Estimating the impact of understaffing on sales and profitability in retail stores

Vidya Mani
Smeal College of Business, Pennsylvania State University, State College, PA 16802
vmani@psu.edu

Saravanan Kesavan
Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
skesavan@unc.edu

Jayashankar M. Swaminathan
Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
msj@unc.edu

Abstract

In this paper we use hourly data on store traffic, sales, and labor from 41 stores of a large retail chain to identify the extent of understaffing in retail stores and quantify its impact on sales and profitability. Using an empirical model motivated from queueing theory, we calculate the benchmark staffing level for each store, and establish the presence of systematic understaffing during peak hours. We find that all 41 stores in our sample are systematically understaffed during a 3-hour peak period. Eliminating understaffing in these stores can result in a significant increase in sales and profitability in these stores. Also, we examine the extent to which forecasting errors and scheduling constraints drive understaffing in retail stores and quantify their relative impacts on store profits for the retailer in our study.

Key words: retail staffing, data analytics, store performance

To appear in POM journal
1. Introduction

The ability to match store labor with variable customer demand in a timely and cost-effective manner is an important driver of retail store performance. The right amount of labor in the store leads to a shopping experience that is satisfying for customers and profitable for retailers. Store labor drives 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). While additional store labor can increase store profits through increased sales, it can reduce store profits due to increased expenses as well. Labor related expenses account for a significant portion of a store’s operating expense (Ton, 2009). Hence, retailers have to walk a fine line between balancing the costs and benefits of store labor in order to maximize their profits.

While retailers invest heavily in store technologies to ensure that there is the right amount of labor in the store at the right time, it is unclear to what extent they 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 (Baker Retail Initiative, May 2007). Such statistics highlighting lost sales opportunities due to understaffing can be vexing for retailers as they spend a substantial amount of their budget on marketing activities to draw customers to their stores. While substantial agreement exists that understaffing would result in lower store performance, the extent of understaffing in retail stores has not been studied rigorously.

We focus on understaffing for the following reasons. First, studies have shown that understaffing could lead to poor service quality that can result in lower customer satisfaction (Oliva and Sterman 2001). Dissatisfied customers may switch to competitors resulting in a loss of lifetime value from those customers (Heskett et al. 1994). In addition, such customers may express their dissatisfaction 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). Second, understaffing has been found to be negatively associated with store associate satisfaction (Loveman 1998) and decline in employee satisfaction has been found to be linked to decline in store’s financial performance (Maxham et al. 2008). Yet prior literature has not documented the extent of understaffing in retail stores. So, it is unclear whether understaffing
exists for reasons beyond the normal fluctuations in demand (customer arrivals) and supply (labor). Hence, we examine whether systematic understaffing exists in retail stores, and if so, determine the impact of understaffing on store performance.

We perform this analysis using data collected from 41 stores of a large specialty apparel retailer (Alpha) over a one-year period. The identity of the retailer is disguised to maintain confidentiality. An important attribute of our data is the availability of hourly traffic data, collected from traffic counters installed at this retailer’s stores, that allows us to estimate the extent of understaffing in this retailer’s stores at a micro level. We combine traffic data with point-of-sale (POS) data on sales and transactions, and labor data, to estimate the extent of understaffing in these stores and estimate its short-term impact on sales and profitability.

We follow two different approaches to identify understaffing and study its impact. The first approach uses reduced-form estimation of an empirical model, motivated by queueing theory, to generate benchmark staffing levels. Hours when actual labor is less than model predicted labor are identified as periods of understaffing in a store and the magnitude of those deviations capture the amount of understaffing during those hours. We then quantify the impact of understaffing on lost sales and profits. In the second approach, we obtain benchmark staffing levels by replacing predicted staffing levels with optimal staffing levels obtained from a structural estimation approach.

We have the following results in our paper. We find that stores are understaffed 40.21% of the time and the average magnitude of understaffing is 2.10 persons (or 33.27% of the predicted labor). We also find that this understaffing is not driven by randomness in demand or supply factors as all the 41 stores in this chain exhibit systematic understaffing during a three-hour peak period. We show that understaffing during peak hours is significantly correlated with decline in conversion rate (defined as the ratio of number of transactions to incoming traffic), and we estimate the impact of understaffing on lost sales to be 8.56%, and that on profitability to be 7.02% in our sample. In other words, if this retailer were able to completely eliminate understaffing in its stores, then its sales and profits would have been higher by 8.56% and 7.02% respectively. We find that understaffing during peak hours is associated with a 1.95% drop in conversion rate for the stores in our sample.

Next, we quantify the impact of forecast errors and scheduling constraints on understaffing. Both factors limit the ability of retailers to reduce understaffing and consequently the
profitability improvement that may be achieved from such reductions is bounded. For example, we find that the profitability improvement achievable with understaffing reduction diminishes from 7.02% in the case of perfect foresight to 4.46% when traffic is forecasted a week in advance. Similarly, we find that the profitability improvement that retailers may achieve through understaffing reduction diminishes by 2.74% when minimum shift length increases from 2 to 4 hours.

This paper makes the following contributions to the growing research on retail operations. First, while prior literature has studied the relationship between labor and financial performance of retail stores (e.g., Fisher et al. 2007; Ton and Huckman 2008; Netessine et al. 2010; Perdikaki et al. 2012), we document the extent of understaffing in retail stores, as well as explore the effect of common drivers of understaffing like forecasting errors and scheduling problems. Second, there is a long line of work in operations management that has employed queueing theory for staffing decisions (Gans et al. 2003; Green et al. 2007). Implicit in these papers is the assumption that understaffing is costly to organizations. As far as we are aware, empirical evidence supporting this assumption have been absent in the retail sector. Our paper documents the impact of understaffing on lost sales and profitability for 41 stores of one retail chain. Third, prior theoretical operations management literature has suggested that forecast errors and scheduling constraints result in understaffing (He et al. 2012). Our paper is the first to quantify the impacts of these two factors on understaffing, lost sales, and profitability in a retail setting.

Our paper has the following managerial implications. First, the traffic counting technology is a nascent one and many retailers are still in the process of evaluating the value of this technology for labor planning. We show that staffing decisions based on traffic can help identify instances of systematic understaffing in the store and quantify the potential for improvement in sales and profits by removing understaffing during those hours. Second, many retailers cite a need for sophisticated software to produce accurate forecasts as one of the most critical components of store operations. Our simulation experiment shows that although accurate forecasts are valuable, they alone would not help retailers to significantly increase store profitability. In addition to investing in centralized technologies that can improve forecasting, retailers also need to increase the flexibility of their workforce to achieve the maximum sales lift and profitability improvement.
2. Literature Review

Labor planning problems have long been studied in operations management. Starting with the seminal papers by Dantzig (1954) and Holt et al. (1956), several papers have developed mathematical models to improve staffing decisions. The objective of these papers is to minimize costs by minimizing the level of over- and understaffing. A popular staffing model based on queueing theory and used in a variety of service settings like call centers and hospitals is the square-root staffing model. This model is easy to implement and has been shown to achieve the desired balance between operations efficiency and service quality in call center settings (Borst et al. 2004). Our paper contributes to this literature by using an empirical model based on the square-root staffing model in a retail setting to quantify the extent of understaffing in retail stores and its impact on lost sales and profitability.

Empirical research examining the impact of labor on retail store performance has been gaining importance in the recent years. Several researchers have examined the impact of labor on store financial performance. Using data from a small-appliances and furnishing retailer, Fisher et al. (2007) find that store associate availability (staffing level) and customer satisfaction are among the key variables in 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, and demonstrate a negative association between labor mismatches at the stores and basket value. 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, Ton (2009) finds evidence that increasing labor leads to higher store profits primarily through higher conformance quality. Our paper differs from the above in its research question, data, and methodology. We investigate the prevalence of understaffing using hourly data on traffic, labor, and sales, and quantify its impact on store profitability.

