

# Retail Store Scheduling for Profit

Nicolas Chapados<sup>a</sup>, Marc Joliveau<sup>b</sup>, Pierre L'Ecuyer<sup>a</sup>, Louis-Martin Rousseau<sup>b</sup>

<sup>a</sup>Université de Montréal, Montréal, Canada

<sup>b</sup>École Polytechnique de Montréal, Montréal, Canada

---

## Abstract

In spite of its tremendous economic significance, the problem of sales staff schedule optimization for retail stores has received relatively scant attention. Current approaches typically attempt to minimize payroll costs by closely fitting a staffing curve derived from exogenous sales forecasts, oblivious to the ability of additional staff to (sometimes) positively impact sales. In contrast, this paper frames the retail scheduling problem in terms of operating profit maximization, explicitly recognizing the dual role of sales employees as sources of revenues as well as generators of operating costs. We introduce a flexible stochastic model of retail store sales, estimated from store-specific historical data, that can account for the impact of all known sales drivers, including the number of scheduled staff, and provide an accurate sales forecast at a high intra-day resolution. We also present solution techniques based on mixed-integer (MIP) and constraint programming (CP) to efficiently solve the complex mixed integer non-linear scheduling (MINLP) problem with a profit-maximization objective. The proposed approach allows solving full weekly schedules to optimality, or near-optimality with a very small gap. On a case-study with a medium-sized retail chain, this integrated forecasting–scheduling methodology yields significant projected net profit increases on the order of 2-3% compared to baseline schedules.

*Keywords:* Shift Scheduling, Constraint Programming, Mixed Integer Programming, Non-Identical Workforce, Statistical Forecasting, Retail

---

## 1. Introduction

The retail sector accounts for a major fraction of the world's developed economies. In the United States, retail sales represented about \$3.9 trillion in 2010, over 25% of GDP, employing more than 14M people at some \$300B annual payroll costs (U.S. Census Bureau, 2011). Given these figures, it stands to reason that effective sales staff scheduling should be of critical importance to the profitable operations of a retail store, since staffing costs typically represent the second largest expense after the cost of goods sold (Ton, 2009): as a result, all efficiencies coming from better workforce deployment translate into an implicit margin expansion for the retailer, which immediately accrues to the bottom line.

---

*Email addresses:* chapados@iro.umontreal.ca (Nicolas Chapados), marc.joliveau@gmail.com (Marc Joliveau), lecuyer@iro.umontreal.ca (Pierre L'Ecuyer), louis-martin.rousseau@polymtl.ca (Louis-Martin Rousseau)

The best retailers today rely on a staff schedule construction process that involves a decomposition into three steps (Netessine et al., 2010). First, the future sales over the planning horizon are forecasted, usually a few weeks to one month ahead at a 15- to 60-minute resolution. Second, this forecast is converted into labor requirements using so-called “labor standards” established by the business (e.g., every \$100 in predicted sales during a given 15-minute period requires an additional salesperson that should be working during that period). Finally, work schedules are optimized in a way that attempts to match those labor requirements as snugly as possible, while meeting other business and regulatory constraints (e.g., one may have that an employee cannot be asked to come to work for less than three consecutive hours or for more than eight hours in total, must have breaks that follow certain rules, etc.).

It may be somewhat unsettling that nowhere does this process acknowledge, explicitly or tacitly, that in many retail circumstances, salespeople *actively contribute to revenue* by doing their job well—advising an extra belt with those trousers, or these lovely earrings with that necklace—and not only represent a salary cost item to store operations. In other words, the presence of an additional staff working at the right time drives up expected sales, a crucial dynamics ignored when sales forecast tranquilly descends “from above”.

This paper, in contrast, formalizes the retail staff scheduling problem by formulating it as one of *expected net operating profit maximization*. It introduces a modeling decomposition that allows representing the expected sales during a time period as a function of the number of salespersons working, thereby capturing the impact of varying staffing hypotheses on the expected sales. The profit-maximizing labor requirements are obtained, for a given time period, as the number of employees beyond which the marginal payroll cost exceed the marginal revenue gain from having an additional staff working. Finally, it introduces new solution techniques that are able to capture the profit maximization aspects of the problem. On a case-study with a medium-sized Canadian clothing and apparel chain, this integrated forecasting–scheduling methodology yields significant projected net profit increases on the order of 2 to 3% with respect to baseline schedules.

Given the retail industry’s overall economic significance, the impact of staffing decisions has received some attention in the literature, albeit perhaps not any commensurate with the gains that are to be expected from improved planning. Thompson (1995) proposed a scheduling model that takes into accounts a linear estimate of the marginal benefit of having additional employees, but not accounting for understaffing costs. Lam et al. (1998) and Perdikaki et al. (2012) have shown that store revenues are causally and positively related to staffing levels, opening the door to staffing rules based on sales forecasts. Moreover, using data from a large retailer, Ton (2009) finds that increasing the amount of labor at a store is associated with an increase in profitability through its impact on conformance quality but, surprisingly, not its impact on service quality. Mani et al. (2011) find systematic understaffing during peak hours in a study with a large retail chain, using a structural estimation technique to estimate the contribution of labor to sales. Additionally, in what is probably the most potent case to date for improving scheduling practices, Netessine et al. (2010) find—with another large retailer—strong cross-sectional association between labor practices at different stores and basket values, and observe in their examples that small improvements in employee scheduling and schedule execution can result in a 3% sales increase for a moderate (or zero) cost.

Perhaps the contribution closest in flavor to what we suggest in this paper comes from the work of Kabak et al. (2008): similarly to the present paper, these authors introduce a two-stage approach to retail workforce scheduling based on a statistical forecasting model of sales incorporating an explicit labor effect. The profit-maximizing number of salespeople under this model is

then computed to obtain the desired hourly staffing levels in the store, assuming workforce cost homogeneity. This is followed by an optimization phase, which assigns individual employees to a set of pre-existing shifts. The approach is then validated through simulation to demonstrate its effectiveness on a Turkish apparel store.

As we show later, our paper extends this approach by suggesting the use of complete *sales curves* within the schedule optimization stage, better reflecting the uncertainty surrounding the sales response model (which often exhibits a rather flat shape around the optimal number of salespeople), and easily allows optimizing schedules with non-homogenous staff costs. To achieve this goal, we first propose a stochastic model of retail store sales in terms of a revenue decomposition that is particularly amenable to robust and modular modeling. This decomposition allows to obtain an accurate 15-minute-resolution forecast of sales, conditional on any desired determinant, opening the path to operating profit maximization.

Furthermore, in contrast to Kabak et al. (2008), the proposed schedule optimization does not use a set of predefined shifts, but rather builds those that, while meeting labor and union regulations, allow to maximize profit. The main difficulty is to efficiently combine two distinct sets of objectives and constraints: those of the retailer, defined in terms of the number of working employees in each time period, and those of the workers, defined in terms of the “friendliness” of the working shifts assigned to them. We formulate the profit-maximization problem as a mixed integer non-linear problem (MINLP) where the decision variables specify the work status (working, break, lunch, rest) of each employee, and we examine two ways of solving this problem. The first approach linearizes the problem and turns it into a mixed integer program (MIP), which is then solved either directly or approximately using piecewise linear functions. The second approach expresses the original MINLP as a constraint program (CP), which is solved directly. It provides better performance, especially on larger problems. The proposed problem formulations are based on the use of a regular language to specify the admissibility of shifts with respect to union agreements. Schedules are constructed both over a day and a week.

The rest is organized as follows. In Section 2 we formulate the problem, setting forth overall schedule construction objectives. In Section 3, we explain the stochastic models we developed for sub-hour sales forecasts, and we evaluate the forecasting ability of these models on out-of-sample sales at several locations of a mid-sized clothing and apparel chain of retail stores. We continue in Section 4 with two scheduling problem formulations, one as a MIP and the other as a CP, whose solutions are employee schedules that maximize expected profits, over a day and over a week (as shown in Section 5). Section 6 concludes. This paper extends work previously presented by Chapados et al. (2011).

## 2. Problem Formulation

Here we present an overview of the proposed approaches. The global schedule construction process has two main steps, summarized in Figure 1:

1. Analyzing the historical workforce sales performance to construct a stochastic model of store behavior and to build *sales curves*, which give the expected sales at each store for each time period, as a function of the number of staffs assigned to sales during that period.
2. Using these sales curves to construct admissible work schedules, for the employees, that maximize the expected profit.

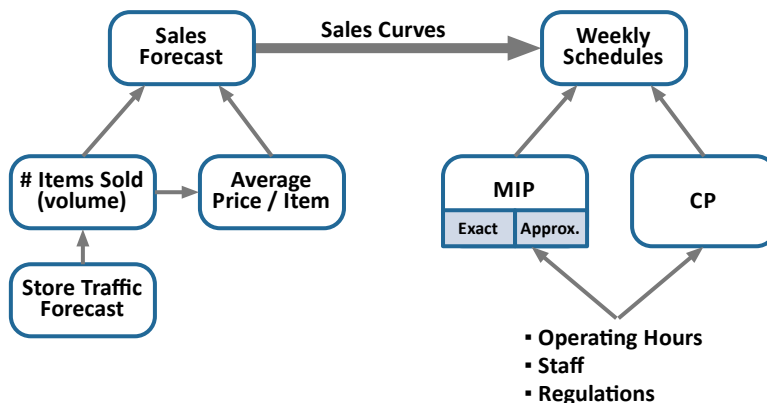


Figure 1: Overview of the proposed methodology to address the problem of retail scheduling for profit.

### 2.1. Forecasting Sales as a Function of Staffing

Each sales curve provides a functional characterization of how the expected sales at a given store and time period are impacted by varying the staffing. A classical staffing *demand curve*, in contrast, indicates the desired number of employees at a given store, for each time period.

As outlined earlier and in Fig. 1, we model store sales through a decomposition into two simpler quantities: (i) the number of items sold during a time period, and (ii) the average price per item. Of course, sales could be forecast without such a decomposition, by applying standard statistical time series approaches directly. But at the high intra-day resolutions (15-minute periods) that we are considering, we found forecasts based on the decomposition to be markedly more accurate than standard time series models having similar parametric complexity, not only in expectation but with respect to the whole distribution as well.

An important driving factor affecting the number of items sold is the *in-store traffic*, of which we require a precise forecast at a high intra-day resolution over the entire planning horizon, accounting for all known explanatory variables and capable of yielding sensible forecasts for stores for which little amounts of historical data are available. To this end, we introduce a traffic forecasting methodology that allows sharing statistical power across multiple locations (for a chain of stores), yet allowing store-specific effects to be accurately handled.

To forecast sales, a key aspect of our contribution is the ability to parsimoniously model the joint distribution between the two factors in the decomposition; to achieve this, we require a model of the full posterior distribution of the number of items sold during a time period, conditional on the known sales drivers. This distribution exhibits significantly higher dispersion than a conditional Poisson, and a standard Poisson Generalized Linear Model (McCullagh and Nelder, 1989)—which linearly relates the intensity of a conditional Poisson random variable to a set of covariates—proves insufficient for the purpose. We adapt the statistical technique of *ordinal regression* to obtain an accurate estimator of the posterior distribution of the number of items sold.

The aforementioned decomposition is related to one often used by retail practitioners (e.g., Lam et al. 2001) wherein sales during a given period are written as the product of three factors: the number of shoppers in the store during the period, the conversion rate (how many shoppers be-

come buyers during that period), and the average basket (given that a shopper buys, how much does s/he buy on average). However, this conventional decomposition is very sensitive to having a precise measure of the number of shoppers in the store at any time—an information that many retailers cannot measure precisely, or do not have at all.

