Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X | EISSN 1526-5463 | 00 | 0000 | 0001

Integer Programming Approaches for Appointment Scheduling with Random No-shows and Service Durations

Ruiwei Jiang
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, ruiwei@umich.edu

Siqian Shen
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, siqian@umich.edu

Yiling Zhang
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, zyiling@umich.edu

We consider a single-server scheduling problem given a fixed sequence of appointment arrivals with random no-shows and service durations. The probability distribution of the uncertain parameters is assumed to be ambiguous and only the support and first moments are known. We formulate a class of distributionally robust (DR) optimization models that incorporate the worst-case expectation/conditional Value-at-Risk (CVaR) penalty cost of appointment waiting, server idleness, and overtime as the objective or constraints. Our models flexibly adapt to different prior beliefs of no-show uncertainty. We obtain exact mixed-integer nonlinear programming reformulations, and derive valid inequalities to strengthen the reformulations that are solved by decomposition algorithms. In particular, we derive convex hulls for special cases of no-show beliefs, yielding polynomial-sized linear programming models for the least and the most conservative supports of no shows. We test various instances to demonstrate the computational efficacy of our approaches, the results and solution performance of various DR models given perfect or ambiguous distributional information.

Key words: appointment scheduling; no-show uncertainty; distributionally robust optimization; mixed-integer programming; valid inequalities; totally unimodularity; convex hulls

1. Introduction

We consider an appointment scheduling problem that involves a single server and a set of appointments following a fixed order of arrivals. A system operator needs to schedule an arrival time for each appointment with random no-shows and service duration. This problem is fundamental for establishing service quality and operational efficiency in many service systems, and has been studied in the context of surgery planning in hospitals (see, e.g., Denton and Gupta 2003), call-center staffing (see, e.g., Gurvich et al. 2010), and cloud computing server operations (see, e.g., Shen and Wang 2014). Random no-shows are observed and analyzed in outpatient appointment scheduling (e.g., Lee et al. 2005, Berg et al. 2014), which may cause equipment and personnel idleness and losses of opportunities of serving other appointments. A common goal is to minimize the expected cost associated with appointment waiting time, server idle time, and overtime if the distributional information is fully accessible. In Section 1.1, we provide an extensive review of the literature on variants of stochastic appointment scheduling under specific objectives, metrics, and applications.

In reality, it is challenging to accurately estimate the probability distribution of no-shows and service durations. The data of no-shows could be limited because of low probability of occurrence and the heterogeneity of appointments. In view of a wide range of plausible substitutes (e.g., Log-Normal, Normal, and Uniform) for modeling the service-time uncertainty, one could easily misspecify its distribution. Then with ambiguous estimates of no-show and service-duration distributions, we could schedule unnecessarily long (respectively, short) time in between appointments, resulting in significant server idleness (respectively, appointment waiting or server overtime). To address the distributional ambiguity issue, Kong et al. (2013) propose a distributionally robust (DR) model using a cross-moment ambiguity set that consists of all distributions with common mean and covariance of the random service durations. They obtain a copositive cone programming reformulation and solve a semidefinite program to approximate the optimal results. The most relevant to this paper, Mak et al. (2015) study a DR model using a marginal-moment ambiguity set of the random service durations. They obtain tractable reformulations by successfully solving a nonconvex optimization problem based on a binary encoding of its feasible region.

In this paper, we generalize the DR appointment scheduling model in Mak et al. (2015) by incorporating heterogeneous no-shows and their distributional ambiguity. We aim to produce appointment schedules with good out-of-sample performance, even only given a few historical data. To the best of our knowledge, this paper is the first to consider both discrete (no-shows) and continuous (service durations) randomness for DR appointment scheduling. This generalization results in a challenging mixed-integer nonlinear program (MINLP), to which the approach by Mak et al. (2015) fails to solve. The main contribution of the paper is to derive effective integer programming approaches for solving the generalized DR model, including valid inequalities that effectively accelerate the computation of the MINLP (see our computational studies in Section 5). We also show that these valid inequalities recover the convex hulls for two important special cases, leading to polynomial-sized linear programming (LP) reformulations that are computationally tractable and can be implemented in desktop solvers to benefit practitioners.

1.1. Literature Review

The studies of stochastic appointment scheduling (see, e.g., Gupta and Denton 2008, Erdogan and Denton 2013, Berg et al. 2014) often assume uncertain service durations following known distributions. Denton and Gupta (2003) formulate a two-stage stochastic LP model for appointment scheduling and demonstrate that the optimal time intervals allocated in between appointments form a "dome shape" if the unit idleness costs are high relative to the unit waiting costs. Mittal et al. (2014), Begen and Queyranne (2011), Begen et al. (2012), and Ge et al. (2013) develop approximation algorithms for deriving near-optimal solutions to various stochastic or robust appointment scheduling problems. Pinedo (2012) provides a comprehensive survey of various scheduling problems including their models, theories, and applications.

Ho and Lau (1992) are among the first to take into account no-show uncertainty in scheduling problems. They propose a heuristic approach to double book the first two arrivals and subsequently schedule the remaining appointments. Erdogan and Denton (2013) incorporate no-shows into a stochastic LP model by Denton and Gupta (2003), and also discuss a stochastic dynamic programming variant of the problem. Cayirli and Veral (2003), Hassin and Mendel (2008), Liu et al. (2010), Robinson and Chen (2010) demonstrate the impact of no-shows on static and dynamic appointment scheduling, and discuss general policies to mitigate negative effects such as system idleness. A number of heuristic policies and approximation algorithms have been proposed to schedule appointments under uncertain no-shows (see, e.g., Muthuraman and Lawley 2008, Zeng et al. 2010, Cayirli et al. 2012, Lin et al. 2011, Luo et al. 2012, LaGanga and Lawrence 2012, Zacharias and Pinedo 2014, Parizi and Ghate 2015, Kong et al. 2016). To our best knowledge, no papers have handled no-shows in a DR framework, which could lead to intractable models due to the discrete nature of 0-1 no-shows (see our computational studies later in Section 5.)

In this paper, we assume a fixed sequence of appointment arrivals. We refer to Denton et al. (2007), Gupta and Denton (2008), Mak et al. (2015), Mancilla (2009), Mak et al. (2014), He et al. (2015) for studies that also involve sequencing decisions, and Denton et al. (2010), Gurvich et al. (2010), Shylo et al. (2012) for studies that optimize server allocation under random service durations. For generic DR optimization using moment-based ambiguity sets, we refer to Scarf et al. (1958), Bertsimas and Popescu (2005), Bertsimas et al. (2010), Delage and Ye (2010).

1.2. Contributions of the Paper

We summarize the main contributions of this paper as follows.

1. Depending on system operators' risk preferences, we formulate DR models that incorporate the worst-case expected/conditional value-at-risk (CVaR) of waiting, idleness, and overtime costs as objective or constraints. Meanwhile, the DR models can flexibly adapt to different prior beliefs of the maximum number of consecutive no-shows, covering from the least conservative case (i.e., no consecutive no-shows) to the most conservative case (i.e., arbitrary no-shows).
2. We develop effective solution approaches for each DR model. The exact reformulations of the DR models result in mixed-integer trilinear programs. We linearize and derive valid inequalities

to strengthen the reformulations, which can significantly reduce computational time of solving various instances by using decomposition algorithms. For special no-show beliefs, our derivation leads to polynomial-sized LP reformulations that can readily be implemented in LP solvers.

3. We test diverse instances to show the computational efficacy and demonstrate the performance of DR models under various uncertainties and levels of conservativeness. We provide guidelines for choosing appropriate DR models and no-show beliefs, depending on historical no-show rates, computation budget, and targeted tradeoffs between quality of service and operational cost.

1.3. Structure of the Paper

The remainder of the paper is organized as follows. Section 2 formulates the DR expectation/CVaR models, and their variants based on different risk preferences. In Section 3, we derive an MINLP of the DR expectation model, as well as valid inequalities for accelerating a generic cutting-plane algorithm. In Section 4, we derive polynomial-sized LP reformulations for special cases of no-show beliefs. In Section 5, we test various instances to demonstrate the computational efficacy and solution performance of different DR models. Section 6 summarizes the paper and provides future research directions. In the e-companion (EC), we describe models and approaches for problems under a general waiting-time cost and a DR CVaR setting, respectively. We present all the proofs, as well as optimal solution patterns of different models.

Notation: The convex hull of a set X is denoted by $\text{conv}(X)$. The abbreviation "w.l.o.g." represents "without loss of generality." We follow the convention that $\sum_{k=i}^{j} a_k = 0$ if $i > j$.

2. Formulations of DR Appointment Scheduling

We consider n appointments arriving at a single server following a fixed order of arrivals given as $1, \ldots, n$. Each appointment i has a random service duration s_i. We interpret the possibility of random no-show for appointment i by a 0-1 Bernoulli random variable q_i such that $q_i = 1$ if appointee i shows up, and $q_i = 0$ otherwise. The goal is to optimize the decision of scheduling an arrival time for each appointment, or equivalently, assigning time intervals between appointments i and $i+1$ for all $i = 1, \ldots, n-1$.

2.1. Modeling Waiting, Idleness, and Overtime under Uncertainty

Let variable x_i represent the scheduled time interval between appointments i and $i+1$, $\forall i = 1, \ldots, n-1$. Under random no-shows and service durations, one or multiple of the following three scenarios can happen: (i) an appointment cannot start on time due to overtime operations of previous appointments, (ii) the server is idle and waiting for the next appointment due to an early finish or no-shows of previous appointments, and (iii) the server cannot finish serving all appointments within a given time limit, denoted by T. For all $i=1,\dots,n$, let variable w_i represent the waiting time of appointment i, and variable u_i represent the server idle time after finishing appointment i. Also, let variable W represent the server's overtime beyond the fixed time limit T to finish all n appointments. The feasible region of decision x is defined as

$$X=\left\{x: x_i\geq 0,\ \forall i=1,\dots,n,\ \sum_{i=1}^{n} x_i = T\right\}, \tag{1}$$

to ensure that we assign nonnegative time in between all consecutive appointments, and appointment n is scheduled to arrive before the end of the service horizon T. The dummy variable $x_n\geq 0$ represents $T-\sum_{i=1}^{n-1} x_i$, i.e., the time from the scheduled start of the last appointment to the server time limit.

Given decision $x\in X$ and a realization of uncertain parameters (q,s), the appointment waiting time $w=[w_1,\dots,w_n]^\mathsf{T}$ and server idleness $u=[u_1,\dots,u_n]^\mathsf{T}$ are given by

$$w_i=\max\{0, q_{i-1}s_{i-1}+w_{i-1}-x_{i-1}\},\ \text{and}\ u_{i-1}=\max\{0, x_{i-1}-q_{i-1}s_{i-1}-w_{i-1}\},\ \forall i=2,\dots,n. \tag{2}$$

We denote nonnegative parameters c_i^{w}, c_i^{u}, and c^{o} as unit penalty costs of waiting, idleness of appointment i, and server overtime, respectively, which satisfy $c_{i+1}^{\mathrm{u}}-c_i^{\mathrm{u}}\leq c_{i+1}^{\mathrm{w}}$ for all $i=1,\dots,n-1$. This assumption is standard (see Denton and Gupta (2003), Ge et al. (2013), Kong et al. (2013), Mak et al. (2015)). In fact, if this assumption does not hold and $c_{i+1}^{\mathrm{u}}>c_i^{\mathrm{u}}+c_{i+1}^{\mathrm{w}}$ for some i, then the system operator would rather enforce idleness even if appointment $i+1$ has arrived and keep it waiting, which is not realistic due to practical concerns. Under this assumption, we can formulate a linear program to compute the total cost of waiting, idleness, and overtime for given x,q,s:

$$Q(x,q,s) := \min_{w,u,W} \sum_{i=1}^{n}(c_i^{\mathrm{w}}w_i+c_i^{\mathrm{u}}u_i)+c^{\mathrm{o}}W \tag{3a}$$

$$\text{s.t.}\quad w_i-u_{i-1}=q_{i-1}s_{i-1}+w_{i-1}-x_{i-1}\quad \forall i=2,\dots,n \tag{3b}$$

$$W-u_n=q_ns_n+w_n-x_n \tag{3c}$$

$$w_i\geq 0,\ w_1=0,\ u_i\geq 0,\ W\geq 0,\quad \forall i=1,\dots,n. \tag{3d}$$

The objective function (3a) minimizes a linear cost function of the waiting, idleness, and overtime. Constraints (3b) yield either the waiting time of appointment i or the server's idle time after finishing appointment $i-1$, both of which will have the same solutions values as given by (2) (see Proposition 1 in Ge et al. (2013)). Similarly, constraint (3c) yields either the over time W or the idle time u_n. Since appointment 1 always arrives at time 0, we have $w_1=0$ and all the waiting, idleness, and overtime variables are nonnegative according to constraints (3d).

In (3), note that the waiting time costs $c_i^{\mathrm{w}}w_i$ are modeled from the perspective of servers (e.g., operating rooms). In particular, we assume that appointment no-shows take place after the server has been set up for serving the appointments. Hence, the waiting time costs stem from equipment and personnel idleness, as well as from the losses of opportunities of serving other appointments, and they are incurred regardless whether the appointments show up. From the perspective of appointments, the waiting time costs should be modeled as $c_i^{\mathrm{w}} w_i q_i$, i.e., the waiting time costs are waived if an appointment does not show up. In this paper, we focus on the DR models and solution methods for the former case, i.e., server-based waiting time costs. In EC.1, we elaborate how our DR approaches can adapt for a more general setting that incorporates both server-based and appointment-based waiting time costs.

2.2. Supports and Ambiguity Set

The classical stochastic appointment scheduling approaches seek an optimal $x \in X$ to minimize the expectation of random cost $Q(x,q,s)$ subject to uncertainty (q,s) with a known joint probability distribution denoted as $\mathbb{P}_{q,s}$. We assume that $\mathbb{P}_{q,s}$ is only known belonging to an ambiguity set $\mathcal{F}(D,\mu,\nu)$ that is determined by the support D of (q,s) and the mean values $\mu = [\mu_1,\ldots,\mu_n]^\mathsf{T}$ and $\nu = [\nu_1,\ldots,\nu_n]^\mathsf{T}$, where μ_i represents the mean $\mathbb{E}[s_i]$, and ν_i represents the average show-up probability $\mathbb{E}[q_i]$ of appointment i for each $i = 1,\ldots,n$. We consider support $D = D_q \times D_s$ where D_q is the support of random no-show parameter q and D_s is the support of random service duration parameter s. We assume upper and lower bounds of the duration of each appointment i and accordingly

$$D_s \ := \ \{s \geq 0 : s_i^{\mathrm{L}} \leq s_i \leq s_i^{\mathrm{U}}, \ \forall i = 1,\ldots,n\}.$$

The full support $D_q = \{q : q \in \{0,1\}^n\}$ contains all no-show scenarios. However, it often leads to over-conservative schedules. In this paper, we parameterize the no-show support by an integer $K \in \{2,\ldots,n+1\}$ such that $D_q = D_q^{(K)}$ rules out consecutive no-shows in any K consecutive appointments. Accordingly,

$$D_q^{(K)} \ := \ \left\{ q \in \{0,1\}^n : \sum_{j=i}^{i+K-1} q_j \geq 1, \ \forall i = 1,\ldots,n-K+1 \right\}.$$

Note that (i) when $K = 2$, $D_q^{(2)}$ rules out all consecutive no-shows, and (ii) when $K = n+1$, we have $D_q^{(n+1)} = \{0,1\}^n = D_q$ as the full support. Also, the parameterized supports $D_q^{(2)} \subset D_q^{(3)} \subset \cdots \subset D_q^{(n+1)}$ form a spectrum of conservativeness levels, with $D_q^{(2)}$ being the least conservative and $D_q^{(n+1)}$ being the most general/conservative. In practice, the system operator has the flexibility to select parameter K according to her targeted conservativeness, regardless whether the ruled-out realizations may still occur. The conservativeness refers to the trade-off between optimality and robustness (see, e.g., Ben-Tal and Nemirovski 2000, Bertsimas and Sim 2004). If we select $K = n+1$, then support $D_q^{(n+1)} \equiv \{0,1\}^n$ contains all possible no-show scenarios and so is most robust.

Meanwhile, $D_q = D_q^{(n+1)}$ leads to the largest value of $\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x,q,s)]$ among all $K \in \{2,\dots,n+1\}$. In this sense, $D_q^{(n+1)}$ is the most conservative. On the contrary, $D_q^{(2)}$ is the least conservative because it leads to the smallest values of $\mathbb{P}\{q \in D_q^{(K)}\}$ and $\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x,q,s)]$ among all $K \in \{2,\dots,n+1\}$.

Although $D_q^{(K)}$ with $K \neq n+1$ does not contain all possible no-shows, we can select a value of K such that $\mathbb{P}\{q \in D_q^{(K)}\}$ exceeds a sufficiently high probability. In Section 2.4, we provide a practical and rigorous guideline for how to select the value of K. More importantly, $D_q^{(K)}$ can lead to better out-of-sample performance. For example, the DR schedules given by $D_q^{(2)}$ outperform those using the full support in the out-of-sample simulations, in which arbitrary consecutive no-shows may still happen (see Section 5.4).

We specify the ambiguity set $\mathcal{F}(D,\mu,\nu)$ as

$$\mathcal{F}(D,\mu,\nu) := \left\{ \mathbb{P}_{q,s} \geq 0 : \begin{array}{l} \int_{D_q\times D_s} d\mathbb{P}_{q,s} = 1 \\ \int_{D_q\times D_s} s_i \; d\mathbb{P}_{q,s} = \mu_i \quad \forall i = 1,\dots,n \\ \int_{D_q\times D_s} q_i \; d\mathbb{P}_{q,s} = \nu_i \quad \forall i = 1,\dots,n \end{array} \right\}, \tag{4}$$

where $\mathbb{P}_{q,s}$ matches the mean values of service durations and no-shows. The ambiguity set $\mathcal{F}(D,\mu,\nu)$ does not incorporate higher moments (e.g., variance and correlations) of service time and no-shows for several reasons. First, with a small amount of data, it is often unclear whether/how the service time and no-shows are correlated. Second, the introduction of higher moments undermines the computational tractability of the DR models, which can be achieved by using $\mathcal{F}(D,\mu,\nu)$ and the solution algorithm derived later. Finally, as we find in the computational study (in Section 5), the DR models based on $\mathcal{F}(D,\mu,\nu)$ already provide near-optimal results, and the benefit of incorporating higher moments is not significant in our case.

2.3. DR Models with Different Risk Measures

We consider DR appointment scheduling models that impose a min-max DR objective and/or DR constraints. Specifically, given $x \in X$, we consider a risk measure $\varrho(Q(x,q,s))$ of $Q(x,q,s)$ where

(i) a *risk-neutral* system operator sets $\varrho(Q(x,q,s)) = \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x,q,s)]$, i.e., the expected total cost of waiting, idleness, and overtime;

(ii) a *risk-averse* system operator sets $\varrho(Q(x,q,s)) = \text{CVaR}_{1-\epsilon}(Q(x,q,s))$, i.e., the CVaR of the total cost with $1-\epsilon \in (0,1)$ confidence.

Then, the DR models impose a generic min-max DR objective in the form

$$\min_{x\in X} \sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \varrho(Q(x,q,s)), \tag{5a}$$

and/or generic DR constraints in the form

$$\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \varrho(Q(x,q,s)) \leq \overline{Q}. \tag{5b}$$

where $\overline{Q} \in \mathbb{R}$ represents a bounding threshold for the risk measure from above. Both DR objective (5a) and constraints (5b) protect the risk measure by hedging against all probability distributions in $\mathcal{F}(D,\mu,\nu)$. A DR model can impose either or both of DR objective (5a) and constraints (5b), and can use either expectation or CVaR as risk measures in (5a)–(5b), i.e., $\varrho(Q(x,q,s)) = \mathbb{E}_{\mathbb{P}_{q,s}}[Q(x,q,s)]$ or $\varrho(Q(x,q,s)) = \text{CVaR}_{1-\epsilon}(Q(x,q,s))$. Furthermore, the system operator can tune the cost parameters c_i^{w}, c_i^{u}, and c^{o} to let $Q(x,q,s)$ represent different consequences (e.g., performance metric, quality of service, resource opportunity cost, etc.) associated with waiting, idleness, and overtime in (5a)–(5b). For example, by setting $c^{\text{o}} = 1$ and $c_i^{\text{u}} = c_i^{\text{w}} = 0$ for all i, we use

$$\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \text{CVaR}_{1-\epsilon}(Q(x,q,s)) \le \overline{Q} \tag{6}$$

to constrain the CVaR of overtime W below threshold $\overline{Q}$. The CVaR constraints provide a safe guarantee on the performance metrics with high probabilities. For this particular cost parameter setting, constraint (6) guarantees that $\inf_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \mathbb{P}_{q,s}\left\{W \le \overline{Q}\right\} \ \ge\ 1-\epsilon$, i.e., the overtime W is controlled under threshold $\overline{Q}$ with the smallest possible probability being no less than $1-\epsilon$. This provides the service team with an appropriate "end-of-the-day" guarantee (see, e.g., Shylo et al. 2012, Wang et al. 2014, Zhang et al. 2015). For presentation brevity, we have relegated the discussions on the CVaR-based model to EC.2.

2.4. Guideline of Selecting Parameter K

In practice, a system operator may evaluate the probability of the random variables $q = (q_1,\dots,q_n)$ belonging to $D_q^{(K)}$, i.e., $\mathbb{P}(q \in D_q^{(K)})$. Then, she can select a value of K such that $\mathbb{P}(q \in D_q^{(K)})$ exceeds a given threshold such as 90%. To this end, she can gradually increase K from 2 until $\mathbb{P}(q \in D_q^{(K)})$ exceeds the threshold for the first time.

