Labor Planning, Execution, and Retail Store Performance: an Exploratory Investigation[1]

Serguei Netessine Marshall L. Fisher

Operations and Information Management Department
The Wharton School, University of Pennsylvania
netessine@wharton.upenn.edu, fisher@wharton.upenn.edu

Jayanth Krishnan
International Monetary Fund, Washington, DC
jkrishnan@imf.org

February 2009, revised February 2010

Abstract: After the cost of goods sold, store labor expense is the largest cost component in the retailing industry. As a result, developing and executing labor plans is a key task for retailers. In this paper we use private data provided by a large retail chain to assess the relationship between labor planning and execution practices and the average transaction (basket) value of a retail store location. We find that consumer basket value varies greatly from store to store and that there is a strong cross-sectional association between labor practices at different stores and basket values. In particular, matching labor deployment to store traffic (rather than to the forecast of sales, as many retailers currently do) is associated with larger basket values. We further separate the task of labor management into the planning component (i.e., creating a labor plan that matches traffic patterns) and the execution components (i.e., deploying part-time employees, full-time employees and managers to match the labor plan), and we find that stores with better plans and stores with better execution of these plans for full-time employees (but not for part-time employees) demonstrate significantly higher basket values. We obtain these results after controlling for customer demographics and product variety, which are also significant in explaining basket values. Our findings suggest that modest improvements in employee scheduling and in execution of the schedule can result in a 3% sales lift at moderate, or even no, additional cost.

[1] We thank the three anonymous referees and the associate editor for their constructive suggestions, and the Fishman-Davidson Research Center, P&G-Gillette Corporation and the International Commerce Institute for their generous financial support for this project. We are also grateful to Andy Buteux, Nicole DeHoratius, the participants of the 2007 COER conference in Boston, and numerous individuals interviewed for this project for their valuable insights and ideas.

"No matter how you slice it in the retail business, payroll is one of the most important parts of overhead, and overhead is one of the most crucial things you have to fight to maintain your profit margins. That was true then, and it is still true today."

Sam Walton: Made in America

1. Introduction

After the cost of goods sold, expenses related to hiring, training, and employing store labor constitute by far the largest component of a retailer's costs, often accounting for 10 to 20% of sales and sometimes for more than 50% of operating costs. Furthermore, labor costs are currently rising at a higher rate than revenues or gross margins, mainly due to increases in employee benefits (IHL Consulting 2005). Not surprisingly, issues of payroll are often on the minds of retail CEOs and leading retail companies expend significant resources to ensure that their stores are efficiently staffed to serve fluctuating customer demands. For example, many prominent retailers including RadioShack, Sears, and J. Crew have moved to computerized store labor scheduling systems (Maher 2007) and, according to AMR research, estimated revenues for work-force management systems are growing by 15 to 20% each year. While traditional approaches of scheduling employees are typically based on previous year's sales figures, there is also a new wave of software products that matches labor to store traffic. For example, Wal-Mart has announced the introduction of a new computerized scheduling system that "will start moving many of its 1.3 million workers from predictable shifts to a system based on the number of customers in stores" (see Maher 2007). Likewise, Kronos (2006) describes a similar system at Payless ShoeSource.

Attempts to optimize the planning and execution of store labor are much publicized and a variety of optimization tools exist to help managers with this task. Nevertheless, in our experience, even the best in-class labor scheduling systems lack one crucial ingredient: the ability to determine the impact of workforce management policies on revenue. Typically, labor planning is done at retail headquarters, with a lot of thought but little input from store managers going into it. Then the plan is given to the store manager, who must do whatever is necessary to comply, despite dealing with such factors as employee absenteeism, vacations and sick leaves, a lack of qualified candidates willing to work at minimum wage, the inflexibility of working parents' schedules, and the necessity to cover sudden surges in the demand for labor. All of these issues must be handled within the overall budget constraints imposed by the corporate office. Additionally, the store manager needs to attend to a variety of other issues within the store that are not related to labor scheduling (e.g., customer complaints, financial reporting, community outreach, etc.). For all these reasons, both the planning process and the execution of the labor deployment plan are far from straightforward.

In our conversations and research engagements with a number of major retailers we found that, while labor deployment is often tightly planned, there is currently no agreement on whether this planning should be done using a forecast of sales or a forecast of store traffic, and both approaches are currently used. Further, few retail executives know or even have ways to assess how store managers execute these plans, since the oversight of store managers' execution of the plan is often limited to making sure that stores do not exceed the allocated payroll budget. Moreover, many store managers are provided little information about how the planning was done and hence may doubt whether the plan is worth following. Others may simply follow the plan as closely as possible and focus their attention on other issues. This often leads to inconsistency in the execution of labor deployment. For example, Fisher and Krishnan (2005) discuss how one manager at the Wawa convenience store chain was particularly careful about monitoring store traffic and acting accordingly, whereas managers of other Wawa stores were not.

Much of the academic literature on retailing has focused on issues such as selecting products, advertising, forecasting demand, and managing supply chains, whereas issues related to practical impediments to labor planning and execution of the plan remain relatively unaddressed. The goal of this paper is to assess and quantify the importance of store labor planning and execution using empirical data. We hypothesize that the basket values of a retail store should be positively associated with the quality of labor planning and execution of the plan. We propose novel measures of the quality of store labor planning and execution. Specifically, we measure the quality of store labor planning using the month-to-month deviations (mismatches) between forecasts of store traffic and planned labor hours: the higher the deviation, the lower the quality of planning. Furthermore, we measure the quality of store labor execution using the month-to-month deviations between planned labor and actual labor deployment. Finally, we suggest that matching full-time labor to the plan is more important than matching either part-time labor or managers. We test our hypotheses using data from a large retail chain. We analyze the performance of each location as captured by basket value (i.e., the dollar value of each transaction) because, we argue, basket value is more likely to be directly affected by store labor policies than other performance metrics. Basket value is clearly of great interest to retailers, since it is directly linked to revenue.

We find significant cross-sectional variation in basket values that, in part, is explained by the quality of labor planning and execution. In particular, our econometric analysis supports the hypotheses that better store labor planning based on traffic through the store and better execution of the plan are positively associated with basket values. Specifically, we find that matching store labor to traffic (or its forecast) is superior to matching store labor to sales and is associated with higher basket values. Further, we find that matching full-time labor to the plan is also associated with higher basket values. Interestingly, the same is not true for either part-time labor or managers: we explain this finding by the fact that these two employee groups usually perform back-office functions while full-time employees perform front-office functions that directly affect basket value. We estimate the sales lift that the chain is likely to generate if corrective measures to adjust short-term staffing levels are taken and we find that the total sales lift for the entire chain could be approximately 1.4% through modest improvements in labor planning and 1.7% through better labor execution. Combined, these estimates exceed annual comparable-store sales growth reported by the company in the year of our study.

The key contributions of this paper are (1) to propose new metrics for store labor planning and execution, (2) to provide evidence that better labor planning and execution are associated with better basket values, (3) to show that labor planning against traffic is superior to labor planning against sales and (4) to estimate the financial significance of better planning and execution as well as to recommend specific steps that are likely to improve the performance of retail stores. The remainder of the paper proceeds as follows. In Section 2 we review the literature. In Section 3 we formulate our research hypotheses, in Section 4 we describe our data and preliminary analysis, and in Section 5 we provide the econometric specifications of the model. Section 6 discusses the results of the estimation and Section 7 provides robustness analysis of our results. Finally, Section 8 concludes with a discussion of the managerial implications and limitations of our findings.

2. Literature Review

The literature on retailing is vast and the science of retailing combines a variety of topics (see Fisher and Raman 2010). The literature stream relating most closely to our work is that addressing issues of workforce planning and scheduling. Salmon (1989) was the first to note the importance of retail store execution. There is a rich literature in operations research which uses mathematical programming models (see, e.g., Bechtold and Jacobs 1990), but there are few papers that empirically analyze the role of the workforce in retail store performance. Lusch and Serpkenci (1990) study the relationship between four personal difference variables and job outcomes for retail store managers. Hise et al. (1983) conduct a cross-sectional analysis of a retail chain using survey data and find that the number of employees and the store managers' experience are significant in explaining the financial performance of a retail location. Ton and Huckman (2008) study the impact of employee turnover on process conformance within retail stores and find that the negative effect of turnover is most pronounced in stores with low process conformance (less discipline in process execution and poorer adherence to quality standards). DeHoratius and Raman (2007) analyze the relationship between incentives provided to store managers and monthly sales and shrinkage across a chain of stores. Fisher et al. (2007) analyze the drivers of a retail store's performance and find that inventory availability, store associate availability (staffing level), and customer satisfaction are key variables in explaining month-to-month variations in sales. Ton (2008) finds that increasing labor at a store results in higher conformance quality and service quality, but only the former improves profitability. Our study differs from these in several respects. First, rather than studying the impact of the number of employees upon performance, we study two entirely different aspects of employee deployment: labor planning and execution of the plan. Second, we possess a wealth of demographic data that allows us to control for local population characteristics. This data allows us to distill cross-sectional differences among stores which, in our research setting, are much more pronounced than variation over time. Third, we introduce new metrics that capture the quality of the planning and execution processes.

