1-1-2012

Empirical Essays in The Entertainment and Hospitality industries

Fangyun Tan
University of Pennsylvania, fangyun.tan@gmail.com

Empirical Essays in The Entertainment and Hospitality industries

Abstract

In the first part of the dissertation, we empirically examine the impact of expanded product variety due to the adoption of the Internet on demand concentration, taking endogeneity into consideration. We analyze two large data sets from the movie rental industry at both movie-level and consumer-level. We find that product variety diversifies the demand for niches more significantly than for hits in absolute terms (e.g., the top/bottom 1,000 titles). However, using relative terms (e.g., the top/bottom 10% of titles) to dynamically adjust for the changing product variety, we find that product variety increases the demand for the hits and reduces the demand for the niches. We further find that product variety increases monthly Gini Coefficients, a measure of demand concentration. We propose that new products appear much faster than consumers discover them. Finally, we find no evidence that niche titles satisfy consumer tastes any better than popular titles and that a small number of heavy users are more likely to venture into niches than light users.

In the second part of the dissertation, we analyze a large, detailed operational data set from a restaurant chain to shed new light on how workload (defined as the hourly average number of diners assigned to a server) affects servers' performance (measured as hourly sales). We use an exogenous shock - implementation of a labor scheduling software - to disentangle the endogeneity between demand and supply in this setting. We find that when the overall workload is low, an increase in the workload leads to higher server performance. However, there is a saturation point after which any further increase in the workload leads to degradation of performance. In the focal restaurant chain we find that this saturation point is generally not reached and, counter-intuitively, the chain can reduce the staffing level and achieve both significantly higher sales (an estimated 35% increase) and lower labor costs (an estimated 20% decrease).

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Operations & Information Management

First Advisor

Lorin Hitt

Keywords

Empirical Operations Management, Entertainment and Hospitality Industries, Long Tail, Performance

Subject Categories

Business | Operational Research | Statistics and Probability

EMPIRICAL ESSAYS IN THE ENTERTAINMENT AND HOSPITALITY INDUSTRIES

Fangyun Tom Tan

A DISSERTATION

in

Operations and Information Management

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

____

Lorin Hitt, Professor, Operations and Information Management

Graduate Group Chairperson

____

Eric Bradlow, Professor, Marketing, Statistics, and Education

Dissertation Committee:

Gérard Cachon, Professor, Operations and Information Management

Serguei Netessine, Professor, Technology and Operations Management

Sergei Savin, Associate Professor, Operations and Information Management

To Mom and Dad, and Heavenly Father.

ACKNOWLEDGEMENTS

Dear Lord, I would like to first offer my thanks to you for your grace and mercy that keep and guide me along my academic and personal journey.

I would also like to thank Mr. Z.Y. Fu, Columbia Engineering School benefactor for his generous scholarship that allows me, a student of a humble family in China to obtain an excellent undergraduate education from Columbia University. There, Professor Ward Whitt, your incisive understanding of my personality and your decisive advice on pursuing academia eventually brought me to the PhD program at Wharton.

During the past five years at Wharton, I have been so blessed with many exemplary mentors and good friends. First and foremost, I am beholden to Serguei Netessine, my advisor for being my professional and personal role model. Serguei, your inspiration and encouragement taught me that there are always opportunities embedded in challenges and that I should keep a long-term vision in mind. Your dedication to our profession imbued me with strong self-discipline. More importantly, your generosity and big heart (larger than your biceps) educated me on how to become a respectable man. Serguei, Спасибо.

I would also like to thank Lorin Hitt, my co-advisor for being my mentor. Your down-to-earth personality made me feel as if I were talking with my uncle. I would also express my profound gratitude to Noah Gans, our father-like PhD coordinator for helping me overcome my most difficult and confusing period during my PhD program. Noah, I really appreciate your approach to identifying and solving a problem. Let me thank Gérard Cachon, our department chair and my committee chair. Under your leadership, our OPIM Department provides a very nourishing and collegial environment for PhD students like me to grow and think freely. Also thanks be to Sergei Savin, my committee member. Sergei, I appreciate you because your high work standard always encourages me to pay attention to details. Ich möchte mich bei Christian Terwiesch herzlich bedanken, für Ihre Freundschaft und die Tür, die Sie für mich öffnen. I would also specially thank Josh Eliashberg, Peter Fader, Kartik Hosanagar and Katy Milkman for your suggestions and advice to my work.

Besides those beloved mentors, many thanks to my superb fellow students. In particular, Hessam Bevafa, thank you for your time and those “academically arousing” conversations.

Jun Li, thank you for being such a patient and open-minded colleague and friend. Jaelynn Oh, thank you for your quiet but substantial help during my job search process. In addition, Shounan Ho and Dominique Ding, thank you for your support!

Finally, I would like to express my deepest gratitude to my family for your unconditional love. You do not mind how high I fly; rather, you care how exhausted I am during the flight. Your infinite love and self-sacrifice have prepared the man that I am today. I apologize that I have not been able to love as much in return. I deeply appreciate your love. Teach me how to love you more.

To all the people, named and unnamed, who love and care about your Tom, Flying Tom and Huan Huan, thank you. Xie xie.

ABSTRACT

EMPIRICAL ESSAYS IN THE ENTERTAINMENT AND HOSPITALITY INDUSTRIES

Fangyun Tom Tan

Lorin Hitt

In the first part of the dissertation, we empirically examine the impact of expanded product variety due to the adoption of the Internet on demand concentration, taking endogeneity into consideration. We analyze two large data sets from the movie rental industry at both movie-level and consumer-level. We find that product variety diversifies the demand for niches more significantly than for hits in absolute terms (e.g., the top/bottom 1,000 titles). However, using relative terms (e.g., the top/bottom 10% of titles) to dynamically adjust for the changing product variety, we find that product variety increases the demand for the hits and reduces the demand for the niches. We further find that product variety increases monthly Gini Coefficients, a measure of demand concentration. We propose that new products appear much faster than consumers discover them. Finally, we find no evidence that niche titles satisfy consumer tastes any better than popular titles and that a small number of heavy users are more likely to venture into niches than light users.

In the second part of the dissertation, we analyze a large, detailed operational data set from a restaurant chain to shed new light on how workload (defined as the hourly average number of diners assigned to a server) affects servers' performance (measured as hourly sales). We use an exogenous shock - implementation of a labor scheduling software - to disentangle the endogeneity between demand and supply in this setting. We find that when the overall workload is low, an increase in the workload leads to higher server performance. However, there is a saturation point after which any further increase in the workload leads to degradation of performance. In the focal restaurant chain we find that this saturation point is generally not reached and, counter-intuitively, the chain can reduce the staffing level and achieve both significantly higher sales (an estimated 35% increase) and lower labor costs (an estimated 20% decrease).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Entertainment and Hospitality industries have substantial impact on people's daily life. For example, the Entertainment industry in the U.S. contributes to approximately 11.12% of GDP and 8.49% of employment (WIPO, 2003). Restaurants alone, as part of the hospitality industry, employ approximately 10% of the workforce in the United States, second only to the military (Hickok and Lazarus, 2003).

Despite of the large economic and social impact of these two industries, they rarely attract the attention of Operations Management researchers. However, in my opinion, these two industries offer a number of opportunities for process improvement. For example, the productivity in restaurants is only half as much as the manufacturing industries (Mill, 2004). Moreover, more operational data from these industries start to become available for empirical researchers. The goal of this dissertation is to use operational data and advance our understanding of the new business challenges in today's Entertainment and Hospitality industries.

The dissertation is structured as follows. In Chapter 2, we use two novel data sets from the movie rental industries to understand the impact of expanded product variety on demand concentration. In Chapter 3 we use a large operational data from a restaurant chain to shed new light on how workload affects servers' performance, taking endogeneity into consideration.

Multiple studies suggested that expanding product variety due to the adoption of the Internet will satisfy consumers' increasingly heterogeneous tastes, thus causing the so-called

Long Tail effect (i.e., increasing demand for niche products). In Chapter 2 we empirically examine the impact of product variety on demand concentration. We use two large data sets from the movie rental industry and analyze the data at both movie-level and consumer-level. We employ multiple measures to understand the changing demand concentration and incorporate the potential endogeneity of product variety.

Multiple models analyzed in our study consistently suggest that larger product variety is likely to increase the demand for hits and decrease the demand for niche products. We propose that new products appear much faster than consumers discover them. Finally, we find no evidence that niche titles satisfy consumer tastes any better than popular titles and that a small number of heavy users are more likely to venture into niches than light users.

In order to understand the multi-dimensions of a business, we move from the consumer's side in Chapter 2 to the service provider's side in Chapter 3. Understanding employees' productivity is an integral part of labor management. Classical Operations Management studies tended to implicitly assume that servers are homogeneous and independent of their work environment. In Chapter 3, we examine the effects of workload on employees' performance, taking endogeneity into consideration. We analyze a large, detailed operational data from a restaurant chain to shed new light on how workload (defined as the hourly average number of diners assigned to a server) affects servers' performance (measured as hourly sales). We use an exogenous shock - implementation of a labor scheduling software - to disentangle the endogeneity between demand and supply in this setting. We find that when the overall workload is low, an increase in the workload leads to higher server performance. However, there is a saturation point after which any further increase in the workload leads to degradation of performance. In the focal restaurant chain we find that this saturation point is generally not reached and, counter-intuitively, the chain can reduce the staffing level and achieve both significantly higher sales (an estimated 35% increase) and lower labor costs (an estimated 20% decrease).

Chapter 2

Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration

2.1 Introduction and Related Literature

Chris Anderson, editor-in-chief of *Wired Magazine*, coined the term "Long Tail effect" (Anderson, 2004) suggesting that, due to the introduction of the Internet, niche products will comprise higher and higher market share, while the demand for hit products will continue to decrease. As a result, he predicted that the old Pareto rule, stating that 20% of all the products generate 80% of the revenues, will no longer hold: hit movies will constitute a smaller and smaller proportion of demand. His predictions of the Long Tail effect were motivated by observations in the media, entertainment and other industries. For example, Anderson (2006) finds that the top 50 best-selling albums of all time were produced in the 70s and 80s; none of them were recorded in recent years. He also observes that the ratings of the top TV shows have gradually decreased and that the top show today would not have ranked among the top ten in 1970. Part of the reason, according to Anderson, is that niche products will better and better satisfy consumer preferences because consumers will continue to have more and more varying preferences while the Internet will make even the most obscure products available to the masses.

The potential for the existence of the Long Tail effect is of great importance for product assortment decisions in a variety of industries, for advertising dollars spent on supporting this variety, and for supply chain management of these products on the Internet. For example, Blockbuster stocks 3,000 DVDs per store on average, while 20% of Netflix rental revenues come from outside the top 3,000 titles (Anderson, 2004). In addition, Ecast, a digital jukebox company, sold 98% of its 10,000 albums available online at least one track per album per quarter (Anderson, 2006), while brick-and-mortar music stores only stock a fraction of this variety. If demand is indeed shifting toward more obscure titles, managers should ensure that these titles are available and that they are advertised properly. Further, Anderson explains that the new online recommendation systems help the niche products quickly find their demand in the market once they are made available. As a result, he asserts that "the tail of available variety is far longer than we expected", and that the combined market share of the niches can outgrow the hits (Anderson, 2006). This comment about the increasing demand for the niches seems to be consistent with Varian's opinion in light of the cheaper technology in the media industry. Specifically, Varian (2006) notes that this "creative, inexpensive and compelling semiprofessional content available via the Internet" has an increased demand particularly among young people, so that the salaries of celebrities, such as Tom Cruise, may decrease. He explains, "It is true that there is only one Tom Cruise, but it is equally true that there are only 24 hours in a day. The more time young people spend watching Lonelygirl15, the less time they will have to watch Mr. Cruise."

Is popular celebrity Tom Cruise really threatened by Lonelygirl15? Although arguments and evidence in favor of the Long Tail effect appeared pervasive at first, there are also indications that hits still drive some markets, and may even become more popular over time, whereas the rising demand for niches is, at best, overestimated. In particular, some evidence suggests that new products appear so quickly that consumers have no time to discover them. Gomes (2006) discloses that at Ecast, the quarterly no-play rate increased from 2% to 12% as product variety has grown. In addition, an even stronger demand for hits is found in the motion picture industry, where both the number of movies that generate box-office revenues of over $50 million and their percentage of the total revenues increased from 14 and 14% in 1998 to 19 and 22% in 2003, respectively (Eliashberg et al., 2006). Finally, Orlowski (2008) reports on an industry study which discovered that 80% of the digital song inventory sold no copies at all - and the 'head' of the frequency distribution was far more concentrated than expected. Given this conflicting evidence, whether or not the Long Tail effect exists remains a hotly debated issue among practitioners.

The Long Tail effect has also recently generated widespread interest in academic circles. Brynjolfsson et al. (2006, 2010, 2011) present plausible factors that may drive the Long Tail effect, including both supply-side and the demand-side effects. On the supply side, they suggest that the Internet reduces the production and distribution costs of niche products, creating more products available. These products can satisfy consumer's heterogeneous needs, thus driving the Long Tail effect. On the demand side, they note that both the active and the passive search and personalization tools lower the search costs and hence facilitate finding niche products. Consistent with this view, Cachon et al. (2008) find that the lowered search costs have market-expansion effect, which encourages firms to enlarge their assortment. As a result, consumers are more likely to find niche products, thus causing demand for them to increase. Moreover, Tucker and Zhang (2011) suggest that product popularity information, such as the number of people who have browsed the product, can increase the appeal of niche products disproportionately. In addition, Kumar et al. (2011) find that broadcasting movies on pay-TV can increase the the awareness of unpopular movies, thus reducing the demand concentration of DVD sales.

However, several studies have shown opposite results of the Long Tail effect. Ghose and Gu (2006) argue that search costs are even lower for popular products than for niches, which may limit the Long Tail effect. Hervas-Drane (2009) provides an analytical model to show that different search processes have mixed impacts on demand concentration. Moreover, Fleder and Hosanagar (2009) suggest that sales diversity can be reduced by selection-biased recommendation systems because these systems tend to recommend products with sufficient historical data (i.e., hits). From a field experiment, Fleder et al. (2009) further find that consumers buy a more similar mix of music after receiving recommendations than before. Dellarocas and Narayan (2007) also find that online consumers are more likely to review popular products, and therefore, online reviews may exhibit "tall heads" instead of "long tails". Bockstedt and Goh (2008) analyze the data of consumer-created custom CDs to examine whether people tend to bundle the hits or the "long tail" music and suggest that managers should sell unbundled information goods to meet the demand from the mainstream consumer. Elberse and Oberholzer-Gee (2008) find further evidence that online retailing triggers demand to shift toward the tail of the distribution, although they also find that a substantive part of demand is concentrated on an even smaller portion of products.

So far, both academic theories and the empirical evidence provide what can probably be described as conflicting evidence for the existence and the magnitude of the increased demand for niche products: while there are many anecdotal examples of its presence, there are fewer than a handful of rigorous and large empirical studies both at the product and consumer levels. Furthermore, previous research has implicitly assumed that expanding product variety will satisfy consumer's increasingly heterogeneous tastes, which causes the Long Tail (e.g., Brynjolfsson et al. 2006, 2010). Little research has empirically confirmed the impact of product variety on the demand concentration. For example, the product variety is the same between online and offline channels in Brynjolfsson et al. (2011), which excludes the impact of product variety. Tracking weekly video sales from 2000 to 2005, Elberse and Oberholzer-Gee (2008) define hits as the top percentiles of all the movies across the entire six years. In other words, they create a static definition of hits for all the movies over a span of six years, which implicitly excludes the impact of dynamic product variety. In a relevant study, Hinz et al. (2011) use aggregate video-on-demand data in Germany and study the effect of product variety on the share of purchased products. Nevertheless, the study does not examine the demand concentration impact.

Although little research has been done to explicitly examine the impact of product variety on the demand concentration, whether or not the demand for niche products increases amid an ever-changing product variety is a fundamental question for decision-makers in operations, marketing and finance, particularly when they face the prospect of further penetration of the Internet channel, which offers expanding product variety and new recommendation systems to help manage it.

In this paper we empirically evaluate the impact of product variety on the demand concentration. We use large data sets from the movie rental industry and report results at both movie-level and consumer-level. The first data set (Dataset I) contains rental activity of about 30% of the entire U.S. market from January 2001 to July 2005. The other data (Dataset II) includes 100 million online ratings of 17,770 movie titles by 480,000 users on Netflix during the same period of time. The movie rental industry, particularly Netflix, is a key example in Anderson's evidence for the Long Tail effect and he primarily refers to the popularity of products in absolute terms, e.g., the top 100 or the top 1,000 for hits. In his own words, "number one is still number one, but the sales that go with that are not what they once were" (Anderson, 2006). Following this example, we first measure movie popularity in absolute terms. We find that, when movie popularity is measured in absolute terms, there is only partial evidence to support the Long Tail effect: product variety diversifies the demand for hits, nevertheless less significantly than for niches .

The above definition of the Long Tail effect and movie popularity is static, which implicitly excludes the impact of an increasing product variety. This definition would certainly reflect product popularity in a channel where product variety is relatively stable and where all products are consumed, such as in a brick-and-mortar store. However, product variety has skyrocketed during the Internet age, and more products than ever are not being discovered by consumers (Brynjolfsson et al., 2003). Such a dramatic increase in product variety is likely to create demand diversification. For example, given a choice set of only five movies, people may tend to concentrate their demand on one movie whose popularity rank is number 1 or equivalently in the top 20%. However, out of a wider choice set of 500 movies, the demand may be concentrated on 100 movies whose popularity ranks in the top 100 or also in the top 20%. This example causes a conflicting definition of hits and niches amid different sizes of product variety at different points in time. Should we classify the top 20%, which is respectively the top one out of five movies and the top 100 out of 500 movies as the hits, or should we restrict the label of hits to only the top one movie regardless of the total variety?

Naturally, when the product variety is large, the demand for any one product tends to be smaller than when the product variety is small. Likewise, when the consumer base is large, learning about new products is faster than when the consumer base is small. In this case, two competing effects might be observed: 1) consumers discover the obscure products as they appear and 2) new products appear, possibly so quickly that most consumers have no time to discover them. Which effect dominates is an empirical question that we aim to address in this paper. Therefore, we propose a dynamic definition of product popularity which adjusts for active product variety. The active product variety excludes the titles that have no current rentals or ratings and therefore reflects the dynamics of both product variety and consumer base. As one specific case, we use the A-B-C classification, which traditionally groups products into three classes in inventory management literature (Silver et al., 1998). In particular, Class A products, which comprise the most popular 10% of the product variety, usually contribute to the most of the total demand, thus requiring the highest priority of managerial attention. Class B items, which comprise the products that rank between top 10% and top 40% of total products, usually account for the rest of the total demand. Class C products, which are the largest class including the remaining 60% of all products. Using A-B-C classification not only allows us to account for the dynamic product variety, but also provides insights to the demand for the "mid-tail" products, which are not as easy to identify as the "head" and the "long tail" products (Jiang et al., 2011).