Observation 1. *$\mathbb{P}(q \in D_q^{(K)}) = 1$ if $K > n$. If $2 \le K \le n$ and the components of q are jointly independent, then $\mathbb{P}(q \in D_q^{(K)}) = 1 - [Q^n]_{1,m(K)+1}$, where $m(i) := \frac{1}{2}i(2n-i+3)$ for $i = 0,\dots,K$ and Q represents a $(m(K)+1)\times(m(K)+1)$ matrix such that (i) $Q_{m(i)+j,i+j+1} = \nu_{i+j}$ for all $i = 0,\dots,K-1$ and $j = 1,\dots,n-i$, (ii) $Q_{m(i)+j,m(i+1)+j} = 1-\nu_{i+j}$ for all $i = 0,\dots,K-2$ and $j = 1,\dots,n-i$, (iii) $Q_{m(K-1)+j,m(K)+1} = 1-\nu_{K-1+j}$ for all $j = 1,\dots,n-K+1$, and (iv) $Q_{m(K)+1,m(K)+1} = Q_{m(i),m(i)} = 1$ for all $i = 1,\dots,K$.*

Proof of Observation 1 If $K > n$, it is clear that we cannot have K consecutive no-shows and so $\mathbb{P}(q \in D_q^{(K)}) = 1$. If $2 \le K \le n$, we construct a Markov chain with $m(K)+1$ states, where state $m(i)+j$ represents i consecutive no-shows in the first $i+j-1$ appointments for all $i = 0,\dots,K-1$ and $j = 1,\dots,n-i+1$, and state $m(K)+1$ represents that K consecutive no-shows happen. By construction, matrix Q is the one-step transition matrix of this Markov chain, where states

$m(K)+1$ and $m(i)$ for all $i=1,\ldots,K$ are absorbing. Thus, $[Q^n]_{1,m(K)+1}$, representing component $(1,m(K)+1)$ of matrix Q^n, equals to the probability of having K consecutive no-shows in the n appointments. $\square$

Using Observation 1, the selection of K can be conveniently done in a spreadsheet. In Figure 1, we display an example of $\mathbb{P}(q\in D_q^{(K)})$ with $n=10$, $K=1,\ldots,11$, and $\nu_i=0.1,\ldots,$ or 0.9 for all $i=1,\ldots,10$. We observe that $K=2$ is sufficient for $\mathbb{P}(q\in D_q^{(K)})\geq 90\%$ when $\nu_i\leq 0.1$, i.e., when the no-show probability for each appointment is no greater than 0.1. This observation motivates us to select $D_q=D_q^{(2)}$ for scheduling appointments with low no-show probabilities.

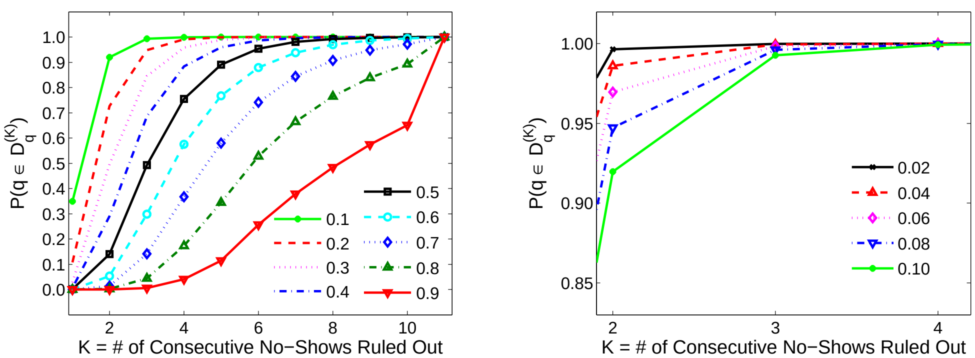


Figure 1 An example of $\mathbb{P}(q\in D_q^{(K)})$ for $n=10$ appointments

Next, we develop reformulations and solution methods. For presentation brevity, we only analyze DR expectation models in Sections 3 and 4. We present the results of DR CVaR models in EC.2. All the proofs are organized in EC.3 (for DR expectation models) and EC.4 (for DR CVaR models).

3. Cutting-Plane Approach and Valid Inequalities for DR Expectation Models

We analyze the DR expectation models by specifying a generic objective form (5a) as

$$\min_{x\in X}\ \sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)}\ \mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)\right], \tag{7}$$

which minimizes the worst-case expected cost of waiting, idleness, and overtime. We first consider the inner maximization problem $\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)}\ \mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)\right]$ for a fixed $x\in X$, where $\mathbb{P}_{q,s}$ is the decision variable. It can be detailed as a linear functional optimization problem

$$\max_{\mathbb{P}_{q,s}\geq 0}\ \int_{D_q\times D_s} Q(x,q,s)d\mathbb{P}_{q,s} \tag{8a}$$

$$\text{s.t.}\ \int_{D_q\times D_s} s_i\ d\mathbb{P}_{q,s}=\mu_i\quad \forall i=1,\ldots,n \tag{8b}$$

$$\int_{D_q\times D_s} q_i\ d\mathbb{P}_{q,s}=\nu_i\quad \forall i=1,\ldots,n \tag{8c}$$

$$\int_{D_q \times D_s} d\mathbb{P}_{q,s} = 1, \tag{8d}$$

where $D_q = D_q^{(K)}$ for some $K \in \{2, \ldots, n+1\}$. Letting ρ_i, γ_i, and θ be dual variables associated with constraints (8b), (8c), and (8d), respectively, we present problem (8) in its dual form as

$$\min_{\rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n, \theta \in \mathbb{R}} \quad \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \theta \tag{9a}$$

$$\text{s.t.} \quad \sum_{i=1}^n s_i \rho_i + \sum_{i=1}^n q_i \gamma_i + \theta \geq Q(x, q, s) \quad \forall (q, s) \in D_q \times D_s. \tag{9b}$$

Here $\rho = [\rho_1, \ldots, \rho_n]^\mathsf{T}$, $\gamma = [\gamma_1, \ldots, \gamma_n]^\mathsf{T}$, and θ are unrestricted, and (9b) are associated with primal variables $\mathbb{P}_{q,s}, \ \forall (q, s) \in D_q \times D_s$. Under the standard assumptions that μ_i belongs to the interior of set $\{\int_{D_q \times D_s} s_i \ d\mathbb{Q}_{q,s} : \mathbb{Q}_{q,s}$ is a probability distribution over $D_q \times D_s\}$, and that ν_i belongs to the interior of set $\{\int_{D_q \times D_s} q_i \ d\mathbb{Q}_{q,s} : \mathbb{Q}_{q,s}$ is a probability distribution over $D_q \times D_s\}$ for each appointment i, strong duality holds between (8) and (9). Furthermore, for a fixed (ρ, γ, θ), constraints (9b) are equivalent to $\theta \geq \max_{(q,s) \in D_q \times D_s} \{Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma q_i)\}$. Thus, due to the objective of minimizing θ, the dual formulation (9) is equivalent to

$$\min_{\rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n} \left\{ \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \max_{(q,s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \right\}. \tag{10}$$

3.1. MINLP Reformulation and a Generic Cutting-Plane Approach

Note that $Q(x, q, s)$ is a minimization problem and thus in (10) we have an inner max-min problem. We next analyze the structure of $Q(x, q, s)$ for given solution x and realized value of (q, s). We formulate $Q(x, q, s)$ in (3) in its dual form as

$$Q(x, q, s) \ = \ \max_y \quad \sum_{i=1}^n (q_i s_i - x_i) y_i \tag{11a}$$

$$\text{s.t.} \quad y_{i-1} - y_i \leq c_i^{\mathrm{w}} \quad \forall i = 2, \ldots, n \tag{11b}$$

$$-y_i \leq c_i^{\mathrm{u}} \quad \forall i = 1, \ldots, n \tag{11c}$$

$$y_n \leq c^{\mathrm{o}}, \tag{11d}$$

where variable y_{i-1} represents the dual associated with each constraint i in (3b) for all $i = 2, \ldots, n$, and variable y_n represents the dual of constraint (3c). Constraints (11b), (11c), (11d) are related to primal variables $w_i, \ i = 2, \ldots, n, \ u_i, \ i = 1, \ldots, n$, and W in (3), respectively. Therefore, formulation (10) is equivalent to

$$\min_{\rho, \gamma} \left\{ \sum_{i=1}^n \mu_i \rho_i + \sum_{i=1}^n \nu_i \gamma_i + \max_{(q,s) \in D_q \times D_s} \left\{ Q(x, q, s) - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i) \right\} \right\} \tag{12a}$$

$$= \min_{\rho,\gamma} \left\{ \sum_{i=1}^{n} \mu_i \rho_i + \sum_{i=1}^{n} \nu_i \gamma_i + \max_{y \in Y} \; h(x, y, \rho, \gamma) \right\}, \tag{12b}$$

where Y represents the feasible region of variable y in (11) given by (11b)–(11d), and

$$h(x, y, \rho, \gamma) := \max_{(q,s) \in D_q \times D_s} \left\{ \sum_{i=1}^{n} (q_i s_i - x_i) y_i - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \right\}. \tag{12c}$$

The derivation of $h(x, y, \rho, \gamma)$ follows that we can interchange the order of $\max_{(q,s)\in D_q \times D_s}$ and $\max_{y \in Y}$ in (12a). Combining the inner problem in the form of (12b) with the outer minimization problem in (7), we derive a reformulation of the DR expectation model (7) as:

$$\min_{x \in X, \rho, \gamma, \delta} \quad \sum_{i=1}^{n} \mu_i \rho_i + \sum_{i=1}^{n} \nu_i \gamma_i + \delta \tag{13a}$$

$$\text{s.t.} \quad \delta \; \geq \; \max_{y \in Y} \; h(x, y, \rho, \gamma) \equiv \max_{y \in Y, (q,s) \in D_q \times D_s} \left\{ \sum_{i=1}^{n} (q_i s_i - x_i) y_i - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \right\}. \tag{13b}$$

Next, we analyze structural properties of $\max_{y \in Y} h(x, y, \rho, \gamma)$ as a function of variables x, ρ, and γ.

Lemma 1. *For any fixed variables x, ρ, and γ, $\max_{y \in Y} h(x, y, \rho, \gamma) < +\infty$. Furthermore, function $\max_{y \in Y} h(x, y, \rho, \gamma)$ is convex and piecewise linear in x, ρ, and γ with a finite number of pieces.*

We refer to EC.3.1 for a detailed proof. Lemma 1 indicates that constraint (13b) essentially describes the epigraph of a convex and piecewise linear function of decision variables in model (13). This observation facilitates us applying a separation-based decomposition algorithm to solve formulation (13) (or equivalently, the DR expectation model (7)), presented in Algorithm 1. This algorithm is finite because we identify a new piece of the function $\max_{y \in Y} h(x, y, \rho, \gamma)$ each time when the set $\{L(x, \rho, \gamma, \delta) \geq 0\}$ is augmented in Step 7, and the function has a finite number of pieces according to Lemma 1.

The main difficulty of the above decomposition algorithm lies in solving the separation problem (14). In general, this problem is a mixed-integer trilinear program because of the integrality restrictions on variables q_i and the trilinear terms $q_i s_i y_i$ in the objective function. This creates obstacles for optimally solving the separation problem if presented in its current form. In Section 3.2, we linearize and reformulate the separation problem (14) as a mixed-integer linear program (MILP) that can readily be solved by optimization solvers. Moreover, we will derive valid inequalities to strengthen this MILP, and test their computational efficiency later.

3.2. MILP Reformulation of the Separation Problem and Valid Inequalities

Our approach is inspired by Mak et al. (2015), where the authors point out that an optimal solution y^* to a similar separation problem but not involving no-shows, exists at an extreme point of

Algorithm 1 A decomposition algorithm for solving DR expectation model (7).

1: Input: feasible regions X, Y, and $D_q \times D_s$; set of cuts $\{L(x,\rho,\gamma,\delta)\geq 0\} = \emptyset$.

2: Solve the master problem

$$\begin{aligned} \min_{x\in X,\rho,\gamma,\delta} \quad & \sum_{i=1}^{n}\mu_i\rho_i + \sum_{i=1}^{n}\nu_i\gamma_i + \delta \\ \text{s.t.} \quad & L(x,\rho,\gamma,\delta)\geq 0 \end{aligned}$$

and record an optimal solution $(x^*,\rho^*,\gamma^*,\delta^*)$.

3: With (x,ρ,γ) fixed to be (x^*,ρ^*,γ^*), solve the separation problem

$$\max_{y\in Y} h(x,y,\rho,\gamma) \equiv \max_{y\in Y,(q,s)\in D_q\times D_s}\left\{\sum_{i=1}^{n}(q_is_i - x_i)y_i - \sum_{i=1}^{n}(\rho_is_i + \gamma_iq_i)\right\} \tag{14}$$

and record an optimal solution (y^*,q^*,s^*).

4: **if** $\delta^* \geq \sum_{i=1}^{n}(q_i^*s_i^* - x_i^*)y_i^* - \sum_{i=1}^{n}(\rho_i^*s_i^* + \gamma_i^*q_i^*)$ **then**

5: stop and return x^* as an optimal solution to formulation (7).

6: **else**

7: add the cut $\delta \geq \sum_{i=1}^{n}(q_i^*s_i^* - x_i)y_i^* - \sum_{i=1}^{n}(s_i^*\rho_i + q_i^*\gamma_i)$ to the set of cuts $\{L(x,\rho,\gamma,\delta)\geq 0\}$ and go to Step 2.

8: **end if**

polyhedron Y. They then successfully decompose the separation problem by appointment for each $i=1,\ldots,n$ and reformulate it by using the extreme points of Y. Different in this paper, for fixed x, ρ, and γ, our separation problem is a mixed-integer trilinear program involving binary variables $q_i,\ i=1,\ldots,n$. Moreover, except for the case $D_q = D_q^{(n+1)} = \{0,1\}^n$, $h(x,y,\rho,\gamma)$ is not decomposable by appointment in view of the cross-appointment nature of D_q. Therefore, the approach in Mak et al. (2015) is no longer applicable, and $\max_{y\in Y} h(x,y,\rho,\gamma)$ becomes much more challenging.

Our analysis consists of the following steps. We start by showing the convexity of $h(x,y,\rho,\gamma)$ in variable y. Then, it follows from fundamental convex analysis that maximizing convex function $h(x,y,\rho,\gamma)$ on polyhedron Y will yield an optimal solution at one of the extreme points of Y. Also considering the cost of idleness, we extend the result of extreme-point representation in Mak et al. (2015) and reformulate the separation problem (14) using a polynomial number of binary variables to replace the continuous variables $y_i,\ i=1,\ldots,n$.

LEMMA 2. *For fixed x, ρ, and γ, function $h(x,y,\rho,\gamma)$ is convex in variable y.*

We refer to EC.3.2 for a proof. According to Lemma 2, an optimal solution y^* to the separation problem (14) exists at one of the extreme points of Y having linear constraints (11b)–(11d). Consider

$$Y = \left\{c^{\mathrm{o}} \geq y_n \geq -c_n^{\mathrm{u}},\ y_n + c_n^{\mathrm{w}} \geq y_{n-1} \geq -c_{n-1}^{\mathrm{u}}, \cdots, y_2 + c_2^{\mathrm{w}} \geq y_1 \geq -c_1^{\mathrm{u}}\right\}. \tag{15}$$

It can be observed that any extreme point $\hat{y}$ of Y satisfy (i) either $\hat{y}_n = -c^{\mathrm{u}}_n$ or $\hat{y}_n = c^{\mathrm{o}}$, and (ii) for all $i = 1, \ldots, n-1$, dual constraint $\hat{y}_{i+1} + c^{\mathrm{w}}_{i+1} \geq \hat{y}_i \geq -c^{\mathrm{u}}_i$ is binding at either the lower bound or the upper bound.

This observation motivates us to establish an alternative formulation of (14) using new binary variables. For notation convenience, we define a dummy variable y_{n+1}, which always takes the lower-bound value $-c^{\mathrm{u}}_{n+1} := 0$. There is a one-to-one correspondence between an extreme point of Y and a partition of the integers $1, \ldots, n+1$ into intervals. For each interval $\{k, \ldots, j\} \subseteq \{1, \ldots, n+1\}$ in the partition, y_j takes on the lower bound value $-c^{\mathrm{u}}_j$ and other y_i equal to their upper bounds, i.e., $y_i = y_{i+1} + c^{\mathrm{w}}_{i+1}$, $\forall i = k, \ldots, j-1$. As a result, for each interval $\{k, \ldots, j\}$ in the partition and $i \in \{k, \ldots, j\}$, the value of y_i is given by:

$$y_i = \pi_{ij} := \begin{cases} -c^{\mathrm{u}}_j + \sum_{\ell=i+1}^{j} c^{\mathrm{w}}_\ell & 1 \leq i \leq j \leq n, \\ c^{\mathrm{o}} + \sum_{\ell=i+1}^{n} c^{\mathrm{w}}_\ell & 1 \leq i \leq n,\ j = n+1, \end{cases} \tag{16}$$

and $y_{n+1} = \pi_{n+1,n+1} := 0$. Define binary variables t_{kj} for all $1 \leq k \leq j \leq n+1$, such that $t_{kj} = 1$ if interval $\{k, \ldots, j\}$ belongs to the partition (i.e., $t_{kj} = 1$ if $y_i = \pi_{ij}$) and $t_{kj} = 0$ otherwise. For a valid partition, we require each index i belonging to exactly one interval, and thus $\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1$, $\forall i = 1, \ldots, n+1$. For notation convenience, we define $x_{n+1} = q_{n+1} = s_{n+1} := 0$. Using binary variables t_{kj}, we reformulate the separation problem (14) as

$$\max_{t} \max_{(q,s) \in D_q \times D_s} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(\sum_{i=k}^{j} (q_i s_i - x_i) \pi_{ij} \right) t_{kj} - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \tag{17a}$$

$$\text{s.t.} \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i = 1, \ldots, n+1 \tag{17b}$$

$$t_{kj} \in \{0, 1\}, \qquad \forall 1 \leq k \leq j \leq n+1. \tag{17c}$$

Note that the objective function (17a) contains trilinear terms $q_i s_i t_{kj}$ with binary variables q_i, t_{kj}, and continuous variables s_i. To linearize formulation (17), we define $p_{ikj} \equiv q_i t_{kj}$ and $o_{ikj} \equiv q_i s_i t_{kj}$ for all $1 \leq k \leq j \leq n+1$ and $k \leq i \leq j$. Also, we introduce the following McCormick inequalities (18a)–(18b) and (18c)–(18d) for variables p_{ikj} and o_{ikj}, respectively.

$$p_{ikj} - t_{kj} \leq 0, \tag{18a}$$

$$p_{ikj} - q_i \leq 0,\ p_{ikj} - q_i - t_{kj} \geq -1,\ p_{ikj} \geq 0, \tag{18b}$$

$$o_{ikj} - s^{\mathrm{L}}_i p_{ikj} \geq 0,\ o_{ikj} - s^{\mathrm{U}}_i p_{ikj} \leq 0, \tag{18c}$$

$$o_{ikj} - s_i + s^{\mathrm{L}}_i (1 - p_{ikj}) \leq 0,\ o_{ikj} - s_i + s^{\mathrm{U}}_i (1 - p_{ikj}) \geq 0. \tag{18d}$$

Thus, the separation problem (14) is equivalent to an MILP as:

$$\max_{t,q,s,p,o} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^{j} (\pi_{ij} o_{ikj} - x_i \pi_{ij} t_{kj}) - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \tag{19a}$$

$$\text{s.t.}\quad (17b)\text{–}(17c),\ (18a)\text{–}(18d), \tag{19b}$$

$$s_i \in [s_i^{\mathrm{L}}, s_i^{\mathrm{U}}],\ q \in D_q \subseteq \{0,1\}^n. \tag{19c}$$

We can replace Steps 3–8 of Algorithm 1 proposed in Section 3 based on this MILP reformulation:

3: With (x,ρ,γ) fixed to be (x^*,ρ^*,γ^*), solve formulation (19) and record an optimal solution (t^*,q^*,s^*,p^*,o^*).
4: **if** $\delta^* \geq \sum_{k=1}^{n+1}\sum_{j=k}^{n+1}\sum_{i=k}^{j}(\pi_{ij}o^*_{ikj} - x^*_i\pi_{ij}t^*_{kj}) - \sum_{i=1}^{n}(\rho^*_i s^*_i + \gamma^*_i q^*_i)$ **then**
5: stop and return x^* as an optimal solution to formulation (7).
6: **else**
7: add the cut $\delta \geq \sum_{k=1}^{n+1}\sum_{j=k}^{n+1}\sum_{i=k}^{j}\left(\pi_{ij}o^*_{ikj} - \pi_{ij}t^*_{kj}x_i\right) - \sum_{i=1}^{n}(s^*_i\rho_i + q^*_i\gamma_i)$ to the set of cuts $\{L(x,\rho,\gamma,\delta)\geq 0\}$ and go to Step 2.
8: **end if**

Remark 1. We note that Algorithm 1 applies to various types of no-show support D_q. For example, we can specify $D_q = \{q \in \{0,1\}^n : \sum_{i=1}^{n}(1-q_i) \leq Q_{\max}\}$, where $Q_{\max}$ represents the maximum number of no-shows. In this case, we only need to replace the definition of D_q in (19c) when applying Algorithm 1. In fact, Algorithm 1 is general regardless of the specific form of set D_q, to select which we take into account the operator's beliefs and/or preferences, and the computational tractability. In this paper, we specify $D_q = D_q^{(K)}$ due to its flexibility (see Section 2.4) and computational tractability (see Proposition 1 and Theorem 2).

We further identify a set of valid inequalities to strengthen formulation (19). We summarize the valid inequalities in the following proposition and delegate its proof in EC.3.3. The inequalities (20a)–(20f) can be added to the MILP (19) solved in Step 3, to strengthen the reformulation.

