos Bidiagonalization."
- Bates, T. 1990. "Entrepreneur Human Capital Inputs and Small Business Longevity," *The Review of Economics and Statistics*), pp. 551-559.
- Bayus, B. L., and Agarwal, R. 2007. "The Role of Pre-Entry Experience, Entry Timing, and Product Technology Strategies in Explaining Firm Survival," *Management Science* (53:12), pp. 1887-1902.
- Bell, D. R., and Lattin, J. M. 2000. "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity," *Marketing Science* (19:2), pp. 185-200.
- Bensoussan, A., Mookerjee, R., Mookerjee, V., and Yue, W. T. 2009. "Maintaining Diagnostic Knowledge-Based Systems: A Control-Theoretic Approach," *Management Science* (55:2), pp. 294-310.
- Bertrand, M., Duflo, E., and Mullainathan, S. 2002. "How Much Should We Trust Differences-in-Differences Estimates?," National Bureau of Economic Research.
- Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., and Venkatraman, N. 2013. "Digital Business Strategy: Toward a Next Generation of Insights," *Mis Quarterly* (37:2), pp. 471-482.
- Biswas, A., Bhowmick, S., Guha, A., and Grewal, D. 2013. "Consumer Evaluations of Sale

- Prices: Role of the Subtraction Principle," *Journal of Marketing* (77:4), pp. 49-66.
- Blattberg, R. C., Briesch, R., and Fox, E. J. 1995. "How Promotions Work," *Marketing science* (14:3_supplement), pp. G122-G132.
- Blattberg, R. C., and Neslin, S. A. 1990. *Sales Promotion: Concepts, Methods, and Strategies*. Prentice Hall Englewood Cliffs, NJ.
- Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), pp. 77-84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *The Journal of Machine Learning Research* (3), pp. 993-1022.
- Bloomberg. 2015. "Improve Bloomberg Visual Data." 2016, from <http://www.bloomberg.com/visual-data/best-and-worst//most-eateries-per-capita-us-cities>
- BrightLocal. 2014. "Local Consumer Review Survey 2014."
- Brüderl, J., Preisendörfer, P., and Ziegler, R. 1992. "Survival Chances of Newly Founded Business Organizations," *American Sociological Review*, pp. 227-242.
- Buchholz, U., Run, G., Kool, J., Fielding, J., and Mascola, L. 2002. "A Risk-Based Restaurant Inspection System in Los Angeles County," *Journal of Food Protection*® (65:2), pp. 367-372.
- Buhrmester, M., Kwang, T., and Gosling, S. D. 2011. "Amazon's Mechanical Turk a New Source of Inexpensive, yet High-Quality, Data?," *Perspectives on psychological science* (6:1), pp. 3-5.
- Buttle, F. A. 1998. "Word of Mouth: Understanding and Managing Referral Marketing," *Journal of Strategic Marketing* (6:3), pp. 241-254.

- Buzzell, R. D., and Gale, B. T. 1987. *The Pims Principles: Linking Strategy to Performance*. Simon and Schuster.
- Byers, J. W., Mitzenmacher, M., and Zervas, G. 2012. "Daily Deals: Prediction, Social Diffusion, and Reputational Ramifications," *Proceedings of the fifth ACM international conference on Web search and data mining*: ACM, pp. 543-552.
- Byers, J. W., Mitzenmacher, M., and Zervas, G. 2012. "The Groupon Effect on Yelp Ratings: A Root Cause Analysis," *Proceedings of the 13th ACM conference on electronic commerce*: ACM, pp. 248-265.
- Cambria, E., Schuller, B., Xia, Y., and Havasi, C. 2013. "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems*:2), pp. 15-21.
- Cao, Q., Duan, W., and Gan, Q. 2011. "Exploring Determinants of Voting for the ``Helpfulness" of Online User Reviews: A Text Mining Approach," *Decision Support Systems* (50:2), pp. 511-521.
- Card, D., and Krueger, A. B. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply," *American Economic Review*), pp. 1397-1420.
- Carman, J. M. 1990. "Consumer Perceptions of Service Quality: An Assessment of T," *Journal of Retailing* (66:1), p. 33.
- Carroll, G. R., and Khessina, O. M. 2005. *The Ecology of Entrepreneurship*. Springer.
- Carter, N. M., Stearns, T. M., Reynolds, P. D., and Miller, B. A. 1994. "New Venture Strategies: Theory Development with an Empirical Base," *Strategic Management Journal* (15:1), pp. 21-41.
- Castrogiovanni, G. J. 1991. "Environmental Munihcence; a Theoretical Assessment,"

Academy of Management Review (16:3), pp. 542-565.