A “demand curve” can be obtained from the sales curve by subtracting from the latter the wage cost incurred at each staffing level—yielding the profit curve—and finding, on the profit curve for each store/time-period pair, the staffing that maximizes profit. This demand curve is often useful for management purposes, providing a concise and interpretable summary of how many employees “should” be working at any particular time. As mentioned previously this type of optimal staffing for profit maximization is studied by Kabak et al. (2008) and also by Koole and Pot (2011) in the context of call centers where successful calls can bring revenue. However, as we show in Section 4, it is generally preferable to incorporate the sales curves directly into the mathematical program that constructs the schedules, rather than trying to fit the demand curve. This overcomes two limitations of traditional demand curves. Firstly, it allows true staff costs to be accounted for, at the individual employee level, allowing to dispense with a cost homogeneity assumption. Secondly, it frequently happens that the profit curve is fairly flat around its maximum, in which case there is a large confidence band around the “truly optimal” number of employees; this allows for more flexibility during the scheduling process to produce better schedules with only minor impact on the cost.

## 2.2. *Scheduling for profit*

In the retail store context, finding the optimal number of salespersons for each time period (the optimal staffing) is not sufficient, because there is generally no way to match these optimal numbers exactly given the various constraints on the admissible shifts of employees. We could think of a two-stage approach that computes the optimal staffing in a first stage and then finds a set of admissible shifts that cover these required staffings, say by having a number of salespersons larger or equal to the optimal number in each time period, at minimal cost. But such a procedure is generally suboptimal, as illustrated for example in Avramidis et al. (2010) in the context of work schedule construction for telephone call centers.

Here we solve the optimization problem in a single stage. The decision variables select a set of admissible working shifts, one for each employee, over the considered time horizon (one day or one week, for example). The objective is to maximize the total expected profit, rather than merely minimizing salary costs under staffing constraints as usually done. That is, we recognize explicitly the dual role played by employees, who generate costs from their salaries but also revenues by driving up sales. This gives a novel problem formulation.

In the objective, the expected revenue as a function of the number of working employees, for each time period of each day, is given by the sales curves, which are generally nonlinear, and these expected revenues are added over the considered time horizon. The cost of each shift is assumed to be decomposable as a sum of costs over the time periods it covers, where the cost over a time period depends on the day, the period, and the activity performed over that period. These costs are added up over all selected shifts, and subtracted from the total expected revenues.

This gives an integer programming problem with a nonlinear objective. The nonlinearity can be handled by various techniques, such as extending the formulation by using boolean variables for every possible number of working employees in each period, approximating the sales curves with piecewise linear functions, and modeling the nonlinear objective directly with a constraint program (see Section 4 for the details).

Table 1: Summary statistics of the data used in the case studies: the traffic refers to the number of clients entering the store as measured by people counters and the staff-hours refers to the staffing levels historically scheduled for the stores during the period covered.

	Daily Traffic (persons)		Daily Sales (C\$)		Daily Items Sold		Daily Staff-Hours	
	Avg	Std. Dev.	Avg	Std. Dev.	Avg	Std. Dev.	Avg	Std. Dev.
Smallest store	199	77	2486	1560	29	20	19	5.6
Average store	402	299	4329	2865	54	49	26	8.1
Largest store	955	681	8494	5891	163	129	42	8.5

The constraints are of two kinds. Some address operational issues of employers, such as imposing a minimal number of working employees at any given time, constraints on the number of full-time and part-time employees, and operating hours of the store. Other constraints deal with union regulations or labor law, such as minimum and maximum number of hours of work per day for an employee, minimum and maximum number of hours of work per week, break and meal placement regulations, minimum rest time (delay) between successive shifts for an employee, etc. Each shift specifies a sequence of activities over the entire time horizon (e.g., a day or a week). In our formulation, employees can have different availabilities, but are assumed homogenous in terms of selling efficiency.

### 2.3. Data for the Case study and Experimentation Context

The data used for our case study was provided by a medium-sized chain of upscale clothing and apparel retail stores. It comes from 15 stores, located in major Canadian cities, which obey normal retail opening hours and sales cycle, including increased holiday activity and occasional promotional events, such as a mid-year sale and an after-Christmas sale. Stores are of two types: regular “signature” stores in standard shopping locations such as malls, and discount outlets, which exhibit substantially higher volume.

All stores are open seven days a week, with varying hours per day and per store; they are closed only a small number of days per year on major holidays (Christmas, New Year, Easter). A total of 20 months of historical data was used, covering the January 2009–August-2010 period. For the purpose of estimating and evaluating the statistical forecasting models, a total of 8038 store–days and 167,725 store–intra-day (30-minute) intervals were used. The available traffic data was aggregated by 30-minute intervals, so we were able to construct sales curves for those intervals. On the other hand, the work schedules are defined based on 15-minute time intervals (the breaks last 15 minutes, for example). So for the scheduling, each 30-minute period was split in two 15-minute intervals with the same sales curve. To evaluate the schedule optimization methods, the 5089 store–days covering the latter portion of the data were used. Summary statistics of the average traffic, sales, items sold, and historical staffing are shown in Table 1. Confidentiality agreements preclude from giving additional details.

## 3. Stochastic Modeling and Forecasting for Sales Curves Construction

### 3.1. Model Overview and Decomposition of Expected Sales

In this section, we develop stochastic models to construct the sales curves. For this, we decompose the expected sales in any given time period as the expectation of the product of the number of items sold by the expected average price per item conditional on the first number.

Then we develop a stochastic model for the cumulative distribution function (cdf) of the number of items sold as a function of the number of working employees on the floor, the estimated in-store traffic (or traffic proxy), and certain explanatory variables that account for seasonal effects and special events. This model uses ordinal regression with cubic spline basis functions. The estimated daily in-store traffic is modeled in log scale by a mixed-effects linear model, whose fixed effects are indicators that account for the month of the year, day of the week, interactions, and special events, and whose residuals follow a SARMA process. This estimated traffic is assumed to be distributed across opening hours of the day according to a multinomial distribution, estimated by multinomial regression. The overall model is summarized on the left of Figure 1. Sections 3.3 to 3.5 detail the submodels for the traffic proxy, the sales volume, and the conditional average item price, respectively. Reports on experimental results and validation are given at the same time.

Simple ways of measuring the in-store traffic in a given time period could be by counting the number of people entering the store during that period, or perhaps by the average number of people in the store during that period. However, in practical settings, this type of information is often unavailable, and one must rely on some indirect measure of store activity which we shall call a *traffic proxy*. In our case study, each store was equipped with people counters that register a “tick” each time a customer enters or leaves the store (the device does not distinguish the direction), and simply count the number of ticks in each period. This recorded data does not tell precisely how many people are present in the store at any given time, or how many entered in a given time period, but nonetheless correlates well with store activity and was the best available information in our case. Our traffic proxy was thus the “tick count” within each given time period.

We consider  $j^*$  stores indexed by  $j$ , over  $d^*$  days indexed by  $d$ . Each day is divided in  $t^*(d, j)$  time periods indexed by  $t$  (the number of time periods generally depends on the day of the week and on the store). For period  $t$  of day  $d$  in store  $j$ , let  $y_{d,t,j}^w$  be the number of employees assigned to sales on the floor (the *selling employees*),  $\tilde{T}_{d,t,j}$  denote the traffic proxy,  $\mathbf{X}_{d,t,j}$  a vector of selected explanatory variables such as calendar regressors (time of the day, day of the week, month of the year) and indicators for special events,  $V_{d,t,j}$  the number of items sold (the *sales volume*),  $P_{d,t,j}$  the average price per item, and  $R_{d,t,j} = V_{d,t,j}P_{d,t,j}$  the total amount of sales. The number  $y_{d,t,j}^w$  of selling employees excludes the employees that (we assume) have no direct impact in turning shoppers into buyers; for instance cashiers that do not directly assist customers on the floor and employees tasked with replenishing shelves, as well as those on break. In this paper, when scheduling the employees, we assume that there is only one type of work activity: each considered employee is assigned to sales. That is, we do not consider the situation where an employee works on the floor for some periods and as a cashier or something else on other periods. To cover this type of situation we could extend our framework to a multi-activity setting, which would make the resolution more complicated (but still feasible).

The sales curves express the expected sales  $r_{d,t,j}(y) = \mathbb{E}[R_{d,t,j} \mid y_{d,t,j}^w = y]$  for period  $t$  of day  $d$  in store  $j$  as a function of the number  $y$  of selling employees in the store for that period. To estimate the sales curves, we decompose the expectation as follows (the dependence of each

term on the deterministic explanatory variables  $\mathbf{X}_{d,t,j}$  is implicit):

$$\begin{aligned}
r_{d,t,j}(y) &= \mathbb{E}[\mathbb{E}[R_{d,t,j} \mid \tilde{T}_{d,t,j}, y_{d,t,j}^{\mathbf{w}} = y]] \\
&= \mathbb{E}[\mathbb{E}[V_{d,t,j} \mathbb{E}[P_{d,t,j} \mid V_{d,t,j}] \mid \tilde{T}_{d,t,j}, y_{d,t,j}^{\mathbf{w}} = y]] \\
&= \sum_{v=0}^{\infty} v \mathbb{E}[\mathbb{P}(V_{d,t,j} = v \mid \tilde{T}_{d,t,j}, y_{d,t,j}^{\mathbf{w}} = y)] \mathbb{E}[P_{d,t,j} \mid V_{d,t,j} = v]. \quad (1)
\end{aligned}$$

We shall construct separate models for the two expectations that appear in (1): the expected conditional probability (the conditional distribution of the volume) and the expected average price per item as a function of the volume. The first expectation is with respect to  $\tilde{T}_{d,t,j}$ , so we need a model for this random variable, and also for the conditional probability inside the expectation. The reason for introducing such a decomposition is that it reduces significantly the estimation error in comparison with a direct model of the expected sales (see Section 3.6). With the decomposition, our model for  $\mathbb{E}[P_{d,t,j} \mid V_{d,t,j} = v]$  can use a single parameter vector estimated from all the data, which makes the model much more accurate. This idea of decomposing the sales in terms of more fundamental quantities is traditional in retail sales modeling. Different decompositions have been proposed; e.g., a popular one in retail expresses the sales as the product of the traffic, conversion rate, and average basket value (Lam et al. 2001). But with the traffic proxy measurements available in our case, we found the latter decomposition to be less robust to statistical estimation errors than the one we have adopted.

To speed up computations, we will in fact replace  $\mathbb{E}[\mathbb{P}(V_{d,t,j} = v \mid \tilde{T}_{d,t,j}, y_{d,t,j}^{\mathbf{w}} = y)]$  in (1) by the approximation  $\mathbb{P}(V_{d,t,j} = v \mid \mathbb{E}[\tilde{T}_{d,t,j}], y_{d,t,j}^{\mathbf{w}} = y)$ , in which  $\tilde{T}_{d,t,j}$  is replaced by its expectation  $\mathbb{E}[\tilde{T}_{d,t,j}]$ . This introduces an approximation error due to the nonlinearity of the conditional probability as a function of  $\tilde{T}_{d,t,j}$  in (1), resulting in slightly less probability mass in the tails and more mass near the mean for the random variable  $V_{d,t,j}$ . The accurate (but more expensive) option would be to integrate numerically with respect to the density of  $\tilde{T}_{d,t,j}$  to compute  $\mathbb{E}[\mathbb{P}(V_{d,t,j} = v \mid \tilde{T}_{d,t,j}, y_{d,t,j}^{\mathbf{w}} = y)]$ . In our case study, we found that using the cruder approximation did not bring significant changes on the sales curves and scheduling solutions.

### 3.2. Model Comparison Methodology

In this subsection, we explain how we will test the goodness of fit of our proposed models. To account for the sequential nature of the data and possible non-stationarities, our testing experiments are conducted in a *sequential validation* framework (also known as “simulated out-of-sample”): the models are initially trained on the 2009/01–2010/02 time interval, and tested out-of-sample on the one-month period of 2010/03. This month is then added to the training set, the model is retrained and tested on the subsequent month; this whole process is repeated until 2010/08, the last month in our data set. The reported performance results are the average across all such training/testing runs.

#### 3.2.1. Measures of Forecasting Accuracy

The following measures are used to evaluate the statistical accuracy of model forecasts. We consider a generic time series of random variables  $\{Y_i, i = 1, \dots, n\}$  with observed realizations (“targets”)  $\{y_i, i = 1, \dots, n\}$ , and corresponding model responses giving the forecasts  $\{\hat{y}_i, i = 1, \dots, n\}$ . For example, when predicting the daily traffic  $\tilde{T}_{d,j}$  at store  $j$ , we have  $i = d$  and  $Y_i = \tilde{T}_{d,j}$ .