Proposition 1. *The following inequalities are valid for set* $F = \{(t,q,s,p,o) : (19b)\text{–}(19c)\}$*:*

$$\sum_{k=1}^{i}\sum_{j=i}^{n+1} p_{ikj} \;=\; q_i \quad \forall i = 1,\ldots,n+1, \tag{20a}$$

$$s_i - \sum_{k=1}^{i}\sum_{j=i}^{n+1}(o_{ikj} - s_i^L p_{ikj}) \;\geq\; s_i^L \quad \forall 1 \leq i \leq n+1, \tag{20b}$$

$$s_i - \sum_{k=1}^{i}\sum_{j=i}^{n+1}(o_{ikj} - s_i^U p_{ikj}) \;\leq\; s_i^U \quad \forall 1 \leq i \leq n+1, \tag{20c}$$

$$\sum_{\ell=i}^{i+K-1} p_{\ell kj} \;\geq\; t_{kj} \quad \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-K+1, \tag{20d}$$

$$\sum_{k=1}^{i-K+2}\sum_{\ell=i-K+2}^{i} p_{\ell ki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \;\geq\; \sum_{k=1}^{i-K+2} t_{ki} \quad \forall i = K-1,\ldots,n, \tag{20e}$$

$$\sum_{k=1}^{i} p_{iki} + \sum_{\ell=i+1}^{i+K-1} \sum_{j=i+K-1}^{n+1} p_{\ell(i+1)j} \ \geq \sum_{j=i+K-1}^{n+1} t_{(i+1)j} \quad \forall i = 1, \ldots, n-K+2. \tag{20f}$$

Remark 2. Note that the above inequalities hold valid for all $K = 2, \ldots, n+1$. We also note two features that (i) valid inequalities (20a)–(20f) are *polynomially* many, and (ii) all coefficients of these inequalities are in *closed-form*. Features (i) and (ii) can significantly accelerate Algorithm 1, because Feature (i) ensures that model (14) (i.e., (19) after reformulation) remains small by incorporating these inequalities, and Feature (ii) implies that we do not need to separate these inequalities.

4. LP Reformulations of the DR Expectation Model

In this section, we present the main results of the paper as the derivation of convex hulls of the separation problem (14) for $D_q = D_q^{(2)}$ (i.e., no conservative no-shows) and $D_q = D_q^{(n+1)}$ (i.e., arbitrary no-shows). This leads to polynomial-sized LP reformulations of the DR expectation model (7).

Case 1. (No Consecutive No-Shows) Recall that F represents the mixed-integer feasible region of formulation (19), i.e., $F = \{(t, q, s, p, o) : \text{(19b)–(19c)}\}$. We show that the valid inequalities identified in Proposition 1 are sufficient to describe conv(F). We first notice that when $K = 2$: (i) inequalities (20d) are equivalent to $p_{ikj} + p_{(i+1)kj} \geq t_{kj}$ for all $1 \leq k < j \leq n+1$ and $k \leq i \leq j-1$, and (ii) inequalities (20e) and (20f) are identical and equivalent to

$$\sum_{k=1}^{i} p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \ \geq \sum_{k=1}^{i} t_{ki} \quad \forall i = 1, \ldots, n. \tag{21}$$

This leads to the following theorem, of which a proof is relegated to EC.3.4.

Theorem 1. *Polyhedron* $CF := \{(t, q, s, p, o) : \text{(17b), (18a), (18c), (20a)–(20d), (21)}\}$ *is the convex hull of set* F, *i.e.,* $CF = conv(F)$.

Therefore, we can reformulate the separation problem (14) as an LP model:

$$\begin{aligned} \max_{t,q,s,p,o} \quad & \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \sum_{i=k}^{j} (\pi_{ij} o_{ikj} - x_i \pi_{ij} t_{kj}) - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \\ \text{s.t.} \quad & (t, q, s, p, o) \in CF. \end{aligned}$$

To combine the separation problem with the outer minimization problem in (13), we present the above reformulation in its dual form:

$$\min \ \sum_{i=1}^{n+1} (\alpha_i + s_i^{\text{U}} \tau_i^{\text{U}} - s_i^{\text{L}} \tau_i^{\text{L}}) \tag{22a}$$

$$
\begin{aligned}
\text{s.t.}\quad & \sum_{i=k}^{j}(\alpha_i-\sigma_{ikj})+\sum_{i=k}^{j-1}\lambda_{ikj}+\sum_{i=j}^{\min\{j,n\}}\phi_i \geq -\sum_{i=k}^{j}\pi_{ij}x_i && \forall 1\leq k\leq j\leq n+1, && \text{(22b)}\\
& \zeta_i\leq\gamma_i && \forall 1\leq i\leq n, && \text{(22c)}\\
& \tau_i^{\mathrm{L}}-\tau_i^{\mathrm{U}}\leq\rho_i && \forall 1\leq i\leq n, && \text{(22d)}\\
& \sigma_{ikj}+s_i^{\mathrm{L}}\varphi_{ikj}^{\mathrm{L}}-s_i^{\mathrm{U}}\varphi_{ikj}^{\mathrm{U}}+\zeta_i-s_i^{\mathrm{L}}\tau_i^{\mathrm{L}}+s_i^{\mathrm{U}}\tau_i^{\mathrm{U}}-\sum_{\ell=\max\{k,i-1\}}^{\min\{j-1,i\}}\lambda_{\ell kj}\\
&\qquad -\sum_{\ell=\min\{2i-k-1,i\}\vee 1}^{\max\{2i-j,i-1\}\wedge n}\phi_\ell\geq 0 && \forall 1\leq k\leq j\leq n+1,\forall k\leq i\leq j, && \text{(22e)}\\
& -\varphi_{ikj}^{\mathrm{L}}+\varphi_{ikj}^{\mathrm{U}}+\tau_i^{\mathrm{L}}-\tau_i^{\mathrm{U}}\geq\pi_{ij} && \forall 1\leq k\leq j\leq n+1,\forall k\leq i\leq j, && \text{(22f)}\\
& \varphi_{ikj}^{\mathrm{L}},\varphi_{ikj}^{\mathrm{U}},\tau_i^{\mathrm{L}},\tau_i^{\mathrm{U}},\lambda_{ikj},\phi_i,\sigma_{ikj}\geq 0 && \forall 1\leq k\leq j\leq n+1,\forall k\leq i\leq j, && \text{(22g)}
\end{aligned}
$$

where we denote $a\vee b:=\max\{a,b\}$ and $a\wedge b:=\min\{a,b\}$ for $a,b\in\mathbb{R}$ for notation convenience. Here the dual variables α_i, σ_{ikj}, $\varphi_{ikj}^{\mathrm{L/U}}$, ζ_i, $\tau_i^{\mathrm{L/U}}$, λ_{ikj}, and ϕ_i are associated with constraints (17b), (18a), (18c), (20a), (20b)–(20c), (20d), and (21) respectively (after transforming all "$\geq$" inequalities into the "$\leq$" form), and constraints (22b)–(22f) are associated with primal variables t_{kj}, q_i, s_i, p_{ikj}, and o_{ikj} respectively. In (22b), the term $\sum_{i=j}^{\min\{j,n\}}\phi_i$ becomes ϕ_j for all $1\leq j\leq n$, and will disappear for $j=n+1$. In (22e), when $k\leq i<j$, the term $-\sum_{\ell=\max\{k,i-1\}}^{\min\{j-1,i\}}\lambda_{\ell kj}$ becomes $-\lambda_{ikj}-\lambda_{(i-1)kj}$; when $k<i=j$, it becomes a singleton $-\lambda_{(i-1)kj}$; and when $k=i=j$, it does not appear. Similarly, when $2\leq k=i=j\leq n$, the term $-\sum_{\ell=\min\{2i-k-1,i\}\vee 1}^{\max\{2i-j,i-1\}\wedge n}\phi_\ell$ becomes $-\phi_i-\phi_{i-1}$; when $j>i=k$ or $k=i=j=n+1$, the term only contains $-\phi_{i-1}$; when $k<i=j$ or $1=k=i=j$, the term only contains $-\phi_i$; and in all other cases, i.e., when $1\leq k<i<j\leq n+1$, the term does not appear. We can then reformulate the DR expectation model in an LP form as follows.

THEOREM 2. *Under no-consecutive no-show assumption, i.e., $D_q=D_q^{(2)}$, the DR expectation model* (7) *is equivalent to the following linear program:*

$$
\begin{aligned}
\min\quad & \sum_{i=1}^{n}\mu_i\rho_i+\sum_{i=1}^{n}\nu_i\gamma_i+\sum_{i=1}^{n+1}(\alpha_i+s_i^U\tau_i^U-s_i^L\tau_i^L)\\
\text{s.t.}\quad & \text{(22b)–(22g)},\quad \sum_{i=1}^{n}x_i=T,\quad x_{n+1}=0,\quad x_i\geq 0\quad \forall i=1,\ldots,n.
\end{aligned}
$$

Case 2. (Arbitrary No-Shows): Given $D_q=\{0,1\}^n$ and $D_s=\prod_{i=1}^n[s_i^{\mathrm{L}},s_i^{\mathrm{U}}]$, the optimization problem defining function $h(x,y,\rho,\gamma)$ (see (12c)) is separable by each appointment, i.e.,

$$
\begin{aligned}
h(x,y,\rho,\gamma) &= \max_{(q,s)\in D_q\times D_s}\left\{\sum_{i=1}^{n}(q_is_i-x_i)y_i-\sum_{i=1}^{n}(\rho_is_i+\gamma_iq_i)\right\}\\
&=\sum_{i=1}^{n}\max_{q_i\in\{0,1\},\ s_i\in[s_i^{\mathrm{L}},s_i^{\mathrm{U}}]}\left\{(q_is_i-x_i)y_i-(\rho_is_i+\gamma_iq_i)\right\}.
\end{aligned}
$$

To reformulate separation problem (14), recall the observations on polyhedron Y in Section 3.2 and again we represent the extreme points of Y based on variables t_{kj}. It follows that

$$\max_{y\in Y} h(x,y,\rho,\gamma) = \max_{t\geq 0} \sum_{k=1}^{n+1}\sum_{j=k}^{n+1}\left(\sum_{i=k}^{j} \max_{q_i\in\{0,1\},\ s_i\in[s_i^{\mathrm{L}},s_i^{\mathrm{U}}]} \{(q_is_i - x_i)\pi_{ij} - (\rho_is_i+\gamma_iq_i)\}\right) t_{kj} \tag{23a}$$

$$\text{s.t.} \sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i=1,\dots,n+1 \tag{23b}$$

$$t_{kj}\in\{0,1\}, \quad \forall 1\leq k\leq j\leq n+1. \tag{23c}$$

Because the constraint matrix formed by (23b)–(23c) is totally unimodular (TU), we can relax the integrality constraints (23c) without loss of optimality. Hence, formulation (23a)–(23c) is an LP model in variables t_{kj} and we can take its dual as:

$$\min_{\alpha,\beta} \sum_{i=1}^{n+1}\alpha_i \tag{24a}$$

$$\text{s.t.} \sum_{i=k}^{j}\alpha_i \geq \sum_{i=k}^{j}\beta_{ij}, \quad \forall 1\leq k\leq j\leq n+1 \tag{24b}$$

$$\beta_{ij} \geq \max_{q_i\in\{0,1\},\ s_i\in[s_i^{\mathrm{L}},s_i^{\mathrm{U}}]} \{(q_is_i - x_i)\pi_{ij} - (\rho_is_i+\gamma_iq_i)\}, \quad \forall i=1,\dots,n,\ \forall j=i,\dots,n+1 \tag{24c}$$

$$\beta_{n+1,n+1} = 0, \tag{24d}$$

where dual variables α_i, $i=1,\dots,n+1$ are associated with constraints (23b), constraints (24b) are associated with variables y_{kj}, each variable β_{ij} represents the value of $\max_{q_i\in\{0,1\},s_i\in[s_i^{\mathrm{L}},s_i^{\mathrm{U}}]}\{(q_is_i-x_i)\pi_{ij}-(\rho_is_i+\gamma_iq_i)\}$, and $\beta_{n+1,n+1}=0$ because $q_{n+1}=s_{n+1}=\pi_{n+1,n+1}=0$. Finally, for each $i=1,\dots,n$, the related objective function $(q_is_i-x_i)\pi_{ij}-(\rho_is_i+\gamma_iq_i)$ is linear in variables q_i and s_i, and thus each of constraints (24c) is equivalent to

$$\beta_{ij} \geq -\pi_{ij}x_i - s_i^{\mathrm{L}}\rho_i \tag{25a}$$

$$\beta_{ij} \geq -\pi_{ij}x_i - s_i^{\mathrm{U}}\rho_i \tag{25b}$$

$$\beta_{ij} \geq -\pi_{ij}x_i - s_i^{\mathrm{L}}\rho_i - \gamma_i + s_i^{\mathrm{L}}\pi_{ij} \tag{25c}$$

$$\beta_{ij} \geq -\pi_{ij}x_i - s_i^{\mathrm{U}}\rho_i - \gamma_i + s_i^{\mathrm{U}}\pi_{ij}, \tag{25d}$$

because $q_i\in\{0,1\}$ and $s_i\in\{s_i^{\mathrm{L}},s_i^{\mathrm{U}}\}$ at optimality. It follows that formulation (13) (i.e., the DR expectation model (7)), is equivalent to the following LP model when $D_q = D_q^{(n+1)}$:

$$\min_{x,\rho,\gamma,\alpha,\beta} \sum_{i=1}^{n}\mu_i\rho_i + \sum_{i=1}^{n}\nu_i\gamma_i + \sum_{i=1}^{n+1}\alpha_i$$

$$\text{s.t. (24b), (24d), (25a)–(25d),} \quad \sum_{i=1}^{n}x_i = T, \quad x_{n+1}=0, \quad x_i\geq 0, \quad \forall i=1,\dots,n.$$

5. Computational Results

We conduct numerical experiments on the three variants of the DR expectation model (7), namely, E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and E-$D_q^{(K)}$ ($K=3,\dots,n$), yielded by $D_q = D_q^{(2)}$, $D_q^{(n+1)}$, and $D_q^{(K)}$, respectively. For benchmark, we also solve a stochastic linear program (SLP) that minimizes the expected total cost of waiting, server idleness, and overtime via the sample average approximation (SAA) approach (Kleywegt et al. 2002). We briefly describe the key computational procedures as follows. First, we follow a distribution belief to generate N i.i.d. samples, of which we randomly pick a small subset of data to compute the empirical mean and support information, and use them to compute the (in-sample) optimal solutions to the DR models.

For the SLP, we solve an LP model:

$$\textbf{SLP:} \quad \min_{x,w,u,W} \ \frac{1}{N}\sum_{m=1}^{N}\sum_{i=1}^{n}(c_i^{\mathrm{w}} w_i^m + c_i^{\mathrm{u}} u_i^m) + c^{\mathrm{o}} W^m \tag{26a}$$

$$\text{s.t.} \quad w_i^m - u_{i-1}^m = q_{i-1}^m s_{i-1}^m + w_{i-1}^m - x_{i-1} \quad \forall i=2,\dots,n,\ m=1,\dots,N \tag{26b}$$

$$W^m - u_n^m = q_n^m s_n^m + w_n^m + \sum_{i=1}^{n-1} x_i - T \quad \forall m=1,\dots,N \tag{26c}$$

$$\sum_{i=1}^{n-1} x_i \le T \tag{26d}$$

$$x_i \ge 0 \quad \forall i=1,\dots,n-1 \tag{26e}$$

$$w_i^m \ge 0,\ w_1^m = 0,\ u_i^m \ge 0,\ W^m \ge 0, \quad \forall i=1,\dots,n,\ m=1,\dots,N \tag{26f}$$

where q_i^m and s_i^m are realizations of parameter q_i and s_i of appointment i in scenario m, respectively, for all $i=1,\dots,n$ and $m=1,\dots,N$. Variables w_i^m, u_i^m, and W^m represent recourse waiting time of appointment i, server idle time after serving appointment i, and server overtime in scenario m, respectively, for all $m=1,\dots,N$. Constraints (26b) and (26c) obtain the waiting time/idle time/overtime values for each appointment dependent on values of x_i and (q_i^m, s_i^m).

Section 5.1 describes how to set the parameter for the above models; Section 5.2 compares the CPU time and details of solving E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, E-$D_q^{(K)}$, and SLP. In Section 5.3, we illustrate optimal objective values given by E-$D_q^{(K)}$ with $K=2,\dots,n$ for different settings of time limit T and no-show probabilities. In Section 5.4, we compare the performance of optimal schedules of E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP via out-of-sample simulation tests. Specifically, we follow a certain distribution to generate N' data samples, which represent realizations of random service durations and no-shows. The distributions used for generating the in-sample and out-of-sample data could be different, and when they are the same and N is sufficiently large, the SLP is considered being optimized under the "perfect information" (Birge and Louveaux 2011). In reality, it is hard to know the exact true distribution, and thus we also test the case where the distribution is "misspecified."

5.1. Experiment Setup

We follow procedures in the appointment scheduling literature (e.g., Denton and Gupta 2003, Mak et al. 2015) to generate random instances as follows. For most instances, we consider $n = 10$ appointments, each having a random service duration s_i with the mean $\mu_i \sim U[36, 44]$ (i.e., uniformly sampled in between the values below and above 10% of 40 minutes) and the standard deviation $\sigma_i = 0.5\mu_i$. We set $T = \sum_{i=1}^n \mu_i + R \cdot \sqrt{\sum_{i=1}^n \sigma_i^2}$ where scalar R adjusts the length of time limit T. Each appointment i has a probability ν_i of showing up, and we test $\nu_1 = \cdots = \nu_n = 0.8$ or $\nu_1 = \cdots = \nu_n = 0.6$. To approximate the upper and lower bounds s_i^{U} and s_i^{L} of each service duration s_i, we respectively use the 80%- and 20%-quantile values of the N in-sample data. We set the ratio $c_i^{\mathrm{w}} : c_i^{\mathrm{u}} : c^{\mathrm{o}} = 1 : 0.5 : 10$ in all the DR models and in SLP.

We sample $N = 1000$ realizations $(q_1^m, s_1^m), \ldots, (q_n^m, s_n^m),\ m = 1, \ldots, N$ by following Log-Normal distributions with the given means and standard deviations of s_i and probability $1 - \nu_i$ of no-shows for each $i = 1, \ldots, n$. (The Log-Normal distribution possesses the long-tail property and has been shown accurately describing the shape of service-time distributions in many service systems (see, e.g., Gul et al. 2011, for a study of five-year outpatient surgical data in Mayo Clinic).) We optimize the SLP model by using all the N data points, and only use 20 randomly picked data samples from the N-data set to calculate the first moments of service durations and no-shows, used in all the DR models. Given the optimal schedules produced by different models, we generate $N' = 10,000$ i.i.d. data samples from certain distributions with details given in Section 5.4, to evaluate the performance of each schedule.

We increase the size of instances with $n = 10, 15, \ldots, 50$ appointments in Section 5.2 to compare the CPU time of different models and approaches. For each instance, we generate $N = 1000$ i.i.d. data samples of service durations and no-shows by following the same procedures as above. All LP (i.e., E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP) and MILP (i.e., E-$D_q^{(K)}$ with $K = 3, \ldots, n$) models are computed in Python 2.7.10 using Gurobi 5.6.3. The computations are performed on a Windows 7 machine with Intel(R) Core(TM) i7-2600M CPU 3.40 GHz and 8GB memory. The CPU time limit is set as 3 hours for solving each instance.

5.2. CPU Time and Computational Details

In this section, we increase the problem size from $n = 10$ to $n = 50$ appointments, and compare CPU time of optimizing different DR models and the SLP. We first vary $R = 0, 0.25, 0.5, 0.75, 1$, and find that the CPU time of all the models are similar for different R values and no-show probabilities. Thus, we fix $R = 0$ and use $T = \sum_{i=1}^n \mu_i = 380.24$ minutes. We consider $\nu_i = 0.6$ for all appointments $i = 1, \ldots, n$, and test ten replications for each instance. Table 1 reports the average CPU time (in second) of solving models E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, SLP, and E-$D_q^{(K)}$ with $K = 0.3n$ and $K = 0.7n$. Note that the first three are LP models, and specifically, there are $\mathcal{O}(n^3)$ variables and $\mathcal{O}(n^3)$ constraints in the two LP models E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$, but $\mathcal{O}(nN)$ variables and constraints in SLP. The E-$D_q^{(0.3n)}$ and E-$D_q^{(0.7n)}$ models are solved via Algorithm 1, and we present the average time for solving the MILP models with (see columns "Ineq.") and without (see columns "w/o") the valid inequalities in Proposition 1. For instances that take longer than 3 hours to solve, we instead report the optimality gap values (in %) achieved at the end of the computation process.

Table 1 Average CPU time (in second) of solving DR models and SLP with $R = 0$ and $1 - \nu_i = 0.4$

n	E-$D_q^{(2)}$	E-$D_q^{(n+1)}$	SLP	E-$D_q^{(0.3)}$		E-$D_q^{(0.7n)}$	
				w/o	Ineq.	w/o	Ineq.
10	0.03	0.00	3.10	6.70	10.68	3.44	1.95
15	0.16	0.01	7.24	52.74	47.56	20.28	4.23
20	0.38	0.01	10.44	158.50	106.67	72.96	15.12
25	2.85	0.02	16.97	409.88	270.60	266.34	47.18
30	5.12	0.05	20.91	1000.38	187.81	823.14	101.68
35	11.50	0.05	28.18	10658.06	401.91	7994.79	340.76
40	28.87	0.15	32.32	(5.76%)	808.38	(10.98%)	614.75
45	26.55	0.20	39.07	(12.49%)	1739.49	(10.24%)	1491.17
50	24.64	0.45	44.49	(31.77%)	3271.83	(43.63%)	3393.62

In Table 1, the CPU time of both E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ are shorter than the one of SLP, especially after $n \geq 40$. The time for solving E-$D_q^{(2)}$ is longer than solving E-$D_q^{(n+1)}$, due to the many more constraints involved in the former. Note that the time presented in Table 1 is only for solving all the models, but does not include the time spent on reading in data and constructing the constraints, which is negligible for all the DR models, but grows quickly for SLP, ranging from 30 seconds to 60 seconds when $n \geq 35$.

All the DR LP models and SLP are efficiently solved for $n = 10, \ldots, 50$, while the MILP models E-$D_q^{(K)}$ with either small or large K-values are computationally intractable, reflected by the long CPU seconds taken by instances with $n = 35, 40, 45, 50$, especially when no valid inequalities were added. The addition of valid inequalities in Proposition 1 drastically speeds up the decomposition algorithm, and the effect is much more significant when $n \geq 35$. For instance, none of the $n = 40, 45, 50$ cases were solved within 3 hours without the valid inequalities, and the average optimality gaps could be as large as $30\% \sim 45\%$ when $n = 50$. In contrast, after adding the valid inequalities, the decomposition algorithm quickly converges, and on average, it only takes no more than 15, 30, and 60 minutes to optimize the MILPs over instances with $n = 40, 45$, and 50, respectively.

