EMPIRICAL STUDY OF LINK BETWEEN OPERATIONS AND FINANCIAL PERFORMANCE FOR RETAILERS

Vidya Mani

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Kenan-Flagler Business School (Operations, Technology, and Innovation Management).

Chapel Hill
2011

Approved by:

Dr. Jayashankar M. Swaminathan, Co-Chair
Dr. Saravanan Kesavan, Co-Chair
Dr. Tarun L. Kushwaha, Committee Member
Dr. Ann Marucheck, Committee Member
Dr. Bradley Staats, Committee Member

ABSTRACT

VIDYA MANI: Empirical Study of Link between Operations and Financial Performance for Retailers
(Under the direction of Dr. Jayashankar M. Swaminathan and
Dr. Saravanan Kesavan)

Retailers continually try to improve their store operations in order to achieve better financial performance. However, there appears to be limited empirical research that shows the influence of operations management on financial performance. We conduct an empirical study of the link between operations management and financial performance of retailers by investigating at drivers of store level operations in a single retail chain, and studying the relative firm level performance of US public retailers. We utilize data from two sources; individual proprietary store level traffic data and publicly available financial data for this study. In addition, we complement our datasets by extracting information on demographics from publicly available databases. In the first chapter, we use detailed traffic data to study whether there is understaffing at a heterogeneous group in retail stores belonging to the same retail chain. We then look at some of the underlying causes for this understaffing, including traffic forecast errors and scheduling constraints, and quantify their impact on store profits. In the second chapter, we characterize the underlying distribution of hourly traffic data that is obtained with help of traffic counters at each of the retail stores and study the impact that competition and location demographics have on the observed variability in traffic. We then explore the managerial implications of having detailed traffic information on labor planning by deriving better forecasts of traffic that would aid staffing decisions. Finally, in the third chapter, we conduct a firm level analysis of US public retailers with help of benchmarking metrics developed from operations management. We demonstrate an inverted-U relationship between abnormal inventory growth and one-year ahead earnings. We also show that equity analysts are systematically biased in their earnings forecasts as they fail to incorporate information contained in abnormal inventory growth and further, an investment strategy based on abnormal inventory growth can yield significant abnormal returns.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the constant inspiration, support and encouragement from the faculty members at University of North Carolina, Chapel Hill, fellow colleagues, members of my family and close friends.

I would like to begin by thanking my co-advisors, Dr. Jayashankar M. Swaminathan and Dr. Saravanan Kesavan for their continuous guidance during this process. Dr. Swaminathan has been an enduring source of inspiration all through my doctoral program. His guidance has been instrumental in helping me develop the ability to identify and critically analyze the different facets to each research problem. As an advisor, a teacher and a mentor, he has shown me, by example, the perseverance that is required to be a good researcher. Through various interactions and project engagements, I have learnt to appreciate that there is as much value in the journey itself, as in the ultimate goal that is to be reached at the end of each project.

Under the guidance of Dr. Saravanan Kesavan, I have had the advantage of not only learning the intricate and advanced tools and techniques required in the field of empirical research, but also to conduct research that would bring theoretical insights to practical applications. He has instilled in me a strong sense of the rigor and discipline that is required from academicians. I would also like to express my gratitude to our department chair and member of my dissertation committee, Dr. Ann Marucheck, for providing me extensive professional guidance throughout the dissertation process. Her encouragement for the PhD program and support for various conferences have given me the advantage to constantly keep in touch with recent research advancements, and interact with other academicians and practitioners in the field of retail operations. I would also like to thank my dissertation committee members, Dr. Bradley Staats and Dr. Tarun Kushwaha, for their advice, suggestions, and support throughout my dissertation.

My appreciation also goes to Dr. Wendell Gilland, with whom I have had the opportunity of working on a research project and who has offered kind encouragements throughout my study; to Dr. Harvey Wagner who has encouraged me to think deeply on the practical relevance of different research approaches, and to other faculty members with whom I have had the opportunity to interact on varied occasions and learn about unique research methods in the field of operations management. I would also like to extend my gratitude to my senior colleagues - Almula Camdereli, Sriram Narayanan, Adelina Gnanlet, Muge Yayla-Kullu and Olga Perdikaki who were always available to help me navigate through the different stages of the program, and my close friends Gokce Esenduran and Yen-Ting Ling who were unwavering in their support and were always present to bolster my confidence during tough times. I would also like to wish the upcoming PhD students - Aaron Ratcliffe, Karthik Natarajan, Adem Orsdemir, Gang Wang, Laura Fink, Hsing-Ping Kuo and Zhe Wang the very best in successfully completing the PhD program and very much look forward to hear about their research topics. I would like to extend my thanks to my friends and fellow classmates Elizabeth Nielson, Kaicheng Zhang and Paul Rowe, for their sincere friendship and support. I am also indebted to the firm from whom we obtained the data for the empirical analysis in my dissertation. Last, but not the least, I would like to say a word of appreciation to the efforts put in by Sharon Parks, Kim Scales, Erin Rimmer, and Holly Guthrie in making sure that there was always adequate administrative support during my time here.

Meeting the unique challenges of a doctoral program would not have been possible without the steadfast love and guidance of my parents, Dr. Uliyar V. Mani and Dr. Indirani Mani. Their unequivocal belief in my potential to succeed in the doctoral program, and confidence in my capabilities has helped me remain resolute in the pursuit of my dreams. Through their own

hard work and accomplishments they have been my role models throughout my professional career. Finally, I would like to extend my deepest gratitude to Dr. Sunil Guptan, for his unstinted support, his unfailing commitment in ensuring that I continue to aim for a higher goal than I thought possible to achieve, and whose steady encouragement has propelled me to constantly look for ways to learn and improve myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Retailers today, face myriad challenges in sustaining adequate profitability levels. Intense competition and declining margins have forced many retailers to critically examine and redesign their operations in an effort to improve their performance. The volatile market dynamics no longer support traditional growth models of rolling out more stores and adding more SKUs to maintain the return on investment. The focus instead has shifted to strategies that would enable retailers to retain their existing customers and earning a bigger share of the customer's wallet, while continuing to operate in a cost efficient manner[1]. This is not possible without good operations management. Operational decisions taken at both the store and at the firm level, that enable the retailers to maintain a high level of customer service and retain customers, together with managing the cost of operations, have been found to be a key driver in driving profitability.

Since achieving a high level of productivity and profitability through good operations management is a top concern for many retailers, a research study into the different factors that would aid them in achieving these objectives at the individual store level, and provide a metric for comparing performance with peers at the firm level would provide valuable insights. In this empirical study, we take one step in this direction by analyzing and modeling traffic flow, developing a framework to determine optimal staffing levels and demonstrate the consequences of understaffing on store profitability. Towards the end, we extend this link between good

---

[1] The Changing Nature of Retail 2006. Deloitte Consulting LLP

operations management and healthy financial performance through a broader study of inventory growth and earnings for different retailers in the industry. Thus, moving from the retail store front to the strategic link between operations and financial indicators at firm level, our empirical study aims to connect the different aspects of store and firm level operations and financial performance.

In the following two chapters, we address two key challenges in the context of retail store operations – first, determining the extent of understaffing in retail stores and how an optimal staffing plan that takes into account the individual store characteristics can lead to better store performance, and second, characterizing the traffic or demand distribution at these stores that form the basis for these staffing plans. A pressing concern today, for most retailers, is to find ways to effectively manage the climbing workload to satisfy increased customer service demands with lower budgets[2]. Store managers are increasingly turning to sophisticated technology and software packages that would help them in this process. In this context, there has been significant interest in leveraging customer data to make operational decisions like labor planning and forecasting traffic. These are critical to store performance as staffing decisions have a direct impact on customers' in-store experience, and in many cases, play a deciding factor in customer's eventual purchase decision.

Finally, in the third chapter, we shift our focus to firm level performance and look at the link between operations and financial performance across a cross section of US retailers. We compare the relative firm level performance of these retailers based on benchmarking metrics obtained from operations management. In particular, we examine the relationship between inventory levels and one-year ahead earnings of retailers using publicly available financial data and demonstrate the economic significance of this relationship by investigating if an investment strategy based on these metrics generates significant returns.

A brief outline of the main focus in each of these chapters is given below.

---

[2] The state of the store manager. 2010. Chain Store Age

1.1 Causes and Consequences of Understaffing in Retail Stores

In the first chapter, we conduct an econometric study of labor planning decisions and explore the problem of understaffing in retail stores. This is a critical area of store operations as all too often, retailers might end up spending millions of dollars in promotional activities that drive customers to stores, only to lose them due to the inadequate level of sales assistance provided within the store. Many consumer reports and shopper satisfaction surveys consistently find one of the chief shopping annoyances to be the difficulty in finding a useful sales person in helping them with their purchase decision[3].

It is crucial for retailers to determine the right amount of labor to have in stores as it impacts sales directly by affecting the level of sales assistance provided to shoppers, and indirectly, through execution of store operational activities such as stocking shelves, tagging merchandise, and maintaining the overall store ambience (Fisher and Raman, 2010). On the other hand, store labor expenses account for a significant portion of a store's operating expense (Ton, 2009). Hence, to maximize profits retailers have to walk a fine line between balancing the costs and benefits of store labor.

In recent years, retailers have invested heavily in technologies like traffic counters and work force management tools to aid store managers in labor planning, conducting training programs for their store managers, and providing incentives for the store managers to have the right amount of labor in the stores. However, it is unclear to what extent the retailers are successful in their efforts. While substantial agreement exists that understaffing would result in lower store performance, the extent of understaffing in retail stores, and its impact on store profitability, has not been studied rigorously.

In this chapter we use data collected from 41 stores of a large specialty apparel retailer to investigate if there is understaffing and quantify the impact of understaffing on store profitability.

[3] Where to shop: August 2010. Consumer Reports Magazine.

We use hourly data on store labor, store traffic, transactions, and sales collected over 365 days to estimate the contribution of labor to sales and expenses for each store. Since these contributions could vary by store and time, our estimation is performed for each individual store and for different time periods to allow for heterogeneity across stores and time. We use the Generalized Method of Moments (GMM) approach to estimate our structural equations model, as it is a semi-parametric technique that produces consistent estimates without making any distributional assumptions. Using a given store's estimates of contribution of labor to sales and cost of labor; we construct the optimal labor plan for the store and study deviations of the actual labor from the optimal plan to check for understaffing.

We find that the stores differ widely in the contribution of labor to sales and their imputed cost of labor. For example, the average hourly imputed cost of labor in our study was found to be $30.47, with a range from $10.50 to $54.92. Furthermore, this cost is significantly higher than the average hourly wage rate of $10.05 for retail salespersons, which can be explained partly by systematic factors based on individual store and local market area characteristics. Second, we find that on average although the stores appear to have the required amount of labor relative to the optimal labor plan at the daily level, there is significant and consistent understaffing during peak hours in most stores (and overstaffing at other times). Third, we show how forecasting errors and scheduling constraints could cause the observed understaffing, and demonstrate that the negative impact due to forecasting errors are exacerbated when there is very little schedule flexibility.

Our results provide one possible explanation for the recent moves by many retailers like Wal-Mart and Payless ShoeSource towards more flexible work schedules (Maher, 2007). We also show that it is important to consider the heterogeneity amongst the different stores, even within the same retail chain, when making staffing decisions that would in-turn impact the resultant service level within the stores.

1.2 Improving Store Operations through Better Traffic Forecasts

In the second chapter, we study how utilizing information on hourly store traffic data can improve staffing decisions with help of better store traffic forecasts. Traffic forecasting is a critical activity for retailers as it drives both stocking and labor planning decisions in the store. Despite significant investments in forecasting technologies (e.g. installation of traffic counters at different stores) and a long line of research in operations management that has looked at improvement in forecasts through use of more recent information on customer demand, in practice, there exists a significant gap in practice between capturing traffic data and leveraging it in the planning process.

In the context of store operations, there has been almost a double digit growth in adoption of workforce management solutions that incorporate customer demand information based on point-of-sale data or traffic counters in generating forecasts of future traffic and create staffing plans, using some underlying algorithms based on these initial traffic forecasts. A key assumption driving many of these algorithms is the distribution of traffic. Hence, it would be useful to characterize the distribution of traffic, how it may differ from some of the common assumptions that are used to drive these algorithms, and the usefulness of this information to store managers in making their labor planning decisions.

We have two main objectives in this chapter. First, we characterize the distribution of traffic based on detailed traffic data obtained from traffic counters from 60 stores of a women's specialty apparel retail chain. Next, we explore the usefulness of this information to retailers in terms of improving their ability to plan and schedule employees and study the impact on store performance.

Towards this objective, we first construct and estimate the parameters of multiple statistical models, like the negative binomial model, the poisson model and the normal model for the store traffic data. We find that the rate of traffic varies considerable across different times of the day, the variance in traffic is considerably higher relative to the mean level of traffic during peak hours and there exist both inter- and intra-day correlations in store traffic. Our results show that a negative binomial distribution, that captures many of these characteristics, provides a better fit with the observed data, as opposed to a Poisson or normal distribution, and that the level of competition is negatively associated with the observed variation in traffic. We find that the forecasts based on a negative binomial model significantly outperform forecasts from other models due to its ability to produce more accurate prediction intervals than other models. Finally, we show that as requirements for service availability increase, the labor forecasts from negative binomial model perform significantly better than those from Poisson and normal models as well as from the time-series forecasts.

Thus, our results show that using the right distribution of traffic would allow retailers to generate staffing plans that would more closely meet their desired service level during different time periods of the day and prevent any systemic understaffing during peak hours. This result is of practical relevance as there is an increasing trend towards integrating demand information from traffic counters with workforce management solutions to plan labor based on traffic (Store, 2010).

1.3 The relationship between abnormal inventory growth and future earnings for U.S public retailers

In the third chapter, we move from studying how operational decisions impact store level performance to exploring the link between operations and financial performance at the firm level. Here we conduct an empirical analysis of the relationship between firms' inventory levels and their one-year ahead earnings.

Earnings- per-share (EPS) is considered as one of the important indicators of financial performance for firms as it is a summary measure of firm profitability and a closely watched metric by many equity analysts and investors. Forecasts based on the reported firm earnings indicate the prospects for future growth and profitability and form a key input to investment decisions. However, current evidence on the relationship between inventory and one-year ahead earnings, at the firm level, for retailers is weak. For example, in the accounting literature, Abarbanell and Bushee (1997) do not find evidence of this relationship for retailers while Bernard and Noel (1991) find that inventory levels predict earnings, wherein they assume a linear relationship between inventory and earnings.

Since earnings are a measure of profitability of the firm, based on insights from operations management, one might expect a negative impact on profits when a firm has too high or too low inventory growth, as compared to optimal inventory growth, i.e. one expects an inverted-U relationship between inventory growth and earnings at the firm level as well.

There are several challenges in testing the relationship between inventory and earnings at the firm-level. First, raw inventory levels cannot be used to determine the relationship since it is correlated with number of stores, sales etc. For example, inventory for a retailer could have grown either due to presence of stale inventory or as a result of opening new stores. While the former would be associated with lower earnings in the future, the latter would not. Second, service level information of retailers is not publicly available. So, it is difficult to figure out whether a retailer's inventory level is high because it is carrying excess inventory or if it is providing a high service level (Lai 2006). The former would be a negative signal of future earnings but the latter would not.

In this chapter, we use the expectation model from Kesavan et al (2010) to obtain the expected inventory growth, calculate abnormal inventory growth as the deviation of actual inventory growth from expected inventory growth, and use it as the benchmarking metric to investigate the relationship between inventory and one-year ahead earnings. We investigate the economic significance of the information content in abnormal inventory growth by examining if equity analysts' earnings forecasts incorporate information contained in abnormal inventory growth and test if an investment strategy based on abnormal inventory growth would yield significant abnormal returns.

We use quarterly and annual financial data for the fiscal years 1993-2009, along with data on comparable store sales, total number of stores and earnings per share for a large cross-section of U.S. retailers listed on NYSE, AMEX, or NASDAQ from Standard & Poor's Compustat database for our analysis. Equity analysts' earnings forecasts are collected from Institutional Brokers Estimates System (I/B/E/S). Stock returns inclusive of dividends are obtained from CRSP. These are supplemented with hand-collected data from financial statements.

We find that there exists an inverted-U relationship between abnormal inventory growth and one-year ahead earnings. These results are robust to the metric used to measure abnormal inventory growth. We also show that equity analysts do not fully incorporate the information contained in past inventory resulting in systematic bias in their earnings forecasts; this bias is predicted by previous year's abnormal inventory growth. Finally, we demonstrate that an investment strategy based on abnormal inventory growth yields significant abnormal returns. Thus, we show that benchmarking metrics possess information useful to predict earnings and serve as a basis for investment strategies.

CHAPTER 2

Causes and Consequences of Understaffing in Retail Stores

2.1 Introduction

In the battle to win retail customers, the importance of labor planning cannot be overemphasized. Having adequate store labor is critical as it impacts sales directly by affecting the level of sales assistance provided to shoppers, and indirectly, through execution of store operational activities such as stocking shelves, tagging merchandise, and maintaining the overall store ambience (Fisher and Raman, 2010).

Store labor affects store profitability not only through its impact on sales but also on expenses. Labor-related expenses account for a significant portion of a store's operating expense (Ton, 2009). Hence, to maximize profits retailers have to walk a fine line between balancing the costs and benefits of store labor. They try to achieve this balance by investing in technologies such as traffic counters and work force management tools to aid store managers in labor planning, conducting training programs for their store managers, and providing incentives for the store managers to have the right amount of labor in the stores. However, it is unclear to what extent the retailers are successful in their efforts. Anecdotal evidence suggests that about 33% of the customers entering a store leave without buying because they were unable to find a salesperson to help them[1]. Such statistics suggest that understaffing can be particularly vexing for retailers since they often spend millions of dollars in marketing activities to draw customers to their stores.

---

[1] Baker Retail Initiative, May 2007.

While substantial agreement exists that understaffing would result in lower store performance, the extent of understaffing in retail stores has not been studied rigorously.

This issue is important for several reasons. First, studies have shown that understaffing could lead to poor service quality that can result in lower customer satisfaction (Loveman 1998; Zeithaml 2000). Such customer dissatisfaction could lead to customer complaints that are expressed in many forums, including social networking websites such as Facebook and Twitter, causing retailers to worry about the word-of-mouth effect (Park et al. 2010; Zeithaml et al. 1996). In a survey of shoppers in the specialty apparel retail segment, shoppers highlighted service-related attributes as being among the top factors that drive them back to stores[2]. Dissatisfied customers may switch to competitors resulting in a loss of lifetime value for those customers (Heskett et al. 1994; Jain and Singh, 2002). Second, understaffing issues have been found to be associated negatively with store associate satisfaction which in turn can negatively impact customer in-store experiences leading to customer dissatisfaction and ultimately lower store financial performance (Maxham et al. 2008; Oliva and Sterman, 2001). Hence, it is important to examine whether understaffing exists in retail stores, and if so, determine the causes and consequences of this understaffing.

In this chapter we study the following research questions: 1) Are retail stores systematically understaffed?, 2) If yes, what are the drivers of this understaffing, for example, how do errors in forecasts and scheduling constraints contribute to this observed understaffing, and 3) what is the impact of this understaffing on store performance. We use data collected from 41 stores of a large specialty apparel retailer to investigate if there is understaffing and quantify the impact of understaffing on store profitability. We use hourly data on store labor, store traffic, transactions, and sales collected over 365 days to estimate the contribution of labor to sales and expenses for each store. Since these contributions could vary by store and time, our estimation is

---

[2] Booz & Company. 2008. Winning in retail with a targeted service model.

performed for each individual store and for different time periods to allow for heterogeneity across stores and time. We use the Generalized Method of Moments (GMM) approach to estimate our structural equations model, as it is a semi-parametric technique that produces consistent estimates without making any distributional assumptions. Using a given store's estimates of contribution of labor to sales and cost of labor; we construct the optimal labor plan for the store and study deviations of the actual labor from the optimal plan to check for understaffing. A store is said to be understaffed in a given time period when the actual labor is less than the optimal labor for that time period. Finally, we investigate causes of understaffing, if any, in retail stores and the consequences of understaffing on store profitability.

We have the following results in our study. First, we find that the stores differ widely in the contribution of labor to sales and their imputed cost of labor. For example, the average hourly imputed cost of labor in our study was found to be $30.47, with a range from $10.50 to $54.92. Furthermore, this cost is significantly higher than the average hourly wage rate of $10.05 for retail salespersons, which can be explained partly by systematic factors based on individual store and local market area characteristics. Second, we find that on average, the stores appear to have the required amount of labor relative to the optimal labor plan at the daily level. However, significant understaffing is observed during peak hours in most stores (and overstaffing at other times). Third, we identify forecast errors and scheduling constraints as the underlying causes of understaffing in these retail stores and quantify their relative impact on store profitability.

This chapter makes the following contributions to the growing research on labor planning in retail operations (e.g., Fisher et al. 2007; Netessine et al. 2010; Ton and Huckman 2008). We document the presence of understaffing during peak hours across multiple stores of a retail chain and quantify the impact of understaffing on store profitability. Our study is also the first to use structural estimation techniques in the context of labor planning. This approach enables us to impute the cost of labor for each store. Several studies in the operations management literature (Gino and Pisano, 2008; Schweitzer and Cachon, 2000) have advocated using intrinsic costs as opposed to accounting costs for decision making. Our approach of imputing the labor costs enables us to capture this intrinsic cost used by store managers in their labor planning decisions. In addition, our results show the significant heterogeneity in the imputed costs across stores, even within the same chain. This heterogeneity indicates that local characteristics play a key role in labor-planning decisions. Prior theoretical literature (Anand and Mendelson 1997; Chang and Harrington 2000) on centralized versus decentralized decision-making has posited that decentralized decision-making is more advantageous when local knowledge is important to balance trade-offs between cost and benefits of a decision. Our results suggest that workforce management tools that are increasingly being deployed in corporate offices should not ignore the heterogeneities in the imputed cost of labor across stores. Else, they could lead to misalignment between the recommendations of the centralized workforce management tool and what the store managers need. This could result in store managers spending considerable time overriding the decisions of the centralized planning tools as documented by van Donselaar et al. (2010) and Netessine et al. (2010).

This chapter is organized as follows. §2.2 reviews the background literature and §2.3 explains our research setup, and the data and variables used in the chapter. In §2.4 we outline the methodology and estimation procedure for imputing the parameters that are used to develop the optimal labor plan. We report our main results in §2.5, explore some of the drivers of differences in store managers' imputed labor costs and discuss their implications in §2.6, and finally present our conclusions in §2.7.

2.2 Literature Review

Labor planning is an integral part of retail store operations and critical to ensure successful retail store execution. Research in labor planning has gained significant interest in recent years. Using data from small appliances and furnishing retailer, Fisher et al. (2007) find that store associate availability (staffing level) and customer satisfaction are among the key variables explaining month-to-month sales variations. Netessine et al. (2010) find a strong cross-sectional association between labor practices at different stores and basket values for a supermarket retailer. The authors demonstrate a negative association between labor mismatches at the stores and basket value. Lu et al. (2011) use purchase history of supermarket customers to a deli-counter to study how waiting in queue affects customer purchasing behavior. With help of price and labor data, they are able to study the impact of different service levels on customer buying behavior and find significant heterogeneity in customer sensitivity to waiting, and that the degree of waiting sensitivity is negatively correlated with customer's sensitivity to price.

Several researchers have looked at the impact of labor decisions on profitability as well. Ton (2009) investigates how staffing level affects store profitability through its impact on conformance and service quality for a large specialty retailer. Using monthly data on payroll, sales and profit margins, she finds evidence of understaffing, and that increasing labor leads to higher store profits primarily through higher conformance quality. Borucki and Burke (1999) find that improved sales personnel service performance has a direct positive impact on store financial performance and suggests removing human resource obstacles like inadequate staffing during peak times as one of the managerial interventions that can help improve sales personnel service performance. Our study adds to this literature by studying if there exists understaffing during the different hours of the day through use of hourly labor, traffic and sales data. Our structural estimation approach allows us to quantify the improvement in store profitability by increasing labor during the hours when store is understaffed.

There are very few papers that have utilized detailed store traffic information in the study of labor planning decisions at retail stores. Exceptions to this are Lam et al. (1998) and Perdikaki et al. (2010). Perdikaki et al. (2010) study the role of traffic and labor on store performance and show that store traffic exhibits diminishing returns to scale with respect to store sales performance. In this chapter, we have a different objective, wherein we use the information on store traffic, sales and labor to study if the stores are understaffed and the consequence of this understaffing on store profitability. Our study is closest to Lam et al. (1998) who show how sales-force scheduling decisions can be made based on a forecast of store traffic and quantify the impact these decisions have on store profits. The authors conduct this analysis for a single store and thus do not consider the heterogeneity across stores in making these decisions or systematic factors that might explain these differences. Further, they elicit information about the compensation, bonus, insurance, and benefits for store labor from the store manager to measure the cost of labor and use it to compute the optimal labor for the store. This approach assumes that store manager's implicit cost of labor is the same as the accounting cost of labor as stated by the store manager.

We follow a more general approach of imputing the labor costs that the store manager uses in making their labor planning decisions. This approach is advantageous as several studies in decision making have shown that the managers' perceptions of costs can be very different from traditional cost calculations (Cooper and Kaplan, 1998; Thomadsen, 2005; Olivares et al. 2008) and that the managers tend to make decisions according to these intrinsic costs (Gino and Pisano, 2008; Schweitzer and Cachon, 2000). Also, when asked, even experts at times tend to underestimate or overestimate the actual costs that should be considered in decision making (Hogarth and Makridakis, 1981; Kahneman and Lovallo, 1993). While it might be possible to explicitly gather information on the cost of labor for a single store, it becomes considerably more challenging to do so for a large group of stores, especially where the store managers may differ in their emphasis on the different parameters that impact the cost of labor. Through our structural estimation techniques, even without having the data on cost of labor for each individual store, we are able to capture these intrinsic costs used by the store manager in labor planning. We show that these costs are heterogeneous among the different stores and that they could depend, in part, on local characteristics like competition, median household income, and availability of labor, factors that have not been considered in prior literature.

The use of structural estimation techniques to impute the underlying costs considered by managers in decision-making has only recently been adopted in operations management literature. This approach to estimate cost parameters from observed decisions in operations management has been utilized by Cohen et al. (2003), Olivares et al. (2008), and Pierson et al. (2010). Cohen et al. (2003) impute the underlying cost parameters of a supplier's problem in the semiconductor industry, where a supplier optimally balances his cost of delay with the holding cost and cost of cancelation in deciding the time to begin order fulfillment. Olivares et al. (2008) look at cost parameters of the newsvendor problem in the context of hospital operating room capacity decisions, where the optimal capacity decision is obtained by balancing the cost of overutilization with the cost of underutilization. Pierson et al. (2010) impute the cost placed by consumers on waiting time in a study of fast food drive-through restaurants, and implications for the firm's market shares. One of their key findings is that the cost customers place on waiting time is much higher than the earnings rate commonly assumed in prior literature. In the instance of online trading, Hann and Terwiesch (2003) present an economic model of consumer behavior that captures the tradeoff between the total frictional cost a consumer incurs and the desire to pay a price as close as possible to the threshold price to make an offer to the retailer. Using transaction data, they impute the frictional costs and find that consumers differ substantially in their frictional costs, which directly impacts their bidding process.

We follow similar approaches in our study and show how the imputed parameters of contribution and cost of labor can be used to calculate the optimal labor for each store. In addition, our panel dataset also allows us to study if there are any systematic factors that explain the variation in these parameters for different stores belonging to the same retail chain.

2.3 Research Setup

We obtained proprietary store-level data for *Alpha*[3], a women's specialty apparel retail chain. As of 2010, there were over 200 *Alpha* stores operating in 35 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Canada. These stores are typically in high-traffic locations like regional malls and shopping centers.

Alpha had installed traffic counters in 60 of its stores located in the United States during 2007. This advanced traffic-counting system guarantees at least 95% accuracy of performance against real traffic entering and exiting the store. This technology also has the capability to distinguish between incoming and outgoing shopper traffic, count side-by-side traffic and groups of people, and differentiate between adults and children, while not counting shopping carts or strollers. The technology also can adjust to differing light levels in a store and prevent certain types of counting errors. For example, customers would need to enter through fields installed at a certain distance from each entrance of the store in order for their traffic to be included in the counts, thus preventing cases in which a shopper enters and immediately exits the store from being included in actual traffic counts. It also provides a time stamp for each record that enables a detailed breakdown of data for analysis. This technology allowed us to obtain hourly data on traffic flow in each of the stores.

2.3.1 Definition of Variables

Let i be the index for a store and t be the time period. Here, a *time period* refers to a specific hour on a specific day of a month for the year, e.g., 10 a.m. to 11 a.m. on January 2nd. We denote for store i in time period t, $Store\_Sales_{it}$ as the dollar value of sales, $Actual\_Labor_{it}$ as the number of labor hours in the store, $Transactions_{it}$ as the number of transactions, and $Traffic_{it}$ as

[3] The name of the store is disguised to maintain confidentiality.

the store traffic or number of customers entering the store. CR_{it} and BV_{it} denote, respectively, conversion rate and basket value for store *i* during time period *t*.

2.3.2 Data Description

Alpha's stores were open during this time 7 days a week. Operating hours differed based on location as well as time period, e.g., weekdays and weekends. We obtained operating hours for each store and restricted our attention to normal operating hours. Of the 60 stores, five stores were in free-standing locations and five stores were in malls that did not have a working website to provide additional information needed to determine their operating hours. Moreover, there were nine stores, for which we did not have complete information for the entire year as they were either opened during the year or did not install traffic counters at the beginning of the year. Hence, we discard data from these 19 stores and focus on the remaining 41 stores that had complete information. These 41 stores were all located in malls/shopping centers and had a similar retail format. For example, a typical *Alpha* store would be approximately 4000 sq. feet in size. These stores are located across 17 states in the U.S.