The *mean squared error* (MSE) is defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

the *root mean squared error* (RMSE) is defined as  $\sqrt{\text{MSE}}$ , and the *mean absolute percentage error* (MAPE) is defined as

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

The MAPE is appropriate when we are interested in relative errors for strictly positive target quantities. To avoid numerical instability in the results, we eliminate from the error computation the very small targets in absolute value ( $|y_i| < 10^{-4}$ ).

For  $Y_i = V_{d,t,j}$  (the sales volume), which has a discrete distribution conditional on  $(\tilde{T}_{d,t,j}, \mathbf{y}_{d,t,j}^{\mathbf{w}})$ , we also evaluate the *negative log likelihood* (NLL), defined as

$$\text{NLL} = - \sum_{i=1}^n \log \hat{\mathbb{P}}(Y_i = y_i \mid \tilde{T}_{d,t,j}, \mathbf{y}_{d,t,j}^{\mathbf{w}}),$$

where  $\hat{\mathbb{P}}$  denotes the conditional probability under the considered model. This NLL measures the fit of the entire distribution; it evaluates the ability of a model to put conditional probability mass in the “right places”, as opposed to just getting the conditional expectation right (the latter being evaluated by the MSE, since the conditional expectation minimizes the MSE; see Lehmann and Casella, 1998).

Recall that the  $q$ -quantile of a random variable  $Y$  is defined as  $y^q = \inf\{y \in \mathbb{R} : \mathbb{P}[Y \leq y] \geq q\}$ . For continuous-response models that yield prediction intervals (traffic and item-price models), we evaluate *interval coverage* by measuring at which actual quantiles (in the test data) are mapped nominal quantiles given by the prediction intervals. For example, when we compute a 95% prediction interval, we estimate the 0.025 and 0.975 quantiles, and we would like to know how good are these estimates. We estimate a  $q$ -quantile  $y^q$  by the  $q$ -quantile  $\hat{y}^q$  of the model’s predictive distribution, and we test the accuracy of this estimate by computing the fraction  $f_q$  of observations  $y_1, \dots, y_n$  that fall below  $\hat{y}^q$ . Ideally,  $f_q$  should be very close to  $q$ . We will report the values of  $f_q$  for  $q = 0.025, 0.1, 0.9, \text{ and } 0.975$ . To a certain extent, these values also measure the goodness of our models in terms of distributional forecasts.

### 3.2.2. Significance of the Predictive Performance Comparisons

When we compare models against each other and we want to test the significance of the difference in predictive ability between any two models, we proceed as follows. Let  $\{\hat{y}_{1,i}, i = 1, \dots, n\}$  and  $\{\hat{y}_{2,i}, i = 1, \dots, n\}$  be the forecast time series of two models to be compared, and let  $g(y_i, \hat{y}_i)$  be a measure of error between the realization  $y_i$  and the forecast  $\hat{y}_i$ ; for example  $g(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$  if we use the square error. We define the sequence of *error differences* between models 1 and 2 by  $d_i = g(y_i, \hat{y}_{1,i}) - g(y_i, \hat{y}_{2,i}), i = 1, \dots, n$ . Let  $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$  be the empirical mean difference. It is important to note that these  $d_i$  are generally *dependent*, so to test if there is a difference in forecasting performance, one cannot just apply a standard  $t$ -test as if the  $d_i$  were independent. For this purpose, to test the statistical significance of  $\bar{d}$ , we use the Diebold-Mariano (D-M) test (Diebold and Mariano, 1995), which estimates the variance of  $\bar{d}$  by

estimating the correlations as follows. If  $\{d_i, i \geq 1\}$  is assumed to be a stationary time series with lag- $k$  autocovariance  $\gamma_k$  for  $k \geq 0$ , then

$$\text{Var}[\bar{d}] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(d_i, d_j) = \frac{1}{n^2} \sum_{k=-n+1}^{n-1} (n-k)\gamma_k.$$

The  $\gamma_k$ 's are assumed negligible outside a certain window; that is, one assumes that  $\gamma_k = 0$  for  $|k| > k_0$  for some window length  $k_0$ . If  $\hat{\gamma}_k$  is a consistent estimator of  $\gamma_k$  for  $|k| \leq k_0$ , and  $n$  is large compared with  $k_0$ , then we can estimate  $\text{Var}[\bar{d}]$  by  $\hat{v}_{\text{DM}} = \frac{1}{n} \sum_{k=-k_0}^{k_0} \hat{\gamma}_k$ . Under (standard) mild conditions, the D-M statistic  $\bar{d}/\sqrt{\hat{v}_{\text{DM}}}$  is asymptotically distributed as a  $\mathcal{N}(0, 1)$  random variable under the null hypothesis that there is no predictive difference between the two models, and a classical test can be carried out based on this. In this paper we use  $k_0 = 5$ ; this was determined from the observed autocorrelation structure in the results. In our results, we will report the  $p$ -value  $p(\text{D-M})$  for each D-M (two-side) test that we perform.

### 3.3. Traffic Estimation

#### 3.3.1. Data on Traffic

The traffic at retail stores is characterized by yearly, weekly, and intraday seasonalities, as illustrated in Figure 2, which shows the evolution of the traffic proxy for one clothing and apparel store in our case study over 8 months (left) and over 5 days (right). On the left, one can observe a significant traffic increase during the end-of-year shopping season (including after-Christmas sales). Traffic is also much higher over the weekends, which correspond to the regular ‘‘spikes’’ in the traffic pattern. On the right, one can observe the patterns at the individual-day level. This plot is based on counts per 30-minute periods while the store is open and these counts have a large relative variance.

Although not a traffic characterization *per se*, one endemic problem with store traffic data is the prevalence of missing values (shown as gaps in the solid traffic curve in the left panel of Figure 2). These may be due to a number of factors: sensor failure, downtime of the recording computer, staff changing store mannequins and obstructing sensors for a long time, etc. It is important that the forecasting methodology be robust to the regular occurrence of such missing data.

The goal of traffic modeling is to characterize the customer arrival process to a store and to make forecasts about future traffic (or its proxy) over the planning horizon. For scheduling purposes, we need forecasts at a high intraday resolution, such as 15- or 30-minute periods. These forecasts should account for all known characteristics of retail traffic patterns, such as seasonalities, autocorrelations, impact of special events and other covariates. We also want to share statistical power across stores, namely use historical data from other stores to help predictions made about any given store. This is particularly valuable when new stores can open on a regular basis, having little historical data of their own.

Our strategy is to first model the total traffic (proxy) at any given store during the whole day using a mixed-effects log-linear model, whose covariates have a multiplicative effect on the response variable. Then, a store-specific SARMA process is fitted to the residuals of this log-linear daily model, to capture the dependence across successive days at a given store. Finally, an intraday distribution model spreads the daily traffic across the intraday periods that correspond to the store-specific opening hours for that day, according to a multinomial distribution. These three model pieces are summarized in Fig. 3 and detailed in Subsections 3.3.2 to 3.3.4.

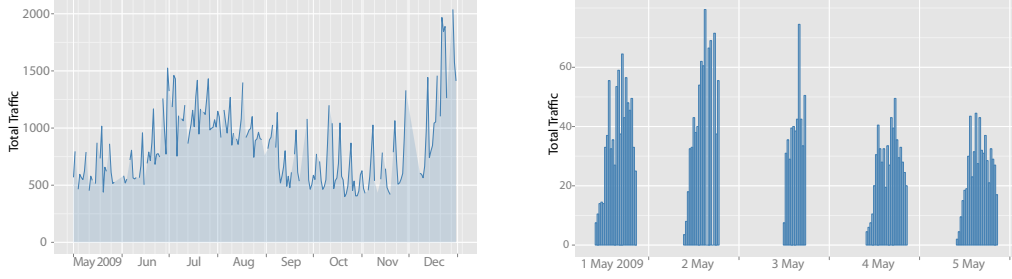


Figure 2: **Left:** Daily traffic at a retail store, illustrating seasonalities at the weekly and yearly horizons. **Right:** Intraday (30-minute) traffic, showing some seasonal behavior but also high noise levels. Throughout, there are missing data, indicated by gaps in the curve.

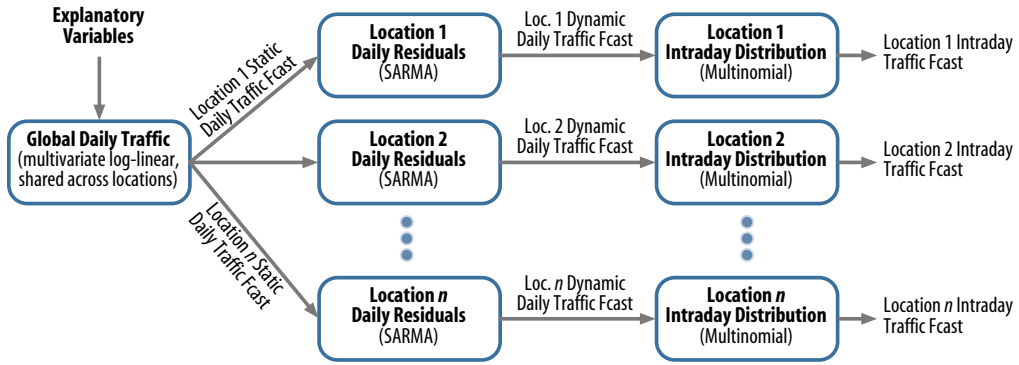


Figure 3: Overall modeling strategy for forecasting the intraday traffic at a number of retail locations, while being capable of exploiting the commonalities across locations.

### 3.3.2. Static Total Daily Traffic Model

We model the total daily traffic  $\tilde{T}_{d,j}$  on day  $d$  at store  $j$ , for  $j = 1, \dots, j^*$ , with a log-linear mixed-effects model (McCulloch et al., 2008; Pinheiro and Bates, 2000) of the general form

$$\log \tilde{T}_{d,j} = \beta_{0,j} + \beta' \mathbf{x}_{d,j} + \epsilon_{d,j}, \quad (2)$$

where  $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,j^*})$  is a vector of random store-specific intercepts,  $\beta$  is a vector of regression coefficients shared among all stores,  $\mathbf{x}_{d,j}$  is a vector of store-specific indicator variables for day  $d$ , and  $\epsilon_{d,j}$  is a zero-mean residual with variance  $\sigma_{d,j}^2$ . These residuals are discussed in Subsection 3.3.3.

For our dataset, we ended up with the specific formulation:

$$\log \tilde{T}_{d,j} = \beta_{0,j} + \beta'_D D_d + \beta'_M M_d + \beta'_{DL} DL_{d,j} + \beta'_E E_{d,j} + \epsilon_{d,j}, \quad (3)$$

where  $D_d$ ,  $M_d$ ,  $DL_{d,j}$ , and  $E_{d,j}$  are vectors of 0–1 indicators (dummy variables) for the day of the week, month of the year, interactions between the day of week and the store, and special events occurring at store, for store  $j$  on day  $d$ . For example, each  $D_d$  is a six-dimensional unit vector with a 1 at the position that corresponds to the day type of day  $d$  (the effect of the first day is absorbed in the intercept  $\beta_{0,j}$ , so we have six remaining daily effects instead of seven),

and zeros elsewhere, while  $\beta_D$  is a six-dimensional vector of coefficients. The vector  $\beta$  is the concatenation of  $\beta'_D$ ,  $\beta'_M$ ,  $\beta'_{DL}$ , and  $\beta'_E$ . The coefficients in  $\{\beta_D, \beta_M, \beta_E\}$  are shared across stores, which allows sensible forecasts to be carried out even for stores which have relatively little data. Obviously, for store chains with greater cross-sectional variability than in our sample, additional interaction terms could be incorporated to better capture store specificities; in our case, such refinements did not help. Although not used for the results presented in this paper, additional variables can be added, such as meteorological conditions (or forecasts thereof), significant sports events, macroeconomic variables or a continuous time variable for a long-term trend. Preliminary investigations suggest that weather-related variables can provide material improvements in traffic forecast quality in a retail context.

