Optimization Methods and Software

ISSN: 1055-6788 (Print) 1029-4937 (Online) Journal homepage: https://www.tandfonline.com/loi/goms20

Distributionally robust workforce scheduling in call centres with uncertain arrival rates

S. Liao , C. van Delft & J.-P. Vial

Optimization Methods & Software, 2013
Vol. 28, No. 3, 501–522, http://dx.doi.org/10.1080/10556788.2012.694166

Distributionally robust workforce scheduling in call centres with uncertain arrival rates

S. Liao[a,b]*, C. van Delft[c] and J.-P. Vial[d]

[a]*Ecole Centrale Paris, Chatenay Malabry, 92290, France;* [b]*EPFL – TOM, Odyssea 1.18, Station 5, 1015 Lausanne, Switzerland;* [c]*HEC, Jouy-en-Josas, Paris, France;* [d]*Ordecsys, Chêne-Bougeries, Switzerland*

(*Received 10 June 2011; final version received 28 April 2012*)

Call centre scheduling aims to determine the workforce so as to meet target service levels. The service level depends on the mean rate of arrival calls, which fluctuates during the day, and from day to day. The staff schedule must adjust the workforce period per period during the day, but the flexibility in doing so is limited by the workforce organization by shifts. The challenge is to balance salary costs and possible failures to meet service levels. In this paper, we consider uncertain arrival rates, that vary according to an intra-day seasonality and a global *busyness* factor. Both factors (seasonal and global) are estimated from past data and are subject to errors. We propose an approach combining stochastic programming and distributionally robust optimization to minimize the total salary costs under service level constraints. The performance of the robust solution is simulated via Monte-Carlo techniques and compared to the solution based on pure stochastic programming.

Keywords: call centres; uncertain arrival rates; robust optimization; ambiguity; staff-scheduling; totally unimodular

1. Introduction

Over the past decades, call centres have emerged as an essential component of the customer relationship management strategy for many large companies. For instance, Brown et al.[9] report that in 2002 more than 70% of all customer–business interactions were handled by call centres. This customer service has become a key factor in gaining or maintaining market shares for many companies. As a consequence, call centre performance indices, such as customer waiting times, are considered now as important assets to be optimized, in particular through efficient workforce management of skilled operators. For this sector of the service industry, the staffing cost is a major component in the operating costs and can represent 70% of the labour cost (see [17]).

Due to the importance of the sector, an abundant literature has focused on call centre operations analysis, management and optimization. We refer the reader to the comprehensive surveys in [1,17]. A central feature in call centres performances is the significant randomness and uncertainty plaguing the call arrival and service processes. In this paper, we propose an approach combining stochastic programming (SP) and distributionally robust (DR) optimization for call

*Corresponding author. Email: shuangqing.liao@epfl.ch

centre multi-shift staffing problems under service level constraints, in a setting of call centres with uncertain calls arrival rates and intra-day seasonality.

Most call centre models in the literature assume that the calls arrive according to a Poisson process with known and constant mean arrival rates. However, data from practice often reveal that the process parameters are themselves subject to fluctuations, with a significant impact on the different performance measures, see [2,18,21]. In this paper, we take the mean arrival rate of calls to be uncertain and assume that it follows a compound of seasonal pattern and purely random variation. The call arrival process is thus a doubly non-stationary stochastic process, with random mean arrival rates that vary according to intra-day seasonality and a random parameter called *busyness factor* of the day. The theoretical number of agents required to efficiently handle the inbound calls can be computed as a function of the arrival rates, and thus vary according to the period in the day and magnitude of the busyness factor. Due to the randomness of the busyness factor under-staffing may occur, but a constraint on the total expected under-staffing across the day is set to limit its negative impact. The staff-scheduling problem is modelled as a cost optimization-based problem with staffing constraints and an expected value constraint. The cost criterion function is the agent salary cost. Our objective is to find the optimal shift scheduling which minimizes the salary cost subject to the total expected under-staffing limit. The stochastic programming model of [24] is a rather straightforward application of the standard formulation [7]. The discrete distribution on the arrival rates is obtained as the product of a discretization of the busyness factor and the intra-day seasonal factors. Note that in [24], the total expected under-staffing is treated as a component of the objective function through a penalty parameter rather than a constraint.

Errors in estimating the discrete probabilities entering the expected value may affect the quality of the solution. This issue has been addressed in the pioneering work of Scarf [26]. There, the simple newsboy-inventory problem is modelled as a continuous stochastic programming problem with random demand. The only known characteristics of the distribution are the mean and the variance. The suggested strategy is the one of minimizing the expected cost with respect to the most unfavourable distribution having the prescribed mean and demand. This minimax strategy with respect to a class of distributions defined by their moments has been used in many subsequent papers, e.g. [8,11,13,14,16,29]. In the light of the revival of the concept of robust optimization in the late 1990s [4,15,23], minimax solutions to stochastic programming with uncertain distributions are now named *distributionally robust*. In a recent paper [6], the authors suggest another approach. Constraints involving expectations with respect to finite uncertain distributions could be viewed as standard constraints affected by uncertain coefficients (the probabilities) in an affine way. A constraint affected by uncertain parameters is called *uncertain constraint* as in [3]. The usual methodology in Robust Optimization consists in associating an uncertainty set to these parameters and look for solutions that enforce the uncertain constraint for all parameter values (in our case, probabilities) in the uncertainty set. The key point in this approach is the design of efficient and meaningful uncertainty sets. These sets should possess two essential properties: they should encapsulate the available knowledge on the distribution of the parameters; and they should lead to numerically tractable formulations. In the case of uncertain probabilities, the knowledge often comes from a statistical estimation procedure and takes the form of point estimates (the nominal probabilities) and some kind of confidence set around those values. Those features are well captured by ϕ-divergence functionals, in particular by using them to define distances from alternative distributions to a reference one [25].

In the present paper, we use the well-known chi-squared statistic, which is a special case of a ϕ-divergence statistic, to define uncertainty regions for the unknown demand distribution. The equivalent counterpart of the robust constraint with respect to such an uncertain region is conic quadratic. Because the staffing problem involves integer decision variables, we chose to approximate the ℓ_2 norm of the conic quadratic constraint with an ℓ_1 norm. The resulting problem can be recast into a mixed integer linear programming problem, which is easily solved by available optimization software. The chi-squared statistic has been used by Klabjan *et al.* [22] in the context of lot-sizing, while Calafiore [10] uses the Kullback–Leibler divergence. In [27], the derivation of the uncertainty set is based on the likelihood ratio. The paper [6] provides an extensive list of ϕ-divergences leading to tractable equivalent robust counterparts.

One of the attractive feature of robust optimization is that even though the uncertainty sets do not involve probabilities and expectations in their definition, nice probabilistic results can still be proved for the chance-constrained formulation [12] of the uncertain constraint. These results are based on safe convex approximations of chance constraints [5, Chapter 4]. Similar results for uncertainty sets based on ϕ-divergence do not seem to be available. In particular, we have not been able to derive the result for the chi-squared (or modified chi-squared) statistic, mainly because the proof techniques rely on an independence property that the probabilities do not possess by essence (by construction they sum to 1). To get around this difficulty, we have proposed an alternative construction of the uncertainty set for which we could derive a bound on the probability of the constraint satisfaction. Unfortunately the latter set is not related to a statistical definition of a confidence region.

The paper is organized as follows. In Section 2, we describe the call centre model under consideration and formulate the associated staff-scheduling problem. The stochastic programming model of this problem is given. In Section 3, we propose two distributionally robust models for the staff-scheduling problem. In Section 4, we conduct a numerical study to evaluate these alternative formulations. We exhibit the impact of the uncertainty of the distributional probability.

2. Problem formulation

We consider a call centre with a single type of inbound calls in a multi-period multi-shift setting. The service level depends on the current workforce (number of servers) and of the inbound call arrival process. The latter is of the Poisson type; it essentially depends on the mean arrival rate, which varies during the day and according to the random busyness of the day. To account for these variations, Liao *et al.* [24] proposed a stochastic programming formulation of the single shift problem. We present here a closely related formulation with an extension to the multi-shift problem. A main difference with the paper quoted above concerns the handling of understaffing. In the present paper, we put the constraint that understaffing does not exceed a fraction of the required staff, while in [24] understaffing was simply part of the objective with a penalty factor.

2.1 *The inbound call arrival process*

Several characteristics of the call arrival process have been underlined in the recent call centre literature. First, it has been observed that the total daily number of calls has an over-dispersion relative to the classical Poisson distribution. Second, the mean arrival rate considerably varies with the time of day. Third, there is a strong positive correlation between arrival counts during the different periods of the same day. We refer the reader to [2,9] for more details.