Much academic interest in retail store operations has focused on improving the availability of products on the store shelves rather than on labor-related issues. Van Donselaar et al. (2010) find that store managers systematically make corrections on automated order advices either by shifting orders from peak days to non-peak days or by changing order size. Further, the literature on missing inventory and inventory record inaccuracy in retailing (see Raman et al. 2001) finds empirically that, because of execution failures, customers often do not find the products they seek, even if these products are within the store. This literature is closely related to our study because (1) it studies retail settings (2) it examines the issue of operational failures (both in planning and execution) and (3) it finds that labor is the key driver of phantom stockouts which is a type of operational failure. Raman et al. (2001) report that over 65% of the inventory records at a retailer they studied were inaccurate at the store-SKU level, and that over 16% of in-stock SKUs were reported as stockouts when customers asked for them. This study reports that such issues arise mainly due to store and distribution center replenishment processes, poor merchandising, poor inventory management, and high employee turnover. DeHoratius and Raman (2003) outline three approaches to the inaccurate inventory problem: preventing and eliminating root causes (using methods similar to the Ishikawa process of just-in-time, or JIT, principles), identifying and correcting errors through inspection policies, and, last, implementing software solutions that integrate the source of errors into the inventory management system. In a follow-up study, Ton and Raman (2010) find that greater product variety and higher inventories lead to a higher incidence of phantom stockouts (whereby inventory is in the back room but does not reach the shelf) and lost sales. Corsten and Gruen (2003) study the root causes of retail inventory stockouts and point to mechanisms that address the issue of stockouts and improve sales. Just like our paper, all of these studies emphasize the role of store labor in the execution process and they attempt to link labor deployment to store's performance.

Empirical studies of execution span other industries as well. In service industries, customer-employee interaction is at least as important as in retail. Frei and Harker (1999) quantify the inefficiencies in process execution due to process design in banking services using Data Envelopment Analysis. Frei et al. (1999) study the impact of aggregate process performance and process variation on financial outcome using a sample of 135 bank branches. They report that process variation negatively affects financial performance. Studies of execution in the healthcare industry have focused on operational failures in the execution process (Tucker 2004) as well as on learning through these failures (Tucker and Edmondson 2003). Ren and Wang (2006a) empirically link process consistency and service quality and Ren and Wang (2006b) further show how service quality affects volume at U.S. hospitals. On the other hand, manufacturing has long been the focus of studying execution in tasks that can be compared to some back-office functions in retail (e.g., shelf restocking, inventory replenishment) as argued in Fisher (2004). In this context the role of process design and conformance has long been debated, and the virtues of the Toyota Production System are well documented. Womack et al. (1991) show that Toyota's competitive advantage arises from a combination of employee motivation, training, process designs, and JIT techniques. Fisher and Ittner (1999) study the impact of product variety on automotive assembly plant operations and find that increased option content variability in car assembly has an adverse effect on plants' operational performance, which is manifested in higher total labor hours, overhead hours, downtime hours, rework, and inventory levels. MacDuffie et al. (1996) find that complexity of parts persistently impairs productivity. Again, in all these studies the role of labor is paramount, although both the planning and the execution processes and associated challenges differ from our retail setting.

3. Research Hypotheses

Our study is partially theory driven and partially exploratory: while there is quite a bit of theory regarding how to forecast demand for labor and how to schedule labor given this forecast (e.g., Bechtold and Jacobs 1990), there are very few rigorous studies on labor plan execution (e.g., Fisher et al. 2007) and, to our knowledge, there are currently no guidelines with respect to how to evaluate either the planning or the execution stages of labor management. The several testable hypotheses we state below originate primarily from our knowledge of the industry, discussions with companies' managers, and industry reports, but also from extant literature cited below.

The goal of the workforce/labor management process in retail organizations is to match demand for labor with labor supply. On the demand side, requirements for labor come from a wide variety of processes within the store which include a mixture of actions that can be roughly allocated into what we will call "back-office" and "front-office" tasks. The back-office tasks include inventory unloading and backroom logistics, restocking shelves, posting correct prices and signage, cleaning the store and managing the parking space/external structures, bagging purchases, administrative tasks, etc. The front-office tasks include customer interaction through managing check-out registers, servicing customers in a variety of store departments, and handling customer requests which may include returns, loyalty program questions, etc. through the customer service desk (see Fisher and Krishnan 2005). Of course, many of these tasks can cross the line between back-office and front-office: an employee restocking shelves can be asked a question by a customer and an employee who normally handles the customer service desk might be called to unload a truck in cases of labor shortage or inclement weather. Additionally, there are typically a number of administrative tasks which have to do with planning and managing the workforce, financial reporting, store audits and inventory counts plus interactions with the head office of the retail chain, and many employees will take part in these tasks from time to time.

When deploying store labor, the retail organization must first estimate the demand for labor and then match this demand with labor supply since labor shortages will impact execution of all tasks. To our knowledge, there are currently two main methodologies for estimating demand for labor. The first, and we believe the more common approach, is to forecast sales, and then deploy labor in proportion to the sales forecast. This is the approach described in Workplace Systems (2009) and this is the same approach that is used by the retailer in our study and in the study by Fisher et al. (2007). An alternative approach is to forecast traffic in the store and allocate labor accordingly. This approach is taken to schedule labor at Payless ShoeSource (Kronos 2006) and at Wal-Mart (Maher 2007), among others. After forecasting sales or traffic, retailers use internally developed standards which dictate how many labor hours need to be allocated per unit of sales or traffic, while leaving some provisions for store tasks that may not be associated with either sales or traffic. The key question is: which approach is better?

The advantages of the first approach (forecasting sales) are evident: sales clearly drive demand for many tasks in the store, including both back-office and front-office tasks. However, using a sales forecast to deploy labor suffers from the problem that staffing levels influence sales. As Fisher et al. (2007) demonstrate, staffing level drives sales (in other words, sales are determined in large part by the staffing level) and, in fact, sales is a concave increasing function of the staffing level. This finding is supported by industry reports (Retail Systems Research 2007, Figure 6) which cite sales improvements as the key driver of retail workforce management strategy and by Hise et al. (1983), who find that the number of employees is a variable that has one of the greatest impacts on sales. Although this suggestion seems intuitive, with the exception of Fisher at al. (2007), we are not aware of any studies that are able to estimate the sales lift from adding an hour of labor and, based on our interactions with numerous retailers, we do not believe that they go through such exercises. Thus, it becomes extremely challenging to forecast sales: if sales are driven by staffing levels, any past observations of sales will be dependent on past staffing levels and there are, to our knowledge, no existing tools that take this issue into consideration. To make an analogy with inventory management, past sales observations might be censored due to inventory unavailability. While there are sophisticated econometric tools to deal with the censoring issue in inventory management (e.g., Conlon and Mortimer 2009), the same tools do not yet exist for workforce management, where the problem is arguably more difficult because the relationship between sales and labor does not have a simple functional form (unlike the relationship between inventory and sales).

The advantages of the second approach (forecasting traffic) is that, as we learned from conversations with numerous retail companies, labor management within the store is unlikely to change traffic in the short term. Of course, in the long term, poor labor management may lead to lower customer satisfaction and, as a result, to lower traffic. Most likely, though, store traffic is largely determined by the store location, customer demographics around the store, and possibly by brand reputation of the store chain itself. Furthermore, demand for labor in many parts of the store is clearly driven by store traffic: more people in the store require more cashiers, more frequent cleaning, more customer-employee interactions, etc. Consistent with these observations, a recent industry benchmarking study (Retail Systems Research 2007, Figure 21) finds that the best-performing retail organizations are almost twice as likely to deem foot traffic to be valuable in scheduling workforce as the average retail company does. Furthermore, a new wave of companies are installing traffic counters at their stores (Kronos 2006), which seems to indicate that this practice (although expensive) is becoming more and more accepted and valued.

To finalize our first hypothesis, we discuss the dependent variable (performance metric). In this study, we select to focus on basket (average transaction) value, since it is likely to be directly affected by labor within the store because the retail format in our study can be characterized as a relatively "high-touch" environment in which employees play a key role in providing service. This service includes directly helping customers with a purchase, advising on the quality and price of a product, substitution, product location, restocking shelves that are often emptied during the day, ensuring that checkout lines are short etc. At the same time, as we described above, store traffic is largely unaffected by labor decisions within the store, so we cannot use it as an outcome variable. Furthermore, the retail format that we study is such that almost all customers who visit the store buy something (to be discussed and demonstrated shortly), so store labor is unlikely to affect conversion of traffic into buyers either. Finally, average transaction value is one of the most frequently used metrics to monitor store performance (see Retail Systems Research 2007, Figure 15). Therefore, basket value is the key performance metric of interest: we postulate that labor mismatches relative to traffic lead to lack of interactions with customers, less frequent shelf restocking, a lower chance of finding the product, etc., all of which decreases the basket value at the store. To summarize, our first hypothesis is:

Hypothesis 1. *Matching labor to store traffic rather than matching labor to a sales forecast is associated with higher basket values.*

Labor deployment in retail organizations is often the result of a relatively long and complicated process, and the process of matching labor supply with labor requirements is separated into two distinct parts: labor planning and labor execution. At the labor planning stage, plans are developed at headquarters for each store, often using operations research and scientific management techniques. Typically, such plans begin with forecasting sales or traffic at each store using past activity, local population characteristics, planned sales promotions, forecasts of future consumer spending, expected promotions and competitors' moves. Then data on task standards (ringing time, bagging time, stock room time, cleaning time, etc.) are incorporated into the schedule using a bottom-up approach. A large stream of operations research literature (see, for instance, Bechtold and Jacobs 1990) focuses on the quantitative aspects of labor planning. But while some of the input data (such as past sales and planned sales promotions) is quantitative, a lot of other information is qualitative in nature and is used in an ad-hoc manner through managerial judgment and rules of thumb (see Retail Systems Research 2009). The resulting labor plan, which is often detailed weekly by store department, is then used to create a labor budget that is subsequently uploaded into the planning system and is given to the store manager for execution. During this entire process, the store manager has limited input into the plan.

The store manager is given the labor budget and tasked with implementing it. Normally, the manager is given discretion regarding how many people to employ provided that the labor budget for the year is not exceeded. In trying to adhere to the labor budget and labor deployment recommendations from corporate, the store manager faces a number of obstacles: absenteeism, sick leaves, vacations, inflexibility of employees, turnover (both voluntary and involuntary) and lack of qualified candidates. Furthermore, external factors such as bad weather, delays in inventory delivery to the store, random disruptions in store operations, etc., may lead to deviations from the labor plan. Finally, store managers may intentionally deviate from the plan if they do not understand what goes into the plan and if they do not believe the plan is correct. For example, Retail Systems Research (2009, page 8) indicates that "… store managers did not believe that a corporate team understood what went on in stores well enough to define either labor standards or the basis for calculating labor budgets…" and that "…store managers spent very little time on the schedule prior to the implementation … but the impact on employees was terrible." Therefore, when implementing this plan, store managers often comply with only the most rigid constraints (e.g., an overall budget) but might essentially ignore other aspects of labor deployment, although adherence to the plan may vary dramatically depending on manager's preferences and convictions (Fisher and Krishnan 2005). To our knowledge, next to nothing is known about how well store managers follow these plans and what are the consequences of these actions.

