Chapter 6
An Overview of Industry Practice and Empirical Research in Retail Workforce Management

Saravanan Kesavan and Vidya Mani

1 Introduction

In the highly competitive retail environment, many retailers consider in-store experience critical to converting incoming traffic into sales and future visits. Superior in-store experience requires having not only inventory in place but also a skilled store workforce to ensure an efficient and pleasant visit for the customers. Numerous studies in marketing have shown that store associates play a critical role in driving customer satisfaction (Parasuraman et al. 1988; Zeithaml et al. 1996). Anecdotal evidence of financial distress resulting from mismanagement by retailers of their labor force is abundant. One recent example involves Circuit City, a consumer electronics company, which undertook several drastic changes under its new management, including revamping its store labor by letting-go of its highest paid sales associates. Retail observers claim that firing such experienced sales associates caused customer satisfaction to decline precipitously and contributed to Circuit City's subsequent bankruptcy (Mui 2007).

While retailers care deeply about providing high service levels to customers through increased labor in their stores, they are also mindful about the expenses associated with this practice. Payroll-expenses are about 10 % of sales in the retail industry and can often be the largest component of a store's variable costs. Kesavan et al. (2013) study a big-box retailer whose labor expenses account for 85 % of total

S. Kesavan (✉)
Kenan-Flagler Business School, University of North Carolina at Chapel Hill,
Campus Box 3490, McColl Building, Chapel Hill, NC 27599-3490, USA
e-mail: skesavan@unc.edu

V. Mani
Supply Chain and Information Systems, Penn State Smeal College of Business,
461 Business Building, University Park, PA 16802, USA
e-mail: vmani@psu.edu

N. Agrawal, S.A. Smith (eds.), *Retail Supply Chain Management*,
International Series in Operations Research & Management Science 223,
DOI 10.1007/978-1-4899-7562-1_6

controllable expenses[1] in the store. As a consequence, retailers need to balance the need to drive sales by using more labor against the need to control expenses that can increase commensurately. This task is challenging and requires careful workforce management. In this chapter, we review the literature on workforce management; provide a detailed overview of labor planning practice at one retailer; review new and upcoming technologies in the retail landscape that can potentially impact labor practices; and conclude with areas of future research.

The raison d'être for store labor is fairly similar across most retailer settings. First, stores need sufficient labor to ensure customer service. Service entails dealing directly with the customers during the purchase process: answering customer questions about the product and any services or warranty associated with them, and indirectly affecting their in-store experience by ensuring a neat and clean store. Second, store labor needs to manage the inventory in the store. Managing inventory involves receiving merchandise from delivery trucks while ensuring that it complies with the bill-of-materials, stocking the shelves to ensure that customers can find the products they are looking for, and finally, keeping the price current so it reflects the discounts or pricing changes that the corporate office may mandate. Third, store labor is required to maintain the signage within a store. Corporate office announcements of a new promotional event require that store labor update store signage to be consistent with the marketing activity. Finally, store labor is required for cashiering.

Broadly, retail labor falls into three categories depending upon the employment contract with the retailer: full-time workers, part-time workers, and temporary or seasonal workers. According to the Bureau of Labor Statistics (BLS), only 70 % of the estimated 15 million strong retail workforce in 2013 is full-time. Further, the retail industry added more than 700,000 seasonal employees for the holiday season in 2013 (BLS). Full-time workers are year-round employees who are typically employed for fixed hours per week, typically 35–40 h. They can be employed for a few more hours with overtime pay. Part-time employees are also year-round employees but face variable hours of employment in a week. BLS defines part-time workers as those who usually work less than 35 h per week. For example, the retail organization studied in Kesavan et al. (2013) guaranteed 10 h of employment per week for its part-time employees and deployed them for an average of 22 h per week. While retailers may increase the hours of the part-time employees to 40 h per week, they can do so only for a short-period of time before these workers get reclassified as full-time employees. Finally, temporary employees, sometimes called seasonal employees, are deployed for shorter-periods of time to manage seasonality or short-term demand fluctuation. Seasonal workers can be a large proportion of the total workforce for retailers during the peak period. For example, Home Depot planned to hire 70,000 seasonal employees to augment its 320,000 regular employees to meet seasonal demand in Spring 2012. Seasonal employees

[1] This retailer had identified the controllable component of each of the expenses based on historical data for each store.

have not only varying lengths of employment but also can be deployed for varying hours in a week. Typically, full-time employees are provided with other benefits, such as sick pay, vacation pay, and health care benefits. Some retailers tend to provide benefits to part-time employees but temporary employees rarely receive such benefits.

These different classes of workers offer various advantages to retailers to manage their stores. Typically, full-time workers are considered to have the highest capability amongst the three classes of workers since full-time employment often draws the most qualified candidates.[2] Further, literature on learning curve effects have shown that performance improves with cumulative experience (Lapré and Nembhard 2011) so full-time workers who spend more time in their jobs compared to part-time and seasonal workers are likely to have greater capabilities. Finally, full-time employees' incentives may be better aligned with that of the organization compared to those of part-time and seasonal workers. So, apart from playing an important role in driving sales through superior customer service, they may also reduce organizational costs by having lower turnover compared to part-time and seasonal workers. Annual turnover for the retail sector can be as high as 100 % (National Retail Foundation) but the break-down for part-time and seasonal workers is not available.

Part-time and seasonal workers, on the other hand, provide other important benefits to retailers. The wage rates and other benefits tend to be lower than that of full-time workers. In addition, they provide volume flexibility (upside flexibility and temporal flexibility) (Kesavan et al. 2013) to retailers that could enable them to manage demand less expensively, at least up to a certain point.

Labor planning involves determining the right number of full-time, part-time, and seasonal workers in the stores and allocating the forecasted hours across those workers. We observe considerable differences in the way labor planning is performed in the retail industry. One important dimension in which retail organizations can vary is the level of sophistication used to manage payroll. At one end of the spectrum, payroll decisions are completely driven by store managers without the support of decision-making tools. This practice is typical of smaller retailers, but we have observed that even retailers with annual revenues exceeding a billion dollars may follow such an ad hoc process. At the other end of the spectrum, several retailers have invested millions of dollars in workforce management tools that plan how much labor each store must carry. Some examples of firms developing workforce management tools are RedPrairie, Kronos, Reflexis, and Ceridian.

Another area of difference is the degree to which different departments within a retail organization are involved in labor planning. Several departments within retail organizations commonly want a say in the amount of labor in the store. Sometimes these departments have different goals. For example, the finance department in a

[2] There are exceptions to this generalization. For example, it is common to witness well qualified plumber or a sales associate who pursues part-time opportunities to balance non-work related activities. About 65 % of part-time workers choose to work part-time (BLS).

retail organization cares about controlling labor expenses in the stores, so it sets a ratio of sales to labor as a target for store managers to achieve. Merchandising departments, on the other hand, have their incentives tied to sales of product categories. Since sales of certain product categories, such as appliances or shoes, would be sensitive to labor, the merchandising department may want appropriate coverage of those departments with labor presence. Finally, store operations care about having sufficient labor to cover the large number of non-customer-facing tasks in a store. While the different groups provide feedback on the amount of labor in the store, many retailers ultimately let the store manager determine the right amount of labor in their stores. By tying the store managers' bonuses to profits, the corporate office tries to overcome the classic agency problem that arises in these situations.

In this book chapter, we present an overview of industry practice around workforce management and empirical research on this topic. Even with such a narrowly defined goal, it was necessary to add further restrictions to strike a balance between the depth and breadth of the topics covered. This book chapter is largely restricted to U.S. public retailers. The industry practice explained here is based on our experience with several specialty and big-box retailers and workforce management software providers, and has been validated through presentations to numerous retail practitioners. However, there are likely to be deviations between the labor planning practices described in this chapter and those followed in other retail settings. Consistent with the contemporaneous nature of the empirical research in this area, the literature survey weighs recent papers more.

Next we explain workforce management planning practice in detail for one of the retailers in Sect. 2. In Sect. 3, we review the literature around labor planning, with emphasis on empirical research in retail labor in response to the emerging interest in this area. In Sect. 4, we discuss some of the new technologies shaping the retail landscape that have implications for retail labor. We conclude with directions for future research in Sect. 5.

2 Labor Planning in Practice: Case Study of HomeRetail

In this section, we explain the labor planning practice at HomeRetail, a pseudonym for the retailer with whom we interacted. HomeRetail is a big-box retailer with annual revenues exceeding $1 billion. This retailer is in the home goods industry and carries over 10,000 items in its stores. This retailer employs year-round full-time and part-time employees and seasonal employees for a shorter duration of time to meet its annual sales spike. The labor planning practice is similar to that of many other big-box retailers with whom we have interacted. Specialty retailers tend to have smaller stores, and their labor planning process tends to be much simpler than the one described in this section.

Due to the large sizes of its stores (over 100,000 sq. feet with more than 100 employees), this retailer has a deep organizational structure for each of its store. Each store was divided into multiple departments based on the product category, and each of those departments have a department manager who is responsible for managing labor associated with that department. The labor within each department is divided into various roles such as sales associates, specialists, cashiers, backend delivery, and assembly, etc. The department managers are incentivized based on sales and profits in their department. The different department managers report to assistant store managers, who in turn, report to the store manager. Store managers also had a human resources (HR) manager to help them with recruiting workers. HR managers play a vital role during the peak season, when they need to hire a large number of seasonal employees for the store, train them, and manage their exit at the end of the season.

Next we explain the labor planning process at HomeRetail in detail. We divide the labor planning process into long-term and short-term planning, where long-term planning refers to planning for 1 year and short-term planning is the planning for near term, such as the next month or two.

Long-term planning: Long-term planning is typically done at the beginning of the fiscal year when retailers revisit the organizational structure for each store and the minimum staff required to manage a store. HomeRetail groups stores based on their sales volume into different tiers. Stores in each tier are allocated base hours, that is the minimum hours per week, for different roles, such as assistant store managers, human resources (HR) manager, cashiers, sales associates, and department managers. These base hours guide the store managers to determine the number of full-time and part-time workers to have in the store on an ongoing basis. Though store managers are given some direction on the proportion of part-time to full-time workers to have in their stores, we observe that they have considerable leeway to deviate from this suggested proportion. If store managers need to recruit additional workers for their stores, they do so with the help of the HR manager in the store.