The model parameters  $\beta_0$  and  $\beta$  are estimated by ordinary least squares (OLS) regression. For the purposes of estimating the model parameters, the residual variance  $\sigma_{d,j}^2$  is assumed to be a common constant  $\sigma^2$ , but when using the model for forecasts, we take the variance obtained from the residual model introduced in section 3.3.3. Under this model, the predictive distribution of  $\tilde{T}_{d,j}$  is lognormal, so the traffic cannot take negative values. If we denote  $\mu_{d,j} \equiv \mathbb{E}[\log \tilde{T}_{d,j} \mid \mathbf{x}_{d,j}]$ , the conditional expected traffic can be written as

$$\mathbb{E}[\tilde{T}_{d,j} \mid \mathbf{x}_{d,j}] = \exp(\mu_{d,j} + \sigma_{d,j}^2/2). \quad (4)$$

Note that this model is reasonable only in situations where the traffic level  $\tilde{T}_{d,j}$  is distributed over sufficiently large values, i.e., not concentrated over just a few non-negative integers close to 0, and with no significant point mass at zero. Were this not the case, an alternative is to use a Poisson regression model (McCullagh and Nelder, 1989).

### 3.3.3. Residual Model

The residual model aims at extracting the dynamic structure present in the process of daily residuals  $\epsilon_{d,j}$  that remains unexplained by (3). This model operates at the *individual store level*, each store yielding a univariate time series of daily residuals  $\epsilon_{d,j}$ —where the store  $j$  is kept constant and the day  $d$  varies—modeled by a SARMA( $p, q$ )  $\times$  ( $P, Q$ ) $_S$  form (seasonal autoregressive moving average; see, e.g., Brockwell and Davis, 1991), specified as

$$\Phi_P(L^S)\phi_p(L)\epsilon_{d,j} = \Theta_Q(L^S)\theta_q(L)\nu_{d,j}, \quad j \text{ held fixed}, \quad (5)$$

where  $\phi_p$  and  $\theta_q$  are polynomials of orders  $p$  and  $q$  in the lag operator  $L$ ,  $\Phi_P$  and  $\Theta_Q$  are polynomials of orders  $P$  and  $Q$  in the seasonal lag operator  $L^S$ , respectively, and  $\{\nu_{d,j}, d = 1, \dots, d^*\}$ , is an i.i.d. zero-mean innovations process with finite variance  $\sigma_{\nu,j}^2$ , for each  $j$ . In our context, the seasonality  $S$  is taken to be 7 days. The model parameters (coefficients of the polynomials  $\Phi_P$ ,  $\phi_p$ ,  $\Theta_Q$ , and  $\theta_q$ ) are estimated by maximum likelihood. The choice of model order (values of  $p, q, P, Q$ ) is carried out automatically using an AIC criterion; see, e.g., Makridakis et al. (1997). This procedure is implemented with the help of the `forecast` package in the R statistical language (Hyndman et al., 2012). We emphasize that since a distinct model is estimated for each store, the model order across stores can be completely different.

### 3.3.4. Comparison of Daily Traffic Models

We will compare three variants of our daily traffic model: (a) the static model in (3) with the indicator covariates only, with the residuals assumed to be independent (termed *covariates only*); (b) the model for the SARMA residuals only, without the covariates (termed *SARMA residuals*)

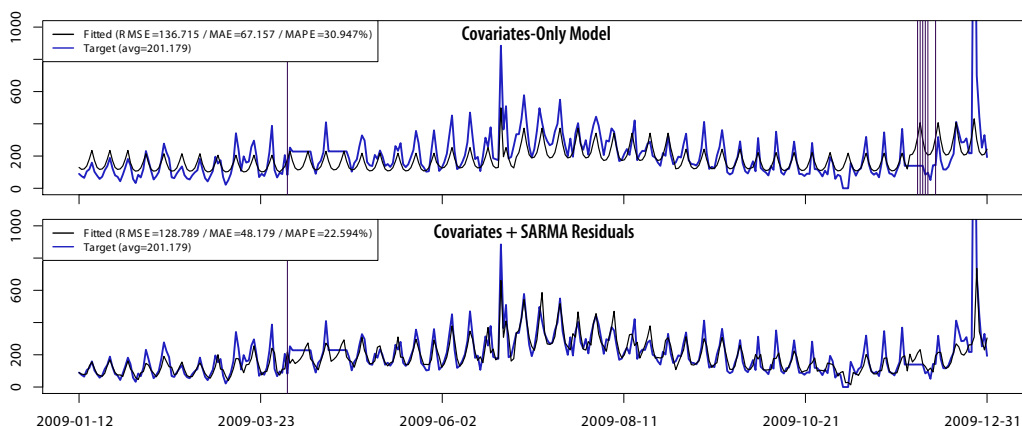


Figure 4: Training fit of the daily traffic forecasting model at a single retail location for the year 2009, using the model with covariates only (top panel) and the model with covariates + SARMA residuals (bottom panel). The latter gives a much better fit. In both cases, the blue line indicates the realized (target) traffic, and the grey line gives the fitted values. Vertical lines indicate days where the relative error exceeds 100%.

Table 2: In- and out-of-sample performance results for several specifications of the daily traffic model. The D-M column compares a given model to the ‘full model’ in the first row on the MSE basis. The other performance statistics are explained in the text.

Model	RMSE (Persons)		MAPE (%)		Quantiles				D-M
	Train	Test	Train	Test	0.025	0.1	0.9	0.975	
Covariates + SARMA resid.	181.65	213.56	18.34	22.74	0.019	0.068	0.935	0.977	
Covariates only	223.64	216.88	25.30	24.91	0.008	0.041	0.950	0.977	1.240
SARMA residuals only	197.69	247.95	21.55	27.42	0.003	0.013	0.921	0.973	<b>2.440 **</b>

only); and (c) the model that combines the previous two, with the covariates at the highest level and a SARMA model for its residuals (termed *covariates + SARMA residuals*).

When reporting the results and test the goodness of fit, we convert all model responses from the logarithmic scale back to their original scale in “daily persons count.” For the conditional expectation, this involves accounting for the conditional variance due to Jensen’s inequality, namely for a random variable  $X \sim \mathcal{N}(\mu, \sigma^2)$ , we have  $\mathbb{E}[e^X] = e^{\mu + \sigma^2/2}$ .

Figure 4 shows the training fit (in-sample) for one of the 15 stores, under the model with covariates only (top) and the model with covariates + SARMA residuals (bottom). The latter gives a much better fit.

Table 2 shows the performance results of the three models. The combined model exhibits the best performance, both in-sample (the Train column) and out-of-sample (the Test column), and on both the RMSE and MAPE measures. This is corroborated by the D-M statistic, which compares each model with the combined one (although the difference is statistically significant only for the SARMA model). The combined model also exhibits the most accurate prediction interval coverage.

Table 3: In- and out-of-sample performance results for the intraday traffic model, given the several possible specifications of the daily traffic model. The D-M column compares a given model to the ‘full model’ in the first row on the MSE basis. The other performance statistics are explained in the text.

Model	RMSE (Persons)		MAPE (%)		Quantiles				D-M
	Train	Test	Train	Test	0.025	0.1	0.9	0.975	
Covariates + SARMA resid.	15.31	16.00	41.51	43.73	0.111	0.207	0.781	0.863	
Covariates only	17.03	16.14	46.53	45.81	0.083	0.185	0.808	0.889	<b>2.037 **</b>
SARMA residuals only	15.61	17.10	43.14	43.47	0.043	0.103	0.782	0.881	<b>5.073 ***</b>

### 3.3.5. Intraday Patterns

Our intraday traffic model takes a forecast for the total traffic within the day at a given store (obtained from the daily model), and *distributes* it across the store opening hours, given the explanatory variables. An appropriate model for this is the *multinomial regression* model, which assumes that a set of random variables is jointly multinomially distributed given explanatory variables. For time period  $t$  of a given day, let  $y_{d,t,j} = \beta_t' \mathbf{z}_{d,j}$ , where  $\beta_t$  is a set of regression coefficients and  $\mathbf{z}_{d,j}$  is a vector of day- and store-specific explanatory variables. The intraday probability attributed to interval  $t$  is

$$\mathbb{P}(t \mid \mathbf{z}_{d,j}) = \frac{\exp y_{d,t,j}}{\sum_{t'} \exp y_{d,t',j}}.$$

The coefficients  $\beta_t$  can be estimated by maximum likelihood. We then have

$$\mathbb{E}[\tilde{T}_{d,t,j} \mid \mathbf{x}_{d,j}, \mathbf{z}_{d,j}] = \mathbb{P}(t \mid \mathbf{z}_{d,j}) \mathbb{E}[\tilde{T}_{d,j} \mid \mathbf{x}_{d,j}],$$

with  $\mathbb{E}[\tilde{T}_{d,j} \mid \mathbf{x}_{d,j}]$  given in (4).

### 3.3.6. Comparison of Intraday Traffic Models

Goodness-of-fit results for the intraday traffic model (30-minute periods) are given in Table 3. The only difference between the tested models lies in the choice of daily model. We observe that the best daily model yields the best intraday forecasts, both in- and out-of-sample. The significance of the performance advantage is confirmed by the D-M test, whose results are more significant than for the daily models, largely because there are considerably more intraday test observations.

The interval coverage is narrower than the nominal intervals since none of the intraday models account for intraday autocorrelations in traffic patterns; these autocorrelations are indeed significant, but they do not impact significantly the *ex ante* traffic expectation for any given time period, which are the relevant quantities for making schedules.

### 3.4. Volume Modeling

At a typical small- to medium-sized retail store, the number of items sold during an intraday period (e.g., 30 minutes) will be a small integer. Due to the decomposition in (1), one needs to estimate the entire distribution of sales volume, not only the expected value. Empirically, we find that standard parametric forms, such as a conditional Poisson distribution, generally provides imperfect fits to the data distribution: the latter is significantly more dispersed and exhibits richer shapes than the former generally admit. We propose to adapt a more flexible framework, that of *ordinal regression* to estimate the volume distribution.

Ordinal regression models (McCullagh, 1980; McCullagh and Nelder, 1989) attempt to fit measurements that are observed on a categorical *ordinal* scale. Let  $Z \in \mathbb{R}$  be an unobserved variable and  $V \in \{1, \dots, K\}$  be defined by discretizing  $Z$  according to ordered cutoff points

$$-\infty = \zeta_0 < \zeta_1 < \dots < \zeta_K = \infty.$$

We observe  $V = k$  if and only if  $\zeta_{k-1} < Z \leq \zeta_k$ ,  $k = 1, \dots, K$ . The *proportional odds ordinal regression* model assumes that the cumulative distribution of  $V$  on the logistic scale is modeled by a linear combination of explanatory variables  $\mathbf{x}$ , i.e.

$$\text{logit } \mathbb{P}(V \leq k | \mathbf{x}) = \text{logit } \mathbb{P}(Z \leq \zeta_k | \mathbf{x}) = \zeta_k - \boldsymbol{\theta}' \mathbf{x},$$

where  $\boldsymbol{\theta}$  are regression coefficients and  $\text{logit}(p) \equiv \log(p/(1-p))$ . Model parameters  $\{\zeta_i, \boldsymbol{\theta}\}$  can be estimated by maximum likelihood (e.g., McCullagh 1980). Although ordinal regression is usually employed in contexts where discrete-valued observations only have ordering constraints (without explicit magnitude information), it is important to emphasize that in this work the ordered values are the non-negative integers themselves (up to  $K$ ): hence magnitude information is preserved, and ordinal regression can be used as a *bona fide* discrete distributional model, with observations spanning a finite range. Within this range, it provides more flexible distributions than typical discrete regression models (e.g. Poisson generalized linear models), while remaining reasonably easy to fit.

In our application,  $V = V_{d,t,j}$ , and we use the following explanatory (input) variables:

- Indicators for the store, the month, and day of week;
- Cubic B-spline basis of the traffic  $T_{d,t,j}$ , with 6 degrees of freedom, whose knots are placed automatically for a uniform coverage of the quantiles of the training data up to the specified number of degrees of freedom;
- Product (interaction) between the store indicators and the traffic B-spline traffic basis;
- Cubic B-spline basis of the number of employees working during the time period, with 6 degrees of freedom.