In order to address uncertain and time-varying mean arrival rates coupled with significant correlations, we model the inbound call arrival process by a doubly stochastic Poisson process (see [2,20,28]) as follows. We assume that a given working day is divided into n distinct periods of equal length T, so that the overall horizon is of length nT. The period length is 15 or 30 min in practice.

The inbound calls arrive following a stochastic process with a random arrival rate in each period i, denoted by Λ_i. Furthermore, using the modelling in [2,28], we assume that the arrival rate Λ_i

is of the form

$$\Lambda_i = \Theta f_i, \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where Θ is a positive real-valued random variable. The random variable Θ can be interpreted as the unpredictable level of busyness of a day. A large (small) outcome of Θ corresponds to a busy (not busy) day. The factors f_i model the intra-day seasonality, i.e. the shape of the variation of the arrival rate intensity across the periods of the day, and they are assumed to be known. Formally, if a sample value in a given day of the random variable Θ is denoted by θ, given the seasonality factor f_i, the corresponding outcome of the arrival rate over period i for that day is defined by $\lambda_i = \theta f_i$.

We assume that service times for inbound calls are independent and exponentially distributed with rate μ. The calls arrive to a single infinite queue working under the the first come, first served discipline of service. Neither abandonment, nor retrials are allowed.

2.2 *Shift setting*

We denote the set of periods of the day by I. Let J be the set of all the feasible work schedules, each of which dictates if an agent answers calls in period $i \in I$. For $i \in I$ and $j \in J$, we define the $|I| \times |J|$ matrix $\mathbf{A} = [a_{ij}]$, where

$$a_{ij} = \begin{cases} 1, & \text{if agents in schedule } j \text{ answer calls during period } i, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore we assume that each agent works during consecutive periods, without breaks. Under this assumption, it is immediate that every column of matrix $\mathbf{A}$ has contiguous ones and this kind of matrix is totally unimodular (see [17]), i.e. for any integral vector $\mathbf{b}$, every extreme point of the feasible region $\{\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$ is integral and thus the feasible region is an integral polyhedron.

2.3 *Service level and staffing methods*

Queuing models are used to determine how many agents must be available to serve calls over a given period. The M/M/N (Erlang C) queuing model is widely used to estimate stationary system performance of short half-hour or hour-periods. A standard service level constraint is introduced for each time period, through which the waiting time is kept in convenient limits. For period i, let the random variable WT_i denote the waiting time of an arbitrary call. The probability distribution of the waiting time of calls is computed using the classical results of the Erlang C model. In doing so, the mean arrival rates and service rates are assumed to be constant in each period of the day, as well as the system achieves a steady state quickly within each period. It is well known [19] that for a given staffing level N, which only handles inbound calls, one has for period i,

$$\begin{aligned} F_{\theta f_i}(N) = P\{\mathrm{WT}_i \leq \mathrm{AWT} | \theta\}(N) = 1 - \left(\sum_{j=0}^{N-1} \frac{(\theta f_i/\mu)^j}{j!} + \frac{(\theta f_i/\mu)^N}{N!(1 - \theta f_i/\mu/N)} \right)^{-1} \\ \times \frac{(\theta f_i/\mu)^N}{N!(1 - (\theta f_i/\mu)/N)} \, \mathrm{e}^{-(N\mu - \theta f_i)\mathrm{AWT}}, \end{aligned} \tag{2}$$

where AWT represents the acceptable waiting time. For a given value of the objective service level in period i, say $\mathrm{SL}_i\%$, and a given sample value of the arrival rate, θf_i, this formula is used in the reciprocal way in order to compute the staffing level which guarantees the required service level,

$$F^{-1}_{\theta f_i}(\mathrm{SL}_i). \tag{3}$$

2.4 *Stochastic programming models for an optimal staffing*

2.4.1 *Model with deterministic seasonality factors f*

We assume first that the f_i are certain and that Θ follows a discrete probability distribution, defined by the sequence of outcomes θ_l, $l \in L$ with L as the outcomes set. The assumed probability distribution is presented by q_l, with constraint $\sum_{l\in L} q_l = 1, q_l \geq 0$. For period $i \in I$, the parameters $N_{il} = F^{-1}_{\theta_l f_i}(\mathrm{SL}_i)$, estimated via (3), represent the required number of agents in period i associated with a particular *busyness factor* value θ_l.

Let x_j, $j \in J$, be the decision variables representing the numbers of agents assigned to the various schedules implemented before the start of the day. Each agent assigned to shift j gets a salary c_j for the day. In order to optimize the call centre operational cost, Liao *et al.* [24] proposed the following stochastic programming model

$$\begin{aligned} \min \quad & \sum_{j\in J} c_j x_j \\ \text{s.t.} \quad & \sum_{l\in L}\sum_{i\in I} q_l M_{il} \leq \bar{M} \\ & \sum_{j\in J} a_{ij}x_j + M_{il} \geq N_{il}, \quad i \in I,\ l \in L \\ & x_j \in \mathbb{Z}^+, \quad j \in J \\ & M_{il} \geq 0, \quad i \in I,\ l \in L. \end{aligned} \tag{4}$$

The objective of Problem (4) is to minimize the agents salary cost. The variables M_{il} represent the amount of under-staffing at period i in event l. The first constraint states that the total expected under-staffing should not exceed the prescribed limit $\bar{M}$. The second constraint bounds from below the understaffing amount at each period of the day and for each level of the busyness factor. When the first constraint on the expected understaffing is active at the optimum, the second constraint will also be active and is equivalent to defining the understaffing as $M_{il} = \max\{0, N_{il} - \sum_{j\in J} a_{ij}x_j\}$. The last two sets of constraints define the non-negativity and integer conditions for programme variables.

It is possible to take advantage of the totally unimodular structure of matrix $\mathbf{A} = (a_{ij})$ and make Problem (4) computationally much easier by adding auxiliary variables ($y_i \in \mathbb{Z}^+, i \in I$) to represent the available work force $\sum_j a_{ij}x_j$. Indeed, the variable x appears in equation $\sum_j a_{ij}x_j = y_i$ and in the objective, but nowhere else. Hence, the integrality condition on y is sufficient to enforce integrality of the x in any solution produced by the Simplex algorithm. The new formulation is

$$\begin{aligned} \min \quad & \sum_{j\in J} c_j x_j \\ \text{s.t.} \quad & \sum_{l\in L}\sum_{i\in I} q_l M_{il} \leq \bar{M} \end{aligned}$$

$$\sum_{j \in J} a_{ij} x_j = y_i, \quad i \in I$$
$$y_i + M_{il} \geq N_{il}, \quad i \in I,\ l \in L \tag{5}$$
$$y_i \in \mathbb{Z}^+,\ i \in I$$
$$x_j \geq 0, \quad j \in J$$
$$M_{il} \geq 0, \quad i \in I,\ l \in L.$$

Clearly, Problem (5) is equivalent of Problem (4). Notice that the integer constraints on x_j are relaxed, Problem (5) contains $|I|$ integer variables and $|J| + |I| \times |L|$ continuous variables while Problem (4) contains $|J|$ integer variables and $|I| \times |L|$ continuous variables.

Problem (5) turns out to be numerically much easier than (4) even if $|J| \leq |I|$. The reason is that the more important constraint on the y's is a simple lower bound. On small problems (e.g. 10 periods and 9 shifts), we already observed a time saving by a factor 3. So we only used formulation (5) on the larger instances used in our experimental study.

2.4.2 *Model with uncertain seasonality factors f*

The seasonality factors may not be known with certainty. Their value is usually estimated through some statistical scheme, and their true value may differ from the estimated one. We can represent the true f_i in period i as its estimator $\hat{f}_i$ plus a white noise ϵ_i:

$$f_i = \hat{f}_i + \epsilon_i.$$

We assume that θ and the noises ϵ_i are independent. Our choice is motivated by the statistical study in [9], which shows that the estimation of θf_i, where θ is random, is prone to an additional error ε_i. Our model is slightly different but captures the same idea.