We find that an increase of 1,000 new movies per month in the product variety increased the demand for Class A movies by about 0.51%, whereas it decreased the demand for Class B movies by 1.67% and for Class C movies by 0.53%. This finding suggests that product variety intensifies demand concentration, which is also manifested by changes in monthly Gini coefficients as product variety changes: 1,000 new movies per month may increase demand inequality by 0.19%. These results are consistent after we take endogeneity of product variety into consideration.

In order to explain why product variety intensify the demand concentration, we analyze the data at both movie and consumer level. At the movie-level, we find that the demand diversification effect due to expanded product variety is admittedly stronger for a Class A movie than for either a Class B or Class C movie. Nevertheless, this discrepancy of the demand diversification effect is not large enough to cause the demand for all the obscure movies to increase. Expanded product variety introduces more Class A movies, which strengthen the demand for the Class A movies altogether. Similarly, at the consumer level, we find that new products appear so quickly that most consumers have no time to discover or consume them, causing them to watch more and more hits. Furthermore, while Anderson (2004; 2006) argued that more and more consumers will choose niche products because they will tend to satisfy consumer preferences better, we reveal that consumers tend to be less satisfied with niche and less popular movies that with popular ones. We also find that it is mostly the heavy users or the "movie buffs", a small fraction of all consumers, that venture into niche movies.

To summarize, the contributions of this paper are three-folds. First, we examine the implicit assumption about the impact of product variety on demand concentration and provide new counter-evidence against the Long Tail. Second, we propose to delineate two effects: demand diversification due to expanding product variety and consumers learning about new products. We suggest that, when measuring product popularity, one has to use relative measures such as A-B-C classification to adjust for instantaneous active product variety. With this definition, we find that product variety increases demand concentration on hits, which is against the Long Tail prediction. Third, we study demand at the consumer level and find that new movies appear so quickly that most consumers have no time to discover them, and that niche movies do not satisfy consumer tastes better than hit movies.

The remainder of this paper is organized as follows. We develop our hypothesis in Section 2.2. In Section 2.3, we describe our data sets and variables. We present the results of our empirical analysis in Section 2.4. We conclude with a discussion of our results, limitations, and future research opportunities in Section 2.5.

2.2 Hypotheses Development

In this section, we develop our hypotheses about the impact of product variety on demand concentration. In particular, we discuss how expanded product variety limits the assumptions of some conventionally-held drivers that would predict the Long Tail effect.

Previous studies argue that producers and retailers have increasing incentives to produce and stock niche products because of lowered production and inventory costs (e.g., Brynjolfsson et al. 2006, 2011). This increased product variety is further assumed to better and better satisfy consumer preferences because consumers will continue to have more and more varying preferences, thus leading to the Long Tail effect (e.g., Anderson 2006).

While there is little doubt that product variety generally increases and that technology such as the Internet and drop-shipping techniques allows companies to offer an even wider variety of products economically, it is less clear that consumers necessarily quickly discover these products let alone actually consuming these products. Large product variety is likely to make it more difficult to notice new products. In other words, a change from two to three options in the choice set can be easily noticed, but it takes a lot more effort to notice a change from 2,000 to 2,001 options. New products that have limited associated advertising budgets to create "buzz", particularly those niche products, may disproportionately emerged unnoticed, thus jeopardizing their demand.

In addition, although consumers have varying tastes and like to seek variety, they are less likely to examine all choices to find their "true" fit of tastes when they are faced with large product variety. Too many choices require more cognitive efforts to evaluate the attractiveness of alternatives in the large variety (see Kuksov and Villas-Boas 2010 for a review), thus increasing search cost. When search cost is high, consumers tend to restrict their choice consideration to the products for which they have ex ante knowledge (Stigler, 1961; Rothschild, 1974). These products tend to be popular products because they are more likely to have louder buzz from advertising, promotion and word-of-mouth. As a result, demand is likely to be even more concentrated on those hits.

Furthermore, good-quality products may seem even more attractive in the large product variety. Simonson and Tversky (1992) suggest that adding extremely low-quality products into a consideration set can increase the attraction of the higher-quality products. Therefore, introducing more lower-quality products, which tend to be niches, is likely to make higher-quality products, which tend to be hits, even more popular. In addition, when product variety is large, consumers are found to be more discriminating in terms of product quality (Bertini et al., 2011), which can further increase the demand for the hits.

For these reasons, we hypothesize that

HYPOTHESIS: *Product variety is positively associated with demand for hits and negatively associated with demand for niches.*

2.3 Data

2.3.1 Research Setting and Data Collection

To examine our research hypothesis, we gathered data available from a company that leases and delivers movies to rental retailers. Its clients represent approximately 30% of the entire U.S. movie rental retailers. This company also collects the related rental information for the movies on a revenue-sharing basis. The data that we possess consist of the DVD rental turns and movie characteristics at the movie level from January 2001 to July 2005.

In addition to the movie-level data (Dataset I), we also collected consumer-level data from Netflix (Dataset II), a major U.S. online movie/TV series[1] rental service with annual revenues in excess of $1 billion in 2008. Dataset II was made available to the public during the Netflix Prize competition, which offered $1 million to the team that could use the data to create the most accurate movie recommendation system. The data set consists of the movie ratings submitted by consumers through the Netflix web site from 2000 to 2005, from which we gathered the data from January 2001 to July 2005 to match with our movie-level rental data. Netflix encourages its users to rate the movies that they have watched both outside and within Netflix to improve its recommendations for them, so users have direct incentives to provide truthful and complete ratings. As a result, Shih et al. (2007) suggest that Netflix has the world's largest collection of movie ratings.

We believe that our data provide rich and strong evidence to study the impact of varying product variety on market concentration patterns. First, Dataset I is one of the most extensive source of information on the movie rental industry among all the related studies. Second, the revenue-sharing contract ensures the accuracy of the reported movie rental turns through considerable computer monitoring and external verification of the results. Finally, Dataset II allows us to observe temporal changes in the attraction of the movies

[1] In the rest of the paper, we refer to the movies and the TV series as well as other DVD content as movies for simplicity.

and the evolving customer preference. The combination of the two data sets allows doing so both at the movie and the consumer level, which is quite rare.

It is important to stipulate here that Dataset II only reflects the number of movies rated, but not all customers rate all movies that they watch. On the other hand, customers do not have to watch the movie at Netflix to be able to rate it. Admittedly, using rating data as proxy for actual rental demand at Netflix can be inconclusive, but the rating data can provide insights into what movies consumers are aware of and are interested in. First, previous literature (see Chen et al. 2004) suggested a strong connection between product demand and the number of consumer reviews. In our own data, we find a correlation of about 0.5 between the monthly number of ratings and the rental turns among the matched movies. Second, unlike other review data that are known to have selection bias because users tend to review items they extremely like or dislike (see Hu et al. 2007; Dellarocas and Narayan 2007; Dellarocas and Wood 2008 and citations therein), pure ratings may avoid this bias because giving a rating is much less costly to a user than writing a review. In our data, we plot the histogram of the rating values on a scale from one to five and find the rating of four to be the most frequent, followed by the ratings of three, five, two and one (more on, see Figure 2.3.3). The bell-shaped histogram seems to suggest that Netflix users may be unbiased toward rating the movies that they extremely like or detest. Third, the recommendation system of Netflix directly incentivizes and facilitates its users to reveal their truthful and complete preference for movies to improve their recommendations. Finally, the movie-level results of the Netflix data endorse the main results of the rental data. For all these reasons, we proceed with utilizing ratings as proxy for the attraction of a movie and the consumer preference to complement Dataset I at the consumer-level, although it should be understood that we imply the number of ratings.

2.3.2 Measures and Controls

In much of our analysis, we first elect to work with monthly (instead of weekly or yearly) data and therefore aggregate all variables at the monthly level. By doing so, we ensure both an adequate sample size in each month for each movie and enough observations over time for statistically significant estimates.

We are interested in studying varying demand for movies having different popularity levels. To reflect how popular movie j is at time t within a particular product offering set, we first rank the rental turns of each movie within each month in a descending order and use this rank as a proxy for movie popularity. Note that a higher (lower) rank indicates a less (more) popular movie. Then we propose two measures to categorize whether or not a movie is a hit. The first measure is the absolute ranking, e.g., the top 100, the top 1,000 movies, which is used in the previous literature including Anderson (2004). In this study, we define variable Hit_{jt} as a dummy variable, with one referring to movie j that ranks in the top 1,000 during month t and zero otherwise. Alternatively, we rank movies in relative terms, e.g., the top 10%, the top 20%, thus adjusting for current product variety (the total number of movies rented this month). Based on the relative rankings, we follow the A-B-C classification and classify the movies as follows: the movies that rank in the top 10% of all the movies rented that month are classified as Class A, i.e., highly popular movies. The movies that rank between top 10% and top 40% are classified as Class B. Finally, the movies that rank below top 40% are classified as Class C. For movie j during month t, we define a categorical variable $Class_{jt}$, with zero to be a Class A movie, one a Class B movie, and two a Class C movie.

In addition, we define $Variety_t$ as the total number of different DVD movies that were rented during month t. Note that this variable reflects the active product variety as many movies are not rented in a given month. We believe that the active product variety is a more relevant variable than the total variety that includes DVDs with no rentals because 1) products that are not discovered by consumers (or that are discovered but forgotten) should not be taken into account when ranking popularity and 2) it accounts for both the product offering and consumer demand. Clearly, by using active rather than total product variety, we are more likely to find evidence of the Long Tail effect. Furthermore, we measure the demand for individual movies with a proxy $Share_{jt}$, which reflects movie j's market share of the rental turns among all the rented movies within month t. This measurement allows us to adjust for the possible change in the consumer base. To analyze the changing distribution of the cumulative demand, we also compute the monthly Gini coefficient $Gini_t$, which is often used in social sciences as a measure of inequality of a distribution (e.g., Yitzhaki 1979;

Lambert and Aronson 1993). A $Gini_t$ of zero indicates total equality during month t, while a value of one suggests maximal inequality.

In addition to these main variables of interest, we consider the following control variables. We include a categorical variable $Trend_t$ to control for monthly characteristics.

To summarize, all variable definitions at the movie level are presented in Table 2.1.

Table 2.1: Movie-level Analysis Variable Definitions

Variable	Definition
Hit_{jt}	Categorical variable, with one indicating a movie that ranks among the top 1,000 during month t, zero otherwise.
$Class_{jt}$	Categorical variable, with zero indicating that movie j is a Class A movie, one a Class B movie, and two a Class C movie during month t.
$Variety_t$	Total number of rented movies during time period t.
$Share_{jt}$	Share of the number of times that movie j is rented among all the rented movies during time period t.
$Gini_t$	The Gini coefficient of demand distribution during month t.
$Trend_t$	Categorical variable of the 55 months in study.

In addition to the movie-level variables, we further define variables for our consumer-level analysis. We define $NicheSeeking_{it}$ as the average absolute ranking of the movies that consumer i rates in a given month t. In essence, *NicheSeeking* is a summary statistics of the movie level popularity. A high value of $NicheSeeking_{it}$ means that this particular consumer i tends to watch more niche movies. We calculate the mean, the median, the top 10%, and the bottom 10% of the rankings to obtain more complete information about consumer choices. Furthermore, we divide these metrics by monthly product variety to obtain relative measurements. These relative measurements adjust for both the increasing product variety and the skewness of demand distribution.

In order to control for consumer heterogeneity over time, we define $Frequency_{it}$ as the number of movies that user i rated in month t. In marketing, certain theoretical constructs such as the Dirichlet model suggest a strong link between purchase frequency and brand choice. In particular, it is often found that most consumers of a brand are low-frequency buyers (Chatfield and Goodhardt, 1975; Goodhardt et al., 1984). These light buyers often constitute the majority of the customers who purchase the popular brand (McPhee, 1963) because of the “super-star” effect (Rosen, 1981). In addition, McPhee (1963) explains that consumers who are familiar with the alternatives tend to consume the niche products. Therefore, consumers with high-consumption frequency are likely to consume more niche products than those with low-consumption frequency because the former may be better informed of the variety of products than the latter.

Furthermore, we define $RatingPropensity_{it}$ and $RatingVariance_{it}$ as the average and the variance of the ratings that user i gives in month t. These two measurements are likely to reflect people's tastes and movie acceptance. For example Clemons et al. (2006) demonstrate the relationship between variance of ratings and demand for products. Further, Hu et al. (2007) recommend controlling for the standard deviation of ratings as well as for two modes to overcome consumer under-reporting bias. Since in our case the distribution of ratings is symmetric, we do not control for the modes. All user-level variables are defined in Table 2.2.

Table 2.2: Consumer-level Analysis Variable Definitions

Variable	Definition
$NicheSeeking_{it}$	Popularity of the movies that user i rated in month t, measured as the mean and the median of the movie rankings, both in absolute and relative terms.
$Frequency_{it}$	Number of movies rated by user i in month t.
$RatingPropensity_{it}$	Average rating given by user i in month t.
$RatingVariance_{it}$	Variance of the ratings given by user i in month t.

2.3.3 Descriptive Statistics

Table 2.3 presents the descriptive statistics of the rentals by year. The product variety, which is the number of distinct movies available in the market, substantially increased from 7,246 in 2001 to 25,488 in 2005, up approximately two and half times. The total rentals also increased over twice from 162 million turns in 2001 to 546 million turns in 2004. For each title, the average turns seems to be stable. The skewness of the turns increased from 5.37 in 2001 to 6.94 in 2005, suggesting that the most popular movies are likely to contribute to an increasing market share. Furthermore, we observe that the minimum yearly turns per title dropped from 23 in 2001 to one in the following years, while the maximum yearly turns per title seem to be increasing from 728,526 in 2001 to over 1 million in 2004. This difference may suggest that the popular movies may be even more popular, while the obscure ones become even more obscure.

Table 2.3: Descriptive Statistics of Movie Rentals

Year	Product Variety	Total Rental Turns (in MN)	Avg Turns per Title	Skewness of Turns	Min Turns	Median Turns	Max Turns
2001	7,246	162	22,381.79	5.37	23	895	728,526
2002	10,975	245	22,319.38	5.40	1	845	933,998
2003	15,681	369	23,502.76	5.62	1	1,054	1,122,852
2004	23,255	546	23,459.19	6.02	1	2,461	1,019,122
2005*	25,488	352	13,812.03	6.94	1	1,555	682,397
Mean/year	16,529	335	21,095	5.87	5	1,362	897,379
Stdev/year	7,798	145	4,110	0.65	10	676	188,235

* We only observe seven months in 2005.

Figure 2.3.1 shows that the monthly product variety increased exponentially from January 2001 to July 2005 and that the rental turns increased linearly during the same period. A relevant question to ask is whether the product variety is growing because a lot of brand new movies are being released or because consumers keep discovering previously released titles. Our data indicate that the number of brand new titles increased from 1,639 in 2001 to 3,879 in 2004, while the newly rented back catalog titles decreased from 5,607 in 2001 to 3,707 in 2004. These observations suggest that the product variety growth is primarily due to introduction of brand new products. The more precise answer to this question is complicated by the fact that many movies are released on DVDs later than in theaters, but his gap continues to decrease over time. Further, most movies are released in several DVD versions at different points in time which makes it hard to exactly delineate rentals of "old" vs. "new" movies.

Figure 2.3.1: Monthly Product Variety and Rentals

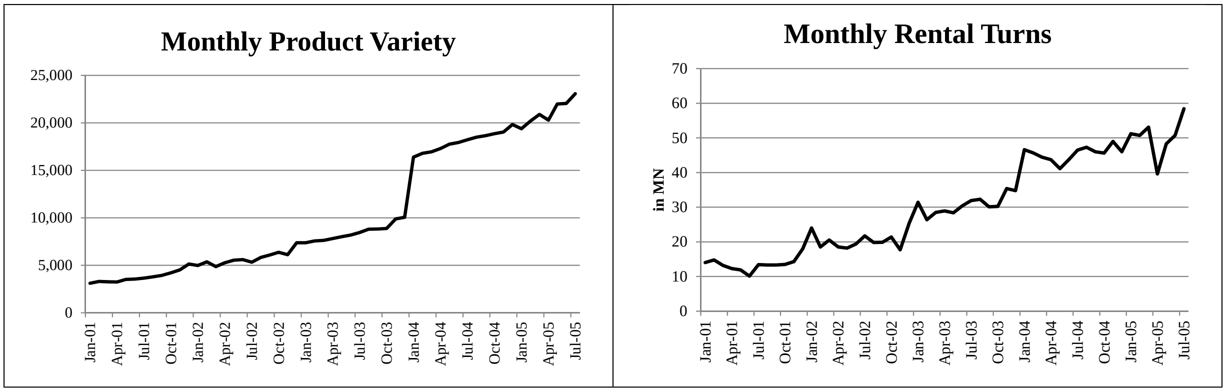


Figure 2.3.2 illustrates the distribution of the rental turns after pooling all the observations from 2001 to 2005. The demand is highly concentrated on a few titles, i.e., the hits: the top 10% of movies constitute close to 80% of the total rental turns; the top 20% of movies contribute to slightly over 92%.