This provides a flexible distributional forecast of the sales volume. As mentioned earlier, to reduce the computations in the construction of the sales curves, we replace  $T_{d,t,j}$  in  $\mathbf{x}$  by its expectation. But to estimate the models parameters, we use the observed traffic.

Figure 5 illustrates two different operation points for a store that was part of our case study. We see that for low traffic and a single employee on the floor, most of the probability mass is concentrated towards selling zero items, whereas when the traffic or the number employees increases, the mass shifts rightwards, reflecting an increase in sales volume.

The prediction errors for this *ordinal regression* volume model are in Table 3.4. For comparison, we report the corresponding results for a Poisson generalized linear model (GLM; e.g., McCullagh and Nelder (1989)) and a negative binomial (NB) GLM (Hilbe, 2011) with the same input variables. These models are frequently used in integer-valued observations models, with the negative binomial distribution allowing to represent “overdispersion” (variance greater than the mean) of the response with respect to the Poisson. For these models, the predictive distribution is respectively a Poisson or NB with the log-expectation given by a linear mapping between the input variables and the regression coefficients.

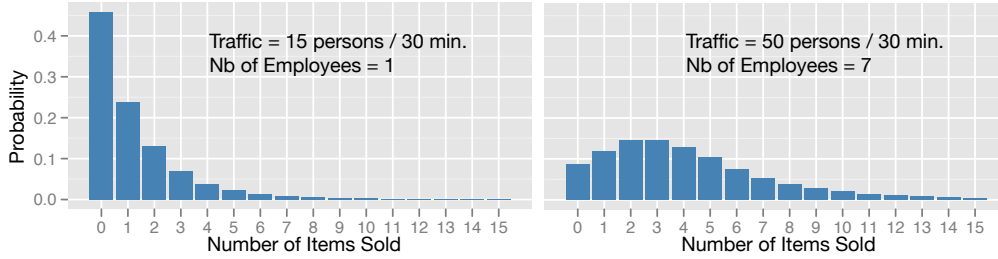


Figure 5: Predictive distribution of the number of items sold during a period of 30 minutes at a retail store under the intraday sales volume model, at two operation points. *Left*: With a single employee in the store, and relatively low store traffic, the most likely number of items sold during a 30-minute period is zero (with 45% probability). *Right*: With seven employees and higher traffic, the distribution shifts rightward, exhibiting a mode between two and three items sold.

Table 4: In- and out-of-sample performance results for several specifications of the ‘volume’ (number of items sold) model. The D-M column compares a given model to the ‘full model’ in the first row on the NLL basis. The other performance statistics are explained in the text.

Model	RMSE (Nb Items)		MAPE (%)		NLL	D-M	
	Train	Test	Train	Test	Test		
Ordinal Regression	4.172	5.222	47.554	46.447	2.015		
Poisson GLM	3.932	4.887	48.372	48.112	2.417	<b>30.600</b>	***
Negative Binomial GLM	3.823	4.722	48.326	47.888	2.021	<b>2.550</b>	**

From Table 3.4, the Poisson and NB models produce better estimates of the expected number of items sold than the ordinal regression model. However, the latter—owing to its greater flexibility—is better at assigning probability mass in the ‘right places’, as ascertained by the NLL criterion. The D-M test statistic confirms the significance of the difference. This is of consequence in practice inasmuch as the complete probability distribution for the number of items sold is necessary in (1) to compute the conditional expectation of sales.

### 3.5. Item Price Modeling

Our item-price model takes the simple log-linear form,  $\log P_{d,j} = \beta_p' \mathbf{w}_{d,j} + \eta_{d,j}$ , where  $\beta_p$  is a vector of regression coefficients (a single vector for all the data),  $\mathbf{w}_{d,j}$  is a vector of item-price explanatory variables at day  $d$  for store  $j$ , and  $\eta_{d,j}$  is a residual. Two specifications of this model are compared: a linear and a log-linear form. They both use of the following explanatory variables:

- Indicators for the store, the month of the year, and interactions thereof;
- Cubic B-spline basis of the number of items sold, with 8 knots placed on a power scale between 1 and 30 (so that small integers are more finely covered than larger ones), impacting the average item price non-linearly.

The number of working employees was not found to be a significant determinant of the average item price.

Results are given in Table 5. The log-linear model performs slightly but significantly better than the purely linear one. It also has the advantage of never allowing negative prices to occur.



Table 5: In- and out-of-sample performance results for two specifications of the average item price model. The D-M column compares a second to the first model on the MSE basis. The other performance statistics are explained in the text.

Model	RMSE (\$)		MAPE (%)		Quantiles				D-M
	Train	Test	Train	Test	0.025	0.1	0.9	0.975	
Log-linear	38.94	34.05	26.69	28.15	0.037	0.094	0.943	0.986	
Linear	38.83	34.21	26.94	28.41	0.010	0.054	0.951	0.983	<b>5.954 ***</b>

If product returns occur frequently, it would make sense to handle them separately, perhaps by allowing negative quantities in the volume distribution of the previous section. In our data, product returns were very infrequent and were ignored.

### 3.6. Comparison with Other Models

We compare in Table 6 the sale curves model based on our proposed decomposition (denoted *Decomposition (full expectation)*) to two alternatives:

1. A model based on the same decomposition, but which treats the volume and average item price as independent quantities. This amounts to making the following approximation in (1):

$$\begin{aligned}
 r_{d,t,j}(y) &= \mathbb{E}[\mathbb{E}[V_{d,t,j}P_{d,t,j} \mid \mathbb{E}[\tilde{T}_{d,t,j}], y_{d,t,j}^w = y]] \\
 &\approx \mathbb{E}[\mathbb{E}[V_{d,t,j} \mid \mathbb{E}[\tilde{T}_{d,t,j}], y_{d,t,j}^w = y]] \mathbb{E}[\mathbb{E}[P_{d,t,j} \mid \mathbb{E}[\tilde{T}_{d,t,j}], y_{d,t,j}^w = y]].
 \end{aligned}$$

This is referred to as *Decomposition (independence assumption)* in the results table.

2. A model that attempts to *directly* model intraday sales in log scale. It is based on the same daily/intraday break-down that is used for traffic modeling, but using the log of sales as the target quantity instead of traffic. In addition, the intraday model makes use of the following variables:
  - Indicators for the store, month, day of the week, and time period;
  - Cubic B-spline basis for the traffic, with 6 degrees of freedom;
  - Interaction terms between the store and the traffic B-spline basis;
  - Cubic B-spline basis for the number  $y_{d,t,j}^w$  of working employees during the time period, with 6 degrees of freedom.

Overall, the direct model has about the same number of free parameters as all the models that are part of the decomposition. It is named *Direct* in the table.

On an out-of-sample basis, the model based on the decomposition, with the full expectation computation, is seen to outperform both the expectation approximation, and much more significantly the direct model. Our subsequent results are based on this model.

Table 6: In- and out-of-sample performance results for revenue models: the first based on the proposed decomposition, and the second based on a direct approach. The D-M column compares a second to the first model on the MSE basis. The other performance statistics are explained in the text.

Model	RMSE (\$)		MAPE (%)		D-M
	Train	Test	Train	Test	
Decomposition (full expectation)	289.80	306.31	67.75	69.04	
Decomposition (independence assumption)	286.92	306.86	67.55	69.53	<b>1.8526 *</b>
Direct	441.55	398.48	105.98	99.19	<b>13.8742 ***</b>

## 4. Scheduling Models

The traditional cost-driven interpretation of shift scheduling problems led to a variety of classical models and formulations. Recent innovative formulations based on concepts from language theory enable one to efficiently represent the shifts that can be assigned to employees. In this section, we first review the principles of these problem formulations, before extending them to our specific profit-based formulation. This extension is achieved by linearizing the objective function and formulating the problem of scheduling for profit as a MIP. We also propose another formulation of the same problem as a CP, which has the advantage of handling the non-linearity in a more natural way. We consider a single store, so we can drop the index  $j$ .

### 4.1. Traditional Models for Cost-Driven Scheduling

We are given a set  $E$  of employees, a set  $T$  of time periods of equal length, a set of activities  $A = \{\mathbf{w}, \mathbf{r}, \mathbf{m}, \mathbf{b}\}$  (which stand for  $\{work, meal, break, rest\}$ ), and a set  $\Omega$  of feasible shifts, each one defined by a finite sequence of activities from  $A$ .

In the classical, cost-driven, shift scheduling problem, a minimal number  $b_t^{\mathbf{w}}$  of employees that perform activity  $\mathbf{w} \in A$  during period  $t$  is specified for each  $t \in T$ , and the goal is to select a multiset of shifts in  $\Omega$  (with multiple selection of the same shift allowed) that cover the personnel requirement at a minimum cost. A set covering formulation of this problem was already formulated by Dantzig (1954) for the toll-booth operator shift scheduling model of Edie (1954). In this model, each shift  $o \in \Omega$  has a cost  $C_o$  and is defined by listing explicitly the activities assigned for each time period. The cost  $C_o$  is often assumed to be additive, i.e., it can be written as a sum of costs over the periods that it covers:

$$C_o = \sum_{a \in A} \sum_{t \in T} \alpha_{t,o}^a C_t^a,$$

where  $\alpha_{t,o}^a = 1$  if shift  $o$  assigns activity  $a$  in period  $t$ ,  $\alpha_{t,o}^a = 0$  otherwise, and  $C_t^a$  is a cost for doing activity  $a$  in period  $t$ . In real-life situations, the set  $\Omega$  is typically very large and the corresponding MIP becomes very hard to solve directly, even via column generation techniques as in Mehrotra et al. (2000), Vatri (2001), Bouchard (2004), and Demassej et al. (2006).

Alternative mathematical reformulations have also been proposed to reduce the number of variables. The most popular ones are so-called *implicit formulations* (Bechtolds and Jacobs 1990, Aykin 1996, Rekik et al. 2004, Rekik et al. 2005), which provide an interesting alternative to the model of Dantzig (1954). The qualifier *implicit* comes from the fact that these models do not assign breaks to shift a priori. They rather introduce the notion of shift types, which are characterized only by starting and ending times and do not contain any details about what

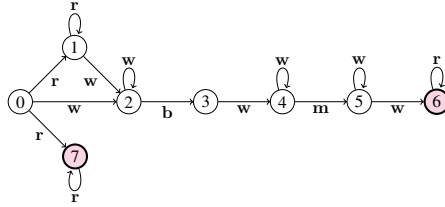


Figure 6: A DFA  $\Pi$  with each state represented as a *circle*, the two final states are shaded, and each transition as an *arc*.

happens in between. Typically, these models capture the number of employees assigned to each shift type and to each break with different sets of variables. From an optimal solution to such an implicit model, one can retrieve the number of employees assigned to each shift type and each break, and construct an optimal set of shifts through a polynomial-time procedure.

More recently, Côté et al. (2011a,b) showed how to exploit the expressiveness of automata and context-free grammars to formulate MIP models in which any feasible shift is represented by a word in a regular language, and the set of admissible staffing solutions (which correspond to an admissible collection of shifts) is defined implicitly by an integral polyhedron that can be constructed automatically even for complex work regulations such as those mentioned earlier. In what follows, we will adapt this approach to our profit-maximizing problem. Our use of automata together with CP for this type of scheduling problem is novel.

#### 4.2. Regular Language and Workforce Scheduling

Following Côté et al. (2011a,b), we will represent each admissible shift by a word defined over the alphabet  $A = \{w, r, m, b\}$ , where the  $i$ th letter in the word indicates the *activity* of the employee during the  $i$ th time period of its shift. In our case study, those time periods have 15 minutes. As an illustration, the word “*wwbwwmmwwww*” symbolizes a 3-hour shift where the employee works for 30 minutes, takes a 15 minutes break, gets back to work for 45 minutes, takes a lunch break of 30 minutes, and goes back to work for the remaining hour.

