ISE

Industrial and
Systems Engineering

Distributionally Robust Home Service Routing and Appointment Scheduling with Random Travel and Service Times

Man Yiu Tsang[1] and Karmel S. Shehadeh[1]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA

ISE Technical Report 21T-011

LEHIGH UNIVERSITY

Distributionally Robust Home Service Routing and Appointment Scheduling with Random Travel and Service Times

Man Yiu Tsang[a,*], Karmel S. Shehadeh[a]

[a]*Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA*

Abstract

In this paper, we study an integrated routing and appointment scheduling problem arising from home service practice. Specifically, given a set of customers within a service region that an operator needs to serve, we seek to find the operator's route and time schedule. The travel time between customers and the service time of each customer are random. The probability distributions of these random parameters are unknown and only a possibly small set of historical realizations is available. To address distributional ambiguity, we propose and analyze two distributionally robust home service routing and scheduling (DHRAS) models that search for optimal routing and scheduling decisions to minimize the worst-case expectation of operational costs over all distributions residing within an ambiguity set. In the first model, we use a moment-based ambiguity set. In the second model, we use an ambiguity set that incorporates all distributions within a 1-Wasserstein distance from a reference (e.g., empirical) distribution. We derive equivalent mixed-integer linear programming reformulations of both models. In an extensive numerical experiment, we investigate the proposed models' computational and operational performance, and derive insights into DHRAS.

Keywords: Home service, scheduling and routing, uncertainty, distributionally robust optimization, mixed integer programming

1. Introduction

In this paper, we study an integrated routing and appointment scheduling problem arising from home service practice. Specifically, given a set of customers within a service region that an operator needs to serve, the home service routing and appointment scheduling (HRAS) problem seeks to determine the operator's route and customers' appointment times. Travel time between customers and their service times are random. The distributions of travel and service times are unknown, and only a possibly small data set on these random parameters is available. If the operator arrives at the customer's location before the scheduled service start time, then the operator needs to wait until the scheduled start time (i.e., remains idle). On the other hand, the customer needs to wait

*Corresponding author.
Email addresses: mat420@lehigh.edu (Man Yiu Tsang), kas720@lehigh.edu (Karmel S. Shehadeh)

if the operator arrives after the scheduled start time. The operator has fixed service hours beyond which one experiences overtime. The quality of routing and scheduling decisions is a function of the total operational cost, consisting of customers' waiting time, and the operator's travel time, idle time, and overtime.

The home service industry has been rapidly growing worldwide due to emergent changes in family structures, work obligation and extended work hours, aging population, and outspread of chronic diseases, among others. As pointed out by Zhan et al. (2021) and Lunden (2019), the United States alone annually spends about \$82 billion on various forms of home healthcare services and \$300 billion on home repairs and maintenance, and these numbers are rapidly growing. In 2018, the global home service market was valued at around \$282 billion and is expected to grow by \$18.9% annually from 2019-2026, reaching \$1133.4 billion by 2026 (Verified Market Research, 2019). Therefore, the development of efficient home service scheduling and planning tools is essential to support decision-making in all areas of the home service industry.

Home service planning process often includes the following steps (Eveborn et al., 2006; Lanzarone and Matta, 2014; Zhan et al., 2021). First, customers request their home services on their preferred dates through a preliminary booking system. Second, given each date's requested services, a home service provider assigns the available providers to customers. Third, the home service provider sets the appointment time for each customer and determine each operator's route. Fourth, the service provider communicates the appointment times to the customers. On the day of service, the operator executes the schedule and visits customers one by one (and exactly once). In this paper, we consider the perspective of a home service provider who needs to make the following critical planning decisions for each operator and assigned set of customers: (1) how to route the operator, and (2) how to assign appointment times to customers. Henceforth, for simplicity, we use the term "*routing*" to refer to routes that specify the operator movement between customers or the sequence of customer visits on the operator's schedule, and "*appointment time*" to refer to the scheduled start time of the service of a customer. The set of appointment times of customers represent a schedule.

HRAS is a challenging optimization problem for several reasons. First, it is a complex combinatorial optimization problem that requires deciding the operator's route and assigning appointment times to each customer in the route simultaneously. Second, suppose we fix the appointment times. In this case, HRAS becomes similar to the traveling salesman problem (TSP) and the single-vehicle routing problem with time constraints, which are known to be challenging optimization problems (see Cook, 2011). Third, for a fixed route, HRAS reduces to the well-known challenging single-server stochastic appointment scheduling (AS) problem with random service time, which seeks a sequence of appointment times that minimizes customers' waiting times and the provider's idling time and overtime (Ahmadi-Javid et al., 2017; Robinson and Chen, 2003). The rationale for considering travel time or travel distance as an additional optimization criterion in HRAS is to minimize service delivery or operational cost by controlling operator's amount of compensation for the distance or time traveled (Grieco et al., 2020; Kandakoglu et al., 2020).

Assuming deterministic travel times, Zhan et al. (2021) recently proposed the first two-stage stochastic programming approach (SP) for HRAS, which seeks first-stage routing and scheduling decisions that minimize the operator's traveling cost and the expected second-stage cost of the operator's idling cost and customers' waiting cost. Zhan et al. (2021) take the expectation with respect to (assumed) known probability distributions of random service duration. Although SP is a powerful and theoretically mature approach to model uncertainty, its applicability is limited to the case in which we know the distribution of uncertain parameters or we have a large data sample. In most real-world applications, it is challenging to estimate the actual distribution of random parameters accurately, especially with limited data during the planning process. If we optimize according to a biased data sample from a misspecified distribution, the resulting (*optimistically biased*) optimal routing and scheduling decisions may have a disappointing out-of-sample performance, i.e., performing poorly under the true distribution (Delage and Ye, 2010; Mohajerin Esfahani and Kuhn, 2018; Smith and Winkler, 2006).

Distributionally robust optimization (DRO) is another approach to model uncertainty when the distribution of random parameters is unknown. In DRO, we assume that the distribution of random parameters resides in a so-called ambiguity set (i.e., a family of distributions that share some partial information about the random parameter). We then optimize based on the worst-case distribution residing within the ambiguity set, i.e., distribution is a decision variable (Delage and Ye, 2010; Mohajerin Esfahani and Kuhn, 2018; Rahimi and Gandomi, 2020). As such, DRO alleviates the unrealistic assumption that the decision-makers have complete knowledge of the distributions, and we have seen many successful applications of DRO to real-world problems (Klabjan et al., 2013; Mak et al., 2015; Pflug and Pohl, 2018; Shang and You, 2018; Wang et al., 2019; Zhang et al., 2017).

The ambiguity set is a key ingredient in DRO. Most of the early work on DRO is based on moment ambiguity sets, which incorporate all distributions sharing certain moments (e.g., mean and variance). Many tools have been developed to derive tractable and solvable reformulations of moment-based DRO models. However, moment-based DRO models often do not have asymptotic properties (i.e., optimal solution and values do not converge as more data becomes available) because the ambiguity sets only incorporates descriptive statistics (i.e., moment information). New modern DRO studies have shifted toward data-driven, distance-based DRO approaches that construct ambiguity sets in the vicinity of a reference (e.g., empirical) distribution (Rahimi and Gandomi, 2020). The main advantage of distance-based ambiguity sets is that they allow the incorporation of possibly small observed data on random parameters directly in the optimization problem, as well as asymptotic guarantees (e.g., Mohajerin Esfahani and Kuhn, 2018). This inspires us to investigate and compare the value and performance of moment-based DRO and distance-based DRO approaches to address distributional ambiguity of service and travel times in this specific HRAS problem.

1.1. Contributions

In this paper, we address the distributional ambiguity of random travel and service times in HRAS. Specifically, we define two *distributionally robust home service routing and scheduling* (DHRAS) models that seek optimal routing and scheduling decisions to minimize the worst-case (maximum) sum of the operater's travel time, idle time, and overtime costs, and customers' waiting time costs. In the first model (M-DHRAS), we take the worst-case expectation with respect to the distributions from an ambiguity set based on the mean values and the range of travel and service times. In the second model (W-DHRAS), we take the worst-case expectation over a 1–Wasserstein ambiguity set (i.e., a set containing all distributions that are sufficiently close to a reference distribution with respect to the 1–Wasserstein metric). We summarize our main contributions as follows.

- Using duality theory and HRAS structural properties, we derive equivalent mixed-integer linear programming (MILP) reformulations of the min-max M-DHRAS and W-DHRAS models. The reformulations can be implemented and solved by off-the-shelf optimization software.
- To investigate the value of the DRO approach for HRAS, we conduct an extensive numerical experiment comparing the SP, M-DHRAS and W-DHRAS models' operational and computational performance. Our results demonstrate that: (1) W-DHRAS yields robust decisions with both small and large data size, and enjoys both asymptotic consistency and finite-data guarantees; (2) M-DHRAS produces the most conservative schedules and it performs well when the distribution of travel time changes dramatically (e.g., actual travel times are longer); (3) W-DHRAS solutions have better operational performance than the SP solutions both under perfect and misspecified distributions, even when only a small data set is available; (4) DRO models produce more reliable solutions than the SP model.
- We show that under the same assumption made by Zhan et al. (2021), our SP formulation for HRAS is more computationally efficient than their pioneering SP formulation.
- To the best of our knowledge, according to the recent review of operational research in home health care applications Grieco et al. (2020) and our literature review in Section 2, our paper is the first to attempt to investigate and compare theoretically and numerically a data-driven DRO with Wasserstein ambiguity and a moment-based DRO model for this specific HRAS problem.

1.2. Structure of the paper

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we formally define the DHRAS models, present and analyze the M-

DHRAS and W-DHRAS models. In Section 4, we present our computational results and managerial implications. Finally, we draw conclusions and discuss future directions in Section 5.

2. Relevant Literature

In this section, we focus primarily on the literature mostly relevant to our problem, i.e., papers that apply stochastic optimization to address HRAS problems similar to ours. For comprehensive recent surveys of operations research methods applied to decisions in home health care, we refer readers to Fikar and Hirsch (2017); Grieco et al. (2020); Gutiérrez and Vidal (2013). If we fix the appointment time, our problem reduces to the single-vehicle routing problem. That is, we need to design the optimal route of a single vehicle to visit all the customers by minimizing the costs of travel. Recent survey of the literature on these problems includes Costa et al. (2019); Oyola et al. (2018). On the other hand, if we fix the route, our problem is similar to the single-server appointment scheduling problem. That is, we need to decide the optimal appointment time for the customers to minimize the total cost of customer waiting and server idling and overtime costs. Appointment times can be fixed time points (see, e.g., Braekers et al., 2016; Demirbilek et al., 2019; Milburn, 2012) or time intervals/windows (see, e.g., Lee et al., 2013; Mankowska et al., 2014; Yuan et al., 2015; and references therein). Herein, we focus on the former case under random service and travel times. For recent studies on appointment scheduling problems in health care applications, we refer to Creemers et al. (2021); Kuiper et al. (2021); Restrepo et al. (2020); Zhou et al. (2021) and references therein. We refer to Pinedo (2016) for a comprehensive survey of scheduling theory, applications and methods.

SP approaches for HRAS include Liu et al. (2019); Yuan et al. (2015); Zhan and Wan (2018). The recent work of Zhan et al. (2021) is closely related to ours. Specifically, Zhan et al. (2021) consider a single operator HRAS problem in which they assume a deterministic travel time between customers and that each customer's service time follows a fully known probability distribution. Accordingly, Zhan et al. (2021) propose a two-stage stochastic mixed-integer linear program, which seeks first-stage routing and scheduling decisions that minimize the operator traveling cost and an expectation of the second-stage cost of the operator idling and customer waiting. In contrast to Zhan et al. (2021), we consider both stochastic travel and service time and address their distributional ambiguity. Furthermore, we incorporate the operator overtime cost in the objective.

The SP approach assumes that the decision-maker is risk-neutral and knows the random parameters' distributions with certainty or can fully estimate them. In practice, data on random parameters is often unavailable or very small, and we may only have a few distributional information on the random service and travel times. Given distributional ambiguity, if we calibrate a model to a misspecified (i.e., biased) distribution, the resulting optimal SP routing and scheduling decisions may have a disappointing out-of-sample performance, i.e., performing poorly under the true distribution.

In practice, it is implausible that decision-makers have sufficient high-quality data to infer the true distribution of random parameters, especially in health care applications. Various robust approaches have been proposed to model the risk-averse nature of decision-makers and uncertain parameters based on partial information of their distributions. The classical robust optimization (RO) approach assumes that uncertain parameters resided on an uncertainty set of possible outcomes with some structure (e.g., polyhedron, see, Ben-Tal et al., 2015; Bertsimas and Sim, 2004; Soyster, 1973). Accordingly, optimization in RO models is based on the worst-case scenario within the uncertainty set.

Recently, Shi et al. (2019) proposed an RO approach for home health care service. They consider the problem with multiple caregivers and the goal is to find the optimal route for each caregiver and appointment time for each patient. They propose a robust model on both unknown service and travel time and discuss the heuristic solution approaches, with the objective function being the travel cost and delay cost with respect to the schedules. They found that, as compared with the SP model in Shi et al. (2018), solutions from RO model has a better out-of-sample performance. For instance, the probability of visiting the customers timely is higher. In this paper, we incorporate the provider's idling costs in the objective function, together with the use of DRO models. Instead of providing heuristic solution approaches, we reformulate our DRO models into MILP, which could be solved by off-the-shelf optimization software.

By focusing the optimization on the worst-case scenario, classical RO approaches often yield over-conservative solutions and suboptimal decisions for other more-likely scenarios (Chen et al., 2020; Delage and Saif, 2018; Thiele, 2010). Distributionally robust optimization (DRO) is another approach for modeling uncertainty that bridges SP and RO by overcoming their shortcoming and has recently gained significant attention. Unlike the risk neutral SP approach and conservative RO approach, DRO adopts a middle-ground approach. In DRO, we assume that the distribution of random parameters resides in an ambiguity set. We then optimize based on the worst-case distribution within the ambiguity set (Rahimian and Mehrotra, 2019).

There are two common types of ambiguity sets: moment-based and distance-based ambiguity sets (see, e.g., Rahimian and Mehrotra, 2019). In this paper, we address the distributional ambiguity of service time and travel time in the single-operator HRAS using these two types of ambiguity set, namely the M-DHRAS and W-DHRAS models. We reformulate them into MILPs, which could be solved by off-the-shelf optimization software. To the best of our knowledge and according to the recent surveys Fikar and Hirsch (2017); Grieco et al. (2020), there is no DRO approach for the specific HRAS problem that we address. Moment-based DRO models for single-server appointment scheduling includes (Jiang et al., 2017; Kong et al., 2013; Mak et al., 2015). Recently, Jiang et al. (2019) proposed the distance-based DRO model based on the Wasserstein metric for single-server appointment scheduling for a fixed sequence of customers with random service time and no-show.

Table 1: Notation.