- Centers for Disease Control (CDC). 2016a. "Foodborne Germs and Illnesses." Retrieved April 10, 2016, <http://www.cdc.gov/foodborneburden/trends-in-foodborne-illness.html>.
- Centers for Disease Control (CDC). 2016b. "Trends in Foodborne Illness." Retrieved April 10, 2016, <http://www.cdc.gov/foodborneburden/trends-in-foodborne-illness.html>.
- Chase, R. B., and Apte, U. M. 2007. "A History of Research in Service Operations: What's the Big Idea?," *Journal of Operations Management* (25:2), pp. 375-386.
- Chen, Y., and Xie, J. 2008. "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," *Management Science* (54:3), pp. 477-491.
- Chen, Y.-J., and Kirmani, A. 2015. "Posting Strategically: The Consumer as an Online Media Planner," *Journal of Consumer Psychology*.
- Chen, Z., and Lurie, N. H. 2013. "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research* (50:4), pp. 463-476.
- Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of marketing research* (43:3), pp. 345-354.
- City University of New York (CUNY). 2011. Restaurant Letter Grading: The First Year.
- City University of New York (CUNY). 2012. Restaurant Grading in New York City at 18 Months.
- Clark, J. R., Huckman, R. S., and Staats, B. R. 2013. "Learning from Customers: Individual and Organizational Effects in Outsourced Radiological Services," *Organization Science* (24:5), pp. 1539-1557.
- Clemons, E. K., Gao, G. G., and Hitt, L. M. 2006. "When Online Reviews Meet

- Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems* (23:2), pp. 149-171.
- Cox, D. R. 1972. "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187-220.
- Cronin, J. J., Jr, and Taylor, S. A. 1992. "Measuring Service Quality: A Reexamination and Extension," *The Journal of Marketing*), pp. 55-68.
- Cronin, J. J., Jr, and Taylor, S. A. 1994. "Servperf Versus Servqual: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *The Journal of Marketing*), pp. 125-131.
- Dai, W., Jin, G. Z., Lee, J., and Luca, M. 2012. "Optimal Aggregation of Consumer Ratings: An Application to Yelp. Com," National Bureau of Economic Research.
- Davenport, T. H., and Beck, J. C. 2013. *The Attention Economy: Understanding the New Currency of Business*. Harvard Business Press.
- Decker, R., and Trusov, M. 2010. "Estimating Aggregate Consumer Preferences from Online Product Reviews," *International Journal of Research in Marketing* (27:4), pp. 293-307.
- Della Bitta, A. J., Monroe, K. B., and McGinnis, J. M. 1981. "Consumer Perceptions of Comparative Price Advertisements," *Journal of Marketing Research*), pp. 416-427.
- Dellarocas, C. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* (49:10), pp. 1407-1424.
- Dellarocas, C., and Narayan, R. 2006. "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science* (21:2), pp. 277-285.

- Dellarocas, C., Zhang, X. M., and Awad, N. F. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive marketing* (21:4), pp. 23-45.
- Department of Health and Mental Hygiene of New York City (DOHM). 2010a. What to Expect When You're Inspected.
- Department of Health and Mental Hygiene of New York City (DOHM). 2010b. Inspection Worksheet.
- Department of Health and Mental Hygiene of New York City (DOHM). 2011. Guide to Condition Levels.
- Department of Health and Mental Hygiene of New York City (DOHM). 2012a. How We Score and Grade.
- Department of Health and Mental Hygiene of New York City (DOHM). 2012b. Enforcement Guidelines.
- Department of Health and Mental Hygiene of New York City (DOHM). 2013. Letter Grading for Restaurants.
- Department of Health and Mental Hygiene of New York City (DOHM). 2014. "Letter Grading Workshop." Retrieved April 10, 2016, <http://www.nyc.gov/html/doh/html/environmental/food-service-grading.shtml>.
- Department of Health and Mental Hygiene of New York City (DOHM). 2015. Practicing A-Grade Food Safety.
- Department of Health and Mental Hygiene of New York City (DOHM). 2016. "Health Code & Rules." <https://www1.nyc.gov/site/doh/about/about-doh/health-code-and-rules.page>.

- Dholakia, U. M. 2010. "How Effective Are Groupon Promotions for Businesses?," *Available at SSRN 1696327*).
- Dholakia, U. M. 2011. "How Businesses Fare with Daily Deals: A Multi-Site Analysis of Groupon, Livingsocial, Opentable, Travelzoo, and Buywithme Promotions," *Livingsocial, Opentable, Travelzoo, and BuyWithMe Promotions (June 13, 2011)*).
- Dholakia, U. M. 2011. "What Makes Groupon Promotions Profitable for Businesses?," *Available at SSRN 1790414*).
- Dholakia, U. M. 2012. "How Businesses Fare with Daily Deals as They Gain Experience: A Multi-Time Period Study of Daily Deal Performance," *Available at SSRN 2091655*).
- Dholakia, U. M., and Kimes, S. E. 2011. "Daily Deal Fatigue or Unabated Enthusiasm? A Study of Consumer Perceptions of Daily Deal Promotions," *A Study of Consumer Perceptions of Daily Deal Promotions (September 11, 2011)*).
- Dholakia, U. M., and Tsabar, G. 2011. "A Startup's Experience with Running a Groupon Promotion," *Available at SSRN 1828003*).
- Dickinger, A., and Mazanec, J. 2008. "Consumers' Preferred Criteria for Hotel Online Booking," *Information and Communication Technologies in Tourism 2008*), pp. 244-254.
- Dimoka, A., Hong, Y., and Pavlou, P. A. 2012. "On Product Uncertainty in Online Markets: Theory and Evidence," *Mis Quarterly* (36).
- DrivenData. 2016. "Competition: Keeping It Fresh: Predict Restaurant Inspections." from <https://www.drivendata.org/competitions/5/>
- Duan, W., Gu, B., and Whinston, A. B. 2008. "Do Online Reviews Matter?—an Empirical Investigation of Panel Data," *Decision support systems* (45:4), pp. 1007-1016.