Clearly, failure to deploy the right amount of labor may be due to both failures at the planning stage (e.g., due to reliance on sales forecasts rather than traffic forecasts) and at the execution stage (e.g., due to poor adherence to the plan by the store manager). Thus, we further subdivide Hypothesis 1 into two parts:

Hypothesis 2a. *Matching the labor plan to traffic is associated with higher basket values.*

Hypothesis 2b. *Matching labor deployment to the labor plan is associated with higher basket values.*

Going deeper into the labor deployment part of the process, there are typically three types of employees available to cover demand for labor who differ both in skills and in the tasks that they execute. Full-time employees work a normal work week (40 hours) and are entitled to a variety of benefits which may include paid vacation, medical insurance, in-store discounts, year-end bonuses, etc., and they receive significant training. Any time in excess of 40 hours is paid at higher overtime rates, but full-time employees typically have regularly scheduled hours (e.g., 8 AM to 5 PM with a lunch break) that are relatively constant throughout the year. These employees are often assigned to perform front-office tasks since they have certain tenure and skills in the organization. Moreover, certain positions within the store may require expertise, license and education which necessitate a full-time employee. Examples of such positions are pharmacists, in-store health clinic personnel, in-store banking/credit union, etc. On the other hand, part-time employees work irregular working weeks and they might be scheduled by the store manager in advance or called on the spot. They often have very high turnover rates, may hold more than one temporary job and receive limited training with almost no benefits. Thus, part-time employees usually perform back-office tasks but can also perform certain front-office tasks (e.g., part-time cashiers are quite common) if necessary. This is especially true during periods of high seasonal demand when there are simply not enough full-time employees. Finally, there are also managers and, depending on the store size, there could be one manager for each large department of the store as well as one store manager. Store managers typically perform a variety of administrative tasks but they may also become involved in customer interactions, especially if the problem is escalated to them or if the transaction requires a higher level of responsibility (e.g., in cases of complex product returns or when the customer needs to be compensated for an inconvenience).

As is evident from the above discussion, different types of employees perform different tasks within the store and therefore the store manager does not only have to deploy the right number of people, but also the right mix of people. Among the three labor types, we can clearly expect that full-time employees perform the most critical tasks that directly affect basket value. Given that up to 70% of buying decisions are made in the store (Workplace Systems 2009), full-time employees who most often interact with customers can drive substitution behavior by explaining features of products and pointing to the right product in the store. Of course, these employees are more expensive and less flexible than part-time employees, so store managers are unable to rely on them exclusively. We would further expect that part-time employees interact less with customers and hence do not affect basket values to the same extent. Even when a part-time employee replaces a full-time employee in a customer-facing function, we expect him/her to have less experience and therefore to be an imperfect substitute for a full-time employee. Thus, we expect part-time labor to have a smaller impact on basket values. Finally, store managers are also unlikely to influence basket value to nearly the same extent as full-time employees since they are not directly involved in buying decisions. These considerations lead us to further separate Hypothesis 2b by labor type as follows:

Hypothesis 3. *Matching deployment of full-time labor to a labor plan has a stronger association with basket value than matching deployment of either part-time labor or managers.*

4. Data Description and Preliminary Analysis

A large national retail chain provided data for this study under conditions of nondisclosure and anonymity. Hence, we are unable to disclose the retailer's name, location or retailing segment. We were provided data from stores in two closely located metropolitan areas. The retailer selected these two regions because they had been a part of the company for a long time, they experienced few store openings and closings, and management generally perceived them to be relatively advanced in their service quality and store execution, whereas some of the other regions were going through mergers and acquisitions and were thus less stable. Such store selection is likely to bias our results against finding any mismatches in labor deployment. For the two regions, we initially had observations for 311 stores over two years (we later added an additional preceding year of data to develop forecasts and verify the robustness of our results). However, a few stores went through closings and openings and did not have sufficient data to develop forecasts. Further, a number of stores had missing and inconsistent data, especially data on labor planning which was obtained from a separate database. As a result, we decided to drop a number of stores from the data, resulting in a dataset of 230 stores. The retailer divides each year into 13 fiscal months, each containing 4 calendar weeks (in the case of a 53-week year, the 13^{th} month can have 5 weeks). The top two panels in Table 1 summarize the data that we obtained. We use subscript *i* to denote each store, and we use subscript *t* ranging from 1 to 26 to denote each fiscal month.

$SALES_{it}$ is the revenue in U.S. dollars during the fiscal month. While sales averaged about \$1.48M per month per store, there was considerable cross-sectional variation (between \$296K and \$4.4M) because stores differed in size and location. Additionally, the retailer tracked $TRANSACTIONS_{it}$, which measures the number of customer transactions recorded at the checkout counters of the store, and the $BASKET\_VALUE_{it}$, which is the value of each transaction. On average, a store in a given month had 66,240 transactions and a basket value of \$22.30. For reasons that are evident from our hypotheses, we developed a simple procedure to forecast store transactions at a monthly level as follows. First the data were seasonally adjusted; then forecasts were generated for the seasonally adjusted data via linear exponential smoothing; and finally the seasonally adjusted forecasts were "re-seasonalized" to obtain forecasts for the original series. We used an additional year of data on transactions to generate the forecast but since many other variables were not available for this extra year, we used it only for constructing the forecast. This procedure was applied to obtain a forecast, denoted by

TXN_FCAST_{it}. Clearly, any conclusions drawn from this forecast are conservative since the retailer would possess additional data (such as changes in local competition, religious composition of the population etc.) to build a much more precise forecast.

$PLAN\_HOURS_{it}$ is the total number of employee hours budgeted for a store in a given month. The plan was created by the corporate office in several steps. First, the corporate office forecasted sales at the retail location ($PLAN\_SALES_{it}$) based on past sales and other information. The corporate office recommended the number of labor hours to the manager. For defined activities such as unloading a truck, labor hours were computed based on the level of activity (e.g. how many trucks would need to be unloaded next month at a given store) combined with "time and motion" industrial engineering studies to measure the labor time required to perform the activity. To this calculation was added an allowance for general customer interaction that was computed as a defined percentage of forecasted sales. The store manager could then suggest changes based on his/her local information to a limited extent. The manager could deviate from the prescribed staffing level as long as monthly deviations balanced out within the year.

EE_EARNS_{it} is the total wages (including benefits) paid to non-managerial employees in a given month. As a fraction of average $SALES_{it}$, EE_EARNS_{it} amounts to between 11% and 14%, which is quite typical in retailing. MGR_EARNS_{it} describes the total wages paid to managerial employees in a given month, and it varies between 1.6% and 3% of $SALES_{it}$.

Employees at this retailer are divided into full-time and part-time workers. Full-time employees receive bonuses and additional healthcare benefits, whereas part-time employees do not. We obtained data on full-time employee equivalents (i.e., the number of hours worked by full-time employees divided by 40) as designated by FT_EE_{it} (averaging about 20 per month per store) and part-time employee equivalents (i.e., number of hours worked by part-time employees divided by 20) as designated by PT_EE_{it} (averaging about 65 per month per store). An average store in an average month had 5.27 full-time managers (FT_MGR_{it}). To control product availability, store managers conducted internal audits to count the number of SKUs that were out of stock in a given month (i.e., not on the shelves when weekly shelf inventory audits were conducted). The number of out-of-stock items is captured by the variable $STORE\_AUDIT_{it}$, which varied between 48.75 and 363.75 in a given month.

We also obtained several store characteristics that did not vary over time. First, there is a dummy DIV_1_i, which indicates that the store is located in the metropolitan area 1 out of the two areas covered by our data. Store selling space ($SQFT_i$) ranged from a minimum of 11,096 ft^2 to a maximum of 98,056 ft^2 with an average of 47,054 ft^2, and the company varied product assortment ($SKUCOUNT_i$) depending on store size. The assortment was set by the corporate office and did not vary over time. $SKUCOUNT_i$ averaged 38,387 products and ranged from a minimum of 20,012 products to a maximum of 49,081.

We also obtained extensive demographic data from the retailer on the population within a two-mile radius of each store. This data utilized U.S. Census information compiled by the retailer just before the initial date of our study. $HOUSEHOLD_i$ denotes the number of households in the area around each store, which ranged from 7 to 61,389, with a mean of 12,685, indicating that some stores were located in densely populated urban areas while others were probably located in rural strip malls. Further descriptors available to us were the average household size ($AVG\_HH\_SIZE_i$), which averaged 2.49 persons, and the percentage of households with no children ($HH\_NOCHILD_i$), which averaged 68.1%. MED_AGE_i reflects the median age of the population, which was 37 years. We also obtained several descriptors of the racial makeup of the population as captured by the self-explanatory variables $WHITE_i$, $BLACK_i$, $AMER\_IND_i$, $ASIAN_i$, $HISPANIC_i$, and $OTHER_i$ to measure representation of each racial group in percentages. $MED\_HH\_INC_i$ and $PER\_CAP\_INC_i$ capture, correspondingly, median household income (\$50,863 on average) and per capita income (\$24,773 on average). Finally, we obtained five variables indicating education levels: $GRADE\_SCHOOL_i$, HS_i, HS_GRAD_i, $COLLEGE_i$ and $COLLEGE\_GRAD_i$ which capture, correspondingly, the percentage of the population with the highest education level of the grade school, some high school, high school graduates, some college, or college graduates.