Sales associates at *Alpha* are trained to provide advice on merchandise to customers, help ring up customers at the cash register, price items, and monitor inventory to ensure that the store is run in an orderly fashion. There is no differentiation in task allocation amongst the different store associates and they receive a guaranteed minimum hourly compensation as well as incentives based on sales. In contrast, an average Wal-Mart store is approximately 108,000 square feet in size and store associates are typically associated to specific product areas like electronics, produce and apparel, monitoring cash registers etc. *Alpha's* store managers were responsible for labor planning decisions as part of their day-to-day operations and the store managers' bonuses were derived as a percentage of store profits.

Working with data from one retail chain allows us to implicitly control for factors such as incentive schemes, merchandise assortments and pricing policies across stores. Data on factors such as employee training, managerial ability, employee turnover and manager tenure that could impact store performance are not available to us. We also have no information from any existing model on planned values of labor that are available to managers in this study, and the amount by which managers override these recommendations. However, as managers are compensated on both sales and payroll costs, we believe they would override model-based decisions largely in cases where they believe they can improve on them. So, by looking at the actual labor, we are implicitly looking at the planned values of labor.

We obtained additional demographic information like the number of women apparel retail stores, total number of clothing stores, population, median rental values, and median household income from EASI Analytics and Mediamark Research, Inc., which provide market research data collated from the Bureau of Economic Analysis (BEA), Bureau of Labor Statistics (BLS), and U.S. Census Bureau at the zip code level for each store. We augmented this with the average hourly wage rate of retail salespersons by Metropolitan Statistical Area (MSA) from the BLS.

2.3.3 Sample description

Staffing decisions could vary widely between weekdays, weekends, and holidays in accordance with the different traffic patterns and labor requirements. We excluded known holidays and the holiday season from our data set to avoid any impact of promotional sales during those days. Prior research and anecdotal evidence suggest that availability of store associates and consumer profile could differ between weekdays and weekends. For example, retailers typically tend to hire more part-time staff on weekends (BLS, 2009; Lambert 2008). Additionally, the consumer profile as well as buying behavior could differ between weekdays and weekends (Roy, 1994; Ruiz et al. 2004). Both of these factors could in turn impact the contribution of labor to sales and the cost of labor. In order to take these differences into account, we grouped our data into two subsamples: Weekdays (Monday – Thursday) and Weekends (Friday – Sunday) for our analysis. At this stage, we have 180 days in the Weekday data set and 130 days in the Weekend data set for each store.

Name	Definition	Weekdays				Weekends			
		Avg.	Std. dev	Min	Max	Avg.	Std. dev	Min	Max
$Store\_Sales_{it}$	Store sales	686.1	243.1	94.5	11020.5	1127.58	918.64	141.25	13067.46
$Actual\_Labor_{it}$	Actual labor	4.71	1.81	1.0	16.0	6.27	2.24	2.0	32.0
$Transactions_{it}$	Store trans-actions	7.14	4.59	1.0	46.0	11.71	7.08	1.0	72.0
$Traffic_{it}$	Store traffic	48.99	29.31	5.0	437.0	95.51	56.40	17.0	630.0
CR_{it}	Conversion Rate	16.79	2.43	9.40	20.19	13.38	4.14	1.85	25.89
BV_{it}	Basket Value	90.93	42.42	10.31	1371.26	94.58	50.11	15.50	1448.56

Table 2.1: Store variable names, definitions and summary statistics

Name	Definition	Average	Std Dev	Min	Max
$Stores_i$	Number of clothing stores in the zip code scaled by population (in thousands)	.064	.056	.001	.207
HHI_i	Median House Household Income for the zip code scaled by population(in thousands)	65.15	31.641	31.510	212.989
HHR_i	Median House Rent for the zip code scaled by population(in thousands)	1.05	.085	.102	3.15
$Comp_i$	Number of competing retailers in the zip code scaled by population (in thousands)	.028	.023	.002	.100
$MSAwage_i$	Average hourly wage rate for retail sales persons ($)	10.05	.634	8.96	11.67

Table 2.2: Demographic variable names, definitions and summary statistics

Our unit of observation is an operating hour for any given store. After removing outliers, we had a total of 73,800 hourly observations for weekdays and 53,300 hourly observations for weekends. All further analysis was conducted on these datasets. Tables 2.1 and 2.2 give a description of variable names, their definitions, and summary statistics of all store-related variables and demographic variables used in this study while Table 2.3 lists the known holidays that were excluded from our analysis.

Date	Holiday
Monday, January 1	New Year's Day
Monday, January 15	Birthday of Martin Luther King, Jr.
Monday, February 19	Washington's Birthday
Monday, May 28	Memorial Day
Wednesday, July 4	Independence Day
Monday, September 3	Labor Day
Monday, October 8	Columbus Day
Monday, November 12	Veterans Day
Thursday, November 22	Thanksgiving Day
Tuesday, December 25	Christmas Day
Sunday, April 8	Easter
Sunday, May 13	Mother's Day

Table 2.3: List of known holidays

2.4 Methodology and Estimation

In this section we explain the methodology used to determine if retail stores are understaffed. We determine that store i in time period t is understaffed if it carries less labor than that dictated by the optimal labor plan. We consider the time period of one hour in this analysis as it has been observed in practice that many retailers tend to have some flexibility in changing staffing levels on an hourly basis with use of part-time flexible workers. We explore the impact having schedule constraints that might prevent store managers from changing staffing levels on an hourly basis in later sections. The optimal labor plan is derived based on a model that captures the manager's past labor decisions, which we assume are rational and maximize store profits.

Several factors influence a store manager's decision about how much labor to have in store, including the availability of labor, the contribution of labor to sales, the direct and indirect costs associated with labor including compensation, bonus, insurance, medical benefits etc., the store manager's experience and skill in managing labor that could also include costs related to hiring and training the employees, managing the employee turnover etc., and constraints on flexibility in scheduling labor – all of which impact the staffing decisions and are not directly observable by the econometrician. Hence we intend to impute these parameters by using store managers' past labor decisions. In §2.4.1 we explain the decision model, in §2.4.2 outline the GMM estimation procedure, and in §2.4.3 provide the estimation details on the test and fit sample that we use for our analysis.

2.4.1 Optimal Labor Plan

We utilize a sales response and profit maximization model from prior literature that captures the tradeoff between cost incurred by the store manager to have labor in the store, and the contribution of labor to sales.

Sales response model:

From queuing theory, we know that an increase in the number of servers, or salespeople in our context, causes fewer customers to renege and consequently results in higher sales. For example, Wernerfelt (1994) shows that an increase in number of salespeople results in more interactions with customers that in turn results in higher sales. However, in a retail setting, it has often been observed that incremental increase in sales decreases during times of high traffic. Some causes for this include the negative effects of crowding on customers, having more browsers than buyers during peak hours and not having enough labor to satisfy the customer service requirements (Grewal et al. 2003). Theoretical literature in service settings has assumed that the relationship between revenue and labor would be concave (Hopp et al. 2007; Horsky and Nelson 1996). This insight is reflected in recent empirical research as well. Both Fisher et al. (2007) and Perdikaki et al. (2010) provide evidence supporting this assumption and find sales to be a concave increasing function of the staffing level. The following modified exponential model, proposed by Lam et al. (1998), captures these relationships between store sales (S_{it}), store traffic (N_{it}), and number of sales associates (l_{it}) in a store *i* at time *t:*

$$S_{it} = \alpha_i N_{it}^{\beta_i} e^{-\gamma_i / l_{it}} \tag{2.1}$$

where β_i is the traffic elasticity, γ_i captures the responsiveness of sales to labor (indirectly measuring labor productivity), and α_i is a store-specific parameter that captures the sales potential in the store. Here, overall store sales are positively associated with labor, but an increase in traffic and labor increases sales at a diminishing rate, i.e., $0 < \beta_i < 1, \gamma_i > 1$.

Profit-maximization model:

We use a linear profit function that adds sales force incrementally as long as the contribution to gross profit exceeds the incremental cost. The rationale behind our model is motivated by practice and literature that has studied staff scheduling problems. Lodish et al. (1988) studied the problem of sales force sizing for a large pharmaceutical company and found that a sizing model that trades off sales force expense against marginal returns was able to significantly improve the company's sales revenue. Lam et al. (1998) use a similar model to schedule retail staff but assume the wage rate is exogenously determined. Gross profit can be expressed as

$$\pi_{it} = S_{it} * g_i - l_{it} * d_i \tag{2.2a}$$

where π_{it} is the gross profit net of labor costs, S_{it} is the overall dollar value of sales, g_i is the average gross margin, l_{it} is the number of salespeople, and d_i is the hourly wage rate.

Deriving the labor decision rule:

As we do not have information on gross margin, we divide equation (2.2a) by gross margin, g_i, and use this as our objective function. Note that maximizing (2.2a) is the same as maximizing

$$\pi_{it} = S_{it} - l_{it} * w_i \tag{2.2}$$

where $w_i = d_i / g_i$ represents the adjusted hourly imputed cost of labor for each store, since pricing and labor decisions are independent. We refer to w_i as the implicit labor cost and to d_i as the unadjusted labor cost. Each store is expected to maximize the profit function in (2.2), yielding the following first-order condition for amount of labor to have in each store:

$$\gamma_i \alpha_i N_{it}^{\beta_i} e^{-\gamma_i / l_{it}} = w_i l_{it}^2 \tag{2.3}$$

Equation 3 is the decision rule for labor, and captures the way each store manager optimally balances the marginal cost and marginal revenue of having labor in the store. The optimal labor plan (l_{it}^*) is the value of labor that is a solution to Equation (2.3), given α_i, β_i, γ_i, w_i and store traffic (N_{it}). In reality, a store manager would not have access to real-time information on store traffic and would instead plan labor based on a forecast of store traffic. We discuss in appendix 6.1.2 the implication of this assumption for our estimate of imputed cost of labor (w_i).

Our method of structural estimation, described below, is advantageous in that it allows us to determine optimal labor even in the absence of store profit data. If we did have store profit data at the individual hourly level, joint estimation of equations (2.1) and (2.2) would have yielded the estimates required to calculate optimal labor for the store. However, store profit data, especially at the individual hourly level, is rarely collected. Moreover, even daily or monthly store profit data are usually difficult to obtain, as these are considered to be of high strategic value, so retailers tend to be reluctant in disclosing this information.

2.4.2 Estimating the contribution of labor to sales and cost of labor

To estimate the sales response parameters and impute the cost of labor, we follow the generalized method of moments (GMM) technique. This approach is similar to that used in Pierson et al. (2010) and Thomadsen (2005). We choose this technique for reasons similar to that described by these authors. In particular, use of GMM estimation method is advantageous as it needs no additional assumptions concerning the specific distribution of the disturbance terms, and it allows us to handle any endogeneity issues that may arise in our estimation. A detailed explanation of GMM estimation can be found in Hall (2005).

The sales response function and labor decision rule serve as moment conditions for GMM estimation. As the parameters $\alpha_i, \beta_i, \gamma_i, w_i$ are specific to each store, and we have year-long hourly data for each store, we estimate these parameters for each store separately to account for any fixed effects that might be present in our dataset. We augment the sales model to control for day-of-week effects by including indicator variables for each day of the week (Monday to Thursday for weekdays and Friday to Sunday for weekends).

Our sales response function for store *i* during time period *t* is given by:

$$S_{it} = \alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i / l_{it}} \varepsilon_{1it} \tag{2.4a}$$

where d denotes the day of week and $a_d = 1$ if day of week *d* = 1, 0 otherwise. Similarly, the labor decision rule is given by:

$$\gamma_i \alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i / l_{it}} = w_i l_{it}^2 \, \varepsilon_{2it} \tag{2.4b}$$

where $\varepsilon_{1it}, \varepsilon_{2it}$ represent unit mean residuals for the sales response function and labor decision rule, i.e., $E[\varepsilon_{1it}] = E[\varepsilon_{2it}] = 1$. Then, based on equations 2.4a and 2.4b, using a log-transform, we have the following two moment conditions:

$$E[z_{1it} \left\{\log (S_{it}) - \log (\alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i / l_{it}})\right\}] = 0 \quad \text{i.e.} \quad E[z_{1it}\, \vartheta_{1it}] = 0$$

$$E\left[z_{2it} \left\{\log (\gamma_i \alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i / l_{it}}) - \log (w_i l_{it}^2)\right\}\right] = 0 \quad \text{i.e.} \quad E[z_{2it}\, \vartheta_{2it}] = 0 \tag{2.4c}$$

where $Z_{it} = \{z_{1it}, z_{2it}\}$ represents the set of instruments and $\Theta = \{\alpha_i, \alpha_{id}, \beta_i, \gamma_i, w_i\}$ represents the vector of parameters to be estimated. The above two equations are also known as the population moment conditions.

An important estimation issue that needs to be tackled is that of possible endogeneity between store sales (S_{it}) and labor (l_{it}). Endogeneity between these two variables can arise due to a few reasons. First, it is commonly assumed that store managers determine store labor based on expected (or forecast) demand, where demand could be measured as sales or traffic. Since actual sales and expected demand are typically highly correlated, the coefficient of labor will suffer from endogeneity bias if we do not explicitly control for expected demand. In our setting, we possess the actual traffic data that allows us to mitigate this bias as we expect actual traffic to be correlated with expected demand. Second, unobserved factors such as store size could be correlated with both sales and labor, and result in endogeneity between sales and labor. However, our use of store fixed-effects helps us mitigate this bias. Finally, use of aggregate data for sales and labor will cause simultaneity bias. For example, in a regression of weekly sales against weekly labor, not only can labor drive sales, but also sales may drive labor as managers can observe sales in the early part of the week and change labor accordingly. Our use of hourly data removes this bias as there is not enough reaction time to change labor. To statistically validate our assumption that endogeneity bias is not present in our setting, we performed an endogeneity test called C-statistic test (Hayashi, 2000) and found that our null hypothesis that labor may be treated as exogenous cannot be rejected (p-value > 0.25). Hence, we use $z_{1it} = z_{2it} = \{N_{it}, l_{it}, a_d\}$. We also conducted an additional robustness check, wherein following past literature (Bloom and Van Reenen 2007, Siebert and Zubanov 2010) we used lagged labor as instruments and found similar results. One possible reason for our estimates to remain unchanged is that we estimate our coefficients separately for each store, which, in turn, allows us to effectively control for any unobservable store-manager characteristics. Unfortunately, we have no information about store-manager turnover in our sample, so cannot confirm if any store managers changed during our observation period.

Based on the population moment conditions, we must have for each store i the sample average of the vector of random variables Z,

$$G_i(\theta_i) = \frac{1}{T}\sum_{t=1}^{T} Z_{it}\ \varepsilon_{it}(\theta_i)$$

as close to zero as possible (where T = total number of individual hourly observations for store i). The GMM estimator determines a parameter vector $\widehat{\theta_i}$ that minimizes a quadratic function of this sample average. More specifically, the GMM estimate is the vector $\hat{\theta}_i$, which optimizes

$$\min_{\theta_i} G_i(\theta_i)' A_i G_i(\theta_i)$$

where A is a weighting matrix for the two moments. We use a commonly followed two-step estimation method. In the first step, we use GMM with the pre-specified weighting matrix $A_{1i} = I$, the identity matrix that gives an initial estimate, $\hat{\theta}_{1i}$, which is also consistent. We use $\hat{\theta}_{1i}$ to estimate the asymptotic variance–covariance matrix of the moment conditions:

$$A_{2i} = (E\left[G_i(\hat{\theta}_{1i})G_i(\hat{\theta}_{1i})'\right])^{-1}$$

The same GMM procedure is now run a second time with this new weighting matrix to arrive at our parameter estimate, $\hat{\theta}_{2i}$.

2.4.3 Estimation results

Our objective in imputing the parameters $\alpha_i, \beta_i, \gamma_i, w_i$ is to use them towards determining the optimal labor for each of the 41 stores. This estimation framework is described graphically in Figure 2.1a.

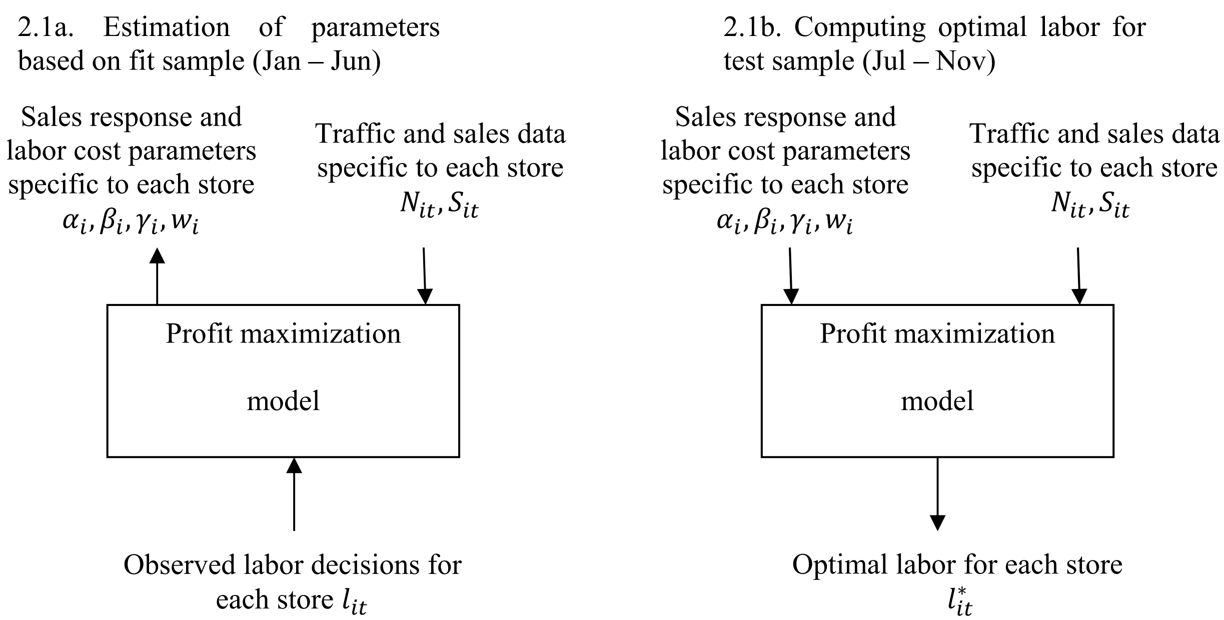


Figure 2.1: Methodology to compute optimal labor

We first use the GMM method of estimation to estimate the parameters. In order to prevent any look-ahead in our estimation process, we also divide both our weekday and weekend samples into a fit sample and a test sample. The fit sample (which includes data from months of Jan – June) is used to estimate $\alpha_i , \beta_i , \gamma_i , w_i$. These estimates are summarized in Table 2.4. For ease of comparison, we also compute the average unadjusted labor cost, d_i, using a gross margin value of 0.48 (this value of gross margin is obtained from the company's 10k report for 2007, the year of our observations). Estimates of the model for each store specification are given in appendix 6.1.1.

	Weekdays				Weekends			
Parameter	Average	Std Dev	Min	Max	Average	Std Dev	Min	Max
$\boldsymbol{\alpha_i}$	36.96	10.39	17.8	56.72	51.50	9.59	33.35	74.45
$\boldsymbol{\beta_i}$	0.29	0.08	0.13	0.42	0.21	0.07	0.11	0.34
$\boldsymbol{\gamma_i}$	12.07	2.93	6.84	19.66	36.64	7.23	24.15	53.58
$\boldsymbol{w_i}$ $(\$/\boldsymbol{hr})$	63.49	22.35	21.88	114.42	40.61	17.83	18.95	79.58
$\boldsymbol{d_i}$ $(\$/\boldsymbol{hr})$	30.47	10.73	10.50	54.92	19.74	7.06	9.10	38.2

Table 2.4: Estimates of model from fit data set: $S_{it} = \alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}}$, $\gamma_i \alpha_i {\alpha_{id}}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}} = w_i l_{it}^2$

These estimates were found to be significant (p<0.05) for each of the 41 stores in our data set. We also conducted a robustness test by comparing the estimates of the parameters $\alpha_i, \beta_i, \gamma_i, w_i$, obtained from the fit sample with that obtained from the full sample (i.e. including all observations from Jan – Nov) and are summarized in Table 2.5. Both our estimates and results based on these estimates based on the fit sample were not significantly different from those obtained based on the full sample. This indicates that the parameters are stable across time and justifies our approach of using a fit sample for estimating the parameters and a test sample to test our predictions.

	Weekdays				Weekends			
Parameter	Average	Std Dev	Min	Max	Average	Std Dev	Min	Max
$\boldsymbol{\alpha_i}$	31.16	10.54	17.2	55.22	49.50	9.51	32.31	73.58
$\boldsymbol{\beta_i}$	0.29	0.09	0.12	0.45	0.20	0.07	0.11	0.34
$\boldsymbol{\gamma_i}$	11.78	2.95	5.81	18.61	35.41	7.25	23.11	52.51
$\boldsymbol{w_i}$ ($\$/\boldsymbol{hr}$)	61.98	20.32	20.55	110.14	38.28	17.81	19.98	80.54
$\boldsymbol{d_i}$ ($\$/\boldsymbol{hr}$)	29.75	9.75	9.86	52.86	18.37	8.54	9.59	38.65

Table 2.5: Estimates of model from full sample: $S_{it} = \alpha_i \alpha_{id}{}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}}$, $\gamma_i \alpha_i \alpha_{id}{}^{a_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}} = w_i l_{it}^2$

The average unadjusted imputed cost of labor, d_i, across 41 stores based on data from weekdays is \$30.47, while the standard deviation, minimum and maximum values are \$10.73, \$10.50, and \$54.92 respectively. We find qualitatively similar results for weekdays and weekends, and hence describe all results based on the weekdays subsample. The corresponding values for weekends are shown in the respective tables. This average unadjusted imputed cost of labor, d_i, is directly comparable to the average hourly wage rate of retail salespersons (*MSAwage$_i$*) and allows us to determine if store managers associate greater or the same costs to labor relative to average hourly wage rate for retail salespersons. We find that the average value, \$30.47, is significantly higher than the average hourly wage rate of \$10.05. A one-tailed *t*-test of d_i>*MSAwage$_i$* for each store showed this difference to be statistically significant (p<0.001).

Furthermore, we find significant difference between weekdays and weekends in the estimates of the parameters α_i , β_i , γ_i , that capture the sales potential in the store, the amount of traffic converted to sales and the contribution of labor to sales respectively. The average traffic elasticity, (β_i), for each of the 41 stores, was found to be lower during weekends as compared to weekdays (p<0.1). Similarly, the responsiveness of labor to sales, (-γ_i) was found to be significantly lower on weekends than on weekdays for each of the 41 stores (p<0.05). This supports anecdotal evidence that there a relatively higher number of browsers who tend to visit the stores during weekends as compared to weekdays leading to lower conversion of traffic to sales. We also find that the average unadjusted imputed cost of labor was significantly higher on weekdays than on weekends (p<0.01), supporting prior literature that has documented the use of higher usage of lower wage part-time labor on weekends in other retail organizations (Lambert, 2008). The estimates of parameters α_i , β_i , γ_i and the imputed cost of labor, w_i , also vary widely across different stores indicating heterogeneity in the way labor contributes to sales and to cost at each of these different stores. We explore if there are any systematic factors that might explain these differences in §2.6.

Thus, our results show that it is important to take into account the heterogeneity amongst different store locations when estimating the contribution of labor to sales and cost.

2.5 Results

In order to determine if a store is understaffed, we use equation 2.3 to compute the optimal labor plan, l_{it}^{*}, for the test sample (Jul- Nov) using estimates of $\alpha_i, \beta_i, \gamma_i$, w_i, the actual store traffic N_{it}, and then compute the deviation of actual labor from the optimal labor plan (i.e. $\Delta l_{it} = l_{it}^{*} - l_{it}$). We use the term labor deviation, as opposed to labor mismatch, as labor mismatch was defined as the difference between actual labor and planned labor in prior literature. This procedure is shown graphically in Figure 2.1b. Positive deviations would represent understaffing, while negative deviations would represent overstaffing relative to the optimal labor plan. All results presented hereon are for the test sample (Jul- Nov).

It is possible to compute the extent of understaffing in a store for different levels of granularity, viz., hourly, daily, weekly, monthly, etc. We first discuss deviations at the daily level and then at the hourly level, as these capture the main aspects of our analysis. Deviations at the daily level help us determine if stores are understaffed or overstaffed for majority of the days. Deviations at the hourly level help us understand if stores are systematically understaffed or overstaffed for certain hours of the day. Deviations at the daily level are calculated as the average deviations across different hours of each day for each store; while deviations at the hourly level are computed as average deviation for a given hour across different days for each store. For each store i let d represent a day and h represent an operating hour, i = 1……41, d = 1….D (D = 85 for weekdays and 60 for weekends) and h = 1……H (H = total operating hours). Then, for each store i, daily deviations, $\Delta l_{id} = \{\sum_{h=1}^{H}(\Delta\, l_{idh})\}/H$ and hourly deviations $\Delta l_{ih} = \{\sum_{d=1}^{D}(\Delta\, l_{idh})\}/D$.

We have 3,485 total store-days (85 days at each of 41 stores) in our weekdays test sample and 2,460 total store-days (60 days at each of 41 stores) in our weekend test sample. We describe results here for the weekdays but find qualitatively similar results for weekends as well. We find that the stores are understaffed 44.2% (1,540 store-days) and overstaffed 55.8% (5,205 store-days) of the time. We test for statistical significance in the following way. For each store, we perform a one-tailed binomial test of to determine if the proportion of days the store is understaffed exceeds 0.5 (or 50%). We find that this proportion is not statistically different from 0.5 for 37 of the 41 stores at $p<0.1$. The remaining 4 stores were found to be understaffed ($p<0.05$). If we look at the magnitude of deviations, the average understaffing at the daily level is 0.48 labor-hr (5.4% of the optimal labor), and the average overstaffing is 0.23 labor-hr (2.6% of the optimal labor). Thus, we find no evidence for understaffing at the daily level and, in fact, find that most stores appear to have the right amount of labor.

Because it is possible for stores to be systematically understaffed in certain hours and overstaffed during the other hours and still appear to have the right amount of labor at the daily level, we repeat our analysis at the hourly level to detect any systematic understaffing or overstaffing. There are 850 total store-hours (~10 operating hours and 85 days). At the hourly-level we find that stores appear to be understaffed only 36.5% of the time. A one-tailed binomial test shows that stores are significantly overstaffed during most hours ($p<0.05$). Thus, we find that even though stores might have the right amount of labor at the daily level, they may be overstaffed most hours during the day. This counterintuitive result could be explained if the understaffing, when it occurs, has a large magnitude compared to overstaffing. To test this, we look at the magnitude of the deviations. Average understaffing at the hourly level is 4.94 labor-hr (23.1% of optimal labor), and the average overstaffing is 2.07 labor-hr (10.5% of optimal labor).

Thus, even though the stores appear to have the right amount of labor at the daily level, there are certain hours of the day when they suffer from large understaffing.

Interestingly, we find that in most cases the stores appear to be understaffed during the same hours of the day. Thus we can rule out understaffing being driven by randomness in the arrival process across hours of the day. Further analysis of traffic flow into the stores reveals that understaffing typically occurs during peak hours, where peak hours are defined as the three-hour duration when atleast 70% of the daily traffic arrives. We confirm this by running a logistic regression and find statistical support to show that understaffing occurs during peak hours (p<0.05). Figure 2.2 shows the plot of actual and optimal labor during peak and non-peak hours to depict the widespread prevalence of understaffing during peak hours.

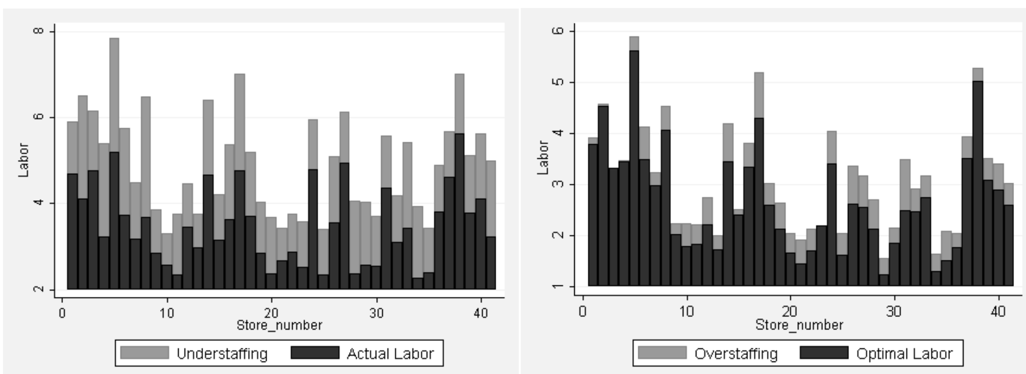


Figure 2.2: Comparison of actual labor and optimal labor for stores during peak and non peak hours

2.5.1 Causes of understaffing and its consequence on store profitability

Next we want to understand the impact of understaffing on profitability and the sensitivity of profitability to factors that drive understaffing at this retail chain. We do so by first calculating the theoretical upper-bound of the profits that this retailer could have achieved with the optimal labor plan. Because it is essential to reduce understaffing without increasing overstaffing, we measure the impact of the optimal labor plan on the profitability for all the hours (and not limited to hours when understaffing occurs). Such an optimal plan would not be realistic as it assumes that retailers would have perfect foresight of the incoming traffic and be able to change labor on an hour-to-hour basis. In §2.5.3 and §2.5.4, we relax both these assumptions and study the impact of forecast error and scheduling constraints on store profitability.

2.5.2 Quantifying improvement in store profitability from the optimal labor plan

Our procedure to quantity the improvement in store profitability from the optimal labor plan is as follows. First, we calculate the sales lift for each store *i* in each time period *t* (in the test sample) using equation 2.5 as shown below.

$$S_{it}^{o} = \hat{\alpha}_{\iota}\hat{\alpha}_{id}{}^{a_d} N_{it}{}^{\hat{\beta}_i} e^{-\hat{\gamma}_i / l_{it}^{*}} \tag{2.5}$$

Here l_{it}^{*} is the optimal labor plan that was generated as explained in the previous section and S_{it}^{o} is the sales generated using the optimal labor plan.