Next, we present more details of solving the E-$D_q^{(K)}$ MILPs. Table 2 illustrates the number of constraints ("**# of Cons.**"), the total number of branching nodes ("**# of Nodes**"), the average CPU seconds taken by the master problem and the subproblem in each iteration, and the number

Table 2 Computational details of solving the MILP models E-$D_q^{(0.3)}$ and E-$D_q^{(0.7n)}$

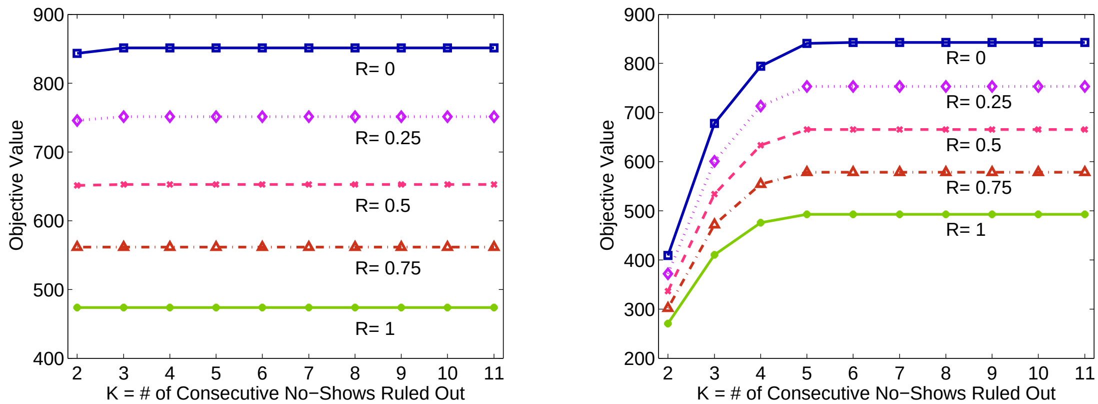
Models	n	with Ineq. # of Cons.	with Ineq. # of Nodes	with Ineq. Avg. Time (s) Master	with Ineq. Avg. Time (s) Sub	with Ineq. # of Cuts	w/o Ineq. # of Cons.	w/o Ineq. # of Nodes	w/o Ineq. Avg. Time (s) Master	w/o Ineq. Avg. Time (s) Sub	w/o Ineq. # of Cuts
E-$D_q^{(0.3)}$	10	2239	790	0.00	0.22	48	2023	1810	0.00	0.13	49
	15	6271	1363	0.01	0.61	77	5742	8510	0.01	0.62	83
	20	13346	2572	0.03	1.04	99	12435	19112	0.03	1.52	102
	25	24637	2865	0.08	2.38	110	22979	31602	0.11	3.83	104
	30	40686	32	0.23	2.00	84	38247	54697	0.22	9.04	108
	35	62932	0	0.51	3.95	90	59116	1245441	0.53	86.12	123
	40	91602	0	1.05	7.28	97	86459	677713	2.63	632.67	17
	45	128499	0	1.60	16.52	96	121153	387066	7.30	1342.76	8
	50	173437	0	3.09	27.48	107	164071	324952	17.57	3582.45	3
E-$D_q^{(0.7n)}$	10	2097	0	0.01	0.06	29	2019	525	0.00	0.13	25
	15	5882	0	0.01	0.11	35	5736	1975	0.02	0.56	35
	20	12626	0	0.04	0.28	48	12427	6674	0.05	1.41	50
	25	23287	0	0.11	0.71	58	22969	19704	0.13	3.79	68
	30	38636	0	0.23	1.50	59	38235	52000	0.27	10.02	80
	35	59691	0	0.48	3.63	83	59102	848133	0.42	82.86	96
	40	87154	0	0.93	6.06	88	86443	880188	3.90	1538.97	7
	45	122121	0	2.07	15.07	87	121135	385058	7.89	1192.12	9
	50	165207	0	3.06	28.36	108	164051	268133	18.05	10786.87	1

of iterations in the decomposition algorithm before it converges to the optimum or reaches the time limit ("**# of Cuts**") for solving both E-$D_q^{(0.3n)}$ and E-$D_q^{(0.7n)}$, with or without the valid inequalities.

In Table 2, we observe that the valid inequalities in Proposition 1 slightly increase the number of constraints, but significantly tighten the MILPs, directly reflected by the significantly reduced branching-and-bound nodes in all the instances. In particular, the decomposition algorithm obtains integer solutions at the root node in each iteration for solving E-$D_q^{(0.3n)}$ when $n \geq 35$, and for E-$D_q^{(0.7n)}$ given any values of n we test. Moreover, the valid inequalities significantly reduce the CPU seconds of computing the separation subproblem in each iteration, especially for instances with $n = 35, 40, 45, 50$. Lastly, by adding the valid inequalities to the MILP models, the decomposition algorithm takes almost constant number of iterations to converge (i.e., being around 100 iterations when $n \geq 20$). However, if no valid inequalities were added, the number of iterations first increases as we increase n from 10 to 35, and drastically decreases as we continue increasing n to 50.

5.3. Optimal Objective Values and Scheduling Solution Patterns

We compare the optimal objective values of the E-$D_q^{(K)}$ models with $K = 2, \ldots, n+1$, and plot their value changes for instances with $R = 0, 0.25, 0.5, 0.75, 1$ and no-show probabilities $1 - \nu_i = 0.2$, 0.4 for all $i = 1, \ldots, n$ when $n = 10$. Recall that parameter K represents the minimum number of consecutive appointments in which consecutive no-shows are ruled out. Therefore, as K increases, the support $D_q^{(K)}$ becomes larger, which leads to smaller feasible region for the scheduling decision vector x, and thus the optimal objective value is nondecreasing for $K = 2, \ldots, n+1$. Figure 2 illustrates the optimal objective values, in which Figure 2(a) corresponds to $1 - \nu_i = 0.2, \ \forall i = 1, \ldots, n$ and Figure 2(b) corresponds to larger no-show probabilities $1 - \nu_i = 0.4, \ \forall i = 1, \ldots, n$.



(a) $n=10,\ 1-\nu_i=0.2,\ \forall i=1,\dots,n$ (b) $n=10,\ 1-\nu_i=0.4,\ \forall i=1,\dots,n$

Figure 2 Optimal objective values of E-$D_q^{(K)}$ for different settings of parameter R (time limit) and $1-\nu_i$ (no-show probability)

Table 3 presents the detailed optimal objective values in Figure 2. Note that the optimal objective values of E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ respectively provide valid lower and upper bounds for the optimal objective value of any E-$D_q^{(K)}$ models with $K=3,\dots,n$. For each combination of R and $1-\nu_i$, we mark the first K-value when the optimal objective value of E-$D_q^{(K)}$ equals to the upper bound, i.e., the optimal objective value of E-$D_q^{(n+1)}$.

Table 3 Optimal objective value changes according to the value of K and no-show probabilities

No-show	R	E-$D_q^{(2)}$	E-$D_q^{(K)}$, with $K=$ 3	4	5	6	7	8	9	10	E-$D_q^{(n+1)}$
$1-\nu_i=0.2$	0	843.38	**851.08**	851.08	851.08	851.08	851.08	851.08	851.08	851.08	851.08
	0.25	745.74	**751.32**	751.32	751.32	751.32	751.32	751.32	751.32	751.32	751.32
	0.5	651.18	**652.54**	652.54	652.54	652.54	652.54	652.54	652.54	652.54	652.54
	0.75	**561.32**	561.32	561.32	561.32	561.32	561.32	561.32	561.32	561.32	561.32
	1	**473.69**	473.69	473.69	473.69	473.69	473.69	473.69	473.69	473.69	473.69
$1-\nu_i=0.4$	0	409.11	677.66	794.41	840.70	**842.91**	842.91	842.91	842.91	842.91	842.91
	0.25	371.47	600.99	713.55	**752.87**	752.87	752.87	752.87	752.87	752.87	752.87
	0.5	336.21	534.22	633.12	**665.24**	665.24	665.24	665.24	665.24	665.24	665.24
	0.75	301.89	472.07	553.93	**578.61**	578.61	578.61	578.61	578.61	578.61	578.61
	1	270.55	410.02	475.67	**492.69**	492.69	492.69	492.69	492.69	492.69	492.69

In Table 3, when the no-show probability is smaller (i.e., $1-\nu_i=0.2,\ \forall i=1,\dots,n$), the differences between the upper and lower bounds are very small, indicating that the two LP models E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ can already provide tight approximations for the MILPs of E-$D_q^{(K)}$ with $K=3,\dots,n$. Considering the long CPU time of solving the MILPs in Table 1, one can avoid directly solving E-$D_q^{(K)}$ ($K=3,\dots,n$), and instead use $K=2$ or $K=n+1$. On the other hand, when the no-show probability is larger (i.e., $1-\nu_i=0.4,\ \forall i=1,\dots,n$), the differences between the results of E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ become larger when R is smaller. In such a case, the choices of different K values will lead to significantly different objective costs and also solution patterns (which we illustrate in EC.5). A decision maker can choose either E-$D_q^{(2)}$ or E-$D_q^{(n+1)}$ to optimize the schedule x based on his/her risk preference. Alternatively, he/she can firstly use the two LP models to quickly compute the bounds of the optimal objective value for a general E-$D_q^{(K)}$ for any $K = 3, \ldots, n$, and then optimize the E-$D_q^{(K)}$ model for some K by employing the valid inequalities in Proposition 1 and the decomposition algorithm.

5.4. Results of Out-of-Sample Performance

We compare the out-of-sample simulation performance of the optimal schedules of E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP (see Figure EC.1) under (i) "perfect information" and (ii) misspecified distributional information. Note that E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ models produce solutions that differ the most under large no-show probabilities, and thus we focus on the case when $1 - \nu_i = 0.4,\ \forall i = 1, \ldots, n$. We examine three cases of the time limit T, by using $R = 0, 0.5, 1$.

We generate two sets of $N' = 10,000$ i.i.d. out-of-sample data $(q_1^m, s_1^m), \ldots, (q_n^m, s_n^m),\ m = 1, \ldots, N'$ of the random vector (q, s) following the procedures as follows.

- **Perfect Information:** We use the same distribution (i.e., Log-Normal) and parameter settings as the ones for generating the N in-sample data to sample the N' data points.
- **Misspecified Distribution:** We keep the same mean values μ_i of the random s_i, ν_i of the random q_i, and standard deviation σ_i of the random s_i for each appointment $i = 1, \ldots, n$. Therefore, the moment information used in all the DR models and in the SLP remain the same, but we vary the distribution type, as well as correlations among the random service durations and no-shows. Specifically, we follow positively correlated truncated Normal distributions with supports $[0, s_i^{\mathrm{U}}],\ \forall i = 1, \ldots, n$ to generate realizations $s_1^m, \ldots, s_n^m$ and follow positively correlated Bernoulli distributions to generate realizations $q_1^m, \ldots, q_n^m$ for $m = 1, \ldots, N'$. The parameters of the truncated Normal distributions and the Bernoulli distributions are designed by following standard statistical methods[1], in order to yield positive data correlations and also to keep the first two moments of the N' out-of-sample data the same as the ones of the N in-sample data.

To measure the out-of-sample performance of each solution given by E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP, we fix x as an interested solution in Model (26), but use parameters $(q_1^m, s_1^m), \ldots, (q_n^m, s_n^m),\ m = 1, \ldots, N'$. We then compute $w_i^m,\ u_i^m,\ W^m$ as the waiting time (WaitT), idle time (idleT), and and overtime (OverT) in each scenario m, for $m = 1, \ldots, N'$. Table 4 displays means and quantiles of WaitT, IdleT, and OverT, yielded by the optimal solution of each model under perfect distributional information.

[1] See `https://en.wikipedia.org/wiki/Truncated_normal_distribution`.

Table 4 Out-of-sample performance of optimal schedules given by E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP under perfect information with no-show probabilities $1-\nu_i = 0.4,\ \forall i = 1,\dots,n$

Metrics	Model	$R=0$ (in minute)			$R=0.5$ (in minute)			$R=1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	E-$D_q^{(2)}$	10.06	13.97	16.62	7.33	11.26	19.31	5.50	9.18	22.06
	E-$D_q^{(n+1)}$	6.62	49.27	20.15	4.87	33.59	21.54	3.33	28.11	23.96
	SLP	11.20	3.50	15.57	8.70	2.29	18.41	6.70	1.64	21.31
Median	E-$D_q^{(2)}$	0.00	0.00	12.27	0.00	0.00	14.68	0.00	0.00	20.47
	E-$D_q^{(n+1)}$	0.00	44.92	15.47	0.00	29.22	19.29	0.00	24.93	21.79
	SLP	0.00	0.00	8.19	0.00	0.00	12.95	0.00	0.00	17.45
75%-QT	E-$D_q^{(2)}$	8.40	19.94	33.32	1.79	14.95	34.99	0.00	9.12	40.25
	E-$D_q^{(n+1)}$	0.00	72.90	45.40	0.00	52.22	45.40	0.00	45.27	47.16
	SLP	11.94	0.00	28.83	6.49	0.00	34.38	1.33	0.00	37.08
95%-QT	E-$D_q^{(2)}$	50.88	35.42	46.62	40.08	26.06	48.64	33.52	19.27	50.14
	E-$D_q^{(n+1)}$	40.51	117.78	50.33	30.68	90.84	50.33	22.03	76.66	50.33
	SLP	54.03	23.75	43.20	45.89	12.61	46.05	38.41	7.54	49.85

Based on Table 4, both E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ yield slightly better waiting time on average and at different quantiles than the SLP. The E-$D_q^{(2)}$ model results in overtime that is close to the one of SLP (which has perfect distributional information). But the optimal schedule of E-$D_q^{(n+1)}$ performs badly on average and at all quantiles. For example, when $R = 0$, the schedule by E-$D_q^{(2)}$ lasts 10 minutes longer than the 3-minute average overtime given by the SLP optimal schedule, while the optimal schedule of E-$D_q^{(n+1)}$ lasts about 46 minutes longer on average. This is due to the overly conservative no-show support assumption used by E-$D_q^{(n+1)}$. When the distributional information is accurate, E-$D_q^{(n+1)}$ results in over-conservative schedules that perform badly especially in the overtime metric.

Table 5 illustrates the means and quantiles of WaitT, IdleT, and OverT, yielded by optimal schedules of the three models under misspecified distributional information.

Table 5 Out-of-sample performance of optimal schedules given by E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP under misspecified distribution with no-show probabilities $1-\nu_i = 0.4,\ \forall i = 1,\dots,n$

Metrics	Model	$R=0$ (in minute)			$R=0.5$ (in minute)			$R=1$ (in minute)		
		WaitT	OverT	IdleT	WaitT	OverT	IdleT	WaitT	OverT	IdleT
Mean	E-$D_q^{(2)}$	19.89	42.30	13.93	15.75	34.24	16.08	11.51	27.39	18.36
	E-$D_q^{(n+1)}$	12.14	69.59	16.65	10.20	52.16	17.87	8.26	44.56	20.07
	SLP	29.32	38.36	13.53	24.30	31.04	15.76	19.61	24.54	18.07
Median	E-$D_q^{(2)}$	0.00	0.00	5.52	0.00	0.00	9.62	0.00	0.00	14.15
	E-$D_q^{(n+1)}$	0.00	46.25	8.61	0.00	30.62	11.83	0.00	30.56	15.64
	SLP	0.00	0.00	5.09	0.00	0.00	9.16	0.00	0.00	13.00
75%-QT	E-$D_q^{(2)}$	15.34	41.98	27.16	10.70	27.98	30.95	4.91	19.47	35.42
	E-$D_q^{(n+1)}$	6.95	93.23	33.03	2.44	65.31	34.86	0.00	53.56	39.49
	SLP	27.64	11.57	26.36	21.18	0.00	30.33	15.06	0.00	34.45
95%-QT	E-$D_q^{(2)}$	109.75	241.69	46.59	89.11	212.08	49.77	65.02	182.47	48.78
	E-$D_q^{(n+1)}$	62.06	241.69	49.77	58.20	212.08	49.77	52.26	182.47	50.08
	SLP	147.84	241.63	43.20	130.09	212.02	46.05	110.73	182.41	49.84

From Table 5, we observe that both DR models yield much better (i.e., 30%–70% shorter) waiting time per appointment than the optimal schedule given by the SLP, when the distribution type becomes different but the first two moments remain unchanged from the assumed case. The time reduction is reflected in all the metrics, including the mean and 50% to 95% quantiles of the random WaitT, for both $R=0$ and $R=1$. On the other hand, the three models yield similar IdleT, and the optimal schedule given by E-$D_q^{(n+1)}$ yields slightly longer idle time per appointment, but much longer OverT than both E-$D_q^{(2)}$ and SLP. This observation indicates that the optimal schedules given by SLP can become suboptimal when the probability distributions are misspecified, while E-$D_q^{(2)}$ can produce schedules that are less sensitive to misspecification of distribution types.

6. Conclusions

In this paper, we studied moment-based DR variants of the stochastic appointment scheduling problems. We incorporated both risk-neutral (i.e., expectation-based) and risk-averse (i.e., CVaR-based) objective and/or constraints to restrict random waiting-time, overtime, and idle-time outcomes under random 0-1 no-shows and service durations. Our approaches are suitable for a system operator who has a limited amount of data and considers ambiguous distributions of the two co-existing uncertainties.

By employing integer programming techniques including linearization and valid inequalities, we derived exact reformulations and decomposition algorithms for these DR models. In important special cases of no-shows, our derivation led to LP reformulations that can readily be implemented in LP solvers such as Microsoft Excel. The computational experiments demonstrated the tractability and effectiveness of our approaches. We also derived the following insights: (i) accounting for no-shows in DR models significantly shortens waiting time in out-of-sample tests, (ii) one can improve the DR models' ability of utilizing distributional information by using reasonably conservative supports, and (iii) the DR model with the least conservative support of no-shows obtains near-optimal schedules under perfect information, and outperforms other DR models and the stochastic program if the distributional type is misspecified.

Future research directions include optimizing the length of time limit (i.e., T) and the incorporation of appointment sequencing decisions while considering uncertain no-shows. It is also interesting to investigate the power of integer programming approaches in other related stochastic and robust optimization models.

Acknowledgments

The authors are grateful to the four referees and the Associate Editor for their constructive comments and helpful suggestions.

References

Begen MA, Levi R, Queyranne M (2012) Technical note – a sampling-based approach to appointment scheduling. *Operations Research* 60(3):675–681.

Begen MA, Queyranne M (2011) Appointment scheduling with discrete random durations. *Mathematics of Operations Research* 36(2):240–257.

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming* 88(3):411–424.

Berg BP, Denton BT, Erdogan SA, Rohleder T, Huschka TR (2014) Optimal booking and scheduling in outpatient procedure centers. *Computers & Operations Research* 50:24–37.

Bertsimas D, Doan XV, Natarajan K, Teo CP (2010) Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* 35(3):580–602.

Bertsimas D, Popescu I (2005) Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15(3):780–804.

Bertsimas D, Sim M (2004) The price of robustness. *Operations Research* 52(1):35–53.

Birge JR, Louveaux FV (2011) *Introduction to Stochastic Programming* (New York, NY: Springer).

Cayirli T, Veral E (2003) Outpatient scheduling in health care: a review of literature. *Production and Operations Management* 12(4):519–549.

Cayirli T, Yang KK, Quek SA (2012) A universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management* 21(4):682–697.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

Denton BT, Gupta D (2003) A sequential bounding approach for optimal appointment scheduling. *IIE Transactions* 35(11):1003–1016.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research* 58(4):802–816.

Denton BT, Viapiano J, Vogl A (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science* 10(1):13–24.

Erdogan SA, Denton BT (2013) Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing* 25(1):116–132.

Ge D, Wan G, Wang Z, Zhang J (2013) A note on appointment scheduling with piecewise linear cost functions. *Mathematics of Operations Research* 39(4):1244–1251.

Gul S, Denton BT, Fowler JW, Huschka TR (2011) Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management* 20(3):406–417.

Gupta D, Denton BT (2008) Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 40(9):800–819.

Gurvich I, Luedtke J, Tezcan T (2010) Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science* 56(7):1093–1115.

Hassin R, Mendel S (2008) Scheduling arrivals to queues: A single-server model with no-shows. *Management Science* 54(3):565–572.

He S, Sim M, Zhang M (2015) Data-driven patient scheduling in emergency departments: A hybrid robust-stochastic approach. Available at Optimization-Online `http://www.optimization-online.org/DB_HTML/2015/11/5213.html`.

Ho CJ, Lau HS (1992) Minimizing total cost in scheduling outpatient appointments. *Management Science* 38(12):1750–1764.

Kleywegt AJ, Shapiro A, Homem-de-Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12(2):479–502.

Kong Q, Lee CY, Teo CP, Zheng Z (2013) Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research* 61(3):711–726.

Kong Q, Li S, Liu N, Teo CP, Yan Z (2016) Appointment scheduling under schedule-dependent patient no-show behavior. Available at `http://www.columbia.edu/~nl2320/doc/Noshow-MS-1030c.pdf`.

LaGanga LR, Lawrence SR (2012) Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management* 21(5):874–888.

Lee VJ, Earnest A, Chen MI, Krishnan B (2005) Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases. *BMC Health Services Research* 5(1):51.

Lin J, Muthuraman K, Lawley M (2011) Optimal and approximate algorithms for sequential clinical scheduling with no-shows. *IIE Transactions on Healthcare Systems Engineering* 1(1):20–36.

Liu N, Ziya S, Kulkarni VG (2010) Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management* 12(2):347–364.

Luo J, Kulkarni VG, Ziya S (2012) Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management* 14(4):670–684.

Mak HY, Rong Y, Zhang J (2014) Sequencing appointments for service systems using inventory approximations. *Manufacturing & Service Operations Management* 16(2):251–262.

Mak HY, Rong Y, Zhang J (2015) Appointment scheduling with limited distributional information. *Management Science* 61(2):316–334.

Mancilla C (2009) *Stochastic sequencing and scheduling of an operating room.* Ph.D. thesis, Lehigh University, Department of Industrial and Systems Engineering.

Mittal S, Schulz AS, Sebastian S (2014) Robust appointment scheduling. Jansen K, Rolim J, Devanur N, Moore C, eds., *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, volume 28, 356–37 (Barcelona, Spain: Leibniz International Proceedings in Informatics).