Figure 2.3.2: Rental Distribution (2001-2005)

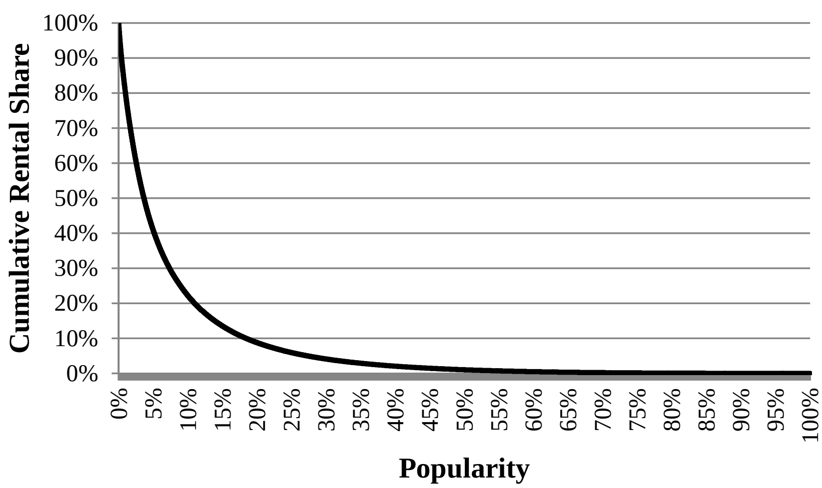


Table 2.4 shows the summary statistics at the consumer level. The number of movies rated per person every month is highly skewed toward the high percentiles, indicating that a small group of people rate a large number of movies each month. It is possible that the large number of movies rated, such as 41 for 90th percentile and 219 for 99th percentile, contain a large number of ratings given by users to train the recommendation system because a user can only watch a limited number of movies every month. Ratings submitted during the

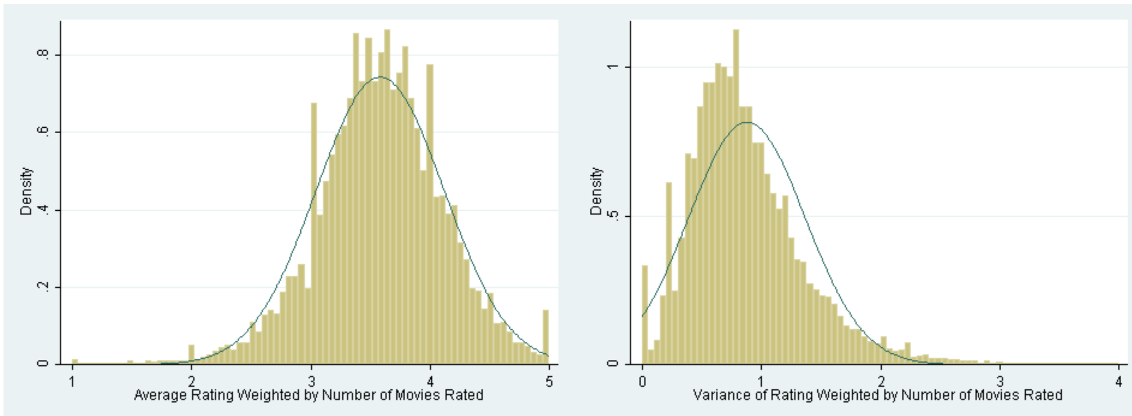
training process can result in a "contamination" of the data because the ratings of previously watched movies may not reflect the current popularity of a movie. In order to alleviate this issue and provide a robustness check, we purged the data with the monthly number of rated movies more than 30. We choose the cutoff point of 30 because watching 30 movies a month is probably the maximum number of movies that a heavy user is technically allowed to watch within Netflix rental system. Our results remain qualitatively and quantitatively similar so in the paper we report results without dropping any ratings.

Furthermore, it does not appear that heavy users have a tendency to give higher or lower ratings because the frequency of ratings very weakly correlates with the average rating (correlation = 0.0038) and with the variance of ratings (correlation = 0.1027). Figure 2.3.3 further shows that consumer ratings are almost normally distributed except that the right tail is censored at the rating of 5 because of the limit of the rating scale. This nearly normal distribution of consumer ratings provides statistical evidence that the users at Netflix may be unbiased toward rating the movies that they extremely like or extremely dislike. Furthermore, from the average variance of ratings (0.7) and the mean average rating (3.57), we compute that the coefficient of variation is approximately 0.23, suggesting that consumers tend to be stable in their ratings.

Table 2.4: Monthly Consumer Breakdown (N=4,041,165)

	Mean	Std	Skewness	1%	10%	25%	50%	75%	90%	99%
Number of Movies Rated	19	49	11	1	1	3	7	16	41	219
Average Rating	3.57	0.72	-0.42	1.33	2.75	3.12	3.6	4	4.5	5
Variance of Ratings	0.70	0.64	1.42	0	0	.22	.6	1	1.55	2.88

Figure 2.3.3: Monthly Rating Distribution



2.4 Estimation and Results

We test our hypothesis about the impact of product variety on demand concentration at two levels. First at the movie-level, we estimate regression models using three measures to examine the shift of demand concentration. We then use an instrumental variable approach to address the potential endogeneity issues. In order to gain insights into the movie-level analysis, we then turn our attention to examining the impact of product variety at the consumer-level. We employ fixed-effect models to understand how the propensity of each consumer to discover niche movies evolves and how the composition of his/her "basket" is affected by product variety.

2.4.1 Movie-level Analysis

We use three different measures to comprehensively examine the changing demand concentration. These measures include Gini coefficients (Model 2.4.1), the absolute measure, i.e., the aggregate market shares of movies having $Hit_{jt} = 1$ or $Niche_{jt} = 1$ (Model 2.4.2), and the relative measure, i.e., the aggregate market shares of Class A, Class B and Class C movies, respectively (Model 2.4.3). We specify the models as follows:

$$
\begin{aligned}
Gini_t &= \theta_0 + \theta_1\, Variety_t/1000 + \upsilon_t && (2.4.1)\\
\log(\sum_{j\in k} Share_{jt}) &= \alpha_0 + \alpha_1\, Variety_t/1000 + \varepsilon_{kt} \quad \forall k \in Hit = 1 \text{ or } Niche{=}1 && (2.4.2)\\
\log(\sum_{j\in l} Share_{jt}) &= \beta_0 + \beta_1\, Variety_t/1000 + \xi_{lt} \quad \forall l \in \text{same } Class && (2.4.3)
\end{aligned}
$$

In these three models, we divide $Variety_t$ by 1,000 and logarithmically transformed $Share_{jt}$ for interpretation purpose. We also use Huber-White estimation to correct standard errors. Although these regressions above provide a useful preliminary and exploratory analysis (Kennedy, 2003), they may be biased by the potential endogeneity of $Variety_t$. Retailers tend to adjust product variety based on sales forecast, thus causing *Share, Gini* and *Variety* to be endogenous. In addition, Hausman tests rejects the null that *Variety* are exogeneous. Other unobserved confounding factors, such as movie's intrinsic quality may also cause omitted variable bias in estimating the dependent variables. In order to address these potential endogeneity and omitted variables issues, we adopt an instrument variable 2SLS approach (Angrist and Krueger, 1994). The 2SLS instrument estimator can provide consistent estimate of the dependent variables. It is also quite robust in the presence of other estimating issues such as multicollinearity. In addition to its relative low computation cost, the 2SLS instrument variable approach is widely used to address the endogeneity issue (Kennedy, 2003).

A valid instrumental variable should satisfy relevance and exclusion restriction assumptions (Wooldridge, 2002). In particular, it should be uncorrelated with the error (i.e., exclusion restriction) and correlated with the endogenous regressor (i.e., relevance). In other words, the instrument should explain the outcome variable only through the endogenous regressor. Following Bloom and Van Reenen (2007), we propose using the lagged value of the independent variable as a candidate for valid instrument. In particular, we compute $Variety_{t-1}$ and $Variety_{t-2}$, previous two months' product varieties to use as instrumental variables for the current month. We expect these lagged *Variety*'s are exogenous because the product variety one month or two months ago should not determine the unobserved factors for the demand during the current month. In other words, the lagged variables are not contemporaneously correlated with the disturbance (Kennedy, 2003), so they should satisfy the exclusion restriction assumption of a valid instrument. Moreover, we expect that the lagged *Variety*'s are correlated with the current *Variety* and therefore satisfy the relevance assumption because retailers generally consider their previous product assortments to assort movies for the current period. We also provide relevant statistics to show the validity of these instruments after the results.

Table 2.5 shows the results of Models 2.4.1 through 2.4.3. As can be seen, the Gini coefficients increases in product variety (coefficients are 0.0019 in both models). The increasing Gini coefficients suggest that product variety may intensify demand concentration, supporting our hypothesis. In absolute measure, product variety is negatively associated with both the demand for top 1,000 movies and bottom 1,000 movies (coefficients are -0.0192 and -0.0188 for top 1,000 movies, -0.2004 and -0.1978 for bottom 1,000 movies). Note that the demand for bottom 1,000 movies drops approximately 10 times as much as the demand for top 1,000 movies, suggesting that the hits are more resistant to demand diversification than the niche movies. Furthermore, in relative measure, the demand for Class A (top 10%) movies increases in product variety (coefficients are 0.0054 and 0.0051). Product variety is further negatively associated with the demand for both Class B and Class C movies (coefficients are -0.0178 and -0.0167 for Class B movies, -0.0049 and -0.0053 for Class C movies). These results imply that expanded product variety may diversify the demand for hits in absolute terms, but intensify the demand for hits in relative terms. Moreover, product variety decreases the demand for niches in both absolute and relative terms, confirming our hypothesis. In particular, interpreting the coefficients of 2SLS results, we find that 1,000 additional products available in the market may increase the demand for Class A movies by 0.51%, but decrease the demand for Class B and Class C movies by 1.67% and 0.53% respectively.

Moreover, the 2SLS models show consistent results with the OLS models. To confirm the validity of the instruments and ensure asymptotic consistence of instrumental variable estimators, we first check the relevance condition. We find that all the F statistics for the joint significance of instruments excluded from the structural model over 1,000, much higher than the suggested rule of thumb 10 for weak instruments. In addition, we verify that the estimates of the instruments at the first stage are all statistically significant at 0.001 level. Hence, we confirm that these instruments have clear expected effects on the endogenous *Variety* and therefore satisfy the relevance condition. As for the exclusion restriction assumption, we acknowledge that there is generally no statistical test. Nevertheless, we would argue that the product variety two months ago should affect the demand only through current product variety. Finally, we conduct Sargan tests of overidentifying restrictions, which are often used as a test of exogenous instruments (Hayashi, 2000). We find that the p-values are over 0.5 and therefore fail to reject the null hypothesis that the error terms are uncorrelated with the instruments. Hence, we conclude that our instrumental variables are likely to satisfy the exclusion restriction assumption.

Table 2.5: Demand Concentration Results

	Gini Coefficient (Model 1)		Absolute Measure (Model 2)			
	OLS (Gini)	2SLS	OLS (Hit)	2SLS	OLS (Niche)	2SLS
Variety	0.0019***	0.0019***	-0.0192***	-0.0188***	-0.2004***	-0.1978***
	(0.0002)	(0.0001)	(0.0005)	(0.0005)	(0.0048)	(0.0044)
Constant	0.8079***	0.8083***	-0.0644***	-0.0697***	-4.7857***	-4.8240***
	(0.0018)	(0.0018)	(0.0065)	(0.0062)	(0.0644)	(0.0555)
Hypothesis supported	Yes	Yes	No	No	Yes	Yes
Observations	55	53	55	53	55	53
R^2	0.770	0.766	0.962	0.966	0.975	0.975

1) Standard errors are in parentheses

2) * p-value<0.05, **p-value<0.01, ***p-value<0.001

	Relative Measure (Model 3)					
	OLS (Class A)	2SLS	OLS (Class B)	2SLS	OLS (Class C)	2SLS
Variety	0.0054***	0.0051***	-0.0178***	-0.0167***	-0.0049**	-0.0053***
	(0.0004)	(0.0004)	(0.0014)	(0.0012)	(0.0016)	(0.0015)
Constant	-0.3581***	-0.3528***	-1.3424***	-1.3590***	-3.1568***	-3.1455***
	(0.0057)	(0.0051)	(0.0174)	(0.0154)	(0.0223)	(0.0194)
Hypothesis supported	Yes	Yes	Yes	Yes	Yes	Yes
Observations	55	53	55	53	55	53
R^2	0.756	0.757	0.780	0.786	0.143	0.227

1) Standard errors are in parentheses

2) * p-value<0.05, **p-value<0.01, ***p-value<0.001

2.4.2 Consumer-level Analysis

Consistent with the movie-level results of the rental data, the Netflix data show that product variety is positively associated with the demand for hit movies and negative associated with niche ones. We now turn to the Netflix data to understand how product variety changes individual consumers' preference for movies in order to gain insights into the movie-level analysis. In particular, we examine how the propensity of each consumer to discover niche movies evolves, while controlling for observed user heterogeneity, such as rating frequency and variance of ratings. The data that we have lack other potentially significant consumer characteristics, such as demographics. In order to cope with this issue, we introduce a time-invariant preference for each consumer's movies through the panel data analysis and we further assume that the preference correlates with the observed characteristics of the consumer. This correlation is likely to be caused by the recommendation systems, which can influence an individual's preference based on his/her observed characteristics. The Hausman test further provides strong evidence of this correlation. Therefore, we employ the following fixed-effect regression to predict consumer propensity to rate movies:

$$\begin{aligned}\log(\mathit{NicheSeeking}_{it}) &= \beta_0 + \beta_1 \mathit{Variety}_t/1000 + \beta_2 \mathit{Frequency}_{it} \\ &\quad + \beta_2 \mathit{RatingPropensity}_{it} + \beta_3 \mathit{RatingVariance}_{it} + \mu_i + \varepsilon_{it}. \qquad (2.4.4)\end{aligned}$$

The top of Table 2.6 shows the results of the absolute movie rankings while the bottom presents the same results of relative movie rankings using Model 2.4.4. As is evident from the top of the table, all $\mathit{Variety}_t$ coefficients are significantly positive, suggesting that the absolute popularity rankings of the movies watched by the average consumer consistently increase except for the very popular movies. In other words, consumers tend to discover more and more niche movies over time when movies are ranked in absolute terms. In particular, the $\mathit{Variety}_t$ coefficient for the bottom 10th percentile of the movies that a person rates (i.e., the obscure titles) is 0.0519, which is approximately 30 times as much as the same coefficient for the top 10th percentile of the movies (i.e., the popular titles). This comparison suggests that consumers are likely to discover niche products much faster than they move away from the hits (again, if popularity is measured in absolute terms).

However, the picture completely changes when the popularity of the movies is measured in relative terms. The bottom part of Table 2.6 shows that $\mathit{Variety}_t$ coefficients are consistently negative, suggesting that, relative to the product variety that is available at that point of time, consumers tend to be interested in more and more popular movies. In particular, the $Variety_t$ coefficient of the top 10th percentile of movies is -0.0919, which is over twice as much as the coefficient of the bottom 10th percentile of movies, suggesting that a consumer's attention shifts toward more popular hits faster than it shifts away from less popular niches.

Taken together, the results of Model 2.4.4 are consistent with the movie-level analysis in Subsection 2.4.1. They suggest that consumers do venture into more niche movies as product variety increases, but the growth rate of product variety is substantially higher than the speed at which consumers discover niche products. We explain that such quickly expanded product variety may cause consumers to anticipate high search costs and watch those movies that they are easily aware of. High concentration of advertising expenditures, word-of-mouth effects and theatrical release allow those popular movies to enjoy more exposure, thus enhancing their demand at the expense of other niche movies. In addition, this finding is consistent with the results of Fleder and Hosanagar (2009) that recommendation systems guide similar consumers to the same products, which does not effectively help consumers discover products at the tail of the distribution.

Table 2.6: Fixed-Effect Model of Nicheseeking Behavior

	Mean	Median	Top 10%	Bottom 10%
Variety	0.0519***	0.0432***	0.0017***	0.0519***
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Frequency	0.0007***	0.0009***	0.0029***	0.0009***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RatingPropensity	-0.1146***	-0.1563***	-0.2609***	-0.0737***
	(0.0010)	(0.0011)	(0.0014)	(0.0011)
RatingVariance	0.2559***	-0.0056***	-0.6601***	0.5156***
	(0.0010)	(0.0012)	(0.0015)	(0.0011)
Overall R^2	0.031	0.012	0.048	0.059

	Relative Mean	Relative Median	Relative Top 10%	Relative Bottom 10%
Variety	-0.0414***	-0.0504***	-0.0919***	-0.0417***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Frequency	0.0006***	0.0009***	0.0028***	0.0008***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RatingPropensity	-0.1184***	-0.1544***	-0.2590***	-0.0718***
	(0.0010)	(0.0011)	(0.0014)	(0.0011)
RatingVariance	0.2440***	-0.0064***	-0.6609***	0.5148***
	(0.0010)	(0.0012)	(0.0015)	(0.0012)
Overall R^2	0.030	0.016	0.064	0.062
Observations	4,740,731	4,740,731	4,740,731	4,740,731

1) Standard errors are in parentheses

2) * p-value<0.05, **p-value<0.01, ***p-value<0.001

Furthermore, the consistently positive and highly significant coefficients of $Frequency_{it}$ indicate that heavier users tend to discover more niche movies. In particular, the coefficients for both absolute and relative means are about 0.0006 and for both medians are 0.0009. In other words, if an average consumer watches five more movies per month, the mean of his/her propensity for niches is likely to increase by 0.3% on average.Accordingly the median is likely to increase by 0.45% on average, holding other factors constant. Thus, it appears that heavy users are the ones that drive toward the demand for niche products, which can cause Long Tail effect. Nevertheless, these heavy users constitute a relatively small segment of the entire population: as we demonstrated earlier, heavy users with a monthly frequency over the mean constitute less than 25% of all users. Although this small group of people tends to discover more niche movies, it does not seem to shift the entire demand from hits to niches. A comparison of coefficient for the top 10th percentile (0.0028) and coefficient for the bottom 10th percentile (0.0008) suggests that heavier users shift away from the hits approximately three times as fast as they discover niches. That is, even heavy users are not as fast in discovering niches as they are in "forgetting" about hits.

Consistently negative and highly significant coefficients of $RatingPropensity_{it}$ suggest that consumers who, on average, give higher ratings and may therefore be more satisfied tend to be interested in more popular movies. For example, the coefficients for both absolute and relative means are around -0.11, which suggests that increasing the average rating by one unit is associated with a 11% increase in the average movie popularity. In other words, the more popular movies generally satisfy people better than the obscure titles. Of course, it is possible that consumers who watch popular movies are systematically different from consumers who watch niche movies in that the former tend to rate all movies higher than the latter. Since we are unable to observe characteristics of individual consumers, our finding is subject to this limitation.

Finally, we note that $RatingVariance_{it}$ is negatively associated with the median and the top 10th percentiles, while this variable is positively associated with the mean and the bottom 10th percentiles (in either absolute or relative terms). We interpret these mixed signs to imply that consumers having highly disperse rating tend to watch extreme hits and niches. In other words, the extreme hits and the extreme niches receive more polarized ratings from those consumers. Further, the popularity of the movies that those consumers watch trends to skew toward the niches. In other words, consumers having highly disperse ratings watch a larger quantity of hits than niches, but the niches that they watch are generally extremely obscure.