Short-term planning: While long-term planning enables store managers to get the right number of full-time and part-time workers in place, short-term planning involves balancing the labor hours required in a given month to the workforce in place. This stage begins with the determination of labor hours that need to be staffed for a given month. At HomeRetail, the store managers, the district managers, and the corporate finance team jointly forecast sales for a month, usually 30 days or more in advance. The sales forecast is then used as an input to a regression model that was estimated using historical sales and labor data to predict the labor hours required to satisfy the forecasted sales. These labor hours are communicated to the store manager, who needs to ensure that a sufficient number of workers exist to cover those hours.

Store managers would then schedule full-time and part-time workers to ensure coverage. Typically, managers use software tools to match worker availability with the workload requirements of the store. The workload requirements are driven based on the number of operational activities they need to perform as well as the labor required to support sales tasks, as predicted by the sales forecast. This tool also takes into account several restrictions imposed by minimum labor requirements set by corporate, local labor laws, union rules, quality of life considerations etc., while determining the final schedule. These schedules are generally posted a week or more in advance so that associates can plan accordingly. At HomeRetail, full-time workers typically worked 8-h shifts for 5 days a week and were asked to work a minimum number of weekend days in a month.[3] So, store managers had some limited flexibility on shift lengths and shift days for full-time employees. Part-time workers offered more flexibility, as they could work for variable shift lengths and for different days of the week.

While the above approach works for most of the year, the forecasted hours could exceed the capacity provided by full-time and part-time workers during peak periods. We find that many stores double their sales during the peak period, so even a fully cross-trained staff would not be able to handle the demand surge necessitating hiring of seasonal workers. While, by convention, peak period coincide with the holiday season, some retailers such as Home Depot and Lowe's begin their peak periods in the spring.

Recruiting and onboarding seasonal workers are challenging tasks for retailers and consume a lot of the attention of store management. Because demand during peak period can be twice as large as that during the non-peak period, stores need to aggressively recruit seasonal workers to maintain service quality. For example, HomeRetail invests heavily in building relationships with local colleges as well as the community as a whole to ensure sufficient supply of seasonal workers to its stores. Many store managers mention that they often start planning for recruitment for the next peak season right at the end of the previous peak season. However, for a majority of stores, the active planning stage for the seasonal workers begins 4 months before the beginning of the peak period, when the area HR manager in consultation with the store manager identifies the approximate number of seasonal workers that the stores may need for the upcoming peak period. This process aligns the corporate managers with the needs of the stores. However, formal recruiting does not begin at this stage. The actual recruiting process takes anywhere between 1 and 3 months for HomeRetail. We explain this process of recruiting seasonal workers next.

When the store manager is ready to recruit seasonal workers, they request approval from the district manager. Once approved, the store's HR manager creates a job description depending upon whether the seasonal worker is required for cashiering, sales, stocking shelves, unloading trucks, or some other role. This job description is posted internally before being communicated to local colleges and other sources of seasonal workers. HomeRetail, for instance, requires its stores to interview three candidates for every position. In addition to multiple rounds of interviews, the candidates also need to undergo drug testing and background checks

[3] At another major apparel retailer that we worked with, full-time workers were asked to work four long shifts of 9 h each and one short shift of 4 h. Shorter shift lengths can increase store profitability significantly (Mani et al. 2014), however associate dissatisfaction could also increase.

before they receive an offer. Thus, even if candidates are readily available, the process of bringing a candidate to a store could take at least a month.

Once workers are recruited, stores follow the essential step of onboarding them by providing appropriate training. The extent of training can vary considerably from retailer to retailer and store to store and by whether workers are full-time, part-time, or seasonal. For example, Fisher and Krishnan (2005) document the case of Wawa convenience stores, where store managers are responsible for training the associates. At HomeRetail, this training is provided partly by the corporate office through centralized web tools and supplemented by store manager and department managers. Unsurprisingly, we find that full-time and part-time workers receive longer periods of training compared to seasonal workers.

3 Literature Review

Labor planning is not new to operations management; indeed, a long history of mathematical models and scheduling algorithms has evolved to optimize staffing requirements. Most of these models have been developed (and successfully applied) in the context of a manufacturing setting. However, some key differences exist between a manufacturing and a retail operation that prevent direct application of these models in a retail store. Since the manufacturing setting is well known to the operations management audience, we begin by highlighting the key differences in labor planning between retail and manufacturing industries. We then discuss the emerging area of empirical research on retail labor in detail.

3.1 Differences Between Manufacturing and Retail Settings

Early work on labor planning in operations management literature concentrated mainly on determining labor requirements in manufacturing environments. The main focus was on determining optimal (or near optimal) solutions to labor requirements in the context of aggregate planning. Aggregate planning is an intermediate-range capacity planning process that typically covers a time horizon of 2–12 months and involves simultaneous determination of a firm's production, inventory, and employment levels over this time horizon to meet the total demand for all products that share the same limited resources. The objective is to minimize the total cost (or expected cost in case of uncertain demand) while taking into consideration constraints on the production rates and changeovers as well as inventory and workforce levels. The cost parameters would include cost of production, inventory and shortage costs, cost of adjusting the production rate through over-time or under-time, and cost of adjusting workforce through hiring and firing employees. In most cases, all available workers were treated as equally productive and cost parameters have to be determined from actual financial data. Subsequent research has dealt with incorporating labor flexibility as well as short-term decisions like workforce scheduling into the aggregate planning framework. Below we highlight a few relevant papers in this domain.

Starting with the seminal paper by Holt et al. (1956), several papers have developed mathematical models to find the aggregate production rate and size of workforce to meet demand. Linear programming and integer programming techniques are used to get the optimal decision rule that minimizes the total cost of regular payroll and overtime, hiring and layoffs, and inventory and shortages incurred during a given planning interval of several months (Lippman et al. 1967). Continuing studies on the problem of determining labor requirements in job shops, later researchers have also used stochastic programming techniques to cope with non-stationary stochastic demand for labor (Dill et al. 1966; Anderson 2001). In many of these papers, quadratic or convex cost functions are used to represent the cost of hiring, firing, and use of overtime (Kunreuther and Morton 1974). Quadratic cost functions are used to penalize deviation of key variables from target levels. The advantage of using quadratic functions is that they result in linear production rules that can be easily applied in a repetitive manner once the constants in the model are determined (e.g. the linear decision rule in Holt et al. 1956). Convex cost structures arise when marginal hiring (firing) costs increase with the number of employees hired (fired). This could arise when there are steep increases in costs with addition of a new shift, technological and productivity changes, labor slowdowns, etc. These cost structures are usually approximated by piecewise linear cost functions and add substantial complexity and computational effort to the problem.

In contrast to continuous assembly line manufacturing environments, job shops are characterized by batch-processing and may require additional skilled labor for specific type of jobs. Thus, all labor units cannot be treated equal in the aggregate planning problem. Subsequent work in this field has looked at incorporating labor flexibility into the aggregate production and workforce planning in the context of job shop planning (Fryer 1974; Brownell and Lowerre 1976). Later work has looked at impact of different cross-training policies on performance of serial production systems with an objective of minimizing the costs of cross-training while meeting staffing requirements (Daniels et al. 2004; Hopp et al. 2004; Bard and Wan 2008). Extensive work also exists on determining detailed shift schedules for employees. A large body of academic literature has developed mixed integer linear programming techniques for scheduling full-time workers to minimize labor hours while satisfying variable workforce requirements of a service delivery system (Dantzig 1954; Morris and Showalter 1983). Considerable work has also been done on modeling workforce requirements based on multiple shifts, by incorporating the effect of constraints on the changing of shifts over the planning period. The common approach in these papers is to use integer programming to determine optimal shift schedules that include flexible rest or meal-breaks, and allow for alternate shift starting times, shift lengths, and break placement (Bechtold and Jacobs 1990; Thompson 1995).

Several differences exist between labor planning processes in manufacturing and retail. The most important source of these differences is that customers often interact with a service provider to jointly produce the outcome, a process known as customer co-production (Karmarkar and Pitbladdo 1995). Co-production requires the real-time involvement of the customer with the store associate for its successful completion; as such, inventorying service in anticipation of future demand is typically not possible. Thus, a key challenge faced by retailers when planning labor is to ensure that their workforce is available when customers walk into their stores. Significant variabilities in customer arrival process occur within a day, across days of week and across months of the year and make it hard for retailers to match labor with demand. The Figs. 6.1, 6.2 and 6.3 depict these three types of variabilities based on traffic data from 41 stores of a women's apparel retail chain. The low, med and high lines depict the 10th, 50th and 90th percentile of average traffic across these 41 stores. Unlike manufacturing settings in which the order lead time enables manufacturers to mitigate the effect of forecast errors by shifting orders across time or facilities, under- or over-staffing in retail settings can have immediate impacts on financial performance of the stores. Mani et al. (2014) find stores to be understaffed about 41 % of the time in their retail setting and find the impact on lost sales and profitability to be managerially significant.