To recognize the valid words, i.e., those that correspond to admissible shifts, we define a *deterministic finite automaton* (DFA) that specifies the rules that regulate the transitions in the sequence of activities that can be assigned to an employee in a shift. The DFA is defined as a 5-tuple  $\Pi = \langle Q, A, \tau, q_0, F \rangle$ , where

- $Q$  is the finite set of states;
- $A$  is the alphabet;
- $\tau : Q \times A \rightarrow Q$  is the transition function;
- $q_0 \in Q$  is the initial state;
- and  $F \subseteq Q$  is a set of final states.

A word is *recognized* by a DFA if, by processing its symbols one by one from the initial state and following the transitions, we end up in a final state after processing the last symbol.

Figure 6 illustrates a DFA  $\Pi$  that recognizes a language defined over the alphabet  $A$ . It has  $Q = \{0, 1, \dots, 6\}$ ,  $q_0 = 0$  and  $F = \{5, 6\}$ . This language contains, for instance, the words “*wwbwwmmww*”, “*rr*” and “*r*”, but it does not include the words “*wwr*” or “*wwbm*” as they are not recognized by  $\Pi$ .



beginning or at the end of the shift.

Let

$$\pi_{q,q'}^a = \begin{cases} 1, & \text{if the transition function } \tau \text{ allows a transition from state } q \in Q \\ & \text{to state } q' \in Q \text{ with label } a \in A \text{ in the automaton } \Pi, \\ 0, & \text{otherwise.} \end{cases}$$

A word of length  $\ell$  is included in the language defined by  $\Pi$  if and only if its corresponding flow variables satisfy the following constraints:

$$\sum_{a \in A} \sum_{q \in Q} s_{q_0,a,q}^1 = 1 \quad (6)$$

$$\sum_{a \in A} \sum_{q' \in Q} s_{q',a,q}^{t-1} = \sum_{a \in A} \sum_{q' \in Q} s_{q,a,q'}^t \quad \forall t \in \{2, \dots, \ell\}, q \in Q \quad (7)$$

$$\sum_{a \in A} \sum_{q \in Q} \sum_{f \in F} s_{q,a,f}^\ell = 1 \quad (8)$$

$$s_{q,a,q'}^t \leq \pi_{q,q'}^a \quad \forall t \in \{1, \dots, \ell\}, a \in A, (q, q') \in Q^2 \quad (9)$$

$$\sum_{(q,q') \in Q^2} s_{q,a,q'}^t = x_t^a \quad \forall t \in \{1, \dots, \ell\}, a \in A \quad (10)$$

$$s_{q,a,q'}^t \in \{0, 1\} \quad \forall t \in \{1, \dots, \ell\}, a \in A, (q, q') \in Q^2. \quad (11)$$

Constraints (6)–(8) ensure that one unit of flow enters the graph by the source and leaves it by a final state, and that the amount of flow entering and leaving each other node in the graph is the same. Constraints (9) guarantee that the transition function  $\tau$  of  $\Pi$  is respected, whereas constraints (10) link the decision variables  $x_t^a$  and the flow variables. Finally, constraints (11) define the domain of the flow variables.

#### 4.3. Mixed Integer Programming for Schedule Construction

We are now ready to formulate our profit-driven workforce scheduling problem as a MIP by building over the network flow formulation of constraints on feasible shifts as given in the previous subsection. We consider a single store  $j$  at a time, so we drop the index  $j$ . We use  $E$  to denote the set of employees,  $D$  for the set of days for which we want to construct the schedule (e.g., Monday to Sunday if it is over a week, or a single element if it is over a single day), and  $T_d$  for the set of time periods when the store is open on day  $d$ . Let  $M = \{0, 1, \dots, |E|\}$  (the possible number of employees that can perform a given task during a given period) and  $\bar{A} = \{\mathbf{w}, \mathbf{b}, \mathbf{m}\}$  (the non-rest activities, which are the only ones considered to determine the real length of a shift). We introduce two sets of binary variables:  $x_{d,t}^{e,a}$  indicates if employee  $e$  performs task  $a$  during period  $t$  on day  $d$ , while  $y_{d,t}^{m,a}$  indicates if exactly  $m$  employees perform task  $a$  during period  $t$  on the day  $d$ . Each flow variable  $s_{q,a,q'}^{e,d,t}$  indicates whether employee  $e$  is transiting from state  $q$

to state  $q'$  (of  $\Pi$ ) during period  $t$  of day  $d$  by performing task  $a$ . Our MIP formulation is:

$$\max \sum_{m \in M} \sum_{d \in D} \sum_{t \in T_d} \left( r_{d,t}(m) y_{d,t}^{m,w} - \sum_{a \in A} m C_{d,t}^a y_{d,t}^{m,a} \right) \quad (12)$$

subject to

$$\sum_{a \in A} \sum_{q \in Q} s_{q_0,a,q}^{e,d,1} = 1 \quad \forall e \in E, d \in D \quad (13)$$

$$\sum_{a \in A} \sum_{q' \in Q} s_{q',a,q}^{e,d,t-1} = \sum_{a \in A} \sum_{q' \in Q} s_{q,a,q'}^{e,d,t} \quad \forall e \in E, d \in D, t \in \{2, \dots, |T_d|\}, q \in Q \quad (14)$$

$$\sum_{a \in A} \sum_{q \in Q} \sum_{f \in F} s_{q,a,f}^{e,d,|T_d|} = 1 \quad \forall e \in E, d \in D \quad (15)$$

$$s_{q,a,q'}^{e,d,t} \leq \pi_{q,q'}^a \quad \forall e \in E, d \in D, t \in T_d, a \in A, (q, q') \in Q^2 \quad (16)$$

$$\sum_{(q,q') \in Q^2} s_{q,a,q'}^{e,d,t} = x_{d,t}^{e,a} \quad \forall e \in E, d \in D, t \in T_d, a \in A \quad (17)$$

$$\sum_{e \in E} x_{d,t}^{e,a} = \sum_{m \in M} m y_{d,t}^{m,a} \quad \forall d \in D, t \in T_d, a \in A \quad (18)$$

$$\sum_{a \in A} x_{d,t}^{e,a} = 1 \quad \forall e \in E, d \in D, t \in T_d \quad (19)$$

$$\sum_{m \in M} y_{d,t}^{m,a} = 1 \quad \forall d \in D, t \in T_d, a \in A \quad (20)$$

$$\delta_e^- \leq \sum_{t \in T_d} \sum_{a \in \bar{A}} x_{d,t}^{e,a} \leq \delta_e^+ \quad \forall e \in E, d \in D \quad (21)$$

$$\Delta_e^- \leq \sum_{d \in D} \sum_{t \in T_d} \sum_{a \in \bar{A}} x_{d,t}^{e,a} \leq \Delta_e^+ \quad \forall e \in E \quad (22)$$

$$s_{q,a,q'}^{e,d,t} \in \{0, 1\} \quad \forall e \in E, d \in D, t \in T_d, a \in A, (q, q') \in Q^2 \quad (23)$$

$$x_{d,t}^{e,a} \in \{0, 1\} \quad \forall e \in E, d \in D, t \in T_d, a \in A \quad (24)$$

$$y_{d,t}^{m,a} \in \{0, 1\} \quad \forall m \in M, d \in D, t \in T_d, a \in A \quad (25)$$

In the objective function, we find the sales curves  $r_{d,t}(\cdot)$  defined in (1) (where we have dropped the store index  $j$ ), which give the expected revenue from sales in each time period as a function of the number of working employees, and the total salary expenses, where  $C_{d,t}^a$  is the per-employee cost for performing activity  $a$  during period  $t$  on day  $d$ . We want to maximize the total expected profit. We have used binary variables  $y_{d,t}^{m,a} \in \{0, 1\}$  instead of integer variables  $y_{d,t}^a \in M$  to represent the number of employees performing each activity in each time period because it provides a linear MIP formulation of the problem, despite the nonlinear sales curves in the

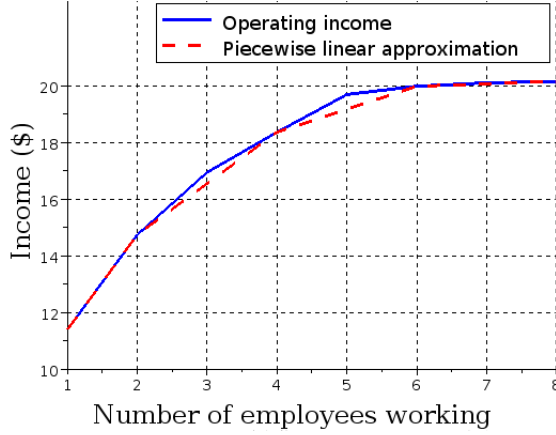


Figure 8: A sales curve defined over  $\{1, \dots, 8\}$  and its PLA with  $p = 4$  pieces separated at 2, 4, and 6.

objective function. Constraints (13)–(17) guarantee that all shifts assigned to employees on day  $d$  correspond to words of length  $|T_d|$  in the language defined by  $\Pi$ , by deriving a flow problem over the variables  $s_{q',a,q}^{e,d,t}$  and linking them to the variables  $x_{d,t}^{e,a}$  (as in subsection 4.2). Constraints (18) link the variables  $x_{d,t}^{e,a}$  and  $y_{d,t}^{m,a}$ , whereas constraints (19) guarantee that a single task  $a$  is assigned to each employee  $e$  for every period  $t$  of a day  $d$ , and constraints (20) ensure that  $y_{d,t}^{m,a}$  captures the number of employees performing task  $a$  in period  $t$  of day  $d$ . Constraints (21)–(22) impose a minimum duration of  $\delta_e^-$  and a maximum duration of  $\delta_e^+$  to the shift length of each employee (excluding the r activity) over any given day  $d$ , as well as a minimum duration of  $\Delta_e^-$  and maximum duration of  $\Delta_e^+$  for the total shift duration over all days covered by the schedule (e.g., over the week). Finally, constraints (23)–(25) define the domain of all variables.

#### 4.4. Estimation of the Revenues by Piecewise Linear Approximations (PLA)

The linearization of the objective function in our MIP formulation has introduced a huge number of binary variables  $y_{d,t}^{m,a}$ , which increases significantly the size of the problem. One way to reduce this size is to approximate the sales curves by piecewise linear functions. If  $p \geq 1$  denotes the number of linear pieces in the approximation, the PLA formulation, which we denote by MIP-PLA $_p$ , use  $p$  binary variables to indicate the piece in which we are, and an integer variable  $y_{d,t}^a \geq 0$  to indicate the number of employees performing activity  $a$  in period  $t$  of day  $d$ . Typically, the sales curves are concave, in which case the PLA will underestimate the expected revenue (see Figure 8). When  $p = |E| - 1$ , MIP-PLA $_p$  is equivalent to the original MIP formulation, whereas if  $p < |E| - 1$ , MIP-PLA $_p$  is an estimation of the MIP, so the optimal solution of MIP-PLA $_p$  might not be optimal for MIP. However, any feasible solution to MIP-PLA $_p$  is feasible for MIP, because the PLA does not modify the constraints.