2.5 Conclusion and Discussion

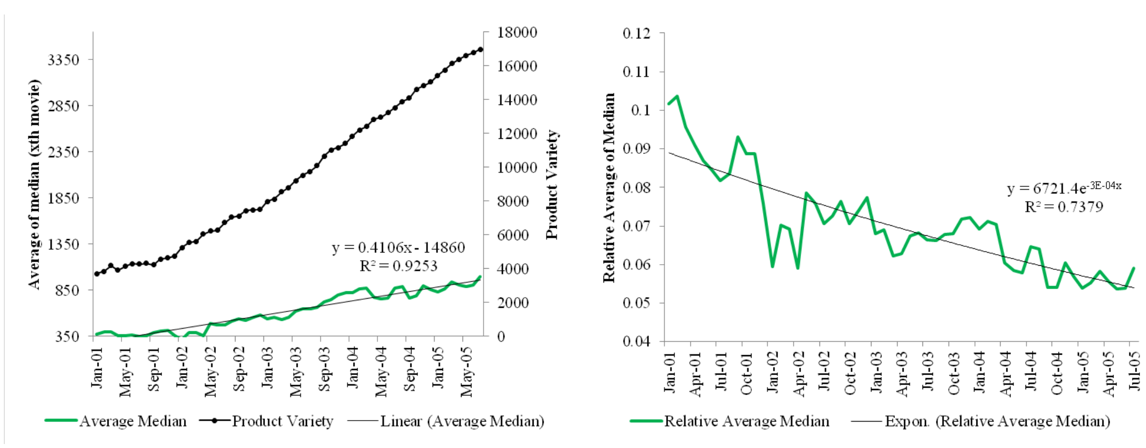
Many studies about the Long Tail effect suggested that expanding product variety due to adoption of the Internet will satisfy consumer's increasingly heterogeneous tastes, thus causing the demand for niches to rise. In this paper we empirically examine the impact of product variety on demand concentration, taking endogeneity into consideration. We argue that one has to be careful about defining hits and niches in the Internet era. In a brick-and-mortar world, where product variety is relatively stable and all products are consumed at some rate, hits and niches are typically defined in absolute terms (e.g., the top 100, the bottom 100 movies). However, product variety has been skyrocketing in the Internet age and therefore more and more products can be left unnoticed by consumers, or are being discovered very slowly, even though the customer base is also expanding. To evaluate the consumer propensity to discover niches and to separate this effect from the entirely different effect of increasing product variety on the Internet, we suggest that product popularity should be measured in relative terms, thus dynamically adjusting for the "active" product variety at that point of time. By doing this, we bring the distribution of demand to a common scale and we analyze how it changes over time.

We use two large data sets from the movie rental industry and analyze the data at both movie-level and consumer-level. In these large data sets we find that, when the popularity of a movie is defined dynamically, i.e., in relative terms, product variety increases the demand for hits, i.e., Class A movies, but it reduces the demand for niches. Additionally, even in the absolute ranking definition, the Long Tail effect is only partially present: product variety reduces the demand for niches more significantly than it reduces the demand for hits. We further find that product variety increases monthly Gini Coefficients, a measure of demand concentration.

In order to gain insights into those movie-level findings, we further examine changes in the demand distribution at the consumer level. Once again, we find that product variety indeed diversified consumers into more niche movies in absolute terms, but we also discover that the rate at which consumers shift demand from the hits to the niches is considerably lower than the growth rate of product variety. Therefore, if we normalize demand for currently active product variety and measure popularity in relative terms, we find that consumers tend to watch more and more hits as product variety grows. In other words, expanded product variety may threaten the attraction of one movie by Tom Cruise because of demand diversification, but it may favor more and more movies by Tom Cruise-like celebrities.

Figure 2.5.1 visually illustrates the comparison between the absolute and relative popularity of the movies that consumers watch. In Figure 2.5.1 (left, bottom line), we observe the median movie that the average consumer discovers over time, which has a linear upward trend, indicating that consumers increasingly discover niches. In particular, at the beginning of our study, the median movie rated by an average consumer was ranked slightly above 350, while at the end of the study the median movie ranking had increased more than twice to over 850. However, Figure 2.5.1 (left, top line) also indicates that product variety increased even faster, which creates an impression of the lengthening tail of demand distribution: there are more and more obscure movies over time. However, once we bring distribution to the common scale through dividing by current product variety, the claim of the increasing demand for niches disappears. Not surprisingly, Figure 2.5.1 (right) shows that, when we look at the median popularity in relative terms, the average consumer gravitates more and more toward hits. In fact, at the beginning of our study the average consumer are interested in the movies in the 10th percentile of product variety while at the end of the study the average consumer is interested in the movies in the 6th percentile. Hence, we conclude that although consumers do venture into niches, new movies appear more quickly than people can actually discover them.

Figure 2.5.1: Average Median and Relative Average Median Popularity



We make a number of additional observations based on our consumer-level analysis. We find evidence that the consumers who give high average ratings tend to watch more popular movies. Hence, we do not find any evidence that niche products satisfy consumer tastes better and better over time, which is suggested by Anderson (2004; 2006). Furthermore, we find that the consumers who discovers the niches tend to be heavy users, constituting only a small part of the entire user base. Light users, however, tend to focus on the popular items and since most users are in this category, hits continue to drive the market.

Our findings have a number of managerial implications as they shed new light on the controversy surrounding the Long Tail effect. First, the promise of the Long Tail effect became a basis for many new business models and business ideas (Anderson, 2006). Our findings suggest that caution needs to be used when assessing the potential benefits of focusing a business on supplying niche products. While it may be true that niche products are much more profitable for companies (e.g., Anderson 2006 rightfully suggests that niche movies cost a fraction of hit movies to make), this argument does not account for the fact that for each niche product that consumers demand, there might be several that are never discovered, thus potentially adding to the costs but not to the revenues. Irrational expansion into niche products will also increase operational difficulties, such as maintaining the level of service (Fisher et al., 1994; Randall and Ulrich, 2001). In fact, to compete against Netflix-like companies that stock a large product variety of niche movies, companies like Redbox successfully remain profitable by focusing only on a selected number of hit movies. In addition, Amazon.com, which is often cited as an example of offering numerous long tail products on its platform, is found to directly sell only a small percentage of all products listed on its website, with the most products being sold by third-party sellers because of insufficient demand for those niche products (Jiang et al., 2011).

Further, a large number of products might take a while to be discovered. This finding seems to suggest that much more attention needs to be paid to recommendation systems, review forums and other means of aiding product discovery. Although Netflix employs what is widely considered to be a sophisticated recommendation system, even this system does not allow numerous consumers to discover titles as fast as they appear. This raises an important issue of carefully forecasting how long will it take for a given title, once it is added to the inventory, to begin accumulating demand. More improvements to the recommendation systems, such through the Netflix Prize and the algorithm proposed by Park and Tuzhilin (2008) should be implemented.

Insights from our consumer-level analysis suggest that consumers are generally much more satisfied by hit products than by niches. This is an important consideration: while

Netflix currently achieves extremely high customer satisfaction, we do not find any evidence to suggest that customers watching obscure titles find them more satisfactory than other movies. We can speculate that many consumers over time will learn that niches are called niches for a reason and might start ignoring them altogether. Our other observation that heavy users tend to venture into more obscure movies suggests that the presence of the Long Tail effect might be moderated by the frequency of service. In the case of Netflix, it is physically impossible to rent more than a few DVDs per month (due to the time that the mailing process takes). However, Netflix and other companies (such as Amazon.com and Hulu.com) have started allowing customers to watch DVDs on their computers at home right away, which may increase the number of heavy users who discover niches. In this case, one will have to re-examine the demand concentration.

It is important to remember the limitations of our findings. First, our study does not directly compare the search costs between brick-and-mortar and Internet companies (e.g., Brynjolfsson et al. 2011) and therefore we are unable to comment on this aspect of the Long Tail effect. Rather, our findings need to be interpreted as a study about the impact of product variety on demand concentration only. Comparing the effects of product variety across channels would be of considerable interest for research. Second, our study has focused on the movie rental industry. It is possible and worthwhile to confirm that demand concentration in other product categories may respond differently to the varying product variety levels. Third, our consumer-level analysis is restricted to the ratings data. To address this issue, we confirm that the main results of rental data are consistent with those of the ratings data and we use the consumer-level rating data to offer additional insights into the movie-level analysis. Admittedly, the consumer-level results cannot be taken as exact evidence for individual rental behavior on Netflix.com. Some consumers probably do not rate the movies that they watched. Nevertheless, consumers also rate movies that they watched elsewhere, providing a richer picture of demand for movies which reflects interest, attention and satisfaction. An interesting venue for research, particularly for behavioral economics, would be to compare ratings data with time-stamped individual-level rental data to understand possible rating behavior biases. In addition, the data set with the time-stamped individual-level information may be potentially used to study consumer purchase patterns and their life-time value.

Further research opportunities also include linking recommendation system metrics, such as product ratings, with operations management and marketing strategies (see Netessine et al. 2006 for some initial work in this direction) is also a fruitful direction. Finally, incorporating the empirical findings of the product variety effects on demand concentration and evolving consumer preferences into the analytical models, such as dynamic assortment (Caro and Gallien, 2007) is highly warranted.

Chapter 3

When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity

3.1 Introduction and Related Literature

Labor is typically one of the largest cost components of service organizations such as retail stores, call centers and restaurants, and labor decisions are known to drive operational performance in services. For example, Zenios et al. (2011) find that hospitals could potentially reduce their staffing costs of nurses by 39% to 49% by deferring staffing decisions until more information about procedure type is available. In a retail setting, Perdikaki et al. (2011) suggest that store staffing levels influence the conversion of traffic into sales, although the sales return on labor increases diminishes. In another retail study, Mani et al. (2011) estimate that an optimal staffing level may improve average store profitability by 3.8% to 5.9%. Not surprisingly, many service companies are increasingly utilizing computerized staffing tools (Maher, 2007). In most of these scheduling systems, however, the employee's productivity is calculated using "grand averages" of historical data, thus overlooking employees' adaptive behavior towards changing work environments, as reflected, for instance, in a call center survey (Gans et al., 2003). In another example, Brown et al. (2005) find several anomalies indicating that some behavioral aspects of labor management may lead to serious staffing errors.

This simplified view of human productivity is inherited from classical operations management (OM) models, which often assume that humans in production or service systems are homogeneous and their productivity is independent from the state of the system, or at best that productivity has random variations (see Boudreau et al., 2003 and Bendoly et al., 2006 for comprehensive reviews). Recent efforts have bridged together OM models and human resource management in order to relax rigid assumptions of the classical OM models and study the impact of external factors on individuals' performance (see Boudreau (2004)). For example, Schultz et al. (1998) challenge the traditional OM assumption that a worker's production rate is independent from the environment. In a production line simulation experiment, they find that individuals' processing times are dependent on the system state, such as the buffer size as well as on the processing speed of co-workers. The experiment reveals less idle time and higher output because people tend to speed up and avoid idle time. Schultz et al. (1999) explain that a low-inventory system improves productivity because it creates more feedback, stronger group cohesiveness and task norms than a high-inventory system. Building on this work, Powell and Schultz (2004) further analyze the effect of line length on the throughput of a serial line. In another lab experiment, Bendoly and Prietula (2008) ask subjects to solve vehicle routing problems. They find a non-monotonic relationship between pressure (induced by the length of queue, i.e., workload) and motivation, which affects performance. They further show that this relationship can change as people receive training and improve skills. While this stream of research is experimental, real-world systems are generally more complex and therefore Boudreau et al. (2003) call for empirical studies to validate the behavioral lab findings in real industrial settings.

A very recent stream of empirical papers have answered this call. For example, Huckman et al. (2009) use detailed data from an Indian software company to study the impact of team composition on performance. They find that team familiarity has a positive impact on performance. In addition, Staats and Gino (2012) analyze data from a Japanese bank's home loan application-processing line to evaluate the impact of task specialization and variety on operational productivity. They show that specialization boosts short-term productivity; however, variety improves long-term productivity.

Closer to the question we pose in this study, several researchers have recently turned to understanding the impact of workload, an integral environmental factor, on individual's performance. These studies often use healthcare services as a test-bed. Kc and Terwiesch (2009) provide a rigorous empirical analysis of the impact of workload on service time and patient safety using operational data from patient transport services in cardiothoracic surgery. They show that workers speed up as workload increases, and that this positive effect of workload may be diminished after long periods of high workload. Kc and Terwiesch (2011) find further evidence that the occupancy level of a cardiac intensive care unit is negatively associated with patients' length of stay because the hospital, faced with high occupancy, is likely to discharge patients early. In addition, Green et al. (2011) establish that nurses tend to be intentionally absent from work if they anticipate a high workload. Powell et al. (2012) find that overworked physicians generate less revenue per patient from reimbursement because of their workload-induced reduction in diligence of paper execution.

Most of these observational studies find linear impacts of workload on performance. Kuntz et al. (2011), as an exception, suggest a non-linear relationship between hospital workload and mortality rates. We contribute to this stream of work by first proposing an inverted U-shaped relationship between workload and performance and testing it using a set of unique and very detailed transaction-level data from a restaurant chain's point-of-sales system that contains 195,311 check-level observations for five restaurants from August 2010 to June 2011. Furthermore, unlike other studies on this topic which focus on the total effect of the workload, we provide a mechanism of direct and indirect effects of workload on performance in a complex operational setting. We go further and demonstrate how to leverage staffing capacity to optimize the workload. After disentangling the endogeneity of demand and supply in this setting using a natural experiment (labor management software implementation) and other instruments, we find that servers react to the workload in the following way. Surprisingly, when overall workload is small, sales increase with the workload. However, after a certain threshold (around 5.76 diners per server) sales start decreasing with the workload. On average, the restaurant chain in our study has about 4.3 diners per server, so we conclude that it is largely overstaffed and, surprisingly, reducing the number of waiters can *both* significantly increase sales and reduce costs. While papers on restaurant management have analyzed the impact of pricing, table mix, table characteristics, food, atmosphere, fairness of wait and staff training on financial performance (see Kimes et al. 1998, 1999; Kimes and Robson 2004; Robson 1999; Kimes and Thompson 2004; Sulek and Hensley 2004), we contribute by showing that staff workload has a major impact on revenue generation.

3.2 Wait Staff Activities and Hypothesis Development

In the USA alone, the restaurant industry employs about 13 million workers, who have to be managed to provide over $500 billion in meals, yet we are not aware of any rigorous empirical studies of operations in this industry. For our analysis we select the restaurant setting because 1) workload in restaurants tends to be highly variable, which provides a great opportunity to study how changes in the workload affect servers' performance; 2) the restaurant industry is labor-intensive, employing approximately 10% of the total workforce in the United States; 3) its productivity is only half of the manufacturing industries, creating many opportunities for productivity improvement (Mill, 2004).

3.2.1 Wait Staff Activities

Waiters and waitresses, also known as servers, are restaurant employees who serve diners when they are seated. In a typical work scenario (Fields, 2007), servers first greet diners shortly after they are seated. They should instantaneously fill water glasses, present the menu and ask diners whether or not they would like anything from the bar. Then they should return to the table to present the specials and take the order. After presenting the food, servers should check on the table during the meal for any special requests or additional drink orders. Finally, they present checks and change on tips trays, thanking diners on their way out of the restaurant. In addition to these typical work activities, high-performance service tends to go beyond: good servers pay attention to diners' requests without being intrusive or letting them wait for too long, they anticipate diners' needs, and might suggest dishes and drinks without appearing aggressive. In sum, servers are an integral part of restaurant service operation. In particular, their performance affects meal duration, sales and guest satisfaction. According to a study by the National Restaurant Association (Mill, 2004), complaints about restaurant service far exceed complaints about food or atmosphere. The biggest complaints are about service speed and inattentive waiters, for example long waits to settle the bill and impatience with answering menu questions.

3.2.2 Hypotheses Development

Hourly sales are of great importance to restaurants which, on average, generate a very small pre-tax profit margin averaging just 4%. In order to increase hourly sales, restaurants typically train their servers to sell as much as possible to diners. Hourly sales clearly depend on meal duration. Controlling for demand, longer meal duration should create more sales opportunity because diners are unable to consume after leaving the restaurant . In this subsection we develop hypotheses about the impact of workload (defined as the number of diners assigned to a server) on servers' performance, which we measure using meal duration and sales. Previous research suggests that employees tend to adapt to the work environment (e.g., Schultz et al., 1998, 1999; Kc and Terwiesch, 2009) and our hypotheses are built on the premise that workload is an important external factor that may influence a server's performance in meal duration and sales as previous literature suggests Furthermore, we wish to distinguish two mechanisms by which servers affect sales: one is the indirect effect through meal duration and the other is the direct effect on sales.

3.2.2.1 Effect on Meal Duration

Naturally, diners' speed of eating primarily determines the meal duration. Nevertheless, one would expect a server's effort and attitude to significantly affect meal duration too: for example, an efficient server tends to quickly present menu and later the bill to expedite the order and check settlement procedures. Occasionally, a server may implicitly rush diners by presenting the check without being asked for it. In addition, s/he may choose to be quick when transporting food from the kitchen to the table. A diligent server may also be more attentive to answer diners' requests without leaving the diner to spend additional time waiting. Furthermore, a server may prolong a meal duration by offering more menu items, such as wine.

We argue that, when the workload is low, a higher level of workload will prolong meal duration. Operationally, when a server serves more diners, his/her attention is divided into smaller portions because of process sharing. Consequently, he/she may not address diners' needs promptly, thus extending meal duration. For example, diner i may need some assistance from his/her server, who is busy serving other diners. Therefore, diner i has to wait to get the server's attention. Furthermore, workload can be seen as a challenge and therefore a motivation stimulus (Deci et al., 1989). As workload increases, motivation also increases, which is shown to improve effort (Locke et al., 1978; Yeo and Neal, 2004). The server may be more motivated to make recommendations and suggest additional menu items. As a result, diners order extra food, which extends the meal.

However, when workload becomes too high, a higher level of workload may encourage servers to speed up. One reason is that servers may want to reduce the costs of customer waiting (e.g., waiting to settle the check) by accelerating service. Kc and Terwiesch(2009; 2011) find empirical evidence in the hospital setting that the higher workload reduces the service time. Moreover, when servers are overworked, they may cut corners, thus reducing service time (Oliva and Sterman, 2001). From a psychological perspective, a high workload may cause servers to become frustrated. Consequently, they may implicitly rush diners by presenting the checks without being asked for it. Similarly, Brown et al. (2005) show that call center agents intentionally hang up on callers to reduce their workload and obtain extra rest time. In sum, based on the arguments above, we propose an inverted U-shaped relationship between workload and meal duration.