Muthuraman K, Lawley M (2008) A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions* 40(9):820–837.

Parizi MS, Ghate A (2015) Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2550560.

Pinedo M (2012) *Scheduling: Theory, Algorithms, and Systems* (Springer).

Robinson LW, Chen RR (2010) A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing & Service Operations Management* 12(2):330–346.

Scarf H, Arrow K, Karlin S (1958) A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, volume 10, 201–209 (Stanford, CA: Stanford University Press).

Shen S, Wang J (2014) Stochastic modeling and approaches for managing energy footprints in cloud computing services. *Service Science* 6(1):15–33.

Shylo OV, Prokopyev OA, Schaefer AJ (2012) Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing* 25(4):682–692.

Wang Y, Tang J, Fung RYK (2014) A column-generation-based heuristic algorithm for solving operating theater planning problem under stochastic demand and surgery cancellation risk. *International Journal of Production Economics* 158:28–36.

Zacharias C, Pinedo M (2014) Appointment scheduling with no-shows and overbooking. *Production and Operations Management* 23(5):788–801.

Zeng B, Turkcan A, Lin J, Lawley M (2010) Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research* 178(1):121–144.

Zhang Z, Denton B, Xie X (2015) Branch and price for chance constrained bin packing. Available at Optimization-Online http://www.optimization-online.org/DB_HTML/2015/11/5217.html.

Extensions and Proofs

EC.1. General Waiting Time Costs

In this section, we discuss a general setting that incorporates both server-based and appointment-based cases to model the waiting costs. We show that the models in Section 2 and solution methods in Section 3 can adapt for this general setting. For presentation brevity, we focus on DR expectation models and the adaptation of methods for DR CVaR models in EC.2 is similar.

To be general, we model the waiting time cost for each appointment i as $c^{\rm s}_i w_i + c^{\rm a}_i w_i q_i = (c^{\rm s}_i + c^{\rm a}_i q_i) w_i$, where $c^{\rm s}_i$ represents the server-based waiting time cost and $c^{\rm a}_i$ represents the appointment-based waiting time cost. Note that this setting applies for server-based and appointment-based cases, because we can set $c^{\rm s}_i := 0$ or $c^{\rm a}_i := 0$ if the corresponding cost does not apply. Accordingly, the general DR expectation model can be formulated as

$$\min_{x\in X} \sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \mathbb{E}_{\mathbb{P}_{q,s}}\left[Q^{\rm G}(x,q,s)\right], \tag{EC.1}$$

where $Q^{\rm G}(x,q,s)$ represents the cost function of the waiting, idleness, and overtime under the general setting. Replacing $c^{\rm w}_i w_i$ with $(c^{\rm s}_i + c^{\rm a}_i q_i) w_i$ in (11), we formulate the dual of $Q^{\rm G}(x,q,s)$ as

$$\begin{aligned}
Q^{\rm G}(x,q,s) \;=\; \max_{y} \;& \sum_{i=1}^{n} (q_i s_i - x_i) y_i && \text{(EC.2a)}\\
\text{s.t.}\;& y_{i-1} - y_i \le c^{\rm s}_i + c^{\rm a}_i q_i \quad \forall i = 2,\ldots,n && \text{(EC.2b)}\\
& -y_i \le c^{\rm u}_i \quad \forall i = 1,\ldots,n && \text{(EC.2c)}\\
& y_n \le c^{\rm o}, && \text{(EC.2d)}
\end{aligned}$$

and we let polyhedron $Y^{\rm G} := \{y : \text{(EC.2b)–(EC.2d)}\}$ represent the feasible region of variable y. As model (EC.2) is a linear program in variables y, there exists an optimal solution to (EC.2) that resides at an extreme point of $Y^{\rm G}$. It can be observed (see, e.g., Zangwill (1966, 1969), Mak et al. (2015)) that any extreme point $\hat{y}$ of $Y^{\rm G}$ satisfy (i) either $\hat{y}_n = -c^{\rm u}_n$ or $\hat{y}_n = c^{\rm o}$, and (ii) for all $i = 1,\ldots,n-1$, dual constraint $\hat{y}_{i+1} + c^{\rm s}_{i+1} + c^{\rm a}_{i+1} q_{i+1} \ge \hat{y}_i \ge -c^{\rm u}_i$ is binding at either the lower bound or the upper bound. Similar to the analysis in Section 3.2, we define binary variables t_{kj} for all $1 \le k \le j \le n+1$ to represent extreme points $\hat{y}$, such that $t_{kj} = 1$ if $\hat{y}_j = -c^{\rm u}_j$ and $\hat{y}_i = \hat{y}_{i+1} + c^{\rm s}_{i+1} + c^{\rm a}_{i+1} q_{i+1}$, $\forall i = k,\ldots,j-1$. It follows that

$$\hat{y}_i = \pi^{\rm G}_{ij} := \begin{cases} -c^{\rm u}_j + \sum_{\ell=i+1}^{j} (c^{\rm s}_\ell + c^{\rm a}_\ell q_\ell) & 1 \le i \le j \le n, \\ c^{\rm o} + \sum_{\ell=i+1}^{n} (c^{\rm s}_\ell + c^{\rm a}_\ell q_\ell) & 1 \le i \le n,\ j = n+1, \end{cases} \tag{EC.3}$$

where $\hat{y}_{n+1} = \pi^{\rm G}_{n+1,n+1} := 0$. Based on this binary representation, we can rewrite $Q^{\rm G}(x,q,s)$ as

$$Q^{\rm G}(x,q,s) \;=\; \max_{t} \sum_{k=1}^{n+1} \sum_{j=k}^{n+1} \left(\sum_{i=k}^{j} (q_i s_i - x_i) \pi^{\rm G}_{ij} \right) t_{kj}$$

$$\equiv \sum_{k=1}^{n}\sum_{j=k}^{n}\sum_{i=k}^{j}\left[(q_i s_i - x_i)\left(-c_j^{\mathrm{u}} + \sum_{\ell=i+1}^{j}(c_\ell^{\mathrm{s}} + c_\ell^{\mathrm{a}} q_\ell)\right)\right] t_{kj} +$$
$$\sum_{k=1}^{n}\sum_{i=k}^{n+1}\left[(q_i s_i - x_i)\left(c^{\mathrm{o}} + \sum_{\ell=i+1}^{n}(c_\ell^{\mathrm{s}} + c_\ell^{\mathrm{a}} q_\ell)\right)\right] t_{k(n+1)} \tag{EC.4a}$$
$$\text{s.t.} \quad \sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} = 1 \quad \forall i = 1, \dots, n+1 \tag{EC.4b}$$
$$t_{kj} \in \{0,1\}, \quad \forall 1 \le k \le j \le n+1. \tag{EC.4c}$$

Note that the objective function (EC.4a) contains multilinear terms $q_i s_i t_{kj}$ and $q_i q_\ell s_i t_{kj}$ with binary variables q_i, q_ℓ, and t_{kj}, and continuous variables s_i. To linearize formulation (EC.4), we define $p_{ikj} \equiv q_i t_{kj}$, $o_{ikj} \equiv q_i s_i t_{kj}$, and $r_{i\ell kj} = q_i q_\ell s_i t_{kj}$ for all $1 \le k \le i \le j \le n+1$ and $i+1 \le \ell \le j$. We then linearize the multilinear terms by applying McCormick inequalities (18a)–(18b) for variables p_{ikj}, (18c)–(18d) for variables o_{ikj}, and (EC.5a)–(EC.5b) as follows for variables $r_{i\ell kj}$.

$$r_{i\ell kj} \ge 0, \; r_{i\ell kj} - q_\ell s_i \le 0, \tag{EC.5a}$$
$$r_{i\ell kj} - o_{ikj} \le 0, \; r_{i\ell kj} - o_{ikj} + s_i^{\mathrm{U}}(1 - q_\ell) \ge 0. \tag{EC.5b}$$

If follows that $Q^{\mathrm{G}}(x,q,s)$ equals to the optimal objective value of the following MILP:

$$\max_{t,p,o,r} \; \sum_{k=1}^{n}\sum_{j=k}^{n}\sum_{i=k}^{j}\left[\left(-c_j^{\mathrm{u}} + \sum_{\ell=i+1}^{j} c_\ell^{\mathrm{s}}\right) o_{ikj} + \left(c_j^{\mathrm{u}} - \sum_{\ell=i+1}^{j} c_\ell^{\mathrm{s}}\right) x_i t_{kj} + \sum_{\ell=i+1}^{j} c_\ell^{\mathrm{a}} r_{i\ell kj} - \sum_{\ell=i+1}^{j} c_\ell^{\mathrm{a}} x_i p_{\ell kj}\right] +$$
$$\sum_{k=1}^{n}\sum_{i=k}^{n+1}\left[\left(c^{\mathrm{o}} + \sum_{\ell=i+1}^{n} c_\ell^{\mathrm{s}}\right) o_{ik(n+1)} - \left(c^{\mathrm{o}} + \sum_{\ell=i+1}^{n} c_\ell^{\mathrm{s}}\right) x_i t_{k(n+1)} + \sum_{\ell=i+1}^{n} c_\ell^{\mathrm{a}} r_{i\ell k(n+1)} - \sum_{\ell=i+1}^{n} c_\ell^{\mathrm{a}} x_i p_{\ell k(n+1)}\right]$$
$$\text{s.t.} \quad (18\text{a})\text{–}(18\text{d}), (\text{EC.4b}), (\text{EC.5a})\text{–}(\text{EC.5b}), \tag{EC.6a}$$
$$t_{kj}, p_{ikj} \in \{0,1\}, \quad \forall 1 \le k \le i \le j \le n+1. \tag{EC.6b}$$

To finish reformulating the general DR expectation model (EC.1), we follow a similar dualization process described in Section 3 and rewrite formulation (EC.1) as

$$\min_{x \in X, \rho \in \mathbb{R}^n, \gamma \in \mathbb{R}^n, \theta \in \mathbb{R}} \; \sum_{i=1}^{n} \mu_i \rho_i + \sum_{i=1}^{n} \nu_i \gamma_i + \theta \tag{EC.7a}$$
$$\text{s.t.} \quad \theta \ge H^{\mathrm{G}}(x,\rho,\gamma) \equiv \max_{(q,s) \in D_q \times D_s} \left\{ Q^{\mathrm{G}}(x,q,s) - \sum_{i=1}^{n}(\rho_i s_i + \gamma_i q_i) \right\}. \tag{EC.7b}$$

Similar to Lemma 1, we observe that for any fixed variables x, ρ, and γ, $H^{\mathrm{G}}(x,\rho,\gamma) < +\infty$. Furthermore, function $H^{\mathrm{G}}(x,\rho,\gamma)$ is convex and piecewise linear in x, ρ, and γ with a finite number of pieces. Hence, we can adapt Algorithm 1 to solve model (EC.1) in a decomposition framework. We present this adaptation in Algorithm 2. Similar to Algorithm 1, we observe that Algorithm 2 is finite. Finally, for the separation problem in Step 3, we remark that (i) feasible region D_q can be set as $D_q^{(K)}$ for $K=2,\ldots,n+1$ based on the scheduler's targeted conservativeness, (ii) the separation problem is an MILP and can be solved by off-the-shelf software, and (iii) we can incorporate the same valid inequalities identified in Proposition 1 to accelerate solving the separation problem and hence the decomposition algorithm.

Algorithm 2 A decomposition algorithm for solving general DR expectation model (EC.1).

1: Input: feasible regions X and $D_q \times D_s$; set of cuts $\{L(x,\rho,\gamma,\theta)\geq 0\}=\emptyset$.

2: Solve the master problem

$$\begin{aligned}\min_{x\in X,\rho,\gamma,\theta}\quad & \sum_{i=1}^{n}\mu_i\rho_i+\sum_{i=1}^{n}\nu_i\gamma_i+\theta\\ \text{s.t.}\quad & L(x,\rho,\gamma,\theta)\geq 0\end{aligned}$$

and record an optimal solution $(x^*,\rho^*,\gamma^*,\theta^*)$.

3: With (x,ρ,γ) fixed to be (x^*,ρ^*,γ^*), solve the separation problem

$$\begin{aligned}\max_{t,p,q,s,o,r}\ & \sum_{k=1}^{n}\sum_{j=k}^{n}\sum_{i=k}^{j}\left[\left(-c_j^{\mathrm{u}}+\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{s}}\right)o_{ikj}+\left(c_j^{\mathrm{u}}-\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{s}}\right)x_it_{kj}+\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{a}}r_{i\ell kj}-\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{a}}x_ip_{\ell kj}\right]+\\ & \sum_{k=1}^{n}\sum_{i=k}^{n+1}\left[\left(c^{\mathrm{o}}+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{s}}\right)o_{ik(n+1)}-\left(c^{\mathrm{o}}+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{s}}\right)x_it_{k(n+1)}+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{a}}r_{i\ell k(n+1)}-\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{a}}x_ip_{\ell k(n+1)}\right]-\\ & \sum_{i=1}^{n}(\rho_is_i+\gamma_iq_i)\\ \text{s.t.}\ & \text{(18a)–(18d)},\text{(EC.4b)},\text{(EC.5a)–(EC.5b)},\\ & t_{kj},p_{ikj}\in\{0,1\},\ \forall 1\leq k\leq i\leq j\leq n+1,\quad (q,s)\in D_q\times D_s\end{aligned}$$

and record an optimal solution $(t^*,p^*,q^*,s^*,o^*,r^*)$.

4: **if** θ^* is greater than or equal to the optimal objective value of the separation problem, **then**

5: stop and return x^* as an optimal solution to formulation (EC.1).

6: **else**

7: add the cut

$$\begin{aligned}\theta\geq & \sum_{k=1}^{n}\sum_{j=k}^{n}\sum_{i=k}^{j}\left[\left(-c_j^{\mathrm{u}}+\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{s}}\right)o_{ikj}^*+\left(c_j^{\mathrm{u}}-\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{s}}\right)t_{kj}^*x_i+\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{a}}r_{i\ell kj}^*-\sum_{\ell=i+1}^{j}c_\ell^{\mathrm{a}}p_{\ell kj}^*x_i\right]+\\ & \sum_{k=1}^{n}\sum_{i=k}^{n+1}\left[\left(c^{\mathrm{o}}+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{s}}\right)o_{ik(n+1)}^*-\left(c^{\mathrm{o}}+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{s}}\right)t_{k(n+1)}^*x_i+\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{a}}r_{i\ell k(n+1)}^*-\sum_{\ell=i+1}^{n}c_\ell^{\mathrm{a}}p_{\ell k(n+1)}^*x_i\right]-\\ & \sum_{i=1}^{n}(s_i^*\rho_i+q_i^*\gamma_i)\end{aligned}$$

to the set of cuts $\{L(x,\rho,\gamma,\theta)\geq 0\}$ and go to Step 2.

8: **end if**

EC.2. Solution Approaches for DR CVaR Models

In this section, we reformulate DR CVaR constraints in (6) with cost parameters $c^{\rm o}$, $c_i^{\rm u}$, and $c_i^{\rm w}$. We first represent CVaR by an alternative definition (Rockafellar and Uryasev (2000, 2002)):

$$\text{CVaR}_{1-\epsilon}\left(Q(x,q,s)\right) \;=\; \inf_{z\in\mathbb{R}}\left\{z+\frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)-z\right]^+\right\},$$

where $[a]^+ := \max\{a,0\}$ for $a\in\mathbb{R}$. It follows that

$$\begin{aligned}\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)} \text{CVaR}_{1-\epsilon}\left(Q(x,q,s)\right) &= \sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)}\inf_{z\in\mathbb{R}}\left\{z+\frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)-z\right]^+\right\}\\ &= \inf_{z\in\mathbb{R}}\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)}\left\{z+\frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)-z\right]^+\right\} &\text{(EC.8a)}\\ &= \inf_{z\in\mathbb{R}}\left\{z+\frac{1}{\epsilon}\sup_{\mathbb{P}_{q,s}\in\mathcal{F}(D,\mu,\nu)}\mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)-z\right]^+\right\}, &\text{(EC.8b)}\end{aligned}$$

where (EC.8a) follows the Sion's minimax theorem (Sion 1958) because $z+\frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}_{q,s}}\left[Q(x,q,s)-z\right]^+$ is convex in z, concave (in particular, linear) in variables $\mathbb{P}_{q,s}$, and $\mathcal{F}(D,\mu,\nu)$ is weakly compact.

EC.2.1. MILP Reformulation and Decomposition Algorithm

Based on a similar dualization process in Section 3 (see the primal and dual formulations (8) and (9)), we reformulate the inner maximization problem in (EC.8b) as a minimization problem, and combine it with the outer minimization problem to obtain

$$\begin{aligned} &\inf_{z,\rho,\gamma,\theta} \quad z+\frac{1}{\epsilon}\left(\sum_{i=1}^n \mu_i\rho_i+\sum_{i=1}^n \nu_i\gamma_i+\theta\right)\\ &\quad\text{s.t.} \quad \sum_{i=1}^n s_i\rho_i+\sum_{i=1}^n\gamma_i q_i+\theta \ge [Q(x,q,s)-z]^+ \quad \forall (q,s)\in D_q\times D_s\\ =\;&\inf_{z,\rho,\gamma,\theta} \quad z+\frac{1}{\epsilon}\left(\sum_{i=1}^n \mu_i\rho_i+\sum_{i=1}^n \nu_i\gamma_i+\theta\right)\\ &\quad\text{s.t.} \quad \sum_{i=1}^n s_i\rho_i+\sum_{i=1}^n\gamma_i q_i+\theta \ge 0 \quad \forall (q,s)\in D_q\times D_s &\text{(EC.9a)}\\ &\qquad\quad\; \sum_{i=1}^n s_i\rho_i+\sum_{i=1}^n\gamma_i q_i+\theta \ge Q(x,q,s)-z \quad \forall (q,s)\in D_q\times D_s, &\text{(EC.9b)}\end{aligned}$$

where constraints (EC.9a) and (EC.9b) are derived based on the definition of $[\cdot]^+$. Thus, the DR CVaR constraint (6) is equivalent to

$$\overline{Q} \;\ge\; z+\frac{1}{\epsilon}\left(\sum_{i=1}^n \mu_i\rho_i+\sum_{i=1}^n \nu_i\gamma_i+\theta\right) \tag{EC.10a}$$

$$\min_{(q,s)\in D_q\times D_s}\left\{\sum_{i=1}^n \rho_i s_i+\sum_{i=1}^n\gamma_i q_i\right\}+\theta \;\ge\; 0 \tag{EC.10b}$$

$$\theta + z \ \geq \max_{(q,s)\in D_q\times D_s} \left\{ Q(x,q,s) - \sum_{i=1}^{n}(\rho_i s_i + \gamma_i q_i) \right\}, \tag{EC.10c}$$

where constraint (EC.10a) is linear, but (EC.10b) and (EC.10c) need further analysis. First, we replace constraint (EC.10b) by equivalent linear constraints in the following proposition, whose proof is relegated to EC.4.1.

Proposition EC.1. *For fixed ρ and γ, and $D_q = D_q^{(K)}$ with $K \in \{2,\ldots,n+1\}$,* (EC.10b) *is equivalent to linear constraints:*

$$\theta + \sum_{i=1}^{n-K+1} \beta_i + \sum_{i=1}^{n} (s_i^L \chi_i^L - s_i^U \chi_i^U - \eta_i) \ \geq 0, \tag{EC.11a}$$

$$-\eta_i + \sum_{j=\max\{i-K+1,1\}}^{\min\{i,n-K+1\}} \beta_j \ \leq \gamma_i \quad \forall 1\leq i\leq n, \tag{EC.11b}$$

$$\chi_i^L - \chi_i^U \ \leq \rho_i \quad \forall 1\leq i\leq n, \tag{EC.11c}$$

$$\beta_i,\ \chi_i^L,\ \chi_i^U, \eta_i \geq 0 \quad \forall 1\leq i\leq n. \tag{EC.11d}$$

Second, note that the right-hand side of constraint (EC.10c) is equivalent to that of constraint (13b) in the reformulated DR expectation model, and so the reformulated separation problem (19) and Algorithm 1 described in Section 3 can be easily adapted to handle constraint (EC.10c). Furthermore, the valid inequalities (20a)–(20f) can be incorporated to accelerate solving the adapted separation problem and implementing the decomposition algorithm.

EC.2.2. LP Reformulations of the DR CVaR Model

We derive LP reformulations for the DR CVaR constraint (6) when $D_q = D_q^{(2)}$ (i.e., no consecutive no-shows) and when $D_q = D_q^{(n+1)}$ (i.e., arbitrary no-shows).

Case 1. (No Consecutive No-Shows) Recall that DR CVaR constraint (6) is equivalent to constraints (EC.10a), (EC.11a)–(EC.11d) with $K = 2$, and (EC.10c). When $D_q = D_q^{(2)}$, we apply Theorem 1 to further reformulate (EC.10c) as linear constraints $\theta + z \geq \sum_{i=1}^{n+1} (\alpha_i + s_i^{\mathrm{U}}\tau_i^{\mathrm{U}} - s_i^{\mathrm{L}}\tau_i^{\mathrm{L}})$ and (22b)–(22g), resulting in the following proposition.

Proposition EC.2. *When $D_q = D_q^{(2)}$, the DR CVaR constraint* (6) *is equivalent to linear constraints* (EC.10a), (EC.11a)–(EC.11d) *with $K = 2$, $\sum_{i=1}^{n+1} (\alpha_i + s_i^U\tau_i^U - s_i^L\tau_i^L) \leq \theta + z$, and* (22b)–(22g).

We remark that the LP reformulation in Proposition EC.2 is of the size $\mathcal{O}(n^3)$ because constraints (22b)–(22g) incorporate $\mathcal{O}(n^3)$ decision variables and linear constraints. In this section, we focus on a specific DR CVaR constraint (6) that restricts overtime only. That is, $c_i^{\mathrm{u}} = c_i^{\mathrm{w}} = 0$ for all $1 \leq i \leq n$ and $c^{\mathrm{o}} = 1$, and $Q(x,q,s) = Q^W(x,q,s) := \min_{w,u,W} W$ subject to constraints (3b)–(3d). Next, we derive a more compact LP reformulation of this DR CVaR constraint with $\mathcal{O}(n^2)$ variables and constraints. To that end, we derive an $\mathcal{O}(n^2)$ LP reformulation for constraint (EC.10c). We begin by specializing the extreme point representation of polyhedron Y for $Q(x,q,s)=Q^W(x,q,s)$.