The theoretical staff size that is required to meet the desired service level in period i also depends on the random noise ϵ_i. We now replace the continuous distribution of the ϵ_i by a discrete one, or equivalently a discrete distribution of the f_i. Let f_{ik}, $k \in K_i$ be the set of discrete values and let π_{ik}, with $\sum_{k \in K_i} \pi_{ik} = 1$, be the associated probabilities. For period $i \in I$, the parameters $N_{ikl} = F^{-1}_{\theta_l f_{ik}}(\mathrm{SL}_i)$, estimated via (3), represent the required number of agents associated with a particular *busyness factor* value θ_l and seasonality factor f_{ik}. We can now formulate an extension of the base model of [24] to account for the stochastic variability of the seasonality factors f_i:

$$\begin{aligned}
\min \quad & \sum_{j \in J} c_j x_j \\
\text{s.t.} \quad & \sum_{l \in L} q_l \sum_{i \in I} \sum_{k \in K_i} \pi_{ik} M_{ikl} \leq \bar{M} \\
& \sum_{j \in J} a_{ij} x_j = y_i, \quad i \in I \\
& y_i + M_{ikl} \geq N_{ikl}, \quad i \in I,\ k \in K_i,\ l \in L \\
& y_i \in \mathbb{Z}^+,\ i \in I \\
& x_j \geq 0, \quad j \in J \\
& M_{ikl} \geq 0, \quad i \in I,\ k \in K_i,\ l \in L.
\end{aligned} \tag{6}$$

3. Distributionally robust model

In the above stochastic programming formulation, the true distribution of θ was assumed to be known, and as a consequence the different constraints of the models are satisfied for any outcome θ_l associated with this distribution.

At the end of the previous section, we proposed an extension of [24] to account for the stochastic variability of the seasonality factors. We now turn our attention to the busyness factor θ. The same argument as for the seasonality factors holds concerning the imperfect knowledge on the true distribution of θ. To make the solutions of models (5) and/or (6) robust with respect to this imperfect knowledge, we substitute to the estimated probabilities of the busyness values θ a family of alternative probabilities distributions compatible with the observed values of θ. The distributionally robust solution is such that it solves the stochastic programming staffing problem against the worst probability distribution in the class of alternative distributions for θ.

A standard problem in such an approach is the size of the probability distribution set. It is well known that too large sets, i.e. in our case, sets including all potential probability distributions, can be extremely conservative in the sense that the robust solution has an objective function value much worse than the objective function value of the solution of the nominal distribution.

It is thus necessary to consider partial uncertainty sets, in the sense that some potential distributions are not included. The idea consists then of introducing, by tuning the size of the uncertainty set, efficient tradeoffs between the probability of constraint violation and the objective function value. Our approach allows thus the modeler to vary the level of conservatism of the robust solutions in terms of probabilistic bounds of constraint violations. Clearly, in such a process, theoretical bounds linking uncertain sets size and constraints violation probabilities are required.

3.1 *Uncertainty set based on a statistical dispersion model*

The true probability distribution of the random factor Θ is not known. It must be estimated by some statistical means. For instance, we can imagine that a set $(\hat{\theta}_1, \dots, \hat{\theta}_N)$ of historical data is available. The maximum-likelihood estimator of the true probability p_l is the observed frequency $q_l = n_l/N$. Moreover, the classical Pearson's test of goodness of fit is based on the quantity

$$X^2 = \sum_l \frac{(n_l - Np_l)^2}{Np_l} = \sum_l N\frac{(q_l - p_l)^2}{p_l}.$$

Asymptotically X^2 follows a χ^2 distribution with $|L| - 1$ degrees of freedom. This asymptotic distribution probability makes it possible to define a first confidence region around q for the true probability p. To this end, we define the dispersion measure $\sum_l N(q_l - p_l)^2/q_l$ and introduce the set of alternative probabilities

$$H_\alpha = \left\{ p \geq 0 : \sum_l N\frac{(q_l - p_l)^2}{q_l} \leq \alpha, \sum_l p_l = 1 \right\} \tag{7}$$

that are somehow compatible with the observed frequencies q_l.

The goal of the present analysis would be to incorporate this formulation into the stochastic programming formulation (4). Namely, we shall try to solve (4) for the worst possible distribution of p in the confidence region (7). The formal implementation of this idea consists of replacing the constraint $\sum_{l\in L}\sum_{i\in I} q_l M_{il} \leq \bar{M}$ in (4) with its robust counterpart

$$\sum_{l\in L}\sum_{i\in I} p_l M_{il} \leq \bar{M}, \quad \text{for all } p \in H_\alpha. \tag{8}$$

Note that (8) is equivalent to

$$\max_{p\in H_\alpha}\left\{\sum_{l\in L}\sum_{i\in I}p_l M_{il}\right\}\le \bar{M}.$$

However, it can be shown that this infinite dimensional robust counterpart has an equivalent formulation as a conic quadratic constraint. Due to the presence of integer variables x, the equivalent robust counterpart leads to a nonlinear mixed integer problem, possibly a difficult one to solve. We shall not use this test in our analysis, but we shall be inspired by it to define a kind of confidence level set for the true probability p. We shall see that we can replace (8) by a more restrictive constraint that is equivalent to a set of linear inequalities. In this way, we remain in the realm of linear programming with integer variables.

To remain in the realm of mixed integer linear programming for which powerful commercial solvers exist, we shall replace the maximization over the confidence region H_α, by the maximization over a larger, but linear, set

$$\mathcal{P}_\beta=\left\{p\ge 0:\sum_{l\in L}p_l=1,\sum_{l\in L}\frac{|p_l-q_l|}{\sqrt{q}_l}\le\beta\right\}. \tag{9}$$

The larger β, the larger the admissible dispersion and the higher is the protection against the unfavourable probability distributions. Clearly, in order to be coherent with (7), the set size factor β has to be chosen to enforce $H_\alpha\subset\mathcal{P}_\beta$. From the simple inequality on norms, we have for any $\theta\in\mathbb{R}^{|L|}$

$$\sum_{l\in L}|\theta_l|\le\sqrt{|L|}\sqrt{\sum_{l\in L}\theta_l^2}.$$

It follows that for $\beta_\alpha=\sqrt{|L|}\sqrt{\alpha/N}$ the set $\mathcal{P}_{\beta_\alpha}$ contains the set H_α. Therefore, one has

$$\beta_\alpha=\sqrt{|L|}\sqrt{\frac{\alpha}{N}}\Rightarrow H_\alpha\subset\mathcal{P}_{\beta_\alpha}.$$

Hence

$$\max_{p\in H_\alpha}\left\{\sum_{l\in L}\sum_{i\in I}p_l M_{il}\right\}\le\max_{p\in\mathcal{P}_{\beta_\alpha}}\left\{\sum_{l\in L}\sum_{i\in I}p_l M_{il}\right\}$$

and (9) implies (8). Let us now derive the equivalent counterpart of (9). Let

$$\begin{aligned}F&=\max_{p\in\mathcal{P}_\beta}\sum_{l\in L}p_l\sum_{i\in I}M_{il}\\&=\max_p\left\{\sum_{l\in L}p_l\sum_{i\in I}M_{il}:\sum_{l\in L}\frac{|p_l-q_l|}{\sqrt{q_l}}\le\beta,\ \sum_{l\in L}p_l=1,\ p_l\ge 0,\forall l\in L\right\}.\end{aligned} \tag{10}$$

We shall now explicit problem (10) as a linear programming problem. Define the new variables

$$\delta_l=p_l-q_l,$$

the problem becomes

$$
\begin{aligned}
\max_{\delta} \quad & \sum_{l\in L} q_l \sum_{i\in I} M_{il} + \sum_{l\in L}\sum_{i\in I} M_{il}\delta_l \\
\text{s.t.} \quad & \sum_{i\in L} \frac{|\delta_l|}{\sqrt{q_l}} \leq \beta \\
& \sum_{l\in L} \delta_l = 0 \\
& \delta_l \geq -q_l, \quad l \in L.
\end{aligned}
\tag{11}
$$

We consider the dual of Problem (11),

$$
\begin{aligned}
\min_{v,w,z} \quad & \sum_{l\in L} q_l \sum_{i\in I} M_{il} + \sum_{l\in L} q_l w_l + \beta z \\
\text{s.t.} \quad & z \geq \sqrt{q_l}\left[\sum_{i\in I} M_{il} + v + w_l\right], \quad l \in L \\
& z \geq -\sqrt{q_l}\left[\sum_{i\in I} M_{il} + v + w_l\right], \quad l \in L \\
& w_l \geq 0, \quad l \in L.
\end{aligned}
\tag{12}
$$

By strong duality, since Problem (11) is feasible and bounded, then the dual Problem (12) is also feasible and bounded and their objective values coincide.