HYPOTHESIS 1 (H1): *As workload increases, the meal duration first increases and then decreases.*

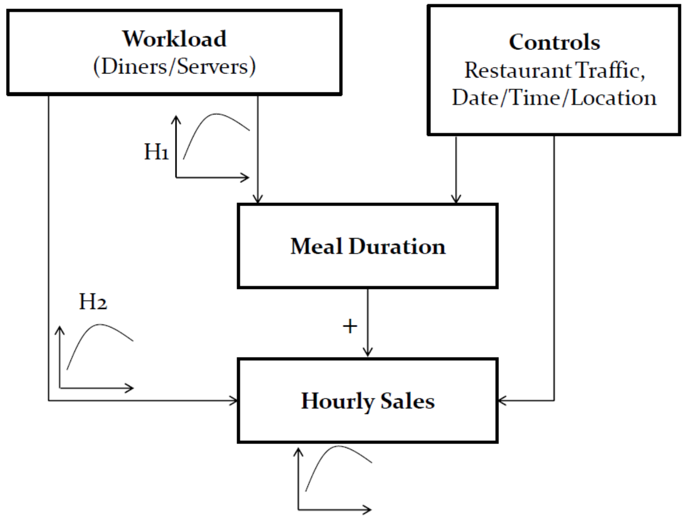
3.2.2.2 Effect on Sales

We expect that the diner's preference for items on the menu is the key factor that determines the sales per check. However, these preferences can be influenced by servers who make suggestive selling efforts which can increase restaurant sales (Fitzsimmons and Maurer, 1991). Such efforts include up-selling low-price menu items and cross-selling items that diners would otherwise not order. Research shows that diners are more likely to purchase a dessert or after-meal drinks if a server makes such suggestions when he/she is clearing the plates of the main course. In addition, servers' suggestions for appetizers, soup, wine and high-margin items are also known to stimulate demand that would otherwise be unexpressed. Thus, when the workload is low, an increasing workload may motivate servers to exert a more suggestive selling effort for psychological reasons similar to those stated above. Similarly, cognitive psychology suggests that workload may trigger the cortex to release hormones that improve cognitive performance (Lupien et al., 2007). Moreover, Parkinson's Law (Parkinson, 1958) suggests that employees tend to fill in the idle time with irrelevant activities, creating inefficiency, such as smoking outside, chatting with each other and folding napkins, as opposed to making effort in selling more items. Hence, increasing the workload may reduce servers' idle time, thus increasing selling effort. However, when workload surpasses a critical level and becomes too high, it is likely to limit sales per check: servers may become so occupied with carrying food that they have no time to conduct suggestive selling. In addition, they may be distracted by other diners, reducing their sales service effectiveness. Fatigue caused by heavy workload may also lead to reduced effort (Cakir et al., 1980; Setyawati, 1995). Servers may also suffer from "contact overload", which refers to the emotional drain from handling too many customers over a prolonged period of time (Mill, 2004). The results of the contact overload can be emotional burnout, which reduces a server's sales effort and effectiveness. For these reasons, we hypothesize that:

HYPOTHESIS 2 (H2): *As workload increases, hourly sales first increases and then decreases.*

Figure 3.2.1 summarizes the mechanism of how workload affects hourly sales. We decompose the effects of workload on hourly sales to an indirect effect via meal duration (i.e., H1) and a direct effect (i.e., H2). As previously hypothesized, both direct and indirect effects are inverted U-shaped. Adding these effects together, the total effect of workload on hourly sales is still inverted U-shaped.

Figure 3.2.1: Summary of Hypothesis Development



3.3 Data

3.3.1 Research Setting and Data Collection

To examine our research hypotheses, we worked closely with a restaurant chain management to collect point-of-sales (POS) data from five restaurants owned and operated by Alpha (the real name is disguised for confidentiality reasons), a restaurant chain that offers family-style casual dining service in the Boston suburbs. We gained access to their sales data as part of implementing a new server scheduling system, which is later used for identification purposes in Subsection 3.4.2. The restaurants are open from 11:30 am to 10:00 pm from Monday to Thursday, and 11:30 am to 11:00 pm from Friday to Sunday. Diners include couples, families, students and their friends. The restaurants have a full-service bar and offer internationally-inspired fusion food. Our study focuses on the main dining room because the bar and the take-out services operate according to a different business model and they would require different operationalization of variables. The data that we possess consists of 11 months of transactions from August 2010 to June 2011. The transaction data includes information about sales, gratuities, party size, when each service started and ended, and who was the server. In order to reduce the influence of the outliers (e.g., very large parties and private events), we drop those transactions which include the day's top and bottom 7.5% of checks. Our final data set includes 195,311 check-level observations. We believe that our restaurant sample represents an appropriate data set to study the impact of workload on restaurant performance because we possess comprehensive temporal and monetary information for each meal service that occurred during both busy and non-busy hours, allowing us to systematically quantify the impact of workload on servers' performance. At the same time, the data set we possess is among the largest and most granular in the extant literature.

3.3.2 Measures and Controls

Restaurants tend to schedule servers on an hourly basis so in our analysis we elect to focus on hourly (instead of check-level or daily) data and therefore we aggregate all variables at the hourly level (comprehensive analysis of this and other assumptions is presented in Subsection 3.4.6). We are interested in studying restaurant performance and therefore we operationalize dependent variables $AvgMealDuration_{tk}$ and $HRsales_{tk}$ to reflect the length of the average meal and how much in total was spent in restaurant k during hour t. We first infer the meal duration of each check from check opening and closing times recorded in our POS data. Naturally, this inferred duration could be slightly inaccurate because diners could arrive before the check was opened and they could leave after the check was closed. Nevertheless, our meal duration measure directly captures server's involvement with the customer (rather than, say, host's involvement before check is open) and it is also consistent with previous literature (Kimes, 2004). We next average each meal duration over the number of checks that were started in hour t.

Furthermore, we define the key independent variable $HRLoad_{tk}$ as the workload during hour t at restaurant k. It is computed as the number of diners who started meals during hour t divided by the number of servers who processed at least one check in the same hour. We provide alternative definitions of workload using the number of tables and menu items in Subsection 3.4.6. Our data only captures how many servers handled checks in an hour, which may be fewer than the actual number of servers available. Thus our workload metric should be understood as "conditional on the server being busy". In addition, different servers may sometimes accommodate a different number of diners in the same hour. This effect, however, is small, since servers are on average assigned a similar number of diners in an hour because the host/hostess is instructed to evenly assign diners to balance workload and potential tips for servers. To verify this, we further check that the coefficient of variation of the hourly number of diners assigned to each server suggests a low variation of about 0.33. Furthermore, we provide a server-level analysis in Subsection 3.4.6.2 as a robustness check.

In addition to these main variables of interest, we consider the following control variables. Variable $HRDiners_{tk}$ is the number of diners who started their meals at restaurant k during hour t. It controls for the demand, which we expect to be positively associated with hourly sales. Variable *HRDiners* helps account for the load on the kitchen and other functions in the restaurants, which affect the average meal duration. Similarly, variable $HRItems_{tk}$ is the number of menu items sold in those meals that started at restaurant k during hour t. These items sold increase the general workload on the kitchen, thus affecting average meal duration during hour t. Finally, we also control for the date/shift/location of hour t. Night shifts usually generate more sales than lunch shifts so we include control variable $Shift_t$. In addition, weekends (from Friday dinner to Sunday lunch) are usually busy hours of the week so we include a day of the week control, $DayWeek_t$. Furthermore, business during summer months is usually slower than during winter months because many residents go on vacation, so we include $Month_t$ as a control. Finally, we account for economic trends using the variable $Trend_t$ and we control for store fixed effects using the variable $Store_{tk}$. To summarize, Table 3.1 presents a list of variable definitions. These data allow us to test our hypotheses while controlling for factors that can affect a restaurant's performance.

Table 3.1: Hourly-level Analysis Variable Definition

Variable	Definition
$AvgMealDuration_{tk}$	Average duration of the meals that started during hour t at restaurant k.
$HRSales_{tk}$	Total sales of the meal services started during hour t at restaurant k.
$HRLoad_{tk}$	Average load of diners on servers during hour t at restaurant k, measured as the division of hourly number of diners and servers.
$HRTableLoad_{tk}$	Average load of tables on servers during hour t at restaurant k, measured as the division of hourly number of tables occupied and servers.
$HRItemLoad_{tk}$	Average load of items on servers during hour t at restaurant k, measured as the division of hourly number of menu items sold and servers.
$HRDiners_{tk}$	Number of diners started meal service at restaurant k during hour t.
$HRItems_{tk}$	Number of menu items sold during the meals started at restaurant k during hour t.
$DayWeek_t$	Categorical variable indicating the day of week of hour t.
$Shift_t$	Categorical variable indicating whether hour t was during lunch or dinner shift.
$Month_t$	Categorical variable indicating the month of hour t.
$Trend_t$	Continuous variable controlling for daily trend.
$Store_{tk}$	Categorical variable indicating store k during hour t.

3.3.3 Descriptive Statistics

Table 3.2 shows the summary statistics of hourly variables. On average, each meal lasts approximately 47 minutes, generating hourly sales of $452.47 for each restaurant. About 26 diners enter a restaurant during an average hour. In addition, each restaurant staffs on average close to six servers per hour, which results in an hourly workload of 4.3 diners per server.

Table 3.2: Summary Statistics of Hourly Variables

	AvgMealDuration	*HRSales*	*HRDiners*	Number of Servers per Hour	*HRLoad*
N	17,432	17,432	17,432	17,432	17,432
Mean	47.03	442.93	26.23	5.63	4.30
Stdev	8.09	335.10	19.43	3.20	1.66
Min	21.85	9.98	1	1	1
P5	34.95	44.17	3	1	2
P25	41.94	154.47	9	3	3
P50	46.69	367.13	22	6	4.14
P75	51.55	681.66	40	8	5.33
P95	59.99	1,068.30	62	11	7.20
Max	109.23	2,045.69	117	18	15.50

Table 3.3 provides the averages of *HRDiners, AvgMealDuration* and *HRSales* by day of week and by shift across the five restaurants in our study. In terms of traffic measured in *HRDiners*, Saturday has the highest hourly number of diners across stores during both lunch shift (the mean is approximately 31 diners) and dinner shift (the mean is close to 38 diners). In contrast, Tuesday is the slowest day for lunch (the mean is slightly below 18 diners), and Monday is the slowest day for dinner (the mean is 23 diners). In terms of average meal duration, dinner duration is typically longer than lunch duration by approximately three minutes. In addition, during each shift, the difference between average meal duration on each day of the week is less than four minutes. Furthermore, Sunday generates highest sales across lunch shifts, while Saturday creates highest sales across dinner shifts.

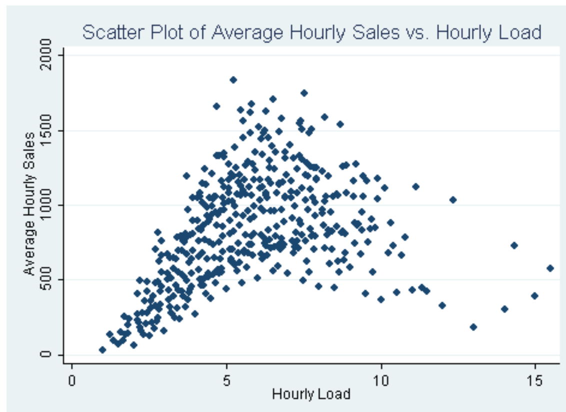
Table 3.3: Average Statistics of Performance by Day of Week and Shift

	Lunch			Dinner		
	HRDiners	*AvgMealDuration*	*HRSales*	*HRDiners*	*AvgMealDuration*	*HRSales*
Sunday	29.97	46.87	489.91	29.43	47.77	533.79
Monday	18.27	44.94	264.89	23.02	48.80	409.01
Tuesday	17.86	45.72	257.31	25.46	48.83	462.26
Wednesday	19.40	45.74	284.35	26.61	48.76	488.53
Thursday	20.24	45.00	295.40	29.29	47.96	541.23
Friday	23.42	46.45	350.42	36.40	48.70	684.17
Saturday	30.69	45.31	466.48	37.97	48.61	715.67
Total	22.79	45.71	343.13	30.07	48.50	554.64

Figure 3.3.1 shows the scatter plot of hourly sales vs. workload. The y-axis is the average

hourly sales for each value of hourly workload. Even visually, it appears that hourly sales is a concave function of the hourly workload, which is already suggestive, but since the graph does not consider other potential confounding factors, we proceed with statistical analysis to identify the effect of workload on hourly sales.

Figure 3.3.1: Scatter Plot of Hourly Sales vs. Load



Before testing our hypotheses, we transform *HRSales*, and *AvgMealDuration* into their natural logarithms in order to linearize the regression model (Kleinbaum et al., 2007). These variables have large standard deviations relative to their means so transforming them is recommended to increase normality prior to model estimation (Afifi et al., 2004). Transforming the monetary variable normalizes the scale to percentages for easier interpretation. We further center *HRLoad* and $HRLoad^2$ at the mean for interpretation purposes.

Table 3.4 shows the correlations of the hourly-level variables. We observe that *HRLoad* is positively associated with log(*HRSales*) (correlation = 0.647), log(*AvgMealDuration*) (correlation = 0.144) and *HRDiners* (correlation = 0.679). In addition, log(*HRSales*) and *HRDiners* are highly correlated (correlation = 0.887), which is not surprising. log(*AvgMealDuration*) and log(*HRSales*) are also positively associated (correlation = 0.306), which is expected. Further, we do not observe systematic trend effects in log(*HRSales*), *HRLoad* and *HRDiners* because the correlations are low and insignificant.

Table 3.4: Correlation Matrix of Hourly-level Variables

	log ($HRSales$)	log($AvgMealDuration$)	$HRLoad$	$HRDiners$	$Trend$
log ($HRSales$)	1				
log($AvgMealDuration$)	0.3055*	1			
$HRLoad$	0.6471*	0.1443*	1		
$HRDiners$	0.8870*	0.2179*	0.6785*	1	
$Trend$	0.0127	0.0810*	-0.0037	0.0182	1

*: Significant at 0.01 level.

3.4 Estimation and Results

First, we estimate a set of multivariate regression models to provide a preliminary and exploratory analysis. Second, we use an instrumental variable approach to address the potential endogeneity issues. Finally, we utilize simultaneous equation modeling to address both the endogeneity and correlated errors issues.

3.4.1 Multivariate Regression

We first specify the following multivariate regression model to provide a preliminary analysis of the relationship between workload and servers' performance:

$$\begin{aligned} \log(AvgMealDuration_{tk}) &= \alpha_0 + \alpha_1 HRLoad_{tk} + \alpha_2 HRLoad_{tk}^2 + \alpha_3 HRDiners_{tk} + \\ &\quad \alpha_4 HRItems_{tk} + \alpha_5 \text{Controls}_{tk} + \varepsilon_{tk}, \quad (3.4.1) \\ \log(HRSales)_{tk} &= \beta_0 + \beta_1 HRLoad_{tk} + \beta_2 HRLoad_{tk}^2 + \beta_3 \log(AvgMealDuration_{tk}) \\ &\quad + \beta_4 HRDiners_{tk} + \beta_5 \text{Controls}_{tk} + \mu_{tk}. \quad (3.4.2) \end{aligned}$$

In this model, Controls_{tk} include $Shift_{tk} \times DayWeek_{tk}$, $Month_{tk}$, $Trend_{tk}$ and $Store_{tk}$ to adjust for the time/date and location factors which is equivalent to a store fixed-effect model because we include store-specific time-invariant factors among controls which help control for unobserved heterogeneity among stores, such as the income level of the neighborhood and other time-invariant omitted variables. Note that the quadratic specification of $HRLoad_{tk}$ allows us to compute the critical points in the regression models. In particular, since the critical point of a quadratic function of the form $f(x) = ax^2 + bx + c$ is $-b/(2a)$, the critical point of, e.g., $\log(HRSales_{tk})$ is expected to be $-\beta_2/(2\beta_3)$.

Models 3.4.1 and 3.4.2 are essentially a set of mediation models which have been widely used in behavioral and social science to understand the implicit nature of the relationship between dependent and independent variables (Baron and Kenny, 1986; Amenta et al., 1992; Preacher and Hayes, 2008). In particular, they allow us to identify the direct and indirect effects of *HRLoad* on log(*HRSales*) via *AvgMealDuration*. In the mediation models, the indirect effect is approximately the coefficients between the independent variables (i.e., *HRLoad* and *HRLoad*^{2}) and the mediator (i.e., *MealDuration*). For example, the indirect effect of *HRLoad* on log(*HRSales*) is approximately $\alpha_1 \cdot \beta_4$, while the indirect effect of *HRLoad*^{2} is approximately $\alpha_2 \cdot \beta_4$. Coefficients β_1 and β_2 are the direct effects. The sum of the direct and the indirect effects are referred to as the total effects. In other words, $(\alpha_1 \cdot \beta_4 + \beta_1)$ is the total effect of *HRSales*, while $(\alpha_2 \cdot \beta_4 + \beta_2)$ is the total effect of *HRSales*^{2}.

Although these regression models are useful as a preliminary estimator (Kennedy, 2003), they may not address two potential issues:

1. Endogeneity: Restaurants tend to schedule servers based on sales and traffic forecast at the hourly level, thus causing *HRLoad, HRLoad*^{2}*,* log(*HRSales*), and log(*AvgMealDuration*) to be endogenously determined. Other unobserved confounding factors, such as customer demographic information, may also cause omitted variable bias in estimating the dependent variables. In order to address these potential endogeneity and omitted variables issues, we first adopt an instrumental variable 2SLS approach (Angrist and Krueger, 1994) and then a 3SLS approach (Zellner and Theil, 1962), which are elaborated in subsections 3.4.2 and 3.4.3.

2. Correlated Errors: Meal duration and sales are two performance metrics. They may be simultaneously affected by an unobserved exogenous demand shock, such as a celebration of a baseball game, but Models 3.4.1 and 3.4.2 assume that errors ε_{tk} and μ_{tk} are uncorrelated, thus eliminating the connection of these two measures via a contemporaneous shock. We propose a simultaneous approach using 3SLS models to allow the errors ε_{tk} and μ_{tk} to be correlated with each other (see subsection 3.4.3).

3.4.2 2SLS Model

We adopt an instrumental variable 2SLS approach (Angrist and Krueger, 1994) to address the endogeneity issue for the following reason. First, the 2SLS instrument estimator can provide consistent estimates of the dependent variables using a large sample. It is also quite robust in the presence of other estimation issues such as multicollinearity. For these reasons, the 2SLS instrument variable approach is widely used to address the endogeneity issues (Kennedy, 2003). A valid instrumental variable should satisfy relevance and exclusion restriction assumptions (Wooldridge, 2002). In particular, it should be uncorrelated with the error (i.e., exclusion restriction) and correlated with the endogenous regressor (i.e., relevance). In other words, the instrument should explain the outcome variable only through the endogenous regressor.

We propose two types of instruments. First, we utilize an exogenous shock in our study period: the implementation of a new staffing system at one of the restaurants. On March 21st, 2011, one of the restaurants adopted a new computer-based scheduling system, while the other four restaurants continued to rely on managers to make demand forecasts and make staffing level decisions. The management chose this particular restaurant as a pilot project to subsequently implement the software chain-wide. The sales performance of this restaurant is similar to the other four restaurants in that they all show stable sales, thus reducing the concern of selection bias. Using historical sales data, the new software forecasts the need for servers. It is reasonable to assume that the system will prescribe different staffing levels from those that managers might suggest because it uses more historical sales data than a manager can handle. In other words, this system should have affected staffing levels after its implementation, thus satisfying the relevance condition. In addition, we expect the implementation of the software to affect meal duration and sales only through the staffing level because the system simply provides a user-friendly interface to schedule servers, perhaps with a different forecast for demand. Diners do not observe the implementation of this labor scheduling system. For these reasons, the implementation of the system should satisfy the exclusion restriction condition.