Table 2 shows correlations among customer demographic variables, some of which exhibit strong correlations. For example, r(AVG_HH_SIZE, HH_NOCHILD) = -.91, which is expected; smaller households are also less likely to have children. To mitigate possible multicollinearity problems among the 17 demographic variables, we attempted to reduce the dimensionality using principal components (see Lattin et al. 2003 for an example of such an approach in a study that also uses demographic data) and step-wise regressions (see Maddala 2001) of variables in our model. The first approach yielded 14 significant principal components that explained only 75% of the variation in the demographic variables. Thus, we did not utilize the principal components in place of the original demographic variables. Step-wise regression results (not reported here) were consistent with the results of the full-fledged estimation that included all variables. In the end, we computed VIF (variance inflation factors) and excluded variables that had VIF>10 including HH_NOCHILD, WHITE, OTHER, MED_HH_INCOME, GRADE_SCHOOL, COLLEGE_GRAD so we report most results with the remaining 11 demographic variables.

We further discovered that FT_EE, FT_MGR, and PT_EE were essentially linear transformations of EE_EARNS and MGR_EARNS, because pay rates for employees and managers were almost identical among stores, so they could not be used simultaneously. Since labor supply is better reflected in the number of employees rather than in their salaries, we focus on the first three variables in the rest of the paper. Table 3 shows the correlations among longitudinal variables. Note that TRANSACTIONS and BASKET_VALUE are essentially uncorrelated (r = .02), indicating that the number of customers going through the store is unrelated to the amount of merchandise that each buys. Furthermore, EE_EARNS are highly correlated with SALES (r = .92), which is quite natural because staffing was done using sales forecasts. Finally, it appears that both PLAN_SALES and TXN_FCAST provide quite accurate forecasts of SALES and TXN, correspondingly (r=.99 in both cases).

5. Econometric Specification

Since we had in our possession several longitudinal variables, we could potentially conduct a classical panel data analysis (Kennedy 2003) using the fixed effects model. However, fixed effects models exclude variables that do not vary over time, and hence are not appropriate in situations in which such variables explain a significant proportion of variation in the data. To check the feasibility of the panel analysis approach, we first ran a simple fixed-effects panel model with dummies for each month and store fixed effects while using monthly sales for each store as the dependent variable. Results (omitted) indicated that 95% of variation in SALES is cross-sectional (explained by store fixed effects) while only 5% of variation is longitudinal. In other words, variations in sales over time are minimal relative to variations in sales from store to store. To verify this observation, we analyzed more closely the cross-sectional and time-series components of the BASKET_VALUE and TRANSACTIONS. Using a simple forecasting time-series model with linear trend and multiplicative seasonality factors, we estimated Forecasted BASKET_VALUE and then calculated predictability for the BASKET_VALUE using the following formula:

$$1-\frac{\sum_{t=1}^{13}\left|\text{BASKET\_VALUE}_{\text{it}}\ \text{- Forecasted BASKET\_VALUE}_{\text{it}}\right|}{\sum_{t=1}^{13}\left|\text{BASKET\_VALUE}_{\text{it}}\right|}.$$

We found that the average predictability for a store in our data set is 95.3%, indicating that within-store variations in BASKET_VALUE were easily predictable. Furthermore, we found that the average BASKET_VALUE explained 94% of the variation in the data set. In other words, the “between-store variation” in BASKET_VALUE overshadowed the “within-store variation”. We conducted similar analysis for TRANSACTIONS and found that seasonality and time trends explained an even higher proportion of within-store variation in transactions. This finding is consistent with the industry segment in which this retailer is competing: the products sold are well-established, satisfying basic needs and not subject to fashion or other highly uncertain influences over time. Furthermore, since we only had a handful of independent longitudinal variables, they explained only a very small proportion of longitudinal variation in sales (R^2 in single digits). We therefore concluded that, although typically there are advantages to conducting a longitudinal study over a cross-sectional study, our data is such that there is little economic significance to be gained from the classical panel data analysis.

To test our hypotheses it is necessary to operationalize the alignment (or the mismatch) between store traffic and planned payroll as well as between planned and actual labor deployment. There are two components to the alignment in question: there is an alignment on average and there is an alignment of deviations (i.e., how well increases/decreases in one variable are aligned with increases/decreases of the other variable). It is difficult to measure the alignment in averages in our setting since we do not know, for example, how many employees are necessary to cover 1,000,000 customers going though the store. To address this issue we controlled for the absolute staffing level, as will be explained shortly, but our main focus is on monthly deviations: since store managers in our data adhered to the annual labor budget quite closely, their discretion was mainly over how to allocate the labor budget from month to month.

It is challenging to operationalize the alignment between deviations: while we have data on planned and actual labor deployment, we do not have direct data on store traffic. In general, obtaining such data would require the installation of traffic counters, but this particular retailer did not have such equipment available. However, the retail format for this company is such that few customers visited a store and left without buying anything. This was independently confirmed by the company's management who indicated to us that the average basket contained 15 items, making it highly unlikely that a customer would visit a store for a single product and leave empty-handed. Furthermore, the average number of SKUs which are out of stock (the variable STORE_AUDIT) is 139 in our data, while the average number of SKUs is 38,387 (variable SKUCOUNT, see Table 1). Together, this leads to the average 0.3% out-of-stock chance (maximum 0.7% out-of-stock)[2]. Thus, even if all customers were to come just for one item, our proxy for store traffic would be 99.7% accurate. Moreover, for an average customer the odds that their entire basket of 15 items is stocked out is $(0.3)^{15}=1.4\times10^{-8}$. Finally, the number of transactions might be an even better measure of traffic than what traffic counters would compute because often an entire family would come to shop, thus inflating the store traffic figure but not necessarily affecting the number of customers that needed the attention of employees. Thus, we use TRANSACTIONS as a proxy for store traffic.

---

[2] In this sense, our transaction data is a much better proxy for traffic than, for example, data in Fisher et al. (2007) because in their setting many customers might come for a single item, and also because their stockout rate is much higher.

We are not aware of any studies that offer guidance in defining the mismatches we seek to quantify and therefore we propose to calculate deviation mismatches as proportional to correlation between two time series of corresponding variables. For this purpose, we define 9 different mismatches, e.g.,

$$TXNvsPLAN\_HOURS_i=1-r(TXN, PLAN\_HOURS)$$

denotes the mismatch between 23 monthly observations of the transactions and the labor plan. Other mismatch variables are listed in the third panel of Table 1. We note that TOTAL_EE denotes the total number of full-time employee equivalents deployed at the store; names of other variables are self-explanatory. As a robustness check, we attempted several different ways to calculate the same mismatches, which are reported in Section 7.

The advantage of these definitions is that they use temporal deviations but convert them into a single number to enable cross-sectional analysis. Thus, while our subsequent analysis is cross-sectional, we use longitudinal data to create mismatch variables. Another advantage of our definitions is that we accommodate variables measured in different units (e.g., hours vs. number of employees). Further, mismatches that we define do not depend on whether the store is, on average, understaffed or overstaffed (for which we control separately): since we calculate correlations between two time series, average staffing levels do not play a role. We believe that this is an advantage. For example, if the plan for one store systematically adjusts the staffing level up/down for a good reason (e.g., local competition), an absolute mismatch measure would say that the store always has a mismatch, which may not be reflective of suboptimal actions. On the other hand, it is hard to envision a justification for lowering the staffing level when traffic increases, or vice versa, so we think our mismatch variables accurately measure the quality of staffing plans.

We note that mismatch variables similar in spirit are frequently used in organizational behavior studies under the name "difference scores" (see, e.g., Tisak and Smith 1994, Edwards 2001). Among the most frequently used difference scores for two variables are: the algebraic difference, the absolute difference, the squared difference, the sum of absolute differences, the sum of squared differences, the square root of squared differences, and the correlation. While the first three difference scores are simpler, they are often criticized for not measuring the dispersion and shape of the dispersion, as they capture only the level differences (Tisak and Smith 1994, Edwards 2001) and furthermore, they confuse the mismatches on average and deviation mismatches. Other more complex scores are based on algebraic differences which presupposes that the two variables are measured in the same units, which is not true in our case (e.g., traffic and labor plan are in different units). Finally, measuring the mismatch in a way other than correlation makes it hard to interpret the results because such a mismatch does not have a natural scale, whereas our proposed mismatch is defined in the interval of [0, 2] enabling easy comparisons across stores. We note also that, from a more general point of view, our mismatch variables measure fit between two constructs similar to Venkatraman (1989). However, existing measures of fit focus on static constructs, while our metric of fit is dynamic (longitudinal), and developing and testing dynamic metrics of fit "…is recognized as a promising area of research …," Venkatraman (1989).

One issue with our proposed operationalization of mismatches is that they are equally affected by positive and negative deviations (e.g., by overstaffing and understaffing). While understaffing should clearly be detrimental to basket value, overstaffing is less likely to have the same effect. We note, however, that in our data store managers quite closely adhere to the annual budget, so overstaffing in any given month leads to understaffing in another month. Thus, in this sense, positive deviations are as detrimental as negative deviations. Finally, to control for the average labor level, we divide the total number of employee equivalents by the number of transactions (in 1,000,000s), since our hypotheses focus on the mismatch between the number of employees and the number of transactions and we obtain variable TOTAL_EE_PER_TXN.

Table 4 shows correlations among resulting cross-sectional variables. Naturally, several of the mismatch variables are strongly correlated, so we make sure that they are never included in the same regression. The single remaining concern here is a very strong correlation between SQFT and SKUCOUNT, indicating that product variety in the store is tailored to store size (r = .79), an observation independently confirmed by company management. For the reminder of the paper we scale product variety by the square footage of the store (SKUCOUNT_PER_SQFT) and we use it along with SQFT in the regressions.