While numerous papers have utilized traffic data on incoming calls to study labor issues in the call center literature, the lack of traffic data has stymied research in labor issues faced by brick-and-mortar retailers. Lam et al. (1998), Lu et al. (2013), and Perdikaki et al. (2012) are notable exceptions. Lam et al. (1998) study sales-force scheduling decisions based on traffic forecast. However, they have data from only one store. Lu et al. (2013) use video-based
technology to compute the queue length in front of a deli counter at a supermarket and show that consumers’ purchase behavior is driven by queue length and not waiting time. In contrast, we use panel data from 41 stores to identify the extent of understaffing in each of these stores and study its impact on lost sales and profitability. We augment the result from the reduced-form regression with structural estimation, where we allow the cost of labor to vary across stores. Using results from the structural estimation, we show that the imputed cost of labor used by store managers is different from the accounting cost used in previous literature, that this cost can vary significantly across stores, and that it is driven by local market characteristics like competition, median household income, and availability of labor.

Our paper is closest to Perdikaki et al. (2012) who characterize the relationships between sales, traffic, and labor for retail stores. They use daily data to show that store sales have a concave relationship with traffic; conversion rate decreases non-linearly with increasing traffic; and labor moderates the impact of traffic on sales. Our paper differs from Perdikaki et al. (2012) in its research question. We examine whether retail stores are understaffed, and if so, what is the impact of this understaffing on lost sales and profitability for these retail stores. Unlike Perdikaki et al. (2012) who examine the relationships at a daily level, we use hourly data to perform a micro-analysis of the extent of understaffing within a day.

3. Research Setting

We obtained proprietary store-level data for Alpha, a women’s specialty apparel retail chain. As of 2012, there were 205 Alpha stores operating in 36 states in the United States, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Canada. These stores are in high-traffic locations like regional malls and shopping centers.

Alpha’s stores are typically less than 3000 sq. ft in size with small backrooms. 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. To emphasize the sales nature of their jobs, these associates are also called stylists. In line with the trendy clothes sold by this retailer, the job requirement states that sales associates need to be fashion forward and should maintain their appearance in a way that represents their brand in a professional and fashionable manner.
We illustrate the typical labor planning process for retailers in Figure 1. First, retailers determine the sales forecast for each store. The sales forecast is generated at an aggregate monthly or weekly level. The sales forecast for a store would take into account, among other factors, store specific characteristics, sales trends, new product introductions, seasonality, and upcoming promotions. Next, these sales forecasts are used to determine the aggregate number of labor hours required in a store. Once the aggregate labor hours are determined, workers’ schedules are created after taking into account several constraints such as worker availability, minimum shift length constraints, and other government and labor union constraints (Quan 2004).

While the labor planning process described above is a typical one, there are wide variations in how this process gets executed. Many large big-box retail organizations use centralized labor-planning tools from companies such as Kronos and RedPrairie to perform these planning activities and the store manager is only responsible for ensuring the compliance to labor plan. On the other hand, smaller retailers such as Alpha use a centralized planning tool that generates daily sales and traffic forecasts for each store. The store managers use these data as inputs to generate the number of labor hours required for each day. This involves the managers’ judgment around how many labor hours would be required to support sales activities in their stores. Because store managers’ bonuses are tied to store profits, they have strong incentives to control payroll expenses in their stores. Thus store managers strive to increase sales while controlling payroll costs.

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 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 to be included in the traffic count, 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. The hourly traffic data along with performance metrics such as
sales volume, conversion rate, basket value (defined as the ratio of sales volume to number of transactions), and the labor in the store were available to corporate headquarters as well as store managers at periodic intervals. We use the same data for our analysis.

*Alpha’s* stores were open 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 are of similar sizes, and are located across 17 states in the U.S. in regional malls and shopping centers.

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 do not possess information on inventory levels and promotions.

4. **Methodology and Estimation**

In this section, we explain the methodology used to identify the extent of understaffing in retail stores and its impact on store sales and profitability. To determine the extent of understaffing, we need to first determine the appropriate benchmark staffing levels so that the deviations from those levels can be used to identify understaffing (and overstaffing). We obtain the benchmark staffing levels using two approaches. The first approach uses reduced-form estimation of an empirical model to obtain predicted staffing levels and the second approach uses a structural estimation methodology to obtain optimal staffing levels.

Each of these approaches has advantages as well as disadvantages. The reduced-form approach is advantageous as it does not make strong assumptions about managerial behavior and merely captures how managers changed store labor based on traffic. Such a rationale is similar to that of Bowman (1963). According to Bowman (1963), experienced managers are aware of the criteria of a system and the system variables that influence these criteria but might implicitly
operate decision rules that relate the variables to the criteria imperfectly. The inconsistency in managerial actions is measured by first calibrating a model to capture managers’ past actions and then measuring the deviations between model predictions and managerial actions. This approach has been followed in other papers such as van Donselaar et al. (2010) who show that a decision rule based on store manager’s ordering behavior can improve over actual performance. The disadvantage of this approach is that it is only possible to examine the improvement in sales and profitability when stores deviate from model predictions but not the total improvement possible by carrying the optimal labor. On the other hand, the structural estimation approach allows us to determine the optimal labor so we can estimate the impact of deviating from optimal labor on store sales and profitability. However, this approach is disadvantaged by the strong assumptions of managerial optimality required to estimate the model using structural estimation techniques.

Next we explain the first approach in detail and present the results from this approach in §5. We discuss the second approach in §8.

4.1 Benchmark staffing levels from reduced-form estimation

In this approach, we build a regression model that can be used to predict how managers make labor decisions based on store traffic. In order to identify an appropriate model specification for the reduced-form approach, we compared three different models: linear expectation model, log-linear expectation model and the square-root expectation model (a model with linear and square-root terms). This last model specification was motivated by queueing theory.

Queueing theory based staffing models have been used in service settings to balance the cost of sustaining a service level standard with the cost of personnel. Retail stores have been modeled in prior literature (Berman and Larson 2004; Berman et al. 2005) as an Erlang C (or Erlang A) queue to calculate the required staffing level subject to service level constraints. In order to use these models directly for empirical estimation, we need detailed information on the arrival rates, service rates, probability of abandonment and rate of abandonment, as well as threshold waiting time, so that we can determine the exact relations between these parameters. Such data on customer waiting time and abandonments are hard to obtain in a retail setting (Allon et al. 2011; Lu et al. 2013). In order to leverage a queueing theory based staffing model for a retail store, we need a model that would require minimal data while being robust to changes in system parameters. The square-root staffing model is such a model; it has the advantages of being derived from queueing theory yet requiring minimal data for its estimation. The advantages of
this model are that it has been shown to support simple and useful rules for staffing (Borst et al.
2004), has been used in practice in different service settings such as workforce planning in call
centers, hospital staffing and police roster planning, and found to be robust to various system
specifications (Green et al. 2007). Hence, we use this model to motivate the square-root
expectation model in the reduced-form regression.

We performed Wald’s Chi-squared goodness-of-fit tests to compare the fit of the three
reduced-form models to our data. In addition, we also tested for model-fit by comparing the
standard error of prediction from each of the three reduced-form models. We found that of the
three reduced-form models, the square-root expectation model had the lowest standard error of
prediction and provided the best model-fit to our data. Hence we use the square-root expectation
model as our primary reduced-form model. Details on model-fit tests for the three reduced-form
models are provided in Appendix A3. Next, we explain in detail the reduced-form estimation
with the square-root expectation model.

There are several factors that could cause the relationship between labor and store traffic to
vary between peak and non-peak hours. For example, higher traffic during peak hours could lead
to higher congestion levels and lower customer service if there are not enough sales associates to
serve the customers. On the other hand, it is possible that workers exert greater effort during
peak hours to manage the service requirements (Kc and Terwiesch 2009). In our setting we find
that almost 60% of the daily traffic arrives during a 3-hour window. We call this 3-hour window
as the peak hours for a store. Some retailers also call it the power hours for a store. In order to
account for peak-hour effects, we include a dummy variable for peak hours in the expectation
model. Also, prior empirical research has suggested including an intercept to capture minimum
labor requirements in the store (Noteboom et al. 1982; Frenk et al. 1991). With these
modifications, we obtain the following expectation model for labor planning for each store $i$ in
time period $t$:

$$N_{it} = \delta_{0i} + \delta_{1i} \lambda_{it} + \delta_{1p} \lambda_{it} * (1_{ip}) + \delta_{2i} \lambda_{it}^{1/2} + \delta_{2p} \lambda_{it}^{1/2} * (1_{ip}) + \xi_{it}$$

(1)

In (1), $N_{it}$ is the staffing level, $\lambda_{it}$ the store traffic and $1_{ip} = 1$ when peak hour $p = 1, 0$ o/w. Let
$\hat{N}_{it}$ denote the predicted labor from (1). We calculate the deviation ($\Delta N_{it}$) of actual labor ($N_{it}$)
from predicted labor ($\hat{N}_{it}$). Positive values of the deviation indicate understaffing while negative
values indicate overstaffing.
Next we explain the sales and profit models that we use to calculate the impact of understaffing on lost sales and profits.