Admittedly, both managers and servers in that particular restaurant may have anticipated the implementation of the new software. They may also have different emotional responses to a computerized scheduling system. For these reasons, they might have readjusted their productivity, which could invalidate using the software implementation as an instrument. In order to address this potential issue, following Bloom and Van Reenen (2007) and Siebert and Zubanov (2010), we supplement our analysis using another type of instrumental variables, the lagged values of the endogenous independent variables. In particular, we compute the $HRLoad$ and $HRLoad^2$ of the same restaurant during the same hour of the previous week to use as instruments for the current week. For example, if the observation tk happened at 8:00 pm on 8/8/2010 at restaurant k, its instrument is the hourly load of the 8:00 pm slot on 8/1/2010 at restaurant k. We then mean-center these instruments for interpretation purposes. We expect that the weekly lagged variables are correlated with the current terms and therefore satisfy the relevance assumption because the restaurants in our study usually consider the load from a week ago to generate staff schedules for the current week. Moreover, we expect these lagged values of the endogenous variables to be exogenous because the staffing decisions from a week ago should not determine the unobserved factors for the average meal duration and sales during the current week, i.e., contemporaneous shocks. In other words, the lagged variables are not contemporaneously correlated with the disturbance (Kennedy, 2003), so they should satisfy the exclusion restriction assumption of a valid instrument. Admittedly, the lagged workload may not be ideal in the event of common demand shocks that are correlated over time. However, these common demand shocks are basically trends (Villas-Boas and Winer, 1999). Trends are controlled for in our models, thus lessening this potential concern. We further provide relevant statistics to show the validity of these instruments in subsection 3.4.5. With both types of instrumental variables, we employ the following 2SLS estimation procedure:

Stage 1: Estimate endogenous independent variables, namely $HRLoad$ and $HRLoad^2$ using OLS and utilizing instrumental variables (i.e., the implementation of the scheduling system and the lagged values) and other exogenous controls (specified in Models 3.4.1 and 3.4.2); compute predicted values of the endogenous independent variables $\widehat{HRLoad}$, and $\widehat{HRLoad}^2$.

Stage 2: Use the predicted values of the endogenous independent variables, namely

$\widehat{HRLoad}$ and $\widehat{HRLoad}^2$ to estimate the coefficients of each equation in the system (Models 3.4.1 and 3.4.2) using OLS.

3.4.3 Simultaneous Equations

The OLS models 3.4.1 and 3.4.2 assume that the unobserved errors of log(*HRSales*) and log(*AvgMealDuration*) are uncorrelated with each other. In order to allow for correlated errors between the number of checks and sales, in addition to addressing the potential endogeneity issues, we adopt a system of simultaneous equations using a three-stage least squares (3SLS) estimation method (Zellner and Theil, 1962) for the following reasons. First, the 3SLS instrument estimation can provide consistent estimates of *HRLoad* and *HRLoad*^{2}. It is also quite robust in the presence of other estimating issues such as multicollinearity. Furthermore, the system of simultaneous equations approach utilizes all available information in the estimates and is, therefore, more efficient than a single equation (Kennedy, 2003). We use the same instruments as described in subsection 3.4.2 and propose the following estimation procedure.

Stage 1: Same as the first stage in the 2SLS approach.

Stage 2: After using the predicted values from Stage 1 to estimate the coefficients of each equation, we use these 2SLS estimates to predict errors in the system of simultaneous equations, i.e., structural equation's errors. These predicted errors are further used to compute the contemporaneous variance-covariance matrix of the structural equation's errors. In other words,

$$\begin{aligned}
\text{Stage 2: } \log(AvgMealDuration_{tk}) &= \alpha_0 + \alpha_1 \widehat{HRload}_{tk} + \alpha_2 \widehat{HRload}^2_{tk} + \alpha_3 HRDiners_{tk} \\
&\quad + \alpha_4 \log(HRItems_{tk}) + \alpha_5 \text{Controls}_{tk} + \varepsilon_{tk}, \\
\log(HRSales)_{tk} &= \beta_0 + \beta_1 \widehat{HRload}_{tk} + \beta_2 \widehat{HRload}^2_{tk} + \beta_3 \log(AvgMealDuration_{tk}) \\
&\quad + \beta_4 HRDiners_{tk} + \beta_5 \text{Controls}_{tk} + \mu_{tk}.
\end{aligned} \tag{3.4.3}$$

where ε_{tk} and μ_{tk} are structural equation's errors.

Stage 3: Compute the General Least Squares (GLS) estimators of the system of Equations 3.4.3.

3.4.4 Results

Table 3.5 shows the results of the impact of *HRLoad* on log(*AvgMealDuration*). The coefficients of $HRLoad^2$ are consistently negative (-0.0025, -0.0117, -0.0116), suggesting that *HRLoad* initially concavely increases the average meal duration of each check and then concavely decreases the meal duration, consistent with H1. The OLS model predicts that the critical point is above the current sample mean, while 2SLS and 3SLS models suggest that the critical point should be below the current sample mean. We acknowledge that the point estimates of *HRLoad* are statistically insignificant at 0.05 level in both 2SLS and 3SLS models. We expect that the instruments will increase standard errors of the estimates because they reduce the variation of the $\widehat{HRDiners}_{tk}$. We also observe that the estimates of *HRLoad* and $HRLoad^2$ by 2SLS and 3SLS are over twice as small as those by OLS, which may indicate weak instruments. We explain why our instruments are not weak instruments in subsection 3.4.5.

Table 3.5: Impact of *HRLoad* on log(*AvgMealDuration*)

	OLS	2SLS	3SLS
HRLoad	0.0110***	-0.0259	-0.0259
	(0.0014)	(0.0221)	(0.0221)
$HRLoad^2$	-0.0025***	-0.0117*	-0.0116*
	(0.0004)	(0.0051)	(0.0051)
HRDiners	-0.0079***	-0.0027	-0.0027
	(0.0003)	(0.0016)	(0.0016)
HRItems	0.0043***	0.0030***	0.0030***
	(0.0001)	(0.0003)	(0.0003)
Controls	Yes	Yes	Yes
Hypothesis Supported	H1	H1	H1
Observations	17,428	16,389	16,389
Prob>Chi-Sq	<0.001	<0.001	<0.001

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

Table 3.6 presents the results of the effects of *HRLoad* on hourly sales, controlling for the demand, i.e., *HRDiners*. As can be seen, the coefficients of *HRLoad* are consistently positive (0.1361, 0.419 and 0.4846). The coefficients of $HRLoad^2$ are consistently negative (-0.0412, -0.1677, -0.1363). Supporting H2, these results suggest that *HRLoad* first concavely increases sales and then concavely decreases sales. In other words, when workload is small, sales increase in the workload because servers exert more selling effort. However, after a certain threshold, workload limits servers' sales effectiveness. Furthermore, as expected, a long average meal duration and the number of diners are both positively associated with higher hourly sales.

Table 3.6: Impact of *HRLoad* on log(*HRSales*)

	OLS	2SLS	3SLS
HRLoad	0.1361***	0.4190***	0.4846***
	(0.0035)	(0.0820)	(0.0916)
$HRLoad^2$	-0.0412***	-0.1677***	-0.1363***
	(0.0025)	(0.0195)	(0.0231)
log(*AvgMealDuration*)	0.4934***	0.2212***	2.4810***
	(0.0264)	(0.0511)	(0.4142)
HRDiners	0.0358***	0.0225***	0.0139*
	(0.0003)	(0.0046)	(0.0058)
Controls	Yes	Yes	Yes
Hypothesis Supported	H2	H2	H2
Observations	17,428	16,389	16,389
Prob>Chi-Sq	<0.001	<0.001	<0.001

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

In sum, Table 3.6 suggests that *HRLoad* has a direct concave effect on log(*HRSales*). Moreover, it has a possible indirect concave effect via log(*AvgMealDuration*). In particular, using the estimates from 3SLS, we find the indirect effect of $HRLoad^2$ is approximately -0.0288 ($-0.0116\times2.481\approx -0.0288$). The indirect effect of *HRLoad* is zero because the estimate is insignificant at 0.05 level. Adding both direct and indirect effects, the total effect of *HRLoad* is approximately 0.4846; the total effect of $HRLoad^2$ is approximately -0.165, which suggests that the total optimal workload is 1.46 diners/server ($\frac{0.4846}{0.33} \approx 1.46$) above the current sample mean. These results allow us to conclude that workload affects hourly sales non-linearly. When workload is small, sales increase in the workload: "The Devil makes work for idle hands"; however, after a certain threshold sales start decreasing in the workload: "Many hands make light work".

3.4.5 Validity of Instrumental Variables

To confirm the validity of the instruments and ensure asymptotic consistence of instrumental variable estimators, we first check the relevance condition. In the 2SLS estimations of the $\log(AvgMealDuration)$ and $\log(HRSales)$, the adjusted R^2's from the first-stage regressions of *HRLoad* are around 0.5. The adjusted R^2's from the regressions of $HRLoad^2$ are 0.08 and 0.06 respectively, which may indicate considerable loss of precision but not so low as to cause a weak-instruments issue. We find that all the F statistics for the joint significance of instruments excluded from the structural model are over 10, the suggested rule of thumb of weak instruments (Staiger and Stock, 1994). In addition, we find that the software implementation is negatively associated with *HRLoad* due to increased staffing levels after implementation. One-week lagged hourly staffing is also negatively associated with the workload in the current week, suggesting that the management adjusted the staffing decisions. Hence, we confirm that these instruments have clear expected effects on the endogenous variables and therefore satisfy the relevance condition.

Unfortunately, there is no generally accepted statistical test for the exclusion restriction assumption. Nevertheless, we would argue that the implementation of the software should affect the restaurant performance only through staffing levels, without affecting demand factors or the service quality of individual servers. Moreover, from our interviews with restaurant managers and industry knowledge, we believe that that hourly staffing levels from one week ago should be independent from the contemporaneous shock to the meal duration and sales of the current week, after controlling for both time varying and time invariant effects. Finally, we conduct Sargan tests of over-identifying restrictions, which are often used as a test of exogenous instruments. We find that the p-values are over 0.05 and therefore fail to reject the null hypothesis that the error terms are uncorrelated with the instruments. Hence, we conclude that our instrumental variables are likely to satisfy the exclusion restriction assumption.

3.4.6 Robustness Check

3.4.6.1 Alternative Definitions of Workload

In addition to the number of diners that a server serves, his/her workload may be measured differently. One alternative measure is the number of tables that a server waits on. Servers tend to perform a set of procedures, such as taking orders and settling bills, one table at a time. In addition, restaurants generally assign a section of tables to a server, so servers are aware of how many tables that they are responsible for. For these reasons, the number of tables served could reflect a server's workload. Another alternative measure is the number of menu items sold. Additional items sold will increase a server's workload because the server needs to carry the item from the kitchen or the bar to the table. We use variables *HRTableLoad* and *HRItemLoad* defined in Table 3.1 to reflect the workload in terms of tables and items. Following the estimation procedures described from Subsection 3.4.1 to 3.4.3, we substitute *HRLoad* with *HRTableLoad* and *HRItemLoad*, respectively, to provide robustness checks of our main results.

Table 3.7 shows the 3SLS estimation results using the alternative workload definitions. These results are largely consistent with our main results using diners as a proxy for workload. In estimating $\log(AvgMealDuration)$, the coefficient of $HRTableLoad^2$ is -0.0917, suggesting that workload measured in tables initially concavely increases the average meal duration of each check and then concavely decreases the meal duration. The coefficient of $HRItemLoad^2$ is also negative ($= -0.0013$), although it is insignificant at 0.05 level. The average meal duration may not be particularly sensitive to the number of items possibly because it sometimes take a server the same amount of time to carry one or two items during one trip from the kitchen.

In estimating $\log(HRSales)$, the coefficients of *HRTableLoad* and *HRItemLoad* are both significant and positive (0.8083 and 0.1244). The coefficients of $HRTableLoad^2$ and $HRItemLoad^2$ are both significant and negative (-0.6567 and -0.0265), consistent with H2 and our main results. In addition, as expected, the average meal duration has a positive impact on hourly sales. Using these 3SLS estimates, we compute that the optimal *HRTableLoad* is about 0.5 table/server above the current sample mean (1.84 tables/server). The optimal *HRItemLoad* is 2.35 items/server above the current sample mean (9.44 items/server). These optimal workload levels using alternative measures suggest that the optimal workload level in diners, i.e., 1.46 diners/server,is quite robust because 2.6 diners on average sit at one table and order about two items per diner in our sample.

Table 3.7: 3SLS Results of Alternative Definitions of Workload

	Table Load		Item Load	
	log(*AvgMealDuration*)	log(*HRSales*)	log(*AvgMealDuration*)	log(*HRSales*)
HRTableLoad	-0.0098	0.8083***		
	(0.0328)	(0.0991)		
$HRTableLoad^2$	-0.0917**	-0.6567***		
	(0.0290)	(0.1079)		
HRItemLoad			-0.0163	0.1244***
			(0.0085)	(0.0226)
$HRItemLoad^2$			-0.0013	-0.0265***
			(0.0011)	(0.0037)
HRDiners	-0.0052***	0.0271***	-0.0102***	0.0257***
	(0.0006)	(0.0012)	(0.0013)	(0.0020)
HRItems	0.0034***		0.0065***	
	(0.0003)		(0.0011)	
log(*AvgMealDuration*)		1.5968***		2.0959***
		(0.2500)		(0.5811)
Controls	Yes	Yes	Yes	Yes
Hypothesis Supported	H1	H2	-	H2
Observations	16,389	16,389	16,389	16,389
Prob>Chi-Sq	<0.001	<0.001	<0.001	<0.001

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

3.4.6.2 Heterogeneous Servers

Servers usually possess idiosyncratic sales skills. These heterogeneous servers form different team compositions during each business hour, which may affect hourly sales. In this subsection, we account for such heterogeneous sales skills to reduce the potential omitted variable bias in estimating the effects of workload. Furthermore, we examine the moderating effects of sales skills to provide insights for scheduling the heterogeneous servers under various workload levels.

Hourly Team Composition

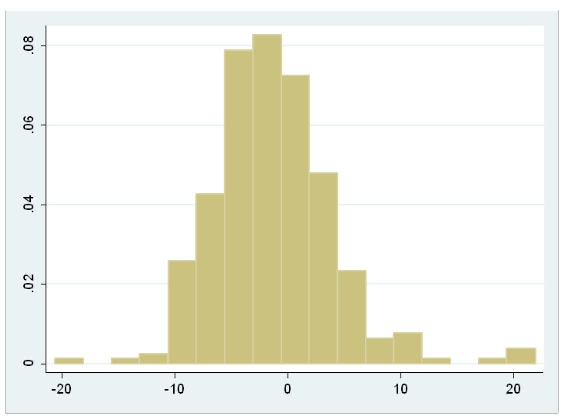
In order to address these issues, we first use the following fixed-effect model to estimate servers' intrinsic sales abilities:

$$Sales_{jt} = \alpha_0 + \alpha_1 MyDiners_{jt} + \alpha_2 \text{Controls}_{jt} + \mu_j + \varepsilon_{jt}.$$

In this model, $Sales_{jt}$ is the total sales that server j generates during hour t. $MyDiners_{jt}$ is the total number of diners that server j starts to serve during hour t. In other words, $MyDiners_{jt}$ spend $Sales_{jt}$ with server j during hour t. In addition, similar to previous models, Controls_{jt} include $Shift_{jt} \times DayWeek_{jt}$, $Month_{jt}$, $Trend_{jt}$ and $Store_{jt}$ to adjust for the the time/date and location factors. The fixed effects μ_j is server j's time-invariant intrinsic sales ability. We use such a fixed-effect model instead of a random-effect model because the intrinsic sales ability μ_j may be correlated with $MyDiners_{jt}$.

Figure 3.4.1 shows the distribution of the intrinsic sales ability μ_j. The distribution of μ_j seems to be symmetrically dispersed around the mean, which is approximately $-1.81.

Figure 3.4.1: Distribution of Servers' Intrinsic Sales Abilities μ_j



We average the intrinsic sales ability μ_j's of the servers working in the same hour t at restaurant k to create a variable $HRAvgSkill_{tk}$. We then include $HRAvgSkill_{tk}$ in our previous 3SLS Model 3.4.3. Table 3.8 shows the results of the robustness check including hourly team composition. The coefficient estimates of $HRLoad$ and $HRLoad^2$ are consistent with the main results shown in Tables 3.5 and 3.6 in terms of both signs and magnitudes. Therefore, H1 and H2 are supported. Note that the $HRAvgSkill$ estimates are insignificant in both log ($AvgMealDuration$) and log ($HRSales$) models. In this study, we do not interpret the effect of $HRAvgSkill$ and use it primarily as a control variable of the team composition. Understanding the causal effect of team composition warrants careful future research.

Table 3.8: Robustness Check of Hourly Team Composition Using 3SLS

	log($AvgMealDuration$)	log($HRSales$)
$HRLoad$	-0.0267	0.4897***
	(0.0223)	(0.0932)
$HRLoad^2$	-0.0116*	-0.1342***
	(0.0052)	(0.0240)
$HRDiners$	-0.0027	0.0133*
	(0.0017)	(0.0061)
$HRItems$	0.0030***	
	(0.0003)	
log($AvgMealDuration$)		2.5397***
		(0.4486)
$HRSalesPerc$	-0.0008	-0.0028
	(0.0010)	(0.0037)
Controls	Yes	Yes
Hypothesis Supported	H1	H2
Observations	16,389	16,389
Prob>Chi-Sq	<0.001	<0.001

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

Moderating Effects of Sales Abilities

We further conduct a server-level analysis to understand the moderating effects of sales skills with respect to workload. A server-level analysis also provides a robustness check to our previous assumption that servers are assigned similar number of diners every hour. Using the previously-estimated intrinsic sales ability μ_j's, we categorize servers into two groups – those servers having $\mu_j \geq -1.81$ (sample mean) are coded as the "high" type and are given a dummy variable $Highsales_j = 1$. Accordingly, those servers having $\mu_j < -1.81$ are the "low" type and are given $Highsales_j = 0$. We can alternatively categorize servers according to different quantiles of the μ_j, which yields similar results. Therefore we only report the categorization based on the sample mean. We include the categorical variable $HighSales_j$

in the following two multivariate regression models:

$$\begin{aligned}
\log(AvgMealDuration_{jt}) &= \alpha_0 + \alpha_1 HighSales_j + \alpha_2 HighSales_j \times MyDiners_{jt} + \\
&\quad \alpha_3 HighSales_j \times MyDiners_{jt}^2 + \alpha_4 Tenure_{jt} + \\
&\quad \alpha_5 \text{SystemLoad}_{tk} + \alpha_6 \text{Controls}_{jt} + \xi_{jt} \qquad (3.4.4) \\
\log(HRSales_{jt}) &= \beta_0 + \beta_1 HighSes_j + \beta_2 HighSales_j \times MyDiners_{jt} + \\
&\quad \beta_3 HighSales_j \times MyDiners_{jt}^2 + \beta_4 \log(AvgMealDuration_{jt}) + \\
&\quad \beta_5 Tenure_{jt} + \beta_6 \text{SystemLoad}_{tk} + \beta_7 \text{Controls}_{jt} + \varepsilon_{jt}. \qquad (3.4.5)
\end{aligned}$$

In these models, $MyDiners_{jt}$, the total number of diners that server j served during hour t, measures the server-level workload. Variable $Tenure_{jt}$, which is the number of days that server j has worked by hour t, controls for servers' learning effect. SystemLoad_{tk} include $HRDiners_{tk}$ and $HRItems_{tk}$ defined in Table 3.1. Controls_{jt} include all the previous time/date and location controls.