- Duan, W., Gu, B., and Whinston, A. B. 2008. "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems* (45:4), pp. 1007-1016.
- Ebben, J. J., and Johnson, A. C. 2005. "Efficiency, Flexibility, or Both? Evidence Linking Strategy to Performance in Small Firms," *Strategic management journal* (26:13), pp. 1249-1259.
- Eckart, C., and Young, G. 1936. "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika* (1:3), pp. 211-218.
- Edelman, B., Jaffe, S., and Kominers, S. D. 2011. "To Groupon or Not to Groupon: The Profitability of Deep Discounts," *Marketing Letters*, pp. 1-15.
- Egan, M., Raats, M., Grubb, S., Eves, A., Lumbers, M., Dean, M., and Adams, M. 2007. "A Review of Food Safety and Food Hygiene Training Studies in the Commercial Sector," *Food Control* (18:10), pp. 1180-1190.
- Farahat, A., Ahmed, N. K., and Dholakia, U. 2012. "Does a Daily Deal Promotion Signal a Distressed Business? An Empirical Investigation of Small Business Survival," *An Empirical Investigation of Small Business Survival* (November 6, 2012).
- Farley, T. 2012. "Testimony of the Commissioner of New York City Department of Health and Mental Hygiene before the New York City Council Committees on Health, Governmental Operations, Oversight and Investigations, and Small Business on Restaurant Grading."
- Feinerer, I., and Hornik, K. 2012. "Tm: Text Mining Package," *R package version 0.5-7.1*.
- Feldman, R. 2013. "Techniques and Applications for Sentiment Analysis," *Communications of the ACM* (56:4), pp. 82-89.
- Fiske, S. T., and Taylor, S. E. 2013. *Social Cognition: From Brains to Culture*. Sage.

- Forman, C., Ghose, A., and Goldfarb, A. 2009. "Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science* (55:1), pp. 47-57.
- Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.
- Friestad, M., and Wright, P. 1994. "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts," *Journal of consumer research*, pp. 1-31.
- Gao, G. G., McCullough, J. S., Agarwal, R., and Jha, A. K. 2012. "A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians over a 5-Year Period," *Journal of Medical Internet Research* (14:1), p. e38.
- Garvin, D. A. 1988. *Managing Quality: The Strategic and Competitive Edge*. Simon and Schuster.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2014. *Bayesian Data Analysis*. Taylor & Francis.
- Ghose, A., and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *Knowledge and Data Engineering, IEEE Transactions on* (23:10), pp. 1498-1512.
- Glynn Mangold, W., Miller, F., and Brockway, G. R. 1999. "Word-of-Mouth Communication in the Service Marketplace," *Journal of Services Marketing* (13:1), pp. 73-89.
- Go, A., Bhayani, R., and Huang, L. 2009. "Twitter Sentiment Classification Using Distant Supervision," *CS224N Project Report, Stanford* (1), p. 12.

- Godes, D., and Mayzlin, D. 2004. "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science* (23:4), pp. 545-560.
- Goldlücke, S., and Schmitz, P. W. 2014. "Investments as Signals of Outside Options," *Journal of Economic Theory* (150), pp. 683-708.
- Goldstein, S. M., Johnston, R., Duffy, J., and Rao, J. 2002. "The Service Concept: The Missing Link in Service Design Research?," *Journal of Operations Management* (20:2), pp. 121-134.
- Goodman, J. K., Cryder, C. E., and Cheema, A. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples," *Journal of Behavioral Decision Making* (26:3), pp. 213-224.
- Gould, L. H., Rosenblum, I., Nicholas, D., Phan, Q., and Jones, T. F. 2013. "Contributing Factors in Restaurant-Associated Foodborne Disease Outbreaks, Foodnet Sites, 2006 and 2007," *Journal of food protection* (76:11), p. 1824.
- Grambsch, P. M., and Therneau, T. M. 1994. "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals," *Biometrika* (81:3), pp. 515-526.
- Grewal, D., Krishnan, R., Baker, J., and Borin, N. 1998. "The Effect of Store Name, Brand Name and Price Discounts on Consumers' Evaluations and Purchase Intentions," *Journal of retailing* (74:3), pp. 331-352.
- Grönroos, C. 1990. "Marketing Redefined," *Management Decision* (28:8).
- Gu, B., Park, J., and Konana, P. 2012. "Research Note-the Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products," *Information Systems Research* (23:1), pp. 182-196.
- Gu, B., and Ye, Q. 2014. "First Step in Social Media: Measuring the Influence of Online