Back to the global formulation of the staffing problem with uncertain busyness daily factors, we obtain the following mixed integer linear programming problem in the original variables (x, M) and the auxiliary variables (v, w, z).

$$
\begin{aligned}
\min \quad & \sum_{j\in J} c_j x_j \\
\text{s.t.} \quad & \sum_{l\in L} q_l \left(\sum_{i\in I} M_{il}\right) + \sum_{l\in L} q_l w_l + \beta z \leq \bar{M} \\
& -z \leq \sqrt{q_l}\left[\sum_{i\in I} M_{il} + v + w_l\right] \leq z, \quad \forall l \in L \\
& \sum_{j\in J} a_{ij} x_j + M_{il} \geq N_{il}, \quad i \in I,\ l \in L \\
& x_j \in \mathbb{Z}^+,\ j \in J \\
& M_{il} \geq 0, \quad i \in I,\ l \in L \\
& w_l \geq 0, \quad l \in L.
\end{aligned}
\tag{13}
$$

Problem (13) is the equivalent robust counterpart of the robust version of Problem (4) with an uncertainty set (9) for the underlying business factor probability distribution. It is worth elaborating on the first constraint in Problem (13). The first term on the left-hand side is the expected under-staffing taken with respect to the reference, or nominal, probability distribution q. The other two components are safety factors the extra under-staffing that could occur when the true probability distribution is the worst possible in the uncertainty set. Note that the safety term βz is proportional to the *immunization* factor β. The larger β, the larger the admissible dispersion and the higher is the protection against the risk of incurring an extra under-staffing if the distance between the true distribution p and the nominal distribution q increases.

Similar to what we proposed in Problem (5), a possible way to make Problem (13) easier to solve is to add some auxiliary variables ($y_i \in \mathbb{Z}^+, i \in I$), Problem (13) can then be reformulated as

$$\begin{aligned}
\min \quad & \sum_{j\in J} c_j x_j \\
\text{s.t.} \quad & \sum_{l\in L}\sum_{i\in I} q_l M_{il} + \sum_{l\in L} q_l w_l + \beta z \le \bar{M} \\
& -z \le \sqrt{q_l}\left[\sum_{i\in I} M_{il} + v + w_l\right] \le z, \quad \forall l \in L \\
& \sum_{j\in J} a_{ij} x_j = y_i, \quad i \in I \\
& y_i + M_{il} \ge N_{il}, \quad i \in I,\ l \in L \\
& y_i \in \mathbb{Z}^+, \quad i \in I \\
& x_j \ge 0,\ j \in J \\
& M_{il} \ge 0, \quad i \in I,\ l \in L \\
& w_l \ge 0, \quad l \in L.
\end{aligned} \tag{14}$$

Problem (14) is equivalent to Problem (13). Notice that the integer constraints on x_j are relaxed, since y_i are restricted to be integers. Thanks to the total unimodularity property of matrix $\mathbf{A}$, x_j are automatically integers. Problem (14) contains $|I|$ integer variables and $|J| + |I| \times |L| + |L| + 2$ continuous variables while Problem (13) contains $|J|$ integer variables and $|I| \times |L| + |L| + 2$ continuous variables.

Model (14) is easily extended to the case with uncertain seasonality factors as it was done in Section 2.4. The equivalent robust counterpart is then

$$\begin{aligned}
\min \quad & \sum_{j\in J} c_j x_j \\
\text{s.t.} \quad & \sum_{l\in L}\sum_{i\in I} q_l \sum_{k\in K_i} \pi_{ik} M_{ikl} + \sum_{l\in L} q_l w_l + \beta z \le \bar{M} \\
& -z \le \sqrt{q_l}\left[\sum_{i\in I}\sum_{k\in K_i} \pi_{ik} M_{ikl} + v + w_l\right] \le z, \quad \forall l \in L \\
& \sum_{j\in J} a_{ij} x_j = y_i, \quad i \in I \\
& y_i + M_{ikl} \ge N_{ikl}, \quad i \in I,\ k \in K_i,\ l \in L \\
& y_i \in \mathbb{Z}^+, i \in I \\
& x_j \ge 0,\ j \in J \\
& M_{ikl} \ge 0, \quad i \in I,\ k \in K_i,\ l \in L \\
& w_l \ge 0, \quad l \in L.
\end{aligned} \tag{15}$$

3.2 *Standard uncertainty set: an alternative formulation*

A statistical dispersion measure, like Pearson's, is a sensible choice for the design of an efficient uncertainty set. Unfortunately, it does not seem possible, via such a measure, to compute a reasonable estimate of the probability that the robust solution satisfies the uncertain constraint. In this subsection, we propose an alternative uncertainty set formulation enabling such a calculation for the constraint violation probability. The derivation is based on the equivalence

$$\begin{cases} \sum_{l\in L} p_l M_l \le \bar{M} \\ \sum_{l\in L} p_l = 1,\ p \ge 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \sum_{l\in L} p'_l (M_l - \bar{M}) \le 0 \\ p' \ge 0, \end{cases} \tag{16}$$

which holds in the following sense: if the left part holds for some p, the right part holds for $p' = p$; if the right part holds for $p' \ne 0$, the left part holds for $p = p' / \sum_{l\in L} p'_l$. This naturally leads to the following uncertainty model

$$\begin{aligned} p'_l &= q_l(1+\xi_l), \quad \forall l \in L \\ \xi_l &\in [-1,1], \quad \forall l \in L \\ p &= \frac{p'}{\sum_{l\in L} p'_l}. \end{aligned} \tag{17}$$

The definition is meaningful if $\max_{l\in L} q_l \le 0.5$ and $p'_l \ne 0$. A sufficient condition for the latter is $\xi_l > -1$ for all $l \in L$.

With this model of probability, the condition on the uncertain constraint (16) becomes

$$\sum_{l\in L} p'_l (M_l - \bar{M}) = \sum_{l\in L} q_l (M_l - \bar{M}) + \sum_{l\in L} \xi_l q_l (M_l - \bar{M}) \le 0.$$

Define the uncertainty set

$$\Xi = \{\xi : \|\xi\|_\infty \le 1, \|\xi\|_2 \le k\}.$$

The robust counterpart of the uncertain constraint is thus

$$\sum_{l\in L} q_l M_l + \sum_{l\in L} \xi_l q_l (M_l - \bar{M}) \le \bar{M}, \quad \forall \xi \in \Xi. \tag{18}$$

The equivalent robust counterpart (see [3]) is the inequality

$$\sum_{i\in I} q_l M_l + k\|Q(M - \bar{M}) + w\|_2 + \|w\|_1 \le \bar{M}, \quad \text{for some } w, \tag{19}$$

where Q is a diagonal matrix with main diagonal $(q_l)_{l\in L}$.

The bound on the probability of constraint satisfaction is given by the following theorem (see [5]).

THEOREM 1 *Assume $\xi_l, l \in L$, are independent random variables with range $[-1,1]$ and common expectation $E(\xi_l) = 0$. Then, for any $z \in \mathbb{R}^{|L|}$*

$$\text{Prob}\left(\sum_{l\in L} z_l \xi_l \ge k\|z\|_2\right) \le \mathrm{e}^{-k^2/2}.$$

The theorem directly applies to a formulation with the ellipsoidal uncertainty set $\{\xi : \|\xi\|_2 \leq k\}$. Because the theorem holds under the hypothesis $\|\xi\|_\infty \leq 1$, we can replace the ellipsoidal uncertainty set by Ξ, which is the intersection of the two balls in the l_2 and l_∞ norms. We thus have

COROLLARY 1 *Assume ξ_l, $l \in L$, are independent random variables with range* $[-1, 1]$ *and common expectation $E(\xi_l) = 0$. Then for any solution to the equivalent robust counterpart* (19)

$$\text{Prob}\left(\sum_{l\in L} p_l M_l \geq \bar{M}\right) \leq \mathrm{e}^{-k^2/2}.$$

Because our problem involves integer variables, it is computationally more efficient (for the time being) to replace the ellipsoidal uncertainty set by one in the l_1-norm. Because the following inequalities hold for any $a \in \mathbb{R}^{|L|}$

$$\frac{1}{\sqrt{|L|}}\|a\|_1 \leq \|a\|_2 \leq \sqrt{|L|}\|a\|_\infty$$

we can replace Ξ by the larger uncertainty set

$$\{\xi : \|\xi\|_\infty \leq 1, \|\xi\|_1 \leq k\sqrt{|L|}\} \supseteq \Xi$$

and the equivalent robust counterpart (19) by the stricter inequality

$$\sum_{l\in I} q_l M_l + k\sqrt{|L|}\|Q(M - \bar{M}) + w\|_\infty + \|w\|_1 \leq \bar{M}, \quad \text{for some } w.$$

Finally, as shown in Proposition 1 of [3], the above inequality is equivalent to the set of inequalities

$$\begin{aligned}
&\sum_{l\in I} q_l M_l + k\sqrt{|L|}z + \sum w_l \leq \bar{M} \\
&\qquad z + w_l \geq q_l(M_l - \bar{M}), \quad l \in L \\
&\qquad z + w_l \geq q_l(\bar{M} - M_l), \quad l \in L \\
&\qquad w \geq 0, \quad z \geq 0,
\end{aligned}$$

where $w \in \mathbb{R}^{|L|}$ and $z \in \mathbb{R}$ are auxiliary variables.