Table 3.9 shows the results of Models 3.4.4 and 3.4.5. Figure 3.4.2 graphically summarizes the moderating effects of sales abilities. First, the coefficient of *HighSales* is -0.0297 in Model 3.4.4 and 0.0776 in Model 3.4.5, suggesting that the "high"-type servers are on average associated with 3% shorter meal duration and 8% higher hourly sales than the "low"-type ones. This result implies that managers might benefit from scheduling more "high"-type servers to work during such high-traffic shifts as Friday dinners or weekends in order to ensure prompt service levels and generate higher total sales.

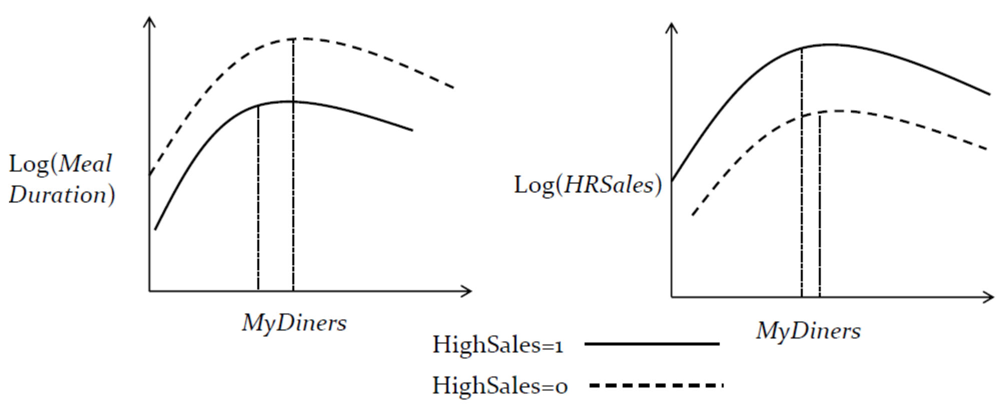
Second, the coefficients of $MyDiners^2$ are significant and negative (-0.001 and -0.0141) in both models and they are indistinguishable between the "high"-type and the "low"-type servers. These results support both H1 and H2, suggesting that workload has an inverted-U shaped relationship with the meal duration and the hourly sales regardless of servers' sales abilities. Furthermore, using the coefficient estimates of the linear term *MyDiners* and its corresponding interactive term, we find that the "high"-type servers seem to start speeding up under a smaller workload than the "low"-type ones. The sales of the "high"-type servers also seem to start decreasing under a lower workload than the "low"-type ones probably because the diners' food consumption constraints limit "high"-type servers' highest sales.

Table 3.9: Server-level Analysis: Moderating Effects of Sales Abilities

	log(*AvgMealDuration*) Model 3.4.4	log(*HRSales*) Model 3.4.5
HighSales = 1	-0.0297***	0.0776***
	(0.0025)	(0.0028)
MyDiners	0.0131***	0.2252***
	(0.0007)	(0.0007)
MyDiners× *HighSales = 1*	-0.0050***	-0.0019*
	(0.0008)	(0.0009)
$MyDiners^2$	-0.0010***	-0.0141***
	(0.0001)	(0.0002)
$MyDiners^2 \times HighSales=1$	0.0002	0.0004
	(0.0002)	(0.0002)
Tenure	-0.0007*	0.0013***
	(0.0003)	(0.0003)
log(*AvgMealDuration*)		0.2117***
		(0.0035)
SystemLoad	Yes	Yes
Controls	Yes	Yes
Hypothesis Supported	H1	H2
Observations	98,018	98,018
Prob>Chi-Sq	<0.001	<0.001

1. Standard errors are shown in the parentheses.
2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

Figure 3.4.2: Summary Plots: Moderating Effects of Sales Abilities



3.4.6.3 Spline Regressions

Previous studies examined linear effects of workload on performance. While we are one of the first studies suggesting a non-linear effect of workload and propose a quadratic function form of *HRLoad*, we are limited to the "non-local" assumption of a quadratic regression. The "non-local" assumption implies that the the fitted dependent variables, i.e., $\log(\widehat{AvgMealDuration})$ and $\log(\widehat{HRSales})$, at a given $HRLoad = HRLoad_0$ depends heavily on *HRSales* values far from $HRLoad_0$. In order to address this issue, we apply spline regressions, choosing n knots that split *HRLoad* into $n+1$ equal-sized groups. In particular, we choose n to be 2 and 3, respectively. Spline regressions then fit piecewise linear functions of *HRLoad.*

Table 3.10 shows the results of the spline regressions. In all models, the coefficients of *HRLoad1* are significant and positive (0.0169, 0.2369, 0.0224 and 0.3590), suggesting that as workload increases, both average meal duration and hourly sales first increase. However, the coefficients of *HRLoad2* are all significant and negative (-0.0046, -0.1232, -0.0074, -0.2595), implying that as workload further rises, both average meal duration and hourly sales then drop. Note that the absolute values of *HRLoad1* coefficients are all larger than the absolute values of *HRLoad2* coefficients, which suggest that the incremental effect of workload may be stronger than the decremental effect of workload on both average meal duration and hourly sales. Furthermore, the coefficient of *HRLoad3* is indistinguishable from zero in estimating $\log(AvgMealDuration)$. In estimating $\log(HRSales)$, the coefficient of *HRLoad3* is significant and negative (-0.0242), but its absolute value is approximately one tenth of the absolute value of *HRLoad2* coefficient (-0.2592). These *HRLoad3* coefficient estimates seem to suggest that the decremental effects of *HRLoad* may be diminishing as workload further increases. All in all, these results from spline regressions further support our H1 and H2, suggesting that as workload increases, both average meal duration and hourly sales first increase and then decrease.

Table 3.10: Robustness Check Using Spline Regressions

	log(*AvgMealDuration*) $n=2$	log(*HRSales*) $n=2$	log(*AvgMealDuration*) $n=3$	log(*HRSales*) $n=3$
HRLoad1	0.0169***	0.2369***	0.0224***	0.3590***
	(0.0019)	(0.0049)	(0.0026)	(0.0060)
HRLoad2	-0.0046**	-0.1232***	-0.0074*	-0.2595***
	(0.0015)	(0.0049)	(0.0031)	(0.0066)
HRLoad3			-0.0005	-0.0242***
			(0.0018)	(0.0047)
HRDiners	-0.0078***	0.0363***	-0.0078***	0.0375***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
HRItems	0.0042***		0.0042***	
	(0.0001)		(0.0001)	
log(*AvgMealDuration*)		0.4862***		0.4618***
		(0.0262)		(0.0246)
Controls	Yes	Yes	Yes	Yes
Hypothesis Supported	H1	H2	H1	H2
Observations	17,428	17,428	17,428	17,428
Prob>Chi-Sq	<0.001	<0.001	<0.001	<0.001

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

3.4.6.4 Quantile Regressions

As an additional robustness check, we adopt quantile regression models, which have been used in the applied economics (Koenker and Bassett Jr, 1978; Buchinsky, 1994). Unlike the OLS Models 3.4.1 and 3.4.2, which estimate the average effects of *HRLoad,* a quantile regression estimates the effect of *HRLoad* on conditional quantiles of *AvgMealDuration* and *HRSales*. Analyzing these quantiles allows us to understand how the covariates including *HRLoad* affects the average meal duration and hourly sales of different levels. In addition, using a quantile regression to estimate the median is more robust to large outliers than using an OLS prediction. We employ the following quantile regression models:

$$\begin{aligned}\log(\mathit{AvgMealDuration}_{tk}) &= A'_{tk}\alpha_\theta + u_{\theta tk} \quad \text{with Quantile}_\theta(\log(\mathit{AvgMealDuration}_{tk}|A_{tk})) = A'_{tk}\alpha_\theta,\\ \log(\mathit{HRSales}_{tk}) &= B'_{tk}\beta_\theta + v_{\theta tk} \quad \text{with Quantile}_\theta(\log(\mathit{HRSales}_{tk}|B_{tk})) = B'_{tk}\beta_\theta,\end{aligned}$$

where A'_{tk} and B'_{tk} are the same group of independent variables as in Models 3.4.1 and 3.4.2. In addition, $\text{Quantile}_\theta(\log(\mathit{AvgMealDuration}_{tk}|A_{tk})$ and $\text{Quantile}_\theta(\log(\mathit{HRSales}_{tk}|B_{tk})$ are the θth conditional quantiles of $\log(AvgMealDuration_{tk})$ and $\log(HRSales_{tk})$.

Table 3.11 shows the results of the quantile regressions at 25%, 50% and 75% quantiles. The left three columns present the results of $\log(AvgMealDuration)$. The coefficients of $HRLoad^2$ are significant and negative (-0.0052 and -0.0028) for the 25% and 50% quantiles, supporting our H1. For the 75% quantile, the inverted-U shaped relationship is inconclusive . The right three columns show the results of $\log(HRSales)$. The coefficients of $HRLoad^2$ are consistently significant and negative (-0.0833, -0.0516 and -0.0231) across quantiles, supporting our H2. In addition, the coefficients of *HRLoad* are consistently positive (0.1720, 0.1083 and 0.0562), consistent with our primary results as shown in Table 3.6.

Table 3.11: Robustness Checks Using Quantile Regressions

	$\log(AvgMealDuration)$			$\log(HRSales)$		
	25 % Quantile	50% Quantile	75% Quantile	25% Quantile	50% Quantile	75% Quantile
HRLoad	0.0177***	0.0111***	0.0053***	0.1720***	0.1083***	0.0562***
	(0.0012)	(0.0016)	(0.0014)	(0.0032)	(0.0037)	(0.0039)
$HRLoad^2$	-0.0052***	-0.0028***	-0.0005	-0.0833***	-0.0516***	-0.0231***
	(0.0005)	(0.0006)	(0.0003)	(0.0023)	(0.0019)	(0.0026)
HRDiners	-0.0056***	-0.0065***	-0.0073***	0.0354***	0.0348***	0.0340***
	(0.0003)	(0.0003)	(0.0004)	(0.0002)	(0.0003)	(0.0003)
HRItems	0.0036***	0.0035***	0.0034***			
	(0.0002)	(0.0002)	(0.0002)			
$\log(AvgMealDuration)$				0.4802***	0.4339***	0.3608***
				(0.0326)	(0.0226)	(0.0140)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Hypothesis Supported	H1	H1	not supported	H2	H2	H2
Observations	17,428	17,428	17,428	17,428	17,428	17,428
Prob>Chi-Sq	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

1. Standard errors are shown in the parentheses.
2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

3.4.6.5 Duration Analysis Models

Finally, we conduct a duration analysis of the average meal duration to estimate how workload affects the hazard rate of ending a meal. The hazard rate of a meal may depend on how long a meal has already elapsed, i.e., time-dependence, exhibiting increasing, decreasing or bell-shaped distributions. We fit a variety of commonly used distributions and check the robustness of the results. In particular, we use the same model specification as in Model 3.4.1 and assume that *AvgMealDuration* follows a Gompertz, a Weibull, a Log-logistic and a Log-normal distribution, respectively. The duration analysis results can be interpreted from a causal perspective (Blossfeld et al., 2007). Nevertheless, unobserved heterogeneity may make the results spurious. In order to address this issue, we include a Gamma-distributed error term in the hazard function, i.e., Gamma mixture. Finally, we also consider a Cox semi-parametric hazard model, but its results violate the proportional hazards assumption. Therefore we elect not to report its results.

Table 3.12 presents the results of the duration analysis models. Note we report the accelerated failure time results of Weibull, Log-logistic and Log-normal distributions, so the signs of the coefficients are consistent with the meal duration. However, we report the log relative-hazard results of Gompertz distribution because Gompertz distribution assumes linear monotonicity in duration. Therefore, a positive coefficient implies a higher hazard rate and a shorter duration.

As can be seen, the coefficient of $HRLoad^2$ is positive in the Gompertz model (0.0426), and negative in the three accelerated failure time models (-0.0008, -0.0008 and -0.0011, respectively). In addition, the coefficient of *HRLoad* is negative in the Gompertz model (-0.1723), and positive in the other three models (0.0032, 0.0033, 0.0042, respectively). These results are all significant and consistent, supporting our H1 that workload has an inverted-U-shaped relationship with meal duration. In addition, the signs of the *HRDiners* and *HRItems* coefficients are the same as our main results shown in Table 3.5. Moreover, the shape parameters of the Gompertz and the Weibull models are both positive (13.3710 and 3.8670), suggesting an increasing hazard rate over time. In the Log-logistic and the Log-normal models, the shape parameters are both negative, suggesting an increasing and then decreasing rate of finishing a meal, namely a bell-shaped hazard rate. In terms of the goodness-of-fit of the four models, the Gompertz model yields the highest log-likelihood (31,546), which suggests that the baseline hazard rate is likely to be linearly increasing.

Table 3.12: Robustness Check Using Duration Analysis

	Gompertz	Weibull	Log-logistic	Log-normal
HRLoad	-0.1723***	0.0032***	0.0033***	0.0042***
	(0.0152)	(0.0003)	(0.0003)	(0.0003)
$HRLoad^2$	0.0426***	-0.0008***	-0.0008***	-0.0011***
	(0.0042)	(0.0001)	(0.0001)	(0.0001)
HRDiners	0.0915***	-0.0018***	-0.0018***	-0.0019***
	(0.0040)	(0.0001)	(0.0001)	(0.0001)
HRItems	-0.0515***	0.0010***	0.0010***	0.0011***
	(0.0019)	(0.0000)	(0.0000)	(0.0000)
Shape Parameter	13.3710***	3.8670***	-3.8689***	-3.3020***
	(0.1710)	(0.0124)	(0.0088)	(0.0080)
Controls	Yes	Yes	Yes	Yes
Hypothesis Supported	H1	H1	H1	H1
Observations	17,428	17,428	17,428	17,428
Prob>Chi-Sq	<0.001	<0.001	<0.001	<0.001
Log-likelihood	31,546	31,425	31,453	30,943

1. Standard errors are shown in the parentheses.

2. *: p-value<=0.05, **: p-value<=0.01, ***: p-value<=0.001

3. Gompertz is interpreted in a log relative-hazard form. A positive sign implies a higher hazard and a shorter duration. Other three models are interpreted in an accelerated failure time form. A positive sign implies a longer duration.

3.5 Managerial Insights and Concluding Remarks

3.5.1 Managerial Insights

Our study underscores several insights for restaurant managers facing the increasing challenges and pressures of managing a diverse workforce in a highly demanding work environment. Making optimal staffing decisions is critical for restaurants to achieve better performance. Perhaps the most counter-intuitive finding of our study is that *reducing* the staffing level might be necessary to improve sales. We find that the optimal workload for hourly sales is approximately 1.46 diners per server above the current sample mean, controlling for the demand. The hourly diners/server ratio in our sample stores is currently on average equals to 4.3, with a standard deviation of 1.66. Our findings indicate that the optimal staffing of 5.76 diners per server may simultaneously increase sales and reduce labor costs. Using the estimates in the 3SLS estimation of log(*HRSales*), we project that the optimal staffing will directly increase the average hourly sales by approximately 41%, controlling for the demand and average meal duration. A sanity check suggests that selling one additional starter or dessert or a good glass of wine through effective up-selling or cross-selling may be worth about 40% of the price of an entree, which makes this estimated sales lift reasonable. In addition, we re-estimate the sales impact after controlling for *HRItems* using 3SLS estimation in Model 3.4.3 to categorize the sales lift into up-selling and cross-selling. It seems reasonable to assume that controlling for *HRItems* leads to isolating the cross-selling effect. We find that the optimal staffing level may result in 12% sales lift in terms of up-selling. In other words, remaining 29% of the sales lift ascribes to the cross-selling. Similar to other restaurants, our focal restaurants offer plenty of add-on options, making 12% up-selling sales lift plausible.

Nevertheless, increasing hourly workload may lead to a trade-off of lost sales due to shortened average meal duration. Using the estimates in the 3SLS models, we calculate that increasing hourly workload from 4.3 to 5.76 may reduce average meal duration by 2.5%, which may incur a marginal loss of 6% of total sales. Note that the marginal loss of sales via shortened meal duration is about one seventh of the marginal gain of sales via direct sales effort, which emphasizes the importance of effective sales training. On aggregate, the total effect of optimal staffing may lift sales by about 35%, controlling for everything else. In other words,

$$35\% \text{ Sales lift} = 12\% \text{ Upselling} + 29\% \text{ Cross selling} - 6\% \text{ Loss from reduced meal duration.}$$

Additionally, we find that over 75% of the time, restaurants tend to over-staff by on average 1.14 servers per hour. Assuming the waiter is paid the minimum hourly wage for tipped workers ($2.63 in Massachusetts), we expect that optimal staffing may save the restaurants about 0.7% of total sales in labor. Moreover, reducing the staffing level by 1.14 servers each hour can save about 20% of current labor costs (the current average hourly staffing level is 5.63 servers). Of course, our model does not allow us to make an entirely accurate estimate of the potential improvement from optimal staffing (e.g., further labor-related non-wage costs), nor can the restaurants perfectly forecast demand. We nevertheless anticipate a significant sales lift and cost saving from optimal staffing because of the benefits from correcting both under-staffing and over-staffing misses.