- Management Responses on Customer Satisfaction," *Production and Operations Management* (23:4), pp. 570-582.
- Guadagni, P. M., and Little, J. D. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing science* (2:3), pp. 203-238.
- Hand, D. J., Mannila, H., and Smyth, P. 2001. *Principles of Data Mining*. MIT press.
- Harrison, C., Jorder, M., Stern, H., Stavinsky, F., Reddy, V., Hanson, H., Waechter, H., Lowe, L., Gravano, L., and Balter, S. 2014. "Using Online Reviews by Restaurant Patrons to Identify Unreported Cases of Foodborne Illness—New York City, 2012–2013," *MMWR* (63:20), pp. 441-445.
- Hart, C. W. 1988. *The Power of Unconditional Service Guarantees*. Harvard Business Review, Reprint Service.
- Hart, C. W., Heskett, J. L., and Sasser, W. E., Jr. 1989. "The Profitable Art of Service Recovery," *Harvard Business Review* (68:4), pp. 148-156.
- Heagerty, P. J., Lumley, T., and Pepe, M. S. 2000. "Time-Dependent Roc Curves for Censored Survival Data and a Diagnostic Marker," *Biometrics*), pp. 337-344.
- Heagerty, P. J., and Zheng, Y. 2005. "Survival Model Predictive Accuracy and Roc Curves," *Biometrics* (61:1), pp. 92-105.
- Hedberg, C. W., Smith, S. J., Kirkland, E., Radke, V., Jones, T. F., Selman, C. A., and Group, E.-N. W. 2006. "Systematic Environmental Evaluations to Identify Food Safety Differences between Outbreak and Nonoutbreak Restaurants," *Journal of Food Protection*® (69:11), pp. 2697-2702.
- Heineke, J., and Davis, M. M. 2007. "The Emergence of Service Operations Management as an Academic Discipline," *Journal of Operations Management* (25:2), pp. 364-374.

- Heskett, J. L., Schlesinger, L., and others. 1994. "Putting the Service-Profit Chain to Work," *Harvard Business Review* (72:2), pp. 164-174.
- Ho, D. E. 2012. "Fudging the Nudge: Information Disclosure and Restaurant Grading," *Yale LJ* (122), p. 574.
- Hofmann, T. 1999. "Probabilistic Latent Semantic Indexing," in: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 50-57.
- Hosmer, D. W., Jr, Lemeshow, S., and Sturdivant, R. X. 2013. *Model-Building Strategies and Methods for Logistic Regression*. John Wiley & Sons, Inc.
- Hu, M., and Liu, B. 2004. "Mining and Summarizing Customer Reviews," *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*: ACM, pp. 168-177.
- Hu, N., Pavlou, P. A., and Zhang, J. 2006. "Can Online Reviews Reveal a Product's True Quality?: Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication," *Proceedings of the 7th ACM conference on Electronic commerce*: ACM, pp. 324-330.
- Iacus, S. M., King, G., and Porro, G. 2009. "CEM: Software for Coarsened Exact Matching," *Journal of Statistical Software* (30:9), pp. 1-27.
- Iacus, S. M., King, G., and Porro, G. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding," *Journal of the American Statistical Association* (106:493), pp. 345-361.
- Iacus, S. M., King, G., and Porro, G. 2011. "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*.

- Imbens, G., and Wooldridge, J. 2007. "Difference-in-Differences Estimation."
- Imbens, G., and Wooldridge, J. 2008. "Recent Developments in Program Evaluation," *NBER Working Paper Series*.
- Irwin, K., Ballard, J., Grendon, J., and Kobayashi, J. 1989. "Results of Routine Restaurant Inspections Can Predict Outbreaks of Foodborne Illness: The Seattle-King County Experience," *American Journal of Public Health* (79:5), pp. 586-590.
- Jabr, W., and Zheng, E. 2014. "Know Yourself and Know Your Enemy: An Analysis of Firm Recommendations and Consumer Reviews in a Competitive Environment," *MIS Quarterly*. (38:3), pp. 635-654.
- Jamieson, L. F., and Bass, F. M. 1989. "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, pp. 336-345.
- Jin, G. Z., and Leslie, P. 2009. "Reputational Incentives for Restaurant Hygiene," *American Economic Journal: Microeconomics* (1:1), pp. 237-267.
- Jin, G. Z., and Lee, J. 2014. "Inspection Technology, Detection, and Compliance: Evidence from Florida Restaurant Inspections," *The RAND Journal of Economics* (45:4), pp. 885-917.
- Journal, W. S. 2015. "Chipotle Says E. Coli Outbreak Will Cut Deeply into Sales, Profit." 2016, from <http://www.wsj.com/articles/chipotle-linked-e-coli-outbreak-widens-to-three-more-states-1449251438>
- Juran, J. M., and Gryna, F. M. 1980. *Quality Planning and Analysis*. McGraw-Hill.
- Kalnins, A., and Mayer, K. J. 2004. "Franchising, Ownership, and Experience: A Study of Pizza Restaurant Survival," *Management Science* (50:12), pp. 1716-1728.