Indices	
i	index of customer, $i = 1, \dots, N$
j	index of service position, $j = 1, \dots, N$
Parameters and sets	
N	number of customers
L	standard work time
c_j^{u}	idling cost associated to the early arrival at the jth customer's location
c_j^{w}	waiting cost of the jth customer
c^{o}	overtime cost
λ	travel time cost
d_i	service time of customer i
$t_{i,i'}$	travel time from customers i to i' (with 0 denoting the depot)
$\underline{d}_i/\overline{d}_i$	lower/upper bound of the service time of customer i
$\underline{t}_{i,i'}/\overline{t}_{i,i'}$	lower/upper bound of the travel time from customers i to i'
First-stage decision variables	
$x_{i,j}$	binary variable equal to 1 if customer i is the jth customer served and 0 otherwise
a_j	appointment time of the jth customer
Second-stage decision variables	
u_j	idling time due to an early arrival at the jth customer's location
w_j	waiting time of the jth customer (w_{N+1} as the overtime)

They derive tractable reformulations of their model under the 1–Wasserstein and 2–Wasserstein ball. In HRAS, we additionally optimize the provider's routing (or sequencing) decision and address the co-existing uncertainty of service and travel time. Setting the travel time to zero, our model generalizes Jiang et al. (2019)'s model by considering appointment sequencing decisions.

3. Formulation and Analysis

In this section, we formulate the DHRAS problem using moment (M-DHRAS) and Wasserstein (W-DHRAS) ambiguity sets. We define DHRAS and provide a two-stage stochastic programming formulation in Section 3.1. We analyze and derive equivalent mixed-integer linear programming reformulations of M-DHRAS in Section 3.2 and W-DHRAS in Section 3.3.

Notation: For $a, b \in \mathbb{Z}$, we define $[a] := \{1, 2, \dots, a\}$ and $[a, b]_{\mathbb{Z}} := \{c \in \mathbb{Z} : a \leq c \leq b\}$, i.e., $[a, b]_{\mathbb{Z}}$ represent the set of running integer indices $\{a, a+1, a+2, \dots, b\}$. For a real number a, we define $(a)^+ = \max\{a, 0\}$. We use boldface notation to denote vectors, e.g., $\boldsymbol{d} := [d_1, d_2, \dots, d_N]^\top$. Table 1 summarizes other notation.

3.1. Definitions and assumptions

As in Zhan et al. (2021), we consider a set of N customers that need to be served within a given day by a single home service provider (operator). Traveling time, $t_{i,i'}$, between each pair of customers i and i' ($i, i' = 0, \ldots, N$) is random, with $i = 0$ representing the service provider's office (depot). Service time of each customer $i \in [N]$ is also random. We assume that the travel time and service time's probability distributions are unknown (ambiguous), and only a possibly small set of historical realizations of these random parameters is available. We seek two sets of decisions: (1) service provider's visiting sequence (routes), (2) customers appointment times. The objective is to minimize the sum of customers' waiting time, and the provider's traveling time, idling time, and overtime.

For all $i \in [N]$ and $j \in [N]$, we let the binary decision variable $x_{i,j}$ equal 1 if customer i is the jth customer in the operator's route/schedule, and zero otherwise. For all $j \in [N]$, we let the continuous variable a_j represent the appointment starting time of the jth customer. The feasible region of variable $\boldsymbol{x}$ is defined in (1) such that each customer is assigned to one position in the operator's route/schedule, and each position is assigned to one customer. The feasible region of variable $\boldsymbol{a}$ is defined in (2) such that all customers' appointments are scheduled within the provider's service hours $[0, L]$.

$$\mathcal{X} = \left\{ \boldsymbol{x} : \begin{array}{l} \sum_{i=1}^{N} x_{i,j} = 1, \ \forall j \in [N] \\ \sum_{j=1}^{N} x_{i,j} = 1, \ \forall i \in [N] \\ x_{i,j} \in \{0,1\}, \ \forall i \in [N], \ \forall j \in [N] \end{array} \right\} \tag{1}$$

$$\mathcal{A} = \left\{ \boldsymbol{a} : \begin{array}{l} 0 \leq a_j \leq L, \quad \forall j \in [N] \\ a_j \geq a_{j-1}, \ \forall j \in [2, N]_{\mathbb{Z}} \end{array} \right\} \tag{2}$$

Due to random travel and service times, one or multiple of the following scenarios may happen: (1) operator arrives at the customer's location before the scheduled service start time (a_j), and thus s/he remains idle until the scheduled start time; (2) the operator arrives after the scheduled start time of the customer, and thus the customer incurs waiting cost; (3) provider works overtime beyond his/her scheduled L to finish all appointments. Let the continuous decision variable w_j represent the waiting time of the jth customer, for all $j \in [N]$ and w_{N+1} represent the operator's overtime. For all $j \in [N]$, let the continuous decision variable u_j represent the provider's idle time before the start time of the jth appointment. For all $i \in [N]$, let the random parameter d_i represent the service duration of customer i. For all $i \in [N]$ and $i' \in [N]$, let the random parameter $t_{i,i'}$ represent the travel time between i and i'. Given a fixed $\boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{a} \in \mathcal{A}$ and a joint realization $\boldsymbol{\xi} := [\boldsymbol{t}, \boldsymbol{d}]^\top$, we can compute the operational costs (as a function of travel time, idle time, and overtime, and customers waiting time), using the following linear program:

$$f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) := \min_{\boldsymbol{u}, \boldsymbol{w}} \left\{ \sum_{j=1}^{N} (c_j^{\mathrm{w}} w_j + c_j^{\mathrm{u}} u_j) + c^{\mathrm{o}} w_{N+1} + \lambda A \right\} \tag{3a}$$

$$\text{s.t.} \quad w_1 - u_1 = \sum_{i=1}^{N} t_{0,i} x_{i,1} - a_1, \tag{3b}$$

$$w_j - w_{j-1} - u_j = a_{j-1} - a_j + \sum_{i=1}^{N} d_i x_{i,j-1} + \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i,j}, \ \forall j \in [2, N]_{\mathbb{Z}}, \tag{3c}$$

$$w_{N+1} - w_N - u_{N+1} = a_N - a_{N+1} + \sum_{i=1}^{N} d_i x_{i,N}, \tag{3d}$$

$$A = \sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i',j} + \sum_{i=1}^{N} (t_{0,i} x_{i,1} + t_{i,0} x_{i,N}), \tag{3e}$$

$$(w_j, u_j) \geq 0, \quad \forall j \in [N+1], \tag{3f}$$

where c_j^{w}, c_j^{u}, c^{o}, and λ are respectively the non-negative unit penalty costs of waiting, idling, overtime, and travel time for all $j \in [N]$. Also, we let $a_{N+1} = L$ and $x_{i,N+1} = 0$ for all $i \in [N]$. The objective function (3a) minimizes a linear cost function of waiting, idling, overtime, and travel time. Constraint (3b) yields either provider's idle time before the scheduled time of the first appointment or the waiting time of the first customer. Constraint (3c) yields either the waiting time of the jth customer or the provider's idle time if s/he arrives at the jth customer, respectively, after or before the scheduled start time a_j of the jth customer. Constraints (3d) yields either the overtime or the schedule earliness. Constraint (3e) computes the operator's total travel time. Finally, constraints (3f) specify feasible ranges of the decision variables.

Classical two-stage SP models seek to find $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$ that minimizes the expectation of the cost $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ subject to random $\boldsymbol{\xi}$ with a known joint probability distribution $\mathbb{P}_{\boldsymbol{\xi}}$ as in (4).

$$Z^* := \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{a} \in \mathcal{A}} \mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})] \tag{4}$$

3.2. *DHRAS over moment ambiguity (M-DHRAS)*

In this section, we assume that $\mathbb{P}_{\boldsymbol{\xi}}$ is not fully known and belongs to a mean-support ambiguity set. Specifically, we assume that we know the mean value $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}(\boldsymbol{\xi}) := \boldsymbol{\mu} = [\boldsymbol{\mu}^{\text{d}}, \boldsymbol{\mu}^{\text{t}}]^\top$, lower bounds $[\underline{\boldsymbol{d}}, \underline{\boldsymbol{t}}]^\top$, and upper bounds $[\bar{\boldsymbol{d}}, \bar{\boldsymbol{t}}]^\top$ of $[\boldsymbol{d}, \boldsymbol{t}]^\top$. Mathematically, we consider the support $\mathcal{S} = \mathcal{S}^{\text{d}} \times \mathcal{S}^{\text{t}}$ of $\boldsymbol{\xi}$, where $\mathcal{S}^{\text{d}}$ and $\mathcal{S}^{\text{t}}$ are respectively the supports of random parameters $\boldsymbol{d}$ and $\boldsymbol{t}$, which are defined as follows:

$$\mathcal{S}^{\text{d}} := \left\{ \boldsymbol{d} \geq 0 : \ \underline{d}_i \leq d_i \leq \bar{d}_i, \ \forall i \in [N] \ \right\}, \tag{5a}$$

$$\mathcal{S}^{\text{t}} := \left\{ \boldsymbol{t} \geq 0 : \ \underline{t}_{i,i'} \leq t_{i,i'} \leq \bar{t}_{i,i'}, \ \forall i \in [N], \ i' \neq i \ \right\}. \tag{5b}$$

Then, we consider the following mean-support ambiguity set $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu})$:

$$\mathcal{F}(\mathcal{S}, \boldsymbol{\mu}) := \left\{ \mathbb{P} \in \mathcal{P}(\mathcal{S}) : \begin{array}{l} \int_{\mathcal{S}} d\mathbb{P} = 1 \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{\xi}) = \boldsymbol{\mu} \end{array} \right\}, \tag{6}$$

where $\mathcal{P}(\mathcal{S})$ represents the set of all probability distributions supported on $\mathcal{S}$. Using ambiguity set $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu})$, we formulate M-DHRAS as follows:

$$\text{(M-DHRAS)} \quad \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{a} \in \mathcal{A}} \left\{ \sup_{\mathbb{P} \in \mathcal{F}(\mathcal{S}, \boldsymbol{\mu})} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})] \right\}. \tag{7}$$

Formulation (7) seeks first stage decisions $(\boldsymbol{x}, \boldsymbol{a})$ that minimizes the worst-case (maximum) expectations of the second-stage operational cost over $\mathcal{F}(\mathcal{S}, \boldsymbol{\mu})$.

*3.2.1. **MILP reformulation of M-DHRAS***

In this section, we use the duality theory to reformulate the min-max M-DHRAS model (7) into an equivalent MILP, with reference to prior DRO studies. First, for a fixed $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$, we rewrite the inner maximization problem $\sup_{\mathbb{P} \in \mathcal{F}(\mathcal{S}, \boldsymbol{\mu})} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})]$ as follows:

$$\max \ \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})] \tag{8a}$$

$$\text{s.t.} \ \ \mathbb{E}_{\mathbb{P}}(\boldsymbol{\xi}) = \boldsymbol{\mu}, \tag{8b}$$

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\mathcal{S}}(\boldsymbol{\xi})] = 1, \tag{8c}$$

where $\mathbb{1}_{\mathcal{S}}(\boldsymbol{\xi}) = 1$ if $\boldsymbol{\xi} \in \mathcal{S}$ and $\mathbb{1}_{\mathcal{S}}(\boldsymbol{\xi}) = 0$ if $\boldsymbol{\xi} \notin \mathcal{S}$. As we show in the proof of Proposition 1 in Appendix A, problem (8) is equivalent to the deterministic problem (9).

Proposition 1. *For any* $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$*, problem* (8) *is equivalent to*

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\rho}} \left\{ \sum_{i=1}^{N} \mu_i^d \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} \mu_{i,i'}^t \alpha_{i,i'} + \max_{\boldsymbol{\xi} \in \mathcal{S}} \left\{ f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) + \sum_{i=1}^{N} -d_i \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} -t_{i,i'} \alpha_{i,i'} \right\} \right\}. \tag{9}$$

Note that $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ is a minimization problem, and thus in (9) we have an inner max-min problem, which is not suitable to solve using standard solution methods. For a given solution $(\boldsymbol{x}, \boldsymbol{a})$ and realized value of $\boldsymbol{\xi}$, $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ is a linear program. We formulate $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ defined in (3) in its equivalent dual form as

$$\begin{aligned} f(x, a, \boldsymbol{\xi}) := \max_{\boldsymbol{y} \in \mathcal{Y}} \Bigg\{ \left(\sum_{i=1}^{N} t_{0,i} x_{i,1} - a_1 \right) y_1 + \sum_{j=2}^{N+1} \left(a_{j-1} - a_j + \sum_{i=1}^{N} d_i x_{i,j-1} + \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i,j} \right) y_j \\ + \lambda \left[\sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i',j} + \sum_{i=1}^{N} \Big(t_{o,i} x_{0,i} + t_{i,0} x_{i,N} \Big) \right] \Bigg\} \end{aligned} \tag{10a}$$

$$\text{s.t.} \ \ \boldsymbol{y} \in \mathcal{Y} := \Big\{ 0 \leq y_{N+1} \leq c^{\text{o}}, \ -c_j^{\text{u}} \leq y_j \leq c_j^{\text{w}} + y_{j+1}, \ \forall j \in [N] \Big\}, \tag{10b}$$

where $y_1, \ldots, y_{N+1}$ are the dual variables associated with constraints (3b)–(3d). In view of the dual formulation (10), we re-write $\max_{\boldsymbol{\xi} \in \mathcal{S}} f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ as

$$\begin{aligned} \sup_{\boldsymbol{\xi} \in \mathcal{S}} \sup_{\boldsymbol{y} \in \mathcal{Y}} \Bigg\{ \left(\sum_{i=1}^{N} t_{0,i} x_{i,1} - a_1 \right) y_1 + \sum_{j=2}^{N+1} \left(a_{j-1} - a_j + \sum_{i=1}^{N} d_i x_{i,j-1} + \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i',j} \right) y_j \\ + \lambda \left[\sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i',j} + \sum_{i=1}^{N} \Big(t_{0,i} x_{i,1} + t_{i,0} x_{i,N} \Big) \right] \Bigg\}. \end{aligned} \tag{11}$$

Given $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$ and $\boldsymbol{\xi}$ we define function H as

$$H(\boldsymbol{y}) := -a_1 y_1 + \sum_{j=2}^{N+1} (a_{j-1} - a_j) y_j + \sum_{j=2}^{N+1} \sum_{i=1}^{N} d_i y_j x_{i,j-1} + \sum_{i=1}^{N} t_{0,i} y_1 x_{i,1} + \sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i \neq i'} t_{i,i'} y_j x_{i,j-1} x_{i',j}. \tag{12}$$

It is straightforward to verify that function H is linear (convex) in variable $\boldsymbol{y}$. Hence, the inner maximization problem $\max_{\boldsymbol{y} \in \mathcal{Y}} H(\boldsymbol{y})$ in (11) is a convex maximization problem. It follows from the fundamental convex analysis that it suffices to consider the extreme points of polytope $\mathcal{Y}$ to compute $\max_{\boldsymbol{y} \in \mathcal{Y}} H(\boldsymbol{y})$. This motivates us to apply a similar approach as in Shehadeh et al. (2020) and Jiang et al. (2017), which leverages the properties of the extreme points of $\mathcal{Y}$ in deriving an equivalent linear programming (LP) reformulation of problem (11). We formally prove this in Proposition 2 (see Appendix B for a detailed proof).