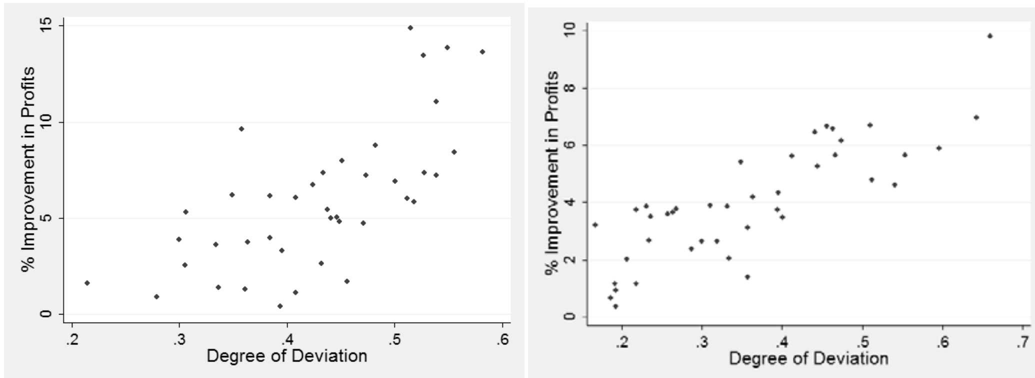
Next, we use the imputed cost w_i to compute optimal profit as:

$$\pi_{it}^{o} = S_{it}^{o} - \hat{w}_i * l_{it}^{*} \tag{2.6}$$

Since actual profit data are not available at the hourly level, we substitute actual sales and actual labor in equation 2.6 to compute the actual profits. The difference between optimal profit and actual profit represent the improvement in store profitability from using an optimal labor plan.

We find that the average improvement in profitability to be 5.8% in the weekdays sample and 3.85% in the weekend sample. Further, we also observe that about 60% of the improvement in profitability can be attributed to increasing staffing levels during times when the stores were understaffed. To examine if the improvement in profitability is larger for stores whose actual labor deviated more from the optimal labor we do the following. We plot the deviations against improvements in profits as shown in Figure 2.3. Our results show that stores that currently deviate most from the optimal labor plan will have the greatest improvement in profitability, as expected. This improvement can be as high as 8.1% in the weekdays sample for stores that fall in the top quartile based on their labor deviation.

As a robustness test, we also plot the deviation between actual and optimal labor against the average conversion rate and basket values of the 41 stores as shown in Figures 2.4a and 2.4b.



To capture the extent of both understaffing and overstaffing and to facilitate comparison across stores, we define the degree of deviations as $\Delta l_{1i} = \{\sum_{t=1}^{T} |\Delta\, l_{it}|\}/\{\sum_{t=1}^{T}(\, l_{it})\}$.

Figure 2.3: Scatter plot of percentage improvement in profits against degree of deviation across stores for weekdays and weekends

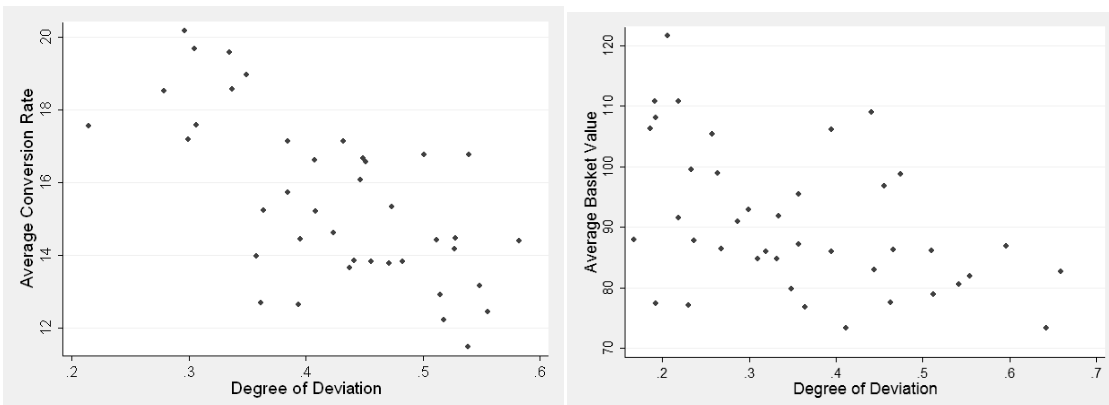


Figure 2.4a: Scatter plot of average conversion rate and basket value against degree of deviation across stores for weekdays

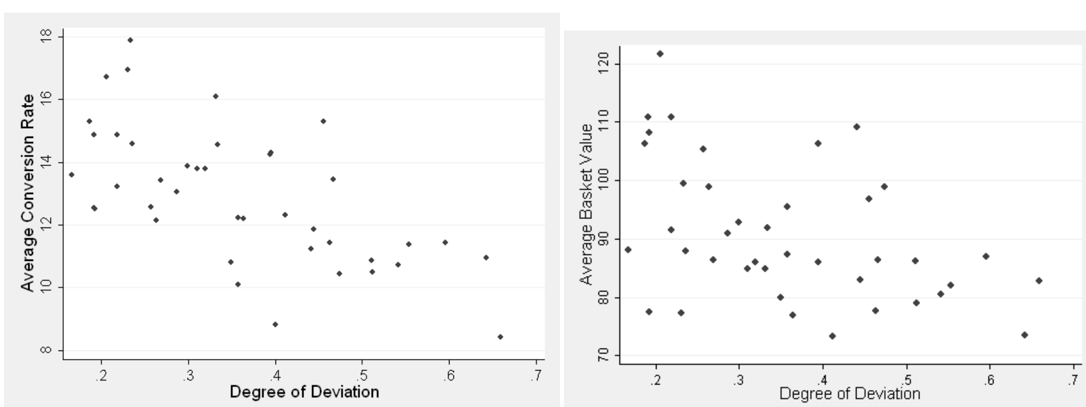


Figure 2.4b: Scatter plot of average conversion rate and basket value against degree of deviations for different stores – weekends

We find that stores having low deviations also have higher CR and BV. These differences are statistically significant as shown in Table 2.6. Thus, our results are consistent with prior literature (Netessine et al. 2010) that has shown that greater mismatches in labor[4] are associated with lower basket values.

	Weekdays		**Weekends**	
	Low deviation	High deviation	Low deviation	High deviation
Mean CR	17.37	13.49	15.28	12.17
Difference in mean CR (t-stat[b])	3.9(.827***)		3.11(.931***)	
Mean BV ($)	96.21	89.48	101.89	91.20
Difference in mean BV ($) ($t$-stat)	6.73(1.181***)		10.69(2.279***)	
Mean Store Profits ($)	643.56	301.72	1092.18	628.17
Difference in mean Store Profits ($) ($t$-stat)	341.84(2.524***)		464.01(3.046***)	

[a] Degree of deviation= Δl_{1i}, [b] Paired one tailed test that mean of CR, BV and store profits for stores with low deviations is higher than for stores with high deviations. ***denotes statistically significant at p<0.001, **at p < 0.05 and *at p < 0.1 level

Table 2.6: Comparison of conversion rate, basket value and store profits for stores with higher and lower degree of deviation[a]

2.5.3 Contribution of traffic forecast errors to understaffing and its consequence on store profits

Next we examine the impact of not having perfect information on incoming traffic on store profitability. We do so in the following manner. Instead of generating the optimal labor plan with actual traffic as described in the previous section, we generate an optimal labor plan based on forecasted traffic. We generate traffic forecasts by using a standard time series Newey-West model. The forecasts are generated one to three weeks in advance, as this is the typical time

[4] We note that this literature has measured labor mismatch as the deviation of actual labor from planned labor.

period for scheduling labor[5]. In this setting, we find that as the forecast horizon increases from 1 week to 3 weeks, forecast errors increase from 12% to 25%. These forecast errors result in labor plans that cause both understaffing and overstaffing. However, the extent of understaffing and overstaffing is still lower than the current labor plan as shown in Table 2.7. Thus we find that labor plan in these cases also generate higher profits (3.3% to 4.0%) than that from the current labor plan. Recall that the improvement in store profits with perfect information about traffic was 5.8%. Thus while common wisdom might indicate that the lack of ability to have real time information on traffic is the major cause of understaffing (and overstaffing), we find that it only partially contributes to the improvement in store profitability.

Labor plan		Weekdays			Weekends		
		% Profit improvement	% under-staffing	% over-staffing	% Profit improvement	% under-staffing	% over-staffing
Optimal		5.8	0.0	0.00	3.85	0.00	0.00
Actual		0.0	23.1	10.5	0.0	25.6	8.5
Generated with traffic forecast[a]	1 wk	4.0	5.17	3.26	2.75	7.58	2.12
	2 wk	3.7	8.16	4.16	2.31	9.57	3.18
	3 wk	3.3	10.75	5.29	1.54	12.36	4.56
With scheduling constraint requiring constant labor for	2 hr	3.4	6.51	5.23	1.25	8.43	3.16
	3 hr	2.1	10.78	6.51	0.95	12.07	4.67
	4 hr	1.5	15.14	8.71	0.66	17.14	7.11
	5 hr	0.8	22.50	11.80	0.06	24.13	9.55

[a]1 week, 2 week and 3 week ahead forecasts correspond to an average forecast error of 12%, 17% and 25% respectively.

Table 2.7: Result of % improvement in profits from incorporating traffic forecasts and constraints in labor scheduling

.

[5] A New Approach to Retail Workforce Forecasting, RedPrairie, 2010

2.5.4 Contribution of scheduling constraints to understaffing and its consequence on store profits

We now look at another possible reason—scheduling constraints—for the understaffing observed at the hourly level. Many retail organizations prefer to schedule employees for a certain minimum number of hours per shift to ensure employee welfare and/or meet government or union regulations. In many organizations, this minimum is 4 hours per shift (Quan 2004). Such a constraint could lead to understaffing in some shifts.

To examine how much of the observed understaffing is explained by this scheduling constraint, we do the following. We compute the optimal labor plan as explained in §2.5.1 to get the optimal labor for each hour, assuming perfect information about future traffic. Next we impose the constraint requiring labor to be constant for a block of time by taking the average labor for the hours in that block and using it for that block of time. Other heuristics such as peak labor for those hours in a block or minimum labor during the hours in a block do not increase profitability. We consider 2-hour, 3-hour, 4-hour, and 5-hour blocks of time[6] in our analysis.

We find that the improvement in profits achieved with the optimal labor plan is dissipated with decrease in scheduling flexibility as shown in Table 2.7. The improvement in store profits drop from 5.8% (in the case of the optimal labor plan with a 1 hour scheduling constraint) to 1.5% when a 4 hour constraint is imposed. Many retailers plan labor 2 weeks in advance and schedule labor in 4 hour blocks. For such retailers, our study shows that their profits are impacted more by their scheduling constraint than by their lead time for labor planning. Thus our results appear to support the recent moves by many retailers like Wal-Mart and Payless ShoeSource towards more flexible work schedules (Maher, 2007).

[6] We did not include the first hour of operation (8am) in shift scheduling as even though the optimal labor may indicate lower labor requirements due to low traffic flow, stores may actually require additional employees for store opening related activities. Including this first hour would make our results even stronger.

Our results from §2.5.3 and §2.5.4 quantity the impact of reducing forecast errors and increasing scheduling flexibility on improvement in store profitability. We now look at the interaction of forecast errors and scheduling constraints on store profits with help of a simulation (details in appendix 6.1.3). The percentage loss from optimal profits with increasing forecast errors and scheduling constraints is shown in Figure 2.5. We observe the following effects of interaction of forecast errors and scheduling constraints. First, we see that scheduling constraints exacerbate the impact of forecast error. This can be seen from the rapid increase in loss in profits from optimal for higher values of forecast error and tighter scheduling constraints. For example, with a 2 hour scheduling constraint, doubling traffic forecast error from 10% to 20% leads an increase in loss from 2.5% to 5.4%. On the other hand, with a 4 hour scheduling constraint, the concomitant increase in loss is from 8% to 13%, i.e. the impact of increase in forecast error is almost doubled. Second, effects of lack of sophisticated technology in forecasting can be mitigated with schedule flexibility and vice versa depending on which of these are more easily implemented at the different stores. For example a 30% forecast error, with 2 hour scheduling constraint (9.9%), yields the same loss in profits as a 15% forecast error, with 4 hour scheduling constraints. This result is of practical interest, as many retailers often cite a need for sophisticated software to produce accurate forecasts as one of the most critical components of store operations[7]. Our simulation experiment here shows that although accurate forecasts are valuable, they alone would not help store managers to significantly increase store profits without increasing the schedule flexibility. On the other hand, given that scheduling in block is a practical constraint that many retailers face to attract and retain employees, reducing forecast errors with help of new technologies might mitigate the impact of these constraints on store profitability.

[7] Integrated Solutions for Retailers. December 2010. Retail Tech 2010/2011: Where We've Been, And Where We're Headed from Here.

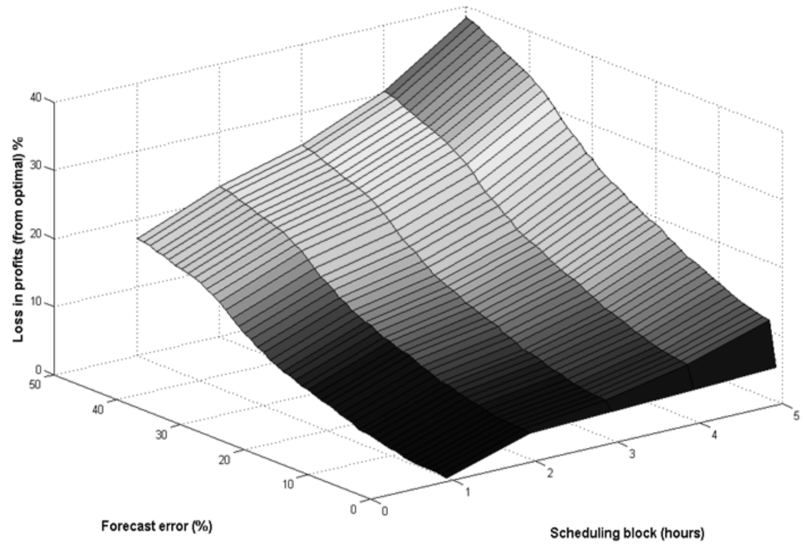


Figure 2.5: Impact of forecast errors and scheduling constraints on store profits

2.6 Discussion

In this section we explore the broader implications of our results for retail store operations. While several authors (Lam et al. 1998; Thomadsen, 2005) have acknowledged that the cost of labor is driven by many store specific factors, we are not aware of any study that has measured these costs at the store-level. Gino and Pisano (2008) emphasize that managers tend to make decisions based on intrinsic costs and not accounting costs. Similarly, Schweitzer and Cachon (2000) find that managerial decisions could systematically deviate from traditional assumptions when managers weight the parameters in the decision making process differently. A major finding of this study is that the imputed cost of labor varies significantly among the different stores, even though they belong to the same retail chain. Hence, we investigate if there are any systematic factors, based on local market characteristics, which influence the differences in cost of labor across these stores as it would indicate if store managers take local market characteristics into account in their labor decisions.

In a retail bank setting, Campbell and Frei (2010) find that operating managers take local market characteristics into account when deciding on the number of tellers to schedule. They identify the cost that customers place on high service time to be one such local market characteristic and show competition and median household income to be suitable proxies for this cost. Thus, if store managers perceive that customers place higher costs on service time in their locations, they might aim to provide a higher service level and place relatively lower emphasis on the cost of labor. Examples where managers place lower emphasis on cost while placing higher emphasis on service level have also been found in other settings (Png and Reitman, 1994; Ren and Willems, 2009). We investigate whether the implicit costs that customers place on high service time can help explain the differences in imputed cost of labor in our setting as well. We use the number of women's clothing stores as a proxy for competition ($Comp_i$) and median household income (HHI_i) as a proxy for high value that customers place on waiting time in the area. In addition, labor cost is dependent on the demand for labor. Hence, we include the number of local clothing stores ($Stores_i$) as a proxy for employment opportunities in the area. Since sales associates' skills may be fairly generic so that other types of stores may increase demand for the associates' labor as well, we repeat our analysis with the total number of retail stores as a proxy for employment opportunities and find no qualitative difference in our results. Finally, rental expenses for the different stores may vary across different locations, especially in cases where these rental expenses are calculated as a percentage of overall sales. As the gross margin reported in the 10-k statement is inclusive of store occupancy costs, it is possible that the gross margin (g) in our profit model might differ across stores based on these rental expenses and indirectly influence the imputed cost of labor. We proxy these rental expenses by the median household rent (HHR_i) to control for differences in imputed cost of labor that may arise out of these rental expenses. Finally, we used average store sales volume to control for store size. We run a cross-sectional regression where for each store i,

$$w_i = \alpha_0{}^w + \alpha_1{}^w Stores_i + \alpha_2{}^w Comp_i + \alpha_3{}^w HHI_i + \alpha_4{}^w HHR_i + \alpha_5{}^w Store\_Sales_i + \varepsilon_i{}^w \quad (2.7)$$

Table 2.8 displays results of this regression. In line with our expectations, we find that a higher imputed cost is negatively associated with higher values of household income and competition, i.e., $\alpha_2{}^w < 0$ and $\alpha_3{}^w < 0$, and is positively and significantly related to higher opportunities for employment and higher rental values, i.e., $\alpha_1{}^w > 0$ and $\alpha_4{}^w > 0$ (significant at p<0.05). When we include the average hourly wage rate for retail salespersons, the coefficient is insignificant and does not change our results. This could be driven by the lack of sufficient heterogeneity in wage rate as a large number of stores fall in the same MSA and hence have the same average hourly wage rate. These results suggest that store managers take local market characteristics into account when determining the amount of labor required in their stores.

Variable	Weekdays	Weekends
Intercept	$22.16^{***}(5.92)$	$20.15^{***}(3.05)$
$Stores_i$	$37.45^{***}(5.11)$	$29.41^{***}(2.01)$
$Comp_i$	$-105.47^{**}(21.82)$	$-115.16^{***}(12.21)$
HHI_i	$-.175^{**}(.07)$	$-.118^{**}(.06)$
HHR_i	$12.11^{**}(.17)$	$11.18^{**}(.07)$
$Store\_Sales_i$	.01(.01)	0.003(0.01)
Adjusted R^2	0.31	0.27
n	41	41

***denotes statistically significant at p<0.001, **at $p < 0.05$ and *at $p < 0.1$ level

Table 2.8: Regression of imputed cost of labor on local market area characteristics

Our finding that the imputed cost of labor is driven by local market characteristics has implications for labor planning in the retail setting. There has been considerable debate over the merits and de-merits of centralized and decentralized decision-making in operations management literature. Theoretical literature (Anand and Mendelson 1997, Chang and Harrington 2000) indicates that decentralized decision-making can lead to better performance when local knowledge is important and centralized decision-making leads to better outcomes when local knowledge is of little value. In practice, many retailers deploy workforce management tools centrally and input the cost of labor to produce the optimal labor plan for their stores. It is unclear to what extent the true imputed cost of labor is used in these calculations and the implication for store profitability. This could result in misalignment between corporate office and store manager regarding labor decisions resulting in sub-optimal solutions or valuable store manager time spent in overriding the corporate office decisions. In fact, there is a lot of evidence that store managers do not always follow recommendations from a centralized planning system (van Donselaar et al. 2010; Campbell and Frei, 2010; Netessine et al. 2010). Our methodology may be used to measure the imputed cost of labor for each store and use it to drive labor decisions for each store. Future research may investigate if the centralized decisions become more aligned with store manager's decisions as a consequence.

2.7 Conclusion

In this chapter, we examine whether or not retail stores are understaffed based on the traffic flow, sales volume as well as the contribution and cost of labor at each of these stores. We find that, on average, at the daily level, managers seem to have the required amount of labor in the store. However, our results also indicate the stores are consistently understaffed at the individual hourly level, especially during peak hours, which negatively impacts store performance. These results support Fisher's (2010) suggestion that an analysis of the contribution of store labor to store profit is best done hour by hour for each store.

Our study also shows that decreasing forecast errors and increasing schedule flexibility would reduce understaffing and lead to higher profits for retailers. These results support the recent move by several retailers who invest heavily in emerging technologies that integrate traffic information with workforce management (Stores, Jan 2010)[8]. At the same time, we also find

---

[8] Scheduled Improvements, Stores Jan 2010.

instances where some workforce management tools recommend changing schedules every fifteen minutes. Such drastic changes in schedules transfers the risk onto hourly workers (Lambert et al. 2008) and leads to variability and unpredictability into the schedules of these workers (Henly et al. 2006). Hence retailers have to be cautious in their choice of strategies to improve forecast errors and scheduling flexibility as some of their actions may lead to employee dissatisfaction and lower long-term profitability.

CHAPTER 3

Improving Store Operations through Better Traffic Forecasts

3.1 Introduction

Retailers invest heavily in customer-facing technologies in order to provide a differentiated in-store experience that forms a critical component in their strategy to gain and retain customers. Prominent among these technologies are the increasing adoption of customer measurement systems, including traffic counting, which would provide retailers with vital information for measuring, managing and improving the customer's in-store experience.

While traffic counting systems have been growing in sophistication and scope, two major applications of traffic data have emerged in the recent years. First, customer traffic data provides retailers with the ability to calculate the ratio of shoppers who end up purchasing an item to overall store traffic (also known as closing ratio or conversion rate) and continues to be a key metric for retailers to monitor and manage. Second, availability of historic traffic information at a granular level (e.g. hourly traffic data) provides retailers with the opportunity to closely tie their staffing requirements in each store to the incoming demand (Stores, 2010).

Prior to this, many of the staffing applications relied on historical sales to forecast future demand and generated weekly schedules on the basis of this forecast. However, this method tends to underestimate the actual demand, especially during times of peak traffic. Without the availability of traffic data, retailers did not have any systematic methods to keep track of the actual demand volume in store. There is anecdotal evidence that in some cases, store managers would take initiatives to note changes in demand trends, but this practice was largely qualitative and subject to individual bias (Fisher et al 2005).

As traffic forecasts are the building block of staffing decisions, we feel it would be worthwhile to study to what extent retailers can leverage information on actual traffic and use it to capture the true customer sales opportunity within their store. In this chapter, we analyze the fit of different demand distributions like normal, Poisson, and negative binomial, to the traffic data and seek to answer the following research questions: 1) Is there a general distribution that can characterize the traffic to different retail stores, 2) Are there any systematic factors based on local market characteristics that explain variation in traffic, and 3) Does this information help generate traffic forecasts that improve over traditional forecasting methods. If yes, how do these forecasts help store managers plan and schedule labor requirements and improve store performance by maintaining the targeted service level requirements.

In this chapter, we investigate the distribution of traffic and its implication for labor planning by analyzing hourly traffic data from 60 stores of a women's apparel retail chain over a one-year period. This research environment offers several advantages. First, the traffic data are available at an hourly level that allows us to analyze traffic variability across different hours of the day and its impact on determining staffing requirements. Second, we utilize the heterogeneity in the locations of the 60 stores by separately collecting data on the locations' market characteristics like household income, apparel spending, and competition to study the impact of these factors on the resulting traffic distribution.

Our study makes the following contributions to the operations management literature. To the best of our knowledge, this is the first study on investigating the underlying distribution of retail store traffic and its implication for generating forecasts for labor planning using panel data on a large number of stores. Through this study we validate theoretical assumptions on customer arrival process in context of a retail store setting. Our study also provides empirical support to findings in the theoretical literature (Lariviere and Van Mieghem, 2004) on use of Poisson distribution for modeling customer arrivals under competitive settings.

This paper is organized as follows. In §3.2 we review the background literature and in §3.3 explain the models used for analyzing the traffic distribution. In §3.4 we explain the data and variables used in the chapter. In §3.5 we report the findings on fit of different models and discuss some of the systematic factors that could explain variation in traffic in §3.6. In §3.7 we demonstrate an application of knowledge of traffic distribution to labor planning, and finally present our conclusions in §3.8.

3.2 Literature Review

In this section, we briefly review two streams of literature relevant to our work. The first is from retail operations where several researchers have studied the different demand distributions that best fit retail demand (item level demand or sales) and aggregate customer purchases. The second stream of literature that is related to our work is from queuing theory. Specifically we refer to the modeling of customer arrival process in the call center literature where, in recent years due to availability of detailed call center data, several researchers have studied the validity of assumptions made in the queuing theory literature in practice. The main focus of the empirical analysis in these papers is to find a model that would best fit the data on hand and study the implications on service level measures based on resulting staffing schedules. We discuss the different papers relevant to our study from these two streams of literature in detail below.

In context of retail operations few researches have studied the relationship between store traffic and store operations. Exceptions are Lam et al. (1998) who find that sales-force scheduling decisions based on a time-series model of past traffic leads to better performance. However they ignore the role of traffic variability on these decisions. More recently, Perdikaki et al. (2010) find that store traffic exhibit considerable inter-day and intra-day variability and they identify traffic variability as one of the important factors that affect store sales performance. We contribute to this stream of literature by studying how models that incorporate this variability in traffic can be useful in making store-labor planning decisions.

Agrawal and Smith (1996), using sales data, infer that retail demand exhibits much more variation than can be captured by a Poisson process alone, and that a negative binomial model fits the sales data significantly better than the Poisson or normal distribution. Similarly, Eppen and Iyer (1998) analyze the fit of Poisson, normal and negative binomial distributions for demand data on fashion goods for a big-book catalog retailer and find that the negative binomial distribution provides a better fit to the data. The authors also observe that store merchandisers typically report that variability increases with the level of demand, a finding that is consistent with Hausman (1973). All these papers have looked individual item level demand distributions but not at retail traffic.

In the marketing literature, a negative binomial model has often been used to model aggregate customer purchases (Morrison and Schmittlein, 1988). Here, studies have shown that customers differ in their purchase rates based on differences in need, attitudes and loyalties, and a negative binomial distribution model is able to capture this heterogeneity in customer purchases (Gupta and Morrison, 1990). More recently, Fader et al. (2005) have found that the negative binomial distribution model with slight modifications to have good performance and to be easily implementable in simulation experiments.

A long standing assumption in the literature on management of service facilities is that customers arrive according to a well-understood process. For example, one might assume that the time between arrivals is given by a renewal process, and the customers arrive according to a Poisson Process. Recommendations for managing the service facilities are based on models built with these assumptions. This is still one of the most common methods used in many workforce management solutions for call centers. Recent empirical work in call center settings has revealed several important characteristics underlying the arrival process of telephone calls that cannot be handled by a Poisson process, including time variability and overdispersion. For example, it has been found that arrival rates vary temporally over the course of a day (Tanir and Booth et al 1999) and that the peak-hour arrival rate can be significantly higher than the level of the average daily arrival rate (Brown et al 2005). Similarly, there is also evidence to suggest that arrival counts exhibit variance that substantially dominates the mean value (Avramidis et al 2004). This implies the assumption that the arrival process is Poisson may be invalid. A mechanism that accounts for time varying arrival rate and incorporates over-dispersion was suggested by Jongbloed and Koole (2001). They propose a Poisson mixture model which incorporates a stochastic arrival rate process to generate the additional variability.

The most common application of studying the fit of different demand distributions is to generate accurate forecasts of future demand. Both Agrawal and Smith (1996), and Eppen and Iyer (1998) study the resultant stocking decisions and service level obtained with help of forecast of demand based on the negative binomial distribution. Another application of these models is to use them to develop a predictive distribution of future demand for determining staffing requirements. This entails forecasting both the arrival counts, as well as their distribution that are necessary to plan for staff schedules to ensure adequate customer coverage. Hence, it is important to also understand the variability around the point forecasts. As suggested by Steckley et al (2009), higher variability in demand necessitates higher staffing targets to achieve the targeted service level goals.

Technological advances in recent years have enabled researchers to employ sophisticated techniques to call forecasting and generate agent requirements to meet service level agreements. Although the operation of a call center can be quite different from that of a physical store where customers can observe the staffing level in the stores and accordingly update their decisions, we look at the this literature to understand how a forecast of traffic can be used for generating staffing plans that would meet target service level requirements. In this context, Soyer and Tarimcilar (2008) analyzed the effect of marketing strategies on call arrivals. Their Bayesian analysis is based on the Poisson distribution of arrivals over different time periods measured and conclude that the data cannot be adequately described by assuming a fixed model without some additional random variability source. Taking a non-Bayesian approach, Shen and Huang, (2008) create a prediction model which provides inter-day forecasts and an intra-day updating mechanism for the arrival rate profiles. For a detailed bibliography on the subject of forecasting telephone call arrivals, as well as other call centers related papers, readers are referred to (Gans et al. 2005). We contribute to this stream of literature by showing how similar forecasting approaches can be used in context of a retail store setting to generate staffing levels.

3.3 Models for traffic distribution

In this section, we characterize the traffic distribution by developing and study the fit of statistical models of the arrival process. In particular, we seek to develop models for retail settings that would incorporate traffic variability and may guide the development forecasting models that aid managers in labor planning and scheduling.

There are some practical reasons why store traffic may be more variable than the Poisson. Random variations may occur in the underlying Poisson arrival rate due to the weather and competitors' promotions, or special events that are not captured by the forecasting system. Most forecasting methods use either the historical traffic or sales information to forecast labor requirements (Netessine et al. 2010). Many theoretical papers have also assumed that traffic follows a Poisson or the normal distribution and use these to forecast traffic as they are analytically convenient distributions for modeling store traffic (Agrawal and Smith, 1996; Lariviere and Van Mieghem, 2004). In case of specialty apparel retail, conversion rates tend to be much lower (in the range of 15 to 25%) as compared to grocery stores (where conversion rates are usually 90% and above). In such cases, use of historical sales data for planning purposes could potentially underestimate the true demand potential. Since we have information on actual traffic, we can use this information directly for estimating the demand parameters. Typically store managers determine weekly staffing schedule one to two weeks in advance (RedPrairie, 2010). A key input to this staffing schedule is a forecast of distribution of traffic for that particular week. Thus, a model that incorporates the high variability should help retailers generate better traffic forecasts and consequently plan labor that more closely matches the demand and obtain higher service levels.