4.2 Calculation of lost sales and profit

Sales model

From queueing 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. Theoretical literature in service settings has assumed a concave relationship between revenue and labor (Horsky and Nelson 1996; Hopp et al. 2007). Additionally, in a retail setting, it has often been observed that sales increase at a decreasing rate with traffic. Some causes for this include the negative effects of crowding on customers and not having enough labor to satisfy customer service requirements (Grewal et al. 2003). These insights are reflected in recent empirical research as well (Fisher et al. 2007; Perdikaki et al. 2012). The following modified exponential model, adapted from Lam et al. (1998), captures these relationships between store sales ($S_{it}$), store traffic ($\lambda_{it}$), and number of sales associates ($N_{it}$) in store $i$ in time period $t$:

$$S_{it} = \alpha_{i}\lambda_{it}e^{-\gamma_{i}/N_{it}}$$

where $\beta_{i}$ is the traffic elasticity, $\gamma_{i}$ captures the responsiveness of sales to labor (indirectly measuring labor productivity), and $\alpha_{i}$ 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. $0 < \beta_{i} < 1$, $\gamma_{i} > 1$. We calculate the sales lift for each store $i$ in each time period $t$ that our model identifies the store to be understaffed as follows. Let $1_{\Delta N_{it} > 0}$ be an indicator function which takes the value of 1 when the store is understaffed ($\Delta N_{it} > 0$), 0 otherwise. Then the lost sales in time period $t$ when the store is understaffed can be represented as:

$$\Delta S_{it} = \left[\hat{\alpha}_{i}\hat{\lambda}_{it}\hat{\beta}_{i}(e^{-\hat{\gamma}_{i}/\hat{N}_{it}}) - S_{it}\right] * (1_{\Delta N_{it} > 0})$$

In (3) “$\hat{\cdot}$” indicates the coefficients estimated from (2). Intuitively, our estimation of lost sales is based on the sales lift that the store would have experienced if it carried the predicted labor ($\hat{N}_{it}$) from (1).
**Profit model**

Because our estimation of lost sales assumes that the stores would increase their labor, we also consider the impact on profit that accounts for the increase in cost due to the increase in labor. Assuming a linear cost function for labor, we obtain the following profit function:

\[
\pi_{it} = S_{it} \times g_{it} - N_{it} \times d_{i}
\]

where \(\pi_{it}\) is the gross profit net of labor costs, \(S_{it}\) is the overall dollar value of sales, \(g_{it}\) is the gross margin, \(N_{it}\) is the number of salespeople, and \(d_{i}\) is the marginal cost of labor. Similar profit functions have been used in prior literature (Lodish et al. 1988). This profit function may be used to determine the impact of understaffing on profitability as:

\[
\Delta\pi_{it} = (\Delta S_{it} \times g_{it} - \Delta N_{it} \times d_{i}) \times (1_{\Delta N_{it} > 0})
\]

### 5. Estimation Results

In this section we report our main estimation results. In §5.1 we describe the sampling procedure and in §5.2 we provide the estimation details.

#### 5.1 Sampling procedure

Our data set consists of hourly observations for each of the 41 stores from January to December, 2007. The retail industry displays significant seasonality in traffic patterns during the year (BLS, 2009) and the traffic pattern also varies considerably between weekdays and weekends. Such variations in traffic could be driven by changes in customer profile visiting the stores (Ruiz et al. 2004). Since this may result in differences in parameter estimates across time periods, we identify sub-samples in our dataset where we expect these parameters to be similar using hierarchical clustering analysis (Punj and Stewart 1983).

Hierarchical clustering begins with each observation being considered as a separate group (N groups each of size 1). The closest two groups are combined (N - 1 groups, one of size 2 and the rest of size 1), and this process continues until all observations belong to the same group. This process creates a hierarchy of clusters. We use the average linkage method (Kaufman and Rousseuw 1990) to cluster days-of-week and month-of-year observations on traffic. The results of the hierarchical cluster analysis based on mean traffic for each weekday and mean traffic for each month for a representative store are shown in Figures 2a and 2b. We found similar results with a hierarchical cluster analysis based on mean sales in place of mean traffic. These patterns were observed for rest of the stores in our sample as well. As shown in Figure 2a, there are two different clusters based on different days of the week; the first cluster corresponding to days of
week, Monday-Thursday, and the second cluster corresponding to the days of week, Friday-Sunday. Based on the different months of the year, as shown in Figure 2b, we observe two clusters, the first cluster consisting of months of January–November, and the second cluster with the month of December.

Since we did not have sufficient observations in December to treat it as a separate subsample, we drop data from this month for the rest of our analysis. Next, we create two subsamples using data from January-November. The *weekdays* sample comprises of data from Monday, Tuesday, Wednesday, and Thursday and the *weekends* sample comprises of data from Friday, Saturday, and Sunday. At this stage, we have 190 days in the weekdays sample and 143 days in the weekends sample for each store. We use the following notations: for store \( i \), on day \( d \) and hour \( h \), \( S_{i,dh} \) denotes the dollar value of sales, \( N_{i,dh} \) denotes the number of sales-persons per hour in the store, and \( \lambda_{i,dh} \) denotes the store traffic or number of customers entering the store while \( T_{x,ih}, C_{R,ih} \) and \( B_{V,ih} \) denote the number of transactions, conversion rate and basket value respectively. After removing outliers based on top and bottom 5 percentile of sales and traffic, we had a total of 73,800 hourly observations for weekdays and 53,300 hourly observations for weekends. All further analyses were conducted on these two datasets. Table 1 gives the summary statistics of all the above store-related variables for both samples.

5.2 Estimation results

We calculate the predicted labor for each store for each hour in the following manner. Since we have data for one year, we use hourly data on traffic, sales, and labor from the first six months to estimate (1). Then we use the coefficient estimates of \( \delta_{0,i}, \delta_{1,i}, \delta_{1,p}, \delta_{2,i}, \delta_{2,p} \) to compute the predicted labor \( \tilde{N}_{it} \) for each hour for the seventh month. We repeat this process six more times to calculate predicted labor for the remaining months on a rolling horizon basis by shifting the six month window by one month each time.

We convert (2) into a log-log form as shown in (6) before estimating the parameters. Using \( \sim \) to denote the logarithm of the variables \( S_{it}, \alpha_i, \) and \( \lambda_{it} \), our sales equation is:

\[
\tilde{S}_{it} = \tilde{\alpha}_i + \beta_i \tilde{\lambda}_{it} - \frac{\gamma_i}{N_{it}} + \xi_{2it}
\]  

The parameters \( \alpha_i, \beta_i \) and \( \gamma_i \) are estimated on a rolling horizon basis as above by holding a six month fixed window in our estimation sample. The estimation of (6) deserves further explanation. The error term in (6) captures the statistical fluctuations in sales across stores and time periods. There are many factors that could affect sales in the stores through changes in
conversion rate and basket value that are unobservable to the researcher. Some of these factors include promotional activities and weather changes. Promotional activities could bring more customers into the store, induce more customers to make a purchase and/or increase basket size. Similarly, weather changes could influence customer purchasing behavior. So, we use the error term in (6) to capture the changes in sales due to unobservable factors.