We follow extant literature and utilize all other variables available to us as controls. For example, broader assortments of products are known to increase sales (Kok and Fisher 2007), and there is vast literature on the impact of product variety on consumers' basket values (e.g., Kahn 1995 and citations therein). Hence, we control for product variety. Furthermore, Ackerman and Tellis (2001) provide evidence of socioeconomic and cultural factors affecting the choice of products and the shopping behavior of customers, justifying our control for customer demographics. Following previous studies of retail execution (Fisher et al. 2007), inventory availability is another relevant variable to consider. However, inventory availability itself is probably driven by labor policies so it might be co-determined with basket value. Unfortunately, we have no instruments to account for the effect of STORE_AUDIT using, for example, IV estimation. We attempted to include STORE_AUDIT as simply an independent variable and it was never significant, probably because the number of stockouts was very small in our sample. Thus, we exclude it from the main regression but utilize it later to verify our results. We begin with Model 1, which simply relates the total mismatch between transactions and total labor to test our first hypothesis:

$$BASKET\_VALUE_i = \alpha_0 + \alpha_1 TXNvsTOTAL\_EE_i + \varepsilon_i. \qquad \text{Model 1}$$

Next, we separate the total mismatch into the planning mismatch and the execution mismatch to test Hypotheses 2a and 2b:

$$\begin{aligned} BASKET\_VALUE_i = {} & \alpha_0 + \alpha_1 TXNvsPLAN\_HOURS_i \\ & + \alpha_2 TOTAL\_EEvsPLAN\_HOURS_i + \varepsilon_i. \end{aligned} \qquad \text{Model 2}$$

Next, we separate the execution mismatch into mismatches for full-time labor, part-time labor and managers to test Hypothesis 3:

$$
\begin{aligned}
BASKET\_VALUE_i = \ & \alpha_0 + \alpha_1 TXNvsPLAN\_HOURS_i \\
& + \alpha_2 FT\_EEvsPLAN\_HOURS_i \\
& + \alpha_3 FT\_MGRvsPLAN\_HOURS_i \\
& + \alpha_4 PT\_EEvsPLAN\_HOURS_i + \varepsilon_i .
\end{aligned} \qquad \text{Model 3}
$$

We further include all available controls in our next specification to ensure that our results are not explained away by the control variables:

$$
\begin{aligned}
BASKET\_VALUE_i = \ & \alpha_0 + \alpha_1 TXNvsPLAN\_HOURS_i \\
& + \alpha_2 FT\_EEvsPLAN\_HOURS_i \\
& + \alpha_3 FT\_MGRvsPLAN\_HOURS_i \\
& + \alpha_4 PT\_EEvPLAN\_HOURS_i \\
& + \alpha_5 TOTAL\_EE\_PER\_TXN_i \\
& + \alpha_6 SKUCOUNT\_PER\_SQFT_i \\
& + \alpha_7 SQFT_i + \alpha_8 DIV\_1_i \\
& + \alpha_9 HOUSEHOLD_i + \alpha_{10} AVG\_HH\_SIZE_i \\
& + \alpha_{11} MED\_AGE_i + \alpha_{12} BLACK_i \\
& + \alpha_{13} AMER\_IND_i + \alpha_{14} ASIAN_i \\
& + \alpha_{15} HISPANIC_i + \alpha_{16} PER\_CAP\_INC_i \\
& + \alpha_{17} HS_i + \alpha_{18} HS\_GRAD_i \\
& + \alpha_{19} COLLEGE_i + \varepsilon_i .
\end{aligned} \qquad \text{Model 4}
$$

Finally, we also estimate Model 5, which excludes mismatch variables but includes all control variables to ensure that our main Model 4 is robust (e.g., coefficients do not jump around in sign, size or significance). We note that in all these specifications we retain the names of variables, although it should be understood that longitudinal variables denote the means of corresponding variables.

6. Results

We summarize the results of OLS estimation for the linear model in Table 5. Starting with Model 1, we see that the mismatch between store traffic and the total number of employees is significant at the 5% level and in the right direction (α=-1.585), indicating that larger mismatches are associated with lower basket values, rendering support for Hypothesis 1. Moving on to Model 2, after separating the total mismatch into the planning mismatch and the execution mismatch, we observe that the planning mismatch is significant at the 1% level (α=-3.097) but the execution mismatch is not significant, thus rendering support for Hypothesis 2a but not Hypothesis 2b. We further proceed to separate the execution mismatch by labor type in Model 3 and we observe that the planning mismatch is still highly significant (α=-3.132) and the execution mismatch for full-time labor is significant as well at the 10% level (α=-1.207), while execution mismatches for part-time labor and for managers are insignificant. Thus, Hypothesis 3 is supported. The four mismatch variables explain about 6.5% of variation (adjusted R^2) in basket value.

We continue with Model 4 where we include all the control variables and we observe that the planning mismatch and the execution mismatch for full-time labor are still significant but the sizes of the coefficients decrease (α=-1.624 and α=-.769, correspondingly). Among the control variables we see that the absolute level of labor (TOTAL_EE_PER_TXN) is highly significant (α=.014), indicating that adding labor per transaction is beneficial (in terms of basket value), and this benefit is above and beyond the separate benefit of matching deviations in labor to the deviations in traffic. Not surprisingly, increasing product variety (SKUCOUNT_PER_SQFT) appears to be beneficial (α=6.633) and larger stores (SQFT) have larger basket values (α=.000) even after controlling for the number of SKUs in the store. Demographically, stores located in the metropolitan area 1 have larger basket value (α=1.403) than stores in the area 2, while a larger number of households (HOUSEHOLD) is associated with lower basket values (α=-.000). This second result might reflect that customers in densely populated areas tend to buy less on a single trip due to less storage space in homes, a larger percentage of foot traffic, a propensity to eat out more, or other factors. Continuing with the demographic variables, higher median age (α=-.087), higher proportion of American Indians (α=6.884), lower proportion of Asian population (α=-15.911) and Hispanics (α=-4.662), higher per capita income (α=.000) and lower proportion of college-education population (α=-14.212) are all associated with larger basket values. The result related to per capita income is expected. Further, the Hispanic population is known to shop three times more often relative to other ethnic groups in the U.S. and therefore Hispanics may buy less per visit (see Food Market Institute study, 1998). We do not have very good theories regarding other demographic variables. We note that Model 4 explains about 67% of variations in basket value and provides a better fit than the three previous models.

Although our results indicate that execution mismatches for full-time labor are significant and execution mismatches for part-time labor are not, it is not immediately obvious from our analysis why this happens. In Section 3 we advanced two possible reasons for this observation: part-time labor performs back-office tasks which may not be directly related to store traffic and even when part-time labor does perform front-office tasks, it provides a lower quality of service. While we cannot test these explanations directly, there are two indirect tests that we can perform. One simple test is to use the number of out-of-stocks as a dependent variable: since shelf replenishment is largely performed by part-time labor, we expect it to be more impactful upon out-of-stocks than full-time labor. We run analysis which is identical to Model 4 but with STORE_AUDIT as a dependent variable and we indeed find that the execution mismatch for part-time labor is highly significant (α=-14.54, p=.015) while other mismatches are not significant. Another test can be used to verify that part-time labor is indeed used to substitute for full-time labor. In this test,, we use a simple panel analysis with store fixed effects and dummies for each fiscal month. We use FT_EE as a dependent variable and PT_EE as an independent variable and we find that the coefficient for part-time labor is negative (α= -.016) and highly significant (p=.000). Another indirect confirmation of this theory is that execution mismatches for full-time labor and part-time labor are larger on average than execution mismatches for total labor (Table 1). Thus it indeed appears that part-time labor is used as a substitute for full-time labor: increases in part-time labor are associated with decreases in full-time labor and vice versa. We conclude that both of these tests are consistent with our theoretical predictions and Hypothesis 3.

To ensure that the results of our regression are stable and consistent, we verified them in Model 5 using only control variables. We observed that sizes, signs and significances of all coefficients are preserved. We further conducted a number of diagnostic tests on our main Model 4. First, we checked for multi-collinearity by calculating variance inflation factors. The average VIF is around 3 and the highest VIF is below 8, which is considered acceptable (Kennedy 2003). Second, we assessed the overall model fit visually by plotting predicted BASKET_VALUE against actual observations and did not encounter any outliers or other irregularities. We also used the Ramsey RESET test to check for omitted variables and it supported our model specification (p=0.18). Next, we checked for heteroskedasticity visually (the residuals plot did not indicate any issues) and then using the Breusch-Pagan test and again did not find any support for it (p=0.23). Nevertheless, we re-ran all models using robust standard errors and our results remained unchanged. We further tested residuals of the regression both visually and using the Kolmogorov-Smirnov test which could not reject Normality. To check for outliers, we computed standardized residuals and the smallest residual was -3.14 and the largest residual was 2.79, which is within the recommended bounds of [-3.5, 3.5] (Kennedy 2003). Finally, we analyzed the leverage plot and did not find any observations which had both high leverage and large normalized squared residuals. We therefore concluded that our model passes all the classical diagnostic tests.

7. Robustness

We attempted a number of robustness checks in order to ensure that our metrics of mismatches are indeed valuable and that our hypotheses tests are convincing. First and foremost, we attempted several alternative ways of calculating mismatches. We tried calculating mismatches for each year of data separately rather than for two years of data simultaneously. This results in twice as many observations (two for each store) and even if we cluster standard errors by store, we obtain stronger results. We also added together two years of data and calculated correlations among the resulting 13 observations (this is similar to calculating seasonal coefficients). This approach again yielded stronger results for the planning and full-time labor execution mismatches (in terms of both significance and size of coefficients). We also tried to de-trend all the data (both within each year and for a two-year trend), then to calculate seasonal factors for each time series and finally to calculate mismatches among seasonal factors. The results of this analysis were again qualitatively similar to those reported in the paper, both in size of coefficients and in significance. We omit these results for space considerations.