Next, we explain each of the models that are used in our analysis for estimating the hourly distribution of traffic below. We consider three models based on prior literature: Modeling traffic distribution by a Poisson distribution, a negative binomial distribution and a normal distribution. Let N be the random traffic (or demand or customer arrivals) per hour on a given day at a particular store, μ the true mean of the traffic distribution and σ the true standard deviation of the traffic distribution.

3.3.1 Model traffic with Poisson distribution

The Poisson distribution, arising from the assumption of independent random arrivals at a steady rate, is often used to describe retail traffic. Assume arrivals follow a Poisson process with a random arrival rate function μ. The probability distribution function is expressed as:

$$P_p\{N = y\} = e^{-\mu} (\mu)^y / y!, \; y = 0,1 \ldots$$

with mean = variance = μ. The key drawback of the Poisson process is that variance is equal to mean, an assumption that is violated in many practical applications. Next, we look at models that allow us to account for greater variability in demand data.

3.3.2 Model traffic with negative binomial distribution

A negative binomial model is capable of capturing higher variation than the Poisson model. This model also has an intuitive meaning in the retailing context. If we model the customer arrivals as a Poisson process, but with random arrival rates, where the randomness is modeled as a gamma distribution, then the resulting distribution is a negative binomial distribution. The randomness in arrival rates is modeled as follows:

$$P_{NB}\{N = y|r, p\} = \int_0^\infty P_P\{N = y; \mu\} g(\mu; r, p) d\mu \tag{3.1}$$

Where the specific functional forms are

$$P_P\{N = y|\mu\} = e^{-\mu} (\mu)^y / y!, \mathrm{y} = 0, 1, 2 \dots,$$

$$g(\mu|r, p) = p^r \mu^{r-1} e^{-p\mu} / \Gamma(r),\ \mu > 0$$

And,

$$P_{NB}\{N = y|r, p\} = \binom{r + y - 1}{y} \left(\frac{p}{p+1}\right)^r \left(\frac{1}{p+1}\right)^y, p > 0, y = 0,1 \dots \tag{3.2}$$

Where $E[\mu] = E[N] = r/p$, $Var[\mu] = r/p^2$, and $Var[N] = r/p + r/p^2$

Thus the variance of the negative binomial distribution has two components, the average Poisson variation (i.e. r/p) and the additional variability in mean traffic (r/p^2). A common re-parameterization in many econometric applications is to model the gamma distribution as Gamma $\left({}^{1}/_{\alpha}, \alpha\right)$ (Cameron and Trivedi, 1998). With this parameterization, a Gamma $\left({}^{1}/_{\alpha}, \alpha\right)$ distribution will have expectation of 1 and a variance of α. α is often referred to as the overdispersion parameter; larger the value of α, greater the overdispersion relative to the Poisson model where $\alpha = 0$.

3.3.3 Model traffic with normal distribution

$$P\{N = y\} = \Phi\left(y + \tfrac{1}{2}\middle|\mu, \sigma\right) - \Phi\left(y - \tfrac{1}{2}\middle|\mu, \sigma\right), y = 0,1,2 \dots \tag{3.3}$$

Where $\Phi(y|\mu, \sigma) =$ normal cumulative distribution with mean μ and variance σ^2. When traffic per period is large, the normal distribution is often used because it is a good approximation for the Poisson model in cases of large mean traffic. This assumption has been especially used in modeling item level demand and in the call center literature for high volumes (Agrawal and Smith, 1996; Gans et al. 2005). The normal distribution however, may fit low-traffic periods poorly as it assigns a probability to negative values and must be symmetric about its mean.

In the following sections, we describe our research setup, and the methodology to study the fit of these different models to the traffic data in our study.

3.4 Research Setup

3.4.1 Description of dataset and data variables

We use a combination of proprietary and secondary data sources for our analysis. Next, we present our data sources and then describe the variables used in this study.

We obtain store-level data for a large retail chain "*Alpha*" [1]. *Alpha* is a women's apparel retail chain that sells affordable luxury products. Its target customers are women in the age group of 21-35 years, and its products span career-wear, evening-wear, and casual-wear. Currently, *Alpha* has over 308 stores in 35 states of the United States, Puerto Rico, the United States Virgin Islands, and Canada. The stores belong to four different retail formats with each format having its own line of merchandise. Most of the retailer's stores are located in regional shopping centers and some of them are present in outlet centers/super regional malls and freestanding street locations. The retailer also sells through online and catalog. The study period was from January 1, 2007 to December 31, 2007.

We obtain the traffic data for each of the stores for the year of 2007. The traffic data was collected with help of traffic counters installed at the entrance of the stores to record the number of visitors to the store. Such traffic counters were installed in 60 of its stores located in the United States during our study period. They use an advanced on-board video sensor with high-speed processing that is able to unobtrusively track customers' movements. It is also able to distinguish between incoming and outgoing shopper traffic; count side-by-side traffic and groups of people; and differentiate between adults and children, while not counting shopping carts or strollers. It also can also adjust to differing levels of light in the store; prevent certain types of counting errors; and time-stamp each record that allows detailed data analysis. The advanced traffic counting system at this retailer is guaranteed to have atleast 95% accuracy of performance against real traffic entering and exiting a store and is validated before the data collection process.

[1] The name of the retail chain is disguised to maintain anonymity

Out of the 60 stores, 7 stores were located in outlet centers/super regional malls. We collected the hours of operation for each store from the store website. We supplement our data set with demographic information from U.S. Census Bureau and ESI Mediamark Research. This included information on population density per square mile for that zip code, median household income, median house values, number of competing women's specialty apparel retail stores, total number of retail stores, and average household apparel spending in women's apparel in the age group of 16 years and above. We also collected data on the presence of direct competitors in the malls/shopping centers for the stores in our sample individually from the competitors' websites. The information on direct competition was obtained from Hoover's company analysis accessible on Lexis-Nexis website. The data for each of these variables were collected based on the zip code in which these stores were located. The summary statistics of these demographic variables are given in Table 3.1

Name	Definition	Average	Std Dev	Min	Max
HHI_i	Median House Household Income for the zip code scaled by population density per sq mile	35.27	60.38	1.57	412.87
HHV_i	Median House Values for the zip code scaled by population density per sq mile	79.31	158.93	4.71	1137.58
HHP_i	Median Household Apparel spending for the zip code scaled by population density per sq mile	9.76	19.40	.563	138.52
$Comp_i$	Number of competing retailers in the zip code scaled by population density per sq mile	.004	.012	.001	.089

Table 3.1: Summary statistics of demographic variables

3.4.2 Preliminary data analysis and sample description

In the retailing literature, it is noted that consumers' shopping trips may follow a weekly cycle, and shoppers' arrival rates at a store may differ by day of week and hour of day (Kahn and Schmittlein, 1989; East et al., 1994; Walters and MacKenzie, 1988). In addition, any retail marketing activity such as store promotions also affects the demand at a given time (Walters and Rinne, 1986). We could not obtain *Alpha's* promotional activities for each store. However, retailers typically run promotions in advance of holidays. As store traffic during periods of major holidays as well during the holiday week of thanksgiving and the Christmas holiday season may be driven by shoppers looking for special deals and promotions, retailers may at times extend hours of operations, run special in-store promotions and hire additional part-time labor to meet demand. As we did not have data on these events that could influence variability in store traffic, we exclude traffic data around major holidays, the week of Thanksgiving and the holiday season of December from our sample. List of holidays is given in Table 3.2

Date	Holiday
Monday, January 1	New Year's Day
Monday, January 15	Birthday of Martin Luther King, Jr.
Monday, February 19	Washington's Birthday
Monday, May 28	Memorial Day
Wednesday, July 4	Independence Day
Monday, September 3	Labor Day
Monday, October 8	Columbus Day
Monday, November 12	Veterans Day
Thursday, November 22	Thanksgiving Day
Tuesday, December 25	Christmas Day
Sunday, April 8	Easter
Sunday, May 13	Mother's Day

Table 3.2: List of known holidays

Given the a priori knowledge that the traffic pattern varies substantially across the days of the week, we began our preliminary data analysis with a multivariate analysis of traffic across different days of the week and across different months of the year for each store in our sample. The statistical decision problem is to cluster the different populations corresponding to each day of the week so that different clusters have a different mean vector based on the average hourly traffic (statistically different means). This is also a common approach in call center scheduling and we follow the same procedure as in Avramidis et al (2004).

Across all stores, we observe the following main results: (a) There are three statistically different populations; and (b) the best clustering of the five populations to three clusters is Saturday, Friday and Sunday, and the aggregate Monday/Tuesday/Wednesday/Thursday. For estimation purposes, we create the following two samples, weekdays–representing the aggregate Monday/Tuesday/Wednesday/Thursday and weekends–representing Friday/Saturday/Sunday. In the remainder of the paper, we mention results corresponding to the aggregate population Monday/Tuesday/Wednesday/Thursday, but report results for both weekdays and weekends in the respective tables where applicable.

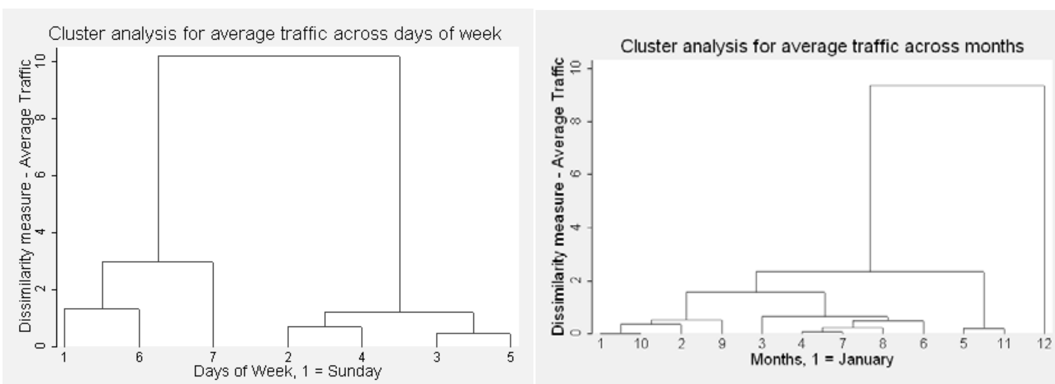


Figure 3.1: Clustering of data

Next, we look at the average number of hourly traffic and variance, and find significant overdispersion in our data. The average traffic across all stores in our sample for weekdays was 48.99 while the variance was 859.07, which indicates that a standard Poisson Process would be a poor fit to the data. Summary statistics are shown in Table 3.3

	Weekdays			Weekends		
	Traffic	Transactions	Sales	Traffic	Transactions	Sales
Average	48.99	8.87	686.11	95.51	12.89	1127.58
Std dev	29.32	4.47	643.56	56.401	6.73	918.64
Min	5.0	2.0	35.18	11.0	3.0	103.25
Max	137.0	36.0	10442.21	230.0	72.0	14123.41

Table 3.3: Summary statistics of variables (value of variable per hour)

3.5 Empirical Analysis

3.5.1 Model Estimation

We are interested in understanding the hourly traffic distribution as it would enable store managers to plan labor to match demand at a more granular level. Hence, our unit of analysis is a “store-hour” i.e. we estimate the three models for each store, for each operating hour. In our dataset, the peak hour traffic was observed to be atleast three times higher than average traffic. Hence, conducting the analysis at the “store-hour” level allows us to incorporate variation in rate of traffic from hour to hour. This approach is similar to Jongbloed and Koole (2001) and is particularly attractive in our context as with use of part-time workers and shift schedules, store managers typically have some flexibility to change the number of associates on an hourly basis (Maher, 2007).

We construct our control variables based on both data availability and factors that have been known to influence customer purchases from prior literature (Gupta and Morrison 1991; Lam et al. 2001). It is possible that traffic to a retail store is influenced by certain long term trends like economic conditions, changes in consumers’ willingness to spend, weather changes, special events and promotions, and so on. Strong correlation in store traffic across different days of the week and across different hours of the day has also observed in practice (Perdikaki et al. 2010, Netessine et al. 2010). We control for these effects in the following manner. To capture correlation across days of the week, we use the lagged values of traffic for the same hour from prior days. The lag length is incorporated by analyzing the correlation in traffic for the same hour across different days. We estimate the lag length for weekdays to be 3, while for weekends to be 2. We follow a similar approach to capture correlation across different hours of the day by using lagged values of traffic for the same day, the lag length in this case being 2 hours. These lag lengths were verified by looking at the Akaike Information Criterion (AIC) and Schwartz Criterion (BIC) for different lag specifications. Using the prior traffic from the day also helps us to account for any day-specific effects like weather changes (e.g. changes in temperature, humidity etc.). Finally, we also control for day of week and month of year effects through use of monthly and day of week dummies. Since the model development and analysis for weekdays and weekends are similar, we explain the details with respect to the weekday sample.

Regression Models

We obtain the following model for Poisson regression

$$y_{idh} \sim \text{Poisson}\ (\mu_{idh}), \mu_{idh} = \exp(\boldsymbol{x}'_{\boldsymbol{idh}}\boldsymbol{\beta}) \tag{3.4}$$

For the observed counts y_{idh} with covariates $\mathbf{x}'_{\mathbf{idh}}$.

Here, $\mathbf{x}'_{\mathbf{idh}} = \{1, \sum_{l=1}^{3} y_{i,d-l,h,} \sum_{l=1}^{2} y_{id,h-l,} a_{dw}, a_{dm}\}$ where and a_{dw} and a_{dm} capture the day-of-week and month fixed effects respectively.

The negative binomial regression models the number of occurrences (counts) of an event when the event has extra-Poisson variation, that is, when it has overdispersion. This derivation can be obtained by assuming that individual units follow a Poisson regression model, but there is uncertainty in the rate at which they arrive which can be modeled as a gamma distribution.

For ease of exposition, let t represent an observation for store i on day d and hour h. Then,

$$y_t \sim \text{Poisson}\ (\mu_t^*), \quad \mu_t^* = \exp(\mathbf{x_t}\boldsymbol{\beta} + \vartheta_t), \quad e^{\vartheta_t} \sim \text{Gamma}\ \left({}^{1}/_{\alpha_t}, \alpha_t\right) \tag{3.5}$$

The dispersion of the t^{th} observation is given by: $1 + \alpha_t \exp\ (\mathbf{x_t}\boldsymbol{\beta})$. The variance of y_t can be derived as follows:

$$\mu_t^* = \text{Gamma}\left({}^{1}/_{\alpha_t}, \alpha_t\mu_t\right) \tag{3.6}$$

$$\mathrm{Var}(y_t) = E\{\mathrm{Var}(y_t|\mu_t^*)\} + \mathrm{Var}\{E(y_t|\mu_t^*)\} \quad (3.7)$$

$$= E(\mu_t^*) + \mathrm{Var}(\mu_t^*) = \mu_t(1 + \alpha_t\mu_t) \quad (3.8)$$

The normal regression model for a generalized linear model is expressed as

$$E(y_{idh}) = \boldsymbol{x}'_{\boldsymbol{idh}}\boldsymbol{\beta}\,,\ y_{idh} \sim Normal \quad (3.9)$$

The estimation of mean and variance for each of the above models is done through maximum-likelihood estimation (MLE) using a Newton-Raphson algorithm. In general, MLE selects values of the model parameters that produce a distribution that gives the observed data the greatest probability (i.e. parameters that maximize the log-likelihood function).

This estimation approach is well defined in case of normal and the linear-exponential family of distributions. MLE possesses many of the desirable asymptotic properties including consistency, asymptotic normality and efficiency. Details on maximum-likelihood estimation for count data can be found in Cameron and Trivedi (1998) and Greene (2003).

3.5.2 Testing for quality of fit

We compared the quality of fit based on each of the above models as follows. First, we employed goodness-of-fit tests based on Pearson's statistic that uses the estimated mean and variance to compute the residuals. We also verified the same with the deviance statistic and the chi-square statistic that look at different measures of model fit. The deviance statistic captures the difference between the fitted log-likelihood and the maximum log-likelihood achievable with the data while chi-square test compares the fitted probabilities with actual frequencies.

We have 600 store-hours in our weekday and weekend samples. We find that the fit of the negative binomial model to actual traffic data was not rejected ($p > 0.23$) for both the weekday and the weekend sample. We find qualitatively similar results for the weekend sample and hence describe our main results for the weekday sample. The normal distribution model was also not rejected in 65% of the cases ($p > 0.15$) while the Poisson distribution model was rejected in 55% of the cases (p< 0.1). We also found the overdispersion parameter (α_{ih}) to be significantly greater than zero in 67% of the instances.

Next, we compared the performance of these models based on the likelihood ratio (LR) test that allows for discrimination among non-nested models based on the log-likelihood values (Cameron and Trivedi, 1998). Based on the likelihood-ratio test, we find that negative binomial distribution performs substantially better than the Poisson distribution for 45 out of the 60 stores ($p < 0.1$). For the remaining 15 stores, the fit from a negative binomial distribution and Poisson distribution were comparable. In all instances, the fit from the negative binomial distribution dominated the fit from the normal distribution ($p < 0.1$).

In general, we find that Poisson distribution tends to understate the traffic in the right tail of the distribution, and although the normal distribution accommodates greater variation than the Poisson distribution, the mode of the normal distribution is shifted to the right, and does poorly during times when there is low traffic volume. Finally, since we expect the log-likelihood to increase as parameters are added to the model (both the negative binomial model and normal model use an additional parameter for computing variance) we compare the models based on Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values. Both these criteria penalize models for additional parameters. We obtained similar results with these tests as well.

Figure 3.2 compares the distribution models based on the empirical cdf from the predicted values obtained from the fitted distributions of the negative binomial model, the Poisson model and normal model, and the actual traffic for a representative store.

Figure 3.2 Comparing distribution models to data based on empirical cdf from predicted values from models

3.6 Relationship between overdispersion in traffic and heterogeneity in market characteristics

Retailers typically recognize that every store is unique and different; consequently, they will have unique traffic patterns. Interestingly, even stores in the same chain and same geographic market show substantial variations in their traffic patterns. Hence it is of interest to retailers to know if there are any systematic factors that may explain this variation in traffic based on their local market characteristics (Ryski, 2005).

Prior theoretical literature on strategic customers (Lariviere and Van Mieghem, 2004) has posited that the arrival process tends to approach a discrete-time Poisson process under competitive conditions. Hence, we investigate if the presence of competition might impact the different values of overdispersion, or greater than Poisson variation in traffic, amongst the different stores in our setting as well. Many industry reports list the direct competitors for different companies based on their product offerings and target consumers. We picked the 15 direct competitors as listed in Hoover's company profiles. Next, we obtained store location information from each of these retailer's websites to calculate the number of direct competitors present in the same malls/shopping centers as the stores in our sample.

Several studies in marketing literature have also explored effects of market and socio-economic characteristics on store performance. For example Reinartz and Kumar (1999) suggest that customers with higher socio-economic status tend to prefer shopping in specialty shops, that these customers are less deal prone, are more regular in their store visits and tend to value service more than other customers. Following this literature, we use average dollar spend on women's clothing (in the age group of 16 and above), median household income and median house values as control variables to capture the socio-economic status of consumers. These variables are scaled by the average population density for that location. In our data set we also have 7 stores located in outlet centers and super-regional malls. It is possible that customers may travel to these outlet centers and malls from places other than the zip code in which these centers are located. To control for these effects, we use an indicator variable that is coded as one if the store is located in an outlet center or regional mall. In addition, we also used the average dollar value of store sales to control for store size.

We use the average overdispersion (α_i) for each store to capture the greater than Poisson variation in traffic across the different stores. To obtain this we run the regression in equation 3.5 for each store i, and control for store-hour fixed effects in addition to the existing exogenous variables. The average α_i across all stores was found to be 0.013 (0.042), the standard deviation 0.018 (0.027) and the minimum and maximum values were 0.002 (0.02) and 0.06 (0.10) respectively for weekdays (weekends). The individual overdispersion values for each of the stores in our sample are given in Appendix 6.2.1.

Next, we run the following cross-sectional regression of α_i against the market characteristics.

$$\alpha_i = \beta_0{}^c + \beta_1{}^c Comp_i + \beta_2{}^c HHP_i + \beta_3{}^c HHI_i + \beta_4{}^c HHV_i + \beta_5{}^c Store\_Sales_i + \beta_6{}^c Outlet_i + \varepsilon_i{}^c \qquad (3.10)$$

Table 3.4 presents the results of this regression. We find $\beta_1 < 0$, (p<0.05) which indicates that higher competition has a negative impact on overdispersion.

As a robustness test, we also used the number of women's clothing stores present in the same location as an alternate measure of competition and find qualitatively similar results. We also find that stores located in outlets/super regional malls tend to have higher variation $\beta_6 > 0$, (p<0.05) as compared to stores located in city shopping centers.

These results indicate that there is heterogeneity in traffic patterns even for stores within the same retail chain based on their local market characteristics.

Market/Store characteristic	Overdispersion (α_i)	
	Weekdays	Weekends
Intercept	.036** (1.01e-02)	.039** (3.01e-02)
Direct competition$_i$	-.082** (2.17e-02)	-.125** (6.12e-02)
Median HH Income$_i$	-.002* (9.48e-04)	-.003* (7.58-04)
Median HH Value$_i$	-0.001* (4.98-04)	-0.001* (4.95-04)
Median HH Apparel Spending$_i$	-.004* (2.01e-03)	-.005* (2.51-03)
Outlet/super regional mall (dummy)	.052** (9.94 e-02)	.065** (11.47 e-02)
Store Size ($Sales)$_i$	0.002 (9.80 e-04)	0.001 (2.80e-04)
R^2	0.38	0.35
n	60	60

Table 3.4 Relationship between variation in traffic and market characteristics

3.7 Application to retail labor planning

In this section we look at some of the practical applications of knowledge of traffic distribution for retail operations. While many retailers have installed traffic counters primarily to assess the closing ratio/conversion rate for their stores and to get an idea of lost sales potential, new technologies are now emerging that aim to integrate the traffic or demand information into workforce management solutions (Tomax, 2009; Stores, 2010). Hence, we investigate if the knowledge on traffic distribution can help retailers in their labor planning decisions through use of traffic forecasts.

3.7.1 Generation of traffic forecasts

We use the months of Jan – June as our fit sample to calibrate our forecast model. We generate one and two week-ahead forecasts on a rolling horizon basis for the test sample of months from July to Nov. To prevent any look-ahead in our data, we only use prior traffic values and control for day-of-week effects in generating the forecasts.

For each store-hour combination, to generate one week ahead forecasts, we run the following regression from equation 3.5:

$$y_{idh} \sim \text{Poisson}\,(\mu_{idh}), \mu_{idh} = \exp\,(\mathbf{x}'_{\mathbf{idh}}\boldsymbol{\beta})$$

where, $\mathbf{x}'_{\mathbf{idh}} = \{1, \sum_{l=7}^{10} y_{i,d-l,h,}\, a_{dw}\}$. The forecast for $\hat{y}_{id+7,h}$ is given by exp $(\mathbf{x}'_{\mathbf{idh}}\hat{\boldsymbol{\beta}})$

For e.g., we use data available till Jun 25th, 2007 to generate one week ahead forecasts for July 2nd, 2007 and the two week ahead forecasts for July 9th, 2007. Similar process was followed for each of the subsequent weeks in the months of July to November. Retailers may actually have access to monthly seasonal factors from prior historical data, but since we did not have this information, we have adopted a conservative approach in generating forecasts based on only information that is available at that point in time. Similar approach is followed for generating forecasts based on a normal distribution. We then obtain the 95% confidence interval around these point forecasts.

Staffing decisions, aimed at guaranteeing minimum service level or traffic coverage, usually require distribution of forecasts and prediction intervals to understand the variability around forecasts. In order to do this, we first generate one week and two week-ahead forecasts based on the three models. We compare the predicted confidence intervals with the actual traffic for that time period to determine the forecast accuracy i.e. for store i and time period t, we compute the coverage probability and the average forecast interval width as follows:

$$COVER_{im} = \left[\frac{1}{T}\sum_{t=1}^{T} I\left(\widehat{N}_t^{2.5} < N_t < \widehat{N}_t^{97.5}\right)\right] \; WIDTH_{im} = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{N}_t^{97.5} - \widehat{N}_t^{2.5}\right)\right]$$

Where m is the model and T the total number of hourly observations (t) for store i.

A good quality forecast would be a forecast that would yield a high coverage with low width (Chatfield, 2000; Shen and Huang, 2008). A low width is attractive for two reasons. First, if we assume that the number of people on staff cannot be adapted to changing conditions, i.e. we cannot increase and decrease the number of people required based on changes in traffic, then, in order to attain a target level of service coverage, this would entail planning for the worst-case scenario. By looking at a prediction interval for the arrival rate, we can give a stochastic guarantee for the coverage that in 95% of the cases, the targeted coverage would be achieved. This is based on using the upper bound of the prediction interval for planning purposes. Second, if we assume that the workforce can be adapted in a flexible way, for example by having flexible contracts that allow the store managers to call for extra personnel when needed, then the lower bound can be used to calculate the fixed number of store associates needed while the difference between the upper and lower bounds would give an idea of the number of store associates that may need to be called in on short notice.

In reality, with shift scheduling constraints and part-time workers, the store managers might face a mix of both these options while making staffing decisions. Some workforce management solution providers (e.g. Red Prairie's enhanced forecasting solution) have recognized the need to closely tie in their scheduling systems with demand information. A popular approach in these software applications for forecasting traffic is the use of a time series method based on prior traffic (Netessine et al. 2010; Tomax, 2009). Hence, we also generate one and two week-ahead time-series forecasts based on traffic from a Newey-West regression as a benchmark for comparison.

We find that, on average, across all stores, the negative binomial model outperforms all other models. For example, in case on one week-ahead forecasts, the coverage probability obtained from a negative binomial model is 0.67 which is higher than that obtained from the other models (as shown in Table 3.5). The Poisson model provides coverage of only 0.45. We present results from here-on for our weekday sample but find qualitatively similar results for the weekend sample as well. The corresponding tables are available in the Appendix 6.2.

We performed a proportion test to test for equality of coverage probabilities provided by each of the models and found that the negative binomial model provided greater coverage for both weekdays and weekends (p<0.05)[2].

	Weekdays					
	One week ahead			Two week ahead		
Model	Coverage (p)	Width [a] (%)	Accuracy	Coverage (p)	Width [a] (%)	Accuracy
Poisson	0.45	17.14	2.60	0.41	18.23	2.25
NB	0.67	20.64	3.22	0.61	22.65	2.69
Normal	0.61	26.11	2.34	0.58	28.76	2.02
Time series (traffic)	0.54	17.66	3.03	0.52	19.67	2.64

[a] Expressed as a percentage of actual traffic

Table 3.5: Forecast accuracy for weekdays

[2] Let p_{itm} be recorded as "1" if the true traffic falls within the forecast interval, 0 o/w. Let $\Delta p_{it} = p_{it,NB} - p_{it,m}$ The proportion test is given as: $H_0 : \Delta p_{it} == 0$, $H_a : \Delta p_{it} > 0$

We also look at an accuracy measure: $Accruacy_{im} = Coverage_{im}/Width_{im}$, to test for a model that gives better coverage as well as lower width.

In order to test if monthly effects would change our results, we conducted two additional robustness tests. First, we generated one week-ahead forecasts including month dummies in the regression model for only the 2nd, 3rd and 4th weeks of the month. The results are as shown in Table 3.6. We find that including the monthly dummies improves the forecasts, but our results on model fit remain qualitatively unchanged, i.e. we still find that the forecasts based on the negative binomial model to have higher coverage and lower width than other forecasts based on other models. Second, we use the prior monthly index of consumer sentiment (ICS) obtained from Thomson Reuters/University of Michigan survey of consumers as this has shown to be a leading indicator of consumer expectations of purchases (Howrey, 2001) and again find qualitatively similar results.

	Weekdays					
	One week ahead			Two week ahead		
Model	Coverage (p)	Width [a] (%)	Accuracy	Coverage (p)	Width [a] (%)	Accuracy
Poisson	0.48	16.38	2.93	0.44	17.14	2.57
NB	0.74	18.47	4.01	0.71	20.58	3.45
Normal	0.65	25.47	2.55	0.60	26.54	2.26
Time series (traffic)	0.57	16.78	3.40	0.55	18.49	2.97

[a] Expressed as a percentage of actual traffic

Table 3.6: Forecast accuracy for weekdays with seasonality factors

Our findings have several practical implications for retailers as forecasts of traffic are used in making assortment decisions, planning promotions and deciding labor requirements. We find that a negative binomial model that accounts for variability in traffic to provide much better forecasts than other models.

3.7.2 Calculation of labor based on service level considerations

Next, we want to determine the impact of using the right distribution of traffic on the service level provided by the retailer by using the forecast of traffic in planning for labor. Traditional measures of service level include the time customers spend in waiting for service, queue length, availability of product etc. However, in case of specialty apparel retail, it is hard to define and measure service level due to the varying requirements of different customers and complexity of the buying decision process. In particular, it has been observed that sales associate availability and help are critical to the buying process (and resulting conversion), but there exists very limited evidence quantifying what constitutes a good or acceptable level of service. For example, good service could entail, among other factors, availability of product, knowledge of store associates, ambience, ease of transaction etc, some of which are very qualitative in nature and differ from customer to customer.

With the availability of traffic data, one of the measures that retailers now monitor is the ratio of customer per staff hour (CSR) for the store to see if they have adequate service availability or staffing coverage for the store (NRF, 2010). For example, traffic of 100 people with a staff of 10 generates a ratio of 10:1. As the ratio increases, service availability declines.

Retailers use the CSR to plan for labor in the following manner. First, depending on the kind of product and the level of sales assistance required by customers, retailers formulate general guidelines for the ideal CSR to be maintained by different stores (Ryski, 2005). Based on each store's individual local and market characteristics, store managers may modify the CSR for their individual stores. Second, the store managers use this CSR and forecast of traffic to plan for the required staffing levels one to two week in advance. This is typically done with help of various workforce management tools. In recent years, these tools have evolved from using just the point forecast of traffic as input for labor planning to analyzing periods of peak traffic separately and planning for them accordingly. The increase in sophistication in these systems is very similar to that seen in call center workforce management solutions which typically consider the 90th (and above) percentile of the traffic distribution of arrivals to make staffing decisions. Thus, we would expect that retailers who value service more, or aim to provide higher service coverage, would thus benefit more from knowing the distribution of traffic and using it in making their labor planning decisions.