Another implication of co-production is that labor affects not only costs but also sales directly. In manufacturing, although labor is a part of the production process, it is not a part of the end-product. Thus, the quality of a product in manufacturing can be made independent of labor through proper oversight and inspection. Defects can be identified and reworked ahead of sale. On the other hand, co-production in retail implies that store labor has a direct impact on sales through the customer-observed service quality. Marketing research shows service quality to be an important determinant of customer satisfaction. Maxham et al. (2008) describe a retail

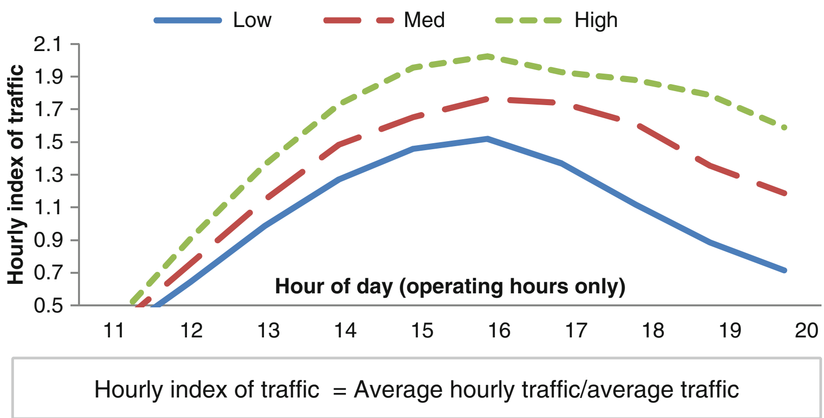

Fig. 6.1 Plot of hourly variation in traffic

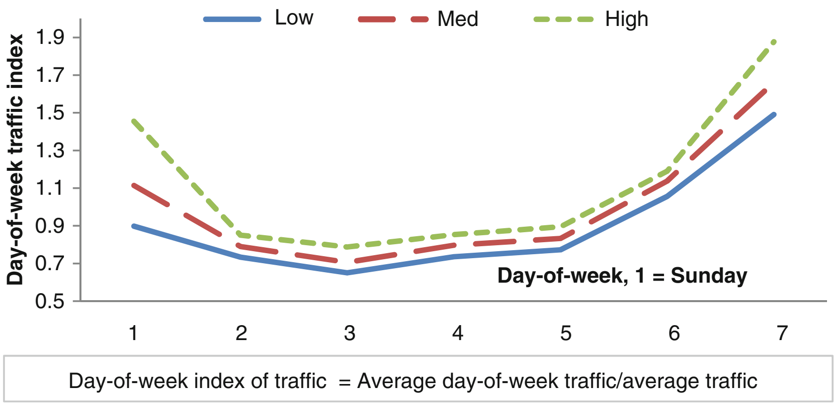

Fig. 6.2 Plot of day-of-week variation in traffic

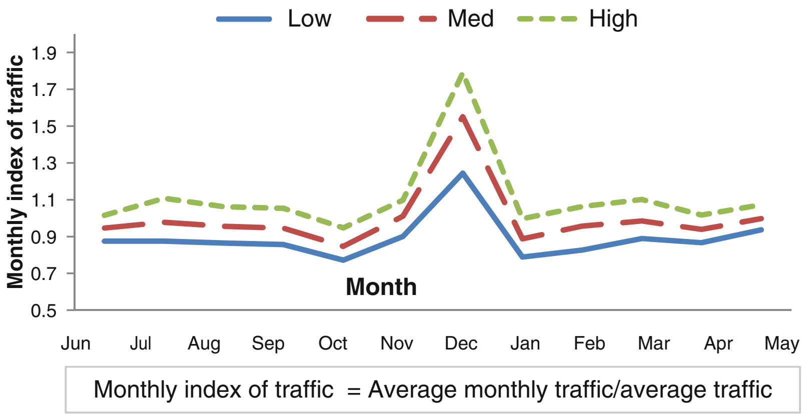

Fig. 6.3 Plot of monthly variation in traffic

value chain wherein perceptions and behaviors of front-line store employees influence customer satisfaction and intent, and ultimately store performance. Through empirical models that examine systems of relationships among employee job perceptions, employee performance, customer evaluations, and store performance, they find that employee perceptions exert a direct influence on customer evaluations, and that customer evaluations affect retail store performance (customer spending and comparable store sales growth). For this reason, retailers need to consider the skills of the associates before letting them perform customer-facing tasks.

Another important difference arising from co-production is that customers in a retail setting impose additional externality costs. While queue length of jobs can be optimized in manufacturing to minimize costs, customers at a retail store may join or leave queues in a way that is sub-optimal to the whole system. Finally, from a skill-set standpoint, retailers can often recruit workers with limited or no background in retail and train them before introducing them in the store. While this practice increases the pool of workers to recruit from, it can also lead to higher turnover, as these positions are often low paying ones. In contrast, the manufacturing setting tends to use smaller proportions of part-time and temporary workers as compared to the retail industry (BLS 2008; Lockard and Wolf 2012).

3.2 Empirical Research on Retail Labor in Operations Management

In this section, we provide a review of the empirical research on retail labor. Empirical research examining the impact of labor on retail store performance has been gaining importance in recent years. In order to empirically investigate the impact of labor on retail store performance, it is necessary to collect data on the amount of labor available in the store, the demand for labor, and store performance measures.

The amount of labor available in the store depends on the type of labor (e.g. full-time, part-time, temporary workers), the job description (e.g. store managers, sales associates, stockroom employees etc.), and the number of hours available for each employee (e.g. maximum number of hours available, break time, vacation time etc.). This information is usually available from personnel records. It is typically gathered at an aggregate level (e.g. bi-weekly or monthly) and used to generate wage-payroll data. More detailed information on the exact number of labor hours within each day would be available from a workforce management system or a scheduling system which uses the number of hours available for each employee and breaks them down into shift schedules. While the data from the workforce management systems would give a detailed breakdown of number of labor hours available, it might be aggregated across employees. For instance, the labor hours available for a given day would be the total manager-hours, full-time labor hours and part-time labor hours available for each hour of the day. Depending on the kind of store operation, the labor hours might be broken down into stockroom labor hours and sales associate hours.

The demand for labor depends on the type of retail store and their product characteristics. For instance, the level of sales assistance provided to shoppers in a specialty furniture store is significantly different from that provided at a discount super store. In addition, the demand for retail labor is also dependent on the level of store execution activities like unloading delivery trucks, stocking shelves, tagging merchandise, and maintaining the overall store ambience. While it is possible to estimate time requirements for standard activities such as unloading delivery trucks and stocking shelves, it is very difficult to set a time for sales-related activities, especially in a service-intensive store (Fisher and Raman 2010). Thus, unlike manufacturing or call-center settings, it is relatively much harder in a retail setting to estimate the exact number of hours required in the store from store activities alone. Hence, it is common practice to estimate the demand for retail labor from the level of sales or the level of traffic in the store. Store managers may then add additional hours required for store-execution activities to determine the total labor hours required in the stores. Store sales data are available from point-of-sale (POS) systems that are installed in almost all retail stores today. Data on traffic are harder to collect and requires a traffic counter to be installed in the stores. Stores that have traffic counters would have customer arrival data that can be aggregated and used for analysis. In some instances (e.g. grocery stores), where almost every customer who enters the store leaves with a purchase, the number of transactions from the POS data can be used as a proxy for traffic. However, in specialty stores (e.g. high-end electronics stores and specialty stores), it would be necessary to have access to traffic data to assess the true demand for retail labor.

Store performance measures include both quantitative measures (like revenue, profits, and conversion rate) as well as qualitative ones (like the level of service provided in the stores). Quantitative measures like sales, expenses, and profits are usually available from financial data. While information on store sales can be obtained from POS data, labor expenses are gathered from wage-payroll data, usually at the monthly level. In order to ascertain other expenses like inventory shrink, administrative expenses etc. it is necessary to have access to stores' financial (P&L) statements. These financial measures can be used to construct additional measures like labor productivity (e.g. sales per labor hour) and level of employee turnover. If traffic data are also available, then additional store performance measures like conversion rate (defined as the ratio of number of transactions to traffic), basket value (ratio of sales volume to number of transactions), and traffic to associate ratio (ratio of number of customers to number of sales associates) can be calculated. Qualitative measures on service quality are obtained from customer surveys.

The two most common challenges encountered in conducting empirical research on retail labor are data availability and dealing with endogeneity issues between labor and store performance.

As mentioned earlier, traditional data on retail labor and store performance have been available from POS transactions and wage-payroll. Due to the sensitive nature of these data, many retailers are reluctant to share them.[4] These data are typically aggregated before archiving. The POS data may be available on a daily basis, or even hourly basis, but payroll data are usually available only at a bi-weekly or monthly level. Retailers who have installed workforce management systems typically have labor data at a disaggregate level that are typically more useful for research on labor planning and scheduling. They usually have information on when

[4] In our experience, we find retailers to be particularly sensitive to sharing age and gender information when providing the payroll data.

and which department each person worked on different days of the week. These data can even be available at 15 min intervals. Retailers who have traffic counters can provide customer arrival data. However, unlike call center data where traffic data are usually accurate, retail traffic counters typically have some errors. While it is common for many technology firms to claim that these errors are less than 5 %, it would be useful to verify these data for their accuracy, if possible, before use in research. Finally, we note that while POS and payroll data are available for a long time period for most retailers, granular data on traffic and labor hours are typically only available for a shorter time period as these systems have not been in place for a long time.

Another important issue to consider when examining store labor is that of unobservable factors that may result in omitted variable bias. A key concern with examining the impact of labor on store performance is that labor typically gets scheduled based on certain anticipated events that the manager knows but is unknown to the empirical researcher. For instance, consider the case of store promotions. When store managers run store promotions, they may hire more labor in advance of these promotions. Store promotions lead to an increase in store traffic and store sales. Without data on store promotions, examining the impact of labor on store performance would lead to misleading inferences. So, it is necessary to either control for sales forecast, which account for store promotions and other anticipated events that affect sales, or use appropriate instruments to overcome the endogeneity bias.

In Sect. 3.2.1, we describe the empirical models that have used sales and payroll data to examine the overall relationship between retail labor and store performance. In Sect. 3.2.2, we describe empirical models that deal with customer traffic and staffing issues in retail stores. In these models, traffic data is used along with store sales and labor data to determine the relationship between demand, availability of labor, and store performance. A recent development in retail labor planning literature is the consideration of the type of retail labor available in the store and its impact on store profits. In Sect. 3.2.3, we highlight empirical models that leverage labor-mix data to examine its impact on store productivity and profits.

3.2.1 Relationship Between Retail Labor, Quality, Sales, and Profits Using Sales and Payroll Data

Store labor is an important driver of retail store performance. The benefits of having store labor include providing an increased level of sales assistance to shoppers and improving execution of store operational activities such as stocking shelves, tagging merchandise, and maintaining the overall store ambience (Fisher and Raman 2010), all of which lead to increased sales. Below, we look at two empirical models that examine the relationship between retail labor, service quality, and store performance.

Retail Labor and Basket Values

Netessine et al. (2010) use sales and payroll data from 311 stores of a large retail chain over a 3-year period to study the relationship between store labor and basket values. They collect monthly level data on sales, number of transactions recorded at checkout, the value of shopping baskets, and the total number of employee hours (full-time, part-time, and manager hours) budgeted for the store in a given month.