Proposition 2. *Let $\Delta d_i = \overline{d}_i - \underline{d}_i$ and $\Delta t_{i,i'} = \overline{t}_{i,i'} - \underline{t}_{i,i'}$. Then, for any $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$, the inner maximization problem in (9) is equivalent to*

$$\min_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}} \sum_{j=1}^{N+2} \beta_j + \sum_{i=1}^{N} \left[(\lambda - \alpha_{i,0}) \underline{t}_{i,0} + \Delta t_{i,0} (\lambda - \alpha_{i,0})^+ \right] x_{i,N} \tag{13a}$$

$$\text{s.t. } \beta_1 \geq -a_1 \pi_{1,1} + \sum_{i=1}^{N} \left[\underline{t}_{0,i} (\pi_{1,1} + \lambda - \alpha_{0,i}) + \Delta t_{0,i} \gamma_{0,i,1,1} \right] x_{i,1}, \tag{13b}$$

$$\begin{aligned} \sum_{j=1}^{v} \beta_j \geq{} & -a_1 \pi_{1,v} + \sum_{j=2}^{\min(v,N+1)} (a_{j-1} - a_j) \pi_{j,v} + \sum_{i=1}^{N} \left[\underline{t}_{0,i} (\pi_{1,v} + \lambda - \alpha_{0,i}) + \Delta t_{0,i} \gamma_{0,i,1,v} \right] x_{i,1} \\ & + \sum_{j=2}^{\min(v,N+1)} \sum_{i=1}^{N} \sum_{i' \neq i} \left[\underline{t}_{i,i'} (\pi_{j,v} + \lambda - \alpha_{i,i'}) + \Delta t_{i,i'} \gamma_{i,i',j,v} \right] x_{i,j-1} x_{i',j} \\ & + \sum_{j=2}^{\min(v,N+1)} \sum_{i=1}^{N} \left[\underline{d}_i (\pi_{j,v} - \rho_i) + \Delta d_i \delta_{i,j,v} \right] x_{i,j-1}, \quad \forall v \in [2, N+2]_{\mathbb{Z}}, \end{aligned} \tag{13c}$$

$$\begin{aligned} \sum_{j=k}^{v} \beta_j \geq{} & \sum_{j=k}^{\min(v,N+1)} (a_{j-1} - a_j) \pi_{j,v} + \sum_{j=k}^{\min(v,N+1)} \sum_{i=1}^{N} \left[\underline{d}_i (\pi_{j,v} - \rho_i) + \Delta d_i \delta_{i,j,v} \right] x_{i,j-1} \\ & + \sum_{j=k}^{\min(v,N+1)} \sum_{i=1}^{N} \sum_{i' \neq i} \left\{ (\pi_{j,v} + \lambda - \alpha_{i,i'}) \underline{t}_{i,i'} + \Delta t_{i,i'} \gamma_{i,i',j,v} \right\} x_{i,j-1} x_{i',j}, \\ & \forall k \in [2, N+1]_{\mathbb{Z}}, \forall v \in [k, N+2]_{\mathbb{Z}}, \end{aligned} \tag{13d}$$

$$\beta_{N+2} \geq 0, \tag{13e}$$

$$\gamma_{0,i,1,v} \geq 0, \quad \gamma_{0,i,1,v} \geq \pi_{1,v} + \lambda - \alpha_{0,i}, \quad \forall i \in [N], \forall v \in [N+2], \tag{13f}$$

$$\begin{aligned} \gamma_{i,i',j,v} \geq 0, \quad \gamma_{i,i',j,v} \geq \pi_{j,v} + \lambda - \alpha_{i,i'}, \quad & \forall i \in [N], \forall i' \in [N] \setminus \{i\}, \\ & \forall j \in [2, N+1]_{\mathbb{Z}}, \forall v \in [j, N+2]_{\mathbb{Z}}, \end{aligned} \tag{13g}$$

$$\delta_{i,j,v} \geq 0, \quad \delta_{i,j,v} \geq \pi_{j,v} - \rho_i, \quad \forall i \in [N],\ \forall j \in [2, N+1]_{\mathbb{Z}},\ \forall v \in [j, N+2]_{\mathbb{Z}}, \tag{13h}$$

where $\pi_{j,v} = -c_v^u + \sum_{l=j}^{v-1} c_l^w$ *for* $1 \leq j \leq v \leq N+2$.

With the use of Proposition 2, we can combine the minimization over $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ in problem (9). Corollary 3 provides the reformulation of problem (9) with the proof in Appendix C.

Corollary 3. *Problem* (9) *is equivalent to*

$$\min_{\boldsymbol{\alpha},\boldsymbol{\rho},\boldsymbol{\beta}} \sum_{i=1}^{N} \mu_i^d \rho_i + \sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i' \neq i} \mu_{i,i'}^t \alpha_{i,i'} x_{i,j-1} x_{i',j} + \lambda \sum_{i=1}^{N} \mu_{i,0}^t x_{i,N} + \sum_{i=1}^{N} \mu_{0,i}^t \alpha_{0,i} x_{i,1} + \sum_{j=1}^{N+2} \beta_j \tag{14a}$$

$$\text{s.t. } (13b) - (13h). \tag{14b}$$

Combining $\sup_{\mathbb{P} \in \mathcal{F}(\mathcal{S},\boldsymbol{\mu})} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{d})]$ in the form of (14) with the outer minimization problem in (7), we can obtain a mixed-integer program, which is nonlinear due to the interaction terms (such as $x_{i,j-1}x_{i',j}$ in the objective, $\alpha_{i,i'}x_{i,j-1}x_{i',j}$, $\gamma_{0,i,1,v}x_{i,1}$ and $\rho_i x_{i,j-1}$ in the constraints). To linearize this MINLP formulation, we define variables $\tau_{i,i',j-1,j} = x_{i,j-1}x_{i',j}$, $\eta_{i,i',j-1,j} = \alpha_{i,i'}\tau_{i,i',j-1,j}$, $\psi_{0,i} = \alpha_{0,i}x_{i,1}$, $\sigma_{0,i,1,v} = \gamma_{0,i,1,v}x_{i,1}$, $\phi_{i,i',j,v} = \gamma_{i,i',j,v}\tau_{i,i',j-1,j}$, $\xi_{i,j,v} = \delta_{i,j,v}x_{i,j-1}$ and $\zeta_{i,j} = \rho_i x_{i,j-1}$. We also introduce McCormick inequalities (15a) to (15k) for these variables.

$$\tau_{i,i',j-1,j} \geq x_{i,j-1} + x_{i',j} - 1,\ \tau_{i,i',j-1,j} \geq 0, \tag{15a}$$

$$\tau_{i,i',j-1,j} \leq x_{i,j-1},\ \tau_{i,i',j-1,j} \leq x_{i',j} \tag{15b}$$

$$\eta_{i,i',j-1,j} \geq \underline{\alpha}_{i,i'}\tau_{i,i',j-1,j},\ \eta_{i,i',j-1,j} \geq \alpha_{i,i'} + \overline{\alpha}_{i,i'}(\tau_{i,i',j-1,j} - 1), \tag{15c}$$

$$\eta_{i,i',j-1,j} \leq \overline{\alpha}_{i,i'}\tau_{i,i',j-1,j},\ \eta_{i,i',j-1,j} \leq \alpha_{i,i'} + \underline{\alpha}_{i,i'}(\tau_{i,i',j-1,j} - 1) \tag{15d}$$

$$\psi_{0,i} \geq \underline{\alpha}_{0,i}x_{i,1},\ \psi_{0,i} \geq \alpha_{0,i} + \overline{\alpha}_{0,i}(x_{i,1} - 1),\ \psi_{0,i} \leq \overline{\alpha}_{0,i}x_{i,1},\ \psi_{0,i} \leq \alpha_{0,i} + \underline{\alpha}_{0,i}(x_{i,1} - 1) \tag{15e}$$

$$\phi_{i,i',j,v} \geq \gamma_{i,i',j,v} + \overline{\gamma}_{i,i',j,v}(\tau_{i,i',j-1,j} - 1),\ \phi_{i,i',j,v} \geq 0 \tag{15f}$$

$$\phi_{i,i',j,v} \leq \overline{\gamma}_{i,i',j,v}\tau_{i,i',j-1,j},\ \phi_{i,i',j,v} \leq \gamma_{i,i',j,v} \tag{15g}$$

$$\sigma_{0,i,1,v} \geq \gamma_{0,i,1,v} + \overline{\gamma}_{0,i,1,v}(x_{i,1} - 1),\ \sigma_{0,i,1,v} \geq 0 \tag{15h}$$

$$\sigma_{0,i,1,v} \leq \overline{\gamma}_{0,i,1,v}x_{i,1},\ \sigma_{0,i,1,v} \leq \gamma_{0,i,1,v} \tag{15i}$$

$$\xi_{i,j,v} \geq \delta_{i,j,v} + \overline{\delta}_{i,j,v}(x_{i,j-1} - 1),\ \xi_{i,j,v} \geq 0,\ \xi_{i,j,v} \leq \delta_{i,j,v},\ \xi_{i,j,v} \leq \overline{\delta}_{i,j,v}x_{i,j-1} \tag{15j}$$

$$\zeta_{i,j} \geq \underline{\rho}_i x_{i,j-1},\ \zeta_{i,j} \geq \rho_i + \overline{\rho}_i(x_{i,j-1} - 1),\ \zeta_{i,j} \leq \overline{\rho}_i x_{i,j-1},\ \zeta_{i,j} \leq \rho_i + \underline{\rho}_i(x_{i,j-1} - 1) \tag{15k}$$

The notations $\bar{\cdot}$ and $\underline{\cdot}$ are the upper and lower bounds for the variable respectively. Accordingly, we derive the following MILP of M-DHRAS.

$$\min \sum_{i=1}^{N} \mu_i^{\mathrm{d}} \rho_i + \sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i' \neq i} \mu_{i,i'}^{\mathrm{t}} \eta_{i,i',j-1,j} + \lambda \sum_{i=1}^{N} \mu_{i,0}^{\mathrm{t}} x_{i,N} + \sum_{i=1}^{N} \mu_{0,i}^{\mathrm{t}} \alpha_{0,i} x_{i,1} + \sum_{j=1}^{N+2} \beta_j \tag{16a}$$

$$\text{s.t. } x \in \mathcal{X},\ a \in \mathcal{A}, (13f) - (13h), (15a) - (15k), \tag{16b}$$

linearized version of (13b) − (13e). (16c)

Note that the McCormick inequalities often rely on big-M coefficients which take large values and can undermine the computational efficiency. In Appendix I, we derive tight bounds of these big-M coefficients to strengthen formulation (16).

3.3. DHRAS over 1-Wasserstein ambiguity

In this section, we assume that the distribution $\mathbb{P}_{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ belongs to an ambiguity set based on the Wasserstein metric to measure the distance between probability distributions. In particular, we construct an ambiguity set based on 1-Wasserstein distance (i.e., we use ℓ_1–norm in the definition of Wasserstein metric), which often admits tractable reformulation in most real-world applications (see, e.g., Duque and Morton (2020); Hanasusanto and Kuhn (2018); Jiang et al. (2019); Saif and Delage (2021) to name a few).

Suppose that two probability distributions $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are defined on a common support $\mathcal{S} \in \mathbb{R}^N$, and let $||\cdot||_p$ represent the p-norm on $\mathbb{R}^N$ with $p \geq 1$. Suppose that random vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ follow $\mathbb{Q}_1$ and $\mathbb{Q}_2$ respectively. Then, the Wasserstein distance between $\mathbb{Q}_1$ and $\mathbb{Q}_2$, denoted as $W_p(\mathbb{Q}_1, \mathbb{Q}_2)$, represents the minimum transportation cost of moving from $\mathbb{Q}_1$ to $\mathbb{Q}_2$, where the cost of moving from $\boldsymbol{\xi}_1$ to $\boldsymbol{\xi}_2$ is measured by the norm $\|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p$. Mathematically,

$$W_p(\mathbb{Q}_1, \mathbb{Q}_2) := \left(\inf_{\Pi \in \mathcal{P}(\mathbb{Q}_1, \mathbb{Q}_2)} \mathbb{E}_{\Pi} \left[\, \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_p^p \, \right] \right)^{\frac{1}{p}}, \tag{17}$$

where $\mathcal{P}(\mathbb{Q}_1, \mathbb{Q}_2)$ is the set of all joint distributions of $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ with marginals $\mathbb{Q}_1$ and $\mathbb{Q}_2$. We also assume that $\mathbb{P}_{\boldsymbol{\xi}}$ is only observed via a finite set $\{\hat{\boldsymbol{\xi}}^1, \ldots, \hat{\boldsymbol{\xi}}^R\}$ of R i.i.d. samples (which may come from the historical realizations of service durations and travel times). Accordingly, we consider the following p-Wasserstein ambiguity set

$$\mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}, \epsilon) = \left\{ \mathbb{Q}_{\boldsymbol{\xi}} \in \mathcal{P}(\mathcal{S}) : W_p(\mathbb{Q}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}}) \leq \epsilon \right\}, \tag{18}$$

where $\mathcal{P}(\mathcal{S})$ is the set of all probability distributions on $\mathcal{S}$, $\hat{\mathbb{P}}^R_{\boldsymbol{\xi}} = \frac{1}{R}\sum_{r=1}^R \delta_{\hat{\boldsymbol{\xi}}^r}$ is the empirical distribution of $\boldsymbol{\xi}$ based on the R i.i.d samples with δ being the Dirac measure, and $\epsilon > 0$ is the radius of the ambiguity set. The set $\mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}, \epsilon)$ can be viewed as the p-Wasserstein ball of radius ϵ centered at the empirical distribution $\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}$. Using the ambiguity set $\mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}, \epsilon)$, we formulate W-DHRAS as follows:

$$\text{(W-DHRAS)} \quad \hat{Z}(R, \epsilon) = \min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{a} \in \mathcal{A}} \left\{ \sup_{\mathbb{Q}_{\boldsymbol{\xi}} \in \mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}, \epsilon)} \mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}} \left[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) \right] \right\}. \tag{19}$$

In data-driven approaches such as W-DHRAS, we often seek asymptotic consistency. Specifically, we expect that as the sample size R increases to infinity, the optimal value of the problem (19) $\hat{Z}(R, \epsilon)$ converges to Z^* (the optimal value of the SP model in (4) with perfect knowledge of

$\mathbb{P}_{\boldsymbol{\xi}}$), and an optimal solution $(\boldsymbol{x}, \boldsymbol{a})$ of (W-DHRAS) converges to an optimal solution of problem (4). Additionally, if $\hat{Z}(R, \epsilon) > Z^*$ almost surely, then W-DHRAS provides a safe upper bound guarantee on the expected total cost with any finite data size R. Recall that our support set $\mathcal{S}$ is non-empty, convex, and compact (note that this a mild assumption which holds in most real-world applications such as healthcare and HRAS). As such, we can use the existing theory in establishing the asymptotic consistency and finite sample guarantee of our W-DHRAS. Specifically, given this assumption on $\mathcal{S}$, Lemma 1 in Jiang et al. (2019) and Theorem 2 in Fournier and Guillin (2015) assure that the p-Wasserstein ambiguity set incorporates the true distribution $\mathbb{P}$ with high confidence. Theorem 1 in Jiang et al. (2019) and Theorem 3.6 in Mohajerin Esfahani and Kuhn (2018) assure asymptotic consistency, and Theorem 2 in Jiang et al. (2019) and Theorem 3.5 in Mohajerin Esfahani and Kuhn (2018) assure finite-data guarantee of W-DHRAS. We refer the readers to these papers for detailed proofs and discussion on these results, and we provide the adapted proofs in Appendix D–Appendix F for completeness.