LEMMA EC.1. *When $c_i^u=c_i^w=0$ for all $1\le i\le n$ and $c^o=1$, the set of extreme points of polyhedron Y defined in (15) is $\{\sum_{\ell=k}^n e_\ell : k=1,\dots,n\}\bigcup\{\mathbf{0}_n\}$, where e_ℓ represents an n-dimensional unit vector with component ℓ equaling to one and any other component equaling to zero; $\mathbf{0}_n$ is an n-dimensional zero vector.*

Recall the observation in Section 3.2 that each extreme point $(y_1,\dots,y_{n+1})$ of Y is associated with a partition of set $\{1,\dots,n+1\}$ into intervals. The result in Lemma EC.1 follows from (16) when the cost parameters take the above specified values. Define binary variables t_k for all $1\le k\le n$ to represent the set of extreme points of Y, such that $t_k=1$ if the extreme point is $\sum_{\ell=k}^n e_\ell$ and $t_k=0$ otherwise. Note that extreme point $\mathbf{0}_n$ is represented by $t_k=0$ for all $1\le k\le n$. For a valid representation, we require $\sum_{k=1}^n t_k \le 1$. It follows that the right-hand side of (EC.10c) (with $Q(x,q,s)=Q^W(x,q,s)$) is equivalent to

$$\begin{aligned}\max_{t,q,s}\quad & \sum_{k=1}^n\left(\sum_{i=k}^n(q_is_i-x_i)\right)t_k-\sum_{i=1}^n(\rho_is_i+\gamma_iq_i)\\ \text{s.t.}\quad & \sum_{k=1}^n t_k \le 1,\quad q\in D_q,\quad s\in D_s,\quad t\in\{0,1\}^n\end{aligned}$$

as a mixed-integer bilinear program with binary vectors q and t, and continuous vector s. We linearize the bilinear terms by defining $p_{ki}\equiv t_kq_i$ and $o_{ki}\equiv t_kq_is_i$ for all $1\le k\le i\le n$. Also, we introduce McCormick inequalities (EC.12b)–(EC.12c) and (EC.12d)–(EC.12e) for variables p_{ki} and o_{ki}, respectively to further reformulate the separation problem as a mixed-integer linear program:

$$\max_{t,q,s,p,o}\quad \sum_{k=1}^n\sum_{i=k}^n(o_{ki}-x_it_k)-\sum_{i=1}^n(\rho_is_i+\gamma_iq_i) \tag{EC.12a}$$

$$\text{s.t.}\quad p_{ki}-t_k\le 0\quad \forall 1\le k\le i\le n, \tag{EC.12b}$$

$$p_{ki}-q_i\le 0,\ p_{ki}-q_i-t_k\ge -1,\ p_{ki}\ge 0\quad \forall 1\le k\le i\le n, \tag{EC.12c}$$

$$o_{ki}-s_i^{\mathrm{L}}p_{ki}\ge 0,\ o_{ki}-s_i^{\mathrm{U}}p_{ki}\le 0\quad \forall 1\le k\le i\le n, \tag{EC.12d}$$

$$o_{ki}-s_i+s_i^{\mathrm{L}}(1-p_{ki})\le 0,\ o_{ki}-s_i+s_i^{\mathrm{U}}(1-p_{ki})\ge 0\quad \forall 1\le k\le i\le n, \tag{EC.12e}$$

$$\sum_{k=1}^n t_k \le 1, \tag{EC.12f}$$

$$q\in D_q,\ s\in D_s,\ t\in\{0,1\}^n. \tag{EC.12g}$$

Similar as before, we aim to derive the convex hull of the feasible region of problem (EC.12), i.e., the mixed-integer feasible region described by constraints (EC.12b)–(EC.12g). We denote the feasible region as set G and derive $\mathrm{conv}(G)$ in the following theorem, whose proof is in EC.4.2.

Theorem EC.1. *When $D_q = D_q^{(2)}$, the following inequalities are valid for set $G = \{(t,q,s,p,o) :$ (EC.12b)–(EC.12g)$\}$:*

$$\sum_{k=1}^{n} p_{kn} \le q_n, \tag{EC.13a}$$

$$p_{ki} + p_{k(i+1)} \ge t_k \quad \forall 1 \le k \le i \le n-1, \tag{EC.13b}$$

$$\sum_{k=1}^{i} (p_{ki} - t_k) \ge q_i - 1 \quad \forall 1 \le i \le n, \tag{EC.13c}$$

$$\sum_{k=1}^{i} \big(p_{ki} + p_{k(i+1)}\big) \le \sum_{k=1}^{i} t_k + q_i + q_{i+1} - 1 \quad \forall 1 \le i \le n-1, \tag{EC.13d}$$

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^L p_{ki}) \ge s_i^L \quad \forall 1 \le i \le n, \tag{EC.13e}$$

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^U p_{ki}) \le s_i^U \quad \forall 1 \le i \le n. \tag{EC.13f}$$

Furthermore, polyhedron $CG := \{(t,q,s,p,o) :$ (EC.12b), (EC.12d), (EC.13a)–(EC.13f)$\}$ is the convex hull of set G, i.e., $CG = conv(G)$.

Theorem EC.1 provides us a compact LP reformulation of the right-hand side of constraint (EC.10c) with $\mathcal{O}(n^2)$ variables and constraints:

$$\max_{t,q,s,p,o} \quad \sum_{k=1}^{n} \sum_{i=k}^{n} (o_{ki} - x_i t_k) - \sum_{i=1}^{n} (\rho_i s_i + \gamma_i q_i) \tag{EC.14a}$$

$$\text{s.t.} \quad (t,q,s,p,o) \in CG. \tag{EC.14b}$$

Finally, by resorting to the dual formulation of (EC.14), we represent constraint (EC.10c) as

$$\sum_{i=1}^{n} (\alpha_i - s_i^{\mathrm{L}} \tau_i^{\mathrm{L}} + s_i^{\mathrm{U}} \tau_i^{\mathrm{U}}) - \sum_{i=1}^{n-1} \phi_i \le \theta + z \tag{EC.15a}$$

$$\sum_{i=k}^{n} (\alpha_i - \sigma_{ki}) + \sum_{i=k}^{n-1} (\lambda_{ki} - \phi_i) \ge -\sum_{i=k}^{n} x_i \quad \forall 1 \le k \le n, \tag{EC.15b}$$

$$\alpha_i - \sum_{\ell=\max\{i-1,1\}}^{\min\{i,n-1\}} \phi_\ell - \sum_{\ell=n}^{\max\{i,n-1\}} \zeta \ge -\gamma_i \quad \forall 1 \le i \le n, \tag{EC.15c}$$

$$\tau_i^{\mathrm{U}} - \tau_i^{\mathrm{L}} \ge -\rho_i \quad \forall 1 \le i \le n, \tag{EC.15d}$$

$$\sigma_{ki} + s_i^{\mathrm{L}} \varphi_{ki}^{\mathrm{L}} - s_i^{\mathrm{U}} \varphi_{ki}^{\mathrm{U}} - \alpha_i - s_i^{\mathrm{L}} \tau_i^{\mathrm{L}} + s_i^{\mathrm{U}} \tau_i^{\mathrm{U}} + \sum_{\ell=n}^{\max\{i,n-1\}} \zeta + \sum_{\ell=\max\{i-1,k\}}^{\min\{i,n-1\}} (\phi_\ell - \lambda_{k\ell}) \ge 0 \quad \forall 1 \le k \le i \le n, \tag{EC.15e}$$

$$-\varphi_{ki}^{\mathrm{L}} + \varphi_{ki}^{\mathrm{U}} + \tau_i^{\mathrm{L}} - \tau_i^{\mathrm{U}} \ge 1 \quad \forall 1 \le k \le i \le n, \tag{EC.15f}$$

$$\sigma_{ki},\ \varphi_{ki}^{\mathrm{L}},\ \varphi_{ki}^{\mathrm{U}},\ \zeta,\ \lambda_{ki},\ \alpha_i,\ \phi_i,\ \tau_i^{\mathrm{L}},\ \tau_i^{\mathrm{U}} \ge 0 \quad \forall 1 \le k \le i \le n, \tag{EC.15g}$$

where dual variables σ_{ki}, $\varphi_{ki}^{\mathrm{L/U}}$, ζ, λ_{ki}, α_i, ϕ_i, and $\tau_i^{\mathrm{L/U}}$ are associated with constraints (EC.12b), (EC.12d), (EC.13a), (EC.13b), (EC.13c), (EC.13d), and (EC.13e)–(EC.13f), respectively (after transforming all "$\geq$" inequalities into the "$\leq$" form), and dual constraints (EC.15b)–(EC.15f) are associated with primal variables t_k, q_i, s_i, p_{ki}, and o_{ki} respectively. This results in an $\mathcal{O}(n^2)$ LP reformulation of the DR CVaR constraint on overtime.

Proposition EC.3. *When $D_q = D_q^{(2)}$, $c_i^u = c_i^w = 0$ for all $1 \leq i \leq n$ and $c^o = 1$, the DR CVaR constraint* (6) *on overtime is equivalent to linear constraints* (EC.10a)*,* (EC.11a)*–*(EC.11d) *with $K = 2$, and* (EC.15a)*–*(EC.15g)*.*

Case 2. (Arbitrary No-Shows) Recall that DR CVaR constraint (6) is equivalent to constraints (EC.10a), (EC.11a)–(EC.11d) with $K = n+1$, and (EC.10c). As $D_q = D_q^{(n+1)}$, we can apply the results in Section 4 (see Case 2) to further reformulate (EC.10c) as linear constraints $\theta + z \geq \sum_{i=1}^{n+1} \alpha_i$, (24b), (24d), and (25a)–(25d). This results in the following proposition.

Proposition EC.4. *When $D_q = D_q^{(n+1)}$, the DR CVaR constraint* (6) *is equivalent to constraints* (EC.10a)*,* (EC.11a)*–*(EC.11d) *with $K = n+1$, $\sum_{i=1}^{n+1} \alpha_i \leq \theta + z$,* (24b)*,* (24d)*, and* (25a)*–*(25d)*.*

EC.3. Proofs for the DR Expectation Model

EC.3.1. Proof of Lemma 1

Proof of Lemma 1 First, feasible regions Y and $D_q \times D_s$ are both independent of x, ρ, and γ, and bounded. Hence, $\max_{y\in Y} h(x,y,\rho,\gamma) \equiv \max_{y\in Y,(q,s)\in D_q\times D_s}\{\sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i)\} < +\infty$. Second, for any fixed y, q, and s, $\sum_{i=1}^n (q_i s_i - x_i) y_i - \sum_{i=1}^n (\rho_i s_i + \gamma_i q_i)$ is a linear function of x, ρ, and γ. It follows that $\max_{y\in Y} h(x,y,\rho,\gamma)$ is the maximum of a set of linear functions of x, ρ, and γ, and hence convex and piecewise linear. Third, it is clear that each linear piece of function $\max_{y\in Y} h(x,y,\rho,\gamma)$ is associated with one distinct extreme point of polyhedra Y, D_q, and D_s respectively. Therefore, the number of pieces of function $\max_{y\in Y} h(x,y,\rho,\gamma)$ is finite because each of these polyhedra has a finite number of extreme points. This completes the proof. □

EC.3.2. Proof of Lemma 2

Proof of Lemma 2 For fixed x, ρ, and γ, in view of the definition of function $h(x,y,\rho,\gamma)$ in (12c), we have $h(x,y,\rho,\gamma) = \max_{(q,s)\in D_q\times D_s} H(q,s,y)$, where $H(q,s,y)$ is a linear function of variable y. It follows that $h(x,y,\rho,\gamma)$ is the supremum of a set of convex functions of y, and hence itself convex in variable y. □

EC.3.3. Proof of Proposition 1

Proof of Proposition 1 First, because $p_{ikj} \equiv q_i t_{kj}$, equality (20a) can be obtained via multiplying equalities $\sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} = 1$ by q_i on both sides.

Second, because $o_{ikj} \equiv q_i s_i t_{kj} \equiv s_i p_{ikj}$, and by equalities (20a) and $s_i \in [s_i^{\mathrm{L}}, s_i^{\mathrm{U}}]$, we have

$$\sum_{k=1}^{i}\sum_{j=i}^{n+1}(o_{ikj} - s_i^{\mathrm{L}} p_{ikj}) \;=\; (s_i - s_i^{\mathrm{L}})\sum_{k=1}^{i}\sum_{j=i}^{n+1} p_{ikj} \;=\; (s_i - s_i^{\mathrm{L}})q_i \;\le\; (s_i - s_i^{\mathrm{L}}),$$
$$\sum_{k=1}^{i}\sum_{j=i}^{n+1}(o_{ikj} - s_i^{\mathrm{U}} p_{ikj}) \;=\; (s_i - s_i^{\mathrm{U}})\sum_{k=1}^{i}\sum_{j=i}^{n+1} p_{ikj} \;=\; (s_i - s_i^{\mathrm{U}})q_i \;\ge\; (s_i - s_i^{\mathrm{U}}),$$

which shows the validity of inequalities (20b) and (20c).

Third, for $1 \le k < j \le n+1$ and $k \le i \le j - K + 1$, because $\sum_{\ell=i}^{i+K-1} q_\ell \ge 1$ by the definition of $D_q^{(K)}$, we have

$$\sum_{\ell=i}^{i+K-1} p_{\ell kj} \;=\; \sum_{\ell=i}^{i+K-1} q_\ell t_{kj} \;=\; \left(\sum_{\ell=i}^{i+K-1} q_\ell\right) t_{kj} \;\ge\; t_{kj},$$

which shows the validity of inequalities (20d).

Fourth, for $i = 1, \ldots, n$, because $\sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} = 1$ and $\sum_{k=1}^{i+1}\sum_{j=i+1}^{n+1} t_{kj} = 1$ by constraints (17b), we have

$$0 \;=\; \sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} - \sum_{k=1}^{i+1}\sum_{j=i+1}^{n+1} t_{kj} \;=\; \sum_{k=1}^{i} t_{ki} - \sum_{j=i+1}^{n+1} t_{(i+1)j}. \tag{EC.16}$$

We show the validity of inequalities (20e) for all $i = K-1, \ldots, n$. If $\sum_{k=1}^{i-K+2} t_{ki} = 0$, then the conclusion holds because each $p_{ikj} \ge 0$. Now suppose that $\sum_{k=1}^{i-K+2} t_{ki} = 1$, then $\sum_{j=i+1}^{n+1} t_{(i+1)j} = 1$ in view of (EC.16). It follows that

$$\begin{aligned}\sum_{k=1}^{i-K+2}\sum_{\ell=i-K+2}^{i} p_{\ell ki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \;&=\; \left(\sum_{\ell=i-K+2}^{i} q_\ell\right)\left(\sum_{k=1}^{i-K+2} t_{ki}\right) + q_{i+1}\sum_{j=i+1}^{n+1} t_{(i+1)j}\\ &=\; \sum_{\ell=i-K+2}^{i+1} q_\ell \;\ge\; 1,\end{aligned}$$

where the last inequality is due to the definition of $D_q^{(K)}$.

Finally, we show the validity of inequalities (20f) for all $i = 1, \ldots, n-K+2$. If $\sum_{j=i+K-1}^{n+1} t_{(i+1)j} = 0$, then the conclusion holds because each $p_{ikj} \ge 0$. Now suppose that $\sum_{j=i+K-1}^{n+1} t_{(i+1)j} = 1$, then $\sum_{k=1}^{i} t_{ki} = 1$ in view of (EC.16). It follows that

$$\begin{aligned}\sum_{k=1}^{i} p_{iki} + \sum_{\ell=i+1}^{i+K-1}\sum_{j=i+K-1}^{n+1} p_{\ell(i+1)j} \;&=\; q_i\left(\sum_{k=1}^{i} t_{ki}\right) + \left(\sum_{\ell=i+1}^{i+K-1} q_\ell\right)\left(\sum_{j=i+K-1}^{n+1} t_{(i+1)j}\right)\\ &=\; \sum_{\ell=i}^{i+K-1} q_\ell \;\ge\; 1,\end{aligned}$$

where the last inequality is due to the definition of $D_q^{(K)}$. □

EC.3.4. Proof of Theorem 1

Recall that polyhedron $CF = \{(t,q,s,p,o) : \text{(17b), (18a), (18c), (20a)–(20d), (21)}\}$ in Theorem 1. We first study the extreme points of polyhedron CF and show their properties as follows.

Proposition EC.5. *Every extreme point (t,q,s,p,o) of CF satisfies the following:*

1. $t_{kj}, p_{ikj} \in \{0,1\}$ *for all* $1 \le k \le j \le n+1$ *and* $k \le i \le j$*;*
2. $q_i \in \{0,1\}$ *for all* $1 \le i \le n+1$*;*
3. $p_{ikj} = q_i t_{kj}$ *and* $o_{ikj} = q_i s_i t_{kj}$ *for all* $1 \le k \le j \le n+1$ *and* $k \le i \le j$*.*

Proof of Proposition EC.5 Consider arbitrary cost coefficients c^{q}_i and c^{s}_i for all $1 \le i \le n+1$, c^{t}_{kj} for all $1 \le k \le j \le n+1$, and c^{p}_{ikj} and c^{o}_{ikj} for all $1 \le k \le j \le n+1$ and $k \le i \le j$. We construct a related linear program

$$\text{(LP-CF)} \quad \min_{t,q,s,p,o} \ \sum_{i=1}^{n+1} (c^{\text{q}}_i q_i + c^{\text{s}}_i s_i) + \sum_{k=1}^{n+1}\sum_{j=k}^{n+1} \left(c^{\text{t}}_{kj} t_{kj} + \sum_{i=k}^{j} \left(c^{\text{p}}_{ikj} p_{ikj} + c^{\text{o}}_{ikj} o_{ikj} \right) \right)$$
$$\text{s.t.} \quad (t,q,s,p,o) \in CF.$$

To prove that each extreme point of CF satisfies properties 1, 2, and 3, we show for any values of c^{t}_{kj}, c^{q}_i, c^{s}_i, c^{p}_{ikj}, and c^{o}_{ikj}, there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CF) that satisfies properties 1, 2, and 3 (cf. Wolsey 1998, Nemhauser and Wolsey 1999).

First, in view of equalities (20a), we can assume that $c^{\text{q}}_i = 0$ for all $1 \le i \le n+1$ w.l.o.g., because we can always replace each c^{p}_{ikj} with $c^{\text{p}}_{ikj} + c^{\text{q}}_i$ so that variables p_{ikj} will carry the cost of decisions q_i. It follows that we can ignore variables q_i in (LP-CF) because they do not contribute to the objective function and their values entirely depend on p_{ikj} by constraints (20a). Also, we note that (i) $s^{\text{L}}_i p_{ikj} \le o_{ikj} \le s^{\text{U}}_i p_{ikj}$ by (18c), and so $p_{ikj} \ge 0$ for all $1 \le k \le i \le j \le n+1$, and (ii) for all $1 \le i \le n+1$, $\sum_{k=1}^{i}\sum_{j=i}^{n+1} p_{ikj} \le \sum_{k=1}^{i}\sum_{j=i}^{n+1} t_{kj} = 1$ by (18a) and (17b).

Second, we rewrite (LP-CF) as a two-stage formulation as follows:

$$\min_{t,p} \ \sum_{k=1}^{n+1}\sum_{j=k}^{n+1} \left(c^{\text{t}}_{kj} t_{kj} + \sum_{i=k}^{j} c^{\text{p}}_{ikj} p_{ikj} \right) + V(p)$$
$$\text{s.t.} \quad (t,p) \in CF_{t,p},$$

where polyhedron $CF_{t,p} := \{(t,p) : \text{(17b), (18a), (20d), (21)}\}$ and $V(p)$ represents a value function of p defined as

$$\text{(LP-CF}(p)) \quad V(p) := \min_{s,o} \ \sum_{i=1}^{n+1} c^{\text{s}}_i s_i + \sum_{k=1}^{n+1}\sum_{j=k}^{n+1}\sum_{i=k}^{j} c^{\text{o}}_{ikj} o_{ikj}$$
$$\text{s.t.} \quad (s,o) \in CF_{s,o}(p),$$

where

$$CF_{s,o}(p) = \Bigg\{ (s,o): \quad o_{ikj} \ge s_i^{\mathrm{L}} p_{ikj} \quad \forall 1 \le k \le j \le n+1,\ \forall k \le i \le j, \tag{EC.17a}$$

$$o_{ikj} \le s_i^{\mathrm{U}} p_{ikj} \quad \forall 1 \le k \le j \le n+1,\ \forall k \le i \le j, \tag{EC.17b}$$

$$s_i - \sum_{k=1}^{i} \sum_{j=i}^{n+1} (o_{ikj} - s_i^{\mathrm{L}} p_{ikj}) \ge s_i^{\mathrm{L}} \quad \forall 1 \le i \le n+1, \tag{EC.17c}$$

$$s_i - \sum_{k=1}^{i} \sum_{j=i}^{n+1} (o_{ikj} - s_i^{\mathrm{U}} p_{ikj}) \le s_i^{\mathrm{U}} \quad \forall 1 \le i \le n+1 \Bigg\} \tag{EC.17d}$$

represents a parametric polyhedron depending on the values of p_{ikj}. We solve (LP-CF(p)) by considering its dual formulation

$$V(p) = \max_{\psi,\omega} \quad \sum_{i=1}^{n+1} \sum_{k=1}^{i} \sum_{j=i}^{n+1} \left(s_i^{\mathrm{L}} p_{ikj} \psi_{ikj}^{\mathrm{L}} - s_i^{\mathrm{U}} p_{ikj} \psi_{ikj}^{\mathrm{U}} \right) + \sum_{i=1}^{n+1} \left[s_i^{\mathrm{L}} \left(1 - \sum_{k=1}^{i} \sum_{j=i}^{n+1} p_{ikj} \right) \omega_i^{\mathrm{L}} - s_i^{\mathrm{U}} \left(1 - \sum_{k=1}^{i} \sum_{j=i}^{n+1} p_{ikj} \right) \omega_i^{\mathrm{U}} \right]$$

$$\text{s.t.} \quad \psi_{ikj}^{\mathrm{L}} - \psi_{ikj}^{\mathrm{U}} - \omega_i^{\mathrm{L}} + \omega_i^{\mathrm{U}} = c_{ikj}^{\mathrm{o}} \quad \forall 1 \le k \le j \le n+1,\ \forall k \le i \le j, \tag{EC.18a}$$

$$\omega_i^{\mathrm{L}} - \omega_i^{\mathrm{U}} = c_i^{\mathrm{s}} \quad \forall 1 \le i \le n+1, \tag{EC.18b}$$

where dual variables $\psi_{ikj}^{\mathrm{L/U}}$ and $\omega_i^{\mathrm{L/U}}$ are associated with primal constraints (EC.17a)–(EC.17b) and (EC.17c)–(EC.17d), respectively (after transforming all "$\le$" inequalities into the "$\ge$" form), and dual constraints (EC.18a) and (EC.18b) are associated with primal variables o_{ikj} and s_i, respectively. Because $p_{ikj} \ge 0$ for all $1 \le k \le i \le j \le n+1$ and $1 - \sum_{k=1}^{i} \sum_{j=i}^{n+1} p_{ikj} \ge 0$ for all $1 \le i \le n+1$, a dual optimal solution to problem (LP-CF(p)) is $\psi_{ikj}^{\mathrm{L}*} = (c_{ikj}^{\mathrm{o}} + c_i^{\mathrm{s}})^+$, $\psi_{ikj}^{\mathrm{U}*} = (-c_{ikj}^{\mathrm{o}} - c_i^{\mathrm{s}})^+$, $\omega_i^{\mathrm{L}*} = (c_i^{\mathrm{s}})^+$, and $\omega_i^{\mathrm{U}*} = (-c_i^{\mathrm{s}})^+$. It follows that

$$V(p) = \sum_{i=1}^{n+1} \left[s_i^{\mathrm{L}} (c_i^{\mathrm{s}})^+ - s_i^{\mathrm{U}} (-c_i^{\mathrm{s}})^+ \right] + \sum_{i=1}^{n+1} \sum_{k=1}^{i} \sum_{j=i}^{n+1} \left[s_i^{\mathrm{L}} (c_{ikj}^{\mathrm{o}} + c_i^{\mathrm{s}})^+ - s_i^{\mathrm{U}} (-c_{ikj}^{\mathrm{o}} - c_i^{\mathrm{s}})^+ - s_i^{\mathrm{L}} (c_i^{\mathrm{s}})^+ + s_i^{\mathrm{U}} (-c_i^{\mathrm{s}})^+ \right] p_{ikj}$$

is a linear function of p. Therefore, (LP-CF) is equivalent to optimizing a linear function of (t, p) on polyhedron $CF_{t,p}$. It follows that there exists an optimal solution (t^*, p^*, s^*, o^*) to (LP-CF) where (t^*, p^*) is an extreme point of polyhedron $CF_{t,p}$.