Based on these data, the authors derive two kinds of mismatches between sales and labor: *Planning mismatch,* which measures the quality of store labor planning using the month-to-month deviations (mismatches) between forecasts of store transactions and planned labor hours, and *Execution mismatch,* which measures the quality of store labor deployment using the month-to-month deviations between planned labor and actual labor deployment. The labor mismatches are calculated as a function of the correlation (r(.)) between two time series of corresponding variables. For example, for store i using monthly observations on transactions (TXN_i), labor plan hours ($PLAN\_HOURS_i$) and total employee hours ($TOTAL\_EE_i$), total labor mismatch ($TXNvsTOTAL\_EE_i$), planning mismatch ($TXNvsPLAN\_HOURS_i$) and execution mismatch ($TOTAL\_EEvsPLAN\_HOURS_i$) are calculated as follows:

$$TXNvsTOTAL\_EE_i = 1 - r(TXN_i, TOTAL\_EE_i);$$
$$TXNvsPLAN\_HOURS_i = 1 - r(TXN_i, PLAN\_HOURS_i);$$
$$TOTAL\_EEvsPLAN\_HOURS_i = 1 - r(TOTAL\_EE_i, PLAN\_HOURS_i)$$

The authors use basket value as a measure of store performance and find that the mismatches between store transactions and the total number of employees are negatively associated with basket value (significant at the 5 % level).

Next, they separate the total labor mismatch into planning mismatch and execution mismatch and find that while planning mismatch is negatively associated with basket values (significant at the 1 % level), the association between execution mismatch and basket values is not significant. They further break down execution mismatches based on type of labor (i.e. full-time labor, part-time labor and managers) and find high statistical significance for an association between full-time employee mismatch and average basket value. However, mismatches for part-time employees and store managers were not statistically significant. The regressions are run on cross-sectional data for each store and include control variables for demographics for each store such as household size, proportion of households with no children, and the proportion of the local population that is Asian or Hispanic.

Finally, the authors find that some stores are consistently better at planning staffing levels to meet traffic, while other stores are consistently better at executing a given plan, but no correlation appears between the ability to plan and the ability to execute well. For the retail chain in their study, they conclude that if managers were able to reduce staff planning mismatches by 50 %, the resulting revenue uplift would be 1.8 % of the current chain-level revenue. Eliminating 50 % of execution mismatches creates an additional revenue uplift of 2.4 %. In conclusion, the authors propose, as ideal, a switch from forecasted sales to forecasted traffic as a basis for labor planning.

Retail Labor, Quality, and Store Profits

Ton (2009) investigates the impact of store labor on store profits through its impact on service quality and conformance quality. In retail stores, increasing the labor level is likely to increase both conformance quality and service quality. For example, when store employees have more time, they are less likely to make errors in activities such as shelving merchandise or placing price tags on display shelves, and more likely to spend time with customers. In turn, sales are likely to be higher when products are shelved properly (Ton and Raman 2010) and salespeople are available to help customers in the purchase process (Fisher et al. 2006). Conformance quality is also expected to increase future sales at retail chains that use centralized merchandise planning systems, as the performance of these systems depends on conformance to in-store merchandising specifications and on accurate point-of-sale and inventory data (Raman et al. 2001). In addition to increasing sales, conformance quality is also likely to improve labor productivity and reduce shrink. Employees can shelve, replenish, and help customers find products more quickly, and fewer products are expected to be damaged or lost. Based on these arguments, Ton (2009) explores the relationship between labor, service and conformance quality, and profitability in retail stores.

Ton (2009) uses monthly data on labor, service quality, and profitability from 1999 to 2002 from 286 stores of large specialty retailer *Beta*. The amount of labor is measured as total labor dollars spent at a store in a given year and includes wages and benefits. The profit margin is defined as the operating income divided by sales. To measure service quality, she uses information from customer surveys that ask questions on five dimensions of service quality: tangibles, responsiveness, assurances, reliability, and empathy (Zeithaml et al. 1990). Ton (2009) also uses three metrics tracked by *Beta* to calculate conformance to the centralized decisions on merchandise planning and display: phantom-products, returns-conformance, and store-conditions. Phantom-products tracks the percentage of products that are in storage areas but not on the selling floor at the time of the physical audit. Returns-conformance tracks whether stores return the products they are supposed to return to the distribution centers. Store-conditions tracks whether stores conform to a wide range of standards related to the flow and storage of products. To create a composite measure of conformance quality, she standardizes each measure of conformance quality for each year by subtracting the mean and dividing by the standard deviation. In the final measure, returns-conformance and store-conditions scores are added and phantom-products scores are subtracted from the average standardized scores.

In the paper, Ton (2009) first tests for the relationship between quality (service and conformance quality) and labor and for the relationship between profit margin and labor. Next, she tests if the relationship between profit margin and labor is mediated by service quality and conformance quality. The regression models include fixed effects for each store and for each year. Store fixed effects control for time-invariant unobserved heterogeneity across stores, which might otherwise affect store labor, conformance quality and service quality, and profitability. The year effects control for factors, such as economic conditions and corporate policies, which if they change over time, will change for all stores. The control variables in the regressions include planning mismatch (measures the degree of mismatch between a store's payroll plans and its actual workload), and execution mismatch (measures the degree of mismatch between payroll plans and actual labor spending) for the different stores. Store monthly sales are used as a proxy for workload. Also included are full-time employees as a percentage of total employees to control for employee mix, employee turnover to control for tacit knowledge lost when employees leave, store manager turnover to control for management changes, units of inventory in a store to control for level of complexity in the operating environment, unemployment rate in a store's MSA (Metropolitan Statistical Area as defined by the Census Bureau) to control for differences in labor market conditions, and the number of competitors in the local market to control for competition.

The results indicate that increasing labor at a store is associated with higher conformance quality and service quality. Increasing employee turnover and departure of store managers are associated with a decrease in conformance quality and service quality. Higher planning mismatch and increased complexity in operating environment are associated with lower conformance quality but have no effect on service quality. Increasing the proportion of full-time employees has no effect on conformance quality but, surprisingly, a negative effect on service quality. Finally, she finds that a 1 standard deviation increase in labor is associated with a 10 % increase in profit margin.

Netessine et al. (2010) and Ton (2009) both conclude that most stores tend to understaff their stores. Fisher and Raman (2010) cite conversations with many retail managers and conclude that most retailers view labor as a cost, not an asset. To the managers, decisions about staffing trade off a known present cost—paychecks written to employees—against an unknown future benefit, namely, the incremental sales that would result from better staffing. Hence, managers tend to focus more on lowering staffing costs. Further, since the negative effect of having too little labor is often difficult to quantify, they posit that many store managers may place greater emphasis on minimizing payroll expenses to meet short-term performance targets. In the next section, we discuss papers that aim to quantify the impact of labor on store sales and profit by using more detailed data on store traffic, labor, and sales.

3.2.2 Relationship Between Store Traffic, Retail Labor, and Store Sales

In an effort to track the true sales potential in their stores, retailers have recently begun to install traffic counters in their stores. Traffic counters enable retailers to collect data on customer traffic and track conversion rate in their stores. The availability of this data has also opened up new avenues for research on retail labor. By combining traffic data with point-of-sale (POS) and labor data, it is now possible to estimate the true customer demand and the lost sales due to inadequate labor. Below we look at two papers that leverage traffic data with sales and labor data to examine these issues.

Effect of Traffic on Retail Sales Performance

Perdikaki et al. (2012) use data from 41 stores of an apparel retailer (*Alpha*) to study the relationship between store traffic, labor, and sales performance. They decompose sales volume into conversion rate and basket value. Increase in traffic would lead to an increase in sales, as higher traffic provides more opportunities for sales conversion. However, in the absence of adequate labor, increase in traffic could lead to higher levels of crowding and a decrease in service quality, both of which could lead to a decrease in sales. Thus, having adequate store labor could moderate the impact of traffic on store sales. Based on the above observations, the authors examine the relationship between traffic, labor and store sales; and study if higher store labor leads to greater positive impact of store traffic on store sales performance.

In addition to studying the impact of traffic, the authors also explore the impact of traffic variability on store sales. Stores with higher inter-day traffic variability may face higher traffic uncertainty, which could result in large errors when forecasting labor requirements for stores. Such large forecast errors would result in large mismatches between store labor required to manage in-store customers and actual store labor. Increased intra-day traffic variability could also lead to higher waiting time in queues and result in higher levels of abandonment. Further, higher levels of intra-day traffic variability could cause difficulties in scheduling labor for different hours of the day. This could lead to understaffing during certain hours of the day, resulting in lower service quality and lower sales performance. Hence, the authors test if greater inter- and intra-day traffic variability could lead to lower store sales performance. Finally, they also explore the implications of lower conversion rate on future sales potential by studying the relationship between conversion rate and traffic growth.

For the year 2007, the authors obtain the following types of data: (1) financial data (i.e., the number of transactions and store sales volume); (2) labor data (i.e., employee hours); and (3) traffic data. Sales performance for store i on day t is measured in two different ways: sales volume in dollars and the number of transactions that occur in the stores. These variables are divided by regular business hours for each store on each day of the week to obtain the average number of transactions $(ATXNS_{it})$ per hour and average sales volume per hour $(ASALES_{it})$. Similarly, total traffic and labor hours are divided by regular business hours to obtain average traffic per hour $(ATRAF_{it})$ and average labor hours per hour

($ALBR_{it}$). The authors calculate intraday traffic variability, using hourly data, as the ratio of standard deviation to mean traffic for that day as shown below:

$$\mu_{it} = \sum\nolimits_{h=1}^{H_{it}} TRAF_{ith}/H_{it}; \sigma_{it} = \sqrt{\sum\nolimits_{h=1}^{H_{it}} \left(TRAF_{ith} - \mu_{it}\right)^2 /H_{it} - 1}; TRAFVAR_{it} = \sigma_{it}/\mu_{it}$$

where $h = 1 \ldots H$ represent the store business hours. Inter-day traffic variability is calculated using the following AR model for traffic where δ_h denote holiday dummies, δ_m denote monthly dummies, and δ_d denote dummies for days of week.

$$TRAF_{it} = b_{i0} + \sum_{l=1}^{7} b_{il} TRAF_{it-l} + b_{i8}\delta_h + b_{i9}\delta_m + b_{i10}\delta_d + \varepsilon_{it} \tag{6.1}$$