3.3.1. ***MILP reformulation of W-DHRAS***

In this section, we derive an equivalent MILP reformulation of our W-DHRAS model. First, we consider the inner maximization problem of W-DHRAS for a fixed $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$.

$$\sup_{\mathbb{Q}_{\boldsymbol{\xi}} \in \mathcal{F}_1(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}}, \epsilon)} \mathbb{E}_{\mathbb{Q}_{\boldsymbol{\xi}}}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})] \tag{20}$$

In Proposition 4, we present an equivalent dual formulation of problem (20) (See Appendix G for a detailed proof).

Proposition 4. *The optimal value of formulation* (20) *equals that of the following formulation:*

$$\inf_{\rho} \left\{ \epsilon\rho + \frac{1}{R} \sum_{r=1}^{R} \sup_{\boldsymbol{\xi} \in \mathcal{S}} \left\{ f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 \right\} \right\} \tag{21a}$$

$$s.t. \;\; \rho \geq 0. \tag{21b}$$

Note that $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1$ is neither convex nor concave in $\boldsymbol{\xi}$. Thus, formulation (21) is potentially challenging to solve because it requires solving R non-convex optimization problems. Given that the support of service duration and travel times are rectangular and finite, we next show that we can recast these problems as linear programs for fixed ρ and $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$. In what follows, we provide a high-level road map of the reformulation, and we relegate the notationally heavy proofs and details to Appendices. First, for fixed $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$ and $\rho \geq 0$, we denote $g_r(\rho, \boldsymbol{x}, \boldsymbol{a}) = \sup_{\xi \in \mathcal{S}} \{ f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 \}$. Given the dual formulation of $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ in (10), we write $g_r(\rho, \boldsymbol{x}, \boldsymbol{a})$ as follows for each r using the dual of $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$.

$$g_r(\rho, \boldsymbol{x}, \boldsymbol{a}) = \sup_{(\boldsymbol{d}, \boldsymbol{t}) \in \mathcal{S}} \sup_{\boldsymbol{y} \in \mathcal{Y}} \Bigg\{ -a_1 y_1 + \sum_{j=2}^{N+1} (a_{j-1} - a_j) y_j + \sum_{j=2}^{N+1} \sum_{i=1}^{N} d_i y_j x_{i,j-1} + \sum_{i=1}^{N} (t_{0,i} y_1 + \lambda t_{0,i}) x_{i,1}$$

$$\sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i\neq i'}(t_{i,i'}y_j+\lambda t_{i,i'})x_{i,j-1}x_{i',j}+\sum_{i=1}^{N}\lambda t_{i,0}x_{i,N}$$
$$\left.-\rho\sum_{j=2}^{N+1}\sum_{i=1}^{N}|d_i-\hat{d}_i^r|x_{i,j-1}-\rho\sum_{j=2}^{N+1}\sum_{i=1}^{N}\sum_{i'\neq i}|t_{i,i'}-\hat{t}^r_{i,i'}|x_{i,j-1}x_{i',j}\right\}. \tag{22}$$

Using the same proof techniques in Proposition 2, we can formulate (22) (see Appendix H for a detailed proof).

Proposition 5. *Problem* (22) *is equivalent to*

$$\min_{\boldsymbol{\beta}^r,\boldsymbol{u}^r,\boldsymbol{\nu}^r}\ \sum_{j=1}^{N+2}\beta_j^r+\sum_{i=1}^{N}u^r_{i,0}x_{i,N} \tag{23a}$$

$$\text{s.t.}\quad \beta_1^r\geq -a_1\pi_{1,1}+\sum_{i=1}^{N}u^r_{0,i,1,1}x_{i,1}, \tag{23b}$$

$$\sum_{j=1}^{v}\beta_j^r\geq -a_1\pi_{1,v}+\sum_{j=2}^{\min(v,N+1)}(a_{j-1}-a_j)\pi_{j,v}+\sum_{i=1}^{N}u^r_{0,i,1,v}x_{i,1}$$
$$+\sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i}u^r_{i,i',j,v}x_{i,j-1}x_{i',j}+\sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\nu^r_{i,j,v}x_{i,j-1},\quad \forall v\in[2,N+2]_{\mathbb{Z}}, \tag{23c}$$

$$\sum_{j=k}^{v}\beta_j^r\geq \sum_{j=k}^{\min(v,N+1)}(a_{j-1}-a_j)\pi_{j,v}+\sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\nu^r_{i,j,v}x_{i,j-1}$$
$$+\sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i}u_{i,i',j,v}x_{i,j-1}x_{i',j},\quad \forall k\in[2,N+1]_{\mathbb{Z}},\ \forall v\in[k,N+2]_{\mathbb{Z}}, \tag{23d}$$

$$\beta^r_{N+2}\geq 0, \tag{23e}$$

$$u^r_{i,0}\geq\lambda\hat{t}^r_{i,0},\quad u^r_{i,0}\geq\lambda\bar{t}_{i,0}-\rho(\bar{t}_{i,0}-\hat{t}^r_{i,0}),\quad \forall i\in[N], \tag{23f}$$

$$u^r_{0,i,1,v}\geq(\pi_{1,v}+\lambda)\underline{t}_{0,i}-\rho(\hat{t}^r_{0,i}-\underline{t}_{0,i}),\quad u^r_{0,i,1,v}\geq(\pi_{1,v}+\lambda)\hat{t}^r_{0,i},$$
$$u^r_{0,i,1,v}\geq(\pi_{1,v}+\lambda)\bar{t}_{0,i}-\rho(\bar{t}_{0,i}-\hat{t}^r_{0,i}),\quad \forall i\in[N],\ v\in[N+2], \tag{23g}$$

$$u^r_{i,i',j,v}\geq(\pi_{j,v}+\lambda)\underline{t}_{i,i'}-\rho(\hat{t}^r_{i,i'}-\underline{t}_{i,i'}),\quad u^r_{i,i',j,v}\geq(\pi_{j,v}+\lambda)\bar{t}_{i,i'}-\rho(\bar{t}_{i,i'}-\hat{t}^r_{i,i'}),$$
$$u^r_{i,i',j,v}\geq(\pi_{j,v}+\lambda)\hat{t}^r_{i,i'},\quad \forall i\in[N],\ i'\in[N]\setminus\{i\},\ j\in[2,N+1]_{\mathbb{Z}},\ v\in[j,N+2]_{\mathbb{Z}}, \tag{23h}$$

$$\nu^r_{i,j,v}\geq\pi_{j,v}\underline{d}_i-\rho(\hat{d}^r_i-\underline{d}_i),\quad \nu^r_{i,j,v}\geq\pi_{j,v}\bar{d}_i-\rho(\bar{d}_i-\hat{d}^r_i),$$
$$\nu^r_{i,j,v}\geq\pi_{j,v}\hat{d}^r_i,\quad \forall i\in[N],\ j\in[2,N+1]_{\mathbb{Z}},\ v\in[j,N+2]_{\mathbb{Z}}. \tag{23i}$$

Let $(C^r)=\{\text{constraints (23b)}-\text{(23i)}\}$. Summing $g_r(\rho,\boldsymbol{x},\boldsymbol{a})$ in the form of Proposition 5 over r and combining it with the outer minimization in (21a) and (19), we derive

$$\min_{\boldsymbol{x},\boldsymbol{a},\rho,\boldsymbol{\beta},\boldsymbol{u}^r,\boldsymbol{\nu}^r}\left\{\epsilon\rho+\frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{N}u^r_{i,0}x_{i,N}+\sum_{j=1}^{N+2}\beta_j^r\right)\right\} \tag{24a}$$

$$
\begin{aligned}
\text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X},\, \boldsymbol{a} \in \mathcal{A}, && \text{(24b)}\\
& (C^r),\ r \in [R], && \text{(24c)}\\
& \rho \geq 0. && \text{(24d)}
\end{aligned}
$$

We can apply McCormick inequalities to linearize the non-linear terms. Here, we reuse the Greek letters in M-DHRAS model for the McCormick inequalities. Let $\tau_{i,i',j-1,j} = x_{i,j-1}x_{i',j}$, $\psi^r_{i,0} = u^r_{i,0}x_{i,N}$, $\sigma^r_{0,i,1,v} = u^r_{0,i,1,v}x_{i,1}$, $\phi^r_{i,i',j,v} = u^r_{i,i',j,v}\tau_{i,i',j-1,j}$ and $\zeta^r_{i,j,v} = \nu^r_{i,j,v}x_{i,j-1}$. In addition to (15a) and (15b), we introduce the following McCormick inequalities.

$$
\psi^r_{i,0} \geq \underline{u}^r_{i,0}x_{i,N},\ \psi^r_{i,0} \geq u^r_{i,0} + \overline{u}^r_{i,0}(x_{i,N}-1),\ \psi^r_{i,0} \leq \overline{u}^r_{i,0}x_{i,N},\ \psi^r_{i,0} \leq u^r_{i,0} + \underline{u}^r_{0,i}(x_{i,N}-1) \quad \text{(25a)}
$$

$$
\phi^r_{i,i',j,v} \geq \underline{u}^r_{i,i',j,v}\tau_{i,i',j-1,j},\ \phi^r_{i,i',j,v} \geq u^r_{i,i',j,v} + \overline{u}^r_{i,i',j,v}(\tau_{i,i',j-1,j}-1) \quad \text{(25b)}
$$

$$
\phi^r_{i,i',j,v} \leq \overline{u}^r_{i,i',j,v}\tau_{i,i',j-1,j},\ \phi^r_{i,i',j,v} \leq u^r_{i,i',j,v} + \underline{u}^r_{i,i',j,v}(\tau_{i,i',j-1,j}-1) \quad \text{(25c)}
$$

$$
\sigma^r_{0,i,1,v} \geq \underline{u}^r_{0,i,1,v}x_{i,1},\ \sigma^r_{0,i,1,v} \geq u^r_{0,i,1,v} + \overline{u}^r_{0,i,1,v}(x_{i,1}-1) \quad \text{(25d)}
$$

$$
\sigma^r_{0,i,1,v} \leq \overline{u}^r_{0,i,1,v}x_{i,1},\ \sigma^r_{0,i,1,v} \leq u^r_{0,i,1,v} + \underline{u}^r_{0,i,1,v}(x_{i,1}-1) \quad \text{(25e)}
$$

$$
\zeta^r_{i,j,v} \geq \underline{\nu}^r_{i,j,v}x_{i,j-1},\ \zeta^r_{i,j,v} \geq \nu^r_{i,j,v} + \overline{\nu}^r_{i,j,v}(x_{i,j-1}-1) \quad \text{(25f)}
$$

$$
\zeta^r_{i,j,v} \leq \overline{\nu}^r_{i,j,v}x_{i,j-1},\ \zeta^r_{i,j,v} \leq \nu^r_{i,j,v} + \underline{\nu}^r_{i,j,v}(x_{i,j-1}-1) \quad \text{(25g)}
$$

Let $(D^r) = \{\text{constraints (25a)} - \text{(25g)}, \text{(23f)} - \text{(23i)}\}$. Then, the MILP reformulation of W-DHRAS model reads

$$
\begin{aligned}
\min \quad & \epsilon\rho + \frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{N}\psi^r_{i,0} + \sum_{j=1}^{N+2}\beta^r_j\right) && \text{(26a)}\\
\text{s.t.} \quad & \boldsymbol{x} \in \mathcal{X},\, \boldsymbol{a} \in \mathcal{A},\, \text{(15a)} - \text{(15b)}, && \text{(26b)}\\
& (D^r),\ \forall r \in [R], && \text{(26c)}\\
& \text{linearized version of (23b)} - \text{(23e)},\ \forall r \in [R]. && \text{(26d)}
\end{aligned}
$$

We derive tight bounds on the variables $\boldsymbol{u}^r$ and $\boldsymbol{\nu}^r$ in Appendix I.

4. Numerical Experiments

Our computational study's primary objective is to compare the performance of the proposed DRO models (M-DHRAS and W-DHRAS) and a sample average approximation (SAA) model. The SAA model solves model (4) with $\mathbb{P}_{\boldsymbol{\xi}}$ replaced by an empirical distribution based on N samples of the random parameters (see Appendix J for the formulation). For simplicity, we call the SAA model as the SP model. We focus on HRAS instances where the sample size is possibly small, which is often seen in healthcare applications. In Section 4.1, we describe the set of HRAS instances that we constructed and discuss other experimental setup. In Section 4.2, we examine the choice of ϵ in W-DHRAS model and the corresponding effect on the out-of-sample simulation performance.

In Section 4.3, we analyze the appointment time structure while in Section 4.4, we analyze the optimal solutions of the models and then compare their out-of-sample simulation performance. We also discuss the reliability of the models in Section 4.5. Finally, in Section 4.6, we compare the computational performance of the three models. In addition, we compare the computational efficiency of our model formulation with the one in Zhan et al. (2021).

4.1. Description of the experiments

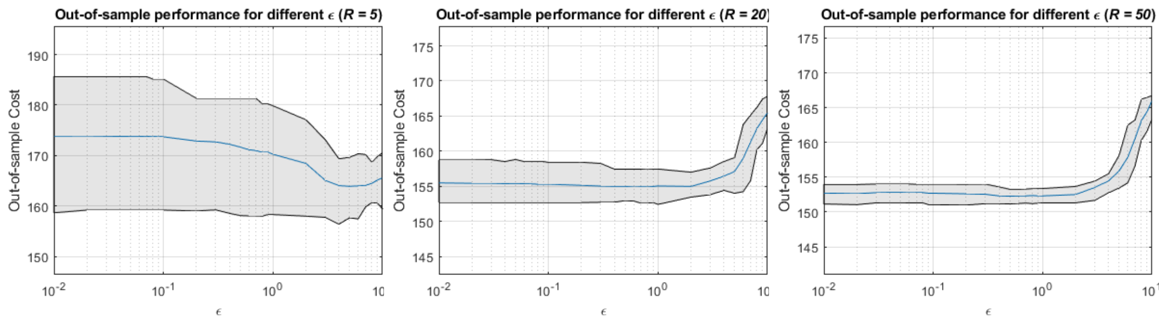
We construct HRAS instances based on the same parameters settings and assumptions made in recent related literature (see, e.g., Jiang et al., 2019, Zhan et al., 2021, Zhan and Wan, 2018). The average number of customers that an operator may visit per day is often less than six in home healthcare and banking and less than 10 in repair service (NAHC, 2010; Yuan et al., 2015; Zhan et al., 2021). Accordingly, we consider problem instances with 6, 8 and 10 customers. In particular, we are interested in the home health care applications.

We consider two different cost structures for waiting, idling and overtime in the objective function: (a) $(c^{\mathrm{w}}_j, c^{\mathrm{u}}_j, c^{\mathrm{o}}) = (2, 1, 20)$ (Jiang et al., 2017, Jiang et al., 2019), and (b) $(c^{\mathrm{w}}_j, c^{\mathrm{u}}_j, c^{\mathrm{o}}) = (1, 5, 7.5)$ (Shehadeh et al., 2020). For the transportation cost, we consider $\lambda \in \{0.5, 1, 2\}$ (Zhan et al., 2021). We set the standard work hours L to 480 as in Zhan and Wan (2018).