- Kang, J. S., Kuznetsova, P., Luca, M., and Choi, Y. 2013. "Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews," *EMNLP: Citeseer*, pp. 1443-1448.
- Katz, E., and Lazarsfeld, P. F. 1955. *Personal Influence, the Part Played by People in the Flow of Mass Communications*. Transaction Publishers.
- Kennedy, M. T. 2008. "Getting Counted: Markets, Media, and Reality," *American Sociological Review* (73:2), pp. 270-295.
- Kimes, S., and Dholakia, U. 2011. "Restaurant Daily Deals: Customers' Responses to Social Couponing," *Cornell Hospitality Report* (11:20), pp. 1-22.
- Kirmani, A. 1990. "The Effect of Perceived Advertising Costs on Brand Perceptions," *Journal of Consumer Research*, pp. 160-171.
- Kirmani, A. 1997. "Advertising Repetition as a Signal of Quality: If It's Advertised So Much, Something Must Be Wrong," *Journal of Advertising*, pp. 77-86.
- Kirmani, A., and Wright, P. 1989. "Money Talks: Perceived Advertising Expense and Expected Product Quality," *Journal of Consumer Research*, pp. 344-353.
- Krasnova, H., Veltri, N. F., Spengler, K., and Günther, O. 2013. "'Deal of the Day' Platforms: What Drives Consumer Loyalty?," *Business & Information Systems Engineering* (5:3), pp. 165-177.
- Kumar, V., and Rajan, B. 2012. "Social Coupons as a Marketing Strategy: A Multifaceted Perspective," *Journal of the Academy of Marketing Science* (40:1), pp. 120-136.
- Kwark, Y., Chen, J., and Raghunathan, S. 2014. "Online Product Reviews: Implications for Retailers and Competing Manufacturers," *Information Systems Research* (25:1), pp. 93-110.

- Lechner, M. 2011. *The Estimation of Causal Effects by Difference-in-Difference Methods*.
Now.
- Lee, L., and Tsai, C. I. 2014. "How Price Promotions Influence Postpurchase Consumption Experience over Time," *Journal of Consumer Research* (40:5), pp. 943-959.
- Lewis, G. 2011. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of Ebay Motors," *The American Economic Review* (101:4), pp. 1535-1546.
- Li, X., and Hitt, L. M. 2008. "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), pp. 456-474.
- Li, X., and Hitt, L. M. 2010. "Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis," *MIS quarterly* (34:4), pp. 809-831.
- Li, X., Hitt, L. M., and Zhang, Z. J. 2011. "Product Reviews and Competition in Markets for Repeat Purchase Products," *Journal of Management Information Systems* (27:4), pp. 9-42.
- Liang, K.-Y., and Zeger, S. L. 1986. "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, pp. 13-22.
- Lin, D. Y., and Wei, L.-J. 1989. "The Robust Inference for the Cox Proportional Hazards Model," *Journal of the American Statistical Association* (84:408), pp. 1074-1078.
- Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
- Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), pp. 74-89.
- Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. 2002. "Text Classification Using String Kernels," *The Journal of Machine Learning Research* (2),

pp. 419-444.

- Lu, X., Ba, S., Huang, L., and Feng, Y. 2013. "Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews," *Information Systems Research* (24:3), pp. 596-612.
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E. C., Wetzels, M., and Pfann, G. 2013. "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing* (77:1), pp. 87-103.
- Luo, T., and Stark, P. B. 2015. "Nine out of 10 Restaurants Fail? Check, Please," *Significance* (12:2), pp. 25-29.
- Luo, X., Andrews, M., Song, Y., and Aspara, J. 2014. "Group-Buying Deal Popularity," *Journal of Marketing* (78:2), pp. 20-33.
- Luo, X., Zhang, J., and Duan, W. 2013. "Social Media and Firm Equity Value," *Information Systems Research* (24:1), pp. 146-163.
- Lynch Jr, J. G., and Ariely, D. 2000. "Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution," *Marketing Science* (19:1), pp. 83-103.
- Maindonald, J., and Braun, J. 2006. *Data Analysis and Graphics Using R: An Example-Based Approach*. Cambridge University Press.
- Mangold, W. G., and Faulds, D. J. 2009. "Social Media: The New Hybrid Element of the Promotion Mix," *Business horizons* (52:4), pp. 357-365.
- Mangold, W. G., Miller, F., and Brockway, G. R. 1999. "Word-of-Mouth Communication in the Service Marketplace," *Journal of Services Marketing* (13:1), pp. 73-89.
- Martens, M. L., Jennings, J. E., and Jennings, P. D. 2007. "Do the Stories They Tell Get Them the Money They Need? The Role of Entrepreneurial Narratives in Resource

- Acquisition," *Academy of Management Journal* (50:5), pp. 1107-1132.
- Mayhew, G. E., and Winer, R. S. 1992. "An Empirical Analysis of Internal and External Reference Prices Using Scanner Data," *Journal of Consumer Research*, pp. 62-70.
- Mazumdar, T., Raj, S., and Sinha, I. 2005. "Reference Price Research: Review and Propositions," *Journal of Marketing* (69:4), pp. 84-102.
- McCallum, A., and Nigam, K. 1998. "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI-98 workshop on learning for text categorization*: Citeseer, pp. 41-48.
- McCulloch, C. E., and Neuhaus, J. M. 2001. *Generalized Linear Mixed Models*. Wiley Online Library.
- McDougall, P., and Robinson Jr, R. B. 1990. "New Venture Strategies: An Empirical Identification of Eight" Archetypes" of Competitive Strategies for Entry," *Strategic Management Journal* (11:6), p. 447.
- McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic," *Biochemia medica* (22:3), pp. 276-282.
- Metters, R., and Marucheck, A. 2007. "Service Management--Academic Issues and Scholarly Reflections from Operations Management Researchers," *Decision Sciences* (38:2), pp. 195-214.
- Meyer, C. D. 2000. *Matrix Analysis and Applied Linear Algebra*. Siam.
- Miller, G. A. 1995. "Wordnet: A Lexical Database for English," *Communications of the ACM* (38:11), pp. 39-41.
- Moe, W. W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research* (48:3), pp. 444-456.

- Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Review? A Study of Customer Reviews on Amazon. Com," *MIS Quarterly* (34:1), pp. 185-200.
- National Restaurant Association. 2012. "Restaurant Industry Forecast."
- Neslin, S. A., Henderson, C., and Quelch, J. 1985. "Consumer Promotions and the Acceleration of Product Purchases," *Marketing science* (4:2), pp. 147-165.
- Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science* (31:3), pp. 521-543.
- New York Times, The. 2012. "Improve Restaurant Report Cards." 2016, from <http://www.nytimes.com/2012/03/07/opinion/the-restaurant-grade-system-is-broken.html>
- New York Daily News. 2015. "Add Legionnaires' Outbreak to Slew of Responsibilities for Overworked City Health Inspectors." 2016, from <http://www.nydailynews.com/new-york/exclusive-nyc-inspectors-overwhelmed-legionnaires-hits-article-1.2319546>
- NYCOpenData. 2016. "NYC Open Data." from <https://nycopendata.socrata.com/>
- Openshaw, E., Belson, G., Wigginton, C., and Derricks, A. 2014. "Digital Democracy Survey," *Deloitte Press*.
- Owen, A. B., and Perry, P. O. 2009. "Bi-Cross-Validation of the SVD and the Nonnegative Matrix Factorization," *The Annals of Applied Statistics*, pp. 564-594.
- Parasuraman, A., Berry, L. L., and Zeithaml, V. A. 1991. "Refinement and Reassessment of the Servqual Scale," *Journal of Retailing* (67:4), p. 420.
- Parasuraman, A., Berry, L. L., and Zeithaml, V. A. 1993. "More on Improving Service Quality Measurement," *Journal of Retailing* (69:1), pp. 140-147.

- Parasuraman, A., Zeithaml, V. A., and Berry, L. L. 1985. "A Conceptual Model of Service Quality and Its Implications for Future Research," *The Journal of Marketing*, pp. 41-50.
- Parasuraman, A., Zeithaml, V. A., and Berry, L. L. 1988. "Servqual," *Journal of Retailing* (64:1), pp. 12-40.
- Parker, G. G., and Van Alstyne, M. W. 2005. "Two-Sided Network Effects: A Theory of Information Product Design," *Management Science* (51:10), pp. 1494-1504.
- Parsa, H., Self, J., Sydnor-Busso, S., and Yoon, H. J. 2011. "Why Restaurants Fail? Part II- the Impact of Affiliation, Location, and Size on Restaurant Failures: Results from a Survival Analysis," *Journal of Foodservice Business Research* (14:4), pp. 360-379.
- Parsa, H., Self, J. T., Njite, D., and King, T. 2005. "Why Restaurants Fail," *Cornell Hotel and Restaurant Administration Quarterly* (46:3), pp. 304-322.
- Paul, H., and Jonathan, Z. 1991. "Flexible Specialization Versus Post-Fordism: Theory, Evidence and Policy Implications," *Economy and Society* (20:1), pp. 5-9.
- Pauwels, K., Hanssens, D. M., and Siddarth, S. 2002. "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity," *Journal of Marketing Research* (39:4), pp. 421-439.
- Pepe, M. S. 2000. "An Interpretation for the Roc Curve and Inference Using GLM Procedures," *Biometrics* (56:2), pp. 352-359.
- Petkova, A. P., Rindova, V. P., and Gupta, A. K. 2013. "No News Is Bad News: Sensegiving Activities, Media Attention, and Venture Capital Funding of New Technology Organizations," *Organization Science* (24:3), pp. 865-888.
- Pine, B. J., and Gilmore, J. H. 1999. *The Experience Economy: Work Is Theatre & Every*

Business a Stage. Harvard Business Press.

Pischke, J.-S. 2004. "Labour Economics for Research Students Differences-in-Differences."

from <http://econ.lse.ac.uk/staff/spischke/ec533/did.pdf>

Pollock, T. G., and Rindova, V. P. 2003. "Media Legitimation Effects in the Market for

Initial Public Offerings," *Academy of Management Journal* (46:5), pp. 631-642.

Pollock, T. G., Rindova, V. P., and Maggitti, P. G. 2008. "Market Watch: Information and

Availability Cascades among the Media and Investors in the Us IPO Market,"

Academy of Management Journal (51:2), pp. 335-358.

Porac, J. F., Mishina, Y., and Pollock, T. 2002. "Entrepreneurial Narratives and the

Dominant Logics of High-Growth Firms," *Mapping Strategic Knowledge*), pp. 112-

136.