Traffic variability is measured using the residuals (Eq. 6.1) as $TRAFUNC_i \equiv sd\,(\varepsilon_{it}/TRAF_{it})$ where sd(.) denotes the standard deviation. The authors include the following control variables. They calculate labor-traffic mismatch as the ratio of traffic to labor $\left(Mismatch_{it} = \frac{\sum\nolimits_{h=1}^{H_{it}} TRAF_{ith}/LBR_{ith}}{H_{it}}\right)$. This ratio is used as a proxy for service level. The authors also collect data on daily average temperature of each store location and the Dow Jones Industrial Average and obtain demographic data like averages on median household income and per capita income for 2007 by location. They use the number of other stores in the mall as a proxy for competition and run the following regression to determine the impact of traffic and labor on store sales. W_{it} denotes the vector of control variables

$$\begin{aligned} ASALES_{it} = {} & \vartheta_0 + \vartheta_i + \vartheta_1 ATRAF_{it} + \vartheta_2 ATRAF^2{}_{it} + \vartheta_3 ALBR_{it} \times ATRAF_{it} \\ & + \vartheta_4 ALBR_{it} \times ATRAF^2{}_{it} + \vartheta_5 TRAFVAR_{it} + \vartheta_6 ALBR_{it} \\ & + \vartheta_7 ALBR^2{}_{it} + \vartheta_8 W_{it} + \xi_{it} \end{aligned} \tag{6.2}$$

The authors find that store sales volume is an increasing concave function of traffic. For values of labor corresponding to mean, and traffic at mean plus 1 sd, increasing average traffic per hour by one unit increases average sales volume by \$8.14. For values of labor corresponding to the mean, and traffic corresponding to mean minus 1 sd, increasing average traffic per hour by one unit increases average sales volume by \$11.80. For values of labor corresponding to mean, mean minus 1 sd, and mean plus 1 sd, the marginal returns to traffic for the store with mean traffic are \$10.00, \$8.68 and \$11.32, respectively. This relationship is shown graphically in Fig. 6.4. Further, the authors find that store sales volume exhibits diminishing returns to labor and increases in intraday traffic are associated with lower sales per hour in stores. Replacing $ASALES_{it}$ with CR_{it} and ABV_{it} in Eq. 6.2 yield similar results, supporting a decreasing nonlinear relationship between traffic and conversion rate. This result is shown graphically in Fig. 6.5.

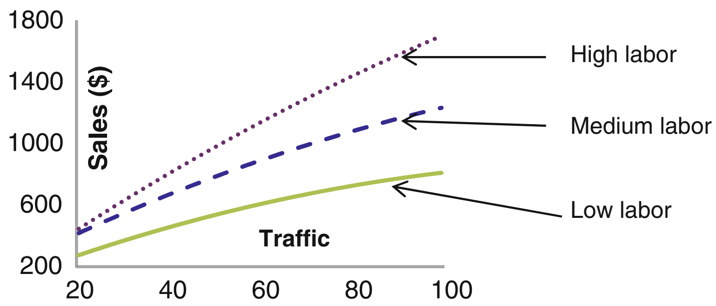

Fig. 6.4 Relationship between store traffic and sales for different levels of labor

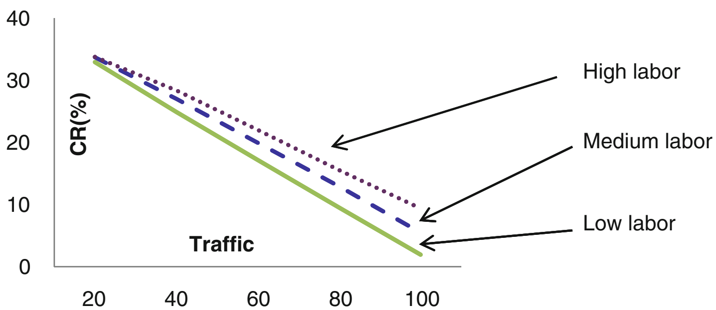

Fig. 6.5 Relationship between conversion rate and store traffic for different levels of labor

In addition, the authors find that increases in inter-day traffic variability are associated with lower sales per hour in stores. Further they find that an increase in conversion rate is associated with an increase in future traffic growth and that this relationship is statistically significant up to 5 months in the future.

Estimating the Impact of Understaffing in Retail Stores

Retailers walk a fine line between having enough labor in their stores to meet service requirements while maintaining low payroll costs. As pointed out by Netessine et al. (2010) and Ton (2009), quantifying the impact of understaffing on store sales and profits is necessary so that managers can make informed decisions on the level of labor to have in their stores. Mani et al. (2014) use detailed traffic data along with labor and sales data to investigate whether retail stores are understaffed and the impact of understaffing on lost sales and profits.

Using hourly traffic data along with POS (point-of-sale) and labor data for 41 stores of an apparel retail chain, the authors first calculate the required amount of labor for each store during each hour. They denote positive deviations from this required labor as understaffing and negative deviations as overstaffing. They follow two different approaches to labor planning. 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.

In the first approach, the authors use an empirical model motivated by the square-root staffing model from queueing theory to calculate staffing levels. For each store i in time period t, let $TRAF_{it}$ the number of customers arriving to the store. Then, the target staffing level (N_{it}) can be determined based on the following equation:

$$\begin{aligned} N_{it} = \delta_{0i} + \delta_{1i} TRAF_{it} + \delta_{1pi} TRAF_{it} \times \left(1_{p=1}\right) \\ +\delta_{2i} TRAF_{it}^{1/2} + \delta_{2pi} TRAF_{it}^{1/2} \times \left(1_{p=1}\right) + \xi_{1it} \end{aligned} \tag{6.3}$$

In the above equation, the authors introduce a dummy variable for peak hours to take into account changes in service rate and quality of service between peak and non-peak hours. The peak hours are determined as a 3-h window during which almost 60 % of store traffic arrives during the day. To quantify the impact of understaffing on sales and profits, they use the following sales and profit functions:

$$S_{it} = \alpha_i TRAF_{it}^{\beta_i} e^{-\gamma_i / N_{it}}; \pi_{it} = S_{it} \times g_{it} - N_{it} \times d_i \tag{6.4}$$

where S_{it} is the store sales, 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, π_{it} is the gross profit net of labor costs, and d_i is the marginal cost of labor. In this model, overall store sales are positively associated with labor, but an increase in traffic and labor increases sales at a diminishing rate, i.e., $0 < \beta_i < 1, \gamma_i > 1$. The difference between required labor and actual labor for each hour is denoted by ΔN_{it}. Let $1_{\Delta N_{it}>0}$ be an indicator function that takes the value of 1 when the store is understaffed $\left(\Delta N_{it} > 0\right)$, 0 otherwise. The lost sales and drop in profits in time period t when the store is understaffed can be represented as:

$$\begin{aligned} \Delta S_{it} &= \left[\hat{\alpha}_i TRAF_{it}^{\hat{\beta}_i} \left(e^{-\hat{\gamma}_i/\hat{N}_{it}}\right) - S_{it}\right] \times (1_{\Delta N_{it}>0}); \\ \Delta \pi_{it} &= (\Delta S_{it} \times g_{it} - \Delta N_{it} \times d_i) \times (1_{\Delta N_{it}>0}) \end{aligned} \tag{6.5}$$

where "^" indicates the coefficients estimated from the sales equation in Eq. (6.4). Thus, the authors' estimation of lost sales is based on the sales lift that the store would have experienced if it carried the predicted labor $\left(\hat{N}_{it}\right)$.

The authors perform a cluster analysis based on average traffic as well as average sales and divide their sample into weekdays and weekends based on similarities in traffic patterns (and sales patterns) across different days of the week.

For the stores in their weekdays sub-sample, the authors find that stores are understaffed 40.21 % of the time. When understaffing occurs, the magnitude of

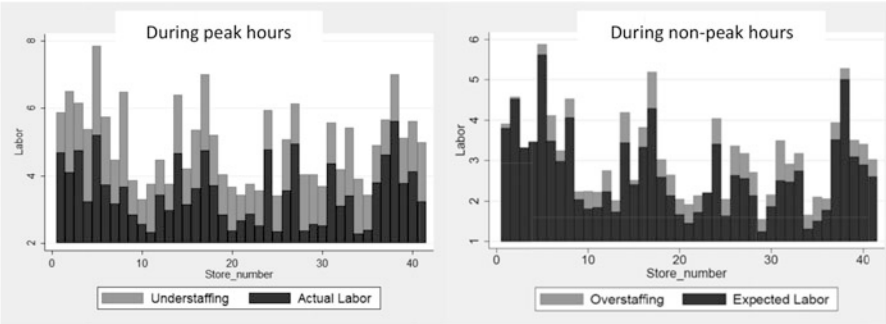

Fig. 6.6 Understaffing during peak and non-peak hours

understaffing is 2.10 persons; this level of understaffing represents a 33.27 % shortage as compared to the predicted labor. During peak hours, they find that the stores are understaffed 64.98 % of the time, and the average magnitude of understaffing is 2.31 persons. Figure 6.6 shows a graphic representation of understaffing during peak and non-peak hours across the 41 stores. Further, they observe a decline of 1.95 % in conversion rate when the store is understaffed (when compared to other peak hours when the store is not understaffed). They determine the average lost sales due to understaffing for the 41 stores during peak hours to be 8.56 %. Approximating g_i by the average gross margin for this retail chain and the labor cost, d_i, by the average wage rate for retail salespersons in that state, the authors find that this retail chain's average profitability will increase by 7.02 % if it eliminates understaffing during peak hours.

Next, the authors investigate the drivers of understaffing by studying the impact of forecast errors and scheduling constraints. They use 1-, 2-, and 3-week-ahead forecasts in place of actual traffic in Eq. (6.3) and calculate the predicted labor. As the forecast horizon increases from 1 to 3 weeks, the magnitude of understaffing as a percentage of predicted labor increases from 5.43 to 17.84 %. The sales lift decreases by 2.61 %, and the profitability improvement lowers by 2.56 % with use of a 1-week-ahead forecast of traffic. To examine how much of the observed understaffing can be explained by scheduling constraints, they consider 2-, 3-, and 4-h shifts in their analysis. They find that when scheduling labor with minimum shift lengths of 4 h, as opposed to 2 h, the magnitude of understaffing as a percentage of predicted labor increases from 7.23 to 28.74 %. The sales lift decreases by 3.76 % and the profitability improvement lowers by 3.52 % when they impose a 2-h shift length constraint. Finally, they explore the impact of the interaction of forecast errors and scheduling constraints on store profitability with the help of a simulation. As shown in Fig. 6.7, scheduling constraints exacerbate the negative impact of forecast error on store profits.