We use the lognormal distribution (Jiang et al., 2019) for the service time d_i truncated on the interval $[10, 50]$ with mean μ and $\sigma = 0.5\mu$, where μ is generated from $U[25, 35]$ ($U[a, b]$ refers to uniform distribution over the interval $[a, b]$). Our model works with any choices of the range $[10, 50]$. Previous studies such as Zhan et al. (2021) assumes deterministic travel time based on Euclidean distance. In our study, we generate the random travel time $t_{i,i'}$ from $U[15, 25]$. That is, we assume that customers are fairly separated and traveling from one place to the other takes 20 minutes on average (though our models can solve instances with any ranges and distribution of travel time). This is also consistent with prior and recent literature. Nikzad et al. (2021) particularly pointed out that customers within a service region form a basic unit or a cluster that share the same distribution of travel time, which is seen in urban areas. We round each generated parameter to the nearest integer.

We use the same upper and lower bounds of service time $[\underline{d}_i, \overline{d}_i] = [10, 50]$ and travel time $[\underline{t}_{i,i'}, \overline{t}_{i,i'}] = [15, 25]$ in M-DHRAS and W-DHRAS. In the M-DHRAS model, the mean parameters $\boldsymbol{\mu}^{\mathrm{d}}$ and $\boldsymbol{\mu}^{\mathrm{t}}$ in the ambiguity set are set as the sample mean of the data. We also introduce symmetry-breaking constraints (see Appendix K). These inequalities leverage the homogeneous nature of customers (i.e., share common distributions of service and travel times) to break the symmetry in the routing decision.

We implement all the models in AMPL programming language and use CPLEX (version 12.7.0.0) solver with the default setting. The relative MIP gap tolerance is set to 0.02 while most of the instances have a terminal relative MIP gap tolerance very close to 0. All the experiments



(a) $R = 5$ (b) $R = 20$ (c) $R = 50$

Figure 1: Performance of W-DHRAS with different choices of ϵ

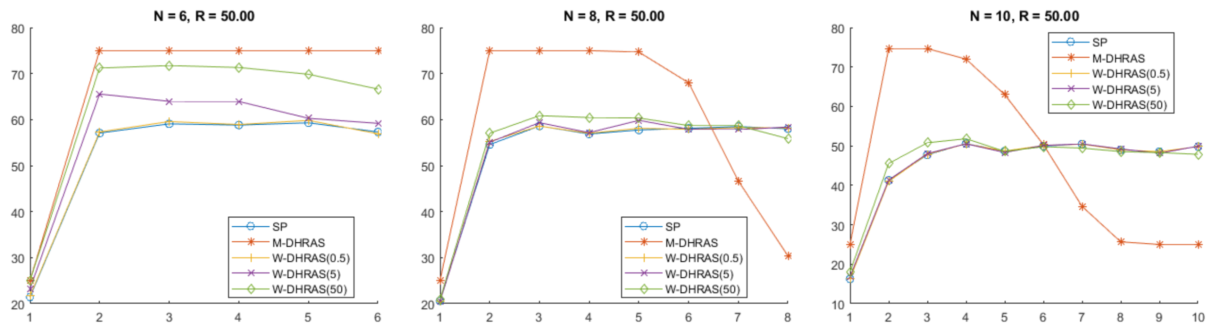
are conducted on a computer with AMD Opteron 2.0 GHz CPU and 16 Gb memory. The time limit for solving each model is set to 2 hours. In some instances where the CPLEX solver run out of memory, we also impose the memory reduction setting (available from CPLEX solver option).

4.2. ***Effect of ϵ in W-DHRAS model***

In the W-DHRAS model, there is one parameter in the uncertainty set that serves as an input: the Wasserstein ball's radius ϵ. In this subsection, we demonstrate the effect of ϵ on the out-of-sample performance of the W-DHRAS's optimal solution, $(\hat{\boldsymbol{x}}(\epsilon, R), \hat{\boldsymbol{a}}(\epsilon, R))$, with respect to the radius ϵ. For illustrative purposes, we focus on one instance of 6 customers with cost structure $(c^{\mathrm{w}}_j, c^{\mathrm{u}}_j, c^{\mathrm{o}}) = (2, 1, 20)$ and $\lambda = 0.5$.

For each $\epsilon \in \{0.01, 0.02, \ldots, 0.09, 0.1, \ldots, 0.9, 1, \ldots, 10\}$ (i.e., log-scaled interval as in Mohajerin Esfahani and Kuhn, 2018 and Jiang et al., 2019), we evaluate the out-of-sample performance as follows. First, for each customer, we randomly generate 30 data sets of service duration and travel time scenarios, each consisting of $R \in \{5, 20, 50\}$ scenarios. We generate these data sets under the same settings described in Section 4.1. Second, we solve the W-DHRAS model in (26) using the generated data sets under each of the candidate Wasserstein radius ϵ. Finally, we fix the first-stage variables to the optimal solution of each instance, and then re-optimize the second-stage of the SP using 10,000 out-of-sample (unseen) data. This is to compute the corresponding out-of-sample overtime, idle time, travel time, and waiting time and hence, the second-stage cost.

Figure 1 illustrates the out-of-sample performance with $R \in \{5, 20, 50\}$ under different choices of ϵ. The blue line represents the mean of the 30 out-of-sample costs, while the shaded region is the area between the 20th and 80th percentiles of the 30 out-of-sample costs. It is quite evident that the out-of-sample performance first decreases with ϵ and then increases after some value of ϵ. This pattern is often observed in the literature (see, e.g., Mohajerin Esfahani and Kuhn, 2018 and Jiang et al., 2019). The values of ϵ which gives the smallest out-of-sample cost with $R = 5$, 20 and 50 are respectively 5 (Figure 1a), 0.6 (Figure 1b) and 0.5 (Figure 1c) respectively. This decrease in ϵ with the increase in R is not surprising. Intuitively, a small sample provides little information



(a) $N = 6$ (b) $N = 8$ (c) $N = 10$

Figure 2: Mean inter-arrival times with $R = 50$ under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $\lambda = 2$

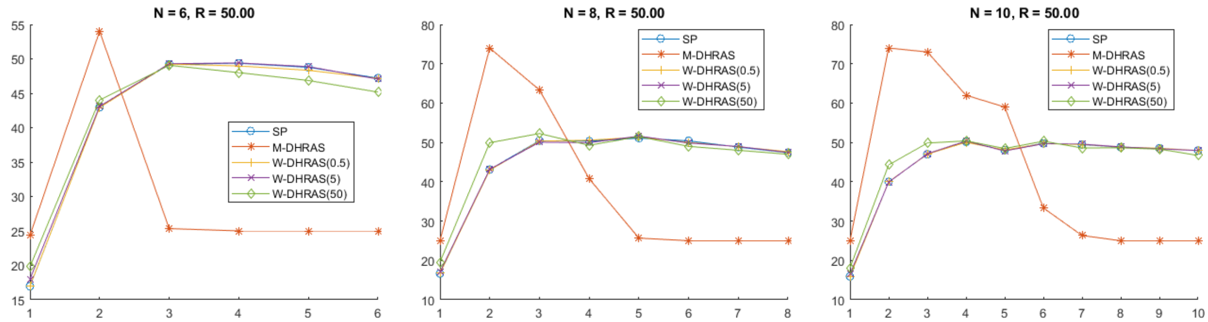
on the true distribution, and thus a larger ϵ produces robust solutions that better hedge against ambiguity. In contrast, with a larger sample, we have more information from the data, and so we can make less conservative decisions using a smaller ϵ value. As such, one should choose a larger (smaller) ϵ with a small (large) sample. Indeed, we can see that the out-of-sample cost for $R = 20$ is smaller than that of $R = 5$ in most of the cases, which indicates that a larger sample size could give a better out-of-sample performance.

In practice, decision-makers do not often have optimization expertise (or time) to run the above procedure (or other iterative or cross validation procedures). Additionally, we do not have full distributional information most of the time but only a small set of data samples. In the following experiments, we pick three different values of ϵ, namely 0.5, 5, and 50, which captures different extents of robustness of the W-DHRAS model. For brevity, we label them as W-DHRAS(0.5), W-DHRAS(5) and W-DHRAS(50) respectively.

4.3. ***Appointment time structure***

In this section, we analyze the optimal appointment structure produced by the SP, M-DHRAS, and W-DHRAS models. For illustrative purposes and brevity, we present results under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ with $\lambda = 2$. We observe similar results with other choices of λ (see Appendix L). Figures 2 and 3 present the optimal schedules of the operator produced by the SP and DRO models under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$, respectively. The point $(x, y) = (i$, inter-arrival time) of every schedule in each subfigure corresponds to the mean optimal inter-arrival time (i.e., differences between the scheduled service start times of two consecutive customers, $I_i = a_i - a_{i-1}$ for $i \in [N]$ with $a_0 = 0$).

We first analyze the results under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ in Figure 2. First, all models assign less time for the first customer than the subsequent customers and more time between the first and second customers. The models distribute (roughly) equal time between the middle customers (e.g., customers 2-5 in Figure 2a and customers 3-7 in Figure 2b), and schedule less time between



(a) $N = 6$ (b) $N = 8$ (c) $N = 10$

Figure 3: Mean inter-arrival times with $R = 50$ under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ and $\lambda = 2$

the last 2 to 3 customers, especially when $N = 10$ for M-DHRAS, though SP allocates (roughly) equal time for both middle and last customers. Second, it is evident that M-DHRAS is the most conservative model and tends to schedule more time between customers when $N = 6$ than the SP and W-DHRAS models. Also, M-DHRAS model yields a dome-shaped inter-arrival time structure (i.e., the time between customers first increases then decreases) when $N = 8$ and $N = 10$. Third, the SP model always schedules less time between the customers than the DRO models (except possibly the last few customers). This is not surprising since the SP model seeks risk-neutral decisions. Fourth, the W-DHRAS model yields a similar pattern as the SP model; however, it schedules more time between customers than the SP as ϵ increases. This makes sense because as ϵ increases, the W-DHRAS model becomes more conservative and schedules more time between appointments. This demonstrates that the W-DHRAS model is more conservative in hedging ambiguity than the SP model but less conservative than the M-DHRAS model. In the next section, we show that the W-DHRAS appointment pattern yields the best out-of-sample performance under various distributions.

Next, we analyze the results under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ in Figure 3. We observe different patterns under this cost structure. All models assign more time for the first few customers but gradually less time for the middle and last few customers. The pattern is different from the cost structure, mainly due to a significantly higher idling cost relative to the waiting cost. However, we can still conclude that M-DHRAS is the most conservative model, but instead of allocating more time, it allocates less time between customers to hedge against scenarios that yields long idling time. Third, we observe that the dome-shaped inter-arrival time structure for all models, which is particularly prominent in $N = 6$. Forth, we note that W-DHRAS models with $\epsilon = 0.5$ and $\epsilon = 5$ roughly allocates the same amount of time between customers as the SP model. However, with $\epsilon = 50$, W-DHRAS model tends to allocate more time for the first few customers but fewer time for the last few customers.

4.4. *Out-of-sample performance*

In this section, we compare the optimal solutions of the three models we analyzed in the previous section. We test the out-of-sample performance (i.e., the objective value obtained by simulating the optimal solution of a model under a larger unseen data) as follows. For each $N \in \{6, 8, 10\}$ and $R \in \{5, 10, 20, 50\}$, we generate 30 SP, M-DHRAS and W-DHRAS instances using the same parameters settings described in Section 4.1. We solve each instance and obtain the optimal first-stage $(\boldsymbol{x}, \boldsymbol{a})$ decision. Fixing the first-stage decision to $(\boldsymbol{x}, \boldsymbol{a})$, we then re-optimize the second-stage of the SP using the following five sets of $10,000$ samples. In the first set, we assume that the data we rely on in the optimization comes from the true distribution. We use data from Sets 2-5 to evaluate the performance of the models under the case when the data we rely on in the optimization may follow a biased or misspecified distribution different from the true distribution.

Set 1. We assume perfect information for the distributions. That is, we generate the $10,000$ samples from the same distribution we use in the optimization as discussed in Section 4.1. This simulation assumes that the data comes from the true unknown distribution.

Set 2. In this set, we assume that we have misspecified the distribution of the travel time in the optimization. We generate $t_{i,i'}$ from $U[25, 35]$ instead of $U[15, 25]$. That is, the average travel time takes 10 minutes longer than usual. This situation (shift in the travel time range) might be seen in practice due to unexpected traffic congestion (e.g., caused by traffic accidents, weather conditions etc.).

Set 3. In this set, we assume that we have misspecified the service time distribution in the optimization (i.e., lognormal is not the true distribution). We simulate d_i from the U-shaped beta distribution $B(0.5, 0.5)$ on $[10, 50]$ (Jiang et al., 2019), which has the same mean and range as the in-sample distribution.

Set 4. In this set, we assume that we have misspecified both the service and travel times distributions in the optimization. Specifically, we follow a similar out-of-sample simulation testing procedure described in Wang et al. (2020) and perturb the support of the random travel and service times by a parameter δ as $[(1-\delta)$ lower bound, $(1+\delta)$ upper bound$]$, where $\delta \in \{0.1, 0.25, 0.5\}$. A higher value of δ corresponds to a higher variation level.

Set 5. We generate data similar to Set 4 but only changing the service time.

For brevity, we focus on HRAS instances of $N = 6$ customers (i.e., the average number of customers that a home service operator often visits per day; see Section 4.1) and present results under $(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$ and $\lambda = 2$. We observe similar out-of-sample performance with $N = 8$ and $N = 10$, and under other values of λ (see Appendix M).

Figure 4 shows the results using Set 1, where the shaded region represents the 20% and 80% percentiles of the out-of-sample cost. Obviously, the optimal solutions of the M-DHRAS model have poor out-of-sample performance. Recall from Section 4.3 that M-DHRAS schedules a long

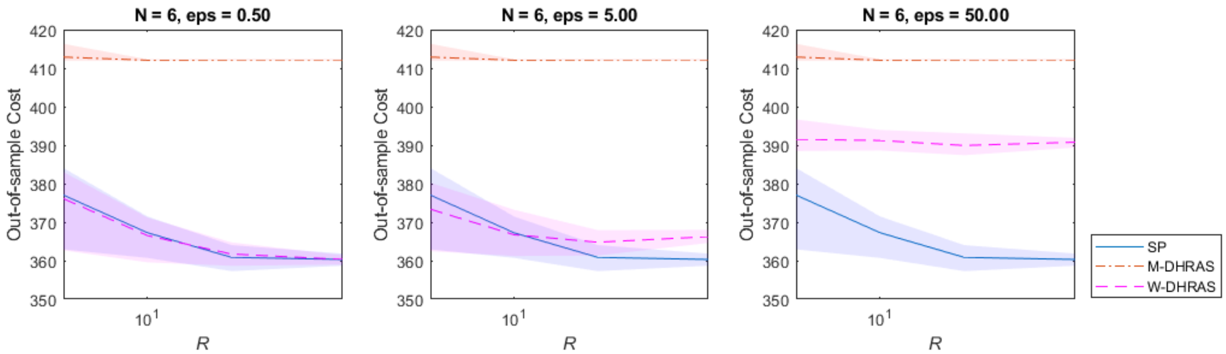


Figure 4: Out-of-sample cost with cost structure (a) and $\lambda = 2$ under Set 1

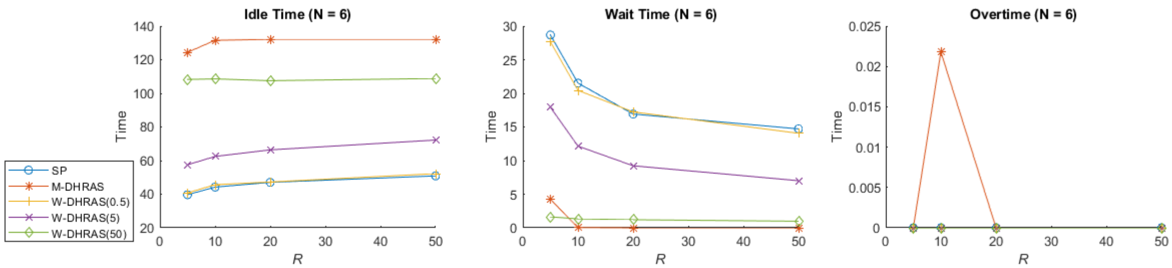


Figure 5: Mean idle time, wait time and overtime with cost structure (a) and $\lambda = 2$ under Set 1

time between customers, which, as shown in Figure 5, yields a significant amount of operator idle time (indicating poor utilization of the operator's time). The out-of-sample performance of the W-DHRAS depends on ϵ, which intuitively follows from the observation that the optimal solutions depend on ϵ (Section 3.3). When $\epsilon = 0.5$, the performance of W-DHRAS is similar to SP, while a more conservative choice of ϵ such as 50 yields a poorer performance as compared to SP. When $\epsilon = 5$, it outperforms SP, especially when the sample size is small. This demonstrates that the W-DHRAS model has a superior performance when the data size is small.