We demonstrate the value of using a negative binomial distribution of traffic in labor planning in the following manner. First, we assume that a retailer aims to maintain a planned CSR of $r{:}\,1$, i.e. maintain a staffing level such that there is atleast 1 sales assistant for every r customers . Second, we assume that the retailer aims to provide service coverage of p%, i.e. we use the upper prediction interval based on this percentage for the distribution based on each of the models as the traffic forecast $\widehat{N}_{itm}$. Third, we calculate the labor forecast for each store as $\hat{l}_{itm} = \widehat{N}_{itm}/r$. Finally, we compute the resultant CSR based on actual traffic as $r_{itm} = N_{it}/\hat{l}_{itm}$. For numerical analysis, we assume r = 5 and p = 95%. We perform a sensitivity analysis with different values of r and p in §3.7.3. In addition, we also compare the performance of these models that are based on a forecast of traffic. We do this as this has also been observed as one of the common approaches to labor planning in the retail industry (RedPrairie, 2010).

We compare the deviation between the actual CSR and planned CSR (5:1) obtained from these different models in Table 3.7. The negative binomial model yields the lowest deviation (11.2%) from the planned CSR while the Poisson model (23.8%) tends to significantly underestimate the staffing requirements.

	Weekdays		Weekends	
Model	One Week ahead	Two Week ahead	One Week ahead	Two Week ahead
NB	11.2	12.2	35.6	43.6
Normal	12.8	16.7	52.8	56.2
Time Series (traffic)	17.2	24.5	59.6	65.2
Poisson	23.8	33.6	63.4	74.2

Table 3.7: % Deviation of actual CSR for different models from planned CSR of 5:1

3.7.3 Sensitivity analysis

In this section, we look at sensitivity of the forecasts when planning for different levels of CSR (r) and service coverage (p). First we look at the sensitivity to different levels of CSR. Note that a higher CSR would indicate lower staffing coverage for customers. We find that as the value of CSR increases (i.e. required service availability decreases) the percentage deviation between the planned and actual CSR decreases due to lower service level requirements

The percentage deviation of actual CSR from planned CSR for different values of CSR for weekdays is as shown in Table 3.8, where we compare the performance of models based on different values of CSR for service coverage of 95%. We see that as the requirements for service availability increase, (i.e. we move from a CSR of 20:1 to 5:1), the percentage of gap between the desired CSR and the actual CSR increases, as well as the gap in performance of the different models also increases. For example, with a CSR of 20:1, the percentage gap between expected and actual service coverage from a negative binomial model and poisson model are 5.5% and 7.6% respectively. On the other hand, for a CSR of 5:1, this percentage gap between expected and actual service coverage increases to 11.2% and 23.8% respectively.

	One Week ahead			
Planned CSR	NB	Normal	Time Series - Traffic	Poisson
5	11.2	12.8	17.2	23.8
10	7.1	8.6	9.5	10.7
15	6.2	7.3	8.3	8.8
20	5.5	6.2	6.9	7.6

	Two week ahead			
Planned CSR	NB	Normal	Time Series - Traffic	Poisson
5	12.2	16.7	24.5	33.6
10	8.3	9.6	10.8	12.6
15	6.3	7.9	9.2	10.1
20	5.6	6.5	7.1	8.5

Table 3.8: Sensitivity analysis of percentage deviation of actual CSR from planned CSR for different values of CSR for weekdays

Second we look at the sensitivity to different levels of service coverage (p). Note that higher service coverage would mean using a higher percentile of the distribution as the forecast of traffic. The percentage deviation of actual CSR from planned CSR for different values of service coverage for weekdays is as shown in Table 3.9. As one would expect, for a given level of CSR (r), as the value of service coverage (p) increases; the percentage deviation of actual CSR from the planned CSR decreases. For example, with targeted service coverage of 90%, the percentage gap between expected and actual CSR from a negative binomial model and a Poisson model are 26.8% and 35.4% respectively. As the targeted service coverage increases to 95%, this percentage gap between expected and actual CSR increases to 11.2% and 23.8% respectively.

	One Week ahead			
Planned Service coverage	NB	Normal	Time Series - Traffic	Poisson
90	26.8	28.2	31.5	35.4
93	17.4	18.9	24.2	30.2
95	11.2	12.8	17.2	23.8
97	6.4	8.6	12.4	21.3
99	4.4	7.0	11.2	20.5

	Two week ahead			
Planned Service coverage	NB	Normal	Time Series - Traffic	Poisson
90	29.0	30.8	33.8	37.8
93	20.2	22.6	27.0	35.4
95	12.2	16.7	24.5	33.6
97	11.0	16.3	23.6	29.6
99	8.6	14.1	21.8	25.9

Table 3.9: Sensitivity analysis of percentage deviation of actual CSR from planned CSR for different values of service coverage for weekdays

These results have practical implications for specialty retailers who operate in service intensive environments. Having a poor model of traffic distribution could yield higher than anticipated congestion level in stores that could in turn yield to loss of revenue from dis-satisfied customers.

3.8 Conclusion

In this chapter we have characterized the distribution of traffic to retail stores for a heterogeneous group of stores belonging to the same retail chain. We find a negative binomial distribution to be a significantly better fit to the traffic data than the Poisson or normal distribution. This choice of the negative binomial distribution is supported by individual goodness-of-fit tests for testing model fit and likelihood ratio tests that demonstrates that the negative binomial distribution is far superior to other distribution choices. Furthermore, we show that the high variability associated with retail traffic can lead to poor performance of some of the commonly assumed models for retail traffic like a Poisson or a normal distribution and time series models based on historical sales.

We demonstrate the application of knowledge of traffic distribution to labor planning by generating traffic forecasts based on each of these distributions, and using these traffic forecasts to generate staffing plans to meet a targeted service level. In our study, the forecasts from a negative binomial model was able to provide service levels closer to the targeted service level as compared to forecasts from other models. A recent survey showed that customers buying luxury goods typically rank service-related attributes as the basis for deciding where to shop (Booz and Hamilton, 2008). In fact almost 33% of customers state poor sales assistance as reasons for leaving without purchase (Baker, 2010). Thus, having the right forecast model that provides the desired service level is critical to store operations for these retailers and can help prevent systemic understaffing during peak hours.

Finally, our results also provide support to some of the observations in theoretical literature on the effects of competition and other local market characteristics on traffic variability. Future research could conduct a more in-depth analysis on the conditions under which these factors influence variability and help retailers in leveraging this information in their planning strategies.

CHAPTER 4

The relationship between abnormal inventory growth and future earnings for U.S. public retailers

4.1 Introduction

Retailers pay close attention to inventory growth in their stores as it can have a significant impact on their future financial performance. Too much of inventory in their stores could result in future markdowns while too little inventory could result in lower demand in the future due to customer dissatisfaction with poor service levels. Numerous anecdotes of poor inventory management leading to decline in financial performance of retailers can be found in the business press. However, there is little empirical evidence on the relationship between current inventory levels and future financial performance of retailers.

In fact, there is growing evidence that even Wall Street investors may have trouble understanding the relationship between inventory levels and future financial performance of retailers. Kesavan et al. (2010) find that even though inventory contains useful information to predict sales for retailers, Wall street analysts fail to incorporate this information in their sales forecasts. Hendricks and Singhal (2009), who examine excess inventory announcements of firms from multiple industry sectors including retail, find that these announcements are associated with negative stock market reactions in a vast majority of those cases. Since excess inventory would get reported only when such inventory problems become large enough, their results suggest that the stock market investors failed to anticipate these announcements even though they had access to past inventory levels of those firms.

In this chapter, we are interested in examining the relationship between inventory and one-year ahead earnings per share. We choose earnings per share because of the following reasons. First, earnings per share is an important financial metric for firms and their forecasts form a key input to investment decisions. Givoly and Lakonishok (1984) find that "*earnings per share emerges from various studies as the single most important accounting variable in the eyes of investors and the one that possesses the greatest information content of any array of accounting variables*." Second, current evidence on the relationship between inventory and one-year ahead earnings[1] for retailers is weak. Accounting literature that examined this question has yielded a mixed response. Abarbanell and Bushee (1997) do not find evidence of this relationship for retailers but Bernard and Noel (1991) do. Even Bernard and Noel (1991), who find inventory predicts earnings for retailers, assume a linear relationship between inventory and earnings and find evidence for the same. Since earnings are a measure of profitability of the firm, one might expect the relationship to be an inverted-U based on the operations management literature. This raises the additional question of whether the inverted-U relationship which forms the building block of inventory models at the SKU-level can be lost at the firm-level?

There are several challenges in testing the relationship between inventory and earnings at the firm-level. First, raw inventory levels cannot be used to determine the relationship since it is correlated with number of stores, sales etc. For example, inventory for a retailer could have grown either due to presence of stale inventory or as a result of opening new stores. While the former would be associated with lower earnings in the future, the latter would not. So, an appropriate method for normalizing inventory is required before we test the relationship between inventory and earnings. Second, service level information of retailers is not publicly available. So, it is difficult to figure out whether a retailer's inventory level is high because it is carrying

---

[1] We use earnings and earnings per share interchangeably

excess inventory or if it is providing a high service level (Lai 2006). The former would be a negative signal of future earnings but the latter would not.

We normalize inventory levels using the expectation model from Kesavan et al. (2010) to obtain the expected inventory growth. Then we calculate abnormal inventory growth as the deviation of actual inventory growth from expected inventory growth and use it as the benchmarking metric to investigate the relationship between inventory and one-year ahead earnings. We investigate the economic significance of the information content in abnormal inventory growth by examining if equity analysts' earnings forecasts incorporate information contained in abnormal inventory growth and test if an investment strategy based on abnormal inventory growth would yield significant abnormal returns.

We use quarterly and annual financial data along with comparable store sales, total number of stores and earnings per share for a large cross-section of U.S. retailers listed on NYSE, AMEX, or NASDAQ from Standard & Poor's Compustat database for our analysis. Equity analysts' earnings forecasts are collected from Institutional Brokers Estimates System (I/B/E/S). Stock returns inclusive of dividends are obtained from CRSP. These are supplemented with hand-collected data from financial statements. Our study period is fiscal years 1993-2009.

We have the following results. First, we demonstrate an inverted-U relationship between abnormal inventory growth and one-year ahead earnings. Our results are robust to the metric used to measure abnormal inventory growth. Second, we find that equity analysts do not fully incorporate the information contained in past inventory resulting in systematic bias in their earnings forecasts; this bias is predicted by previous year's abnormal inventory growth. Third, we find that an investment strategy based on abnormal inventory growth yields significant abnormal returns.

Our analysis is closest to Kesavan et al. (2010) who study if inventory can be used to predict future sales in the retail industry. They find that incorporating inventory and margin information significantly improves sales forecasts. Further they find that analysts do not fully incorporate this information resulting in predictable biases in their sales forecasts. We add to their findings by showing that inventory also contains information useful to predict earnings in the retail industry. Since earnings are a function of sales and expenses, we run several tests to show that inventory predicts earnings not only because it predicts sales but also because it predicts expenses for a retailer. Similarly we show that bias in analysts' earnings forecasts arises not only because analysts ignore information in inventory useful to predict sales, as shown by Kesavan et al. (2010), but also because they fail to consider the impact of inventory on the expenses for retailers. Finally, we analyze stock market data for retailers, not considered in Kesavan et al. (2010).

This study contributes to the operations management literature in the following ways. There is a growing interest among researchers in operations management to examine firm-level inventory (Gaur et al. 2005; Rumyantsev and Netessine 2007; Chen et al. 2005; Rajagopalan 2010). Many of these papers are motivated to develop new benchmarking metrics that are useful to gauge the inventory performance at the firm-level. Our study complements this line of research by demonstrating that such benchmarking metrics possess information useful to predict earnings and serve as a basis for investment strategies. In addition, the research on firm-level inventory has sought to examine if the insights from the analytical models also hold at the firm-level. Rumyantsev and Netessine (2007), for example, argue that this is important to perform such tests to demonstrate to the high-level managers who deal with firm-level inventory that they may benefit from understanding classical inventory models. Ours is the first study to demonstrate that the inverted-U relationship between inventory and profits, that forms the fundamental building block of SKU-level literature, holds at the firm-level as well.

This chapter is organized as follows. In §4.2 we discuss the operations management literature and accounting literature that relates to our work. In §4.3 we discuss existing theory in operations management to argue why changes in inventory levels could be considered as a signal of future earnings, §4.4 outlines our research setup and §4.5 describes the methodology we adopt to calculate abnormal inventory growth. In §4.6, we report results showing the relationship between abnormal inventory growth and one-year ahead earnings while §4.7 investigates the economic significance of ignoring information contained in abnormal inventory growth. Finally, we conclude with limitations and directions for future research in §4.8.

4.2 Literature Review

There has been significant interest in developing benchmarking metrics for firm-level inventory performance. Gaur, Fisher and Raman (2005) study inventory turns and develop a metric, *adjusted inventory turns*, to compare inventory productivity across firms. Rumyantsev and Netessine (2007) show that increase in demand uncertainty, lead times, and margins, and decrease in economies of scale are associated with increase in inventory levels. Chen et al. (2007) benchmark inventory performance using a metric called *abnormal days-of-inventory* or *AbI*, which is defined relative to the segment's average days-of-inventory. Rajagopalan (2010) combines primary and secondary data to show that product variety, along with other factors such as gross margin and economies of scale, affects the firm-level inventory carried by retailers. Our work adds to this literature by testing the efficacy of two of those metrics for prediction purposes. There is some evidence linking inventory performance to stock market performance of firms. Chen et al. (2005) and Chen et al. (2007) find correlation between inventory changes and abnormal stock market returns. Hendricks and Singhal (2005) show that announcement of supply chain glitches, which commonly cause inventory problems, are associated with a negative stock market reaction. However, these papers perform ex post facto analysis and hence do not test the informational content in inventory levels for prediction purposes. Our study adds to this literature by showing that inventory-based benchmarking metrics may serve as basis for investments in the stock market.

Next, we would like to briefly review the accounting literature that relates to our work. The accounting literature has shown some mixed evidence of the predictive power of inventory over earnings in the retail sector. Bernard and Noel (1991) find that inventory predicts earnings in the retail industry but Abarbanell and Bushee (1997) do not. We differ from both these papers in methodology as well as contribution. Both Bernard and Noel (1991) and Abarbanell and Bushee (1997) use a simple expectation model of inventory growth based on sales growth. We use a sophisticated expectation model based on operations management literature that not only considers sales growth but also changes in gross margin, store growth, days-payables, and capital investment. Furthermore, accounting literature has typically assumed a linear relationship between inventory and future earnings. We are motivated by theoretical literature in operations management to test an inverted-U relationship between inventory and one-year ahead earnings and find evidence to support this relationship.

In a seminal paper Sloan (1996) shows that stock market misprices accruals, where accruals are defined as changes in working capital. In other words, hedge portfolios formed based on accruals generate significant abnormal stock returns. This was called the *accruals anomaly* as the stock market fails to process publicly available information causing stocks to be mispriced. Thomas and Zhang (2002) decompose accruals into its components and show that most of the predictive power of accruals is generated by the inventory component in accruals. In our analysis, we show that an investment strategy based on abnormal inventory growth would yield significantly higher abnormal returns compared to a strategy based on inventory growth, as defined by Thomas and Zhang (2002). Thus our result shows that a benchmarking metric for inventory performance derived from operations management literature can improve upon simpler metrics for inventory performance and serve as a basis for an investment strategy.

4.3 Can changes in inventory signal future earnings?

Earnings are a summary measure of a firm's financial performance and are widely used to value shares and determine executive compensation. They are a function of the revenue, cost-of-goods sold, interest expenses, income tax, insurance, etc. (Stickney and Weil 2003). The contemporaneous impact of inventory on earnings is well known. The most recognized component of this impact is the holding cost of inventory, which affects both the capital cost of money tied up in inventory and the physical cost of having inventory (warehouse space costs, storage taxes, insurance, rework, breakage, spoilage, etc.). In addition, there are indirect costs associated with inventory that impact a retailer's earnings as well. These include the risks of lower gross margins and inventory write-offs due to stale inventory. The relationship between inventory and future earnings, however, is unclear.

We argue that the relationship between inventory and future earnings arises because inventory contains incremental information useful to predict both demand and expenses for retailers. Changes in inventory level at a retailer contain two signals. First, they indicate whether a retailer's inventory levels have become leaner or bloated. A retailer's inventory level becomes leaner (bloated) if its inventory level decreases (increases) while providing the same service level to its customers. Second, changes to inventory levels indicate whether a retailer's service level has increased or decreased. Because it is not possible to measure service level of retailers based on public financial data (Lai 2006), one cannot tease out these two effects for a given retailer. So, we argue for the implications of both of these effects in an aggregate sample and perform empirical analysis to determine the dominating effect. Admittedly, there are several reasons why changes in inventory levels may not serve as signals of future earnings so we discuss them as well.

Implications of increase in inventory levels for future earnings

First, consider the arguments for the implications of increase in inventory levels on future earnings when the service level remains same or declines. Increase in inventory level for a retailer could signal lower earnings in the future due to impending markdowns that would drive gross margins lower. Such markdown impacts have been well-studied at the SKU-level. Gallego and van Ryzin (1994), who consider dynamic pricing for a seasonal item, show that the optimal price trajectory is decreasing in the stocking quantity. Smith and Achabal (1998) get a similar result from a model with deterministic demand rate that is a multiplicative function of price and stocking quantity. Thus, when retailers' inventory levels become bloated, they may need to mark down some inventory causing their gross margin to decline which would lead to lower earnings. Such markdowns may often be accompanied by increases in advertisement spending that are required to clear such merchandise. These increases in advertisement expenses would further contribute to decrease in future earnings.

Bloated inventory levels can also signal the presence of stale inventory at a retailer. When such stale inventory is salvaged, it can drive earnings lower. Ferguson and Koenigsberg (2007) state that Bloomingdale's Department Store salvages about 9% ($72M) of its women's apparel by selling it to discount retailers for pennies on the dollar in order to make space for new inventory. Raman et al. (2005) discuss the investment strategy of David Berman, which involves identifying retailers who may be carrying stale inventory, because this would lead to lower earnings in the future.

Bloated inventory levels could also be a negative signal of future demand because high inventory levels may hinder the ability of retailers to introduce new products in their stores. Retailers regularly introduce new products to stimulate demand; such new product introductions are often called the life-blood of retailing. For example, Chico's FAS states that maintaining the newness of its merchandise is a critical factor in determining its future success[2]. Retailers' ability to introduce new products depends upon the availability of shelf-space and financial resources. When retailers carry high inventory levels, they are likely to have less shelf-space available for new products; fewer new products depress the demand for the retailer, leading to lower profits.

[2] http://media.corporate-ir.net/media_files/irol/72/72638/Annual_Report/2004AR.pdf

Also, bloated inventory levels may lead to longer cash conversion cycles that may cause cash-flow constraints for the retailer. Carpenter et al. (1998) find that inventory investment of firms decrease when they face financial constraints. Thus, higher inventory levels may result in lower investment in new products that could also depress the demand faced by the retailer, which could reduce earnings.

Bloated inventory levels could also be symptomatic of operational issues at a retailer that may continue into the future, resulting in higher costs due to supply-demand mismatches. Fisher (1997) states that excess inventory at a retailer is the result of supply-demand mismatches and is associated with poor operational performance. Several operational capabilities have been identified with good inventory management. Some of these capabilities are the ability to forecast accurately (Makridakis and Wheelwright 1987), supply chain responsiveness (Fisher 1997), and reduction in information distortion in supply chain (Lee, Padmanabhan, and Wang 1997). Excess inventory could also be the result of supply chain glitches that can cause operational performance to deteriorate (Hendricks and Singhal 2005). The authors also note that operational performance may not recover to its earlier levels even several years after the supply chain glitch. Hence, a high inventory level could signal operational issues at a retailer that could lead to lower future earnings.

Second, consider the implications of an increase in inventory level that is accompanied with an increase in service level for future earnings. Applying the newsvendor logic to the aggregate setting, we argue that the impact of such an increase in inventory level on future profitability would depend upon the trade-off between the benefits of such increased service level versus the costs of carrying the extra inventory. Thus increase in inventory level could be a signal of higher earnings for some retailers while it is likely to be a signal of lower earnings for the others depending on this trade-off.

Implications of decrease in inventory levels for future earnings

Next we present the implications of leaner inventory levels, i.e., decrease in inventory levels without decrease in service levels, for future earnings. Leaner inventory levels will not only enable retailers to reduce inventory holding costs for retailers but also enable retailers to react more quickly to change in demand. When a retailer carries lean inventory levels, it can procure fresh merchandise for its stores that will stimulate demand.

A decrease in inventory level for a retailer may result in a lower service level in some cases. In such cases, the impact on future profitability would depend on the trade-off between the benefit of having lower inventory levels and the cost of decline in service levels. For example, when customers are willing to substitute in the presence of stockouts, retailers may be able to reduce inventory level and service level without hurting profitability. On the other hand, stockouts may cause customers to switch retailers when the competition is intense. Several papers in operations management including Bernstein and Federgruen (2004), Dana (2001), Gans (2002), and Gaur and Park (2007) have developed analytical models of fill-rate strategies when customers switch to competitors when they experience out-of-stocks. Olivares and Cachon (2009) show that automobile dealers increase inventory levels in the face of competition. So, a decrease in inventory levels may be a signal of higher or lower earnings in the future when it is accompanied by decrease in service levels.

Changes in inventory levels are noisy signals of earnings

There are many cases when retailers' inventory levels change due to the normal course of operations and therefore, may not contain any useful information to signal future earnings. We refer to these changes as *normal* changes in inventory. For example, a retailer may open or close stores resulting in higher or lower inventory levels in its chain. In such cases an increase or decrease in inventory level does not provide any incremental information not contained publicly available data on store openings and closings. Many other factors have also been identified in the operations management literature that are correlated with changes in inventory levels. These factors include gross margin, capital intensity, and sales surprise (Gaur et al. 2005); sales growth and size (Gaur and Kesavan 2007); demand uncertainty and lead time (Rumyantsev and Netessine 2007); competition (Olivares and Cachon 2009); and product variety (Rajagopalan 2010). Thus, changes in these underlying factors may be associated with changes in inventory levels; such normal changes in inventory may not contain any incremental information useful to predict future earnings.

To summarize, normal changes in inventory levels would not contain any useful information to predict earnings. However, the changes in inventory level beyond these normal changes may contain useful information to predict earnings. We call these changes as abnormal changes in inventory levels. As we discussed above, abnormal increases and decreases in inventory levels could be positive or negative signals of future earnings. However, due to the convex cost structure of inventory costs (Zipkin, 1986), we expect that extreme increases or decreases in inventory levels to be associated with lower one-year ahead earnings. Hence we expect an inverted-U relationship between abnormal changes in inventory levels and one-year ahead earnings. We test this using empirical analysis.

4.4 Research Setup

4.4.1 Definition of Variables

The following notations are used in this chapter. For retailer *i* in fiscal year *t*, we denote SR_{it} as the total sales revenue, $COGS_{it}$ as the cost of sales, SGA_{it} as the selling, general and administrative expenses, $LIFO_{it}$ as the LIFO reserve, and $RENT_{it1}, RENT_{it2}, \dots . RENT_{it5}$ as the rental commitments for the next five years, ΔCA_{it} as the change in current assets, $\Delta Cash_{it}$ as change in cash/cash equivalents, ΔCL_{it} as change in current liabilities, ΔSTD_{it} as change in debt included in current liabilities, ΔTP_{it} as change in income taxes payable, Dep_{it} as depreciation and amortization expense, AT_{it} as the total assets and N_{it} as the total numbers of stores open for firm *i* at the end of fiscal year *t*. These are obtained from the Compustat Annual Database. For firm *i* in fiscal year *t* and quarter q, we denote PPE_{itq} as the net property, plant and equipment, AP_{itq} as the accounts payable, and I_{itq} as the ending inventory. These are obtained from the Compustat Quarterly Database.

Next we explain the adjustments that we make to the variables. To ensure that all retailers have similar inventory valuations, we add back LIFO reserve to the ending inventory and subtract the annual change in LIFO reserve from the cost of sales. Similarly, the value of PPE could vary depending on the values of capitalized leases and operating leases held by the retailer. Hence, we first compute the present value of rental commitments for the next five years using $RENT_{it1}, RENT_{it2}, \ldots. RENT_{it5}$ and then add it to the PPE to adjust uniformly for operating leases held by a retailer. Here, we use a discount rate of *d* = 8% per year for computing the present value, and also verify our results with *d* = 10%.We normalize some of the above variables by the number of retail stores in order to avoid correlations that could arise due to scale effects caused by an increase or decrease in the size of a firm. We calculate accruals based on Sloan (1996). Refer to Table 1 for the relevant data fields in the Compustat Database. Using these data and adjustments, we calculate the following variables for each firm *i* in fiscal year *t* and fiscal quarter q:

Average cost-of-sales per store: $CS_{it} = [COGS_{it} - LIFO_{it} + LIFO_{it-1}]/N_{it}$

Average inventory per store: $IS_{it} = \left[\frac{1}{4}\sum_{q=1}^{4} I_{itq} - LIFO_{it}\right]/N_{it}$

Gross Margin: $GM_{it} = SR_{it}/[COGS_{it} - LIFO_{it} + LIFO_{it-1}]$

Average SGA per store: $SGAS_{it} = [SGA_{it}]/N_{it}$

Average capital investment per store: $CAPS_{it} = \left[\frac{1}{4}\sum_{q=1}^{4} PPE_{itq} + \sum_{r=1}^{5} PPE_{itq}\frac{RENT_{itr}}{(1+d)^r}\right]/N_{it}$

Store growth: $G_{it} = [N_{it}]/N_{it-1}$

Accounts payable to inventory ratio: $PI_{it} = \left[\sum_{q=1}^{4} AP_{itq}/4\right] / \left[\left(\sum_{q=1}^{4} I_{itq}/4\right) + LIFO_{it}\right]$

Accruals: $Acc_{it} =$

$[(\Delta CA_{it} - \Delta Cash_{it}) - (\Delta CL_{it} - \Delta STD_{it} - \Delta TP_{it}) - Dep_{it}]/[(AT_{it-1} + AT_{it})/2]$

Definition	Variable Name	Database	Field Name
Cost of Sales	$COGS_{it}$	Compustat Annual Updates - Fundamentals Annual	COGS
Ending Inventory	I_{itq}	Compustat Quarterly Updates – Fundamentals Quarterly	INVTQ
LIFO	$LIFO_{it}$	Compustat Annual Updates - Fundamentals Annual	LIFR
Revenue	SR_{it}	Compustat Annual Updates - Fundamentals Annual	SALE
Selling, General and Administrative Expenses	SGA_{it}	Compustat Annual Updates - Fundamentals Annual	XSGA
Number of stores	N_{it}	Compustat – Industry Specific Annual	RTLNSE
Accounts Payable	AP_{itq}	Compustat Quarterly Updates – Fundamentals Quarterly	APQ
Net Property, Plant and Equipment	PPE_{itq}	Compustat Quarterly Updates – Fundamentals Quarterly	PPENTQ
Rental Commitments	$RENT_{it,1\ldots5}$	Compustat Annual Updates - Fundamentals Annual	MRC1….5
Comparable Store Sales Growth	$Comps_{it}$	Compustat – Industry Specific Annual	RTLCS
Earnings Per Share	EPS_{it}	Compustat Annual Updates - Fundamentals Annual	EPSFI
Analyst Forecast for EPS	$Forecast\_EPS_{it}$	IBES – Detailed History Statistics	Estimate - EPS
Analyst Forecast for EPS	$Forecast\_SAL_{it}$	IBES – Detailed History Statistics	Estimate - SAL
Closing Price	P_{it}	Compustat Annual Updates - Fundamentals Annual	PRCC_F
Total Assets	AT_{it}	Compustat Annual Updates - Fundamentals Annual	AT
Current Assets	CA_{it}	Compustat Annual Updates - Fundamentals Annual	ACT
Cash/Cash Equivalents	$Cash_{it}$	Compustat Annual Updates - Fundamentals Annual	CHE
Current Liabilities	CL_{it}	Compustat Annual Updates - Fundamentals Annual	LCT
Debt	STD_{it}	Compustat Annual Updates - Fundamentals Annual	DLC
Depreciation	Dep_{it}	Compustat Annual Updates - Fundamentals Annual	DP
Income Taxes Payable	TP_{it}	Compustat Annual Updates - Fundamentals Annual	TXP

Table 4.1 Data fields for variables (Retailer i, fiscal year t, quarter q)

The variables obtained after taking the logarithm are denoted by their respective lowercase letters, i.e., cs_{it}, is_{it}, gm_{it}, $sgas_{it}$, $caps_{it}$, g_{it}, and pi_{it}. In addition, we refer to comparable store sales as $Comps_{it}$, actual earnings per share as EPS_{it} and closing share price as P_{it}. All three values are obtained from the Compustat Annual Database.

Finally, we also obtained individual sell side analysts' forecasts for Earnings Per share (EPS) and Sales (SAL) from Institutional Brokers Estimate System (I/B/E/S), and stock market returns from CRSP.

4.4.2 Data Description

We start with the entire population of U.S. retailers that have reported at least one year of financial information during the period 1993-2009. The U.S. Department of Commerce classifies the retailers into eight different categories, identified by the two-digit SIC code as follows: *Lumber and other building materials dealers* (SIC: 52); *general merchandise stores* (SIC: 53); *food stores* (SIC: 54); *eating and drinking places* (SIC: 55); *apparel and accessory stores* (SIC: 56); *home furnishing stores* (SIC: 57); *automotive dealers and service stations* (SIC: 58) and *miscellaneous retail* (SIC: 59).