Next, we tried alternative model specifications. We attempted a semi-log specification in which all dependent and all or some independent variables (except for variables that are in relative terms) are logged. Logarithmic specification is often advantageous when there is significant skewness in variables (see Maddala 2001), as is the case with some of our variables. However, both the goodness of fit and the regression results with the semi-log models remained qualitatively similar, so we do not report them here. We also attempted to add non-linear terms to some variables. For instance, Fisher et al. (2007) find the concave increasing relationship between labor level and sales, so we added to the regression a squared variable TOTAL_EE_PER_TXN. Although its sign was negative and significant (which is consistent with concavity), we ultimately decided to remove this variable because it caused collinearity (high VIF). We also attempted to include a squared term for the SKUCOUNT variable, but it was insignificant and other results did not change.

We further attempted to include all demographic variables into the regression. As mentioned earlier, this resulted in multi-collinearity with extremely high individual VIFs (in the millions). Nevertheless, this did not significantly affect the coefficients of the mismatched variables although some coefficients of control variables were affected. We also attempted to include interaction terms to check whether the combination of planning and execution mismatches has an even greater impact upon basket value, but there were no statistically significant results. Finally, we tried re-running Models 1 and 2 with all the control variables and obtained similar results.

An argument can be made that all mismatches should be calculated against transaction forecasts rather than against transactions themselves. We therefore used the forecast of transactions in Models 6 and 7 (Table 6). In Model 6, we replace the planning mismatch with the mismatch between transaction forecast (TXN_FCAST) and labor plan and we observe that the coefficient decreases in size (α=-1.624 in Model 4 vs. α=-1.262 in Model 6) and significance (p=.01 in Model 4 vs. p=.027 in Model 6), both of which are expected. We also consider another possibility: one may argue that in addition to planning and execution mismatches, there is also a forecasting mismatch that could measure deviations of the forecast from actual transactions. We introduce this mismatch (TXN_FCASTvsTXN) in Model 7 in addition to all other mismatches and control variables, and we find that it is insignificant while other results are largely unaffected. This is hardly surprising given that the forecast of transactions is extremely accurate, as noted earlier. Further evidence of this explanation is that the forecasting mismatch is the smallest (0.17 on average, Table 1) of all mismatches.

While our models so far demonstrate that deploying labor based on transactions is associated with larger basket values relative to the retailer's current methodology of deploying labor based on sales, an argument can be made that, perhaps this retailer's labor plan can simply be better aligned with sales (rather than with transactions). We consider such a possibility in Models 8 and 9, where we include the planning mismatch between sales and labor (PLAN_SALESvsPLAN_HOURS) individually and along with the execution mismatches[3]. In either case, the mismatch against sales is insignificant while other results are largely unaffected. We thus do not find any evidence that planning labor better against sales is associated with larger basket values. Again, this is perhaps unsurprising given that the correlation between sales forecast and labor plan is very high (Table 3) and further, that the above mismatch is quite small (0.27 on average, Table 1) relative to other mismatches. So it appears that although the retailer is quite successful in both forecasting sales and planning labor against them, this process can nevertheless be improved.

8. Implications and Discussion

8.1. Managerial and theoretical implications

Our results suggest that the old adage of having "the right product at the right time and place" could be true of employees as well. While previous literature has found it important to hire the proper

[3] We note here that there is a conceptual problem with measuring mismatches relative to sales. Since sales are a function of basket value, we have basket value appearing on the left and the right sides of the regression. Thus, any relationship we might find between the mismatch against sales and basket value may be due to the mathematical relationship between the two variables. However, since we do not find any relationship, this issue is moot.

number of employees (Fisher et al. 2007, Hise et al. 1983) and to train them better (Fisher et al. 2007, Ton 2008), our current work makes the point that staffing levels should be aligned with store traffic. When store traffic increases, the number of employee-customer contact points goes up and therefore more employees are needed to attend to both front-office and back-office functions. Currently, the retailer we studied uses sales rather than traffic to determine staffing levels in its stores. We argue that this may not be the best way to plan labor. Staffing policies affect sales (they may also affect traffic over a long time horizon, but it is unlikely that poor store execution would cause observable changes in store traffic within a week or even a month) and this endogeneity makes it difficult to use sales to determine correct staffing levels, while traffic in our setting does not suffer from the same problem. There also appears to be no correlation between the ability to plan and the ability to execute better. The reason might be that labor planning is largely driven by the corporate office and execution practices are largely driven by the store manager's experience and biases (see Fisher and Krishnan 2005). Our results suggest that closer interaction between the store manager and corporate might be beneficial to overcome these problems.

If we believe that the mismatches affect basket value, we can estimate the potential effect of improving store labor planning and execution. It would be unlikely that a retail chain could completely eliminate planning mismatch, but even through reducing the current mismatch levels by half, we estimate the sales lift to be (based on Model 6 which uses our crude forecast of transactions and all control variables) 1.4 % of the current chain level of sales, which is a very large number for the retail industry in general and for this retail chain in particular (over the time period surrounding our study, the chain reported approximately 3% comparable store sales[4] increase per year). From a practical point of view, it appears that improving labor planning requires only switching from forecasting sales to forecasting traffic, which should be relatively straightforward to do.

Likewise, by reducing the current mismatch levels by half, we estimate the sales lift to be 1.7 % of the current chain-wide sales. However, it is much harder to advise on specific policies that are likely to

[4] Comparable store sales growth is a measurement of productivity in revenue used to compare the sales of retail stores that have been open for a year or more.

improve store labor execution because there are practical issues associated with managing the seasonal hiring of full-time employees, the most compelling one being compliance with labor union contracts concerning tenure of employment for full-time employees (because much of this retailer's store labor is unionized). Many retailers operate under significant pressure from labor practices groups, which puts pressure on store managers to be attentive toward labor utilization levels within the store. The recent backlash against Wal-Mart's attempt to make labor "more flexible" illustrates this point (Maher 2007). However, given the estimation of this sales lift, it might be advisable to offer full-time employees some incentives to work more flexible hours, for example. Our results also caution against the practice of substituting part-time labor for full-time labor.

8.2. Limitations

Given our data, we cannot claim a causal mechanism whereby reducing the mismatch in payroll planning and execution will definitely result in higher basket values: we can only demonstrate the negative association between the mismatch and basket value. One may argue that, perhaps, the company allocated the best store managers to better performing stores but company executives did not support such a suggestion. It is quite difficult to imagine some other reverse causal effect (that is, that larger basket value drives mismatches), whereas it is quite plausible that mismatches can affect basket value. We also possess a reasonable number of control variables that are likely to capture most of the relevant sources of variation, although, as in any other study, there is always a possibility of the omitted variable bias. For instance, previous research has shown that manager's experience, employee turnover, employee training, and incentives are significant drivers of retail performance. While we do not possess such data, we believe that employee training and incentives are unlikely to affect our results since these were uniform across the entire chain. But certainly manager experience and employee turnover might be able to explain execution mismatches. In our data set there was only miniscule turnover among store managers (about 0.7% per year) so for the vast majority of stores the manager was the same for the duration of our study. Unfortunately, the company was unable to provide detailed data on store managers and therefore we were unable to analyze in depth which traits of managers make them better/worse executers. Since the goal of our study was to relate mismatches to store performance, we do not see this as a serious limitation. However, explaining mismatches remains a fruitful direction for future research.

As in many other empirical studies, our modeling choices and variable definitions are sometimes driven by the data availability and hence some of our results might have alternative explanations that we are unable to rule out. For example, one can imagine that some of the mismatches could be "good" because the store manager uses the latest information to increase/decrease staffing levels relative to the plan. An alternative way to interpret our results is to say that the central plan in our case is better than any local plan so centralization of decisions is superior to decentralized decision-making for our retailer. We attempted to obtain vital data for store managers and employees, including their personnel records that would indicate the reason for employee termination/temporary layoff to differentiate between voluntary and involuntary mismatches. However, all this data was in hard-copy format and coding it did not appear feasible. Moreover, there were privacy concerns. Further, one can argue that positive deviations from the plan (i.e., exceeding the plan) should not have the same impact on performance as negative deviations. We acknowledge this possibility: it is possible that negative deviations from the plan affect basket value much more than positive deviations. For example, if one believes that the relationship between basket value and store labor is concave increasing (which is a reasonable assumption), then positive deviations from the payroll are inherently less impactful than negative deviations. Unfortunately, testing this hypothesis would require panel data analysis which, as we indicated earlier, is not feasible in our case; unlike Fisher et al. (2007), for example, we do not possess a sufficient number of longitudinal variables to achieve any reasonable explanatory power and to control for a variety of longitudinal effects. We did attempt to look at positive vs. negative deviations between seasonal coefficients for each time series. Unfortunately, the sum of positive deviations is equal to the sum of negative deviations, so we did not obtain necessary variation.

Finally, our measures of execution mismatches might be indicative of the general "quality" of the store manager: deviations from the plan, for example, might be symptoms of larger problems (incompetence, lack of training or experience, etc.) which affect other aspects of the retail store operations not directly related to store labor. We acknowledge that our data does not allow us to distill these issues from one another – perhaps this would become possible if the company conducted customer surveys which might indicate specific aspects of service shortfall within the store. One specific test that we did perform is regressing STORE.AUDIT against mismatch variables. The only significant (and negative) mismatch was the part-time labor execution mismatch which makes sense to us: as we described earlier, part-time labor often takes care of the back-office functions in the store such as stocking store shelves. Although we have no evidence to support this theory, we nevertheless acknowledge that our mismatch variables might be partially capturing the overall quality of the store manager.