Next, we discuss results for the case when we have misspecified the distributions in the optimization. Figure 6 shows the performance under Set 2, where a longer travel time is encountered. In contrast to the results in Figure 4, M-DHRAS and W-DHARS(50) outperforms SP, W-DHRAS(0.5) and W-DHRAS(5). This demonstrates that M-DHRAS and W-DHRAS (50), which are more conservative than the other models, have superior performance in environments where travel time increases by a lot. This is not suprising since M-DHRAS and W-DHRAS (50) schedule longer time between customers that could hedge against the increase in travel time. The W-DHRAS model outperforms the SP model and the M-DHRAS model when $\epsilon = 50$ (a conservative choice of ϵ). We can observe from Figure 7 the trade-off between idle and wait time. Under Set 2, SP model results in a significant amount of waiting time but a small idling time compared with the two DRO models.

We also demonstrate one of the cases that the supports are perturbed. Figures 8 to 10 show

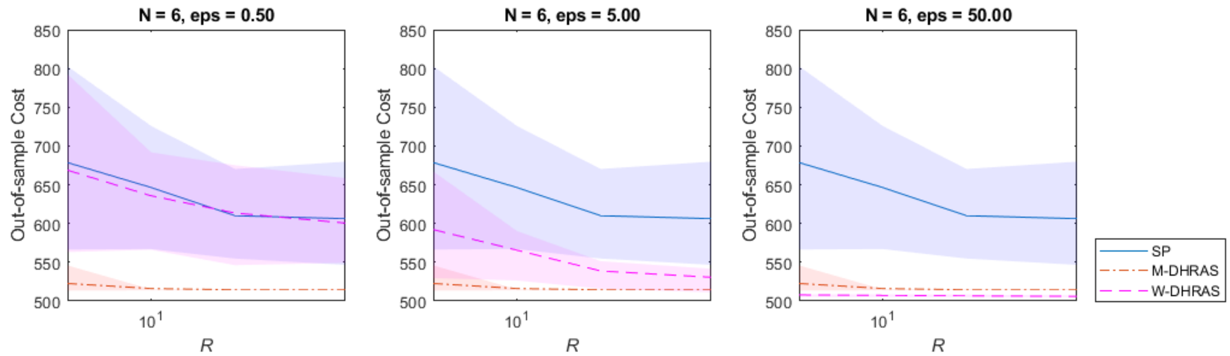


Figure 6: Out-of-sample cost with cost structure (a) and $\lambda = 2$ under Set 2

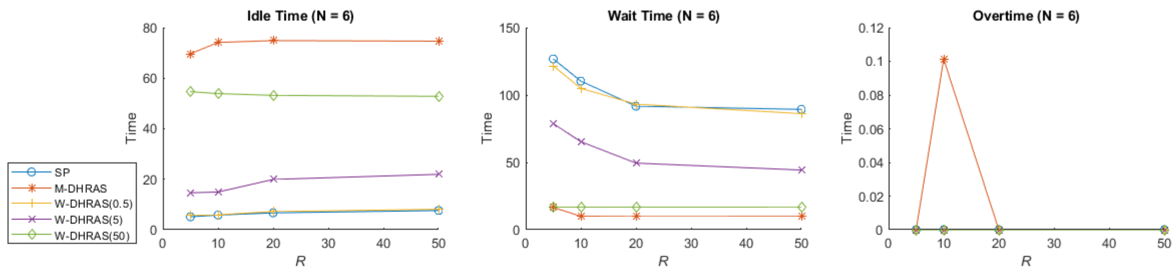


Figure 7: Mean idle time, wait time and overtime with cost structure (a) and $\lambda = 2$ under Set 2

the results for Set 4 with different choices of δ. We observe that M-DHRAS performs better (with respect to SP) when δ increases, which corresponds to a larger deviation. The W-DHRAS(5) model consistently outperforms the SP model. We also observe that the W-DHRAS model has a larger edge over SP when the sample size is small. This is consistent with what we have discussed in previous cases. All the results demonstrate that DHRAS models work better than the SP model when there is a large deviation from the sample distribution.

4.5. ***Reliability of the models***

In this section, we analyze the reliability of the three models. Home service operators often need to estimate the budgets on the operational costs ahead of the actual service provision. The optimal value of an HRAS optimization model serves as a relevant estimate of the associated cost if we implement the corresponding optimal decision $(\boldsymbol{x}, \boldsymbol{a})$. Intuitively, when the estimated cost is smaller than the actual cost, the operator may run into financial problems related to a budget deficit and poor planning decisions. A risk-averse operator would then prefer implementing decisions that provide an upper bound on the estimated cost, i.e., seek solutions with higher reliability. Hence, we make use of the reliability measure. For any given decision $(\boldsymbol{x}, \boldsymbol{a})$, reliability is the probability that the optimal value from the model is greater than or equal to the actual cost $\mathbb{E}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})]$. Mathematically, if $v(R)$ is the optimal value and $(\boldsymbol{x}(R), \boldsymbol{a}(R))$ is the corresponding optimal solution

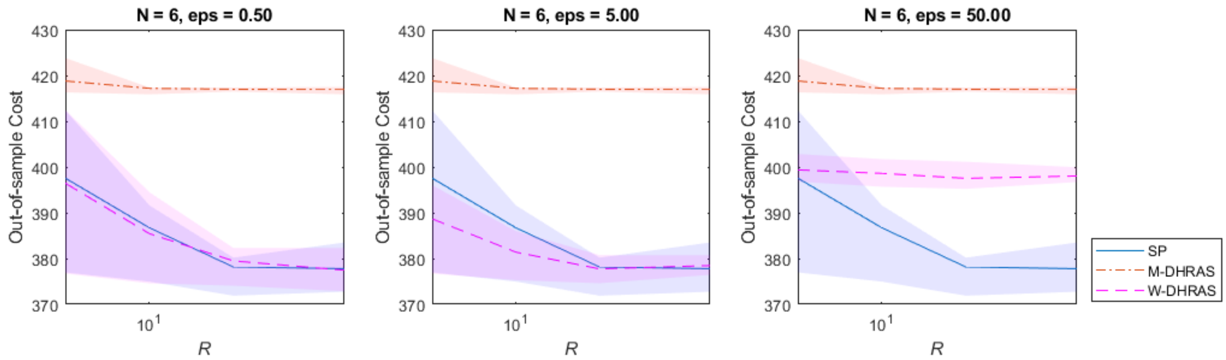


Figure 8: Out-of-sample cost with cost structure (a) and $\lambda = 2$ under Set 4 ($\delta = 0.10$)

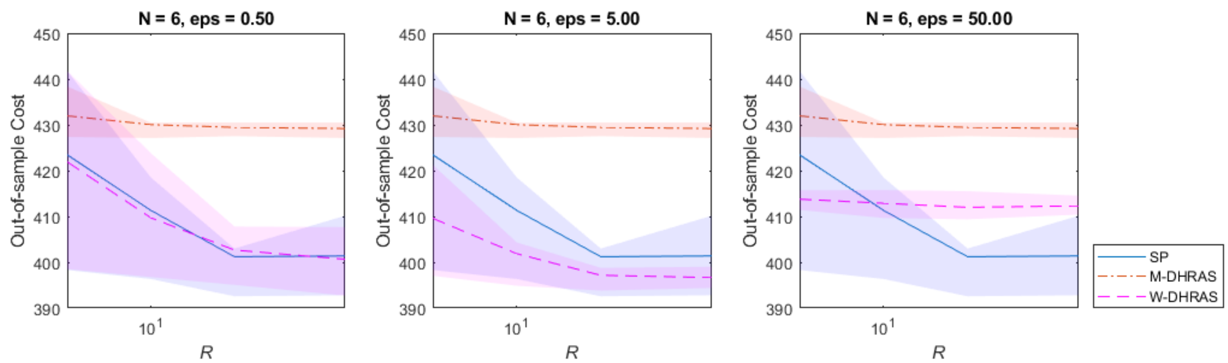


Figure 9: Out-of-sample cost with cost structure (a) and $\lambda = 2$ under Set 4 ($\delta = 0.25$)

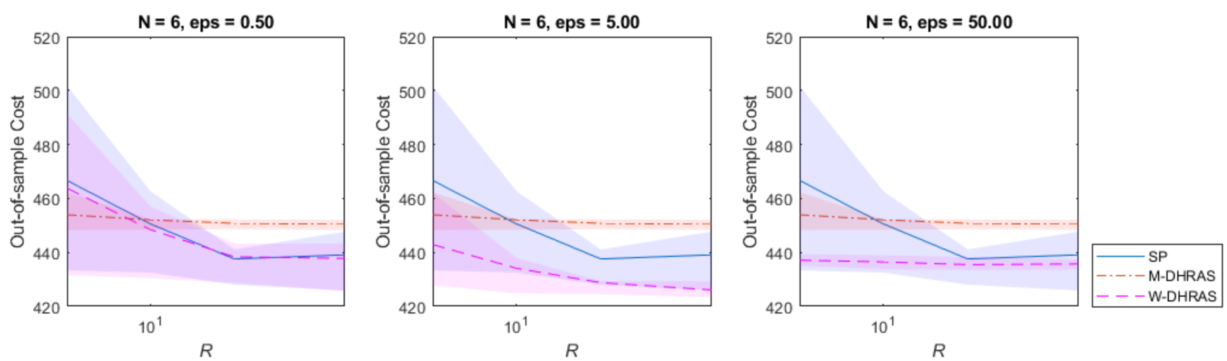


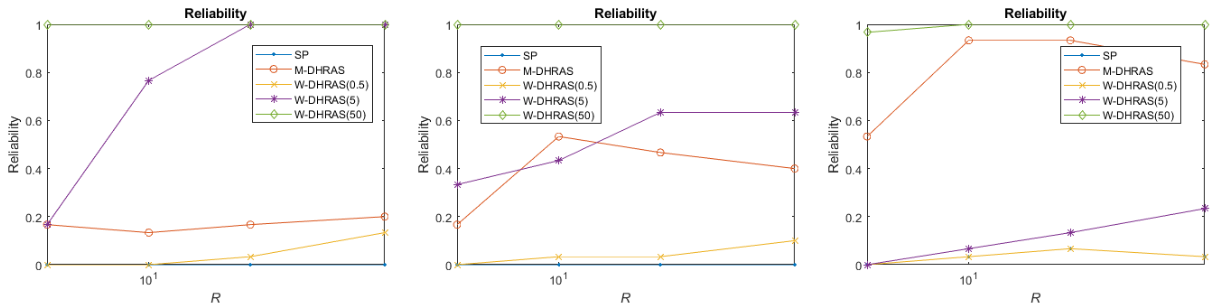
Figure 10: Out-of-sample cost with cost structure (a) and $\lambda = 2$ under Set 4 ($\delta = 0.50$)

with R samples,

$$\text{reliability} = \mathbb{P}_{\boldsymbol{\xi}}^R\Big(v(R) \geq \mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}[f(\boldsymbol{x}(R), \boldsymbol{a}(R), \boldsymbol{\xi})]\Big),$$

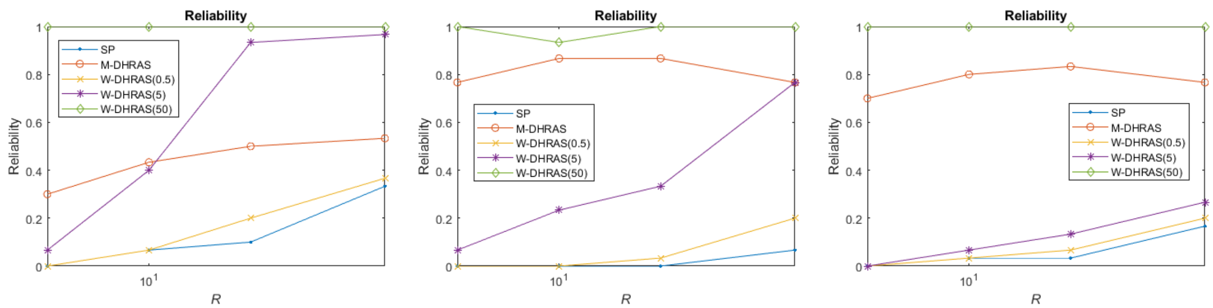
where $\mathbb{P}_{\boldsymbol{\xi}}^R$ is the product measure of R copies of $\mathbb{P}_{\boldsymbol{\xi}}$. The actual cost $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}[f(\boldsymbol{x}(R), \boldsymbol{a}(R), \boldsymbol{\xi})]$ is estimated from 10,000 out-of-sample scenarios and the reliability is computed from 30 instances.

Figures 11 and 12 show the results for the two cost structures $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $(1, 5, 7.5)$ respectively with $\lambda = 2$. We observe that the reliability is, in general, increasing with R. This is reasonable since we have a larger amount of historical data. The reliability of SP is



(a) $N = 6$ (b) $N = 8$ (c) $N = 10$

Figure 11: Reliability under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $\lambda = 2$



(a) $N = 6$ (b) $N = 8$ (c) $N = 10$

Figure 12: Reliability under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ and $\lambda = 2$

the lowest among the three models, which demonstrates that SP with a small number of historical scenarios may not be able to provide a robust cost estimate. On the other hand, the reliability of W-DHRAS(50) consistently gives the best reliability result. This makes sense since $\epsilon = 50$ corresponds to a relatively robust model. It is interesting that although M-DHRAS appears to be the most conservative model from the out-of-sample cost perspective, the reliability result is not the best. Overall, this demonstrates that DRO models have a better reliability compared to SP.

4.6. ***CPU time***

In this section, we analyze the computational time for solving the three models. Specifically, we examine the effect of N (number of customers) and R (number of scenarios) on the solution time for the three models as follows. Table 2 shows the mean CPU time for solving 30 instances using two different cost structures with $\lambda = 2$.