In order to have a model associated with a theoretical bound for the constraint violation probability, we plug these inequalities into our distributionally robust call centre model by replacing M_l by $\sum_{i\in I}\sum_{k\in K_i} \pi_{ik} M_{ikl}$, and we obtain a new model, similar to (15):

$$\begin{aligned}
\min\ & \sum_{j\in J} c_j x_j \\
\text{s.t.}\quad & \sum_{l\in I} q_l \left(\sum_{i\in I}\sum_{k\in K_i} \pi_{ik} M_{ikl}\right) + k\sqrt{|L|}z + \sum w_l \leq \bar{M} \\
& z + w_l \geq q_l \left(\sum_{i\in I}\sum_{k\in K_i} \pi_{ik} M_{ikl} - \bar{M}\right), \quad \forall l \in L
\end{aligned}$$

$$z + w_l \geq q_l \left(\bar{M} - \sum_{i \in I} \sum_{k \in K_i} \pi_{ik} M_{ikl} \right), \quad \forall l \in L \tag{20}$$

$$\sum_{j \in J} a_{ij} x_j = y_i, \quad i \in I$$

$$y_i + M_{ikl} \geq N_{ikl}, \quad i \in I,\ k \in K_i,\ l \in L$$

$$y_i \in \mathbb{Z}^+,\ i \in I$$

$$x_j \geq 0, \quad j \in J$$

$$M_{ikl} \geq 0, \quad i \in I,\ k \in K_i,\ l \in L$$

$$z \geq 0, \quad w_l \geq 0,\ l \in L.$$

We conclude this subsection by showing that the uncertainty set Ξ could be viewed as a form of dispersion measure. Indeed, we define the set of probability distributions

$$\mathcal{P}_k = \left\{ p : p = \frac{p'}{\sum_{l \in L} p'_l}, \sum_{l \in L} \left(\frac{p' - q}{q} \right)^2 \leq k^2, p' \geq 0 \right\}. \tag{21}$$

This definition is compatible with an assumption of independence of the variables p'_l. It leads us to assume that the quantities $(p'_l - q_l)/q_l$ are independent random variables with range $[-1, 1]$. Note that it does not imply that the p_l are independent. Thanks to the independence assumption, we have been able to compute a bound on the probability of satisfaction of the uncertain constraint. The alternative formulation (21) bypasses the difficulty we have met with H_α. There, p only enters the definition and the condition $\sum_{l \in L} p_l = 1$ creates an explicit dependence among the variables.

4. Numerical experiments and results

The numerical results reported in this section aim at assessing empirically the merit of the distributionally robust approach as compared to the plain stochastic programming approach. A robust, or stochastic programming, solution consists in a set of shifts x. The behaviour of this solution is analysed on large samples of daily operations scenarios.

In this section, we conduct a numerical study in order to evaluate and compare between the classic stochastic programming approach and the distributionally robust programming approach. In Section 4.1, we describe the numerical experiments. In Section 4.2, we analyse the results and derive various insights.

4.1 *Setting of the experiments*

We describe in this section the data used in the numerical examples first, and then the design of experiments.

4.1.1 *Parameter values*

Inbound calls. In the experiments, we use real data from a Dutch hospital which exhibits a typical and significant workload time-of-day seasonality. To give an idea of the pattern of the mean arrival rate, we consider three days, a normal one, a busy one and a not so busy one. The solid line in

Figure 1, represents arrival in a normal day, while the dashed lines represent the two other cases. Clearly the three lines have a similar pattern, with low values at the beginning and at the end of the day, with a two peaks one in late morning and one in the afternoon, and a relative decrease in-between during the lunch break. This illustrates the choice of the model, with (almost) fixed seasonality factors and a multiplicative busyness factor. The day starts at 8:00 a.m., finishes at 8:30 p.m., and is divided into $|I| = 50$ periods of 15 min each.

From this observation, we construct an illustrative example as follows. The average rate of arrivals at each period of the day is supposed to have been estimated by statistical analysis on a record of $n = 400$ working days. The estimated seasonal factors are given in Table 1. Note that the seasonal factors could have been normalized because the true arrival rate is obtained by multiplying those values by the busyness day factor.

The uncertain environment of the problem is built as follows. First we consider that each individual seasonal factor is subject to an independent noise. For the sake of the illustration, we selected a discrete distribution for each seasonal factor with three outcomes $f_i - f_i/10, f_i, f_i + f_i/10$, with respective probabilities $0.25, 0.5, 0.25$. This choice is arbitrary, but can be easily replaced by an alternative one. In the numeric experiments, we analyse the general case with uncertain seasonal factors. The seasonal factors known with certainty can be considered as a special case of uncertain seasonal factors.

The second element that introduces uncertainty is the distribution of Θ, the random busyness factor. It is estimated by comparing the record of the mean arrival rate of each working day with

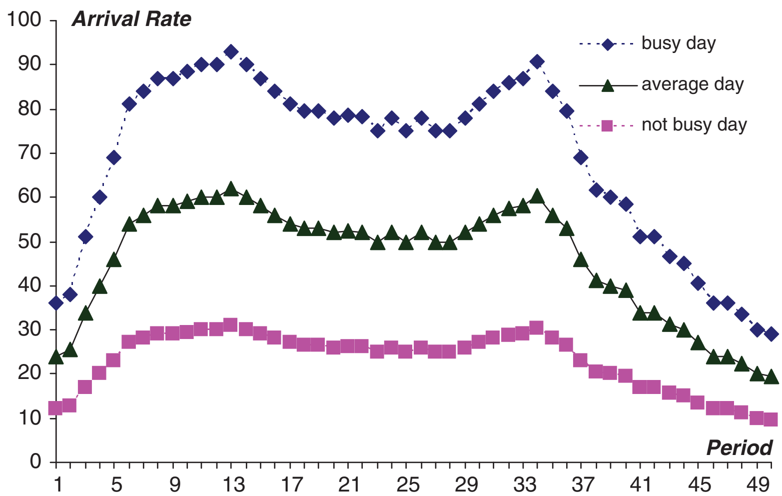

Figure 1. (Solid line), average day; (higher dashed line), a busy day; (lower dashed line), a low busyness day.

Table 1. Average seasonality factors estimated from a sample of $n = 400$ working days.

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}
6	6.35	8.5	10	11.5	13.5	14	14.5	14.5	14.75	15	15	15.5
f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}
15	14.5	14	13.5	13.25	13.25	13	13.1	13.05	12.5	13	12.5	13
f_{27}	f_{28}	f_{29}	f_{30}	f_{31}	f_{32}	f_{33}	f_{34}	f_{35}	f_{36}	f_{37}	f_{38}	f_{39}
12.5	12.5	13	13.5	14	14.35	14.5	15.1	14	13.25	11.5	10.3	10
f_{40}	f_{41}	f_{42}	f_{43}	f_{44}	f_{45}	f_{46}	f_{47}	f_{48}	f_{49}	f_{50}		
9.75	8.5	8.5	7.8	7.5	6.75	6	6	5.6	5	4.85		

the average of all these means. We assume that the distribution of Θ has been estimated from past records by a discrete distribution with $|L| = 41$ outcomes θ_l and probabilities q_l. To construct a plausible distribution, we choose to discretize a continuous distribution. In [2], the authors postulate in their Model 1 that Θ follows a gamma distribution with shape parameter $\gamma > 0$ and scale parameter 1. In this paper, we assume that Θ can take values from interval [0.00, 12.00], we take 41 equidistant points including the two endpoints 0.00 and 12.00, which gives $|L| = 41$ possible values of θ_l. And we consider three types of estimate probability distributions q: distributions A, B and C, which are discretized from a gamma distribution with scale parameter 1 and shape parameter γ (consequentially the mean $E[\Theta]$) as 2, 4 and 6, respectively. For each type of estimate probability distributions q, we have $\sum_{l\in L} q_l = 1, q_l \geq 0$. Figure 2 shows the three probability density functions.

Finally, for each value of the arrival rate at period i with busyness factor θ_l and seasonal factor (given by one of the three values $f_i - f_i/10$, f_i, $f_i + f_i/10$), we compute the staffing requirement that is needed to meet the service level. To this end, we start with the assumption that mean service time is $1/\mu = 5$ min. We use the classical service level corresponding to the well-known $\frac{80}{20}$ rule: the probability that a call waits for less than 20 s has to be larger or equal to 80%. Using Condition (3) and Definition (1), we deduce the required number of agents N_{ikl} during period i, associated to the values θ_l and f_{ik}.

The understaffing bound $\bar{M}$. The quantity is user dependent. We chose it as follows. We compute the average or the size $N = \sum_{i,k,l} q_l \pi_{ik} N_{ikl}$ of ideal staff. We consider three values for $\bar{M}$: $0, 1\% \times N$ and $2\% \times N$. Note that the value $\bar{M} = 0$ imposes that all $M_{ikl} \geq 0$ in the constraint

$$\sum_{l\in L}\sum_{i\in I}\sum_{k\in K_i} p_l \pi_{ik} M_{ikl} \leq \bar{M} = 0$$

are zero. This case corresponds to the conservative position of 100% protection.