We note that store promotions could cause endogeneity in our setting as contemporaneous labor and traffic could be correlated with the error term. For example, when promotions are planned in advance, store managers could add additional temporary labor to meet staffing requirements for a short period of time. To overcome this endogeneity problem, we follow Judge et al. (1985) to use lagged values of labor and traffic as instrument variables. Lagged labor has been used as an instrument variable in retail settings in prior studies as well (Siebert and Zubanov 2010; Tan and Netessine 2012). Specifically, we use labor and traffic that are lagged by 7 days as instruments in estimation of (6) and traffic lagged by 7 days as instrument in estimation of (1). These serve as appropriate instrument variables since they will be correlated with contemporaneous labor and traffic respectively, but will be uncorrelated with contemporaneous error terms. The estimation is done using GMM (generalized method of moments) with a weighting matrix that accounts for any heteroskedasticity and autocorrelation effects that might be present in the data. The estimation results for the weekdays and weekends sample are summarized in Table 3.

We find significant difference in the parameter estimates obtained from the weekdays and weekends sample for each store. The average traffic elasticity ($\beta_i$) and the responsiveness of sales to labor ($\gamma_i$) were found to be lower during weekends as compared to weekdays ($p<0.1$) for each of the 41 stores. Since $\gamma_i$ captures the responsiveness of labor to sales, our analysis provides an estimate of the marginal impact on sales of an additional staff person. For example, based on the average value of $\gamma_i$ from table 3, we find that for a given level of traffic, increasing the labor by 1 person increases sales by 28.5% on average in the weekdays sample. The increase in sales could be due to additional sales associates increasing conversion among shoppers or by sales associates having more opportunities to increase basket value through upselling and cross selling activities.

It is possible that autocorrelation of demand shocks could cause our instrument variable to be correlated with our dependent variable. We test if this might be a concern in our estimation in
two different ways. First, we followed the method proposed in Tan and Netessine (2012) who suggest controlling for trend in the model to overcome this problem. Hence, we re-estimate our parameters after controlling for trend ($T_i$) in (6). We find the difference between the estimates of the parameters from the two models (with and without trend) to be insignificant. Second, for a smaller sample, we use traffic of a co-located retailer ($\textit{Gamma}$) as an instrument. $\textit{Gamma}$ belongs to the same NAICS code as $\textit{Alpha}$, but is a family clothing store and hence the merchandise at $\textit{Gamma}$ is a different assortment than at $\textit{Alpha}$. We obtain similar estimates from use of this instrument as well. The estimation results based on these two tests are reported in appendix A1.

6. Results

In this section we describe our results on the extent of understaffing observed in retail stores and its impact on lost sales and store profitability.

6.1 Extent of understaffing in the retail stores

We have 33620 total store-hours in our weekdays test sample and 20664 total store-hours in our weekends test sample. We describe results here for the weekdays test sample but find qualitatively similar results for the weekends test sample as well. The summarized results for all 41 stores are presented in Table 4. Results for each individual store are presented in appendix A2. As shown in Table 4, we find that stores are understaffed 40.21% of the time. When understaffing occurs, the magnitude of understaffing is 2.10 persons. In other words, the stores were short, on average, by 2.10 persons when there was understaffing. This level of understaffing represents a 33.27% shortage compared to the predicted labor.

Further investigation reveals that peak hours account for most of the understaffing in a store. During peak hours, we find that the stores of $\textit{Alpha}$ are understaffed 64.98% of the time and the average magnitude of understaffing is 2.31 persons (Table 4). In figure 3 we plot the actual and predicted labor during peak and non-peak hours across the 41 stores to depict the widespread prevalence of understaffing during peak hours. We validate our observation with a logistic regression model and find statistical support that peak hours are understaffed ($p<0.05$). The details of this test are explained in appendix A3.1.

Next, we validate our findings on the extent of understaffing during peak hours. Because conversion rate is positively associated with store labor (Perdikaki et al. 2012), we triangulate our findings by examining if conversion rate is lower during the hours when our model predicts
the store to be understaffed. Since conversion rate is an independent metric, if the model prediction is incorrect, then we would not observe a decline in conversion rate during the understaffed hours. This validity check will also help us rule out two alternate explanations where understaffing may be treated as innocuous as far as store performance is concerned. First, if store associates exert greater effort to compensate for the lack of workers during peak hours (Kc and Terwiesch 2009) then stores may not face a negative impact on sales and profitability during those understaffed hours. Second, if there are more browsers during peak hours then understaffing, as predicted by our model, may not necessarily result in lower financial performance for the stores.

For weekdays, we find that the average conversion rate during non-peak hours is 16.45% while the average conversion rate during peak hours is 14.33%. During peak hours, our model predicts that stores are understaffed 64.98% of the time. We observe a decline of 1.95% in conversion rate when the store is understaffed compared to other peak hours when the store is not understaffed. The correlation between decline in conversion rate and magnitude of understaffing is 0.25 ($p<0.05$). Next, consider the analysis of non-peak hours. During non-peak hours, we observe that the stores are understaffed only 17% of the time but we observe a 1.02% decline in conversion rate during those hours. We obtain similar results for the weekends sample as well. These validity checks increase our confidence in our model predictions of understaffing for this retailer and show that our results are not driven by alternate explanations such as workers exerting greater effort when stores are understaffed and the presence of more browsers during those hours.

In the next section we quantify the impact of understaffing on both lost sales and profitability in the stores during peak hours.

**6.2 Impact of understaffing on lost sales and store profitability**

We use (3) to measure the impact of understaffing on lost sales. We divide the lost sales from (3) by actual sales to normalize sales across different time periods. The results are presented in Table 5. We determine the average lost sales due to understaffing for the 41 stores during peak hours to be 8.56%. In other words, if the retailer was able to completely eliminate understaffing in its stores during the 3-hour peak period, then it would experience a sales lift of 8.56% during those hours. The range of values for lost sales across the 41 stores is [4.23%, 17.55%].
Next we measure the impact of eliminating understaffing on profitability using (5). Since we do not possess information on each store’s gross margin, we approximate $g_{it}$ by the average gross margin for this retail chain. Further, we approximate the labor cost, $d_i$, by the average wage rate for retail salespersons in that state to calculate the impact on profitability. Our analysis reveals that this retail chain’s average profitability will increase by 7.02% if it eliminated understaffing during peak hours. We calculate the percentage improvement in profits based on the calculated actual profit (using actual sales ($S_{it}$) and labor ($N_{it}$) in (5)) throughout the paper. Across the 41 stores in our sample, we find the profitability improvement to vary between [3.01%, 15.53%].

We note that our results on sales lift and profitability improvement may be conservative as our estimate of the impact of understaffing does not consider the long-term impact of lost sales on store performance. For instance, customers who did not receive proper service might switch to competitors resulting in a loss of life-time value from those customers (Heskett et al. 1994).

In summary, our results highlight the potentially large sales lift and profitability improvement that this retail chain can obtain by eliminating understaffing during the peak hours. These managerially salient results indicate that retailers should pursue different methods to eliminate, or more pragmatically, mitigate understaffing during peak hours. In §7, we examine some ways in which retailers may do so. We also perform different robustness checks of our key results on lost sales and profitability due to understaffing. These robustness checks are explained in more detail in appendix A3.

6.3 Extent of overstaffing and its impact on sales and profitability
For the sake of completion, we also discuss the results on the extent of overstaffing in these stores. Since stores need to maintain minimum labor in their stores even if there was no traffic, we account for this in our calculation of overstaffing levels. We determine the minimum labor in the stores based on our data and find it to be 1 person. If the predicted labor was less than 1 person, we set the predicted labor for that hour to be 1 person. We consider a store to be overstaffed during a time period if the actual labor in the store was greater than the predicted labor. We find that the stores were overstaffed 40.35% of the time and the average magnitude of overstaffing was 1.01 persons.

Eliminating overstaffing would impact both sales and profitability of these stores. For weekdays, we estimate that removing overstaffing would lead to a 1.05% increase in
profitability. Thus, for the stores in our sample, we find that the effect of removing overstaffing on profits is smaller compared to effect of removing understaffing which we estimated to be 7.02%. However, removing overstaffing would also lead to a decline in sales. For weekdays, our analysis shows that sales will decline by 1.8% if all overstaffing is removed in the retail stores. We find similar results for the weekends sample as well where we estimate that removing overstaffing would lead to a 0.97% increase in profitability and a 1.5% decline in sales. Although our study shows that reducing overstaffing would lead to an increase in profits in the short-term, it is possible that overstaffed stores may have greater long-term sales and profitability due to the higher level of customer service provided in these stores. Thus, further research should examine the long-term versus short-term impact of eliminating overstaffing in retail stores.