We exclude retailers in the categories *eating and drinking places* and *automotive dealers and service stations* from our study as they contain significant service component to their business. There were 670 retailers that reported at least one year of data to the U.S. Securities and Exchange Commission (SEC) for these years. Since data on the number of stores are sparsely populated in Compustat for the period prior to 1999, we obtain store data from Compustat starting from the year 1999 and supplement them with hand collected data for the period before 1999. We find that 208 retailers did not report any store information. To enable us to perform a longitudinal analysis, we only consider retailers that had atleast five years of consecutive data. After removing several observations that had missing data for the variables required for our analysis, we find that

369 of the 462 were left for further analysis. Further we eliminated foreign retailers that are listed as *American Depository Receipt* (ADR) in the U.S. stock exchanges and also removed jewelry firms from the *miscellaneous retail* sector as their inventory levels could be driven by commodity prices and other macroeconomic conditions not captured by our model.

Inventory changes could also happen due to changes in foreign exchange rates, mergers & acquisitions (M&A), and discontinued operations. For some companies, these changes could be substantial. However, retailers do not report these changes separately. Hence, we identify firms that may have undergone substantial changes in inventory due to these reasons in the following way. We follow Lundholm et al. (2010) to identify retailers with non-zero values of *AQC* and *DO* from Compustat[3]. We find that about 35% of observations in our population of retailers from 1993 – 2009 had these values populated with non-zero values. The values of *AQC* and *DO* have a wide range and depend on the relative firm size. Hence, we normalize *AQC* and *DO* by total revenue and drop observations which are more than 3 standard deviations away from the mean as we expect these observations are likely to be cases wherein inventory could have undergone substantial changes due M&A and discontinued operations.

We also identified retailers with non-zero values of *PIFO* from Compustat[4] to account for retailers whose income was affected substantially due to changes in foreign exchange rates and find that about 18% of the observations in our population of retailers from 1993 – 2009 had this variable populated with a non-zero value. We followed a similar procedure as above to drop observations with extreme values of *PIFO* as we expect these would be cases where inventory changes may have been significantly impacted due to fluctuations in foreign exchange rates.

---

[3] The definition for AQC in Compustat is given as "This item represents cash outflow of funds used for and/or the costs relating to acquisition of a company in the current year or effects of an acquisition in a prior year carried over to the current year." The definition for DO in Compustat is given as "This item represents the total income (loss) from operations of a division discontinued or sold by the company and the gain (loss) on disposal of the division, reported after income taxes".

[4] The definition for this variable in Compustat is given as "This item represents the income of a company's foreign operations before taxes as reported by the company."

These data adjustments lead to a loss of 4% of observations from our population of retailers from 1993-2009. Finally, some retailers may combine part of their selling, general, and administrative expenses with cost of goods sold; we identify 15 such firm-year observations in the period of 1993-2009 using the data code *xsga_dc* which is populated as "4" in such cases and drop them from our analysis.

Retail sector	2-digit SIC code	Examples of firms	# of firms	# of firms that reported store information for at least 5 years	Entire data set for 1993-2009		EPS sample 2004-2009		Test sample for Analyst Forecast 2004-2009	
					# of firms	# of obs.	# of firms	# of obs.	# of firms	# of obs.
Lumber and other building materials	52	Home Depot, Lowe's, National Home Centers	29	18	59	487	23	96	21	92
Home furnishing stores	57	Williams-Sonoma	69	47						
General merchandise stores	53	Costco, Dollar General, Wal-mart	78	51	48	413	21	106	25	80
Food stores	54	Safeway, Dairy Mart Convenience stores, Shaws	92	57	51	391	18	65	10	51
Apparel and Accessory Stores	56	Mens Wearhouse, Childrens Place	91	74	70	653	45	195	43	139
Miscellaneous retail	59	Toys R Us, Officemax, Walgreen	311	122	95	709	29	121	21	84
TOTAL			670	369	323	2653	136	583	120	446

Table 4.2: Description of initial, final and test data sets by retail sectors, 1993 - 2009

We combine SIC 52 and SIC 57 as SIC 52 has a smaller number of firms and is closest in match to SIC 57. After removing observations with missing data and making the above adjustments, the resulting dataset had 323 retailers across 5 retail segments, viz. *Apparel and Accessory Stores, Food Stores, General Merchandise Stores, Home and Lumber*, and *Miscellaneous Retail Stores*. This resulted in 2653 observations for the period of 1993 – 2009. We choose the six year time period from 2004-2009 as our test sample for analyzing the relationship between abnormal inventory growth and one year ahead earnings. We have 136 retailers and 583 firm-year observations in this sub-sample, of which 125 retailers had reported comparable store sales data yielding 519 firm-year observations. We conduct further analysis with analysts' forecasts and stock market data. We found individual analysts' forecasts of EPS were available for 446 of the 583 observations from I/B/E/S and obtained stock market returns data from CRSP.

Definitions	Variables	Mean	Standard Deviation	Min	Max
Average cost-of-sales per store ($ M)	CS_{it}	8.130	11.341	.189	115.267
Average inventory per store ($ M)	IS_{it}	1.216	1.955	.011	15.118
Gross margin	GM_{it}	1.584	.235	1.121	2.613
Average SGA per store ($ M)	$SGAS_{it}$	2.196	1.874	.066	16.147
Store growth	G_{it}	1.081	.174	.299	2.459
Accounts-payable-to-inventory ratio	PI_{it}	.511	.305	.121	3.714
Accruals	Acc_{it}	.032	.081	-.319	.727
Comparable store sales growth (%)	$Comps_{it}$	3.117	5.583	-12.60	25.10
Change in gross margin	ΔGM_{it}	-.005	.043	-.165	.087
Earnings per share ($)	EPS_{it}	1.901	1.487	-5.914	10.214
Prior period closing price ($)	P_{it-1}	28.101	19.857	.047	138.75
Change in earnings per share ($)	ΔEPS_{it}	-.095	1.014	-6.170	5.170
Change in earnings per share/price	$\Delta EPS1_{it}$	-.002	.105	-.874	.754

Table 4.3 Definitions and summary statistics of variables for 2004 – 2009. Descriptive statistics are based on sample size = 583 observations

We provide further details on these data in §4.7. The number of retailers in each segment and distribution of retailers are given in Table 4.2. Summary statistics for all variables used in our analysis is as shown in Table 4.3.

4.5 Methodology

In this section, we describe the methodology used to measure abnormal change in inventory levels. It is customary in the operations management literature to determine the normal or expected changes in inventory level based on expectation models for inventory levels for retailers. Deviations of actual inventory levels from such expected inventory levels are then expected to serve as benchmarking metrics.

We use the expectation model from Kesavan et al. (2010) to measure abnormal inventory growth for retailers. We use this model since it subsumes many of the factors identified in past research and it was found to be useful in the context of sales forecasting. This model uses a log-log specification where the inventory per store for a retailer in a given fiscal year depends on firm-fixed effect (J_i), inventory per store in the previous fiscal year ($IS_{i,t-1}$), contemporaneous and lagged cost-of-goods-sold per store ($CS_{it}, CS_{i,t-1}$), gross margin (GM_{it}), lagged accounts payable to inventory ratio ($PI_{i,t-1}$), store growth (G_{it}) and lagged capital investment per store ($CAPS_{i,t-1}$) for that retailer. Using lower-case letters to denote the logarithm of these variables, the logged inventory per store for retailer *i* in fiscal year *t* is given as:

$$is_{it} = J_i + \boldsymbol{\beta}_2 \mathbf{x}'_{\mathrm{it}} + \eta_{it} \tag{4.1a}$$

Where $\mathbf{x}'_{\mathrm{it}}$ is a column vector of all right hand side explanatory variables;

$\mathbf{x}'_{\mathrm{it}} = (1, cs_{it}, gm_{it}, \; cs_{it-1}, is_{it-1}, pi_{it-1}, g_{it}, caps_{it-1})'$ and $\boldsymbol{\beta}_2$ is the row vector of the corresponding coefficients; $\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \; \beta_{24}, \beta_{25,}\beta_{26}\,, \beta_{27}\,)'$ and J_i is the firm fixed effect.

First-differencing the above equation gets rid of the firm fixed-effect J_i and yields the following growth model:

$$\Delta is_{it} = \Delta\mathbf{x}'_{\mathrm{it}}\boldsymbol{\beta}_2 + \Delta\eta_{it} \tag{4.1b}$$

Here Δ denotes the change in logged variable in fiscal year *t* from fiscal year *t-1*.

One may treat all of the coefficients $\boldsymbol{\beta}_2$ in the above regression as being firm specific, i.e. allow the sensitivity of inventory per store to different factors such as cogs per store, gross margin, capital investment per store etc. to vary from retailer to retailer. However, in order to estimate such a model, we would need a long time-series of observations for each retailer. Since we use annual data in our analysis we would need several decades of data for each retailer to estimate such a model. To overcome the paucity of data, we assume that all firms in a given segment are homogenous, i.e. we assume that the coefficients $\boldsymbol{\beta}_2$ are same for all retailers within a given segment and estimate these coefficients at the segment level. Thus our estimation equation is:

$$\Delta is_{it} = \Delta\mathbf{x}'_{\mathrm{it}}\boldsymbol{\beta}_{2,s(i)} + \Delta\eta_{it} \tag{4.1c}$$

Where *s(i)* denotes the corresponding segment specific coefficients for firm *i*.

We can now obtain the expected logged inventory growth from the above equation, $E(\Delta is_{it})$, and then compute abnormal inventory growth in the following way. Let $\left\{\frac{IS_{it}}{IS_{it-1}} - 1\right\}$ denote the actual inventory per store growth and $AIG_{it} = \left(\left\{\frac{IS_{it}}{IS_{it-1}} - 1\right\} - \{exp(E(\Delta is_{it})) - 1\}\right)$ denote the abnormal inventory per store growth or, in short, abnormal inventory growth for a retailer *i* in fiscal year *t*. We estimate (1c) and use the coefficients to compute abnormal inventory growth. Thus, $AIG_{it} > 0$ implies that the retailer *i* has abnormally high inventory growth while $AIG_{it} < 0$ implies that the retailer *i* has abnormally low inventory growth compared to the norm of the segment to which the retailer belongs to, after controlling for firm-level differences.

Kesavan et al. (2010) show that historical gross margin contains information valuable to forecast sales. Further they show that inventory and gross margin are highly correlated for retailers, so we calculate abnormal change in gross margin in a similar manner as AIG and use it as a control variable in our analysis. Similar to equation 4.1c, the first differenced equation for gross margin can be written as:

$$\Delta gm_{it} = \Delta\mathbf{x}'_{\mathrm{it}}\boldsymbol{\beta}_{3,s(i)} + \Delta v_{it} \tag{4.2}$$

Where $\Delta\mathbf{x}'_{\mathrm{it}} = (1, \Delta cs_{it}, \Delta is_{it}, \Delta gm_{it-1})'$. We calculate abnormal change in gross margin for retailer i in fiscal year t as $ACGM_{it} = \left(\left\{\frac{GM_{it}}{GM_{it-1}} - 1\right\} - \{exp\big(E(\Delta gm_{it})\big) - 1\}\right)$.

Next, we explain the data used to obtain AIG_{it-1} from (4.1c) which is then used to predict earnings in fiscal year *t*. We use data till fiscal year *t-2* to estimate (4.1c). We avoid data from fiscal year *t-1* in the estimation since firms announce their financial results at different times of the year that could lead to a potential look-ahead which could bias our results about the relationship between AIG and one-year ahead earnings. Once a retailer's financial results are announced for fiscal year *t-1*, we use the coefficient estimates to measure the AIG_{it-1} for *that* retailer. We follow this process for all retailers in our test sample, i.e., *t*=2004, 2009. We follow a similar approach to obtain $ACGM_{it-1}$.

We considered two different techniques to estimate equations (4.1c) and (4.2). We used the instrument variable generalized least squares (IVGLS) method used in Kesavan et al. (2010) to estimate the equations and also used a simpler single equation technique, a Generalized Least Squares method (GLS), to estimate these equations. We found the results to be similar. Since the IVGLS method requires defining an additional equation containing new variables, we choose to report the results of the GLS technique that is simpler to implement and explain. The GLS method handles heteroskedasticity and panel specific auto-correlation in the data.

Table 4.4 reports sample results of estimation of equations (4.1c) and (4.2) using data from 1993 – 2007. These coefficient estimates were then used to calculate AIG and ACGM for fiscal year 2008 that are then used to predict earnings for fiscal year 2009.

		Retail Industry Segment				
Equation	Variables	General merchandise stores	Food Stores	Apparel and accessory stores	Home furnishing stores	Miscellaneous retail
Inventory Equation	Intercept	-.014** (.007)	-.018*** (.006)	-.011** (.005)	-.012*** (.006)	-.015*** (.004)
	Δis_{it-1}	-.087* (.005)	-.125 (.098)	-.101* (.006)	-.077* (.005)	-.095* (.006)
	Δcs_{it}	.904** (.405)	.815** (.387)	.798*** (.125)	1.012*** (.014)	.874** (.211)
	Δgm_{it}	-3.175** (1.214)	.057 (.125)	.547** (.247)	-.915* (.551)	1.875** (.931)
	Δcs_{it-1}	.046* (.001)	.014 (.198)	.038* (.023)	.098 (.124)	.055* (.033)
	Δpi_{it-1}	-.013** (.005)	-.007* (.004)	-.018** (.008)	-.017** (.009)	-.012* (.007)
	Δg_{it}	-.099** (.006)	-.087** (.005)	-.074** (.006)	-.075* (.005)	-.081* (.005)
	$\Delta caps_{it-1}$	.049* (.003)	.058 (.044)	.061* (.005)	.047* (.004)	.051** (.001)
Gross Margin Equation	Intercept	-.001** (.000)	-.003** (.000)	-.015** (.008)	-.011** (.006)	-.004* (.001)
	Δgm_{it-1}	-.014* (.009)	-.019 (.269)	-.011* (.006)	-.017 (.549)	-.012 (.019)
	Δcs_{it}	.095** (.004)	.078* (.005)	.181** (.009)	.198** (.010)	.121** (.005)
	Δis_{it}	-.077** (.035)	-.041** (0.20)	-.054** (.027)	-.184** (.010)	-.171** (.090)

Table 4.4 Coefficients' estimates for the variables in Equations 1c and 2 for all retail segments, 1993 – 2007. Note: All variables have been first differenced. n=2322. All regressions are run after controlling for year fixed effects and panel specific autocorrelation. ***denotes statistically significant at $p<0.001$, **at $p < 0.05$ and *at $p < 0.1$ level

Figure 4.1a presents the histogram of AIG for all retailers in our EPS sample (n= 583) during the period $t=2004,.., 2009$. We find that 61% of retailers have AIG>0 and 39% of retailers to have AIG<0.We find that the average, lowest, highest and standard deviation of AIG for this time period are 1.71%[5], -19.55%, 26.65%, and 7.70% respectively. These statistics for inventory per store growth during the same period are 3.61%, -55.27%, 179.27%, and 14.33% respectively. The average AIG across the different segments for the same period is 2.81% (apparel), 0.14%

[5] This value corresponds to $0.34 million dollars of abnormal inventory per store for a retailer who carries $20 million of inventory per store.

(food), 1.15% (general), 2.01% (home) and 1.06% (miscellaneous). The magnitude of correlations between AIG and sales per store, sales growth and store growth are less than 0.18. These weak correlations indicate that AIG is not specific to retailer characteristics such as its size or growth rate. We also find that the relative rank of retailers based on AIG varies considerably from year to year indicating that AIG is not persistent. The average, lowest, highest, and standard deviation of ACGM in the same period is -.50%, -14.9%, 63.91%, and 4.1% respectively.

Abnormal days of inventory (AbI)

We use the abnormal days of inventory (AbI) measure proposed by Chen et al. (2007) as an alternate measure of abnormal inventory growth. Chen et al. (2007) define abnormal days of inventory (AbI_{it}) as the normalized deviation of the days of inventory (DOI_{it}) of retailer i in fiscal year t from those of its industry peers.

$$AbI_{it} = \frac{(DOI_{it} - \overline{DOI_{st}})}{\overline{\overline{DOI_{st}}}}$$

Here $\overline{DOI_{st}}$ and $\overline{\overline{DOI_{st}}}$ are the average and standard deviation of days of inventory of all retailers in the segment s to which retailer i belongs to[6]. If $AbI_{it} > 0$ ($AbI_{it} < 0$) then retailer i holds inventory longer (shorter) than the segment norm in year t. The histogram of AbI is shown in Figure 1b. The average, lowest, highest and standard deviation of AbI during 2004-2009 are 0.15, -1.85, 2.11 and 1.21 respectively. The average abnormal days of inventory across the different segments for the same period is 0.21(apparel), 0.08(food), 0.14(general), 0.16(home) and 0.12(miscellaneous).

---

[6] Chen et al. (2007) define industry segment based on the North American Industry Classification System (NAICS). Chen et al. (2005) who also define the same metric use the 3-digit SIC code to identify firms in the same segment. We follow Chen et al. (2007)'s definition of industry as opposed to that of Chen et al. (2005) since the former study included retailers while the latter was restricted to manufacturers.

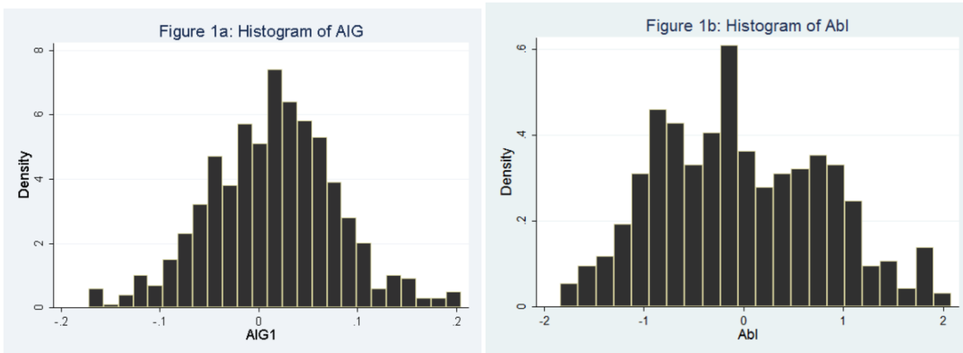


Figure 4.1: Histograms of AIG and AbI

The main difference between the AIG metric and AbI metric is that the former metric controls for factors such as gross margin, capital investment, store growth, and accounts payable that have been identified as important factors that drive inventory levels in operations literature while the AbI metric does not. However, the use of lagged variables in the regression used to estimate AIG means that we need at least three years of data to measure AIG for a retailer. On the other hand, the AbI metric can be computed even for a retailer that has just one year of data. We use both metrics to test the relationship between abnormal inventory growth and one-year ahead earnings.

4.6 Results

In this section, we discuss the results of our statistical tests of the relationship between AIG and one-year ahead earnings. Several researchers in accounting have used a first order autoregressive model for change in one-year ahead earnings. We adopt the same model to test the relationship between AIG and one-year ahead earnings. Accounting literature has also found that accruals, defined as the difference between net income and operating cash flows, predict one-year ahead earnings (Sloan 1996). Since one of the components of accruals is change in inventory, we use accruals as a control variable to examine if AIG has additional information over that contained in accruals to predict earnings. This gives us the following model to test the relationship between AIG and one-year ahead earnings:

$$\Delta EPS1_{it} = \alpha_o{}^{eps} + \alpha_t{}^{eps} + \alpha_1{}^{eps}\Delta EPS1_{it-1} + \alpha_2{}^{eps}\Delta Acc_{it-1} + \alpha_3{}^{eps}AIG_{it-1} + \alpha_4{}^{eps}AIG_{it-1}^2 + \varepsilon_{it}{}^{eps} \tag{4.3}$$

Here, $\Delta EPS1_{it}$ denotes change in EPS deflated by previous fiscal year's ending stock price to homogenize firms when firms are drawn from a broad range of sizes (Durtschi and Easton, 2005). We use a full set of year dummies ($\alpha_t{}^{eps}$) to account for macroeconomic factors that may impact earnings of all retailers. We use coefficient estimates of $\alpha_3{}^{eps}$ and $\alpha_4{}^{eps}$ to determine the relationship between AIG and change in one year ahead earnings.

Model 1 in Table 4.5 reports the results for the base model, i.e. a first order autoregressive model of EPS with accruals. Consistent with the accounting literature, we find that accruals have predictive power over one-year ahead earnings. Model 2 gives the estimated coefficients of Equation 4.3. We find that the coefficients of AIG_{it} and AIG_{it}^2 are negative and significant (*p*<0.001) and provide support for an inverted-U relationship between AIG and change in one-year ahead EPS. We perform a Wald test to confirm that the addition of AIG_{it} and AIG_{it}^2 improves the fit of our model (*p*<0.001). We use the coefficient estimates ($\alpha_3{}^{eps}, \alpha_{34}{}^{eps}$) from Model 2 to graphically illustrate the inverted-U relationship between AIG and change in one-year ahead earnings as shown in Figure 4.2. The mean AIG in our sample is .017 and the standard deviation is 0.07. At the mean, the impact of increasing AIG by 0.01 leads to a decrease in EPS of 0.2 cents. At a higher level of distribution, increasing AIG by 0.01 leads a decrease in EPS of 0.7 cents. At a lower level of distribution, further decreasing AIG by 0.01 leads to a decrease in EPS of 0.3 cents.

Independent Variables	Dependent Variable: Change in EPS1			
	Model 1	Model2	Model3	Model 4
Intercept	-1.118***(.001)	-.077***(.001)	-.026***(.001)	-.020***(.001)
$\Delta EPS1_{it-1}$	-1.014***(.009)	-.345***(.021)	-.120***(.017)	-.110***(.003)
AIG_{it-1}		-.165***(.005)	-.189***(.006)	-.176***(.005)
AIG^2_{it-1}		-1.689***(.078)	-1.775***(.084)	-1.119***(.061)
$ACGM_{it-1}$			.174***(.019)	.159***(.040)
ΔAcc_{it-1}	1.286***(.015)	.315***(.006)	.126***(.015)	.115***(.011)
Segment dummies	No	No	No	Yes
Wald χ^2	1101.02	2877.13	3571.13	4741.69
n	583	583	583	583

Table 4.5 Impact of AIG on change in one-year-ahead EPS1, 2004-2009
Where $\Delta EPS1_{it-1}$ = Previous change in EPS, AIG_{it-1} = Lagged AIG, AIG^2_{it-1} = Lagged AIG^2, $ACGM_{it-1}$ = Lagged ACGM and ΔAcc_{it-1} = Lagged change in Accruals. All regressions are run after controlling for year fixed effects and panel specific autocorrelation. Standard errors are reported in brackets below the coefficients. ***denotes statistically significant at p<0.001, **at p < 0.05 and *at p < 0.1 level

To ensure that outliers are not driving the inverted-U relationship, we follow Aiken and West (1991) to statistically test this relationship. Aiken and West (1991) recommend performing tests for the significance of slopes spanning observations on either side of the inflexion point and verify the change in signs of the slopes. The inflexion point occurs at a value of -0.049 ($-{\alpha_3}^{eps}/2{\alpha_4}^{eps}$) which is less than one standard deviation away from the mean and well within the range of our sample, [-0.195, 0.267]. We find that 16% of our observations lie to the left of the inflexion point. We perform *t*-tests of simple slopes using coefficient estimates of 4.3 and the results are reported in Table 4.6. Since the simples slopes at values of AIG that are two standard deviations below mean and above mean are both statistically significant and opposite in signs, we can conclude that the inverted-U relationship between AIG and one-year ahead earnings is supported within the range of our data and not driven by outliers.

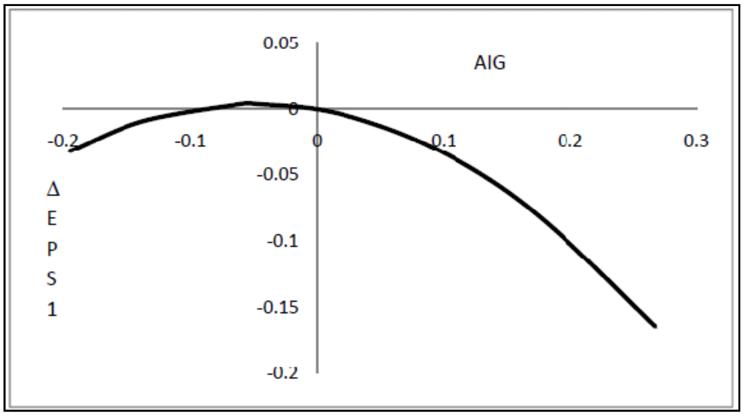


Figure 4.2: Impact of AIG on one-year ahead change in earnings per share (ΔEPS1)

AIG Value[a]	Simple slope	Standard error	Significance
-.195	.493	.021	4.865***
-.137	.297	.015	2.995**
-.06	.037	.007	.915
-.049[b]	.000	.006	.000
.017	-.222	.004	-9.036***
.094	-.483	.010	-10.071***
.171	-.743	.019	-9.035***
.267	-1.067	.028	-8.382***

Table 4.6: t-tests for simple slopes at different values of AIG for the regression equation:
$\Delta EPS1_{it} =$
$\alpha_o{}^{eps} + \alpha_t{}^{eps} + \alpha_1{}^{eps}\Delta EPS1_{it} + \alpha_2{}^{eps}\Delta Acc_{it} + \alpha_3{}^{eps}AIG_{it-1} + \alpha_4{}^{eps}AIG^2_{it-1} + \varepsilon_{it}{}^{eps}$
Where coefficients $\alpha_o{}^{eps}, \alpha_t{}^{eps}, \alpha_1{}^{eps}, \alpha_2{}^{eps}, \alpha_3{}^{eps}$ and $\alpha_4{}^{eps}$ are based on model 2 in table 4.5.[a]These represent the different values of AIG in our sample i.e. (minAIG, mean AIG – 2 sd, mean AIG – 1 sd, mean AIG, mean AIG + 1 sd, mean AIG + 2 sd and maxAIG). [b] inflexion point ***denotes statistically significant at p<0.001, **at p < 0.05 and *at p < 0.1 level

The impact of AIG on one-year ahead earnings per share is measured using estimates of coefficients from Model 2 (table 4.5) in the following way, i.e., Impact $= -.165AIG_{it-1} - 1.689AIG^2{}_{it-1}$.

There are two attributes of the inverted-U relationship that we further elaborate on. First, the inflexion point occurring at a value less than zero is interesting. Since our results are picking up the dominant effect (as we cannot control for service level), we conjecture that the region [-0.049, 0] is dominated by retailers who became leaner. That is, these retailers were able to reduce their inventory levels without substantial reduction in service level. We conjecture that the region

AIG<-0.049 is dominated by retailers whose inventory levels declined so much that the accompanying decline in service level hurt their one-year ahead earnings. To ensure that this result is not an artifact of the AIG metric, we re-test Model 2 by substituting the AIG metric with AbI metric from Chen et al. (2007) and find qualitatively similar results. The coefficients of AbI_{it} and AbI_{it}^2 are -0.006 and -0.007 (p<0.001) respectively indicating the existence of an inverted-U shape relationship between AbI and change in one-year ahead earnings. Similar to the results obtained with the AIG metric, we find that the inflexion point is negative and lies within one standard deviation of the mean. The values of the inflexion point, minimum, and maximum values of AbI are -0.43, -1.85, and 2.11 respectively. Thus our results appear to be robust to the method used to compute abnormal inventory carried by retailers.

Second, consistent with Chen et al. (2007) who examined long-term stock market returns of retailers, we find that our strongest results are for retailers with abnormally high inventory growth who have poor subsequent performance. There are likely to be some retailers in this region (AIG>0) who were able to increase service level such that the new service level is closer to the optimal service level and found a subsequent increase in profitability. However, those retailers appear to be dominated by retailers who had excess inventory.

We perform several robustness checks to confirm the inverted-U relationship. First, we control for ACGM as Kesavan et al. (2010) show that this variable contains information useful to predict sales. The correlation between AIG and ACGM is low (ρ=-0.18) and not significant. We also compute the variance inflation factor (VIF) between AIG and ACGM for Model 3 and find it to be 1.3. This rules out any multicollinearity issues arising due to including these variables together in an equation as the VIF is less than 10 (Maddala 2001). The results, as shown in Model 3 in Table 4.5, continue to show the inverted-U relationship. Second, we add segment dummies to Model 3 and the estimation results are as shown in Model 4. Our conclusions about the inverted-U relationship remain unchanged with the addition of these variables.

Because AIG predicts sales (Kesavan et al. 2010) and earnings are a function of sales, we want to determine if the relationship between AIG and earnings are driven only by AIG's ability to predict sales or if there are additional reasons why AIG might predict earnings. As we argue in §3, AIG might also predict higher expenses such as advertising costs to clear merchandise, holding costs, and inventory write-downs. Since these expenses are not readily available as separate line items in retailers' income statement we test this indirectly in the following way. We add one-year ahead comparable store sales ($COMPS_{it}$) to Model 4 to control for changes in EPS due to change in sales for that retailers in the following way:

$$\Delta EPS1_{it} = \alpha_o{}^{eps} + \alpha_t{}^{eps} + \alpha_1{}^{eps}\Delta EPS1_{it-1} + \alpha_2{}^{eps}AIG_{it-1} + \alpha_3{}^{eps}AIG^2{}_{it-1} + \alpha_4{}^{eps}\Delta Acc_{it-1} + \alpha_5{}^{eps}COMPS_{it} + \alpha_6{}^{eps}ACGM_{i,t-1} + \varepsilon_{it}{}^{eps} \quad (4.4)$$

Thus, significance of AIG would indicate that it contains information about future expenses that would be useful to predict one-year ahead earnings for retailers. Models 5 and 6 in Table 4.7 report the results of the base model and results from equation 4.4. We find that both the linear term and the quadratic terms of AIG are significant (*p*<0.001), even after controlling for contemporaneous values of comparable store sales, indicating that AIG predicts earnings for retailers due to its ability to predict sales and expenses for retailers.