In the second approach, the authors use a staffing model based on a popular practice wherein the cost of labor in store is balanced with the contribution of labor to sales. Assuming that the store managers make optimal labor decisions at an

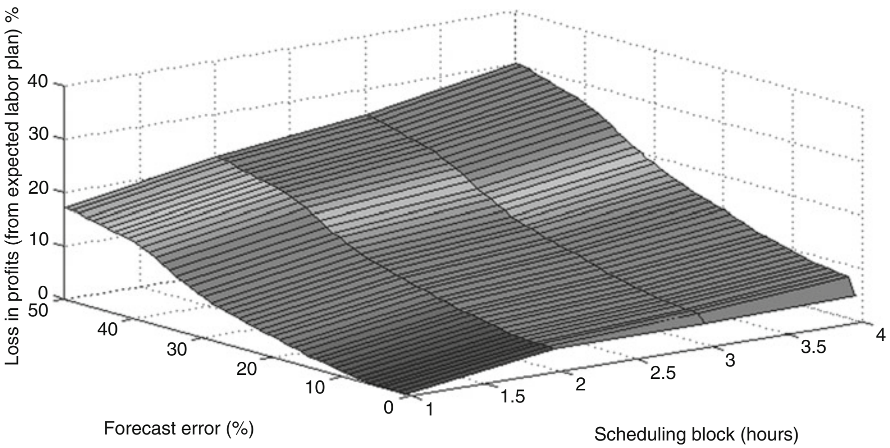

Fig. 6.7 Impact of forecast errors and scheduling constraints on store profits

aggregate daily level, the authors estimate the parameters of the model for each store using historical daily data on sales, traffic, and labor. They use the same sales and profit equations as in Eq. (6.4) but replace d_i with w_i 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 Eq. (6.4), yielding the following first-order condition for amount of labor to have in each store:

$$\gamma_i \alpha_i TRAF_{it}^{\beta_i} e^{-\gamma_i / N_{it}} g_{it} = w_i N_{it}^2 \tag{6.6}$$

The optimal labor plan (N_{it}^*) is the value of labor that is a solution to Eq. (6.6), given α_i , β_i , γ_i , w_i and store traffic ($TRAF_{it}$).

The authors use the generalized method of moments to estimate α_i, β_i , γ_i, and w_i. They 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 \$58.87 and \$22.43, respectively. Even stores within the same state, having the same average wage rate for retail salespersons, had very different imputed costs of labor. Also, the authors find that the imputed cost of labor is significantly higher during weekdays than weekends ($p < 0.001$). Based on this approach, the authors find that during peak hours, the stores were understaffed 68.21 % of the time, the extent of understaffing was 3.52 persons, and removing understaffing would lead to a sales lift of 7.21 % and increase profitability by 5.87 %.

3.2.3 Relationship Between Employee Turnover, Labor Flexibility, and Store Performance

In Sects. 3.2.1 and 3.2.2, labor mismatches were shown to have a negative impact on store performance. One way retailers can reduce these mismatches is to increase labor flexibility—through use of part-time and temporary workers. This proposition is attractive, as part-time and temporary workers generally incur lower payroll expenses since they do not receive the full benefits of full-time workers and can be given non-standard work schedules. Thus labor flexibility helps retailers handle traffic variability and schedule sales associates for a few hours to meet peak demand. However, employing part-time and temporary workers could impact the quality of sales assistance provided. Also, retailers may have to contend with higher turnover rates among part-time and temporary employees as these employees look to move towards more stable working environments. Below, we describe empirical models that investigate the relationship between labor-mix, employee turnover, and store performance in more detail.

Employee Turnover and Store Performance

Ton and Huckman (2008) posit that performance at mature retail chains depends highly on the successful execution of known activities such as processing inventory, shelving merchandise, responding to customer queries, and transacting sales on cash registers. In such a setting, they expect employee turnover to have a negative effect on firm performance due to operational disruption from employee departures, additional work that must be absorbed by remaining employees, and the loss of tacit knowledge and accumulated experience held by departing employees. However, to the extent that stores operate with a high degree of process conformance, they expect that knowledge concerning task performance will more easily transfer to new employees. Based on the above observations, they propose that in case of settings requiring high levels of knowledge exploitation, turnover will have a negative effect on operating performance, and this effect will be moderated by the level of process conformance present in these settings.

Ton and Huckman (2008) conduct their analysis on data collected on 268 stores of Borders Group superstores over 48 months (1999–2002). The average annual full-time employee and part-time employee turnover across Borders stores ranged from 49 to 69 % and from 94 to 114 % respectively. The authors obtained monthly turnover and performance data for each store from 1999 to 2002. Profit margin is defined as operating income divided by sales, and they exclude temporary workers from the analysis. Turnover is calculated as the number of employees who left a store during that period divided by the average number of employees working at the store during that period.

To develop a composite measure of process conformance, they use the average store conditions score and the average return pull list (RPL) score for each store for each year. In this setting, retailers are allowed to return unsold books to the publishers for a full refund minus the costs of shipping and handling. The RPL score is a returns conformance score based on the number of units returned divided by the total number of units that were supposed to be returned. The returns process, described in detail in the policy and process book, involves finding the books, packing them, and shipping them to the distribution centers.

Using these scores, they calculate the mean and standard deviation of store conditions and RPL scores across all Borders stores for each year. For each store, they standardize the yearly store conditions and RPL scores by subtracting the mean and dividing by the standard deviations. They combine these standardized scores to create the composite process conformance measure and divide stores into high and low process conformance stores. The authors run a regression of store performance against employee turnover. The regression model controls for several store level variables that vary over time. These include an indicator for turnover by store managers during the current month (to control for management changes); full-time employees as a percentage of total employees (to control for employee mix); total store payroll (to control for the total amount of labor used by the store); and the number of competitors in the local market and unemployment rate in the store's MSA (Metropolitan Statistical Area as defined by the Census Bureau) to control for labor supply. Specifications also include the fixed effects for each store, each year from 1999 to 2002, and each month of the calendar year. Next, to determine whether process conformance moderates the relationship between turnover and performance, they include an interaction variable between level of employee turnover with two categories of process conformance—high and low in the regression model.

The authors find that on average, turnover is associated with decreased store performance, as measured by profit margin and customer service. An increase of 1 standard deviation in full-time turnover at an average store leads to a reduction of 0.5 % in average customer service score and a 2.41 % decrease in average profit margin. The effect of turnover for low-process conformance stores is negative and significant. This negative effect is offset in stores with high levels of process conformance. Thus they conclude that turnover has a non-linear effect on performance.

Employee Turnover and Labor Productivity

Siebert and Zubanov (2009) examine the impact of turnover on labor productivity under two different work systems for sales assistants in a large UK retail organization. In the first system, known as the secondary system, the part-time employees receive less responsibility and specialist training, and have fewer promotion opportunities. Their pay is flat and is determined by salary surveys of similar occupations in the country. In the second system, known as the commitment system, full-time employees are given more responsibility, receive specialist training, and have their pay linked to performance. Managers expect that more turnover will occur under the secondary system than under the commitment system. The authors suggest that some level of sales assistant turnover is beneficial for performance and test if a negative or an inverted U shaped relationship exists between employee turnover and performance under these systems.

The authors collect data for 325 stores. This data has three parts. In the first part, they collect personnel records of all employees who worked at any time between 1995 and 1999. Individual records include age, gender, date hired, date left employment, and weekly contract hours (measure of hours worked). Since individual productivity was not available, the authors derive employee average data for each store in order to compute store annual average productivity. The second part of the data contains store information such as revenues, store square footage, number of floors, and store environment variables like city center and retail park. The third part of data consists of area-wide wage and unemployment data for the county in which each store was located. They exclude stores that were opened or closed during 1995–1999.

The authors use labor productivity as a measure of store performance. Labor productivity is calculated as annual sales per store, adjusted for inflation, divided by total annual hours worked in the store. Employee turnover is measured as the separation rate. They calculate separations from employee records for the sales assistants working fewer than 30 h per week (defined as part-time separations) and 30 or more hours per week (defined as full-time separations). They calculate the full-time equivalent of separation rate as the number of hours that leavers would have worked had they not left, relative to annual total hours worked. The authors run a regression model of labor productivity against employee turnover and include control variables for capital input, measured as store size in square feet, and labor input, measured as the sum of hours worked by every sales-assistant employed in a given store in a given year. They also include a number of variables relating to store environment and employee characteristics that might also affect productivity. The store environment variables were store location (eight dummies, including city center and retail park), type of product (three dummies, indicating more or less expensive goods), share of children's goods, number of floors, and area wealth and unemployment. By including these variables, the authors aim to control for the fact that it is easier to sell in prime locations, and sales volumes may vary with type of product, store configuration, and customer target group. To control for employee characteristics, they include sales assistants' weekly hours (shares of employees working 0–4, 5–14, 15–29, and 30+ h per week), which determine labor flexibility, the relative wage (sales assistants' pay relative to county average) and sales assistants' average age and tenure which helped control for workforce quality. They use a full-time-equivalent for average age and for employee turnover. Finally, they include 20 regional manager dummy variables to control for possible effects of regional management on store productivity.

The authors obtain two sets of regression results for the turnover-performance link: one—negative—for full-timers, who are managed under a commitment work system, and the other—an inverted U shape—for part-timers, managed under a secondary system. For workers managed under a secondary work system, they find a clear inverted U-shaped relationship between turnover and performance. The initial positive impact of part-time separations on productivity implies that less productive workers are more likely to separate. However, at the inflexion point, the benefits of improved job-worker match and workplace flexibility become offset by dysfunctional turnover and the loss of firm-specific capital. As for employees managed under a commitment system, the link is purely negative. Thus, the costs of core worker turnover appear to be higher than the costs of secondary worker turnover. Finally, they find that the effect of full-time separations is exacerbated by secondary turnover.