Cost parameters. Agents work 4 or 8 h days, with neither break nor overtime. Full-time shifts (8 h) start at the hour or the hour and a half. Part-time shifts (4 h) start at the hour between 8 a.m. and 2 p.m. There are 17 part-time and full-time feasible schedules. Without loss of generality, we use a normalized cost of 1 for each period an agent works at full-time shifts. The part-time

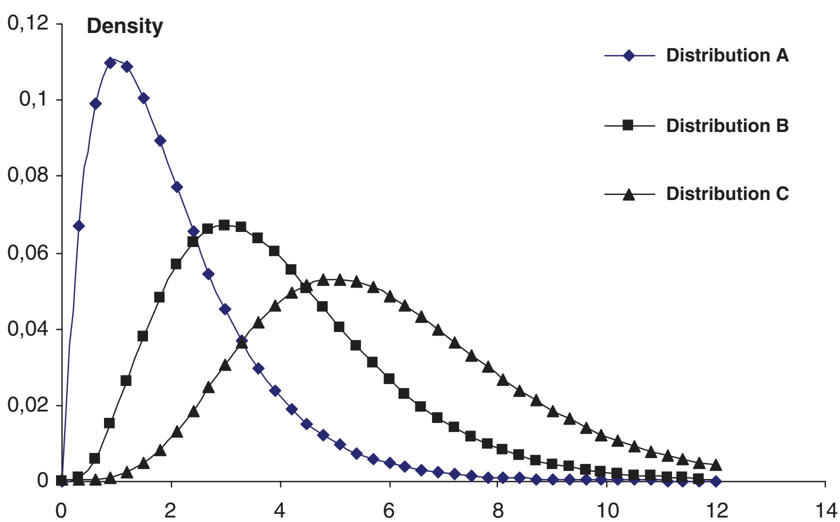

Figure 2. Some probability density functions.

shifts unit cost, assumed to be larger, is equal to 1.4 per period. Therefore, the agent salary is $c_j = \sum_{i \in I} a_{ij}$ for full-time shifts and $c_j = 1.4 \sum_{i \in I} a_{ij}$ for part-time ones.

4.1.2 *The distributionally robust solution*

We proposed two models for the uncertain parameters p_l in the constraint

$$\sum_{l \in L} p_l \sum_{i \in I} \sum_{k \in K_i} \pi_{ik} M_{ikl} \leq \bar{M}. \tag{22}$$

In the fist model, the parameter p_l belongs to the uncertainty set $\mathcal{P}_\beta$, with immunization level β. In the second model constraint (22) is written as (18) with uncertainty set $\{\xi : \|\xi\|_\infty \leq 1,\ \|\xi\|_1 \leq k\sqrt{|L|}\}$. There the immunization level is k. The choice of the proper immunization level should be guided by the goal of achieving constraint satisfaction on (22) with a probability at least equal to $1 - \alpha$, where α is some user-defined target, e.g. 5%. Corollary 1 can be used to determine the value of k, so that the solution of (20) ensures the required probability of constraint satisfaction. We do not have a comparable result for the first model of uncertainty. So the choice of β is made on empirical considerations. We compute solutions of (15) for different values of β between 0 and 1 and compare the performance of the robust solutions on simulations as described in the next subsection.

Even though k can be selected on theoretical grounds as explained above, we have settled for a similar empirical approach as for β. Indeed, it turns out that the value recommended by the theory for k is much too high in practice, leading to over-conservatism. This situation is quite common

Table 2. Models with uncertain f_i, uncertainty set $\mathcal{P}_\beta$ and $\bar{M}$ is 1% of total required workforce.

β	Salary cost	Constraint violation (%)	Expectation $(M - \bar{M} \mid M > \bar{M})$	Worst case $M - \bar{M}$
Set A of probabilities q, $\bar{M} = 64.97$				
0	21,500.8	45.46	28.96	168.88
0.01	21,660.8	43.13	27.82	164.58
0.05	22,313.6	33.55	24.16	147.87
0.1	23,164.8	22.73	20.47	126.14
0.2	24,896.0	8.77	14.96	88.56
0.5	29,372.8	0.11	7.62	21.54
0.8	32,361.6	0.00	NaN	−12.08
1	33,673.6	0.00	NaN	−23.40
Set B of probabilities q, $\bar{M} = 120.77$				
0	27,481.6	47.19	29.73	162.76
0.01	27,555.2	44.76	28.57	157.71
0.05	27,849.6	35.95	24.72	142.81
0.1	28,220.8	25.11	21.42	126.20
0.2	28,953.6	10.25	16.03	94.16
0.5	30,998.4	0.13	7.11	24.62
0.8	32,713.6	0.00	NaN	−20.22
1	33,667.2	0.00	NaN	−39.90
Set C of probabilities q, $\bar{M} = 179.19$				
0	32,752.0	48.97	28.10	150.00
0.01	32,809.6	46.00	26.81	145.28
0.05	33,030.4	34.44	22.64	129.64
0.1	33,302.4	21.56	18.56	110.80
0.2	33,814.4	6.09	12.82	77.82
0.5	35,171.2	0.01	4.29	4.29
0.8	36,243.2	0.00	NaN	−41.17
1	36,841.6	0.00	NaN	−62.68

in robust optimization and the choice of the immunization factors β and k via simulations is more appropriate.

4.1.3 *Simulations*

The idea of simulation is to create K samples of operational days. To this end, we first draw by Monte-Carlo sampling, a value for p. This is done as follows. We perform n independent random trials with respect to the probability distribution q. For each θ_l we record the frequency of occurrence of θ_l; this frequency defines p_l. Next we draw a value for each seasonal factor among the three possibilities with respect to the given probabilities (here, 0.25, 0.5 and 0.25). Given the day operation conditions, we can compute the understaffiing of the DR for that day. We have thus K realizations of the understaffing of the DR solution.

We compute three types of statistics

(1) The proportion of times the constraint on understaffing is violated, i.e. the expected understaffing $M = \sum_{l\in L}\sum_{k\in K_i}\sum_{i\in I} pi_l\pi_{ik}M_{ilk}$ exceeds $\bar{M}$.
(2) The conditional expectation value of $(M - \bar{M})$ conditionally to $M - \bar{M} > 0$.
(3) The worst case for $(M - \bar{M})$.

4.2 *Analysis of the numerical results*

In this section, we comment on the numerical results and derive the main insights. Four criterions are considered in order to evaluate the performance of both SP and DR methods: The salary

Table 3. Models with uncertain f_i, uncertainty set $\mathcal{P}_\beta$ and $\bar{M}$ is 2% of total required workforce.

β	Salary cost	Constraint violation (%)	Expectation $(M - \bar{M} \mid M > \bar{M})$	Worst case $M - \bar{M}$
Set A of probabilities q, $\bar{M} = 129.94$				
0	18,134	45.91	40.39	271.99
0.01	18,227	43.62	39.58	268.08
0.05	18,605	35.20	36.19	249.23
0.1	19,098	26.03	31.90	224.26
0.2	20,157	12.33	25.66	178.00
0.5	23,578	0.40	13.11	62.66
0.8	26,611	0.00	NaN	−3.36
1	28,342	0.00	NaN	−36.14
Set B of probabilities q, $\bar{M} = 241.54$				
0	24,438	47.42	44.80	236.19
0.01	24,493	45.66	43.31	231.43
0.05	24,704	37.55	38.82	212.71
0.1	24,973	28.10	34.18	190.61
0.2	25,504	13.40	27.41	148.93
0.5	27,059	0.38	13.22	49.83
0.8	28,550	0.00	NaN	−20.39
1	29,466	0.00	NaN	−56.78
Set C of probabilities q, $\bar{M} = 358.38$				
0	29,891	48.66	46.11	223.46
0.01	29,939	46.00	44.43	217.43
0.05	30,125	35.34	39.02	197.46
0.1	30,352	23.59	34.13	173.81
0.2	30,800	8.44	26.49	128.40
0.5	32,035	0.05	11.72	20.55
0.8	33,120	0.00	NaN	−59.46
1	33,766	0.00	NaN	−99.46

cost, the probability of violation of the constraint $M \leq \bar{M}$, the conditional expectation value of $(M - \bar{M})$ for M that exceeds $\bar{M}$, and the maximum $(M - \bar{M})$ among all the $K = 10,000$ trials. We compare the performance between SP and DR with different sizes of uncertainty sets $\mathcal{P}_\beta$ (defined by (9)) and $\mathcal{P}_k$ (defined by (21)), for different under-staffing bound $\bar{M}$. We analyse the trade-off between salary cost and the other three criterions, and show the necessity of taking into account the uncertainty in the probability distribution. These comparisons are done based on the three types of estimate probability distributions presented previously.