#### 4.5. Constraint Programming for Schedule Construction

We now consider another efficient approach, based on a problem formulation as a CP. This formulation handles the non-linearity naturally without altering the model by using approxima-

tions and without adding extra variables. Our CP formulation is based on three sets of variables:  $x_{d,t}^e$  specifies the *activity*  $a \in A$  performed by employee  $e$  in period  $t$  of day  $d$ ,  $y_{d,t}^a$  represents the *number* of employees that perform activity  $a$  in period  $t$  of day  $d$ , and  $s_{d,t}^e \in Q$  describes the state of the automaton that led to the assignment of the activity performed by employee  $e$  in period  $t$  of day  $d$ . We represent a transition from state  $q$  to state  $q'$  labeled by  $a$  by the triplet  $\langle q, a, q' \rangle$ , and we define  $\Theta$  as the set of triplets  $\langle q, a, q' \rangle$  that correspond to transitions allowed under the DFA  $\Pi$ . Before stating the CP model, we recall to the reader that in Constraint Programming the expression “(variable = value)”, when used within a constraint, is automatically transformed (known as *reification*) into a boolean variable that takes value 1 when the expression holds and 0 otherwise. The proposed CP model is then:

$$\max \sum_{d \in D} \sum_{t \in T_d} \left( r_{d,t}(y_{d,t}^{\mathbf{w}}) - \sum_{a \in A} C_{d,t}^a y_{d,t}^a \right) \quad (26)$$

subject to

$$s_{d,0}^e = q_0 \quad \forall e \in E, d \in D \quad (27)$$

$$\langle s_{d,t-1}^e, x_{d,t}^e, s_{d,t}^e \rangle \in \Theta \quad \forall e \in E, d \in D, t \in \{1, \dots, |T_d|\} \quad (28)$$

$$s_{d,|T_d|}^e \in F \quad \forall e \in E, d \in D \quad (29)$$

$$\sum_{e \in E} (x_{d,t}^e = a) = y_{d,t}^a \quad \forall d \in D, t \in T_d, a \in A \quad (30)$$

$$\delta_e^- \leq \sum_{t \in T_d} (x_{d,t}^e \neq \mathbf{r}) \leq \delta_e^+ \quad \forall e \in E, d \in D \quad (31)$$

$$\Delta_e^- \leq \sum_{d \in D} \sum_{t \in T_d} (x_{d,t}^e \neq \mathbf{r}) \leq \Delta_e^+ \quad \forall e \in E \quad (32)$$

$$s_{d,t}^e \in Q \quad \forall e \in E, d \in D, t \in T_d \quad (33)$$

$$x_{d,t}^e \in A \quad \forall e \in E, d \in D, t \in T_d \quad (34)$$

$$y_{d,t}^a \in M \quad \forall a \in A, d \in D, t \in T_d \quad (35)$$

The objective function in (26) is the same as in the MIP, but is expressed more concisely, as it can use directly the integer variables  $y_{d,t}^{\mathbf{w}}$ . The constraints (27)–(29) ensure that the shift performed by any employee  $e$  over a day  $d$  is recognized as a word of length  $|T_d|$  in the language defined by  $\Pi$ . Constraints (27) and (29) ensure that the exploration of the automaton begins at its initial state and ends in a final state (in  $F$ ), whereas constraints (28) guarantee that the transition function  $\tau$  of the DFA is respected when moving from state to state. Constraints (30) link the  $x_{d,t}^e$  and  $y_{d,t}^a$  variables. Constraints (31)–(32) guarantee that the length of each employee’s shift respects the minimum and maximum durations defined over a day and over a week, respectively. Finally, constraints (33)–(35) define the domain of all variables. In constraint programming terminology, constraints (28) are known as *table* constraints and constraints (30) as *global cardinality* constraints. Most constraint programming solvers can handle these families of constraints using special-purpose algorithms. This CP is exactly equivalent to the MIP introduced earlier. However, the two formulations give rise to different solutions methods, whose performance can be very different. In particular, the large number of binary variables in the MIP can have a significant impact on the speed of resolution.



## 5. Solving the Scheduling Problem for the Case Study: Numerical Comparisons

We have implemented the MIP and CP formulations of Section 4 for our case study, and solved them using the ILOG CPLEX optimization tool on 2.1 GHz dual core processors with 3.5 GB of memory. In this section, we demonstrate and compare the accuracy and the efficiency of the proposed approaches to construct schedules for one day and one week, while considering the work regulation rules of the retailer. Note that a new labor regulation impacting scheduling was put in place by the retailer in all its stores after the period for which we have data. Although the regulations defined over a day did not change, the new regulations impose a new set of rules common to all stores, over a week. Before the change, each store manager was constructing weekly schedules according to his own set of rules, which could potentially change between weeks. In this sense, we do not have full access to the exact weekly scheduling rules for the period studied, and therefore a fair comparison between our schedules and the schedules historically used in the stores can only be performed over a day. However, our models integrate the new regulation of the retailer and the performance of the schedules they provide for each week can still be measured from the gap between the expected profit they generate and an accurate upper bound of the maximum expected profit that can be generated over a week while respecting the regulations.

### 5.1. Work regulation rules and other information

In addition to the generic constraints given in our MIP or CP formulation, there are other daily and weekly constraints specific to our retailer. In their formulation, two types of employees are distinguished: part-time and full-time.

- Daily regulations:
  - *Work and rest distribution*: each day, the employee can have a day off or work for one shift;
  - *Shift duration*: a shift over a day lasts from 3 to 8 hours;
  - *Break and meal distribution*: non-working stretches (break, meal, rest) have to be separated by at least one hour of work;
  - *Break and meal authorization*: an employee working a shift lasting less than 5 hours must take a single 15-minute break, whereas an employee working a shift of 5 hours or more must take two 15-minute breaks, separated by a 30-minute meal break.
  - *Minimum presence*:
    - \* Stores with external back stores must have a minimum of two employees at all times during opening hours, else a minimum of one employee is required;
    - \* Part-time employees must never work alone, a full-time employee must always be present..
- Weekly regulations:
  - *Work and rest distribution*
    - \* An employee cannot work more than 5 shifts in a week;
    - \* An employee must have two consecutive days off in a week or one week-end day off;

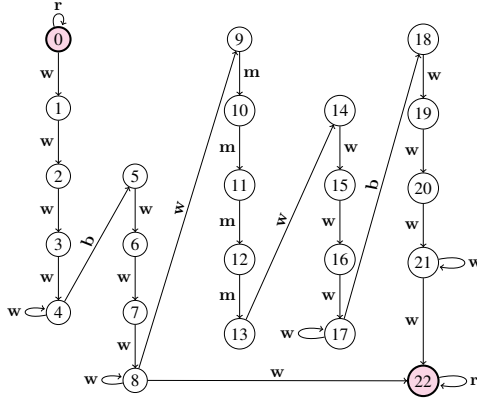


Figure 9: Illustration of the DFA II that defines the formal language used by our model.

– *Shift duration*

- \* A part-time employee must work between 10 hours and 30 hours over a week;
- \* A full-time employee must work between 25 hours and 40 hours over a week.

We used data coming from 15 stores, for 5089 days in total (the same portion of data that was used to evaluate the forecasting models). In every store, half the workforce are full-time employees, and we must have at least three employees with a full-time contract. These stores hire between 4 and 9 employees, and their opening hours vary from 9 hours to 12 hours per day. Table 7 gives additional details on these values. Only stores 14 and 15 have an external back store. Each day is split into 15-minute intervals. All employees earn the same hourly salary, from which the costs  $C_t^a$  are extracted. Breaks and meals are not paid. Note that, although the number of employees of each kind (partial time vs full time) is fixed, the amount of time spent working on the floor over a day and a week by each of them is only constrained to be inside a specific range, namely  $[\delta_e^-, \delta_e^+]$  for a day and  $[\Delta_e^-, \Delta_e^+]$  for a week for every employee  $e \in E$ , so the total cost of employees over a day or a week has to be determined during the optimization process.

5.2. *Automaton for retail store scheduling*

Figure 9 illustrates the automaton II that defines the formal language for our model. This automaton is used to construct a word for each employee, that represents his admissible daily shift, in which the sequence of activities he performs obeys all the sequencing rules (e.g., an employee must work at least one hour before his first break, he must work for a shift of at least five hours to be allowed to take two breaks, etc.). The other regulation rules are incorporated using traditional MIP and CP constraints based on the number of occurrences of each activity in the word of each employee or the states of the automaton explored by an employee.

To formulate these extra constraints, we use  $\bar{E}$  to denote the set of full-time employees,  $\bar{D}$  for the set of the first four days of the week (Monday to Thursday),  $\hat{D}$  for the set of weekend days (Saturday and Sunday), and  $q_1$  to denote state 1 of the automaton of Figure 9. For the extra constraints in the MIP formulation, we introduce two new sets of binary variables:  $v_d^e$  indicates

if employee  $e$  is off for both day  $d$  and the following day, while  $z^e$  indicates if employee  $e$  has a day off over the weekend. The extra constraints for the MIP are:

$$\delta_e^- \sum_{t \in T_d} s_{q_0, \mathbf{w}, q_1}^{e, d, t} \leq \sum_{t \in T_d} \sum_{a \in \bar{A}} x_{d, t}^{e, a} \leq \delta_e^+ \sum_{t \in T_d} s_{q_0, \mathbf{w}, q_1}^{e, d, t} \quad \forall e \in E, d \in D \quad (36)$$

$$\sum_{t \in T_d} y_{d, t}^{0, \mathbf{w}} = 0 \quad \forall d \in D \quad (37)$$

$$\sum_{e \in \bar{E}} x_{d, t}^{e, \mathbf{r}} \leq |\bar{E}| - 1 \quad \forall d \in D, t \in T_d \quad (38)$$

$$\sum_{d \in D} s_{q_0, \mathbf{r}, q_0}^{e, d, |T_d|} \geq 2 \quad \forall e \in E, d \in D \quad (39)$$

$$2v_d^e \leq s_{q_0, \mathbf{r}, q_0}^{e, d, |T_d|} + s_{q_0, \mathbf{r}, q_0}^{e, d+1, |T_{d+1}|} \quad \forall e \in E, d \in \bar{D} \quad (40)$$

$$z^e \leq \sum_{d \in \bar{D}} s_{q_0, \mathbf{r}, q_0}^{e, d, |T_d|} \quad \forall e \in E \quad (41)$$

$$z^e + \sum_{d \in \bar{D}} v_d^e \geq 1 \quad \forall e \in E \quad (42)$$

$$v_d^e \in \{0, 1\} \quad \forall e \in E, d \in \bar{D} \quad (43)$$

$$z^e \in \{0, 1\} \quad \forall e \in E \quad (44)$$

while the extra constraints to be added to the CP are:

$$\delta_e^- \sum_{t \in T_d} (s_{d, t}^e = q_1) \leq \sum_{t \in T_d} (x_{d, t}^e \neq \mathbf{r}) \leq \delta_e^+ \sum_{t \in T_d} (s_{d, t}^e = q_1) \quad \forall e \in E, d \in D \quad (45)$$

$$y_{d, t}^{\mathbf{w}} \geq 1 \quad \forall d \in D, t \in T_d \quad (46)$$

$$\sum_{e \in \bar{E}} (x_{d, t}^e \neq \mathbf{r}) \geq 1 \quad \forall d \in D, t \in T_d \quad (47)$$

$$\sum_{d \in D} (s_{d, |T_d|}^e = q_0) \geq 2 \quad \forall e \in E, d \in D \quad (48)$$

$$\left( \sum_{d \in \bar{D}} \left( \left( \sum_{t \in T_d} (x_{d, t}^e = \mathbf{w}) = 0 \right) \wedge \left( \sum_{t \in T_{d+1}} (x_{d+1, t}^e = \mathbf{w}) = 0 \right) \right) \geq 1 \right) \vee \left( \sum_{d \in \bar{D}} \left( \sum_{t \in T_d} (x_{d, t}^e = \mathbf{w}) = 0 \right) \geq 1 \right) \quad \forall e \in E. \quad (49)$$

The constraints (36) and (45) are reformulation of the generic constraints (21) and (31) that still guarantee that the length of each employee's shift respects the minimum and maximum durations defined over a day, while adding the possibility that an employee does not work over a day (i.e., the employee has a day off). The constraints (37) and (46) ensure that there is at least one

Table 7: Estimated profit gains and gap to best upper bound for the solutions obtained by solving the proposed MIP and CP models, after 30 seconds and after 300 seconds of CPU time.