7. Drivers of understaffing in retail stores
In the previous section we quantified the large impact of understaffing on lost sales and profitability. These impacts represent the upper bound of the improvement that a retailer can expect if it completely eliminates understaffing. We made two assumptions in order to quantify these effects. First, we assumed that retailers would have perfect foresight of incoming traffic. This is a strong assumption since retailers schedule labor at least one or two weeks ahead based on traffic forecasts; forecast errors could limit the amount of sales lift and profitability improvement that may be achieved by retailers. Second, we assumed that retail stores would be able to change labor on an hour-to-hour basis. This assumption may also be unrealistic since retailers typically impose scheduling constraints such as minimum shift lengths in order to reduce the variability in store associates’ working hours. Therefore, we relax each of these assumptions to examine the amount of understaffing that can be realistically reduced in these retail stores. In addition, this analysis would shed light on the value of improving forecasting accuracy and scheduling labor with shorter minimum shift lengths.

7.1 Impact of forecast errors on extent of understaffing
In this section we study the value of improving forecast accuracy to eliminate understaffing. We do so by examining the impact of forecast errors on understaffing by first generating traffic forecasts one to three weeks in advance so that we may obtain realistic forecast errors of different magnitudes for this retail chain. We use one-three weeks as our forecast horizon as this is the typical time period for scheduling labor in retail stores. These forecasts are generated using a Newey-West time-series model. Let \( \hat{\lambda}_{t,f} \) be the forecast of traffic generated \( f \) weeks ahead,
i.e. \( f = 1, 2 \) and 3. In our setting, we find that as the forecast horizon increases from one week to three weeks, the average forecast errors increase from 12% to 25%. Next, we substitute this forecast of traffic \((\hat{\lambda}_{lt,f})\) in place of actual traffic \((\lambda_{lt})\) in (1) to calculate the predicted labor \((\hat{N}_{lt,f})\) as shown below:

\[
N_{lt,f} = \delta_{0_{lt,f}} + \delta_{1_{lt,f}} \hat{\lambda}_{lt,f} + \delta_{1_{ip,f}} \hat{\lambda}_{lt,f} * (1_{ip}) + \delta_{2_{lt,f}} \hat{\lambda}_{lt,f}^{31/2} + \delta_{2_{ip,f}} \hat{\lambda}_{lt,f}^{31/2} * (1_{ip}) + \xi_{4lt}
\]

(7)

The estimation results with a one-week ahead forecast of traffic are shown in Table 6. A positive deviation between the predicted labor from (1), which is based on perfect foresight, and the predicted labor from (7) would capture the extent of understaffing due to forecast errors. These results are reported in Table 7. We find that as we increase the forecast horizon from one to three weeks, the magnitude of understaffing as a percentage of predicted labor increases from 5.43% to 17.84%.

Next, we examine the sales lift and profitability improvement that would be obtained in these stores if they used a forecast of traffic to plan labor. Let \( \Delta N_{lt,f} = \hat{N}_{lt,f} - N_{lt} \) and \( 1_{\Delta N_{lt,f} > 0} \) be an indicator function which takes the value of 1 when the store is understaffed with respect to the predicted labor from (7), i.e. when \( \Delta N_{lt,f} > 0 \), 0 otherwise. Due to forecast errors, the labor plan based on a forecast of traffic will not be able to identify instances as well as magnitude of understaffing to the same extent as the labor plan from (1) where we had perfect information on traffic. Next, we use \( \hat{N}_{lt,f} \) and \( 1_{\Delta N_{lt,f} > 0} \) in (3) to compute the sales lift that would be obtained from using this labor plan. We use the estimated sales lift with traffic forecasts in (5) to compute the profit improvement. As expected, forecast errors lead to a lower sales lift and profitability improvement since all understaffing cannot be eliminated. Recall that the estimated sales lift and profitability improvement with perfect foresight of traffic were 8.56% and 7.02% respectively. The sales lift decreases by 2.61% (from 8.56% to 5.95%) and the profitability improvement is lower by 2.56% (from 7.02% to 4.46%) when we use a one-week ahead forecast of traffic. These results are reported in Table 7. Further increase in forecast horizon leads to higher forecast errors and consequently lower sales lift and profitability improvement. For example, sales lift is lower by 2.07% (from 5.95% to 3.88%) and profitability improvement is lower by 1.68% (from 4.46% to 2.78%) as we move from a one week to a three week forecast horizon.
Anecdotal evidence suggests that retailers typically schedule labor anywhere between one to three weeks ahead and our results quantify the sales lift and profitability improvement that comes with a shorter forecasting horizon.

7.2 Impact of scheduling constraints on the extent of understaffing

We now turn our attention to another possible driver of understaffing, namely scheduling constraints. 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 as retailers will be reluctant to increase labor hours due to the expenses incurred when labor is idle. To examine how much of the observed understaffing is explained by this scheduling constraint, we do the following. First, assuming perfect information on traffic, we calculate the predicted labor for each time period $t$, i.e. $\hat{\eta}_{it}$ from (1). Next, we divide the operating hours of the day into $b = 2$, 3 and 4 hour blocks of time. Then, we compute the average predicted labor for the block of time, i.e. $\hat{\eta}_{it,b} = (\hat{\eta}_{it} + \hat{\eta}_{it+1} + \ldots \hat{\eta}_{it+b-1})/b$. For example, in case of a 2 hour scheduling constraint, we calculate the predicted labor as $\hat{\eta}_{it,2} = (\hat{\eta}_{it} + \hat{\eta}_{it+1})/2$. We fix $\hat{\eta}_{it,2}$ as the predicted labor for timeperiod $t$ and $t + 1$, thus allowing no labor changes in the 2 hour block of time. A positive deviation between the labor plan from (1) and a labor plan with scheduling constraints would represent understaffing driven due to scheduling constraints. We consider 2-hour, 3-hour and 4-hour blocks of time in our analysis. We find that increasing scheduling constraints leads to an increase in understaffing, as shown in Table 7. For example, when we schedule labor with minimum shift lengths of 4 hours, as opposed to 2 hours, the magnitude of understaffing as a percentage of predicted labor increases from 7.23% to 28.74%.

Next, we follow a similar procedure as explained in the previous section to determine the sales lift and profitability improvement that would be obtained in these stores in the presence of scheduling constraints. As one might expect, imposing scheduling constraints on the labor plan leads to a lower sales lift and lower profit improvement. Recall that the estimated sales lift and profitability improvement when we allowed hour-to-hour labor changes were 8.56% and 7.02% respectively. The sales lift decreases by 3.76% (from 8.56% to 4.8%) and the profitability improvement is lower by 3.52% (from 7.02% to 3.5%) when we impose a 2-hour shift length constraint. Further increase in minimum shift length leads to further reduction in sales lift and
profitability improvement. For example, the sales lift is lower by 3.82% (from 4.8% to 0.98%) and profitability improvement is lower by 2.74% (from 3.5% to 0.76%) as we move from a 2-hour to a 4-hour shift length.

We observe several retailers moving towards shorter minimum shift lengths. For example, Wal-Mart and Payless ShoeSource have been trying to move towards more flexible work schedules (Maher, 2007). Though our results offer justification for these recent moves, retailers should carefully tread the path of reducing minimum shift lengths since it can adversely affect worker welfare (Lambert, 2008).