Pullman, M. E., and Gross, M. A. 2004. "Ability of Experience Design Elements to Elicit

Emotions and Loyalty Behaviors," *Decision Sciences* (35:3), pp. 551-578.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing." Vienna,

Austria.

Raghubir, P., and Corfman, K. 1995. "When Do Price Promotions Signal Quality? The Effect

of Dealing on Perceived Service Quality," *Advances in consumer research* (22), pp.

58-58.

Raghubir, P., and Corfman, K. 1999. "When Do Price Promotions Affect Pretrial Brand

Evaluations?," *Journal of Marketing Research*, pp. 211-222.

Ramdas, K., Williams, J., and Lipson, M. 2013. "Can Financial Markets Inform Operational

Improvement Efforts? Evidence from the Airline Industry," *Manufacturing & Service*

Operations Management (15:3), pp. 405-422.

- Reeves, C. A., and Bednar, D. A. 1994. "Defining Quality: Alternatives and Implications," *Academy of Management Review* (19:3), pp. 419-445.
- Reichheld, F. P., and Sasser, W. E. 1990. "Zero Defections: Quality Comes to Services," *Harvard Business Review* (68:5), pp. 105-111.
- Rindova, V. P., Petkova, A. P., and Kotha, S. 2007. "Standing Out: How New Firms in Emerging Markets Build Reputation," *Strategic Organization* (5:1), pp. 31-70.
- Roberts, J. H., and Lattin, J. M. 1997. "Consideration: Review of Research and Prospects for Future Insights," *Journal of Marketing Research*, pp. 406-410.
- Rosenzweig, E. D., Laseter, T. M., and Roth, A. V. 2011. "Through the Service Operations Strategy Looking Glass: Influence of Industrial Sector, Ownership, and Service Offerings on B2b E-Marketplace Failures," *Journal of Operations Management* (29:1), pp. 33-48.
- Ross, S. M. 1996. *Stochastic Processes*. John Wiley & Sons New York.
- Rossi, P. E., and Allenby, G. M. 2003. "Bayesian Statistics and Marketing," *Marketing Science* (22:3), pp. 304-328.
- Rossi, P. E., Allenby, G. M., and McCulloch, R. 2012. *Bayesian Statistics and Marketing*. John Wiley & Sons.
- Roth, A. V., Chase, R. B., and Voss, C. 1997. *Service in the Us: Progress Towards Global Service Leadership*. Severn Trent Plc.
- Roth, A. V., and Jackson III, W. E. 1995. "Strategic Determinants of Service Quality and Performance: Evidence from the Banking Industry," *Management Science* (41:11), pp. 1720-1733.
- Roth, A. V., and Menor, L. J. 2003. "Insights into Service Operations Management: A

- Research Agenda," *Production and Operations Management* (12:2), pp. 145-163.
- Rubin, D. B. 1976. "Multivariate Matching Methods That Are Equal Percent Bias Reducing, I: Some Examples," *Biometrics*, pp. 109-120.
- Rust, R. T., Zahorik, A. J., and Keiningham, T. L. 1994. *Return on Quality: Measuring the Financial Impact of Your Company's Quest for Quality*. Probus Publ.
- Rust, R. T., Zahorik, A. J., and Keiningham, T. L. 1995. "Return on Quality (Roq): Making Service Quality Financially Accountable," *The Journal of Marketing*, pp. 58-70.
- Sasser, W. E., Schlesinger, L. A., and Heskett, J. L. 1997. *Service Profit Chain*. Simon and Schuster.
- Schlesinger, L. A., and Heskett, J. L. 1991. "Breaking the Cycle of Failure in Services," *MIT Sloan Management Review* (32:3), p. 17.
- Schneider, B., and Bowen, D. E. 2010. *Winning the Service Game*. Springer.
- Schoeffler, S., Buzzell, R. D., and Heany, D. F. 1974. *Impact of Strategic Planning on Profit Performance*. Graduate School of Business Administration, Harvard University.
- Schoenfeld, D. 1982. "Partial Residuals for the Proportional Hazards Regression Model," *Biometrika* (69:1), pp. 239-241.
- Sebastiani, F. 2002. "Machine Learning in Automated Text Categorization," *ACM computing surveys (CSUR)* (34:1), pp. 1-47.
- Sethuraman, R., Srinivasan, V., and Kim, D. 1999. "Asymmetric and Neighborhood Cross-Price Effects: Some Empirical Generalizations," *Marketing science* (18:1), pp. 23-41.
- Shapiro, C. 1986. "Investment, Moral Hazard, and Occupational Licensing," *The Review of Economic Studies* (53:5), pp. 843-862.
- Shivendu, S., and Zhang, Z. 2013. "Economics of Daily-Deal Websites: Advertising and