The authors calculate that if all the stores operated at the optimum level of full-time turnover instead of at their observed level, the organization would gain 0.3 % of total sales over the period of 1995–1999. Also, the organization can gain up to 1.4 % in productivity by choosing the optimum levels of full- and part-time turnover, which translates into £0.73 per each hour worked per year on average. Finally, summing up individual productivity gains from moving to optimal part-time turnover for all stores and years, they find that overall gain for the organization would be 0.6 % of the total sales over the period of 1995–1999.

Labor Flexibility and Store Performance

Flexible resources, in the form of part-time or temporary workers, create volume flexibility that can affect profitability through either sales or expenses. Individual flexible labor resources may be less productive than full-time workers, as they might have less capability and fewer qualifications than full-time resources. Having too many flexible resources may lead to an increase in co-ordination costs and poor store execution, which could in turn lower sales. On the other hand, flexible resources may increase sales because they provide firms with a dynamic adjustment option—they offer a greater ability to match staffing within a day or within a week.

Flexible resources might help reduce expenses as flexible labor resources are usually paid less than permanent, full-time labor. They can also be retained for fewer working hours than full-time associates, thereby reducing idle labor expenses. On the flip side, having too many flexible resources could lead to an increase in cost due to more frequent hiring, firing, and training costs. Based on the above, Kesavan et al. (2013) test for the following: an inverted U shaped relationship between flexible labor-mix and store performance measures like store sales and store profitability; and a U shaped relationship between flexible labor mix and store expenses.

The authors obtain data from 445 *RetailCo* stores for the period of July 2009 to August 2011. They collect data from three departments—finance, HR, and store operations—and obtain monthly financial statements data for each of the 445 stores. Statements contain stores' revenues and detailed information about expenses (e.g., fleet expenses, administrative expense, inventory shrink, etc.). They also received 30-day monthly forecasts of sales, labor hours, and payroll for each of the stores. HR data provided information on each employee who worked in the store for each of the 26 months. The information included whether each employee was full-time, part-time, or temporary. The store operations data contained weekly information on the actual hours each employee worked aggregated to monthly level to match the financial data.

The authors measure store performance using sales, expenses, and profits. Monthly store sales are calculated as the total revenue net of returns. Monthly store expenses include all expenses in the store, including labor costs related to salaries and commissions paid, employee related costs connected to relocation and training, occupancy costs resulting from rent and property taxes, administrative expenses related to accidents and insurance, and inventory related costs including insurance shrink, and changes. Labor-related costs account for slightly over half the total expenses in the store.

The store profit represents the before-tax profit for each individual store for that month. It is a function of sales, expenses, and cost of materials. To normalize the performance measures for level of activity to enable comparison across stores and time, the authors divide each of the metrics by average monthly sales for that store. Part-time labor mix is defined as the ratio of part-time to full-time employees and temporary labor mix is defined as the ratio of temporary to full-time employees.

The authors regress the store performance measures (sales, expenses and profits) on the two types of labor mix. They include both linear and square terms of labor mix to capture the non-linear relationship. They also include store fixed effects, region-specific monthly indicator variables to control for seasonality in a year and region-specific time effects to control for seasonal effects that are common to all stores in a given region. The authors also include controls on sales forecast, employee turnover amongst part-time and full-time workers and actual labor hours.

The results show that both part-time and temporary labor-mixes demonstrate an inverted U-shaped relationship with sales. Temporary labor mix has a U-shaped relationship with expenses, while part-time labor mix has a decreasing, concave relationship with expenses. The authors also find an inverted-U shaped relationship between temporary and part-time labor mix and profitability. Based on counterfactual analysis, they show that temporary and part-time workers can increase store sales by 11.5 % over the monthly average during the peak demand period. Further, the volume flexibility offered by these flexible resources can increase profitability by 28 % over the monthly average during the same period.

3.3 Other Relevant Literature

Operations management literature has a long history of studies that use queueing theory-based staffing models to determine service requirements (Hassin and Haviv 2003). These include service settings, such as manned service-desks in retail stores, check-out counters, bank tellers, deli take outs, airport kiosks, theaters, etc., in which people form a queue in front of the counter to wait for service. Using information on arrival rates, service time, and abandonment rates, the models are used to determine labor requirements to satisfy a service level constraint. In the context of a retail setting, most of the papers model retail stores as an Erlang C (or Erlang A) queue to determine the optimal number of retail workers (Berman and Larson 2004; Berman et al. 2005; Terekhov and Beck 2009). However, empirical research that examines staffing models with retail data is limited, as large-scale traffic data have only recently become available. Exceptions include Lu et al. (2013), Mani et al. (2014) and Tan and Netessine (2014). Lu et al. (2013) use queue data from a deli counter in a supermarket to show that customers focus on the length of the queue without adjusting sufficiently for the speed at which it is served. Mani et al. (2014), as explained earlier, use an empirical model motivated by square-root staffing model to determine the extent of understaffing and overstaffing in retail stores. Tan and Netessine (2014) use operational data from a restaurant chain to show an inverted U-shaped relationship between workload and performance and demonstrate how staffing capacity staffing capacity can be leveraged to optimize workload and increase sales.

As more granular data from traffic counters and other new technology become available, further research on the application of detailed staffing models to retail store operations may become possible.

4 Retail Technologies: Past, Present, and Future

For many decades now, the ubiquitous retail store has been identified with sales associates helping customers in their purchase decisions. Recent consumer and retailer research indicates two trends for the near future: first, that the store will continue to be the channel through which retailers receive the largest proportion of their revenue; and second, that, in general, consumers continue to prefer to shop and buy in the store (Gartner 2013a, b). While the storefront itself is slowly evolving from a simple brick-and-mortar presence to a hub of physical and virtual activity, optimization on store labor is beginning to take center stage in store operations.

In 2003, Gartner predicted that by 2013 retail stores will operate with 20 % less labor because of innovations in in-store retail technologies. Advances in customer-assistance technology (like automated merchandising solutions and self-checkout counters) and work management applications (including integration with real-time information on demand) were predicted to be two drivers of this transformation. At the time of this prediction, a high level of emphasis was placed on operational efficiency through both widespread deployment of point-of-sale terminals and electronic data interchange (EDI) linkages that helped retailers share demand information with supply-chain partners, and adoption of workforce management systems to help plan and schedule labor in store. While labor scheduling tools were not new to retail, they were largely deployed independent of other systems.

By 2007, most point-of-sale (POS) technologies had matured and been widely adopted by many mainstream retailers. Emerging technologies now focused primarily on improving store execution. In this context, end-to-end workforce management applications were developed that would help retailers balance workload sent to stores, manage the tasks and activities within stores, monitor store compliance, and quickly collect and analyze store feedback. Alongside the development of workforce management solutions, many stores had also begun to experiment with traffic counters—to count customer traffic in stores—and in-store cameras—to prevent theft in their stores, in an effort to make store operations more efficient. Thermal imaging techniques, borrowed from the defense and manufacturing industries were just being commercialized for application in the retail industry. Non-intrusive intelligent sensors could be used to detect customer movement and hot spots in a retail store. The archival data is used to calculate store performance measures like conversion rate and basket values as well as customer service metrics like average queue length and average wait time. These sophisticated traffic-counting technologies were considered to be a technology trigger or breakthrough with huge potential for improving operational efficiency and customer service at the same time. While, over the last few years, quite a few success stories prove the usefulness of traffic counting and workforce management solutions, most retailers still tend to use these real-time systems as stand-alone applications (for example, focusing exclusively on implementing traffic counters or queue management systems). Examples of vendors providing traffic counters for retail applications include Shopper Trak, SMS, and Sensource. More advanced technologies that use GPS to track customers in store, identify behavior of new and repeat customers, and analyze impact of promotional displays in store are also available today from vendors like Euclid Analytics, Goliath Solutions, and Retail Solutions. During this same period, workforce management systems have also further evolved into labor standards systems, scheduling systems, and task management systems.

The advances in store technologies in the last decade have led to an explosion of data available to retailers; many retailers consequently lament that they are "drowning in numbers but have very little actionable insights" (Fisher and Raman 2010). At a typical retailer, real-time data is now available through multiple touch points, such as POS transaction log, customer traffic counters, video over IP (network video), radio frequency identification devices, location-aware applications, and remote monitoring of appliances, including heating, ventilation, and air conditioning. Understandably, there is now a clamor for analytic applications that would help retailers transform this massive data into useful knowledge. In fact, managing and optimizing on big data continually ranks, along with cost containment, among the top 10 priorities for retailers over the next 3–5 years. Industry research on state-of-the art analytical applications shows a similar trend. Gartner's 2013 analytics applications hype cycle classifies real-time store monitoring platforms as being close to the "peak of inflated expectations" phase of the technology life cycle, i.e., technologies that have shown promise in some early adopters and, if successful, can gain widespread adoption in the next 5–10 years.

Real-time store monitoring platforms deliver store activity monitoring on dashboards by bringing together signals and alerts from real-time data sources available in the retail stores. These include inputs from traffic counters, queue management sensors, point-of-sale transaction logs, remote sensors on in-store devices, and RFIDs.

The next challenge lies in combining these real-time feeds onto a single platform so that a store manager can have a comprehensive view of what is happening in their stores. Here, complex algorithms are used to analyze real-time signals, and exception reports or alerts are generated based on user-defined metrics.

One example of real-time store monitoring technology is use of information from infrared sensors above store checkout lanes to calculate average queue lengths and wait times in real time. When the queue length or wait time reach a particular threshold limit, store staff is either reassigned to open additional counters, or more self-checkout lanes are opened to help reduce congestion. At the back end, based on the enormous amount of data collected, the system can also both determine the optimum number of checkouts needed and project traffic congestion and service requirements in future. When combined with a labor management tool, customer wait-time information is linked to labor scheduling to improve labor efficiencies throughout the store. Another example of advancements in real-time store monitoring technology is the use of digital video surveillance systems, coupled with analytics software, to track customer behavior, such as dwell times. Dwell time—the time customers spend in different points in the store—allows retailers to gauge the effectiveness of displays, signage, and promotions. This system allows retailers to get valuable data and insights on every part of the store, from entrances and aisles to customer service, dressing rooms, and even bathrooms. Recent advancements allow retailers to conduct on-going traffic and conversion-rate analysis not only by store but also by aisle and display, on down to the SKU level. They can also use this information to decide which sections (or product categories) in a store might require more sales assistance and allocate labor accordingly. Some vendors in this space are Brikstream, BVI RetailNext, Irisys, Scopix, Retailigence, and Re Tel Technologies.