Third, we show that all extreme points of $CF_{t,p}$ are integral. To this end, we show that the constraint matrix describing $CF_{t,p}$ is totally unimodular (TU). For presentation convenience, we rewrite the constraints defining $CF_{t,p}$ in inequalities as follows:

$$\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} \ \ge 1 \quad \forall i = 1, \ldots, n+1, \tag{EC.19a}$$

$$-\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} \ \ge -1 \quad \forall i = 1, \ldots, n+1, \tag{EC.19b}$$

$$-t_{kj} + p_{ikj} + p_{(i+1)kj} \ \ge 0 \quad \forall 1 \le k < j \le n+1,\ \forall k \le i \le j-1, \tag{EC.19c}$$

$$-\sum_{k=1}^{i} t_{ki} + \sum_{k=1}^{i} p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j} \ \ge 0 \quad \forall i = 1, \ldots, n, \tag{EC.19d}$$

$$t_{kj} - p_{ikj} \ \geq 0 \quad \forall 1 \leq k \leq j \leq n+1, \ \forall k \leq i \leq j, \tag{EC.19e}$$

and we denote the constraint matrix as

$$\mathcal{CF}^0_{t,p} \ := \ \begin{bmatrix} (\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}), \quad \forall 1 \leq i \leq n+1 \\ (-\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}), \quad \forall 1 \leq i \leq n+1 \\ (-t_{kj} + p_{ikj} + p_{(i+1)kj}), \quad \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \\ (-\sum_{k=1}^{i} t_{ki} + \sum_{k=1}^{i} p_{iki} + \sum_{j=i+1}^{n+1} p_{(i+1)(i+1)j}), \quad \forall 1 \leq i \leq n \\ (t_{kj} - p_{ikj}), \quad \forall 1 \leq k \leq i \leq j \leq n+1 \end{bmatrix},$$

where the five row sub-matrices in matrix $\mathcal{CF}^0_{t,p}$ are associated with the left-hand side of constraints (EC.19a)–(EC.19e), respectively. To show that matrix $\mathcal{CF}^0_{t,p}$ is TU, we conduct pivot operations on the matrix with variables p_{ikj} and t_{kj}. Note that a matrix is TU if and only if it remains TU after pivot operations (Nemhauser and Wolsey 1999). We conduct the following pivot operations in order.

(i) For all $1 \leq k \leq j \leq n+1$ and $k \leq i \leq j$, pivot with variable p_{ikj} based on the component -1 in sub-matrix $(t_{kj} - p_{ikj})$ (corresponding to constraints (EC.19e)). This pivot operation is equivalent to (a) adding $t_{kj} - p_{ikj}$, for all $1 \leq k \leq j \leq n+1$ and $k \leq i \leq j$, to the left-hand side of every constraint (EC.19c)–(EC.19d) in which variable p_{ikj} has coefficient 1, and (b) multiplying the left-hand side of each constraint (EC.19e) by -1. As a result, the matrix after pivoting becomes

$$\mathcal{CF}^1_{t,p} \ := \ \begin{bmatrix} (\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}), \quad \forall 1 \leq i \leq n+1 \\ (-\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}), \quad \forall 1 \leq i \leq n+1 \\ (t_{kj}), \quad \forall 1 \leq k < j \leq n+1, \forall k \leq i \leq j-1 \\ (\sum_{j=i+1}^{n+1} t_{(i+1)j}), \quad \forall 1 \leq i \leq n \\ (-t_{kj} + p_{ikj}), \quad \forall 1 \leq k \leq i \leq j \leq n+1 \end{bmatrix}.$$

Note that sub-matrix $(-t_{kj} + p_{ikj} + p_{(i+1)kj})$ becomes (t_{kj}) because each $-t_{kj} + p_{ikj} + p_{(i+1)kj}$ on the left-hand side of (EC.19c) is summed with $t_{kj} - p_{ikj}$ and $t_{kj} - p_{(i+1)kj}$, and so the coefficient of each t_{kj} changes from -1 to 1 after pivoting.

(ii) For all $1 \leq k < j \leq n+1$, pivot with variable t_{kj} based on any component 1 in sub-matrix (t_{kj}) (note that there are multiple components 1 corresponding to each variable t_{kj} in sub-matrix (t_{kj}) and we can pick any one of them). Since all components in each row of sub-matrix (t_{kj}) are zeros except one equaling 1, these pivot operations (a) make all coefficients of all variables t_{kj} zeros in matrix $\mathcal{CF}^1_{t,p}$ as long as $1 \le k < j \le n+1$, and (b) keep all coefficients of all variables p_{ikj} unchanged. As a result, the matrix after pivoting becomes

$$\mathcal{CF}^2_{t,p} := \begin{bmatrix} (t_{ii}), \quad \forall 1 \le i \le n+1 \\ (-t_{ii}), \quad \forall 1 \le i \le n+1 \\ (t_{kj}), \quad \forall 1 \le k < j \le n+1, \forall k \le i \le j-1 \\ (t_{(i+1)(i+1)}), \quad \forall 1 \le i \le n \\ \begin{cases} (-t_{ii} + p_{iii}), \quad \forall 1 \le i \le n+1 \\ (p_{ikj}), \quad 1 \le k < j \le n+1, \forall k \le i \le j-1 \end{cases} \end{bmatrix}.$$

It follows that matrix $\mathcal{CF}^2_{t,p}$ contains only $\{-1, 0, 1\}$ entries, has no more than two nonzero entries in each row, and the sum of the entries is zero for each row containing two nonzero entries. Hence, matrix $\mathcal{CF}^2_{t,p}$ is TU, and so is matrix $\mathcal{CF}^0_{t,p}$.

Therefore, the extreme points of polyhedron $CF_{t,p}$ are integral and so property 1 is proved.

Fourth, to show property 2, we consider any extreme point (t, q, s, p, o) of polyhedron CF. By constraints (18a) and (20a), we have $q_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} p_{ikj} \le \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1$, and so $q_i \in \{0, 1\}$ because each $p_{ikj} \in \{0, 1\}$ by property 1. This shows property 2.

Finally, to show property 3, we consider any extreme point (t, q, s, p, o) of polyhedron CF. We show $p_{ikj} = q_i t_{kj}$ by discussing the following cases.

(i) If $q_i = 0$, then $p_{ikj} = 0$ for all $1 \le k \le i$ and $i \le j \le n+1$ because $q_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} p_{ikj}$. It follows that $p_{ikj} = q_i t_{kj}$.

(ii) If $q_i = 1$, then there exist $1 \le k^* \le i$ and $i \le j^* \le n+1$ such that $p_{ik^*j^*} = 1$ and any other $p_{ikj} = 0$. It follows that $t_{k^*j^*} = 1$ because $p_{ikj} - t_{kj} \le 0$ by constraint (18a), and any other $t_{kj} = 0$ because $\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1$ given by (17b). Therefore, we have $p_{ik^*j^*} = q_i t_{k^*j^*} = 1$ and $p_{ikj} = q_i t_{kj} = 0$ for all other $1 \le k \le i$ and $i \le j \le n+1$.

For all $1 \le i \le n+1$, since $\sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1$, there exist $1 \le k^* \le i$ and $i \le j^* \le n+1$ such that $t_{k^*j^*} = 1$ and any other $t_{kj} = 0$. Since (t, q, s, p, o) is an extreme point of polyhedron CF, each o_{ikj} satisfies either inequality (EC.17a) or (EC.17b) at equality, and each s_i satisfies either inequality (EC.17c) or (EC.17d) at equality. We discuss the following two cases to show $o_{ikj} = q_i s_i t_{kj}$.

(i) If $q_i = 0$, then $p_{ikj} = q_i t_{kj} = 0$ for all $1 \le k \le i$ and $i \le j \le n+1$. It follows from inequalities (EC.17a)–(EC.17b) that each corresponding $o_{ikj} = 0$. Therefore, we have $o_{ikj} = s_i p_{ikj} = 0$, or equivalently $o_{ikj} = q_i s_i t_{kj} = 0$, for all $1 \le k \le j \le n+1$ and $k \le i \le j$.

(ii) If $q_i = 1$, then $p_{ik^*j^*} = q_i t_{k^*j^*} = 1$ and $p_{ikj} = 0$ for all other $1 \le k \le i$ and $i \le j \le n+1$. Then, inequalities (EC.17a)–(EC.17b) yield $o_{ikj} = s_i p_{ikj} = 0$ for all $1 \le k \le i$ and $i \le j \le n+1$ such that $(k, j) \ne (k^*, j^*)$. Furthermore, inequalities (EC.17c)–(EC.17d) yield

$$s_i - \sum_{k=1}^{i} \sum_{j=i}^{n+1} (o_{ikj} - s_i^{\mathrm{L}} p_{ikj}) \; = \; s_i - o_{ik^*j^*} + s_i^{\mathrm{L}} p_{ik^*j^*} \; \ge \; s_i^{\mathrm{L}},$$

$$s_i - \sum_{k=1}^{i}\sum_{j=i}^{n+1}(o_{ikj} - s_i^{\mathrm{U}} p_{ikj}) \; = s_i - o_{ik^*j^*} + s_i^{\mathrm{U}} p_{ik^*j^*} \; \le \; s_i^{\mathrm{U}}.$$

It follows that $s_i = o_{ik^*j^*}$. Therefore, we have $o_{ik^*j^*} = s_i p_{ik^*j^*}$.

□

We are now ready to show Theorem 1.

Proof of Theorem 1 ($CF \supseteq \mathrm{conv}(F)$) By Proposition 1, since polyhedron CF consists of either trivial equalities/inequalities or valid inequalities of set F, we have $(t,q,s,p,o) \in CF$ if $(t,q,s,p,o) \in F$. It follows that $CF \supseteq \mathrm{conv}(F)$.

($CF \subseteq \mathrm{conv}(F)$) By Proposition EC.5, since each extreme point (t,q,s,p,o) of CF satisfies properties 1, 2, and 3, $(t,q,s,p,o) \in F$. By the Minkowski's Theorem on polyhedron, we have $x \in \mathrm{conv}(F)$ if $x \in CF$. It follows that $CF \subseteq \mathrm{conv}(F)$. This completes the proof. □

EC.4. Proofs for the DR CVaR Model

EC.4.1. Proof of Proposition EC.1

Proof of Proposition EC.1 We analyze the following two cases based on the value of K.

When $K \in \{2,\dots,n\}$: For the embedded minimization problem in constraint (EC.10b), we observe that the constraint matrix of D_q, described by constraints $\sum_{j=i}^{i+K-1} q_j \ge 1$ for all $1 \le i \le n-K+1$, is an interval matrix and thus TU. It follows that $\mathrm{conv}(D_q) = \{q \in [0,1]^n : \sum_{j=i}^{i+K-1} q_j \ge 1, \ \forall 1 \le i \le n-K+1\}$. Because the feasible regions of variables q and s (i.e., D_q and D_s) are disjoint in (EC.10b), we can replace D_q with $\mathrm{conv}(D_q)$ and obtain

$$\theta + \min_{q,s} \left\{ \sum_{i=1}^{n} \rho_i s_i + \sum_{i=1}^{n} \gamma_i q_i \right\} \ge 0 \tag{EC.20a}$$

$$\text{s.t.} \quad s_i^{\mathrm{L}} \le s_i \le s_i^{\mathrm{U}} \quad \forall 1 \le i \le n, \tag{EC.20b}$$

$$\sum_{j=i}^{i+K-1} q_j \ge 1 \quad \forall 1 \le i \le n-K+1, \tag{EC.20c}$$

$$q_i \le 1 \quad \forall 1 \le i \le n, \tag{EC.20d}$$

$$q_i, s_i \ge 0 \quad \forall 1 \le i \le n. \tag{EC.20e}$$

Presenting linear program (EC.20) in its dual form yields (EC.11a)–(EC.11d), where dual variables $\chi_i^{\mathrm{L/U}}$, β_i, and η_i are associated with constraints (EC.20b), (EC.20c), and (EC.20d) respectively, and dual constraints (EC.11b) and (EC.11c) are associated with primal variables q_i and s_i, respectively.

When $K = n+1$: In this case, $D_q = \{0,1\}^n$ and so $\mathrm{conv}(D_q) = [0,1]^n$. It follows that constraint (EC.10b) is equivalent to

$$\theta + \min_{q,s} \left\{ \sum_{i=1}^{n} \rho_i s_i + \sum_{i=1}^{n} \gamma_i q_i \right\} \ge 0$$

$$\text{s.t.}\quad s_i^{\mathrm{L}} \le s_i \le s_i^{\mathrm{U}} \quad \forall 1 \le i \le n,$$
$$q_i \le 1 \quad \forall 1 \le i \le n,$$
$$q_i, s_i \ge 0 \quad \forall 1 \le i \le n,$$

Similar to the case when $K \in \{2, \ldots, n\}$, we can present the embedded LP in its dual form to obtain the following linear constraints:

$$\begin{aligned}
\theta + \sum_{i=1}^{n} (s_i^{\mathrm{L}} \chi_i^{\mathrm{L}} - s_i^{\mathrm{U}} \chi_i^{\mathrm{U}} - \eta_i) \ &\ge 0, \\
-\eta_i \ &\le \gamma_i \quad \forall 1 \le i \le n, \\
\chi_i^{\mathrm{L}} - \chi_i^{\mathrm{U}} \ &\le \rho_i \quad \forall 1 \le i \le n, \\
\chi_i^{\mathrm{L}},\ \chi_i^{\mathrm{U}}, \eta_i \ge 0 & \quad \forall 1 \le i \le n.
\end{aligned}$$

We note that these linear constraints are equivalent to (EC.11a)–(EC.11d) because $\sum_{i=1}^{n-K+1} \beta_i = \sum_{i=1}^{0} \beta_i = 0$ and $\sum_{j=\max\{i-K+1,1\}}^{\min\{i,n-K+1\}} \beta_j = \sum_{j=1}^{0} \beta_j = 0$. The proof is completed. □

EC.4.2. Proof of Theorem EC.1

We take the following three steps to prove Theorem EC.1.

Step 1: We prove the validity of inequalities (EC.13a)–(EC.13f) in the following proposition.

PROPOSITION EC.6. *When* $D_q = D_q^{(2)}$, *inequalities* (EC.13a)–(EC.13f) *are valid for set* $G = \{(t, q, s, p, o) : \text{(EC.12b)–(EC.12g)}\}$.

Proof of Proposition EC.6 First, since $\sum_{k=1}^{n} t_k \le 1$ by constraint (EC.12f) and $q_n \ge 0$, we have

$$\sum_{k=1}^{n} p_{kn} \ = \ q_n \sum_{k=1}^{n} t_k \ \le \ q_n,$$

which shows inequality (EC.13a).

Second, for all $1 \le k \le i \le n-1$, since $q_i + q_{i+1} \ge 1$ by the definition of D_q and $t_k \ge 0$, we have

$$p_{ki} + p_{k(i+1)} \ = \ t_k(q_i + q_{i+1}) \ \ge \ t_k,$$

which shows inequalities (EC.13b).

Third, for all $1 \le i \le n$, since $t_i \ge 0,\ \forall i$ and thus $\sum_{k=1}^{i} t_k \le \sum_{k=1}^{n} t_k \le 1$ by constraint (EC.12f) and since $q_i \le 1 \ \Rightarrow\ q_i - 1 \le 0$, we have

$$\sum_{k=1}^{i} (p_{ki} - t_k) \ = \ (q_i - 1) \sum_{k=1}^{i} t_k \ \ge \ q_i - 1,$$

which shows inequality (EC.13c).

Fourth, for all $1 \le i \le n-1$ because (a) $\sum_{k=1}^{i} t_k \le \sum_{k=1}^{n} t_k \le 1$ by constraint (EC.12f) and $t_k \ge 0$, $\forall k$, and (b) $q_i + q_{i+1} \ge 1$ by the definition of D_q, we have

$$\begin{aligned}\sum_{k=1}^{i}(p_{ki} + p_{k(i+1)}) - (q_i + q_{i+1}) \;&=\; (q_i + q_{i+1}) \sum_{k=1}^{i} t_k - (q_i + q_{i+1}) \\ =\; (q_i + q_{i+1}) \left(\sum_{k=1}^{i} t_k - 1 \right) \;&\le\; \sum_{k=1}^{i} t_k - 1,\end{aligned}$$

which shows inequality (EC.13d).

Finally, for each $1 \le i \le n$, since $\sum_{k=1}^{n} t_k \le 1$ by constraint (EC.12f), and $t_k, q_i \in [0,1]$, we have $\sum_{k=1}^{i} p_{ki} = q_i(\sum_{k=1}^{i} t_k) \le q_i(\sum_{k=1}^{n} t_k) \le 1$. Also, because $s_i \in [s_i^{\mathrm{L}}, s_i^{\mathrm{U}}]$, it follows that

$$\begin{aligned}\sum_{k=1}^{i}(o_{ki} - s_i^{\mathrm{L}} p_{ki}) \;&=\; (s_i - s_i^{\mathrm{L}}) \sum_{k=1}^{i} p_{ki} \;\le\; s_i - s_i^{\mathrm{L}}, \\ \sum_{k=1}^{i}(o_{ki} - s_i^{\mathrm{U}} p_{ki}) \;&=\; (s_i - s_i^{\mathrm{U}}) \sum_{k=1}^{i} p_{ki} \;\ge\; s_i - s_i^{\mathrm{U}},\end{aligned}$$

which shows inequalities (EC.13e)–(EC.13f). □

Step 2: We show the properties of the extreme points of polyhedron CG in the following proposition. Recall that $CG = \{(t,q,s,p,o) : \text{(EC.12b), (EC.12d), (EC.13a)–(EC.13f)}\}$.

Proposition EC.7. *Each extreme point (t,q,s,p,o) of CG has the following properties:*

1. $t_k, q_i, p_{ki} \in \{0,1\}$ *for all* $1 \le k \le i \le n$;
2. $p_{ki} = t_k q_i$ *and* $o_{ki} = t_k q_i s_i$ *for all* $1 \le k \le i \le n$.