Independent Variables	Dependent Variable: Change in EPS1	
	Model 5	Model 6
Intercept	-.024***(.001)	-.024***(.001)
$\Delta EPS1_{it-1}$	-.110***(.017)	-.111***(.003)
AIG_{it-1}		-.171***(.005)
AIG^2_{it-1}		-1.129***(.061)
$ACGM_{it-1}$	.459***(.001)	.119***(.014)
ΔAcc_{it-1}	.156***(.011)	.110***(.011)
Segment dummies	Yes	Yes
$Comps_{it}$	.591***(.015)	.297***(.015)
Wald χ^2	5171.13	7841.69
n	519	519

Table 4.7 Impact of comparable store sales and AIG on change in one-year-ahead EPS1, 2004-2009, Where $Comps_{it}$ = Comparable store sales.

We also replace comparable store sales in period *t* with sales growth in that period and obtain similar results (not reported).

4.7 Economic Significance of Information contained in AIG

In this section, we investigate the economic significance of our finding. First, we examine if equity analysts take the information contained in AIG into account when generating their earnings forecasts. Second, we test if stock prices incorporate this information. The former test would show if even sophisticated investors could benefit from knowing the information contained in AIG and the latter would indicate if AIG can form the basis for an investment strategy.

4.7.1 Do equity analysts ignore information in abnormal inventory growth in EPS forecasts?

We examine if equity analysts ignore information contained in AIG or not in the following way. Analysts issue earnings forecasts at different times during a year and revise those forecasts as more information becomes available. These forecasts are time stamped with the dates that they are issued. Since the financial information for the previous fiscal year are released on the earnings announcement date (EAD), the information required to compute AIG for a retailer is available after its EAD. If analysts' incorporate the information from lagged AIG, then their forecasts issued subsequent to EAD should not generate errors that can be predicted by lagged AIG. On the other hand, if they do not incorporate this information then lagged AIG will have predictive power over their forecast errors.

Our tests are conducted using the IBES detailed (median) forecasts of annual EPS. We perform this analysis using data obtained for fiscal years 2004-2009. We consider analysts' earnings forecasts for the forthcoming fiscal year issued after EAD of the prior fiscal year. In some cases, analysts might have to wait till the retailers file their 10-K statement with SEC to have access to those retailer's financial statements. To be conservative, we drop any analyst forecasts made before the SEC filing date as well. We obtain SEC filing date for each retailer from Morningstar Document Research that is accessible from http://www.10kwizard.com/. If multiple forecasts are made by an analyst for a retailer, we use the most recent forecast as it should contain the latest information available to them.

We find that analysts' forecasts were available for 446 observations out of the 583 overall observations. For each firm-year, we determine the median of analysts' forecasts made for each of the *m*= 1 to 12 months after EAD for fiscal year *t-1* and use it as consensus forecasts to generate analysts' consensus forecast error. We compute forecast errors by subtracting consensus EPS forecast from realized EPS. We also deflate the forecast error by the previous fiscal year's ending stock price (Gu and Wu 2003). The average, standard deviation, minimum, and maximum deflated analysts' forecast error between 2004-2009 are -.005, .036, -.384, and .137 respectively. The average forecast error of -.005 shows that analysts are optimistic on average, which is consistent with prior accounting literature.

Next we statistically test if analysts' forecast errors are biased are predicted by lagged AIG by running the following regression:

$$FE_{itm} = \chi_0 + \chi_t + \chi_1 AIG_{i,t-1} + \chi_2 ACGM_{i,t-1} + \boldsymbol{\gamma} Y_{itm} + \psi_{it} \quad (4.5)$$

Here, FE_{itm} is the deflated forecast error of analysts' consensus forecast generated m months before end of fiscal year *t* for retailer *i*. χ_0 is the bias that is common to all retailers, χ_t is the bias that is specific to a given fiscal year, χ_1 is that bias that is correlated with previous year's AIG; and χ_2 is the bias that is correlated with previous year's ACGM. Y_{itm} is the vector of control variables that were found to be related to forecast bias in the accounting literature (Gu and Wu 2003). These include dispersion among analysts' forecasts, analyst coverage, market value, unexpected earnings from a seasonal random walk model, and a dummy variable to capture ex-ante expectation of loss in earnings.

We estimate (4.5) using the GLS technique. Table 4.8 provides the formal statistical test of the relation between analysts' forecast errors and lagged AIG.

Independent Variables	Dependent Variable: Deflated analyst forecast error m months after EAD_{t-1}			
	m = 1 month	m = 3 months	m = 6 months	
Intercept	-3.5e-3*** (1.5e-5)	-3.3e-3*** (1.1e-5)	-3.1-3** (6.2e-4)	-1.5e-3** (4.1e-4)
AIG_{it-1}	-.031*** (3.9e-5)	-.029*** (0.5e-5)	-.019* (6.1e-3)	-.012* (3.6e-3)
$ACGM_{it-1}$	-.114*** (1.2e-5)	-.111*** (2.1e-5)	-.101* (.006)	-.090* (.005)
$DISP_{itn}$	-1.147*** (.015)	-1.088*** (.013)	-1.001* (.014)	-.861* (.011)
ΔAcc_{it-1}	-8.5e-3*** (1.5e-4)	-7.4e-3*** (2.5e-4)	-2.9e-3* (9.5e-4)	-5.9e-3* (9.1e-4)
FE_SALE_{itm}				1.1e-4* (7.2e-5)
Control Variables	Yes	Yes	Yes	Yes
n	415	446	446	387
Wald (χ^2)	9081.93	5452.25	3412.57	8426.7

Table 4.8 Bias in deflated analysts' EPS forecasts due lagged AIG, 2004 - 2009
$DISP_{itm}$= analyst dispersion i.e. the standard deviation of deflated forecasts for each year and FE_SALE_{itm}= deflated analyst sales forecast error. This regression also includes the following control variables from Gu and Wu (2003) (coefficients not reported): $LGMV_{it-1}$= log (market value), $LGFFW_{it-1}$= log(analyst coverage), $Loss_{itm}$= an ex-ante loss dummy variable which takes a value of 1 if the forecasted current earnings are negative and 0 otherwise, SUE_1_{it-1} and SUE_2_{it-1} are price deflated lag-one and lag-two unexpected earnings from a seasonal random walk model. All regressions are run after controlling for year fixed effects and panel specific autocorrelation. Standard errors are reported in brackets below the coefficients. ***denotes statistically significant at p<0.001, **at p < 0.05 and *at p < 0.1 level.

We find that the bias in analysts' forecasts due to AIG continues to remain significant (p<0.1) up to 6 months from EAD for the prior fiscal year as shown in Table 4.8. The magnitude of the reported coefficients may be interpreted in the following way. Consider the forecasts made m= 6 months after EAD for the prior fiscal year (m = 6 in table 4.8). We find that the coefficient of AIG is -.019 (p<0.1). This implies that for an average retailer with a share price of $28.101 (mean share price in our sample), a one standard deviation increase in AIG is associated with 4.11 cents increase in analysts' forecast error. We also confirm that the direction and significance of our control variables are consistent with prior accounting literature (Gu and Wu 2003).

Next, we want to determine if our result that AIG predicts bias in analysts EPS forecasts is driven due to analysts failing to incorporate information contained in AIG to predict sales, as shown by Kesavan et al. (2010), or to predict expenses as well. We do so by picking analysts' sales forecasts made m=6 months after EAD for prior fiscal year (column 3, table 4.8). Sales forecasts were available for only 387 out of 446 observations in our sample. We add contemporaneous error in analysts' sales forecasts to equation 5 for this sample (m = 6) and find that AIG continues to remain a significant predictor of bias in analysts' forecasts of EPS (p<0.1). We obtain consistent results for m=1 and 3 months as well where the results are stronger (p<0.05).

To summarize, our results show that analysts fail to incorporate information in AIG to predict earnings. These results not only support the results from Kesavan et al. (2010) but add to it by showing that analysts fail to incorporate the information contained in AIG useful to forecast expenses as well.

4.7.2 Does an investment strategy based on AIG yield abnormal returns?

In this section, we examine whether investments based on AIG can yield abnormal returns. We follow the methodology used in Abarbanell and Bushee (1998) and Desai et al (2004) to perform this analysis. This methodology can be used to determine the abnormal returns to a zero-investment strategy based on AIG and other control variables. We use accruals, inventory growth, and book-to-market as control variables for the following reasons. Sloan (1996) showed that investment strategies based on accruals yields significant abnormal returns. Since accruals are comprised of many components, Thomas and Zhang (2002) test the strength of each of those components and find that the inventory growth component has the highest explanatory power. We are motivated to examine if investment strategy based on AIG would generate incremental abnormal returns after controlling for accruals and inventory growth. Finally, we also control for book-to-market as it is a proxy for whether a firm is a value stock or not (Desai et al. 2004).

This methodology involves dividing firms in to different portfolios based on quintile ranks of AIG and quintile ranks of each of the control variables. An important consideration in the construction of these portfolios is that all information required to construct the portfolios for a given fiscal year are available at the time point when the portfolio is created. Since firms file their 10-K statements with SEC at different time points in the year, the information required to construct a portfolio comprising of all the retailers becomes available at different points in time. In our sample, we find that the month of April had the most number of SEC filings (245), followed by March (83) and May (52). The rest of the months had fewer than 20 filings. To ensure that all firms in our portfolio are aligned in calendar time such that their 10-K statements are filed in the same month, we include only firms that file their 10-K statements in April. Thus our portfolios are created on May 1 and we compute the buy-and-hold abnormal returns for the 12 month period from thereon.

We measure the size-adjusted buy-and-hold abnormal return (*BHAR*) of each of the stock in our sample in the following way (Kothari and Warner 2007). The $BHAR_{it}$ for firm *i* cumulated for a period of 12 months for fiscal year *t,* beginning from May of the *prior* fiscal year is:

$$BHAR_{it}(m) = \prod_{m=1}^{12}(1 + R_{imt}) - \prod_{m=1}^{12}(1 + SAR_{kmt})$$

Where R_{imt} is the stock return of firm i in month m and SAR_{kmt} is the return of the value-weighted portfolio of firms in the CRSP size decile to which this firm belongs for that fiscal year. The size deciles are obtained from the distribution of market values of all NYSE/AMEX firms at the beginning of the fiscal year.

Next we compute the quintile ranks based on lagged AIG for each of the firms in each of the fiscal years 2004-2009, and then scale those ranks to obtain new variable $sAIG_{it-1}$. We illustrate using fiscal year 2004 as an example. First, we rank firms from 0 to 4 based on the quintile rank of AIG_{2003}. Retailers with rank of zero (four) are those firms whose AIG were low (high) enough to belong to the bottom (top) quintile in 2003. Next we divide the ranks by 4 to obtain $sAIG_{i2003}$. We repeat the above procedure to obtain variables $sAcc_{it-1}, sInvg_{it-1}, sBM_{it-1}$ which are the scaled quintile ranks based on lagged accruals, lagged inventory growth, and lagged book-to-market respectively. The values of each of these scaled variables will lie between 0 and 1.

After creating all the variables, we run a number of regressions of $BHAR_{it}$ against different combinations of $sAIG_{it-1}$, $sAcc_{it-1}$, $sInvg_{it-1}$, and sBM_{it-1} , after controlling for year-fixed effects. The coefficients on each of the variables can be interpreted as the abnormal return to a zero investment strategy in the respective variable. This test of significance of these coefficients has been found to overcome bias that might otherwise occur due to cross-sectional correlation of the size-adjusted return metric (Bernard 1987).

The regression results are reported in Table 4.9. First, consider models M1a, M1b, and M1c. Model M1a shows that a zero investment strategy based on accruals would yield an abnormal return of 10.1% (p<0.001). This result is consistent with that reported in Sloan (1996). Similar analysis, as shown in Model 1b, shows that the abnormal returns to AIG is 11.8% (p<0.001). Next, we consider accruals and AIG together in Model 1c and find that the incremental returns of AIG decreases from 11.8% to 10.8% but continues to remain significant (p<0.001). Our results show that an investment strategy based on AIG would produce abnormal returns. Furthermore, we find that the information content in AIG is not subsumed in accruals.

	M1a	M1b	M1c	M2	M3
Intercept	.191***(.025)	.167***(.021)	.234***(.024)	.136***(.016)	.212***(.021)
$sAcc_{it-1}$	-.101***(.021)		-.097***(.020)		-.086***(.023)
sBM_{it-1}	-.205***(.024)	-.202***(.022)	-.201***(.026)	-.241***(.025)	-.204***(.027)
$sAIG_{it-1}$		-.118***(.023)	-.108***(.019)		-.103***(.022)
$sInvg_{it-1}$				-.034***(.001)	-.006(.021)
Wald(χ^2)	1622.52	1691.66	2159.62	1079.77	4293.73
n	245	245	245	245	245

Table 4.9 Regression of SAR (BHAR) on zero-investment portfolios based on AIG, Accruals, Book-to-market and Inventory Growth
All regressions are run after controlling for year fixed effects and panel specific autocorrelation. Standard errors are reported in brackets below the coefficients. ***denotes statistically significant at p<0.001, **at $p < 0.05$ and *at $p < 0.1$ level

Next we want to determine if AIG contains more information than inventory growth metric, defined as change in total inventory scaled by average total assets (Thomas and Zhang 2002). So, we run model M2 where we replace accruals ($sAcc_{it-1}$) in M1a by inventory growth ($sInvg_{it-1}$). Consistent with Thomas and Zhang (2002), we find that abnormal returns to inventory growth to be significant ($p< 0.001$). Finally, we run full Model M3 which includes all the variables. We find that incremental return to AIG is 10.3% ($p<0.001$). However, the incremental returns to inventory growth are no longer significant; this is as expected since inventory growth is a component of accruals. Thus we add to the results of Thomas and Zhang (2002) by showing that it is the abnormal component of inventory growth, not the normal component, which generates abnormal returns.

We also perform additional robustness tests to validate our findings. First, we ran cross-sectional regressions of the full model for each of the six years (2004-2009) and find that abnormal returns to AIG are significant ($p<0.1$) in each of those years and they vary between 3.91% ($p<0.1$) and 12.9% ($p<0.001$). Similarly, we find abnormal returns to accruals also to be significant ($p<0.1$) for all 6 years while the abnormal returns to inventory growth is significant ($p<0.1$) in four of the six years.

Next, we increase our sample size by considering retailers whose SEC filing dates were in March ($n=83$) and May ($n=52$) in addition to those who filed in April ($n=245$). Our number of observations increases to 380. In order to ensure that all the information required to create portfolios are available at the time of creating the portfolio, our holding period starts from June 1 and continues for 12 months. Thus, we would wait for atleast 2 months for retailers who release their 10K statements in March and atleast 1 month for those who release their 10K statements in April before including them in our portfolio. We find qualitatively similar results as in Table 9, i.e. abnormal returns to a zero-investment strategy in AIG continue to be significant. For example, when we run the full Model M5 in this sample, we find the abnormal returns to AIG, accruals, and inventory growth to be 8.1%, 7.2% and 2.6% respectively. The abnormal returns to

AIG are significant at 5% level while the abnormal returns to inventory growth are not significant.

We also test the robustness of this finding using the Ibbotson-RATS procedure that we briefly explain in the Appendix 6.3.1 (see Kothari and Warner 2007 for details). This approach allows us to estimate the abnormal returns of the hedge portfolios using the three-factor Fama-French model. We find abnormal returns to AIG, accruals, and inventory growth are 11.65%; 6.57%; and 3.86% respectively. Thus our results are robust to the method used to estimate abnormal returns.

4.8 Conclusion, Limitations, and Future Work

We document an inverted-U relationship between abnormal inventory growth and one-year ahead earnings for retailers using publicly available financial data. To the extent that this relationship is an evidence of causality, our results imply that retailers should avoid abnormal decrease in inventory growth beyond a certain point and abnormal increase in inventory growth, other things being equal. However, caution should be exercised with this interpretation as our model does not guarantee causality in this relationship. Our study has valuable contributions to investors as well. We show that equity analysts do not fully incorporate the information contained in AIG in their earnings forecasts and an investment strategy based on AIG can yield significant abnormal returns. In our sample, we find this return to be 11.8% (p<0.001). Thus a benchmarking metric for inventory performance derived from operations management literature can serve as the basis for an investment strategy.

Next we discuss the limitations of our study. Rajagopalan (2010) use primary data on product variety to show that inventory levels of a firm increase with variety. Olivares and Cachon (2009) show that inventory levels of automotive dealerships increase with competition. AIG and AbI do not account for these factors as details on product variety and competition are not reported at the firm level in financial statements. Future research may investigate the possibility of using proxies for variety and competition to improve the benchmarking metric and test if that results in an increase in abnormal returns. Future research may also use proprietary firm data to study factors that moderate the impact of AIG on earnings. Some of the factors include contracts between retailers and their suppliers that determine how merchandise returns are handled and presence of factory outlet stores or other mechanisms that enable retailers to salvage unsold inventory. Similarly, operational data on customer service levels, product lifecycle, and product assortment would help us better understand the impact of inventory on financial performance. Finally, it would be useful to investigate if benchmarking metrics based on inventory are useful to predict earnings and to form investment strategies in other sectors.

CHAPTER 5

Conclusions and Future Research

Through this dissertation we have examined the link between operations and financial performance for retailers at both the store-level, using store proprietary data, and at the firm-level using publicly available financial data. Together, these highlight the critical role that good operations management plays in achieving higher performance. The empirical studies in the three chapters contribute to the understanding of how operational variables like adequate store labor, good traffic forecasts and having the right level of inventory impact store profitability and firm earnings. They also underscore the crucial role these operational decisions play in driving top-line growth for retailers through increased customer conversion, as well as bottom-line growth by identifying right cost-cutting measures.

In the first chapter, we examine whether or not retail stores are understaffed by utilizing hourly data on the traffic flow, sales volume and labor, and imputing the contribution and cost of labor at each of these stores. Our method of imputing the parameters takes into account the underlying heterogeneity in traffic elasticity, labor productivity and service expectations at each of these stores. We are also able to impute the underlying cost of labor that store managers use in their labor planning decisions that captures not just the direct cost of having labor in the store, but also indirect costs around employee-benefits and training costs. Our study is also the first to use structural estimation techniques in the context of labor planning. Using a given store's estimates of contribution of labor to sales and cost of labor; we construct the optimal labor plan for the store and study deviations of the actual labor from the optimal plan to check for understaffing.

We find that the stores differ widely in the contribution of labor to sales and their imputed cost of labor. For example, the average hourly imputed cost of labor in our study was found to be $30.47, with a range from $10.50 to $54.92. Furthermore, this cost is significantly higher than the average hourly wage rate of $10.05 for retail salespersons, which can be explained partly by systematic factors based on individual store and local market area characteristics. These results suggest that workforce management tools that are increasingly being deployed in corporate offices should not ignore the heterogeneities in the imputed cost of labor across stores. Else, this could lead to misalignment between the recommendations of the centralized workforce management tool and what the store managers need and could in-turn result in store managers spending considerable time overriding the decisions of the centralized planning tools as documented by van Donselaar et al. (2010) and Netessine et al. (2010).

While, at the daily level, managers seem to have the required amount of labor in the store, we find that the stores are consistently understaffed during peak hours, and quantify the impact of this understaffing on store performance. Our study also shows that decreasing forecast errors and increasing schedule flexibility would reduce understaffing and lead to higher profits for retailers. These results support the recent move by several retailers to invest heavily in emerging technologies that integrate traffic information with workforce management (Stores, Jan 2010)[1]. Future research could study the efficacy of different strategies that would aid store managers produce better forecasts of traffic and improve schedule flexibility on store performance.

In the second chapter, we have characterized the distribution of traffic to retail stores for a heterogeneous group of stores belonging to the same retail chain based on hourly traffic data obtained from traffic counters installed in each of these stores. We find that a negative binomial distribution to be a significantly better fit than the Poisson or normal distribution. Furthermore,

---

[1] Scheduled Improvements, Stores Jan 2010.

we show that the high variability associated with retail traffic can lead to poor performance of some of the commonly assumed models for retail traffic like a Poisson or a normal distribution.

We demonstrate the application of knowledge of traffic distribution to labor planning by using the traffic forecasts based on each of these models as input to staffing plans and comparing the difference between the targeted service level and resultant service levels. A recent survey showed that customers buying luxury goods typically rank service-related attributes as the basis for deciding where to shop (Booz and Hamilton, 2008). In fact almost 33% of customers state poor sales assistance as reasons for leaving without purchase (Baker, 2010). Thus, having the right forecast model that provides the required service coverage is critical to store operations for these retailers and can help prevent systemic understaffing during peak hours.

Finally, our results also provide support to some of the observations in theoretical literature on the effects of competition and other local market characteristics on traffic variability. Future research could conduct a more in-depth analysis on the conditions under which these factors influence variability and help retailers in leveraging this information in their planning strategies.

In the third chapter, we document an inverted-U relationship between abnormal inventory growth and one-year ahead earnings for retailers using publicly available financial data. To the extent that this relationship is an evidence of causality, our results imply that retailers should avoid abnormal decrease in inventory growth beyond a certain point and abnormal increase in inventory growth, other things being equal. However, caution should be exercised with this interpretation as our model does not guarantee causality in this relationship.

This study has valuable contributions to investors. We show that equity analysts do not fully incorporate the information contained in AIG in their earnings forecasts. We also show that an investment strategy based on AIG can yield significant abnormal returns. In our sample, we find this return to be -10.7%. Thus our results show that a benchmarking metric for inventory performance derived from operations management literature can serve as the basis for an investment strategy.

This dissertation is only a first step towards understanding the link between operations and financial performance for retailers. The insights developed through our empirical study have opened up several potential avenues for future research that could delve deeper into how these operational decisions may propel retailers towards meeting their objectives.

In context of store level operations, retailers can now exploit the rich information available in their transactional data bases to investigate how to provide a consistent shopping experience to their customers. While many retailers have detailed information on the effectiveness of different marketing activities in driving customers to the stores, they have surprisingly little insight into what causes a sales failure in their stores. Many store managers also report an increasing level of disconnect between corporate retail intent and store-level execution[2]. Future research that looks at integrating localized information on customer demand profiles and preferences into store operations using detailed individual store data on labor productivity, store layout and product offerings as well as service level expectations of customers patronizing their respective stores, would allow the retailers to create corporate guidelines and policies that are more in line with each individual store's requirements.

Similarly, further empirical research using data from more upcoming advanced in-store technologies that analyze customer movement and interactions with different display formats, and capture their behavior during peak and non-peak selling periods can provide store managers with a better understanding of how different operational levers around managing the different queue lengths, placement of promotional and popular items and scheduling employees at different times of the day would impact store sales and profitability.

---

[2] State of Store Manager. 2010. Annual benchmark report, Integrated Solutions for Retailers

Finally, identifying best practices around store operations and transferring them to other stores in the chain can help retailers boost their firm level operations and improve their financial performance. Benchmarking metrics from operations management enable comparison between different retailers and understand industry best practices. Future research in this regard that considers how other operational variables like product assortment and variety decisions, the service level provided at the stores and the level of competition intensity around the individual product categories impact the inventory levels and lead to development of better benchmarking metrics to compare firm performance. Similarly, empirical research into the effect of external factors like contracts between retailers and their suppliers that determine how merchandise returns are handled, presence of factory outlet stores or other mechanisms that enable retailers to salvage unsold inventory could help understand if they have a moderating influence on firm profitability.

In conclusion, this dissertation on the link between operations and financial performance for retailers is only a starting point in providing empirical evidence on many of the theoretical insights around the impact of operational decisions on firm performance that have been long studied in the operations management literature. Incorporating different sources of information using both proprietary store and firm level data can help extending this line of research into many more avenues and further explore the intricate relationship between operational decisions and financial performance that would bring in valuable insights to both academics and practitioners.

APPENDICES

6.1: Appendix I

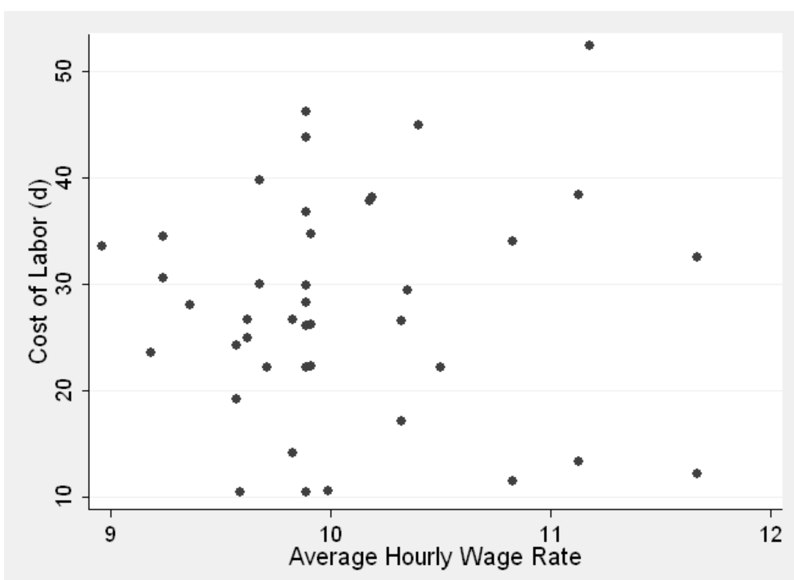
6.1.1: Individual store wise estimates of model: $S_{it} = \alpha_i \alpha_{id}{}^{\alpha_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}}$, $\gamma_i \alpha_i \alpha_{id}{}^{\alpha_d} N_{it}^{\beta_i} e^{-\gamma_i/l_{it}} = w_i l_{it}^2$

	Weekdays				Weekends			
Store	α_i	β_i	γ_i	d_i	α_i	β_i	γ_i	d_i
1	45.58	0.23	13.29	34.18	59.20	0.13	40.87	24.50
2	31.01	0.41	12.87	38.31	48.85	0.30	40.61	25.90
3	53.32	0.21	14.46	52.27	59.40	0.16	46.38	38.20
4	25.85	0.26	11.68	19.60	48.85	0.21	38.04	9.30
5	25.00	0.28	13.32	10.50	41.00	0.26	40.96	9.10
6	56.72	0.17	13.75	54.92	66.95	0.15	43.25	25.40
7	35.11	0.26	9.31	37.84	43.00	0.21	28.93	10.30
8	49.60	0.17	10.88	40.00	53.30	0.12	34.64	17.90
9	37.23	0.21	9.74	37.50	43.36	0.13	32.44	21.41
10	44.89	0.15	9.66	36.56	49.50	0.12	31.98	21.40
11	17.80	0.35	14.95	12.10	37.25	0.30	45.85	11.40
12	46.09	0.21	10.81	39.63	52.25	0.17	32.43	20.80
13	34.88	0.21	9.07	27.82	39.75	0.18	27.21	21.78
14	36.85	0.28	10.89	41.03	53.55	0.20	36.67	27.20
15	54.45	0.38	8.64	10.50	74.45	0.25	28.92	9.60
16	49.70	0.34	6.84	34.00	61.70	0.24	24.52	21.90
17	28.40	0.38	11.37	39.70	57.40	0.34	36.11	26.90
18	27.05	0.34	11.01	25.77	52.25	0.26	35.03	23.30
19	36.56	0.25	10.05	37.94	49.95	0.17	33.15	23.40
20	28.93	0.32	8.05	26.74	50.55	0.26	24.15	25.00
21	31.80	0.36	17.86	38.10	50.55	0.19	53.58	25.30
22	37.77	0.37	10.94	34.23	54.45	0.34	32.82	22.10
23	22.70	0.42	18.4	28.00	49.90	0.31	26.20	14.10
24	25.95	0.38	19.66	13.30	50.95	0.30	42.98	12.30
25	36.05	0.31	15.09	21.70	47.10	0.24	49.27	19.40
26	25.55	0.20	11.22	26.60	38.60	0.14	35.66	13.80
27	25.27	0.17	11.09	34.54	43.15	0.12	37.27	16.10
28	38.10	0.32	11.52	24.30	56.30	0.31	37.56	9.80
29	41.10	0.36	13.1	19.10	69.45	0.28	39.30	14.30

30	27.30	0.35	11.88	14.20	34.80	0.22	39.64	9.90
31	30.55	0.37	12.82	30.40	46.70	0.27	39.46	17.40
32	41.92	0.24	10.96	35.57	57.85	0.22	51.53	27.00
33	24.67	0.42	10.01	32.69	33.35	0.31	44.21	28.90
34	49.85	0.32	18.1	10.60	63.90	0.20	27.30	9.30
35	44.80	0.13	9.52	36.27	52.40	0.11	32.53	21.90
36	37.63	0.20	9.37	32.37	55.75	0.14	31.11	31.20
37	52.95	0.36	15.72	23.60	66.30	0.11	47.16	19.50
38	38.75	0.29	10.38	26.20	48.55	0.11	32.14	17.50
39	54.90	0.20	14.99	39.75	64.10	0.13	36.91	24.60
40	38.07	0.32	11.66	34.18	46.35	0.21	30.76	20.10
41	24.90	0.25	9.96	36.80	38.50	0.18	32.88	10.20

All estimates significant at $p<0.05$. The system represented by equation 4c is over-identified, as there are more exogenous variables than endogenous variables. In order to statistically test the validity of the assumed exogenous variables as instruments, we performed Hansen's over-identification restriction test (Hansen 1982). In all specifications, the validity of these variables as instruments could not be rejected as the *p*-value for Hansen's J-statistic was in excess of 0.10.

6.1.2 Scatter plot of imputed cost of labor(d_i) for weekdays against the average wage rate



6.1.3: Relaxing assumptions in GMM estimation

Here we discuss the implications of relaxing our assumption that store manager has real time information on traffic on the imputed cost of labor.

Assume that the store manager plans labor based on a forecast of traffic $\widehat{N}_{it}$ and cost of labor, say w_{1i} and store specific parameters $\alpha_i, \beta_i, \gamma_i$. The store manager's labor decision rule (analogous to equation 3) then can be written as

$$\gamma_i \alpha_i \widehat{N}_{it}{}^{\beta_i} e^{-\gamma_i / l_{it}} = w_{1i} l_{it}^2$$

Comparing this with our labor decision rule in equation 3

$$\gamma_i \alpha_i N_{it}{}^{\beta_i} e^{-\gamma_i / l_{it}} = w_i l_{it}^2$$

the error in estimation of imputed cost of labor *w* is

$$\frac{w_{1i}}{w_i} = \left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i}$$

Assuming that the error in forecast of traffic is unbiased and independent and identically distributed, let

$$\widehat{N}_{it} = N_{it} * \mu_{it}$$

Where E(μ_{it}) = 1, Then we have (assuming that the error terms are stationary),

$$E\left(\frac{w_{1i}}{w_i}\right) = E\left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i} = E(\mu_{it})^{\beta_i} = 1$$

Thus, we show that our estimates of imputed cost of labor w_i are unaffected by use of actual traffic in our estimation as long as the store manager's traffic forecast is unbiased.