8.3. Summary

To summarize, our results once again suggest the paramount importance of store labor in the financial success of a retailer. Fisher et al. (2007) demonstrate that it is important to staff stores adequately with knowledgeable employees. In the current paper we go further by demonstrating that performance can be further improved by matching staffing levels with store traffic, which is something that a new wave of software products attempts to do (Maher 2007). Our analysis suggests that a sales lift of 1.4% can be achieved by modest improvements in store labor planning. However, we believe that our analysis is quite conservative and is likely to underestimate potential improvements: while we only possess monthly data, current best practices suggest that traffic should be matched with labor in hourly or even shorter time intervals. This makes great sense to us: depending on store location and product profile, different stores will experience increased traffic at different points of time throughout the day. The management of the company we have been working with also commented that, ideally, scheduling should be separated by department within the store as well. We believe that, if such data becomes available, it is likely to show that even higher sales lifts can be realized. Additionally, we demonstrate the importance of the execution of the planned full-time labor deployment schedule. Our results suggest that a 1.7% sales lift can be realized by modestly improving execution. It remains to be studied what specific actions the store manager can take with respect to this observation and what these actions' true impact will be. Given that, according to the management of the company, stores in our study were well-run, it is likely that potential improvements for the rest of the chain would be even higher. Our other results show that customer demographics play an important role in determining the financial success of a store. These results raise the interesting possibility of tailoring product assortments to local customer demographics and store location; see Fisher and Vaidyanathan (2008) for a related approach. Our results on customer demographics could also be used to guide the retailer in choosing store locations.

References

Ackerman, D. and G. Tellis. 2001. Can culture affect prices? A cross-cultural study of shopping and retail prices. *Journal of Retailing*, 77, 57-82.

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

Conlon, C.T. and J.H. Mortimer. 2009. Demand estimation under incomplete product availability. Working paper, Harvard University.

Corsten, D. and T.W. Gruen. 2003. Desperately seeking shelf availability: an examination of the extent, the causes and the efforts to address retail out-of-stocks. *International Journal of Retail and Distribution Management*, 31(12), 605-716.

DeHoratius, N. and A. Raman. 2003. Building on foundations of sand? *ECR Journal,* 3(1), 62-63.

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

Edwards, J.R. 2001. Ten difference score myths. *Organizational Research Methods*, 4(3), 265-287.

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

Fisher M. L. and C. Ittner. 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science*, 45(6), 771-786.

Fisher, M. L. and J. Krishnan. 2005. Gotta Hava Wawa. Case study, The Wharton School, University of Pennsylvania.

Fisher, M.L., J. Krishnan, and S. Netessine. 2007. Retail store execution: An empirical study. Working paper, University of Pennsylvania.

Fisher, M.L. and A. Raman. 2010. *The New Science of Retailing: How Analytics are Transforming the Supply Chain and Improving Performance*. Harvard Business Press, Cambridge, Massachusetts.

Fisher, M.L. and R. Vaidyananthan. 2008. An algorithm and demand estimation procedure for retail assortment optimization. Working paper, University of Pennsylvania.

Food Market Institute. 1998. *Profile of the U.S. Hispanic Shopper*. Research Report.

Frei, F. X. and P. T. Harker. 1999. Measuring the efficiency of service delivery processes: An application to retail banking. *Journal of Service Research,* 1(4), 300-312.

Frei, F. X., R. Kalakota, A. J. Leone, and L. M. Marx. 1999. Process variation as determinant of bank performance: Evidence from the retail banking study. *Management Science,* 45(9), 1210-1220.

Hise, R. T., J. P. Kelly, M. Gable, and J. B. McDonald. 1983. Factors affecting the performance of individual chain store units: An empirical analysis. *Journal of Retailing,* 59(2), 22-39.

IHL Consulting. 2005. The future of workforce management for retailers. Available at http://www.ihlservices.com .

Kahn, B.E. 1995. Consumer variety-seeking among goods and services: An integrative review. *Journal of Retailing and Consumer Services,* 2(3), 139-148.

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

Kök, A.G. and M. L.Fisher. 2007. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55(6), 1001–1021.

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

Lattin, J., J. D. Carroll, and P. E.Green. 2003. *Analyzing multivariate data*. Duxbury Press, Belmont, California.

Lusch, R. F. and R. R. Serpkenci. 1990. Personal differences, job tension, job outcomes, and store performance: A study of retail store managers. *Journal of Marketing,* 54, 85-101.

MacDuffie, J. P., K. Sethuraman, and M. L. Fisher. 1996. Product variety and manufacturing performance: Evidence from the international automotive assembly plant study. *Management Science*, 42(3), 350-369.

Maddala, G. S., 2001. *Introduction to econometrics*. 3rd ed., John Wiley & Sons.

Maher, K. 2007. Wal-Mart seeks flexibility in worker shifts. *The Wall Street Journal*, 3 Jan. 2007, 1.

Raman, A., N. DeHoratius, and Z. Ton. 2001. The Achilles heel of supply chain management. *Harvard Business Review*, May, 136-152.

Ren, J. and X. Wang. 2006a. Process consistency and service quality in U.S. hospitals: An empirical investigation. Working paper, Boston University.

Ren, J. and X. Wang. 2006b. Volume and variability: An empirical investigation of service quality in U.S. hospitals. Working paper, Boston University.

Retail Systems Research. 2007. The state of retail workforce management: Benchmarking study. November. Available at http://www.kronos.com.

Retail Systems Research. 2009. The human side of workforce management: Retail best practices. April. Available at http://www.kronos.com.

Salmon, W. J. 1989. Retailing in the age of execution. *Journal of Retailing*, 65(3), 368-378.

Tisak, J. and C.S. Smith. 1994. Defending and extending difference score methods. *Journal of Management*, 20(3), 675-682.

Ton, Z. 2008. The effect of labor on profitability: The role of quality. Working Paper, Harvard Business School.

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

Ton, Z. and A. Raman. 2010. The effect of product variety and inventory levels on retail store operations: A longitudinal study. Forthcoming, *Production and Operations Management*.

Tucker, A.L. 2004. The impact of operational failures on hospital nurses and their patients. *Journal of Operations Management*, 22(2), 151-169.

Tucker, A.L., and A. Edmondson. 2003. Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *California Management Review*, 45(2), 55-72.

van Donselaar, K. H., V. Gaur, T. van Woensel, R. A. C. M. Broekmeulen, and J. C. Fransoo. 2010. Ordering Behavior in Retail Stores and Implications for Automated Replenishment. Forthcoming, *Management Science.*

Venkatraman, N. 1989. The concept of fit in strategy research: toward verbal and statistical correspondence. *Academy of Management Review*, 14(3), 423-444.

Workplace Systems. 2009. Retail industry workforce management solutions. Report available at http://www.workplacesystems.com.

Table 1: Summary statistics and variable description

Variable	Min	Median	Average	Max	Description
Panel 1: Operational and financial data (longitudinal monthly statistics)					
SALES	296,114	1,425,818	1,484,337	4,405,400	Revenues, $
TXN	14,121	64,612	66,240	164,327	Number of transactions, #
TXN_FCAST	14,236	64,328	66,197	167,225	Forecasted number of transactions, #
BASKET_VALUE	10.20	22.20	22.30	35.1	Average money spent by a customer per visit, $
PLAN_SALES	294,500	1,375,989	1,436,216	4,293,500	Forecast of sales, $ used for labor planning
PLAN_HOURS	1,231	9,070	9,610	26,570	Number of budgeted labor hours, hours
EE_EARNS	27,548	173,279	182,204	521,784	Employee payroll, $
MGR_EARNS	4,812	29,843	29,320	67,051	Manager payroll, $
FT_EE	2.00	20.00	20.56	116.00	Full-time Employees, #
FT_MGR	1.00	5.00	5.27	20.00	Full-time Managers, #
PT_EE	4.00	61.00	65.76	188.00	Part-time Employees, #
STORE_AUDIT	48.75	131.25	139.15	363.75	SKUs out of stock identified during audit, #
Panel 2: Store level characteristics (cross-sectional statistics)					
DIV_1	0	0	.23	1	Indicator variable, stores located in area 1, 0/1
SQFT	11,096	48,086	47,054	98,056	Total sales area, ft^2
SKUCOUNT	20,012	39,155	38,387	49,081	Product variety in the store, #
HOUSEHOLD	7	10,579	12,685	61,389	Number of households, #
AVG_HH_SIZE	1.67	2.47	2.49	3.95	Average number of persons in a household, #
HH_NOCHILD	45.9	68.3	68.1	96.3	Households without children, %
MED_AGE	25.6	36.2	37.0	69.8	Median age, years
WHITE	24.9	88.0	84.5	97.8	Caucasian in population, %
BLACK	0.0	1.9	3.3	23.5	African-American in population, %
AMER_IND	0.1	1.2	3.0	57.9	American Indians in population, %
ASIAN	0.0	3.3	5.8	46.8	Asians in population, %
OTHER	0.5	3.2	3.4	8.0	Population from other ethnic groups, %
HISPANIC	1.8	5.7	10.1	94.7	Hispanics in population, %
MED_HH_INC	0.0	49,644	50,863	120,615	Median household income, $
PER_CAP_INC	0.0	23,878	24,773	50,101	Per capita income, $
GRADE_SCHOOL	0.2	3.1	4.3	34.8	Population in grade school, %
HS	2.0	9.8	9.9	25.6	Population in high school, %
HS_GRAD	7.8	26.7	25.9	44.0	Population graduated from high school, %
COLLEGE	14.6	35.8	35.1	47.8	Population in college, %
COLLEGE_GRAD	3.9	21.6	24.6	60.9	Population graduated from college, %
Panel 3: Mismatch variables (cross-sectional statistics)					
TXN_FCASTvsTXN	0.01	0.14	0.17	0.74	Transactions forecast to transactions mismatch
TXN_FCASTvsPLAN_HOURS	0.02	0.35	0.42	1.48	Transactions forecast to planned hours mismatch
TXNvsPLAN_HOURS	0.00	0.25	0.33	1.56	Transactions to planned hours mismatch
FT_EEvsPLAN_HOURS	0.10	0.93	0.97	1.84	Full time employee to planned hours mismatch
FT_MGRvsPLAN_HOURS	0.20	0.86	0.87	1.77	Full time managers to planned hours mismatch
PT_EEvsPLAN_HOURS	0.04	0.51	0.57	1.62	Part time employee to planned hours mismatch
TOTAL_EEvsPLAN_HOURS	0.05	0.52	0.58	1.61	Total employee to planned hours mismatch
TXNvsTOTAL_EE	0.04	0.43	0.52	1.89	Total employee to transactions mismatch
PLAN_SALESvsPLAN_HOURS	0.00	0.13	0.27	1.55	Planned sales to planned hours mismatch

Table 2. Correlations among customer demographic variables.