As with the adoption of any new technology, it is important that retailers consider not only the immediate costs and benefits but also how such a technology will aid various functional roles in their businesses. For example, although traditionally conversion has been the responsibility of the marketing (and merchandising) department, real-time conversion data can aid in judiciously allocating sales associates in the store. This emphasis on integrating demand information with staffing decisions is even more important today when store labor costs run between 10 and 13.5 % of the typical retailer's revenues and are set to rise dramatically as a result of changes in labor supply and the increasing volume of store tasks to be performed in the store (Forrester Research 2009). The biggest worry for many store managers is that tasks are loaded onto the store without visibility about the amount of labor required to execute them. Thus, technologies that help tie labor requirements with store activities and customer demand present retailers with a tremendous potential to streamline labor decisions while maintaining a high level of customer engagement in the stores. These technologies may have the potential to transform the retail store into a data-rich enterprise and thus pave the way for the use of more analytical models and decision support tools to improve store operations.

5 Future Research and Conclusions

Retail labor is an emerging area of research and holds exciting prospects for several reasons. First, the intense competition with online retailers has led many brick and mortar retailers to take a closer look at the in-store experience offered to their customers and find ways to distinguish themselves based on customer service. While retailers have traditionally cared about retail labor because of its huge impact on sales and expenses, surprisingly we find the penetration of analytical techniques for workforce management to be limited in the retail industry. It is unclear why retailers who invest millions of dollars to drive traffic into the stores through marketing activities would not invest sufficiently in labor planning to ensure that the incoming traffic is converted to sales. However, this situation is likely to change. We observe many new start-ups in the area of traffic counting and in-store technology that offer retailers new, hitherto unavailable, data. These data present an excellent opportunity for retailers to transform their store operations to enhance productivity and compete effectively with online retailers.

Second, the availability of new data could make it possible to answer questions that were difficult to do so earlier. For example, the availability of store traffic data now allows researchers to examine the impact of labor on conversion rate (Perdikaki et al. 2012), a metric that has been long been tracked in other settings such as online retailers. Other performance metrics, such as dwell time, the amount of time spent by customers in the store, and frequency of customer visits, could be examined in future research. In addition, the integration of online and brick-and-mortar operations raises new and interesting questions around the role of store labor and the design of its incentives. Prior research on store manager behavior has shown that change in incentives can have significant impact on store performance (DeHoratius and Raman 2007).

Third, enormous scope exists for applying analytical techniques to improve labor planning. A large body of research has addressed the labor planning issues in manufacturing, but such research is absent in retail. As explained in Sect. 3.1, the presence of significant differences between manufacturing and retailing necessitate studying retail labor as an independent problem. Fisher (2004) state that a retail store is an amalgam of a factory and sales office, so labor planning solutions in retail can potentially build on prior research in manufacturing but would need to additionally account for the differences that arise due to co-production. Another area of research would be to examine how to apply queueing theory to the retail setting. While queueing theory holds large prospects to improve retail store operations, the complexity of retail store operations offers new opportunities for extensions. For example, several aspects—customers being able to complete most activities without the help of sales associate; associates being able to multi-task by dealing with none, one, or multiple customers simultaneously; and associates performing different types of activities such as stocking, cashiering, helping customers, etc.—need to be accounted for appropriately before applying queueing theory to retail stores.

References

Anderson, E. G. (2001). The nonstationary staff-planning problem with business cycle and learning effects. *Management Science, 47*(6), 817–832.

Bard, J. F., & Wan, L. (2008). Workforce design with movement restrictions between workstation groups. *Manufacturing and Service Operations Management, 10*(1), 24–42.

Bechtold, S. E., & Jacobs, L. W. (1990). Implicit modeling of flexible break assignments in optimal shift scheduling. *Management Science, 36*(11), 1339–1351.

Berman, O., & Larson, R. C. (2004). A queueing control model for retail services having back room operations and cross trained workers. *Computers and Operations Research, 31*(2), 201–222.

Berman, O., Wiang, J., & Sapna, K. P. (2005). Optimal management of cross-trained workers in services with negligible switching costs. *European Journal of Operational Research, 167*(2), 349–369.

BLS. (2008). *Involuntary part-time work on the rise. Issues in labor statistics*. Washington, DC: U.S. Bureau of Labor Statistics.

Brownell, W. S., & Lowerre, J. M. (1976). Scheduling of work forces required in continuous operations under alternative labor policies. *Management Science, 22*(5), 597–605.

Daniels, R. L., Mazzolla, J. B., & Shi, D. (2004). Flow shop scheduling with partial resource flexibility. *Management Science, 50*(5), 658–669.

Dantzig, G. (1954). A comment on Edie's traffic delay at toll booths. *Operations Research, 2*(3), 339–341.

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

Dill, W. R., Gawer, D. P., & Weber, W. L. (1966). Models and modeling for manpower planning. *Management Science, 13*(4), 142–167.

Fisher, M. (2004). To you it's a store. To me it's a factory. *ECR Journal, 4*(2), 9–18.

Fisher, M. L., & Krishnan, J. (2005). *Store level execution at Wawa. Case study*. The Wharton School, University of Pennsylvania, Philadelphia.

Fisher, M. L., Krishnan, J., & Netessine, S. (2006). Retail store execution. *Working paper*, University of Pennsylvania, Philadelphia, PA.

Fisher, M. L., & Raman, A. (2010). *The new science of retailing*. Boston, MA: Harvard Business School Press.

Forrester Research. (2009). *Filling the store labor productivity gap*. Cambridge, MA: Forrester Research.

Fryer, J. S. (1974). Labor flexibility in multi-echelon dual-constraint job shops. *Management Science, 20*, 1073–1080.

Gartner. (2003). *Emerging trends and technologies in the retail industry. IR COM-21—358*. Stamford, CT: Gartner.

Gartner. (2013a). *Agenda overview for retail, IR G00246684*. Stamford, CT: Gartner.

Gartner. (2013b). *Hype cycle for analytic applications, IR G00251137*. Stamford, CT: Gartner.

Hassin, R., & Haviv, M. (2003). *To queue or not to queue: Equilibrium behavior in queueing systems*. Norwell, MA: Kluwer.

Holt, C. C., Modigliani, F., & Simon, H. A. (1956). Linear decision rule for production and employment scheduling. *Management Science, 2*(2), 159–177.

Hopp, W. J., Tekin, E., & VanOyen, M. P. (2004). Benefits of skill chaining in production lines with cross-trained workers. *Management Science, 50*(1), 83–98.

Karmarkar, U., & Pitbladdo, R. (1995). Service markets and competition. *Journal of Operations Management, 12*(3), 397–411.

Kesavan, S., Staats, B. R., & Gilland, W. (2013). Volume flexibility in services: The costs and benefits of flexible labor resources. *Working paper*. University of North Carolina.

Kunreuther, H. C., & Morton, T. E. (1974). General planning horizons for production smoothing with deterministic demands. *Management Science, 20*(7), 1037–1046.

Lapré, M. A., & Nembhard, I. M. (2011). *Inside the organizational learning curve: Understanding the organizational learning process*. Norwell, MA: Now Publishers Inc.

Lippman, S. A., Rolfe, A. J., Wagner, H. M., & Yuan, J. S. C. (1967). Optimal production scheduling and employment smoothing with deterministic demands. *Management Science, 14*(3), 127–158.

Lockard, C. B., Wolf, M. (2012). Monthly labor review. *135* (1), 84–108.

Lu, Y., Olivares, M., Musalem, A., & Schilkurt, A. (2013). Measuring the effect of queues on customer purchases. *Management Science, 59*(8), 1743–1763.

Mani, V., et al. (2014). Estimating the impact of understaffing on sales and profitability in retail stores. *Production and Operations Management*. doi:10.1111/poms.12237.

Maxham, J. G., III, Netemeyer, R. G., & Lichenstein, D. R. (2008). The retail value chain: Linking employee perceptions to employee performance, customer evaluations, and store performance. *Marketing Science, 27*(2), 147–167.

Morris, J., & Showalter, M. (1983). Simple approaches to shift, days-off and tour scheduling problems. *Management Science, 29*(8), 942–950.

Mui, Y. Q. (2007). Circuit city cuts 3,400 'Overpaid' workers. Retrieved Dec 1, 2013, from http://www.washingtonpost.com/wp-dyn/content/article/2007/03/28/AR2007032802185.html

Netessine, S., Fisher, M. L., & Krishnan, J. (2010). Labor planning, execution, and retail store performance: An exploratory investigation. *Working paper*. The Wharton School, University of Pennsylvania.

Parasuraman, A., Zeithaml, V., & Berry, L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.

Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of retail store traffic on conversion rate and sales. *Manufacturing and Service Operations Management, 5*(2), 79–141.

Raman, A., DeHoratius, N., & Ton, Z. (2001). Execution: The missing link in retail operations. *California Management Review, 43*(3), 136–152.

Siebert, W. S., & Zubanov, N. (2009). Searching for the optimal level of employee turnover: A study of a large U.K. retail organization. *Academy of Management Journal, 52*(2), 294–313.

Tan, T. F., & Netessine, S. (2014). When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science, 60*(6), 1574–1593. http://dx.doi.org/10.1287/mnsc.2014.1950.

Terekhov, D., & Beck, J. C. (2009). An extended queueing control model for facilities with front room and back room operations and mixed skilled workers. *European Journal of Operational Research, 198*(1), 223–231.

Thompson, G. M. (1995). Improved implicit optimal modeling of the labor shift scheduling problem. *Management Science, 41*(4), 595–607.

Ton, Z. (2009). The effect of labor on profitability: The role of quality. *Working paper*, Harvard University, Boston.

Ton, Z., & Huckman, R. S. (2008). Managing the impact of employee turnover on performance: The role of process conformance. *Organization Science, 19*(1), 56–68.

Ton, Z., & Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management, 19*(5), 546–560.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing, 60*(2), 31–46.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. New York, NY: The Free Press.