We first observe that the SP model takes the shortest time to solve for all instances and most of them can be solved within one second. Moreover, SP requires a longer solution time with a larger number of scenarios. The M-DHRAS model has a slightly longer solution time than the SP model but, since it only depends on the mean and range of the sample, it has a consistent performance under all values of R. Solution times of the W-DHRAS model are longer than those of the M-DHRAS and SP models. Solution times of the W-DHRAS model increase with R and

$N = 6$	$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$				$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (1, 5, 7.5)$			
CPU Time	$R = 5$	$R = 10$	$R = 20$	$R = 50$	$R = 5$	$R = 10$	$R = 20$	$R = 50$
SP	0.25	0.30	0.37	0.42	0.23	0.21	0.33	0.44
M-DHRAS	0.46	0.48	0.47	0.48	0.46	0.46	0.46	0.46
W-DHRAS(0.5)	2.14	5.22	18.51	102.48	1.97	5.06	19.09	110.49
W-DHRAS(5)	2.27	5.98	20.87	125.00	1.98	5.26	20.40	123.04
W-DHRAS(50)	2.27	5.33	22.76	134.45	2.50	7.05	25.17	175.61
$N = 8$	$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$				$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (1, 5, 7.5)$			
CPU Time	$R = 5$	$R = 10$	$R = 20$	$R = 50$	$R = 5$	$R = 10$	$R = 20$	$R = 50$
SP	0.277	0.334	0.392	0.760	0.28	0.32	0.39	0.79
M-DHRAS	1.11	1.05	1.07	1.10	0.90	0.92	0.91	0.93
W-DHRAS(0.5)	5.99	16.63	90.01	913.99	5.35	15.52	82.45	718.70
W-DHRAS(5)	7.25	20.29	100.80	983.56	5.67	18.24	85.17	782.85
W-DHRAS(50)	9.86	27.03	145.14	1369.67	9.08	25.30	114.55	1161.36
$N = 10$	$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$				$(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (1, 5, 7.5)$			
CPU Time	$R = 5$	$R = 10$	$R = 20$	$R = 50$	$R = 5$	$R = 10$	$R = 20$	$R = 50$
SP	0.376	0.43	0.54	1.04	0.35	0.42	0.54	1.07
M-DHRAS	2.05	2.02	2.11	1.97	1.79	1.83	1.86	1.89
W-DHRAS(0.5)	17.45	56.61	278.79	2105.37	15.23	52.14	344.68	2686.72
W-DHRAS(5)	17.81	61.15	291.45	2327.50	15.64	53.85	375.07	2961.26
W-DHRAS(50)	19.55	70.40	384.39	3384.06	16.57	60.25	467.39	3835.71

Table 2: CPU time for solving the three models are shown and the number in parentheses followed by W-DHRAS model is the choice of ϵ. The reported times are for $\lambda = 2$ with two different cost structures.

varies across three choices of ϵ. We attribute the difference in solution times between the SP and W-DHRAS models to their respective sizes (i.e., the number of variables and constraints). For fixed R, the W-DHRAS model has more variables and constraints than the SP model. As pointed out by previous papers (e.g. Artigues et al., 2015; Klotz and Newman, 2013; Shehadeh et al., 2020), an increase in model size often suggests an increase in solution time for the LP relaxation and, thus, the MILP model's overall solution time. Though W-DHRAS model takes the longest time among the three, it can be solved in a reasonable time (within 3 minutes) for the practical case $N = 6$.

Finally, we compare the computational efficiency of our SP formulation with that of Zhan et al. (2021). Zhan et al. (2021) use the traveling salesperson formulation to determine the operator's route (i.e., they use the binary decision $x_{i,j}$, which takes value 1 if the operator travels from i to j directly, with traveling salesperson constraints). In contrast, we determine operator's route using sequencing variables and constraints. We perform the same experiment in Zhan et al. (2021). The costs c^{u}_j and c^{w}_j are generated from $U[0, 10]$ and $U[0, 5]$ respectively with $\lambda = 2$, $N = 6$ and 1000 scenarios ($c^{\text{o}} = 0$ since no overtime costs are considered in their work). Zhan et al. (2021) assume that the service time is random and travel time is deterministic. Therefore, for a fair comparison, we also assume a deterministic travel time in this experiment, and we use the Euclidean distance to represent the travel time (we refer to Zhan et al., 2021 for detailed experiment description). We generate and solve 30 instances with service time generated from a uniform distribution. Zhan et al. (2021) reported that solving one instance with $N = 6$ takes nearly 2 hours. However, the average computational time under our model formulation is about 26 seconds, which is even better than their proposed L-shaped method. This shows that employing our model formulation could significantly improve the computational efficiency in the routing problem.

5. Conclusion

In this paper, we study an integrated routing and appointment scheduling problem arising from home service practices. We assume both the service and travel times are random and we probably have a few data only. Instead of using the classical SP model that usually leads to poor out-of-sample performances, we propose two DRO models. M-DHRAS is based on the moment ambiguity set with known first moment while W-DHRAS is based on the 1-Wasserstein distance with the empirical distribution. We also establish the asymptotic convergence for the W-DHRAS model. Using duality theories, we reformulate the proposed DRO models into MILPs, which are solvable using off-the-shelf optimization software.

In the extensive numerical study, we investigate the performances of the proposed DRO models. We find that both DRO models achieve superior out-of-sample performances, particularly under distributional uncertainty. Moreover, W-DHRAS could outperform the SP model when the data size is small. We conclude that the proposed DRO models are useful if the distributions of the random service and travel time are unknown and only a limited amount of data is available. The resulting optimal solutions are robust to mis-specifications of the underlying distribution with high reliability. The numerical results also suggest that our SP formulation is more computational efficient than the formulation by Zhan et al. (2021).

Our model can serve as a building block for the following future extensions and areas of research in terms of home service aspects, constraints, and various sources of uncertainties. First, we want to generalize our distributionally robust approach to the case when we have multiple service providers and include decisions such as (1) determining the number of service teams to hire, (2) assigning the hired service teams to customers, (3) constructing routes for the service teams, and (4) determining the customers' appointment times. Second, we aim to model other sources of uncertainty, such as last-minute customer cancelations and operator's cancelation on the day of service. Third, we also aim to incorporate customers' preferences on appointment times.

Appendix A. Proof of Proposition 1

Proof. For a fixed $(\boldsymbol{x}, \boldsymbol{a}) \in \mathcal{X} \times \mathcal{A}$, we can formulate problem (8) as the following linear functional optimization problem.

$$\max_{\mathbb{P} \geq 0} \int_{\mathcal{S}} f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) \, d\mathbb{P} \tag{A.1a}$$

$$\text{s.t.} \int_{\mathcal{S}} d_i \, d\mathbb{P} = \mu_i^{\text{d}}, \qquad \forall i \in [N], \tag{A.1b}$$

$$\int_{\mathcal{S}} t_{i,i'} \, d\mathbb{P} = \mu_{i,i'}^{\text{t}}, \qquad \forall i \in [N], \; i' \in [N], \tag{A.1c}$$

$$\int_{\mathcal{S}} d\mathbb{P} = 1. \tag{A.1d}$$

Letting ρ_i, $\alpha_{i,i'}$, and θ be the dual variables associated with constraints (A.1b), (A.1c) and (A.1d) respectively, we present the dual of problem (A.1):

$$\min_{\boldsymbol{\rho}, \boldsymbol{\alpha}, \theta} \sum_{i=1}^{N} \mu_i^{\text{d}} \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} \mu_{i,i'}^{\text{t}} \alpha_{i,i'} + \theta \tag{A.2a}$$

$$\text{s.t.} \sum_{i=1}^{N} d_i \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} t_{i,i'} \alpha_{i,i'} + \theta \geq f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}), \quad \forall \boldsymbol{\xi} \in \mathcal{S}, \tag{A.2b}$$

where $\boldsymbol{\rho}$, $\boldsymbol{\alpha}$ and θ are unrestricted in sign, and constraint (A.2b) is associated with the primal variable $\mathbb{P}$. Under the standard assumptions that (1) μ_i^{d} lies in the interior of the set $\{\int_{\mathcal{S}} d_i \, d\mathbb{Q} : \mathbb{Q}$ is a probability distribution over $\mathcal{S}\}$, and (2) $\mu_{i,i'}^{\text{t}}$ lies in the interior of the set $\{\int_{\mathcal{S}} t_{i,i'} \, d\mathbb{Q} : \mathbb{Q}$ is a probability distribution over $\mathcal{S}\}$, strong duality holds such that (A.1) and (A.2) equal in optimal objective value (Bertsimas and Popescu, 2005; Shehadeh et al., 2020). Note that for fixed $(\boldsymbol{\rho}, \boldsymbol{\alpha}, \theta)$, constraint (A.2b) is equivalent to

$$\theta \geq \max_{\boldsymbol{\xi} \in \mathcal{S}} \left\{ f(x, a, \boldsymbol{d}, \boldsymbol{t}) + \sum_{i=1}^{N} -d_i \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} -t_{i,i'} \alpha_{i,i'} \right\}.$$

Since we are minimizing θ in (A.2), the dual formulation of (A.1) is equivalent to:

$$\min_{\boldsymbol{\rho}, \boldsymbol{\alpha}} \left\{ \sum_{i=1}^{N} \mu_i^{\text{d}} \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} \mu_{i,i'}^{\text{t}} \alpha_{i,i'} + \max_{\boldsymbol{\xi} \in \mathcal{S}} \left\{ f(x, a, \boldsymbol{d}, \boldsymbol{t}) + \sum_{i=1}^{N} -d_i \rho_i + \sum_{i=1}^{N} \sum_{i'=1}^{N} -t_{i,i'} \alpha_{i,i'} \right\} \right\}.$$

This completes the proof. □

Appendix B. Proof of Proposition 2

Proof. In view of the objective function (10a), we consider all the terms involving y and define the function as in (12).

$$H(\boldsymbol{y}) = -a_1 y_1 + \sum_{j=2}^{N+1} \big(a_{j-1} - a_j) y_j + \sum_{j=2}^{N+1} \sum_{i=1}^{N} d_i x_{i,j-1} y_j + \sum_{i=1}^{N} t_{0,i} x_{i,1} y_1 + \sum_{j=2}^{N} \sum_{i=1}^{N} \sum_{i \neq i'} t_{i,i'} x_{i,j-1} x_{i',j} y_j.$$

We are maximizing $H(\boldsymbol{y})$ over a polyhedral set $\boldsymbol{y} \in \mathcal{Y}$. Note that H is a linear (convex) function in $\mathcal{Y}$ and we are maximizing H over a convex compact set $\mathcal{Y}$. From basic convex analysis, we know that there exists an optimal solution at some extreme point of $\mathcal{Y}$. Recall the definition of $\mathcal{Y}$.

$$\mathcal{Y} = \Big\{ 0 \leq y_{N+1} \leq c^{\mathrm{o}},\ -c^{\mathrm{u}}_j \leq y_j \leq c^{\mathrm{w}}_j + y_{j+1},\ \forall j \in [N] \Big\}$$

We can apply a similar technique in Proposition 2 in Jiang et al. (2019) and Proposition 3 in Shehadeh et al. (2020) to derive an equivalent formulation for the maximization problem. To characterize the extreme points of $\mathcal{Y}$, we introduce dummy variables y_{N+2} with $c^{\mathrm{w}}_{N+1} = c^{\mathrm{o}}$, $c^{\mathrm{u}}_{N+1} = 0$, $c^{\mathrm{w}}_{N+2} = 0$ and $c^{\mathrm{u}}_{N+2} = 0$. Then, we rewrite the set $\mathcal{Y}$ as

$$\mathcal{Y} = \{ \boldsymbol{y} \mid y_{N+2} = -c^{\mathrm{u}}_{N+2},\ -c^{\mathrm{u}}_j \leq y_j \leq c^{\mathrm{w}}_j + y_{j+1},\ j \in [N+1] \}.$$

Note that an extreme point of $\mathcal{Y}$ satisfies (i) $y_{N+2} = -c^{\mathrm{u}}_{N+2} = 0$ and (ii) for $j \in [N+1]$, the variable constraint on y_j is binding either at the lower bound or the upper bound. If y_j is binding at the upper bound, it does depend on y_{j+1} but if it is binding at the lower bound, it just takes the value of $-c^{u}_j$. From this, we can construct a one-to-one correspondence between the extreme points in $\mathcal{Y}$ and partitions of the set $\{1, \ldots, N+2\}$. That is, given an interval $[k, v]_{\mathbb{Z}}$ in $[N+2]$, y_v is binding at the lower bound, i.e. $y_v = -c^{u}_v$, and for $j \in [k, v-1]$, y_j is binding at the upper bound, i.e. $y_j = -c^{\mathrm{u}}_j + \sum_{l=j}^{v-1} c^{\mathrm{w}}_l$. Therefore, for notational simplicity, we define $\pi_{j,v} = -c^{\mathrm{u}}_v + \sum_{l=j}^{v-1} c^{\mathrm{w}}_l$ for $1 \leq j \leq v \leq N+2$. We can reformulate our optimization problem over $\mathcal{Y}$ to optimizing over partitions of $\{1, \ldots, N+2\}$.

To do so, let $b_{k,v}$ be the binary variable with 1 indicating that the interval $[k, v]$ belongs to an element of a partition for $1 \leq k \leq v \leq N+2$. The condition

$$\sum_{k=1}^{j} \sum_{v=j}^{N+2} b_{k,v} = 1,\ \forall j \in [N+2]$$

is equivalent to saying that $\{[k, v]_{\mathbb{Z}} \mid b_{k,v} = 1,\ k \in [N+2],\ v \in [k, N+2]_{\mathbb{Z}}\}$ is a partition of $[N+2]$. With the use of the equality for any extreme point y, we have

$$y_j = \sum_{k=1}^{j} \sum_{v=j}^{N+2} y_j b_{k,v} = \sum_{k=1}^{j} \sum_{v=j}^{N+2} \pi_{j,v} b_{k,v}\ ,$$

and the problem maximizing $H(\boldsymbol{y})$ over $\boldsymbol{y} \in \mathcal{Y}$ is equivalent to the following integer program,

$$\begin{aligned}
\max_{\boldsymbol{b}} \sum_{v=1}^{N+2} -a_1\pi_{1,v}b_{1,v} &+ \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(a_{j-1}-a_j)\pi_{j,v}b_{k,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N} d_i x_{i,j-1}\pi_{j,v}b_{k,v} \\
&+ \sum_{v=1}^{N+2}\sum_{i=1}^{N} t_{0,i}x_{i,1}\pi_{1,v}b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'}x_{i,j-1}x_{i',j}\pi_{j,v}b_{k,v} \qquad \text{(B.1a)} \\
\text{s.t.} \ \sum_{k=1}^{j}\sum_{v=j}^{N+2} b_{k,v} &= 1, \qquad \forall j \in [N+2], \qquad \text{(B.1b)} \\
b_{k,v} &\in \{0,1\}, \quad \forall k \in [N+2], \forall v \in [k, N+2]_{\mathbb{Z}}. \qquad \text{(B.1c)}
\end{aligned}$$

We remark that the last summation in (B.1a) is summed from $j = 2$ up to $N + 1$ instead of N since $x_{i,N+1} = 0$ for all i. This makes the last summation term can behave similarly to the previous terms, which simplifies our discussion.