For the three types of estimate probability distributions, the value of the under-staffing bound $\bar{M}$, defined as 1% of the total required workforce is 64.97, 120.77 and 179.19, respectively. That defined as 2% of the total required workforce is $\bar{M} = 129.94$, 241.54 and 358.38. For the models with uncertain seasonal factor f_i, uncertainty set $\mathcal{P}_\beta$, and $\bar{M}$ as 1% (2%) of the total required workforce, Table 2 (Table 3) displays for each type of estimate probability distribution, the four evolutional criterions mentioned above. Table 4 (Table 5) has a similar structure, but it is related to models with uncertainty set $\mathcal{P}_k$.

In order to examine the trade-off between the salary cost and the protection against risk, we consider for DR different values of β (or k), which correspond to uncertainty sets with different sizes. The higher the β (or k) value, the higher the degree of protection against the uncertainty in probability distribution. An extreme case can be considered, namely $\beta = 0$ (or $k = 0$), which can be viewed as equivalent to SP. For information, given the uncertainty set $\mathcal{P}_\beta$ ($\mathcal{P}_k$) chosen in the following tables, we have observed the percentage that the sampled true probability distribution p falls outside the uncertainty set. We find that almost 100% of p falls outside the uncertainty set $\mathcal{P}_\beta$ ($\mathcal{P}_k$), but numeric results show that not all of them lead to constraint violation.

Table 4. Models with uncertain f_i, uncertainty set $\mathcal{P}_k$ and $\bar{M}$ is 1% of total required workforce.

k	Salary cost	Constraint violation (%)	Expectation $(M - \bar{M} \mid M > \bar{M})$	Worst case $M - \bar{M}$
Set A of probabilities q, $\bar{M} = 64.97$				
0	21,500.8	44.03	29.27	161.16
0.10	21,865.6	38.96	26.81	152.16
0.30	22,678.4	27.73	23.04	133.95
0.50	23,574.4	18.28	19.25	114.24
0.80	25,152.0	7.38	14.80	84.10
1.00	26,246.4	3.43	12.43	65.16
1.50	29,456.0	0.13	4.65	17.01
2.00	32,656.0	0.00	NaN	−18.91
Set B of probabilities q, $\bar{M} = 120.77$				
0	27,481.6	47.67	29.62	180.18
0.10	27,686.4	41.20	26.73	168.26
0.30	28,073.6	29.20	22.93	149.06
0.50	28,473.6	19.05	19.83	131.70
0.80	29,152.0	7.56	15.82	102.42
1.00	29,628.8	3.37	14.75	84.59
1.50	30,934.4	0.26	8.15	38.89
2.00	32,361.6	0.00	NaN	−2.81
Set C of probabilities q, $\bar{M} = 179.19$				
0	32,752.0	48.27	27.69	157.67
0.10	33,040.0	32.83	22.35	135.80
0.30	33,577.6	10.82	15.71	97.75
0.50	34,041.6	2.75	12.39	68.12
0.80	34,556.8	0.41	8.59	38.26
1.00	34,819.2	0.10	10.03	24.44
1.50	35,561.6	0.00	NaN	−10.70
2.00	36,403.2	0.00	NaN	−45.40

Table 5. Models with uncertain f_i, uncertainty set $\mathcal{P}_k$ and $\bar{M}$ is 2% of total required workforce.

k	Salary cost	Constraint violation (%)	Expectation $(M-\bar{M}\|M>\bar{M})$	Worst case $M-\bar{M}$
Set A of probabilities q, $\bar{M}=129.94$				
0	18,134.4	45.88	40.41	272.00
0.10	18,483.2	37.71	37.22	254.79
0.30	19,264.0	23.10	31.14	217.21
0.50	20,131.2	12.73	26.21	181.70
0.80	21,673.6	3.50	18.56	122.38
1.00	22,764.8	1.10	15.82	86.56
1.50	26,073.6	0.01	1.32	1.32
2.00	29,593.6	0.00	NaN	−53.38
Set B of probabilities q, $\bar{M}=241.54$				
0	24,438.4	47.41	44.78	236.15
0.10	24,640.0	39.95	40.03	218.17
0.30	25,059.2	25.20	33.18	184.33
0.50	25,513.6	13.28	27.51	149.05
0.80	26,249.6	3.38	19.22	97.65
1.00	26,771.2	0.90	15.76	67.33
1.50	28,195.2	0.00	NaN	−6.78
2.00	29,760.0	0.00	NaN	−69.16
Set C of probabilities q, $\bar{M}=358.38$				
0	29,891.2	48.43	45.44	255.78
0.10	30,137.6	34.78	37.37	226.47
0.30	30,608.0	13.31	27.59	173.67
0.50	31,043.2	3.86	22.42	128.40
0.80	31,619.2	0.45	15.53	72.91
1.00	31,932.8	0.10	19.82	44.92
1.50	32,803.2	0.00	NaN	−26.15
2.00	33,808.0	0.00	NaN	−96.50

From Table 2 to 5, we can observe a trade-off between the salary cost and the other three criterions which present the protection against risk. By increasing the β (or k) value, which increases the uncertainty set size, the constraint violation percentage, the conditional expectation of $(M-\bar{M})$, and the max case $(M-\bar{M})$ are eliminated progressively, with an increase in salary cost. Figure 3 shows the trade-off between salary cost and the constraint violation percentage. And Figure 4 shows the decreasing tendency of the other two criterions in total cost. As expected,

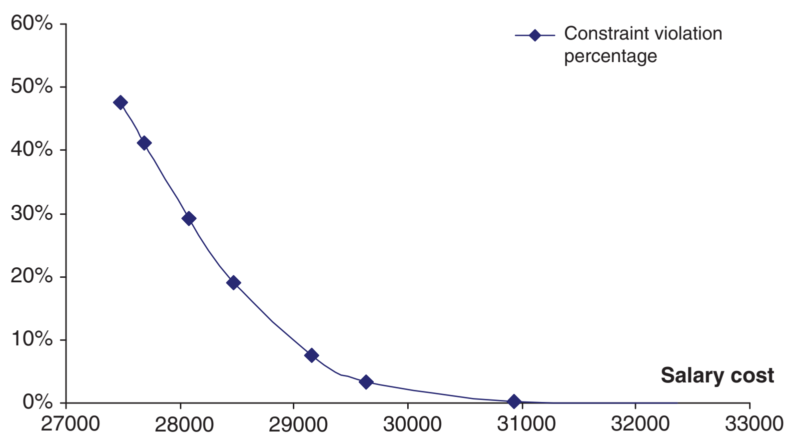

Figure 3. Trade-off between the salary cost and constraint violation percentage.

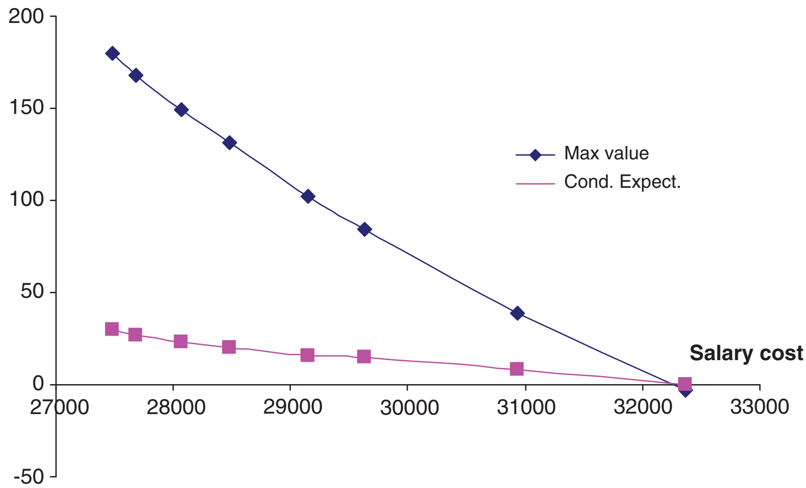

Figure 4. Trade-off between the salary cost and the max and conditional expected $(M - \bar{M})$.

Table 6. $\bar{M} = 0$, the upper bound for the salary cost.

Estimate prob. distribution	Uncertain f_i	
	DR salary	SP salary
A	48,956.8	48,956.8
B	48,956.8	48,956.8
C	48,956.8	48,956.8

SP has the lowest salary cost. However, the constraint violation percentage for the method SP is remarkable. For the three types of estimate probability distribution, the solutions of SP tend to violate the constraints by about half chance. The performance of DR is quite nice. For example, given $\beta = 0.2$, both Tables 2 and 3 show that, for the estimate probability distribution A, B and C, DR reduces the constraint violation percentage more than 33%, 34% and 40% by only increasing about 11%, 4% and 3% the salary cost, respectively. Similar remarks can be found from the results in Tables 4 and 5.