Up until this point, we have quantified the individual impact of reducing forecast errors and scheduling constraints on store sales and profits. In retail labor planning, typically traffic forecasts are used to drive scheduling decisions. Thus, one may expect an interaction of the forecasting errors and scheduling constraints. So, we next look at the impact of the interaction of forecast errors and scheduling constraints on store profitability with help of a simulation. Our baseline for comparison is store profits with the labor plan in (1) that has perfect foresight of traffic and allows for hour-to-hour labor changes. The percentage loss in store profits with increasing forecast errors and scheduling constraints is shown in Figure 4. Our results show that increasing scheduling constraints exacerbates the negative impact of forecast error. This can be seen from the rapid decline in profitability for higher values of forecast error and tighter scheduling constraints. For example, with a 2 hour scheduling constraint, doubling the traffic forecast error from 10% to 20% leads to an additional loss of 2.5% in store profits. On the other hand, with a 4 hour scheduling constraint, the concomitant additional loss in store profits is 6.1%, i.e. the impact of increase in forecast error on store profitability due to tighter scheduling constraints is more than doubled in this case.

This simulation 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 (Integrated Solutions for Retailers, 2010). Our simulation experiment here shows that although accurate forecasts are valuable, they alone would not help retailers to significantly increase store profitability. In addition to investing in centralized technologies that can improve forecasting, retailers also need to increase the flexibility of their workforce to achieve the maximum sales lift and profitability.
8. Alternate Methodology based on Structural Estimation Approach

Our results, so far, were based on predicted staffing levels obtained from a reduced-form estimation of the square-root expectation model. In this section, we consider an alternate approach where the benchmark staffing level is the optimal labor obtained from a structural estimation methodology. Assuming managers’ labor decisions at the daily level to be optimal, we impute the parameters of the sales and expense functions for each store. Then we use these parameters to determine the model-predicted optimal labor at the hourly level for each store and use this labor as the benchmark staffing level for our analysis. These model predictions represent the optimal labor for the store if the manager had perfect foresight of traffic and can freely change labor on an hour-to-hour basis in an unconstrained manner. Under this approach, the actual labor would represent the optimal labor that the store manager chose under several constraints, such as minimum shift lengths and break periods that are unknown to the empirical researcher. So, the deviation between the optimal labor obtained from the model and actual labor captures the magnitude of understaffing due to those constraints.

An advantage of the structural estimation approach is that it allows us to account for the intrinsic costs of labor for each store when determining its optimal labor. Prior literature has shown that managers’ intrinsic costs could be different from accounting costs as they capture the implicit cost-benefit tradeoff each manager faces in making operational decisions (Gino and Pisano 2008; Olivares et al. 2008; Allon et al. 2011). In a retail setting, the intrinsic cost of labor could be different from the accounting costs (i.e. the average wage rate of retail salespersons) due to the following reasons. First, some components of the accounting costs such as minimum wage rate, insurance, and medical benefits could vary across stores based on state laws in the United States. Second, the cost of labor could be driven by local market characteristics such as labor supply and customer expectations. For example, local markets with a tight labor supply might face high employee turnover that could increase labor costs due to increase in costs of hiring and training (Stiglitz 1974). Similarly, managers of stores that are located in markets where customers’ expectation of service is higher might place greater emphasis on service level and assess lower costs to labor (Campbell and Frei 2011). Finally, the cost of labor could depend upon the efficiency of labor and management in each store (Thomadsen 2005).
In §8.1 we explain the model, in §8.2 we discuss our estimation details and results, and in §8.3 we summarize the results on the extent of understaffing and its impact on lost sales and profitability based on this approach.

8.1 Model
We use the sales model and the profit model from §4.2 to capture the contribution of labor to sales and impute the cost of labor as shown below.

\[ S_{lt} = \alpha_i \lambda_i t e^{-y_i / N_{lt}} \]
\[ \pi_{lt} = S_{lt} * g_{lt} - N_{lt} * w_i \]

We replaced \( d_i \) in (4) with \( w_i \) in (9) to capture the intrinsic cost of labor that the store manager uses when deciding the amount of labor to have in the store. Each store manager is expected to maximize the profit function in (9), yielding the following first-order condition for amount of labor to have in each store:

\[ y_i \alpha_i \lambda_i t e^{-y_i / N_{lt}} g_{lt} = w_i N_{lt}^2 \]

The optimal labor plan \((N_{lt}^*)\) is the value of labor that is a solution to (10), given \( \alpha_i, \beta_i, y_i, w_i \) and store traffic \((\lambda_i t)\). In (10) we do not observe gross margin for each store at each time period. Since all stores sell similar assortments, we make a simplifying assumption that \( g_{lt} = g * e_{1lt} \), where \( g \) is the gross margin of the retail chain. Substituting for \( g_{lt} \) in (10) and dropping \( g \) which is a constant, we obtain the following decision rule for labor.

\[ y_i \alpha_i \lambda_i t e^{-y_i / N_{lt}} e_{1lt} = w_i N_{lt}^2 \]

Positive deviations of actual labor \((N_{lt})\) from optimal labor \((N_{lt}^*)\) denote understaffing while negative deviations denote overstaffing. Next we explain the estimation details and results.

8.2 Estimation results
We use the average values of sales, traffic and labor for each day in the estimation. We use structural estimation techniques to estimate the parameters \( \alpha_i, \beta_i, y_i, w_i \) in (11). The estimation results for both weekdays and weekends sample across the 41 stores are summarized in Table 8 and the details of the estimation procedure are provided in appendix A4.

We find considerable heterogeneity in the estimates of the imputed cost of labor across the 41 stores. For example, the average and standard deviation of \( w_i \) are $58.81 and $21.42 respectively. Even stores within the same state, that had the same average wage rate for retail salespersons, had very different imputed costs of labor. We find that the imputed cost of labor is...
significantly higher during weekdays than weekends \((p<0.001)\). This result is consistent with prior literature on higher usage of lower wage part-time labor on weekends in other retail organizations (Lambert, 2008).

Because the findings from a structural estimation approach are critically dependent upon the underlying assumption, we perform further validity checks to determine if the economic behavior implied by the imputed cost parameter is consistent with findings from prior research. To do this, we examine if the findings of Campbell and Frei (2011) hold in our setting as well. Campbell and Frei (2011) find that operating managers take local market characteristics into account when deciding on the number of tellers to schedule in a retail bank setting. 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. Similar examples of managers placing lower emphasis on cost while placing higher emphasis on service level have also been found in other settings as well (Png and Reitman 1994; Ren and Willems 2009). We test if the imputed cost of labor for different store managers also exhibits a similar behavior. Consistent with Campbell and Frei (2011), we find that a higher imputed cost is negatively associated with higher values of household income and competition. This validates our rationale for considering different labor costs for different stores. The details of the test and the results are explained in appendix A5.

**8.3 Extent of understaffing and impact on lost sales and profitability**

We compute the deviation of actual labor from the optimal labor for each hour. We find that stores are understaffed 32.88% of the time. When understaffing occurs, the magnitude of understaffing is 3.22 persons and this level of understaffing represents a 32.55% shortage compared to the optimal labor. The results are presented in Table 9. These results are in line with our results from the reduced-form regression in §6 where we found that stores were understaffed 40.21% of the time and the extent of understaffing was 2.10 persons. During peak hours, the stores were understaffed 68.21% of the time and the extent of understaffing was 3.52 persons. Next, we compute the impact of understaffing on lost sales and profits using the estimated parameters from (11) and the optimal labor \((N^*_{lt})\) in (3) and (5). We find that removing understaffing during peak hours would lead to a sales lift of 7.21% and increase profitability by 5.87%.
Next, we test if our results on peak-hour understaffing are driven by our assumption that store managers make optimal labor planning decisions at the daily level. If this were the case, we might find that stores are understaffed on different hours of the day or different days of the week under alternate assumptions. Hence, we re-estimate $\alpha_i, \beta_i, \gamma_i, w_i$ under an alternate assumption that store managers make optimal labor decisions at the weekly level. We find similar estimates of $\alpha_i, \beta_i, \gamma_i, w_i$ under this alternate assumption as well. The estimates are reported in appendix A6.1. Next, we use these weekly level estimates to calculate average understaffing at both the daily level and hourly level. We find the average understaffing at the daily level in this case to be 0.75 person (7.3% of optimal labor) for weekdays. For weekends, the average understaffing at the daily level is 0.52 person (4.5% of the optimal labor). Thus the deviation between optimal labor and actual labor at the daily level is not significantly different from zero even under a different optimality assumption. However, we still find significant understaffing at the hourly level (appendix A6.2). For example, we find that stores are understaffed 65.66% of the time during peak hours and the magnitude of this understaffing is 3.28 persons (30.05% of optimal labor) with these estimates. These results lead us to believe that our finding on peak-hour understaffing in retail stores is robust.