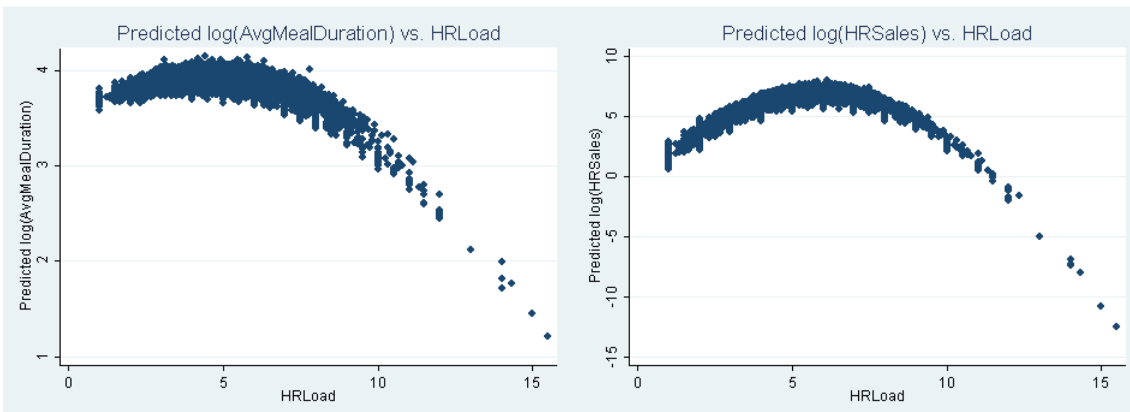
3.5.2 Concluding Remarks

Most studies on staffing decisions in services tend to overlook employees' adaptive behavior to work environments. In this paper we use detailed operational data gathered from a restaurant chain to study the effects of workload, an environmental variable, on servers' performance, taking endogeneity into consideration. We find that, as workload increases, the meal duration first increases and then decreases. We also find that, when the overall workload is low, increasing the workload may motivate servers to generate more sales. When the workload is high, increasing the workload may limit servers' effective sales. Our empirical findings contribute to the existing analytical models on staffing in two aspects. First, the non-linearity of the meal duration impact enriches the analytical papers on staffing that considers workload-dependent productivity. Hasija et al. (2010) have written an important and timely paper on the linearly speeding-up behavior induced by workload to estimate a call center's capacity. Future research in this stream may further assume non-linear productivity induced by workload. In their paper, less adequate is their assumption that the workload does not affect service quality. Our finding further provides empirical evidence to existing research studying the effect of workload on service quality (see e.g., Anand et al. 2011). Higher sales not only benefit the restaurant's bottom line but also may arguably reflect higher service quality. Understanding the trade-off of productivity and quality induced by workload may strengthen the analytical models on staffing.

The drivers of workload effects are initially unclear. On the one hand, a high workload may indicate high demand, which will increase the hourly performance. On the other hand, a high workload may indicate under-staffing, which may result in overloaded servers and diminished performance. We show that optimal staffing decisions, i.e., supply factors, mainly drive the results of our analysis. In particular, optimal staffing can improve sales generation and save labor costs.

Figure 3.5.1 illustrates the relationship between workload and predicted hourly performance. On the left we observe that workload initially slows down servers. After a critical point, servers start to speed up and thus reduce the average meal duration. As shown on the right of Figure 3.5.1, surprisingly, when the workload is small, sales increase with the workload: "The Devil makes work for idle hands". However, after a certain threshold (between five and six diners per server) is reached, sales start to decrease with the workload: "Many hands make light work". We make a seemingly counter-intuitive suggestion that reducing staffing is sometimes necessary to achieve higher sales and lower costs in a situation when restaurants are overstaffed. This is contrary to what most research into retail stores finds, which is significant under-staffing.

Figure 3.5.1: Predicted Hourly Performance



It is important to take into account the limitations of our findings. Although our data set is among the largest in the existing literature on worker performance response to external factors, it misses a few interesting variables. For example, we do not observe exact duration of each service procedure, such as taking the order and settling the bill . An interesting avenue for future research would be to examine the impact of workload on each specific service procedure. In addition, we lacked data about complete tipping information because we only observed tips paid through credit cards. We analyzed tips data that was available to us and found that tips showed very little variation; therefore we did not find robust impact of workload on tips. However, other types of customer satisfaction data, such as customer surveys, would be desirable to study the impact of workload on guest satisfaction. Furthermore, due to data limitations, our study does not examine the impact of other factors, such as kitchen capacity and diner heterogeneity. Although we employed instrumental variables to address this omitted variable issue, these factors would be worthwhile studying in future research. Additionally, our data only shows the number of servers who receive checks, which should cause a downward bias relative to the actual staffing decisions. Nevertheless, we find that the restaurant is already overstaffed so including more precise information in this case would only strengthen our findings. Further research opportunities in this setting include studying other OM/Human Resources interface issues, such as the “chemistry” among team members and team composition. Using our findings about server’s adaptive behavior to environmental constrains to design new workforce scheduling algorithm would offer an interesting and fruitful direction, too. Finally, in our models, in order to separate the supply-side driver of workload effect, we assume exogenous demand, namely the number of diners starting service every hour. In practice, arriving diners may choose to enter the restaurants or leave, depending on the occupancy of the restaurant. For example, when the restaurant is too empty, diners may interpret it as a sign of low restaurant quality, thus deciding to leave. However, when the restaurant is too full, diners may anticipate a long wait, thus balking at the door. It will be very interesting to empirically test how occupancy affects demand.

Bibliography

Afifi, A.A., V. Clark, S. May. 2004. *Computer-aided multivariate analysis*. CRC Press.

Amenta, E., B.G. Carruthers, Y. Zylan. 1992. A hero for the aged? The townsend movement, the political mediation model, and us old-age policy, 1934-1950. *The American Journal of Sociology* **98**(2) 308–339.

Anand, K.S., M.F. Paç, S. Veeraraghavan. 2011. Quality-speed conundrum: Tradeoffs in customer-intensive services. *Management Science* **57**(1) 40–56.

Anderson, C. 2004. The long tail. *Wired Magazine* **12**(10) 170–177.

Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Angrist, J., A.B. Krueger. 1994. Why do world war II veterans earn more than nonveterans? *Journal of Labor Economics* **12**(1) 74–97.

Baron, R.M., D.A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* **51**(6) 1173.

Bendoly, E., K. Donohue, K.L. Schultz. 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management* **24**(6) 737–752.

Bendoly, E., M. Prietula. 2008. The role of evolving skill and transitional workload on motivation and realized performance in operational tasks. *International Journal of Operations & Production Management* **28**(12) 1130–1152.

Bertini, M., L. Wathieu, S.S. Iyengar. 2011. The discriminating consumer: Product proliferation and willingness to pay for quality. *Journal of Marketing Research* **10**(0028) R1.

Bloom, Nicholas, John Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* **122**(4) 1351–1408.

Blossfeld, H.P., K. Golsch, G. Rohwer. 2007. *Event history analysis with Stata*. Lawrence Erlbaum.

Bockstedt, Jesse, Kim Huat Goh. 2008. Unbundling information goods: An empirical analysis of consumer created custom cds. INFORMS Annual Conference Presentation.

Boudreau, J.W. 2004. Organizational behavior, strategy, performance, and design in Management Science. *Management Science* 1463–1476.

Boudreau, J.W., W. Hopp, J.O. McClain, L.J. Thomas. 2003. On the interface between operations and human resources management. *Manufacturing & Service Operations Management* **5**(3) 179–202.

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao. 2005. Statistical analysis of a telephone call center. *Journal of the American Statistical Association* **100**(469) 36–50.

Brynjolfsson, E., Y. Hu, M. D. Smith. 2006. From niches to riches: The anatomy of the long tail. *Sloan Management Review* **47**(4) 67–71.

Brynjolfsson, E., Y. J. Hu, M. D. Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* **49**(11) 1580–1596.

Brynjolfsson, E., Y.J. Hu, M.D. Smith. 2010. Research commentary—long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research* **21**(4) 736–747.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, Duncan Simester. 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* **57**(8) 1373–1386.

Buchinsky, M. 1994. Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica* **62**(2) 405–458.

Cachon, G. P., C. Terwiesch, Y. Xu. 2008. On the effects of consumer search and firm entry in a multiproduct competitive market. *Marketing Science* **27**(3) 461–473.

Cakir, A., DJ Hart, TFM Stewart. 1980. *Visual display terminals: A manual covering ergonomics, workplace design, health and safety, task organization*. Wiley (Chichester Eng. and New York).

Caro, F., J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science* **53**(2) 276.

Chatfield, Christopher, Gerald Goodhardt. 1975. Results concerning brand choice. *Journal of Marketing Research* **12**(1) 110–113.

Chen, Pei-Yu, Shin-yi Wu, Jungsun Yoon. 2004. The impact of online recommndations and consumer feedback on sales. International Conference on Information Systems (ICIS).

Clemons, E. K., G. G. Gao, L. M. Hitt. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems* **23**(2) 149–171.

Deci, E.L., J.P. Connell, R.M. Ryan. 1989. Self-determination in a work organization. *Journal of Applied Psychology* **74**(4) 580.

Dellarocas, C., C.A. Wood. 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* **54**(10) 460–476.

Dellarocas, Chrysanthos, Ritu Narayan. 2007. Tall heads vs. long tails: Do consumer reviews increase the informational inequality between hit and niche products? Working Paper.

Elberse, A., F. Oberholzer-Gee. 2008. Superstars and underdogs: An examination of the long tail phenomenon in video sales. HBS Working Paper.

Eliashberg, J., A. Elberse, M. Leenders. 2006. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science* **25**(6) 638–661.

Fields, Roger. 2007. *Restaurant Success by the Numbers*. Ten Speed Press.

Fisher, Marshall, Janice Hammond, Walter Obermeyer, Ananth Raman. 1994. Making supply meet demand in an uncertain world. *Harvard Business Review* (May-June) 83–93.

Fitzsimmons, J.A., G.B. Maurer. 1991. A walk-through audit to improve restaurant performance. *The Cornell Hotel and Restaurant Administration Quarterly* **31**(4) 94–99.

Fleder, D., K. Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science* **55**(5) 697–712.

Fleder, D., K. Hosanagar, A. Buja. 2009. Recommender systems and their effects on consumers: the fragmentation debate. Working Papers.

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and service operations management* **5**(2) 79–141.

Ghose, Anindya, Bin Gu. 2006. Search costs, demand structure and long tail in electronic markets: Theory and evidence. Working Paper.

Gomes, L. 2006. It may be a long time before the long tail is wagging the web. *The Wall Street Journal* July 26th.

Goodhardt, G. J., A. S. C. Ehrenberg, C. Chatfield. 1984. The dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)* **147**(5) 621–655.

Green, Linda, Sergei Savin, Nicos Savva. 2011. "Nursevendor Problem": Personnel staffing in the presence of endogenous absenteeism. *Working Paper* .

Hasija, S., E. Pinker, R.A. Shumsky. 2010. Work expands to fill the time available: Capacity estimation and staffing under parkinson's law. *Manufacturing & Service Operations Management* **12**(1) 1–18.

Hayashi, F. 2000. Econometrics. 2000. *Princeton University Press. Section* **1** 60–69.

Hervas-Drane, Andres. 2009. Word of mouth and taste matching: A theory of the long tail. Working Paper.

Hickok, Allan, Lana Lazarus. 2003. Restaurant industry review. *U.S. Bancorp Piper Jaffray Equity Research* 9.

Hinz, O., J. Eckert, B. Skiera. 2011. Drivers of the long tail phenomenon: An empirical analysis. *Journal of Management Information Systems* **27**(4) 43–70.

Hu, N., P.A. Pavlou, J. Zhang. 2007. Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication. Working Paper.

Huckman, R.S., B.R. Staats, D.M. Upton. 2009. Team familiarity, role experience, and performance: Evidence from indian software services. *Management Science* **55**(1) 85–100.

Jiang, Baojun, Kinshuk Jerath, Kannan Srinivasan. 2011. Firm strategies in the Mid Tail? of platform-based retailing. *Marketing Science* **30**(5) 757–775.

Kc, D.S., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

Kc, D.S., C. Terwiesch. 2011. An econometric analysis of patient flows in the cardiac ICU. *Manufacturing & Service Operations Management* .

Kennedy, P. 2003. *A guide to econometrics*. The MIT Press.

Kimes, S.E. 2004. Restaurant revenue management implementation at chevys arrowhead. *Cornell Hotel and Restaurant Administration Quarterly* **45**(1) 52.

Kimes, S.E., D.I. Barrash, J.E. Alexander. 1999. Developing a restaurant revenue-management strategy. *Cornell Hotel and Restaurant Administration Quarterly* **40**(5) 18.

Kimes, S.E., R.B. Chase, S. Choi, P.Y. Lee, E.N. Ngonzi. 1998. Restaurant revenue management:: Applying yield management to the restaurant industry. *The Cornell Hotel and Restaurant Administration Quarterly* **39**(3) 32–39.

Kimes, S.E., S.K.A. Robson. 2004. The impact of restaurant table characteristics on meal duration and spending. *Cornell Hotel and Restaurant Administration Quarterly* **45**(4) 333.

Kimes, S.E., G.M. Thompson. 2004. Restaurant revenue management at Chevys: determining the best table mix. *Decision Sciences* **35**(3) 371–392.

Kleinbaum, D.G., L.L. Kupper, K.E. Muller. 2007. *Applied regression analysis and other multivariable methods*. Duxbury Pr.

Koenker, R., G. Bassett Jr. 1978. Regression quantiles. *Econometrica* 33–50.

Kuksov, D., J.M. Villas-Boas. 2010. When more alternatives lead to less choice. *Marketing Science* **29**(3) 507–524.

Kumar, A., M.D. Smith, R. Telang. 2011. The broadcast window effect. Working Paper.

Kuntz, Ludwig, Roman Mennicken, Stefan Scholtes. 2011. Stress on the ward: Linking individual performance to system service quality. *Working Paper* .

Lambert, P.J., J.R. Aronson. 1993. Inequality decomposition analysis and the gini coefficient revisited. *The Economic Journal* 1221–1227.

Locke, E.A., A.J. Mento, B.L. Katcher. 1978. The interaction of ability and motivation in performance: An exploration of the meaning of moderators1. *Personnel Psychology* **31**(2) 269–280.

Lupien, SJ, F. Maheu, M. Tu, A. Fiocco, TE Schramek. 2007. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and cognition* **65**(3) 209–237.

Maher, K. 2007. Wal-mart seeks flexibility in worker shifts. *The Wall Street Journal* January 3rd.

Mani, V., S. Kesavan, J.M. Swaminathan. 2011. Understaffing in retail stores: Drivers and consequences .

McPhee, William N. 1963. *Formal Theories of Mass Behavior*. Glencoe, NY: Free Press.

Mill, R. 2004. Restaurant management: Customers, operations, and employees .

Netessine, Serguei, Serguei Savin, Wenqiang Xiao. 2006. Revenue management through dynamic cross selling in e-commerce retailing. *Operations Research* **54**(5) 893–913.

Oliva, R., J.D. Sterman. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* 894–914.

Orlowski, Andrew. 2008. Chopping the long tail down to size. *The Register* 7 Nov.

Park, Yoon-Joo, Alexander Tuzhilin. 2008. The long tail of recommender systems and how to leverage it. ACM Conference on Recommender Systems.

Parkinson, C. 1958. Northcote: Parkinsons law: The pursuit of progress.

Perdikaki, O., S. Kesavan, J.M. Swaminathan. 2011. Effect of traffic on sales and conversion rates of retail stores .

Powell, Adam, Sergei Savin, Nicos Savva. 2012. Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management* (Forthcoming).

Powell, S.G., K.L. Schultz. 2004. Throughput in serial lines with state-dependent behavior. *Management Science* 1095–1105.

Preacher, K.J., A.F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods* **40**(3) 879.

Randall, Taylor, Karl Ulrich. 2001. Product variety, supply chain structure, and firm performance: An empirical examination of the U.S. bicycle industry. *Management Science* **47**(12) 1588–1604.

Robson, S.K.A. 1999. Turning the tables. *Cornell Hotel and Restaurant Administration Quarterly* **40**(3) 56.

Rosen, Sherwin. 1981. The economics of superstars. *The American Economic Review* **71**(5) 845–858.

Rothschild, M. 1974. Searching for the lowest price when the distribution of prices is unknown. *The Journal of Political Economy* **82**(4) 689–711.

Schultz, K.L., D.C. Juran, J.W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Science* 1664–1678.

Schultz, K.L., D.C. Juran, J.W. Boudreau, J.O. McClain, L.J. Thomas. 1998. Modeling and worker motivation in jit production systems. *Management Science* 1595–1607.

Setyawati, L. 1995. Relation between feelings of fatigue, reaction time and work productivity. *Journal of human ergology* **24**(1) 129.

Shih, Willy, Stephen Kaufman, David Spinola. 2007. Netflix. *Harvard Business School Case* (November) 9–607–138.

Siebert, W.S., N. Zubanov. 2010. Management economics in a large retail company. *Management Science* **56**(8) 1398–1414.

Silver, E.A., D.F. Pyke, R. Peterson, et al. 1998. *Inventory management and production planning and scheduling*. Wiley New York.

Simonson, I., A. Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* **29**(3) 281–295.

Staats, B.R., F. Gino. 2012. Specialization and variety in repetitive tasks: Evidence from a japanese bank. *Management Science* (Forthcoming).

Staiger, D., J.H. Stock. 1994. Instrumental variables regression with weak instruments .

Stigler, G.J. 1961. The economics of information. *The Journal of Political Economy* **69**(3) 213–225.

Sulek, J.M., R.L. Hensley. 2004. The relative importance of food, atmosphere, and fairness of wait. *Cornell Hotel and Restaurant Administration Quarterly* **45**(3) 235.

Tucker, C., J. Zhang. 2011. How does popularity information affect choices? a field experiment. *Management Science, forthcoming* .

Varian, Hal. 2006. Why old media and Tom Cruise should worry about cheaper technology. *The New York Times* October 19th.

Villas-Boas, J.M., R.S. Winer. 1999. Endogeneity in brand choice models. *Management Science* 1324–1338.

WIPO. 2003. National studies on assessing the economic contribution of the copyright-based industrie. *World Intellectual Property Organization Reports* .

Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT press.

Yeo, G.B., A. Neal. 2004. A multilevel analysis of effort, practice, and performance: Effects; of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology* **89**(2) 231.

Yitzhaki, S. 1979. Relative deprivation and the gini coefficient. *The Quarterly Journal of Economics* 321–324.

Zellner, A., H. Theil. 1962. Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica: Journal of the Econometric Society* 54–78.

Zenios, Stefanos, Biyu He, Franklin Dexter, Alex Macario. 2011. The timing of staffing decisions in hospital operating rooms: Incorporating workload heterogeneity into the newsvendor problem. *Manufacturing & Service Operations Management* (Forthcoming).