Store	Instances description			After 30 sec.				After 300 sec.			
				Profit growth (%)		Gap (%)		Profit growth (%)		Gap (%)	
	#Days	#Periods	#Empl.	MIP	CP	MIP	CP	MIP	CP	MIP	CP
1	408	33.8	6	-	3.4	-	0.3	3.5	3.5	0.2	0.2
2	326	32.4	5	-	2.8	-	0.2	2.6	2.6	0.1	0.1
3	372	40.6	5	-	2.6	-	0.2	2.6	2.6	0.1	0.1
4	377	36.7	5	-	2.9	-	0.2	3.0	3.0	0.1	0.1
5	309	40.4	5	-	2.6	-	0.2	2.7	2.7	0.1	0.1
6	204	32.0	6	-	4.0	-	0.2	3.4	4.2	0.8	0.0
7	399	37.5	5	-	2.7	-	0.2	2.8	2.8	0.1	0.1
8	384	41.7	6	-	3.0	-	0.3	2.0	3.2	1.4	0.2
9	241	33.0	5	-	2.9	-	0.2	3.0	3.1	0.1	0.0
10	401	33.4	4	-	1.3	-	0.0	1.3	1.3	0.0	0.0
11	369	42.0	5	-	2.1	-	0.2	2.2	2.2	0.2	0.2
12	381	40.3	5	-	2.2	-	0.1	2.2	2.2	0.1	0.0
13	366	41.4	5	-	2.2	-	0.1	2.1	2.3	0.2	0.1
14	368	42.6	8	-	2.9	-	0.5	2.8	3.2	0.6	0.2
15	184	42.1	9	-	2.7	-	0.8	2.4	3.2	1.1	0.3
<b>Average</b>	<b>339.3</b>	<b>38.0</b>	<b>5.6</b>	-	<b>2.7</b>	-	<b>0.2</b>	<b>2.6</b>	<b>2.8</b>	<b>0.3</b>	<b>0.1</b>

employee working on the floor at any time of the day. Note that if the store have an external back store, these constraints must be replaced by the constraints (50) and (51) given by:

$$\sum_{t \in T_d} y_{d,t}^{0,w} + y_{d,t}^{1,w} = 0 \quad \forall d \in D \quad (50)$$

and

$$y_{d,t}^w \geq 2 \quad \forall d \in D, t \in T_d, \quad (51)$$

respectively. The constraints (38) and (47) guarantee that there is at least one full-time employee working at any time. Constraints (39) and (48) ensure that an employee cannot work more than five shifts (days) over a week. The days off are managed differently in the MIP and CP; the MIP requires the introduction of new variables. For the MIP, constraints (40) and (41) are added to link the new variables  $v_d^e$  and  $z^e$  to the variables  $s_{q,a,q'}^{e,d,t}$ , respectively, so the constraints (42) can guarantee that an employee has two consecutive days off during the week or one weekend day off. Constraints (43) and (44) must also be added to define the domain of variables  $v_d^e$  and  $z^e$ . For CP, only the constraints (49) must be added, as they directly ensure that an employee has two consecutive days off in a week or one weekend day off.

### 5.3. Scheduling over a day

By taking the set  $D$  as a single element, the proposed models can be used to construct work schedules over a day. In this context, the weekly regulations are ignored.

Table 7 compares the efficiency gains provided by solving our models, over the scheduling solutions used by the retailer, over a single day. These numbers are averages over the 5089 days of our data. The table gives two types of informations: (i) the estimated increase in expected profit with the schedules constructed by MIP after 300 seconds of CPU time, and by CP after

Table 8: Evolution of the gap of the solutions identified by CP when building schedules over a week with two hours of CPU time, according to the best upper bound obtained after 24 hours.

Store	Instances description			Gap with upper bound obtained from MIP (in %)		
	#Weeks	#Periods	#Empl.	after 30 min	after 60 min	after 120 min
1	37	236	6	2.2	2.0	1.6
2	23	224	5	1.9	1.7	1.5
3	21	282	5	3.8	3.6	3.4
4	34	256	5	2.5	2.3	2.1
5	17	280	5	4.4	4.0	3.6
6	15	232	6	2.8	2.5	2.4
7	38	262	5	3.1	2.7	2.5
8	38	288	6	3.4	3.3	3.1
9	14	230	5	2.0	1.8	1.6
10	36	234	4	2.0	1.9	1.7
11	22	294	5	3.4	3.3	3.2
12	23	282	5	2.9	2.7	2.6
13	22	288	5	3.0	2.8	2.7
14	18	296	8	1.9	1.7	1.7
15	12	292	9	8.3	7.2	6.9
<b>Average</b>	<b>24.7</b>	<b>265.1</b>	<b>5.6</b>	<b>3.2</b>	<b>2.9</b>	<b>2.7</b>

30 seconds and 300 seconds of CPU time, compared with the schedules used by the retailer; and (ii) the gap between those MIP and CP solutions and the best upper bound extracted from the MIP solution after 300 seconds. As the solutions provided by solving MIP with PLA generate less profit than the solutions of the full MIP while using the same amount of CPU time to construct feasible schedules, we always used the full MIP (no PLA) in this section. For many instances, the MIP failed to return a feasible solution after 30 seconds; for this reason we do not report the results. The solutions returned by either MIP or CP after 300 seconds are very accurate. They are nearly optimal with a respective average gap of only 0.3% and 0.1%, and correspond to estimated profit increases of 2.6% and 2.8%, respectively (in terms of the estimated expected sales given by the sales curves). CP seems to provide a slightly larger profit growth than MIP, and also provides good solutions already after 30 seconds. Unfortunately, CP does not provide an upper bound on the optimal value during the resolution. In practice, the CP approach seems to be the most efficient method overall.

#### 5.4. Scheduling over a week

Only CP was able to construct feasible weekly schedules consistently in less than two hours of CPU time, for the week instances in our data. For MIP, even the combination with PLA with a single linear piece (MIP-PLA<sub>1</sub>) failed to provide feasible solutions for over 30% of the weeks. Table 8 shows the evolution of the gap between the solutions identified by solving the CP model after 30 minutes, 60 minutes, and 120 minutes of CPU time, and the best upper bound extracted after 24 hours of CPU time with the MIP model. Given that the weekly schedules used historically in the stores do not satisfy the new rules imposed by the retailer, we cannot make a fair comparison with our solutions. This is why we only compare with an upper bound. The quality of the schedules constructed by CP is demonstrated by the small average gap (less than 3% after 60 minutes) between the value of these solutions the corresponding upper bounds (which are likely to be larger than the optimal values).

## 6. Conclusion

We have proposed a new methodology to construct schedules for the non-identical workforce of retail stores, to optimize the expected net revenue, based on sales curves that provide the expected revenue from sales as a function of the number of working employees on the floor. We developed two formulations of the profit-maximization problem, as a MIP and as a CP, which incorporate both the operating revenue and the context of a non-identical workforce, and consider employees as a source a revenue, an approach that differs markedly from the traditional approach where employees are considered as a sole expense and have the same specificity (availability, skills, etc.).

Our numerical experiments indicate that although both CP and MIP are able to provide near-optimal solutions in a few minutes for the scheduling problems over one day, CP is faster and provides better solutions. When constructing schedules for larger problem instances (over a week), CP provides a solution that respect all the regulation rules of the retailer in less than an hour of CPU time, with a gap of around 3% with respect to the upper bound given by the solution of the linear relaxation of the equivalent MIP model obtained after 24 hours of computation. For the MIP, we were unable to consistently generate feasible solution within an hour.

Further work will focus on the elaboration of better branching and cutting strategies to help the resolution of CP, as well as the consideration of multiple skills and multiple working activities. We also plan to improve the accuracy of the staffing forecast models by considering inter-period dependence and events, such as customers that leave the store before buying items in order to come back later the same day, when more sellers are available.

## References

- Avramidis, A.N., Chan, W., Gendreau, M., L'Ecuyer, P., Pisacane, O., 2010. Optimizing daily agent scheduling in a multiskill call centers. *European Journal of Operational Research* 200, 822–832.
- Aykin, T., 1996. Optimal shift scheduling with multiple break windows. *Management Science* 42, 591–602.
- Bechtolds, S., Jacobs, L., 1990. Implicit optimal modeling of flexible break assignments. *Management Science* 36, 1339–1351.
- Bouchard, M., 2004. Optimisation des pauses dans le problème de fabrication des horaires avec quarts de travail. *Memoire de maitrise, Ecole Polytechnique de Montreal*.
- Brockwell, P.J., Davis, R.A., 1991. *Time Series: Theory and Methods*. 2nd ed., Springer.
- Chapados, N., Joliveau, M., Rousseau, L.M., 2011. Retail store workforce scheduling by expected operating income maximization, in: Achterberg, T., Beck, J. (Eds.), *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Springer Berlin / Heidelberg. volume 6697 of *Lecture Notes in Computer Science*, pp. 53–58.
- Côté, M.C., Gendron, B., Quimper, C.G., Rousseau, L.M., 2011a. Formal languages for integer programming modeling of shift scheduling problem. *Constraints* 16, 54–76.
- Côté, M.C., Gendron, B., Rousseau, L.M., 2011b. Grammar-based integer programming models for multi-activity shift scheduling. *Management Science* 57, 151–163.
- Dantzig, G., 1954. A comment on Edie's traffic delay at toll booths. *Journal of Operation Research Society of America* 2, 339–341.
- Demasse, S., Pesant, G., Rousseau, L.M., 2006. A cost-regular based hybrid column generation approach. *Constraints* 11, 315–333.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Edie, L., 1954. Traffic delays at toll booths. *Operations Research* 2, 107–138.
- Hilbe, J.M., 2011. *Negative Binomial Regression*. Second edition ed., Cambridge University Press, Cambridge, UK.
- Hyndman, R.J., Razbash, S., Schmidt, D., 2012. forecast: Forecasting functions for time series and linear models. URL: <http://CRAN.R-project.org/package=forecast>. r package version 3.21.
- Kabak, Ö., Ülengin, F., Aktaş, E., Önsel, Ş., Topcu, Y.I., 2008. Efficient shift scheduling in the retail sector through two-stage optimization. *European Journal of Operational Research* 184, 76–90.

- Koole, G., Pot, A., 2011. A note on profit maximization and monotonicity for inbound call centers. *Operations Research* 59, 1304–1308.
- Lam, S.Y., Vandenbosch, M., Hullah, J., Pearce, M., 2001. Evaluating promotions in shopping environments: Decomposing sales response into attraction, conversion, and spending effects. *Marketing Science* 20, 194–215.
- Lam, S.Y., Vandenbosch, M., Pearce, M., 1998. Retail sales force scheduling based on store traffic forecasting. *Journal of Retailing* 74, 61–88.
- Lehmann, E.L., Casella, G., 1998. *Theory of Point Estimation*. Second ed., Springer.
- Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J., 1997. *Forecasting: Methods and Applications*. 3rd ed., Wiley.
- Mani, V., Kesavan, S., Swaminathan, J.M., 2011. Understaffing in retail stores. Working Paper.
- McCullagh, P., 1980. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B* 42, 109–142.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Second ed., Chapman & Hall, London.
- McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. second ed., John Wiley & Sons, Hoboken, NJ.
- Mehrotra, A., Murthy, K., Trick, M., 2000. Optimal shift scheduling: A branch-and-price approach. *Naval Research Logistics* 47, 185–200.
- Netessine, S., Fisher, M.L., Krishnan, J., 2010. Labor planning, execution, and retail store performance: an exploratory investigation. Working Paper, University of Pennsylvania.
- Perdikaki, O., Kesavan, S., Swaminathan, J.M., 2012. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing and Service Operations Management* 14, 145–162.
- Pesant, G., 2004. A regular language membership constraint for finite sequences of variables., in: *Proceedings of the tenth international conference on principles and practice of Constraint Programming (CP 2004)*, pp. 482–495.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed Effects Models in S and S-Plus*. Springer-Verlag, New York, NY.
- Rekik, M., Cordeau, J.F., Soumis, F., 2004. Using Benders decomposition to implicitly model tour scheduling. *Annals of Operations Research* 128, 111–133.
- Rekik, M., Cordeau, J.F., Soumis, F., 2005. Implicit shift scheduling with multiple breaks and work stretch duration restrictions. Publication GERAD-2005-15 .
- Thompson, G.M., 1995. Labor scheduling using npv estimates of the marginal benefit of additional labor capacity. *Journal of Operations Management* 13, 67–86.
- Ton, Z., 2009. The effect of labor on profitability: The role of quality. Working Paper, Harvard Business School.
- U.S. Census Bureau (Ed.), 2011. *Statistical Abstract of the United States: 2012*. U.S. Census Bureau. chapter Wholesale and Retail Trade.
- Vatri, E., 2001. *Integration de la generation de quart de travail et de l'attribution d'activites*. Memoire de maitrise, Ecole Polytechnique de Montreal.