In conclusion, our analysis shows that reduction in understaffing can lead to significant improvements in profitability for this retailer.

9. Conclusions, Limitations, and Future Work
In this paper we use hourly data on store traffic, sales, and labor to examine whether retail stores are understaffed. We find that the stores in our sample are consistently understaffed during peak hours, and estimate the impact of understaffing on lost sales and profitability to be 8.56% and 7.02% respectively. We show that forecast errors and scheduling constraints limit the ability of retailers to reduce understaffing and consequently the profitability improvement they may achieve from such reductions. We add to the growing research on the impact of labor on financial performance (Fisher et al. 2007; Ton and Huckman 2008; Netessine et al. 2010; Ton 2009; Perdikaki et al. 2012) by documenting the extent of understaffing in retail stores, as well as exploring the effect of common drivers of understaffing like forecasting errors and scheduling problems.

One of the managerial implications of our study is the value of traffic information in labor planning to retailers. We show that staffing decisions based on traffic can yield higher profits
since customer traffic captures the true demand potential of a store. Second, an important staffing related decision that retailers need to make is determining how many weeks in advance they need to schedule associates for work. By quantifying the impacts of forecast errors and minimum shift length on store performance, our study informs retailers of the relative impacts of these decisions on store profits. Our experience with several retailers shows that managers know that their stores are understaffed during peak periods. However, they often find it difficult to estimate the impact of understaffing on lost sales and profits for the store. For example, after seeing a presentation of these results, one senior manager commented that, “We always knew our stores were understaffed at some times and overstaffed during others. Culturally we have tended to nod at this issue but this analysis shows how we have undervalued the impact of understaffing on financial performance.”

We have the following limitations in this paper. First, we did not possess information on service time and abandonment rates for the stores in our sample that could impact store labor requirements, sales and profits. For example, it is possible that there are higher abandonment rates during peak hours due to congestion effects. Similarly, the effective service rate during these hours could change not only because of not having enough sales associates, but also due to limited capacity at the fitting room in these stores. Also, we do not have information on the time spent by customers in browsing through products before joining the queue, and hence could not separate the browsing or selection process from the purchase process and calculate the actual waiting time for customers. Future research could look at collecting more detailed information on customer buying process and waiting times as well as information on service rates and abandonments to investigate how these would impact the staffing decisions for the store.

Second, our study is focused on short-term profitability. Prior research has shown that decrease in service quality could result in a decline in customer satisfaction and loyalty (Zeithaml et al. 1996; Oliva and Sterman 2001). Therefore, the impact of understaffing on total profitability could be much higher than what we estimate it to be. Future research can use longer time series of data along with long-term customer satisfaction data for each store to study the impact of understaffing on future profitability.

Third, we did not possess data on store promotions in our sample. A store manager may increase labor in order to change signage and perform associated tasks ahead of the promotion. Absent promotion information and more importantly when such store activities are executed, our
model would treat those instances as overstaffing. We also did not have information on gross margin for each individual store in our sample. Hence we could not control for this in our estimation of improvement in store profitability. Additionally, we did not have detailed data on store labor like employee turnover, absenteeism and the proportion of full-time workers, part-time workers, and temporary workers. Employee turnover and absenteeism could cause unexpected changes in labor and lead to understaffing and overstaffing issues in the store. Similarly, because part-time workers and temporary workers may not have the same amount of cumulative experience as full-time workers, they would possess lower knowledge (Argote 1999) and likely provide lower quality of service. Also, a change in labor-mix could affect store sales and profits in different ways (Kesavan et al. 2012). Future research could look at how these variables impact store staffing decisions and profitability.

Finally, prior research has shown the presence of intentional and unintentional biases in the forecasting processes (Oliva and Watson 2009). It is possible that some of the understaffing we observe is driven by deviations in day-level biases and hour-level biases. Future research may collect additional data to examine how the extent and impact of understaffing on store profitability varies across stores based on the gross margin, labor composition, and magnitude of biases.

References


**Figure 1: Labor planning process**

<table>
<thead>
<tr>
<th>Sales/Traffic forecast</th>
<th>Planned hours</th>
<th>Scheduled hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting</td>
<td>Labor Planning</td>
<td>Labor Scheduling</td>
</tr>
</tbody>
</table>
Figure 2: Cluster analysis of average traffic across days of week and months of year

Figure 3: Comparison of actual labor and optimal labor for stores during peak and non-peak hours

Figure 4: Impact of forecast errors and scheduling constraints on store profits

Table 1: Summary statistics of store variables

<table>
<thead>
<tr>
<th>Name</th>
<th>Weekdays</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Weekends</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg.</td>
<td>Std. dev</td>
<td>Min</td>
<td>Max</td>
<td>Avg.</td>
<td>Std. dev</td>
<td>Min</td>
<td>Max</td>
<td></td>
</tr>
<tr>
<td>$S_{idh}$</td>
<td>686.11</td>
<td>243.12</td>
<td>94.58</td>
<td>11020.52</td>
<td>1127.58</td>
<td>918.64</td>
<td>91.86</td>
<td>141.25</td>
<td>13067.46</td>
</tr>
<tr>
<td>$N_{idh}$</td>
<td>5.41</td>
<td>2.11</td>
<td>1.0</td>
<td>16.0</td>
<td>10.27</td>
<td>2.84</td>
<td>2.0</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>$Txn_{idh}$</td>
<td>7.14</td>
<td>4.59</td>
<td>1.0</td>
<td>46.0</td>
<td>11.71</td>
<td>7.08</td>
<td>1.0</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{idh}$</td>
<td>48.99</td>
<td>29.31</td>
<td>5.0</td>
<td>437.0</td>
<td>95.51</td>
<td>56.40</td>
<td>17.0</td>
<td>630.0</td>
<td></td>
</tr>
<tr>
<td>$CR_{idh}$</td>
<td>15.79</td>
<td>2.43</td>
<td>9.40</td>
<td>20.19</td>
<td>13.38</td>
<td>4.14</td>
<td>1.85</td>
<td>25.89</td>
<td></td>
</tr>
<tr>
<td>$BV_{idh}$</td>
<td>90.93</td>
<td>42.42</td>
<td>10.31</td>
<td>1371.26</td>
<td>94.58</td>
<td>50.11</td>
<td>15.50</td>
<td>1448.56</td>
<td></td>
</tr>
</tbody>
</table>

n = 73800 for weekdays and 53300 for weekends
Table 2: Estimation results for the regression in equation 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weekdays</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Weekends</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Std. Dev</td>
<td>Min</td>
<td>Max</td>
<td>Average</td>
<td>Std. Dev</td>
<td>Min</td>
<td>Max</td>
<td></td>
</tr>
<tr>
<td>$\delta_{0i}$</td>
<td>1.031</td>
<td>.328</td>
<td>.315</td>
<td>2.174</td>
<td>1.104</td>
<td>.435</td>
<td>.585</td>
<td>2.214</td>
<td></td>
</tr>
<tr>
<td>$\delta_{1i}$</td>
<td>.285</td>
<td>.028</td>
<td>.006</td>
<td>.658</td>
<td>.277</td>
<td>.035</td>
<td>.002</td>
<td>.656</td>
<td></td>
</tr>
<tr>
<td>$\delta_{1p}$</td>
<td>-.031</td>
<td>.001</td>
<td>-.110</td>
<td>.054</td>
<td>-.035</td>
<td>.020</td>
<td>-.103</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>$\delta_{2i}$</td>
<td>.166</td>
<td>.075</td>
<td>.064</td>
<td>.937</td>
<td>.155</td>
<td>.102</td>
<td>.052</td>
<td>.851</td>
<td></td>
</tr>
<tr>
<td>$\delta_{2p}$</td>
<td>-.026</td>
<td>.006</td>
<td>-.154</td>
<td>.082</td>
<td>-.048</td>
<td>.011</td>
<td>-.187</td>
<td>.065</td>
<td></td>
</tr>
</tbody>
</table>

n = 73800 for weekdays and 53300 for weekends. The regressions were run for each store separately. The average values of the estimates are reported in the table. $^a$estimates were statistically significant at $p<0.1$ level. $^b$estimates were statistically significant at $p<0.1$ level for 32 out of 41 stores.