Variable Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. HOUSEHOLD	1.00																
2. AVG_HH_SIZE	-.31	1.00															
3. HH_NOCHILD	-.42	-.91	1.00														
4. MED_AGE	-.11	-.57	.59	1.00													
5. WHITE	-.23	-.20	.14	.34	1.00												
6. BLACK	.41	.11	-.01	-.25	-.73	1.00											
7. AMER_IND	-.19	.24	-.20	-.24	-.50	.00	1.00										
8. ASIAN	.32	.04	-.01	-.14	-.77	.61	-.07	1.00									
9. OTHER	.21	.22	-.23	-.45	-.78	.62	.35	.52	1.00								
10. HISPANIC	.03	.46	-.21	-.33	-.10	.17	.19	-.08	-.00	1.00							
11. MED_HH_INC	.14	.25	-.33	-.00	-.11	.01	-.13	.30	.04	-.28	1.00						
12. PER_CAP_INC	.33	-.22	.14	.27	.03	-.02	-.20	.27	-.06	-.37	-.86	1.00					
13. GRADE_SCHOOL	-.07	.45	-.18	-.32	-.22	.20	.30	.00	.10	.87	-.44	-.52	1.00				
14. HS	-.27	.27	-.13	-.17	-.12	.15	.26	-.11	.09	.60	-.68	-.77	.71	1.00			
15. HS_GRAD	-.49	.10	-.16	.02	.05	-.05	.16	-.22	.07	-.04	-.59	-.71	.16	.64	1.00		
16. COLLEGE	-.11	-.07	-.14	-.00	.15	-.07	-.18	-.09	.05	-.53	.16	.04	-.64	-.36	.12	1.00	
17. COLLEGE_GRAD	-.44	-.30	.26	.17	.04	-.07	-.22	.20	-.13	-.31	.66	.85	-.46	-.83	-.89	-.09	1.00

Table 3. Correlations among longitudinal variables.

Variable Name	1	2	3	4	5	6	7	8	9	10	11	12
1. SALES	1.00											
2. TXN	.89	1.00										
3. TXN_FCAST	.88	.99	1.00									
4. BASKET_VALUE	.44	.02	.02	1.00								
5. PLAN_SALES	.99	.89	.88	.44	1.00							
6. PLAN_HOURS	.96	.89	.89	.36	.96	1.00						
7. EE_EARNS	.92	.89	.88	.29	.92	.92	1.00					
8. MGR_EARNS	.56	.50	.50	.29	.56	.55	.58	1.00				
9. FT_EE	.69	.63	.63	.30	.70	.70	.71	.44	1.00			
10. FT_MGR	.52	.48	.48	.23	.52	.53	.53	.87	.42	1.00		
11. PT_EE	.89	.85	.85	.27	.89	.90	.90	.51	.52	.50	1.00	
12. STORE_AUDIT	.32	.27	.27	.18	.31	.29	.20	.15	.14	.16	.32	1.00

Table 4. Correlations among cross-sectional variables

Variable Name	1	2	3	4	5	6	7	8	9	10	11
1. TXN_FCASTvsTXN	1.00										
2. TXN_FCASTvsPLAN_HOURS	.42	1.00									
3. TXNvsPLAN_HOURS	.19	.87	1.00								
4. FT_EEvsPLAN_HOURS	-.09	-.05	-.06	1.00							
5. FT_MGRvsPLAN_HOURS	-.03	.18	.21	.00	1.00						
6. PT_EEvsPLAN_HOURS	.11	.41	.48	-.21	.23	1.00					
7. TOTAL_EEvsPLAN_HOURS	.06	.40	.47	.14	.36	.86	1.00				
8. TXNvsTOTAL_EE	.10	.22	.27	.11	.23	.63	.73	1.00			
9. PLAN_SALESvsPLAN_HOURS	.11	.64	.76	-.04	.17	.45	.43	.22	1.00		
10. SQFT	.13	-.04	-.09	-.01	-.16	-.09	-.14	-.05	-.17	1.00	
11. SKU_COUNT	.18	-.11	-.19	-.01	-.18	-.13	-.15	-.10	-.22	.79	1.00

Table 5. Predicting BASKET.VALUE (using last two years of data for mismatches).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
TXNvsTOTAL_EE	-1.585** 0.019				
TXNvsPLAN_HOURS		-3.097*** 0.001	-3.132*** 0.002	-1.624*** 0.010	
TOTAL_EEvsPLAN_HOURS		-0.198 0.799			
FT_EEvsPLAN_HOURS			-1.207* 0.064	-0.769* 0.059	
FT_MGRvsPLAN_HOURS			1.069 0.154	0.155 0.743	
PT_EEvsPLAN_HOURS			-0.680 0.400	-0.427 0.431	
TOTAL_EE_PER_TXN				0.014*** 0.000	0.013*** 0.000
SKUCOUNT_PER_SQFT				6.633*** 0.001	8.061*** 0.000
SQFT				0.000*** 0.000	0.000*** 0.000
DIV_1				1.403*** 0.007	1.370*** 0.008
HOUSEHOLD				-0.000*** 0.001	-0.000*** 0.000
AVG_HH_SIZE				-0.597 0.443	-0.076 0.923
MED_AGE				-0.087* 0.064	-0.067 0.164
BLACK				9.860 0.108	5.274 0.391
AMER_IND				6.884** 0.014	6.399** 0.024
ASIAN				-15.911*** 0.000	-13.374*** 0.000
HISPANIC				-4.662* 0.087	-5.642** 0.040
PER_CAP_INC				0.000*** 0.003	0.000** 0.022
HS				0.629 0.948	1.094 0.905
HS_GRAD				3.817 0.458	1.418 0.774
COLLEGE				-14.212*** 0.003	-12.959*** 0.007
CONSTANT	23.122 0.000	23.409 0.000	23.925 0.000	6.225 0.227	2.016 0.694
N	230	230	223	223	230
R^2	0.024	0.061	0.081	0.696	0.661
Adjusted R^2	0.020	0.052	0.065	0.667	0.638
F	5.567	7.311	4.835	24.439	27.853
RMSE	3.686	3.624	3.622	2.160	2.241

Coefficient estimates on top, p-values at the bottom. ***, ** and * denote statistical significance at $p<0.01$, $p<0.05$ and $p<0.10$, respectively.

Table 6. Predicting BASKET.VALUE (robustness tests)

Variable	Model 4	Model 6	Model 7	Model 8	Model 9
TXNvsPLAN_HOURS	-1.624*** 0.010		-1.507** 0.018		
TXN_FCASTvsTXN			-1.201 0.287		
TXN_FCASTvPLAN_HOURS		-1.262** 0.027			
PLAN_SALESvsPLAN_HOURS				-0.777 0.150	-0.566 0.333
FT_EEvsPLAN_HOURS	-0.769* 0.059	-0.785* 0.055	-0.826* 0.045		-0.757* 0.067
FT_MGRvsPLAN_HOURS	0.155 0.743	0.165 0.729	0.150 0.751		0.061 0.899
PT_EEvsPLAN_HOURS	-0.427 0.431	-0.649 0.212	-0.468 0.388		-0.824 0.126
TOTAL_EE_PER_TXN	0.014*** 0.000	0.014*** 0.000	0.014*** 0.000	0.013*** 0.000	0.014*** 0.000
SKUCOUNT_PER_SQFT	6.633*** 0.001	6.543*** 0.001	6.662*** 0.001	7.840*** 0.000	6.957*** 0.001
SQFT	0.000*** 0.000	0.000*** 0.000	0.000*** 0.000	0.000*** 0.000	0.000*** 0.000
DIV_1	1.403*** 0.007	1.416*** 0.007	1.408*** 0.007	1.376*** 0.008	1.524*** 0.004
HOUSEHOLD	-0.000*** 0.001	-0.000*** 0.001	-0.000*** 0.002	-0.000*** 0.000	-0.000*** 0.001
AVG_HH_SIZE	-0.597 0.443	-0.408 0.598	-0.513 0.512	-0.191 0.808	-0.316 0.686
MED_AGE	-0.087* 0.064	-0.086* 0.070	-0.091* 0.054	-0.071 0.138	-0.078 0.101
BLACK	9.860 0.108	9.272 0.132	10.125 0.100	6.380 0.302	8.359 0.178
AMER_IND	6.884** 0.014	6.158** 0.028	6.407** 0.024	6.337** 0.025	6.199** 0.029
ASIAN	-15.911*** 0.000	-15.617*** 0.000	-16.136*** 0.000	-14.026*** 0.000	-14.933*** 0.000
HISPANIC	-4.662* 0.087	-5.140* 0.059	-5.161* 0.062	-5.126* 0.064	-5.165* 0.062
PER_CAP_INC	0.000*** 0.003	0.000*** 0.003	0.000*** 0.002	0.000** 0.014	0.000** 0.005
HS	0.629 0.948	2.454 0.799	2.696 0.784	0.260 0.977	1.216 0.902
HS_GRAD	3.817 0.458	3.517 0.496	3.542 0.492	1.956 0.693	3.521 0.500
COLLEGE	-14.212*** 0.003	-14.609*** 0.003	-14.402*** 0.003	-12.805*** 0.007	-14.869*** 0.002
CONSTANT	6.225 0.227	5.958 0.249	6.053 0.240	2.725 0.596	4.991 0.337
N	223	223	223	230	223
R^2	0.696	0.693	0.698	0.665	0.687
Adjusted R^2	0.667	0.664	0.668	0.639	0.658
F	24.439	24.134	23.290	26.375	23.465
RMSE	2.160	2.169	2.159	2.235	2.191

Coefficient estimates on top, p-values at the bottom. ***, ** and * denote statistical significance at p<0.01, p<0.05 and p<0.10, respectively.