- Sampling Effects," *ICIS 2013 Proceedings*.
- Soteriou, A., and Zenios, S. A. 1999. "Operations, Quality, and Profitability in the Provision of Banking Services," *Management Science* (45:9), pp. 1221-1238.
- Spence, M. 1973. "Job Market Signaling," *The quarterly journal of Economics*, pp. 355-374.
- Spence, M. 2002. "Signaling in Retrospect and the Informational Structure of Markets," *The American Economic Review* (92:3), pp. 434-459.
- Starbird, S. A. 2005. "Moral Hazard, Inspection Policy, and Food Safety," *American Journal of Agricultural Economics* (87:1), pp. 15-27.
- Stiglitz, J. 2009. "Regulation and Failure," *New perspectives on regulation*, pp. 11-23.
- Sunstein, C. R. 2002. "What's Available-Social Influences and Behavioral Economics," *Nw. UL Rev.* (97), p. 1295.
- Tang, H., Tan, S., and Cheng, X. 2009. "A Survey on Sentiment Detection of Reviews," *Expert Systems with Applications* (36:7), pp. 10760-10773.
- Trusov, M., Bodapati, A. V., and Bucklin, R. E. 2010. "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research* (47:4), pp. 643-658.
- Tsai, A. C.-R., Wu, C.-E., Tsai, R. T.-H., and Hsu, J. Y.-j. 2013. "Building a Concept-Level Sentiment Dictionary Based on Commonsense Knowledge," *IEEE Intelligent Systems*:2), pp. 22-30.
- Tuten, T. L., and Ashley, C. 2011. "Promotional Strategies for Small Businesses: Group Buying Deals," *Small Business Institute® Journal* (7:2), pp. 15-29.
- Valitutti, A., Strapparava, C., and Stock, O. 2004. "Developing Affective Lexical Resources," *PsychNology Journal* (2:1), pp. 61-83.
- Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., and Verhoef, P. C.

2010. "Customer Engagement Behavior: Theoretical Foundations and Research Directions," *Journal of Service Research* (13:3), pp. 253-266.
- Vermeulen, I. E., and Seegers, D. 2009. "Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration," *Tourism Management* (30:1), pp. 123-127.
- Vorhies, D. W., and Morgan, N. A. 2005. "Benchmarking Marketing Capabilities for Sustainable Competitive Advantage," *Journal of Marketing* (69:1), pp. 80-94.
- Voss, C., Roth, A. V., and Chase, R. B. 2008. "Experience, Service Operations Strategy, and Services as Destinations: Foundations and Exploratory Investigation," *Production and Operations Management* (17:3), pp. 247-266.
- Waldfogel, J., and Chen, L. 2006. "Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers," *The Journal of Industrial Economics* (54:4), pp. 425-449.
- Winer, R. S. 1986. "A Reference Price Model of Brand Choice for Frequently Purchased Products," *Journal of consumer research*, pp. 250-256.
- Wong, M. R., and Matis, B. 2015. "Letter Grading and Transparency Promote Restaurant Food Safety in New York City," *Journal of environmental health* (78:2), p. 46.
- Wright, P. 1986. "Schemer Schema-Consumers Intuitive Theories About Marketers Influence Tactics," *Advances in consumer research* (13), pp. 1-3.
- Wu, B., Petroschius, S., and Newell, S. 2004. "The Impact of Store Image, Frequency of Discount, and Discount Magnitude on Consumers' Value Perceptions and Search Intention," *Marketing Management Journal* (14:1), p. 14.
- Wyckoff, D. D. 1984. "New Tools for Achieving Service Quality," *Cornell Hotel and Restaurant Administration Quarterly* (25:3), pp. 78-91.

- Ye, Q., Law, R., and Gu, B. 2009. "The Impact of Online User Reviews on Hotel Room Sales," *International Journal of Hospitality Management* (28:1), pp. 180-182.
- Yelp. 2014. "Yelp Q2 2014 Investor Deck."
- Yelp. 2016. "Investor Relations - Quarterly Results." 2016, from <http://www.yelp-ir.com/phoenix.zhtml?c=250809&p=irol-irhome>
- Yoo, C. W., Sanders, G. L., and Moon, J. 2013. "Exploring the Effect of E-Wom Participation on E-Loyalty in E-Commerce," *Decision Support Systems* (55:3), pp. 669-678.
- Yoon, S., Oh, S., Song, S., Kim, K. K., and Kim, Y. 2014. "Higher Quality or Lower Price? How Value-Increasing Promotions Affect Retailer Reputation Via Perceived Value," *Journal of Business Research* (67:10), pp. 2088-2096.
- Zagat. 2015. "The State of American Dining."
- Zeithaml, V. A. 1988. "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *The Journal of Marketing*, pp. 2-22.
- Zeithaml, V. A., Berry, L. L., and Parasuraman, A. 1993. "The Nature and Determinants of Customer Expectations of Service," *Journal of the Academy of Marketing Science* (21:1), pp. 1-12.
- Zeithaml, V. A., Berry, L. L., and Parasuraman, A. 1996. "The Behavioral Consequences of Service Quality," *The Journal of Marketing*, pp. 31-46.
- Zeithaml, V. A., Parasuraman, A., and Berry, L. L. 1990. *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. Simon and Schuster.
- Zhao, Y., Yang, S., Narayan, V., and Zhao, Y. 2013. "Modeling Consumer Learning from Online Product Reviews," *Marketing Science* (32:1), pp. 153-169.

Zhou, W., and Duan, W. 2010. "Online User Reviews and Professional Reviews: A Bayesian Approach to Model Mediation and Moderation Effects," *ICIS*, p. 256.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of marketing* (74:2), pp. 133-148.