Table 3: Estimation results for the regression in equation 6

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weekdays</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Weekends</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Std Dev</td>
<td>Min</td>
<td>Max</td>
<td>Average</td>
<td>Std Dev</td>
<td>Min</td>
<td>Max</td>
<td></td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>10.70</td>
<td>5.82</td>
<td>2.67</td>
<td>39.12</td>
<td>16.11</td>
<td>7.05</td>
<td>3.28</td>
<td>45.45</td>
<td></td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.42</td>
<td>0.13</td>
<td>0.08</td>
<td>0.79</td>
<td>0.39</td>
<td>0.11</td>
<td>0.06</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>8.71</td>
<td>2.65</td>
<td>4.10</td>
<td>19.96</td>
<td>12.65</td>
<td>3.08</td>
<td>5.11</td>
<td>27.06</td>
<td></td>
</tr>
</tbody>
</table>

n = 73800 for weekdays and 53300 for weekends. The regressions were run for each store separately. The average values of the estimates are reported in the table. All estimates significant at $p<0.1$.

Table 4: Extent of under- and overstaffing based on reduced-form regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weekdays</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Weekends</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Understaffing</td>
<td>Overstaffing</td>
<td>Understaffing</td>
<td>Overstaffing</td>
<td></td>
<td>Understaffing</td>
<td>Overstaffing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All hours</td>
<td>Peak hours</td>
<td>All hours</td>
<td>Peak hours</td>
<td>All hours</td>
<td>Peak hours</td>
<td>All hours</td>
<td>Peak hours</td>
<td></td>
</tr>
<tr>
<td>Magnitude of under- or overstaffing (persons)</td>
<td>2.10</td>
<td>2.31</td>
<td>1.01</td>
<td>0.04</td>
<td>2.77</td>
<td>3.58</td>
<td>1.31</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Magnitude of under- or over-staffing scaled by predicted labor (%)$^a$</td>
<td>33.27</td>
<td>30.19</td>
<td>26.61</td>
<td>0.98</td>
<td>25.45</td>
<td>26.68</td>
<td>22.36</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>% of time store is under- or overstaffed (%)</td>
<td>40.21</td>
<td>64.98</td>
<td>40.35</td>
<td>9.62</td>
<td>40.66</td>
<td>70.95</td>
<td>35.12</td>
<td>6.98</td>
<td></td>
</tr>
</tbody>
</table>

n = 33620 for weekdays and 20664 for weekends. $^a$For example if in a given hour the predicted labor is 6.31 persons and understaffing is 2.10 persons, then this represents 33.27% of the predicted labor for that hour.

Table 5: Impact of understaffing on lost sales and profitability

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weekdays</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Weekends</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All hours</td>
<td>Peak hours</td>
<td>All hours</td>
<td>Peak hours</td>
<td></td>
<td>All hours</td>
<td>Peak hours</td>
<td></td>
</tr>
<tr>
<td>Drop in CR(%)</td>
<td>1.78</td>
<td>1.95</td>
<td>1.56</td>
<td>1.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact on lost sales (%)</td>
<td>7.26</td>
<td>8.56</td>
<td>5.11</td>
<td>6.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact on profitability (%)</td>
<td>5.81</td>
<td>7.02</td>
<td>4.23</td>
<td>5.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 33620 for weekdays and 20664 for weekends
Table 6: Estimation results using one-week ahead forecast of traffic in the regression in equation 7

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weekdays</th>
<th></th>
<th>Weekends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Std. Dev</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>$\delta_{011}^a$</td>
<td>1.064</td>
<td>.538</td>
<td>.215</td>
<td>3.151</td>
</tr>
<tr>
<td>$\delta_{111}^a$</td>
<td>.272</td>
<td>.035</td>
<td>.002</td>
<td>.558</td>
</tr>
<tr>
<td>$\delta_{11p}^b$</td>
<td>-0.018</td>
<td>.001</td>
<td>-1.09</td>
<td>.084</td>
</tr>
<tr>
<td>$\delta_{211}^a$</td>
<td>.187</td>
<td>.119</td>
<td>.041</td>
<td>1.415</td>
</tr>
<tr>
<td>$\delta_{21p}^b$</td>
<td>-0.021</td>
<td>.004</td>
<td>-1.68</td>
<td>.095</td>
</tr>
</tbody>
</table>

n = 33620 for weekdays and 20664 for weekends for Tables 6 and 7. *estimates were statistically significant at $p<0.1$ level. **estimates were statistically significant at $p<0.1$ level for 32 out of 41 stores.

Table 7: Results with traffic forecasts and constraints in labor scheduling

<table>
<thead>
<tr>
<th>Labor Plan</th>
<th>Weekdays</th>
<th></th>
<th>Weekends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%Under-staffing</td>
<td>% Sales lift</td>
<td>% Profit improvement</td>
<td>%Under-staffing</td>
</tr>
<tr>
<td>Predicted</td>
<td>0.00</td>
<td>8.56</td>
<td>7.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Actual</td>
<td>30.19</td>
<td>0.00</td>
<td>0.00</td>
<td>26.68</td>
</tr>
<tr>
<td>Generated with traffic forecast</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 wk</td>
<td>5.43</td>
<td>5.95</td>
<td>4.46</td>
<td>7.63</td>
</tr>
<tr>
<td>2 wk</td>
<td>7.33</td>
<td>4.97</td>
<td>3.59</td>
<td>9.22</td>
</tr>
<tr>
<td>3 wk</td>
<td>17.84</td>
<td>3.88</td>
<td>2.78</td>
<td>15.04</td>
</tr>
<tr>
<td>With scheduling constraint requiring constant labor for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 hr</td>
<td>7.23</td>
<td>4.80</td>
<td>3.50</td>
<td>9.83</td>
</tr>
<tr>
<td>3 hr</td>
<td>14.03</td>
<td>3.55</td>
<td>1.99</td>
<td>17.05</td>
</tr>
<tr>
<td>4 hr</td>
<td>28.74</td>
<td>0.98</td>
<td>0.76</td>
<td>25.44</td>
</tr>
</tbody>
</table>

Table 8: Structural estimation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Weekdays</th>
<th></th>
<th>Weekends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Std Dev</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>16.15</td>
<td>3.57</td>
<td>11.01</td>
<td>22.62</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.25</td>
<td>0.08</td>
<td>0.16</td>
<td>0.38</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>12.55</td>
<td>3.41</td>
<td>7.02</td>
<td>18.98</td>
</tr>
<tr>
<td>$w_i$ ($/hr$)</td>
<td>58.81</td>
<td>21.42</td>
<td>20.46</td>
<td>113.60</td>
</tr>
</tbody>
</table>

n = 73800 for weekdays and 53300 for weekends for Tables 8 and 9. All estimates significant at $p<0.05$.

Table 9: Extent of understaffing and overstaffing based on structural estimation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weekdays</th>
<th></th>
<th>Weekends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Understaffing</td>
<td>Overstaffing</td>
<td>Understaffing</td>
<td>Overstaffing</td>
</tr>
<tr>
<td></td>
<td>All hours</td>
<td>Peak hours</td>
<td>All hours</td>
<td>Peak hours</td>
</tr>
<tr>
<td>Magnitude of under- or overstaffing (persons)</td>
<td>3.22</td>
<td>3.52</td>
<td>1.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Magnitude of under- or overstaffing scaled by optimal labor (%)</td>
<td>32.55</td>
<td>32.48</td>
<td>24.23</td>
<td>1.08</td>
</tr>
<tr>
<td>% of time store is under- or overstaffed (%)</td>
<td>32.88</td>
<td>68.21</td>
<td>52.85</td>
<td>9.87</td>
</tr>
</tbody>
</table>