Proof of Proposition EC.7 For any c_k^{t}, c_i^{q}, c_i^{s}, c_{ki}^{p}, and c_{ki}^{o} for all $1 \le k \le i \le n$, we consider linear program

$$\text{(LP-CG)} \quad \min_{t,q,s,p,o} \; \sum_{i=1}^{n}(c_i^{\mathrm{q}} q_i + c_i^{\mathrm{s}} s_i) + \sum_{k=1}^{n} \left(c_k^{\mathrm{t}} t_k + \sum_{i=k}^{n} (c_{ki}^{\mathrm{p}} p_{ki} + c_{ki}^{\mathrm{o}} o_{ki}) \right) \quad \text{s.t.} \; (t,q,s,p,o) \in CG.$$

To prove that each extreme point of CG satisfies properties 1 and 2, we show that for any c_k^{t}, c_i^{q}, c_i^{s}, c_{ki}^{p}, and c_{ki}^{o}, there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CG) that satisfies properties 1 and 2. First, we rewrite (LP-CG) as a two-stage formulation as follows:

$$\min_{t,q,p} \; \sum_{i=1}^{n} c_i^{\mathrm{q}} q_i + \sum_{k=1}^{n} \left(c_k^{\mathrm{t}} t_k + \sum_{i=k}^{n} c_{ki}^{\mathrm{p}} p_{ki} \right) + V(p) \quad \text{s.t.} \; (t,q,p) \in CG_{t,q,p},$$

where polyhedron $CG_{t,q,p} := \{(t,q,p) : \text{(EC.12b), (EC.13a)–(EC.13d)}\}$ and $V(p)$ represents a value function of p defined as

$$\text{(LP-CG}(p)\text{)} \quad V(p) := \min_{s,o} \left\{ \sum_{i=1}^{n} c_i^{\mathrm{s}} s_i + \sum_{k=1}^{n} \sum_{i=k}^{n} c_{ki}^{\mathrm{o}} o_{ki} : (s,o) \in CG_{s,o}(p) \right\},$$

where $CG_{s,o}(p) = \Big\{(s,o):\ \text{(EC.12d), (EC.13e), (EC.13f)}\Big\}$

$$= \Big\{(s,o):\ o_{ki} \ge s_i^{\mathrm{L}} p_{ki} \quad \forall 1 \le k \le i \le n, \tag{EC.21a}$$

$$o_{ki} \le s_i^{\mathrm{U}} p_{ki} \quad \forall 1 \le k \le i \le n, \tag{EC.21b}$$

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^{\mathrm{L}} p_{ki}) \ge s_i^{\mathrm{L}} \quad \forall 1 \le i \le n, \tag{EC.21c}$$

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^{\mathrm{U}} p_{ki}) \le s_i^{\mathrm{U}} \quad \forall 1 \le i \le n \Big\} \tag{EC.21d}$$

represents a parametric polyhedron depending on the values of p_{ki}. We solve (LP-CG(p)) by considering its dual formulation

$$V(p) = \max_{\psi,\omega} \quad \sum_{i=1}^{n} \sum_{k=1}^{i} (s_i^{\mathrm{L}} p_{ki} \psi_{ki}^{\mathrm{L}} - s_i^{\mathrm{U}} p_{ki} \psi_{ki}^{\mathrm{U}}) + \sum_{i=1}^{n} \left[s_i^{\mathrm{L}} \left(1 - \sum_{k=1}^{i} p_{ki}\right) \omega_i^{\mathrm{L}} - s_i^{\mathrm{U}} \left(1 - \sum_{k=1}^{i} p_{ki}\right) \omega_i^{\mathrm{U}} \right]$$

$$\text{s.t.} \quad \psi_{ki}^{\mathrm{L}} - \psi_{ki}^{\mathrm{U}} - \omega_i^{\mathrm{L}} + \omega_i^{\mathrm{U}} = c_{ki}^{\mathrm{o}} \quad \forall 1 \le k \le i \le n, \tag{EC.22a}$$

$$\omega_i^{\mathrm{L}} - \omega_i^{\mathrm{U}} = c_i^{\mathrm{s}} \quad \forall 1 \le i \le n, \tag{EC.22b}$$

where dual variables $\psi_{ki}^{\mathrm{L/U}}$ and $\omega_i^{\mathrm{L/U}}$ are associated with primal constraints (EC.21a)–(EC.21b) and (EC.21c)–(EC.21d), respectively (after transforming all "$\le$" inequalities into the "$\ge$" form), and dual constraints (EC.22a) and (EC.22b) are associated with primal variables o_{ki} and s_i, respectively. Because (i) $s_i^{\mathrm{L}} p_{ki} \le o_{ki} \le s_i^{\mathrm{U}} p_{ki}$ by (EC.12d), and so $p_{ki} \ge 0$ for all $1 \le k \le i \le n$, and (ii) $s_i^{\mathrm{L}}(1 - \sum_{k=1}^{i} p_{ki}) \le s_i - \sum_{k=1}^{i} o_{ki} \le s_i^{\mathrm{U}}(1 - \sum_{k=1}^{i} p_{ki})$ by (EC.13e)–(EC.13f), and so $1 - \sum_{k=1}^{i} p_{ki} \ge 0$ for all $1 \le i \le n$, a dual optimal solution to problem (LP-CG(p)) is $\psi_{ki}^{\mathrm{L}*} = (c_{ki}^{\mathrm{o}} + c_i^{\mathrm{s}})^+$, $\psi_{ki}^{\mathrm{U}*} = (-c_{ki}^{\mathrm{o}} - c_i^{\mathrm{s}})^+$, $\omega_i^{\mathrm{L}*} = (c_i^{\mathrm{s}})^+$, and $\omega_i^{\mathrm{U}*} = (-c_i^{\mathrm{s}})^+$. It follows that

$$V(p) = \sum_{i=1}^{n} \left[s_i^{\mathrm{L}} (c_i^{\mathrm{s}})^+ - s_i^{\mathrm{U}} (-c_i^{\mathrm{s}})^+ \right] + \sum_{i=1}^{n} \sum_{k=1}^{i} \left[s_i^{\mathrm{L}} (c_{ki}^{\mathrm{o}} + c_i^{\mathrm{s}})^+ - s_i^{\mathrm{U}} (-c_{ki}^{\mathrm{o}} - c_i^{\mathrm{s}})^+ - s_i^{\mathrm{L}} (c_i^{\mathrm{s}})^+ + s_i^{\mathrm{U}} (-c_i^{\mathrm{s}})^+ \right] p_{ki}$$

is a linear function of p. Therefore, (LP-CG) is equivalent to optimizing a linear function of (t, q, p) on polyhedron $CG_{t,q,p}$. It follows that there exists an optimal solution $(t^*, q^*, s^*, p^*, o^*)$ to (LP-CG) where (t^*, q^*, p^*) is an extreme point of polyhedron $CG_{t,q,p}$.

Second, we show that all extreme points of $CG_{t,q,p}$ are integral. To this end, we show that the constraint matrix describing $CG_{t,q,p}$ is TU. For presentation convenience, we rewrite the constraints defining $CG_{t,q,p}$ as follows:

$$q_i + q_{i+1} + \sum_{k=1}^{i} t_k - \sum_{k=1}^{i} (p_{ki} + p_{k(i+1)}) \ \ge 1 \quad \forall 1 \le i \le n-1 \tag{EC.23a}$$

$$q_n - \sum_{k=1}^{n} p_{kn} \ \ge 0, \tag{EC.23b}$$

$$-q_i - \sum_{k=1}^{i} t_k + \sum_{k=1}^{i} p_{ki} \ \geq \ -1 \quad \forall 1 \leq i \leq n, \tag{EC.23c}$$

$$-t_k + p_{ki} + p_{k(i+1)} \ \geq 0 \quad \forall 1 \leq k \leq i \leq n-1, \tag{EC.23d}$$

$$t_k - p_{ki} \ \geq 0 \quad \forall 1 \leq k \leq i \leq n, \tag{EC.23e}$$

and we denote the constraint matrix as

$$\mathcal{CG}^0_{t,q,p} \ := \begin{bmatrix} (q_i + q_{i+1} + \sum_{k=1}^{i} t_k - \sum_{k=1}^{i}(p_{ki} + p_{k(i+1)})), \quad \forall 1 \leq i \leq n-1 \\ (q_n - \sum_{k=1}^{n} p_{kn}) \\ (-q_i - \sum_{k=1}^{i} t_k + \sum_{k=1}^{i} p_{ki}), \quad \forall 1 \leq i \leq n \\ (-t_k + p_{ki} + p_{k(i+1)}), \quad \forall 1 \leq k \leq i \leq n-1 \\ (t_k - p_{ki}), \quad \forall 1 \leq k \leq i \leq n \end{bmatrix},$$

where the five rows of sub-matrices are associated with constraints (EC.23a)–(EC.23e), respectively. To show that matrix $\mathcal{CG}^0_{t,q,p}$ is TU, we conduct pivot operations on the matrix with variables p_{ki}, t_k, and q_i. Note that a matrix is TU if and only if it remains TU after pivot operations (cf. Nemhauser and Wolsey 1999). We conduct the following pivot operations in order.

(i) For all $1 \leq k \leq i \leq n$, pivot with variable p_{ki} based on the component -1 in sub-matrix $(t_k - p_{ki})$ (corresponding to constraints (EC.23e)). This pivot operation is equivalent to (a) adding $t_k - p_{ki}$, for all $1 \leq k \leq i \leq n$, to the left-hand side of every constraint (EC.23c)–(EC.23d) in which variable p_{ki} has coefficient 1, (b) adding $p_{ki} - t_k$, for all $1 \leq k \leq i \leq n$, to the left-hand side of every constraint (EC.23a)–(EC.23b) in which variable p_{ki} has coefficient -1 and (c) multiplying each left-hand side of constraint (EC.23e) by -1. As a result, the matrix after pivoting becomes

$$\mathcal{CG}^1_{t,q,p} \ := \begin{bmatrix} (q_i + q_{i+1} - \sum_{k=1}^{i} t_k), \quad \forall 1 \leq i \leq n-1 \\ (q_n - \sum_{k=1}^{n} t_k) \\ (-q_i), \quad \forall 1 \leq i \leq n \\ (t_k), \quad \forall 1 \leq k \leq i \leq n-1 \\ (-t_k + p_{ki}), \quad \forall 1 \leq k \leq i \leq n \end{bmatrix},$$

Note that sub-matrix $(-t_k + p_{ki} + p_{k(i+1)})$ becomes (t_k) because the left-hand side of each constraint (EC.23d), $-t_k + p_{ki} + p_{k(i+1)}$, is summed with $t_k - p_{ki}$ and $t_k - p_{k(i+1)}$ and so the coefficient of each t_k changes from -1 to 1 after pivoting. Meanwhile, sub-matrix $(-q_i - \sum_{k=1}^{i} t_k + \sum_{k=1}^{i} p_{ki})$ becomes $(-q_i)$ after pivoting because, for each $1 \leq i \leq n$, $-q_i - \sum_{k=1}^{i} t_k + \sum_{k=1}^{i} p_{ki}$ is summed with $t_k - p_{ki}$ for all $1 \leq k \leq i$.

(ii) For all $1 \leq k \leq n-1$, pivot with variable t_k based on any component 1 in sub-matrix (t_k) (note that there are multiple components 1 associated with each variable t_k in (t_k) and we can pick any one of them). Since all components in each row of sub-matrix (t_k) are zeros except one equaling 1, these pivot operations (a) make all coefficients of all variables t_k zeros in matrix $\mathcal{CG}^1_{t,q,p}$ as long as $1 \le k \le n-1$, and (b) keep all coefficients of all variables q_i and p_{ki} unchanged. As a result, the matrix after pivoting becomes

$$\mathcal{CG}^2_{t,q,p} := \begin{bmatrix} (q_i + q_{i+1}), \quad \forall 1 \le i \le n-1 \\ (q_n - t_n) \\ (-q_i), \quad \forall 1 \le i \le n \\ (t_k), \quad \forall 1 \le k \le i \le n-1 \\ \begin{cases} (p_{ki}), \quad \forall 1 \le k \le i \le n-1 \\ (-t_n + p_{nn}), \end{cases} \end{bmatrix}.$$

(iii) For all $1 \le i \le n$, pivot with variable q_i based on any component -1 in sub-matrix $(-q_i)$ in $\mathcal{CG}^2_{t,q,p}$. Since $(-q_i)$ is an identity matrix, these pivot operations eliminate all coefficients of variables q_i in all other sub-matrices. As a result, the matrix after pivoting becomes

$$\mathcal{CG}^2_{t,q,p} := \begin{bmatrix} (0), \quad \forall 1 \le i \le n-1 \\ (-t_n) \\ (q_i), \quad \forall 1 \le i \le n \\ (t_k), \quad \forall 1 \le k \le i \le n-1 \\ \begin{cases} (p_{ki}), \quad \forall 1 \le k \le i \le n-1 \\ (-t_n + p_{nn}), \end{cases} \end{bmatrix}.$$

It follows that matrix $\mathcal{CG}^3_{t,q,p}$ contains only $\{-1, 0, 1\}$ entries, has no more than two nonzero entries in each row, and the sum of the entries is zero for each row containing two nonzero entries. Hence, matrix $\mathcal{CG}^3_{t,q,p}$ is TU, and so is matrix $\mathcal{CG}^0_{t,q,p}$.

Therefore, the extreme points of polyhedron $CG_{t,q,p}$ are integral and so property 1 is proved.

Third, to show property 2, we consider any extreme point (t, q, s, p, o) of polyhedron CG. Because $\sum_{k=1}^n t_k \le 1$ and each $t_k \in \{0, 1\}$ by property 1, we show $p_{ki} = t_k q_i$ by discussing the following two cases on values of t_k.

(i) If $t_k = 0$ for all $1 \le k \le n$, then $p_{ki} = 0$ for all $1 \le k \le i \le n$ because $p_{ki} \le t_k$. It follows that $p_{ki} = t_k q_i = 0$.

(ii) If there exists $1 \le k^* \le n$ such that $t_{k^*} = 1$, then any other $t_k = 0$. It follows that $p_{ki} = 0$, and so $p_{ki} = t_k q_i = 0$ for all $1 \le k \le i \le n$ and $k \ne k^*$. For all $k^* \le i \le n$, constraints (EC.13c) yield

$$-q_i - \sum_{k=1}^{i} t_k + \sum_{k=1}^{i} p_{ki} \;=\; -q_i - 1 + p_{k^*i} \;\ge\; -1 \quad \Rightarrow \quad p_{k^*i} \ge q_i.$$

Also, for all $k^* \le i \le n-1$, constraints (EC.13d) yield

$$q_i + q_{i+1} + \sum_{k=1}^{i} t_k - \sum_{k=1}^{i} (p_{ki} + p_{k(i+1)}) \;=\; q_i + q_{i+1} + 1 - p_{k^*i} - p_{k^*(i+1)} \;\ge\; 1$$

$$\Rightarrow \; p_{k^*i} + p_{k^*(i+1)} \le q_i + q_{i+1}.$$

It follows that $p_{k^*i} + p_{k^*(i+1)} = q_i + q_{i+1}$ for all $k^* \le i \le n-1$. Furthermore, constraint (EC.13a) implies $q_n - \sum_{k=1}^n p_{kn} = q_n - p_{k^*n} \ge 0$, and so $q_n = p_{k^*n}$. Therefore, $q_i = p_{k^*i}$, or equivalently $q_i = t_{k^*} p_{k^*i}$ since $t_{k^*} = 1$, for all $k^* \le i \le n$.

Since (t, q, s, p, o) is an extreme point of polyhedron CG, each o_{ki} satisfies either inequality (EC.21a) or (EC.21b) at equality, and each s_i satisfies either inequality (EC.21c) or (EC.21d) at equality. We discuss the following cases to show $o_{ki} = s_i p_{ki} = t_k q_i s_i$.

(i) If $t_k = 0$ for all $1 \le k \le n$, then $p_{ki} = 0$ for all $1 \le k \le i \le n$ because $p_{ki} \le t_k$. It follows that $o_{ki} = 0$ by constraints (EC.12d). Therefore, $o_{ki} = s_i p_{ki} = 0$.

(ii) If there exists $1 \le k^* \le n$ such that $t_{k^*} = 1$, then any other $t_k = 0$. It follows that $p_{ki} = 0$, and so $o_{ki} = s_i p_{ki} = 0$ for all $1 \le k \le i \le n$ and $k \ne k^*$. Then, for all $k^* \le i \le n$, inequalities (EC.21c)–(EC.21d) yield

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^{\mathrm{L}} p_{ki}) \; = \; s_i - o_{k^*i} + s_i^{\mathrm{L}} p_{k^*i} \; \ge \; s_i^{\mathrm{L}}, \tag{EC.24a}$$

$$s_i - \sum_{k=1}^{i} (o_{ki} - s_i^{\mathrm{U}} p_{ki}) \; = \; s_i - o_{k^*i} + s_i^{\mathrm{U}} p_{k^*i} \; \le \; s_i^{\mathrm{U}}. \tag{EC.24b}$$

Hence, each s_i satisfies either inequality (EC.24a) or (EC.24b) at equality. We discuss the following two sub-cases to finish the proof.

Sub-case 1. If $p_{k^*i} = 0$, then $o_{k^*i} = 0$ by constraints (EC.12d). Therefore, $o_{k^*i} = s_i p_{k^*i} = 0$.

Sub-case 2. If $p_{k^*i} = 1$, then inequalities (EC.24a)–(EC.24b) imply $s_i = o_{k^*i}$. Therefore, $o_{k^*i} = s_i p_{k^*i}$.

□

Step 3: Finally, we prove Theorem EC.1 based on the previous two propositions.

Proof of Theorem EC.1 ($CG \supseteq \mathrm{conv}(G)$) By Proposition EC.6, since polyhedron CG consists of either trivial equalities/inequalities or valid inequalities of set G, we have $(t, q, s, p, o) \in CG$ if $(t, q, s, p, o) \in G$. It follows that $CG \supseteq \mathrm{conv}(G)$.

($CG \subseteq \mathrm{conv}(G)$) By Proposition EC.7, since each extreme point (t, q, s, p, o) of CG satisfies properties 1, 2, and 3, $(t, q, s, p, o) \in G$. By the Minkowski's Theorem on polyhedron, we have $x \in \mathrm{conv}(G)$ if $x \in CG$. It follows that $CG \subseteq \mathrm{conv}(G)$. This completes the proof. □

EC.5. Optimal Schedule Demonstration

We demonstrate in Figure EC.1 the optimal schedules of instances with $n = 10$ appointments, produced by E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP for $R = 0, 1$ and $1 - \nu_i = 0.2, 0.4$, $\forall i = 1, \dots, n$. The points

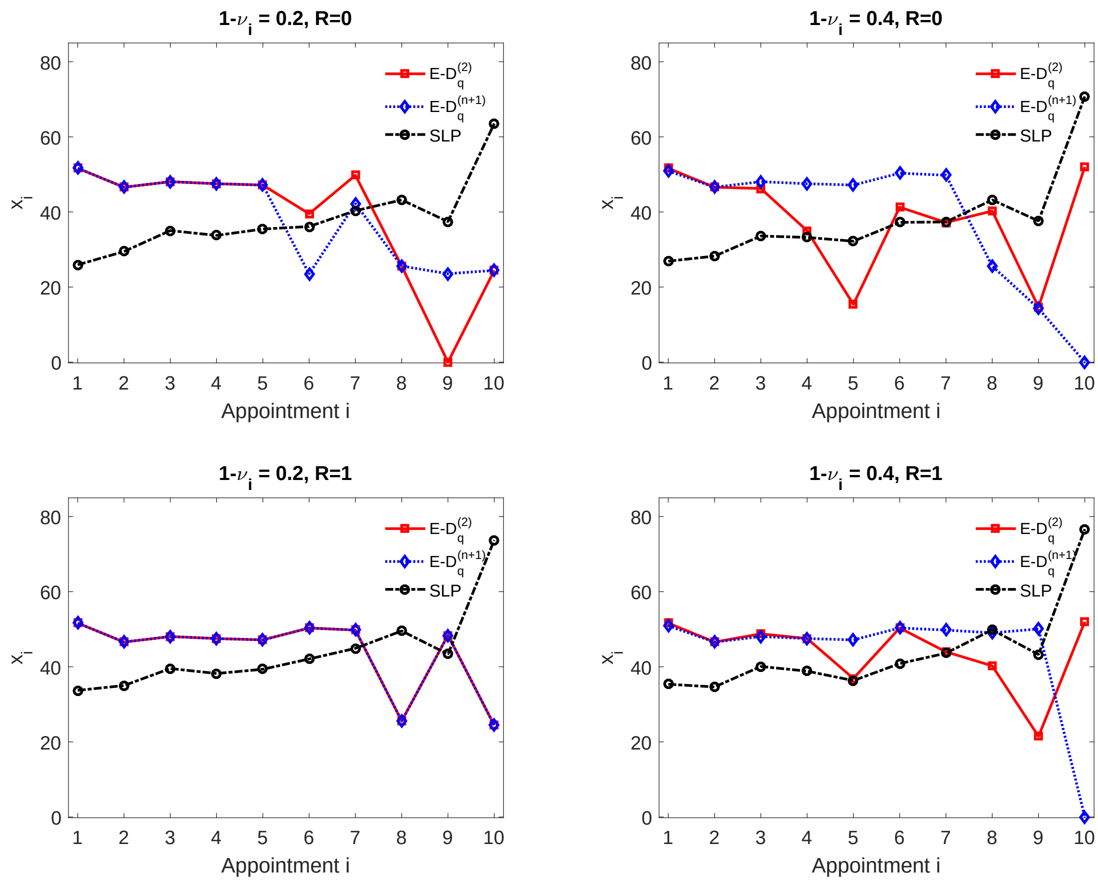


Figure EC.1 Appointment schedules produced by E-$D_q^{(2)}$, E-$D_q^{(n+1)}$, and SLP for different settings of parameter R (time limit) and $1-\nu_i$ (no-show probability)

(i, x_i) of every model in each subfigure correspond to the time interval (in minute) assigned in between the arrivals of appointments i and $i+1$, for all $i = 1, \ldots, 9$.

As shown in Figure EC.1, SLP almost equally distributes the time in between each arrival and schedules a long interval for the last appointment, for all combinations of $1-\nu_i$ and R values. As compared to SLP, both E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ intend to schedule longer inter-arrival time for the early appointments but significantly shorter time towards the end. The purpose is to mitigate the waiting time that may accumulate due to long service durations (also reflected by the shorter waiting time for the DR models in Tables 4 and 5). When the no-show probability is relative small (i.e., $1-\nu_i = 0.2,\ \forall i = 1, \ldots, n$) and the time limit T is sufficiently long (i.e., $R = 1$), both E-$D_q^{(2)}$ and E-$D_q^{(n+1)}$ yield the same optimal schedule (also reflected by the same optimal objective value of the two models in Table 3).

References

Nemhauser GL, Wolsey LA (1999) *Integer and Combinatorial Optimization* (John Wiley & Sons).

Mak HY, Rong Y, Zhang J (2015) Appointment scheduling with limited distributional information. *Management Science* 61(2):316–334.

Rockafellar RT, Uryasev S (2000) Optimization of conditional Value-at-Risk. *Journal of Risk* 2(3):21–42.

Rockafellar RT, Uryasev S (2002) Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance* 26(7):1443–1471.

Sion M (1958) On general minimax theorems. *Pacific J. Math* 8(1):171–176.

Wolsey LA (1998) *Integer Programming*, volume 42 (Wiley New York).

Zangwill WI (1966) A deterministic multi-period production scheduling model with backlogging. *Management Science* 13(1):105–119.

Zangwill WI (1969) A backlogging model and a multi-echelon model of a dynamic economic lot size production system – a network approach. *Management Science* 15(9):506–527.