Let us now consider a case where there exists a bias in store manager's forecasts of traffic. Consider two scenarios: (1) $\widehat{N}_{it} = N_{it}^{\varphi_i} * \mu_{it}$ and (2) $\widehat{N}_{it} = \varphi_i N_{it} * \mu_{it}$. (1) assumes that the bias in forecasts in increasing in the level of traffic while (2) assumes that the bias is independent of the level of traffic

$$\widehat{N}_{it} = N_{it}^{\varphi_i} * \mu_{it}$$

The store manager's labor decision is given by: $\gamma_i \alpha_i \widehat{N}_{it}{}^{\beta_i} e^{-\gamma_i/l_{it}} = w_{1i} l_{it}^2$

Comparing this with our labor decision rule in equation 3: $\gamma_i \alpha_i N_{it}{}^{\beta_i} e^{-\gamma_i/l_{it}} = w_i l_{it}^2$

The error in estimation of imputed cost of labor *w* is

$$\frac{w_{1i}}{w_i} = \left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i}$$

Let E(μ_{it}) = 1, Then we have (assuming that the error terms are stationary),

$$E\left(\frac{w_{1i}}{w_i}\right) = E\left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i} = E\left(N_{it}^{\varphi_i - 1}\mu_{it}\right)^{\beta_i} = E\left(N_{it}^{\varphi_i - 1}\right)^{\beta_i} E(\mu_{it})^{\beta_i} = E\left(N_{it}^{\varphi_i - 1}\right)^{\beta_i}$$

i.e. the error in estimation of w_i is increasing in the level of traffic N_{it} and the bias φ_i, but is moderated by the parameter $\beta_{i.}$

$$\widehat{N}_{it} = \varphi_i N_{it} * \mu_{it}.$$

The store manager's labor decision is given by: $\gamma_i \alpha_i \widehat{N}_{it}{}^{\beta_i} e^{-\gamma_i/l_{it}} = w_{1i} l_{it}^2$

Comparing this with our labor decision rule in equation 3: $\gamma_i \alpha_i N_{it}{}^{\beta_i} e^{-\gamma_i/l_{it}} = w_i l_{it}^2$

The error in estimation of imputed cost of labor *w* is

$$\frac{w_{1i}}{w_i} = \left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i}$$

Let E(μ_{it}) = 1, Then we have (assuming that the error terms are stationary),

$$E\left(\frac{w_{1i}}{w_i}\right) = E\left(\frac{\widehat{N}_{it}}{N_{it}}\right)^{\beta_i} = E(\varphi_i \mu_{it})^{\beta_i} = E(\varphi_i)^{\beta_i} E(\mu_{it})^{\beta_i} = E(\varphi_i)^{\beta_i}$$

i.e. the error in estimation of w_i is increasing in the bias φ_i and is moderated by the parameter $\beta_{i.}$

6.1.4 Simulation details

1. Compute optimal profit

In the first step, we compute the optimal profit for each store assuming full information on traffic and full scheduling flexibility as in 5.1.1. This corresponds to the benchmark case of no forecast error and hourly scheduling.

2. Generate traffic forecasts

Forecast of traffic is calculated in the following manner: For each store *i*, we run the following regression: $N_{ikt} = \tau_{oik} + \tau_{1ik}N_{ik-1t} + \tau_{2ik}t + \tau_{3ik}d + \varepsilon_{ikt}^{N}$, where N_{ikt} refers to traffic for store *i* during week *k*, in hour *t*, and *d* represents the day of week. The coefficient estimates of $\tau_{iko}, \tau_{ik1}, \tau_{ik2}$ and τ_{ik3} are used to generate the two-week-ahead traffic forecast, $\widehat{N}_{ik+2,t} = \hat{\tau}_{oik} + \hat{\tau}_{1ik}N_{ikt} + \hat{\tau}_{2ik}t + \hat{\tau}_{3ik}d$

where $\widehat{N}_{ik+2,t}$ refers to the traffic forecast for hour *t* in week *k+2*.

3. Scheduling constraints

We calculate the optimal labor given the two week-ahead traffic forecasts in presence of scheduling constraints by assuming that available labor cannot be changed within blocks of 2 hrs, 3 hrs, 4 hrs and 5 hrs (which represents half of the operating day in our sample). For e.g. with a 2 hour scheduling constraint, if the optimal labor required was 3 labor-hrs in the first hour and 5 labor-hrs in the second hour, the optimal labor plan with the scheduling constraint is 4 labor-hrs for the two hour block of time.

4. Loss from optimal

We then calculate the resultant store profits obtained for different combinations of forecast error (0 to 50%) and scheduling constraints (1 to 5 hours). These profits are compared with the optimal case where there are no forecast errors and labor is scheduled on an hourly basis to find the percentage loss in profits.

6.2 Appendix II

6.2.1 Overdispersion values for each store

	Weekdays		Weekends	
store	α_i	**p value**	α_i	**p value**
1	0.0140	0.1040	0.0325	0.0200
2	0.0015	0.1015	0.0321	0.0089
3	0.0025	0.1005	0.0316	0.0096
4	0.0033	0.1033	0.0419	0.0089
5	0.0576	0.0424	0.0481	0.0509
6	0.0563	0.1563	0.0370	0.0610
7	0.0229	0.1029	0.0216	0.0128
8	0.0018	0.1018	0.0341	0.0131
9	0.0050	0.1050	0.0409	0.0105
10	0.0150	0.1050	0.0330	0.0167
11	0.0041	0.1041	0.0427	0.0079
12	0.0223	0.1023	0.0329	0.0098
13	0.0042	0.1042	0.0449	0.0134
14	0.0029	0.1029	0.0380	0.0089
15	0.0037	0.1037	0.0480	0.0133
16	0.0041	0.1041	0.0332	0.0147
17	0.0021	0.1021	0.0307	0.0147
18	0.0049	0.1049	0.0441	0.0109
19	0.0233	0.1033	0.0330	0.0127
20	0.0039	0.1039	0.0449	0.0067
21	0.0005	0.1005	0.0432	0.0075
22	0.0002	0.1001	0.0380	0.0079
23	0.0033	0.1033	0.1083	0.0131
24	0.0021	0.1021	0.0348	0.0129
25	0.0028	0.1028	0.0462	0.0147
26	0.0047	0.1047	0.0395	0.0135
27	0.0015	0.1015	0.0398	0.0089
28	0.0595	0.0405	0.0391	0.0491
29	0.0016	0.1016	0.0469	0.0131
30	0.0019	0.1009	0.0385	0.0097
31	0.0006	0.1006	0.0374	0.0050
32	0.0441	0.0559	0.0436	0.0374
33	0.0476	0.0524	0.0596	0.0393

34	0.0515	0.0485	0.0424	0.0438
35	0.0114	0.1014	0.0559	0.0098
36	0.0040	0.1040	0.0428	0.0119
37	0.0505	0.0495	0.0481	0.0421
38	0.0049	0.1049	0.0612	0.0148
39	0.0027	0.1027	0.0498	0.0139
40	0.0477	0.0523	0.0370	0.0382
41	0.0176	0.0824	0.0394	0.0116
42	0.0019	0.1019	0.0359	0.0084
43	0.0032	0.1032	0.0353	0.0127
44	0.0032	0.1032	0.0375	0.0118
45	0.0220	0.1120	0.0435	0.0238
46	0.0111	0.1011	0.0563	0.0099
47	0.0019	0.1000	0.0409	0.0074
48	0.0317	0.1017	0.0343	0.0102
49	0.0019	0.1019	0.0360	0.0098
50	0.0005	0.1005	0.0471	0.0097
51	0.0026	0.1026	0.0460	0.0116
52	0.0009	0.1009	0.0370	0.0059
53	0.0444	0.0556	0.0392	0.0390
54	0.0469	0.0531	0.0407	0.0390
55	0.0115	0.1015	0.0372	0.0051
56	0.0033	0.1033	0.0443	0.0064
57	0.0032	0.1032	0.0658	0.0133
58	0.0048	0.1048	0.0419	0.0085
59	0.0038	0.1038	0.0454	0.0140
60	0.0034	0.1034	0.0482	0.0107

6.2.2 Forecast accuracy for weekends

	Weekends					
	One week ahead			Two week ahead		
Model	Coverage (p)	Width (%)	Accuracy	Coverage (p)	Width (%)	Accuracy
Poisson	0.42	16.45	2.55	0.38	18.35	2.07
NB	0.64	18.42	3.47	0.61	21.62	2.82
Normal	0.57	25.47	2.24	0.54	27.45	1.97
Time series (traffic)	0.51	16.87	3.02	0.48	18.15	2.65

[a] Expressed as a percentage of actual traffic

6.2.3 Forecast accuracy for weekends with seasonality factors

	Weekends					
	One week ahead			Two week ahead		
Model	Coverage (p)	Width (%)	Accuracy	Coverage (p)	Width (%)	Accuracy
Poisson	0.46	15.32	3.00	0.42	16.57	2.53
NB	0.68	17.84	3.81	0.69	20.94	3.30
Normal	0.61	22.47	2.71	0.64	27.41	2.33
Time series (traffic)	0.55	15.49	3.55	0.53	18.14	2.92

6.2.4 Sensitivity analysis of percentage deviation of actual CSR from planned CSR for different values of CSR for weekends

	One Week ahead			
Planned CSR	NB	Normal	Time Series - Traffic	Poisson
5	35.6	52.8	59.6	63.4
10	23.4	34.6	48.9	61.8
15	19.3	22.1	26.2	34.5
20	11.8	14.5	17.3	26.5

	Two week ahead			
Planned CSR	NB	Normal	Time Series - Traffic	Poisson
5	43.6	56.2	65.2	74.2
10	31.8	52.3	59.7	63.4
15	23.5	37.8	40.1	47.3
20	15.9	20.6	22.6	25.4

6.2.5 Sensitivity analysis of percentage deviation between of actual CSR from planned CSR for different levels of service coverage for weekends

	One Week ahead			
Planned Service coverage	NB	Normal	Time Series - Traffic	Poisson
90	44.2	63.8	66.8	78.2
93	42.8	58.8	62.2	75.2
95	35.6	52.8	59.6	63.4
97	28.8	41.4	45.6	59.6
99	20.2	32.6	34.2	44.6

	Two week ahead			
Planned Service coverage	NB	Normal	Time Series - Traffic	Poisson
90	59.6	73.4	86.4	92.4
93	45	66.8	78.2	84.6
95	43.6	56.2	65.2	74.2
97	31.2	51.6	51.6	71.2
99	26.8	41.2	45.2	68.6

6.3 Appendix III

6.3.1 Calculation of abnormal return using the Ibbotson-RATS procedure (or Jensen-alpha approach)

The Jensen-alpha approach to estimating risk adjusted abnormal performance is an alternative to the size-adjusted return (BHAR) calculation using a matched-firm approach to risk adjustment which has since then been advocated by Fama (1998) and Mitchell and Stafford (2000). We use this approach by employing Ibbotson's returns across time and securities (RATS) methodology here that calculates portfolio returns for firms experiencing an event, and calibrates whether they are abnormal in a multifactor (Fama-French three factor) regression as shown below[1].

$$R_{pmt} - R_{fmt} = a_{pmt} + b_{pmt}\left(R_{bmt} - R_{fmt}\right) + s_{pmt}\, SMB_{mt} + h_{pmt} HML_{mt} + e_{pt}$$

Where R_{pmt} is the value weighted return for month m for the portfolio p of event firms that experienced the event in that month, R_{fmt} is the risk free rate, R_{bmt} is the return on the CRSP value weighted market portfolio, SMB_{mt} is the difference between the return on the portfolio of "small" stocks and "big" stocks; HML_{mt} is the difference between the return on the portfolio of "high" and "low" book to market stocks; a_{pmt} is the average monthly abnormal return (Jensen's alpha) on the portfolio of event firms; $b_{pmt}, s_{pmt}, h_{pmt}$ are sensitivities (betas) of the event portfolio to the three factors. The regression is estimated for each month in the event period, i.e. 1 year period following the month after SEC filing date for the prior fiscal year. Since a_{pmt} is the average monthly abnormal performance over the 12 month post event period, it can be used to calculate annualized post-event abnormal performance.

---

[1] The Ibbotson RATS methodology is implemented using the Eventus software accessed through Wharton Research Data Services (WRDS). One of the advantages of the Eventus software is that it automatically accounts for different evaluation periods for each of the stocks due to differences in the SEC filing dates of these stocks.

REFERENCES

Abarbanell, J.S, B.J. Bushee. 1997. Fundamental Analysis, Future Earnings and Stock Prices. *Journal of Accounting Research*. **35** (1) 1-24

Abarbanell, J.S., B. J. Bushee. 1998. Abnormal returns to a fundamental analysis strategy. *The Accounting Review* **73** (1) 9-45.

Aberdeen Research Group. May 2009. Workforce Scheduling: Managerial Strategies.

Agrawal, N. and S. Smith. 1996. Estimating Negative Binomial Demand for Retail Inventory Management with Unobservable Lost Sales. *Naval Research Logistics.* **43** (1) 839-861

Aiken, Leona S. and Stephen G. West 1991. Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: Sage Publications.

Anand, K., H. Mendelson. 1997. Information and organization for horizontal multimarket coordination. *Management Science*. 43(12) 1609–1627.

Avramidis A.N., Deslauriers A. and L'Ecuyer P. (2004). Modeling daily arrivals to a telephone call center. *Management Science,* **50**(7), 896–908.

Bernard, V. 1987. Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research* **25**(1) 1-48.

Bernard, V.L., J.Noel. 1991. Do inventory disclosures predict sales and earnings? *Journal of Accounting, Auditing and Finance*. **6**(2) 145-181.

Bernstein F and Federgruen A. 2004. A General Equilibrium Model for Industries with Price and Service Competition. *Operations Research*. **52** (6) 868-886.

Bloom, N., J. Van Reenen. 2007. Measuring and explaining management practices across firms and countries, *Quarterly. Journal of Economics* **122**(4) 1351-1408.

Borucki, C.C., M.J. Burket. 1999. An Examination of Service-Related Antecedents to Retail Store Performance. *Journal of Organizational Behavior.* **20**(6). 943-962.

Booz & Company. 2008. Winning in retail with a targeted service model.

Brown L.D., Gans N., Mandelbaum A., Sakov A., Shen H., Zeltyn S., and Zhao L. (2005). Statistical analysis of a telephone call center: A queueing-science prespective. *Jasa*, **100**(469) 36–55.

Bureau of Labor Statistics. March 2009. Holiday Season Hiring in Retail Trade

Cameron, C.A, P.K. Trivedi. 1998. Regression Analysis of Count Data. *Cambridge University Press*

Carpenter, R.E., S.M. Fazzri, B.C. Petersen. 1998. Financing Constraints and Inventory Investment: A Comparative Study with High-Frequency Panel Data. *Review of Economics and Statistics, MIT Press*. **80** (4) 513 – 519.

Chatfield, C. 2000. Time-series forecasting. Boca Raton, FL: *Chapman & Hall/CRC*.

Chang, M.,J. Harrington. 2000. Centralization vs. decentralization in a multi-unit organization: A computational model of retail chain as a multi-agent adaptive system. *Management Science*. **46**(11) 1427–1440.

Chen, H., F.Z. Murray, O.Q. Wu. 2005. What actually happened to the inventories of American companies between 1981 and 2000? *Management Science.* **51**(7) 1015-1031.

Chen, H., F.Z. Murray, O.Q. Wu. 2007. U.S Retail and Wholesale Inventory Performance from 1981 to 2004. *Manufacturing & Service Operations Management.* **9** (4) 430-456.

Christen, M., G Iyer, D. Soberman. 2006. Job Satisfaction, Job Performance, and Effort: A Reexamination Using Agency Theory. *Journal of Marketing.* **70** (1) 137–150.

Cohen, M.A., T.H. Ho, J.Z. Ren, C. Terwiesch. 2003. Measuring imputed cost in the semiconductor equipment supply chain. *Management Science*. **49** (12)1653-1670.

Cooper, R., R.S. Kaplan. 1988. Measure Costs Right: Make the Right Decisions. *Harvard Business Review.* **66**(5) 96-103.

Dana J. 2001. Competition in Price and Availability When Availability is Unobservable. *The RAND Journal of Economics*. **32** (4) 497-513.

DeHoratius, N., A. Raman. 2007. Store manager incentive design and retail performance: An exploratory investigation. *Manufacturing & Service Operations. Management* **9**(4) 518–534.

Desai H., S. Rajgopal, M. Venkatachalam. 2004. Value glamour and accrual mispricing, One anomaly or two? The Accounting Review **79** (2) 355–385.

Durtschi C., Easton P. 2005. Earnings Management? The Shapes of the Frequency Distributions of Earnings Metrics Are Not Evidence Ipso Facto, *Journal of Accounting Research*. **43** (4) 557-592.

East et al., 1994. R. East, W. Lomax, G. Wilson and P. Harris. 1994. Decision Making and Habit in Shopping Times. *European Journal of Marketing* **28**(4) 56–71.

Fama, E., 1998. Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* **49** (2) 283-306.

Ferguson, M., O. Koenigsberg. 2007. How Should a Firm Manage Deteriorating Inventory? *Production and Operations Management*. **16** (3) 306 – 321.

Fisher, M. 1997. What is the Right Supply Chain for Your Product ? *Harvard Business Review* 105 – 116.

Fisher, M.L., J.Krishnan. 2005. Store level execution at Wawa. Case Study, The Wharton School, University of Pennsylvania, Philadelphia

Fisher, M.L., A. Raman. 2010. The New Science of Retailing, Harvard Business Press.

Forrester Research. April 2009.Filling the store labor productivity gap.

Fried, D. and D. Givoly. 1982. Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics.* **4**(2) 85 – 107.

Gallego, G., G. van Ryzin. 1994. Optimal Dynamic Pricing of Inventories with Stochastic Demand. *Management Science.* **40** (8) 999-1020.

Gans, N. 2002.Customer loyalty and supplier quality competition. *Management Science* **48** (1) 207-221.

Gartner Industry Research. September 2007. Retail task management and integration challenge.

Gaur, V., M. L. Fisher., A. Raman. 2005. An Econometric Analysis of Inventory Turnover Performance in Retail Services. *Management Science.* **51** (2) 181-194.

Gaur, V. and Y. Park. 2007. Asymmetric Consumer Learning and Inventory Competition. *Management Science.* **53** (2) 227-240.

Gino, F., G. Pisano. 2008. Toward a Theory of Behavioral Operations. *Manufacturing & Service Operations Management.* **10**(4) 676-691.

Givoly, D., J. Lakonishok. 1984. The Quality of Analysts' Forecasts of Earnings. *Financial Analysts Journal.* **40** (5) 40 – 47.

Greene, W. H. 2008. Econometric Analysis. 6th ed. Upper Saddle River, NJ: Prentice–Hall.

Grewal, D., J. Baker, M. Levy, G.B. Voss. 2003. The effects of wait expectation and store atmosphere evaluations on patronage intentions in service-intensive retail stores. *Journal of Retailing*. **79** 259-268.

Gu, Z., & Wu, J. 2003. Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics.* **35** (1) 5–29.

Gupta, S. D.G. Morrison. 1991. Estimating Heterogeneity in Consumers' Purchase Rates. *Marketing Science*. **10**(3) 264-269.

Hall, A.R. 2005. *Generalized Method of Moments.* Oxford University Press, USA.

Hansen, L. P., and K. J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica.* **50** (1) 1269–1286.

Hann, I., and C. Terwiesch. 2003. Measuring the Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel. *Management Science*. **49**(11) 1563-1579.

Hendricks, K.B., V.R. Singhal. 2005. Association between Supply Chain Glitches and Operating Performance. *Management Science.* **51** (5) 695-711.

Hendricks, K.B., V.R. Singhal. 2009. Demand – Supply Mismatches and Stock Market Reaction: Evidence from Excess Inventory Announcement. *Manufacturing & Service Operations Management.* **11** (3) 509 - 524.

Henly, J.R., H.L.Shafer, E.Waxman. 2006. Employer- and Employee-Driven Flexibility in Retail Jobs. *Social Service Review*. **80**(4) 609-634.

Heskett, J.L., T.O. Jones, G.W. Loveman, W.E. Jr Sasser, L.A. Schlesinger. 1994. Putting the service profit chain to work. *Harvard Business Review*. **72**(2) 164-174.

Hogarth, R.M., S. Makridakis. 1981. Forecasting and Planning: An Evaluation. *Management Science*. **27**(2) 115-138.

Hopp, W.J., S.M.R. Iravani, G.Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Science*. **53** (1) 61-77.

Horsky, D. and P. Nelson. 1996. Evaluation of Salesforce Size and Productivity Through Efficient Frontier Benchmarking. *Marketing Science*. **15**(4), 301-320.

Howrey, P. E. 2001. The Predictive Power of the Index of Consumer Sentiment. *Brookings Papers on Economic Activity.* **1**175-216

Integrated Solutions for Retailers. December 2010. Retail Tech 2010/2011: Where We've Been, And Where We're Headed from Here.

Jain, D, S.S. Singh. 2002. Customer Lifetime Value Research in Marketing: A Review and Future Directions. *Journal of Interactive Marketing.* **16**(2) 34-46.

Jongbloed G. and Koole G. (2001). Managing uncertainty in call centers using Poisson mixtures. *Applied stochastic models in bussinesss and industry*, **17** (1) 307–318.

Kahn, B E., D. C. Schmittlein. 1989. "Shopping Trip Behavior: An Empirical Investigation *Marketing Letters*, **1**(12) 55-70.

Kahneman, D., D. Lovallo. 1993. Timid choices and Bold Forecasts. *Management Science*.**39** (1) 17-31.

Kesavan, S. V. Gaur., A. Raman. 2010. Incorporating Price and Inventory Endogeneity in Firm Level Sales Forecasting. *Management Science.* **56** (9) 1519 - 1533

Kothari, S. P., & Warner, J. B. 2007. Econometrics of event studies. Handbook of Corporate Finance: Empirical Corporate Finance. North Holland: Elsevier.

Kronos. September 2006. A workforce optimized, Integrated solutions for retailers.

Lai, R. 2006. Inventory Signals, Harvard NOM Research Paper Series No. 05-15. Boston, MA: Harvard Business School.

Lam, S.Y, M. Vandenbosch, M. Pearce. 1998. Retails sales force scheduling based on store traffic forecasting. *Journal of Retailing*. **74**(1) 61-88.

Lam, S.Y., M. Vandenbosch, J. Hulland, M. Pearce. 2001. Evaluating promotions in shopping environments: Decomposing sales response into attraction, conversion, and spending effects. *Marketing Science.* **20**(2) 194–215.

Lambert, S. 2008. Passing the buck: Labor flexibility practices that transfer risk onto hourly workers. *Human Relations.* **61**(9), 1203-1227.

Lariviere, M., J.A. Van Mieghem. 2004. Strategically Seeking Service: How Competition Can Generate Poisson Arrivals. *Manufacturing & Service Operations Management*. 6 (1) 23 – 40.

Lee H.L., V. Padmanabhan, S. Whang. 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*. **43** (4) 546-558.

Lodish, L., E. Curtis, M. Ness, M.K. Simpson. 1988. Sales Force Sizing and Deployment Using a Decision Calculus Model at Syntex Laboratories. *Interfaces*. **18**(1) 5-20.

Loveman. G. 1998. Employee, satisfaction, customer loyalty, and financial performance: An empirical examination of the service profit chain in retail banking. *Journal of Service Research*. **1**(1) 18-31.

Lu, Y, M. Olivares, A. Musalem, A. Schilkurt. 2011. Measuring the effect of queues on customer purchases. Working paper. Columbia University.

Lundholm, R., S. McVay. 2010. Forecasting Sales: A Model and Some Evidence from the Retail Industry. Working paper, University of Michigan.

Maddala, G.S. 2001. Introduction to Econometrics. John Wiley & Sons, New York.

Makridakis, S., and S. C. Wheelwright. 1987. The Handbook of Forecasting: A Manager's Guide. John Wiley & Sons, New York.

Maxham, J.G. III, R.G. Netemeyer, D.R. Lichenstein. 2008. The Retail Value Chain: Linking Employee Perceptions to Employee Performance, Customer Evaluations, and Store Performance. *Marketing Science.* **27**(2) 147 – 167.

Netessine, S., M. L. Fisher, J.Krishnan. 2010. Labor Planning, Execution, and Retail Store Performance: an Exploratory Investigation, Working Paper, The Wharton School, University of Pennsylvania.

NRF, 2010. There just may be a silver bullet.

Oliva, R., J.D. Sterman. 2001. Cutting corners and working overtime: quality erosion in the service industry. *Management Science*. **47**(7) 894-914.

Olivares, M., C. Terwiesch, L. Cassorla. 2008. Structural Estimation of the Newsvendor Model: An Application to Reserving Operating Room Time. *Management Science*. **54**(1) 41-55.

Olivares, M. G. Cachon. 2009. Competing Retailers and Inventory: An Empirical Investigation of General Motors' Dealerships in Isolated U.S. Markets. *Management Science.* **55** (9) 1586 – 1604.

Park,Y., C.H. Park, V. Gaur. 2010. Consumer Learning, Word of Mouth, and Quality Competition. Working Paper, The Johnson School, Cornell University.

Pierson, A., G. Allon, A. Federgruen. 2010. Does it pay to reduce your customers' wait? An empirical industrial organization study of fast-food drive-thru industry based on structural estimation methods. *Forthcoming in Manufacturing & Service Operations Management.*

Perdikaki, O., S Kesavan, and J. Swaminathan. 2010. Effect of retail store traffic on conversion rate and sales. Working Paper, University of North Carolina – Chapel Hill.

Png, I.P.L., D. Reitman. 1994. Service time competition. *The RAND Journal of Economics*. **25**(4) 619-634.

Quan, V. 2004. Retail Labor Scheduling. *OR/MS Today,* December 2004.

Rajagopalan, S. 2010. Factors driving inventory levels at US retailers. *Working paper.*

Raman, A., V. Gaur., S. Kesavan. 2005. David Berman. *Harvard Business School Case 605-081*.

Ren, J., S. Willems, 2009. An empirical study of inventory policy choice and inventory level decisions. Working paper, Boston University, Boston.

RedPrairie, 2010. A New Approach to Retail Workforce Forecasting.

Reinartz, J.W., V. Kumar. 1999. Store-, Market-, and Consumer – Characteristics: The Drivers of Store Performance. *Marketing Letter.* **10**(1) 5-22

Roy, A. 1994. Correlates of Mall Visit Frequencies. *Journal of Retailing.* **70**(3) 139-161.

Rumyantsev, S., S. Netessine. 2007. What Can be Learned from Classical Inventory Models? A Cross-Industry Exploratory Investigation. *Manufacturing & Service Operations Management*. **9** (4) 409 – 429.

Ryski, M. 2005. When Retail Customers Count: How understanding customer traffic patterns can help good retailers become great retailers. *AuthorHouse Publication.*

Schweitzer, M.E., G.P. Cachon. 2000. Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*. **46**(3) 404-420.

Shen, H. & Huang, J.Z. (2008). Interday forecasting and intraday updating of call center arrivals. *Manufacturing & Service Operations Management.* **10(**3): 391–410.

Siebert, W.S. and N. Zubanov. 2010. Management Economics in a Large Retail Company. *Management Science*. **56** (8) 1398-1414.

Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* **71** (1) 289–315.

Smith, S., D. Achabal. 1998. Clearance Pricing and Inventory Policies for Retail Chains. *Management Science.* **44** (3) 285-300.

Steckley, S., S. Henderson, V. Mehrotra. 2009. Forecast errors in service systems. *Probab. Eng. Inf. Sci.* **23** (2) 305–332.

Stickney, C.P., R.L. Weil. 2003. Financial Accounting: An Introduction to Concepts, Methods, and Uses. Thomson, South-Western.

Stores. Jan 2010.Scheduled Improvements.

Tanir, O. and Booth, R. J. (1999). Call center simulation in Bell Canada*. In Proceedings of the 1999 Winter Simulation Conference* 1640–1647. IEEE Press, Piscataway, NJ.

Thomadsen, R. 2005. The Effect of Ownership Structure on Prices in Geographically Differentiated Industries. *The RAND Journal of Economics*. **36**(4) 908 – 929

Thomas, J.K., H. Zhang. 2002. Inventory Changes and Future Returns. *Review of Accounting Studies.***7** (2) 163 – 187.

Tomax, 2009. Tomax Partners with ShopperTrak to Enhance Workforce Solution.

Ton, Z. 2009. The effect of labor on profitability: The role of quality. Working Paper, Harvard Business School.

Ton, Z. and R.S. Huckman. 2008. Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science.* **19**(1) 56-58.

van Donselaar, K.H, V. Gaur, T. van Woensel, R.A.C.M. Broekmeulen and J.C. Fransoo. 2010. Ordering Behavior in Retail Stores and Implications for Automated Replenishment. *Management Science*. **56**(5) 766–784.

Walters, R. G., S. B. MacKenzie.1988. A structural equations analysis of the impact of price promotions on store performance. *J. Marketing Res*. **25**(1) 51-63.

Walters, Rockney G. and Heikki J. Rinne. 1986. An empirical investigation into the impact of price promotions on retail store performance. *Journal of Retailing.* **62** (3), 237–266.

Wernerfelt, B. 1994. On the Function of Sales Assistance. *Marketing Science* **13**(1) 68-82.

Zeithmal, V.A., L.L. Berry, A. Parasuraman. 1996. The Behavioral Consequences of Service Quality. *Journal of Marketing* **60**(2), 31-46.

Zeithaml, V.A. 2000. Service quality, profitability and the economic worth of customers; what we know and what we need to learn. *Journal of the Academy of Marketing Science.* **28**(1) 67-85.