In general, the method SP which does not take into account the uncertainty on the probability distribution, leads to violation of constraint (infeasibility) with a quite high proportion. While the DR method we proposed avoids this trouble, by only paying a relatively small increase on the salary cost. This illustrates the necessity of taking into account the uncertainty on the probability distribution.

For both $\bar{M}$ equals to 1% and 2% of the required total workforce, we find similar performance for both SP and DR, as presented above. An extreme value of $\bar{M}$ is 0, with all M_{ikl} of both SP and DR equal to zero. Consequently, SP and DR behave the same. The salary cost is the upper bound for all further results. As $\bar{M}$ grows, it is likely that SP and DR diverge more and more. For both models with uncertainty set $\mathcal{P}_\beta$ and $\mathcal{P}_k$, Table 6 displays the upper bound salary cost for the three types of estimate probability. We observe that given the model with uncertain f_i, the upper bound costs are the same for the three types of probability distribution. The reason is simply that in our numeric example, the random variable Θ takes values from the same interval [0.00, 12.00], and all f_{ik} are defined by the same way, then the largest required agents number N_{ikl} are the same for the three types of probability distribution.

5. Conclusion

Our paper is essentially an enhancement of the stochastic programming formulation [24] of the call centre staffing problem to account for the uncertainty of the probabilities of the busyness factor Θ entering the computation of the expected understaffing. We have resorted to a distributionally robust formulation and proposed two different constructions for the uncertainty set. One is based on statistical confidence set; the other one does not make use of probabilistic arguments. Yet the second one makes it possible to derive a probabilistic bound. We submitted both approaches to a validation procedure by simulations of the busyness factor Θ. The results with the two approaches are very similar. They make it possible to build a tradeoff curve between the salary costs and various measures of satisfaction, e.g. average number of times the constraint is satisfied, conditional expectation of the understaffing, maximum understaffing. A good control of the constraint satisfaction is achieved with immunization factors that are significantly smaller than what the theory suggests. This 'conservatism' is often put as a drawback of Robust Optimization. One can object that it is always possible to play with one parameter only, the immunization factor, to achieve the proper goal. What theory brings in is a coherent design of the shape of the uncertainty set.

In our analysis the seasonal factors f_i have also been taken to be uncertain. We have considered that each factor can take different values with given probabilities. Under the assumption that the distribution of the seasonal factors is independent of the busyness factor Θ, the stochastic programming problem is still straightforward. A natural extension would be to consider that the probabilities attached to the realizations of the factors f_i are themselves uncertain. Again under the independence condition, the extension is straightforward with a moderate increase in the model complexity.

The stochastic programming problem may be refined. In the present paper, we have given an equal weight to the agent shortfall in each period. However, the number of required agents strongly differ from period to period. It is not clear that one unit of being short of one agent in a period with a large required staffing has the same weight than when the required staffing is small. To cope with this effect, one can add weights to the understaffing that depend on the values of the required agent number.

In this paper, we mainly focused on problems having shifts without breaks. Under this hypothesis, we could exploit the unimodularity of the shift matrix to make computations numerically easier. We have not yet explored large instances of problems for which this hypothesis does not hold. It is possible that commercial solvers have difficulty in solving the resulting MIP. If the problem instance has only few shifts with breaks and a majority of shifts without breaks, we could still exploit the unimodularity property as follows. If J_1 is the set of shifts without break and J_2 its complement. We could replace $\sum_{j\in J} a_{ij}x_j$ by $y_i + \sum_{j\in J_2} a_{ij}x_j$ and set $y_i = \sum_{j\in J_1} a_{ij}x_j$ with $i \in I$. The variables y_i and $x_j, j \in J_2$, are non-negative integer variable while $x_j, j \in J_1$, is simply non-negative. If this scheme is not implementable, we may have to resort to heuristics to solve the problem approximately. This is a subject of further research.

References

[1] Z. Aksin, M. Armony, and V. Mehrotra, *The modern call-center: A multi-disciplinary perspective on operations management research*, Prod. Oper. Manage. 16 (2007), pp. 665–688.
[2] A. Avramidis, A. Deslauriers, and P. L'Ecuyer, *Modeling daily arrivals to a telephone call center*, Manage. Sci. 50 (2004), pp. 896–908.
[3] F. Babonneau, J.-P. Vial, and R. Apparigliato, *Robust optimization for environmental and energy planning*, in *Handbook on 'Uncertainty and Environmental Decision Making'*, J. Filar and A. Haurie, eds., International Series in Operations Research and Management Science, Springer Verlag, New York, 2010, pp. 79–126.
[4] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res. 23 (1998), pp. 769–805.

[5] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2009.

[6] A. Ben-Tal, D. den Hertog, A. de Waegenaere, B. Melenberg, and G. Rennen, *Robust solutions of optimization problems affected by uncertain probabilities*, Working paper, Tilburg University, The Netherlands, 2011.

[7] J. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer Verlag, New York, 1997.

[8] M. Breton and S. El Hachem, *Algorithms for the solution of stochastic dynamic minimax problems*, Comput. Optim. Appl. 4 (1995), pp. 317–345.

[9] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao, *Statistical analysis of a telephone call center: A queueing-science perspective*, J. Amer. Statist. Assoc. 100 (2005), pp. 36–50.

[10] G.C. Calafiore, *Ambiguous risk measures and optimal robust portfolios*, SIAM J. Optim. 18(3) (2007), pp. 853–877.

[11] G. Calafiore and L. El Ghaoui, *On distributionally robust chance constrained linear programs with applications*, J. Optim. Theory Appl. 130(1) (2006), pp. 1–22.

[12] A. Charnes, W. Cooper, and G. Symonds, *Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil*, Manage. Sci. 4 (1958), pp. 235–263.

[13] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res. 58(3) (2010), pp. 595–612.

[14] J. Dupačová, *The minimax approach to stochastic programming and an illustrative application*, Stochastics 20 (1987), pp. 73–88.

[15] L. El-Ghaoui and H. Lebret, *Robust solutions to least-square problems to uncertain data matrices*, SIAM J. Matrix Anal. Appl. 18 (1997), pp. 1035–1064.

[16] L. El-Ghaoui, M. Oks, and F. Oustry, *Worst-case value-at-risk and robust portfolio optimization: A conic programming approach*, Oper. Res. 51 (2003), pp. 543–556.

[17] N. Gans, G. Koole, and A. Mandelbaum, *Telephone calls centers: A tutorial and literature review and research prospects*, Manuf. Serv. Oper. Manag. 5 (2003), pp. 79–141.

[18] L. Green, P. Kolesar, and W. Whitt, *Coping with time-varying demand when setting staffing requirements for a service system*, Prod. Oper. Manage. 16 (2007), pp. 13–39.

[19] D. Gross and C. Harris, *Fundamentals of Queueing Theory*, 3rd ed., Wiley Series in Probability and Mathematical Statistics, New York, 1998.

[20] J. Harrison and A. Zeevi, *A method for staffing large call centers based on stochastic fluid models*, Manuf. Serv. Oper. Manag. 7 (2005), pp. 20–36.

[21] T. Jimenez and G. Koole, *Scaling and comparison of fluid limits of queues applied to call centers with time-varying parameters*, OR Spectrum 26 (2004), pp. 413–422.

[22] D. Klabjan, D. Simchi-Levi, and M. Song, *Robust stochastic lot-sizing by means of histograms*, Working paper, Evanston, Illinois, 2010.

[23] P. Kouvelis and G. Yu, *Robust Discrete Optimization and its Applications*, Kuwer Academic Publishers, London, 1997.

[24] S. Liao, C. van Delft, G. Koole, and O. Jouini, *Staffing a call center with uncertain non-stationary arrival rate and flexibility*, Working paper, Ecole Centrale Paris, 2010, to appear in *OR Spectrum*.

[25] L. Pardo, *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC, Boca Raton, 2006.

[26] H. Scarf, *A min–max solution of an inventory problem*, in *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, CA, 1958, pp. 201–209.

[27] Z. Wang, P. Glynn, and Y. Ye, *Likelihood robust optimization for data-driven newsvendor problem*, Working paper, Department of Management Science and Engineering, Stanford University, USA, 2009.

[28] W. Whitt, *Dynamic staffing in a telephone call center aiming to immediately answer all calls*, Oper. Res. Lett. 24 (1999), pp. 205–212.

[29] J. Žáčková, *On minimax solution of stochastic linear programming problems*, Časopis pro Peštování 91 (1966), pp. 423–430.