Now, we consider the problem

$$\max_{\boldsymbol{\xi}\in\mathcal{S}} \left\{ f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi}) - \sum_{i=1}^{N} d_i\rho_i - \sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i'\neq 1} t_{i,i'}\alpha_{i,i'}x_{i,j-1}x_{i',j} - \sum_{i=1}^{N} t_{0,i}\alpha_{0,i}x_{i,1} - \sum_{i=1}^{N} t_{i,0}\alpha_{i,0}x_{i,N} \right\}.$$

With the use of (B.1), we can reformulate it as

$$\begin{aligned}
\max_{\boldsymbol{b}} \sum_{v=1}^{N+2} -a_1\pi_{1,v}b_{1,v} &+ \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(a_{j-1}-a_j)\pi_{j,v}b_{k,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N} d_i x_{i,j-1}\pi_{j,v}b_{k,v} \\
&+ \sum_{v=1}^{N+2}\sum_{i=1}^{N} t_{0,i}x_{i,1}\pi_{1,v}b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'}x_{i,j-1}x_{i',j}\pi_{j,v}b_{k,v} + \lambda\sum_{i=1}^{N} t_{0,i}x_{i,1} \\
&+ \lambda\sum_{i=1}^{N} t_{i,0}x_{i,N} + \lambda\sum_{j=2}^{N+1}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'}x_{i,j-1}x_{i',j} - \sum_{i=1}^{N}\rho_i d_i - \sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'}\alpha_{i,i'}x_{i,j-1}x_{i',j} \\
&- \sum_{i=1}^{N} t_{0,i}\alpha_{0,i}x_{i,1} - \sum_{i=1}^{N} t_{i,0}\alpha_{i,0}x_{i,N} \qquad \text{(B.2a)} \\
\text{s.t.} \ \sum_{k=1}^{j}\sum_{v=j}^{N+2} b_{k,v} &= 1, \qquad \forall j \in [N+2], \qquad \text{(B.2b)} \\
b_{k,v} &\in \{0,1\}, \quad \forall k \in [N+2], \forall v \in [k, N+2]_{\mathbb{Z}}. \qquad \text{(B.2c)}
\end{aligned}$$

By the construction of $\mathcal{S}$, we can take maximum over each d_i and $t_{i,i'}$. To achieve this, we first gather all the terms involving d_i and $t_{i,i'}$ separately and take supremum accordingly. We can reformulate the objective function (B.2a) as

$$\sum_{v=1}^{N+2} -a_1\pi_{1,v}b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(a_{j-1}-a_j)\pi_{j,v}b_{k,v} + \sum_{i=1}^{N}\left[\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(\pi_{j,v}-\rho_i)\right] x_{i,j-1}b_{k,v}d_i$$

$$+\sum_{i=1}^{N}\sum_{v=1}^{N+2}[(\pi_{1,v}+\lambda)-\alpha_{0,i}]x_{i,1}b_{1,v}t_{0,i}+\sum_{i=1}^{N}(\lambda-\alpha_{i,0})x_{i,N}t_{i,0}$$

$$+\sum_{i=1}^{N}\sum_{i'\neq i}\left\{\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\left[(\pi_{j,v}+\lambda)-\alpha_{i,i'}\right]x_{i,j-1}x_{i',j}b_{k,v}\right\}t_{i,i'}, \tag{B.3}$$

where we focus on the third to the last term. We first define $\Delta d_i=\overline{d}_i-\underline{d}_i$ and $\Delta t_{i,i'}=\overline{t}_{i,i'}-\underline{t}_{i,i'}$. Note that

$$\sup_{t_{i,0}\in[\underline{t}_{i,0},\overline{t}_{i,0}],\, i\in[N]}\sum_{i=1}^{N}(\lambda-\alpha_{i,0})x_{i,N}t_{i,0}=\sum_{i=1}^{N}\left[(\lambda-\alpha_{i,0})\underline{t}_{i,0}+\Delta t_{i,0}(\lambda-\alpha_{i,0})^{+}\right]x_{i,N}. \tag{B.4}$$

For the forth term, for any given $\boldsymbol{x}\in\mathcal{X}$ and feasible b, since $x_{i,1}=1$ and $b_{1,v}=1$ for exactly one i and v, say i_1 and v_1, we have

$$\begin{aligned}
&\sup_{t_{0,i}\in[\underline{t}_{0,i},\overline{t}_{0,i}],\, i\in[N]}\sum_{i=1}^{N}\sum_{v=1}^{N+2}x_{i,1}(\pi_{1,v}+\lambda-\alpha_{0,i})b_{1,v}t_{0,i}\\
&=\sup_{t_{0,i_1}\in[\underline{t}_{0,i_1},\overline{t}_{0,i_1}]}(\pi_{1,v_1}+\lambda-\alpha_{0,i_1})t_{0,i_1}\\
&=(\pi_{1,v_1}+\lambda-\alpha_{0,i_1})\underline{t}_{0,i_1}+\Delta t_{0,i}(\pi_{1,v_1}+\lambda-\alpha_{0,i_1})^{+}\\
&=\sum_{i=1}^{N}\sum_{v=1}^{N+2}\left[(\pi_{1,v}+\lambda-\alpha_{0,i})\underline{t}_{0,i}+\Delta t_{0,i}(\pi_{1,v}+\lambda-\alpha_{0,i})^{+}\right]x_{i,1}b_{1,v}.
\end{aligned} \tag{B.5}$$

For the third term, notice that for any given $\boldsymbol{x}\in\mathcal{X}$, i and j have a one-to-one correspondence. That is, for any i, we can identify exactly one $j=j_i$ such that $x_{i,j-1}=1$. Then,

$$\left[\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(\pi_{j,v}-\rho_i)\,x_{i,j-1}b_{k,v}\right]d_i=\sum_{k=1}^{j_i}\sum_{v=j_i}^{N+2}(\pi_{j_i,v}-\rho_i)\,b_{k,v}d_i=(\pi_{j_i,v_i}-\rho_i)d_i.$$

since $b_{k,v}=1$ only for one pair of (k,v), say (k_i,v_i). Hence, we have

$$\begin{aligned}
&\sup_{d_i\in[\underline{d}_i,\overline{d}_i],\, i\in[N]}\sum_{i=1}^{N}\left[\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(\pi_{j,v}-\rho_i)\,x_{i,j-1}b_{k,v}\right]d_i\\
&=\sum_{i=1}^{N}\sup_{d_i\in[\underline{d}_i,\overline{d}_i],\, i\in[N]}\left[\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(\pi_{j,v}-\rho_i)\,x_{i,j-1}b_{k,v}\right]d_i\\
&=\sum_{i=1}^{N}\sup_{d_i\in[\underline{d}_i,\overline{d}_i],\, i\in[N]}(\pi_{j_i,v_i}-\rho_i)d_i\\
&=\sum_{i=1}^{N}\left[(\pi_{j_i,v_i}-\rho_i)\underline{d}_i+\Delta d_i(\pi_{j_i,v_i}-\rho_i)^{+}\right]\\
&=\sum_{i=1}^{N}\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\left[(\pi_{j,v}-\rho_i)\underline{d}_i+\Delta d_i(\pi_{j,v}-\rho_i)^{+}\right]x_{i,j-1}b_{k,v}.
\end{aligned} \tag{B.6}$$

The last term involves the quadratic term $x_{i,j-1}x_{i',j}$. For any given pair (i,i'), there exists at most one j, say j_i, such that $x_{i,j-1}x_{i',j}=1$. If there does not exist such an j_i, the triple summation term is just zero. Otherwise, there exists such an j_i and the last term reads

$$\left\{\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\left[(\pi_{j,v}+\lambda)-\alpha_{i,i'}\right]x_{i,j-1}x_{i',j}b_{k,v}\right\}t_{i,i'}=(\pi_{j_i,v_i}+\lambda)-\alpha_{i,i'}.$$

With the use of these two observations, we have

$$\sup_{t_{i,i'}\in[\underline{t}_{i,i'},\bar{t}_{i,i'}],\,i\in[N],i'\neq i}\sum_{i=1}^{N}\sum_{i'\neq i}\left\{\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\left[(\pi_{j,v}+\lambda)-\alpha_{i,i'}\right]x_{i,j-1}x_{i',j}b_{k,v}\right\}t_{i,i'}$$
$$=\sum_{i=1}^{N}\sum_{i'\neq i}\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\left[(\pi_{j,v}+\lambda-\alpha_{i,i'})\underline{t}_{i,i'}+\Delta t_{i,i'}(\pi_{j,v}+\lambda-\alpha_{i,i'})^{+}\right]x_{i,j-1}x_{i',j}b_{k,v}\tag{B.7}$$

Eventually, we have reformulated the objective function using (B.4), (B.5), (B.6) and (B.7).

Note that from (B.2b), the coefficient matrix is totally unimodular and hence, we can relax the integer constraint (B.2c) to $b_{k,v}\geq 0$. This reformulates (B.2) into an LP in b. One can observe that we put $b_{k,v}$ to the outermost position in all the terms so that we are able to collect the terms with the same $b_{k,v}$ by interchanging the summations. Indeed, for some arbitrary terms z_{ijk}, the triple summation can be written as

$$\sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}z_{jkv}=\sum_{v=2}^{N+2}\sum_{j=2}^{\min(v,N+1)}z_{j1v}+\sum_{k=2}^{N+1}\sum_{v=k}^{N+2}\sum_{j=k}^{\min(v,N+1)}z_{jkv}.$$

Then, we are able to take the dual of the resulting LP. Let β_j be the dual variable associated to the constraint (B.2b) for $j=1,\ldots,N+2$. Thus, we arrive at the following program.

$$\min_{\boldsymbol{\beta}}\ \sum_{j=1}^{N+2}\beta_j+\sum_{i=1}^{N}\left[(\lambda-\alpha_{i,0})\underline{t}_{i,0}+\Delta t_{i,0}(\lambda-\alpha_{i,0})^{+}\right]x_{i,N}\tag{B.8a}$$

$$\text{s.t. }\beta_1\geq -a_1\pi_{1,1}+\sum_{i=1}^{N}\left[\underline{t}_{0,i}(\pi_{1,1}+\lambda-\alpha_{0,i})+\Delta t_{0,i}(\pi_{1,1}+\lambda-\alpha_{0,i})^{+}\right]x_{i,1},\tag{B.8b}$$

$$\begin{aligned}\sum_{j=1}^{v}\beta_j\geq&-a_1\pi_{1,v}+\sum_{j=2}^{\min(v,N+1)}(a_{j-1}-a_j)\pi_{j,v}\\&+\sum_{i=1}^{N}\left[\underline{t}_{0,i}(\pi_{1,v}+\lambda-\alpha_{0,i})+\Delta t_{0,i}(\pi_{1,v}+\lambda-\alpha_{0,i})^{+}\right]x_{i,1}\\&+\sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i}\left[\underline{t}_{i,i'}(\pi_{j,v}+\lambda-\alpha_{i,i'})+\Delta t_{i,i'}(\pi_{j,v}+\lambda-\alpha_{i,i'})^{+}\right]x_{i,j-1}x_{i',j}\\&+\sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\left[\underline{d}_i(\pi_{j,v}-\rho_i)+\Delta d_i(\pi_{j,v}-\rho_i)^{+}\right]x_{i,j-1},\quad\forall v\in[2,N+2]_{\mathbb{Z}},\end{aligned}\tag{B.8c}$$

$$\sum_{j=k}^{v}\beta_j \geq \sum_{j=k}^{\min(v,N+1)}\big(a_{j-1}-a_j\big)\pi_{j,v} + \sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\big[\underline{d}_i(\pi_{j,v}-\rho_i)+\Delta d_i(\pi_{j,v}-\rho_i)^+\big]x_{i,j-1}$$
$$+\sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i}\Big\{(\pi_{j,v}+\lambda-\alpha_{i,i'})\underline{t}_{i,i'}+\Delta t_{i,i'}(\pi_{j,v}+\lambda-\alpha_{i,i'})^+\Big\}x_{i,j-1}x_{i',j}$$
$$\forall j\in[2,N+1]_{\mathbb{Z}},\ v\in[k,N+2]_{\mathbb{Z}}, \tag{B.8d}$$
$$\beta_{N+2}\geq 0. \tag{B.8e}$$

Finally, we can introduce auxiliary variables to replace the terms with $(\cdot)^+$. In particular, we let $\gamma_{0,i,1,v}=(\pi_{1,v}+\lambda-\alpha_{0,i})^+$, $\gamma_{i,i',j,v}=(\pi_{j,v}+\lambda-\alpha_{i,i'})^+$, $\delta_{i,j,v}=(\pi_{j,v}-\rho_i)^+$ and introduce the following constraints.

$$\gamma_{0,i,1,v}\geq 0,\quad \gamma_{0,i,1,v}\geq \pi_{1,v}+\lambda-\alpha_{0,i},\quad \forall i\in[N],\ \forall v\in[N+2] \tag{B.9a}$$
$$\gamma_{i,i',j,v}\geq 0,\quad \gamma_{i,i',j,v}\geq \pi_{j,v}+\lambda-\alpha_{i,i'},\quad \forall i\in[N],\ \forall i'\in[N]\setminus\{i\},$$
$$\forall j\in[2,N+1]_{\mathbb{Z}},\ \forall v\in[j,N+2]_{\mathbb{Z}} \tag{B.9b}$$
$$\delta_{i,j,v}\geq 0,\quad \delta_{i,j,v}\geq \pi_{j,v}-\rho_i,\quad \forall i\in[N],\ \forall j\in[2,N+1]_{\mathbb{Z}},\ \forall v\in[j,N+2]_{\mathbb{Z}} \tag{B.9c}$$

This completes the reformulation. □

Appendix C. Proof of Corollary 3

Proof. In view of Propositions 1 and 2, problem (9) (i.e., the inner problem $\sup_{\mathbb{P}\in\mathcal{F}(\mathcal{S},\mu)}\mathbb{E}_{\mathbb{P}}[f(x,a,\boldsymbol{d})]$ in (7)) is equivalent to

$$\min_{\boldsymbol{\alpha},\boldsymbol{\rho},\boldsymbol{\beta}}\ \sum_{i=1}^{N}\mu_i^{\mathrm{d}}\rho_i+\sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i'\neq i}\mu_{i,i'}^{\mathrm{t}}\alpha_{i,i'}x_{i,j-1}x_{i',j}+\sum_{i=1}^{N}\mu_{i,0}^{\mathrm{t}}\alpha_{i,0}x_{i,N}+\sum_{i=1}^{N}\mu_{0,i}^{\mathrm{t}}\alpha_{0,i}x_{i,1}$$
$$+\sum_{j=1}^{N+2}\beta_j+\sum_{i=1}^{N}\left[(\lambda-\alpha_{i,0})\underline{t}_{i,0}+\Delta t_{i,0}(\lambda-\alpha_{i,0})^+\right]x_{i,N} \tag{C.1a}$$
$$\text{s.t. } (13b)-(13h) \tag{C.1b}$$

Noting that the variables $\alpha_{i,0}$ for $i\in[N]$ appear in the objective function only and it is easy to see that the minimizer is given by λ. Indeed, consider the terms involving $\alpha_{i,0}$, i.e.,

$$\mu_{i,0}^{\mathrm{t}}\alpha_{i,0}x_{i,N}+\left[(\lambda-\alpha_{i,0})\underline{t}_{i,0}+\Delta t_{i,0}(\lambda-\alpha_{i,0})^+\right]x_{i,N}.$$

If $\alpha_{i,0}\geq\lambda$, the term reads $[(\mu_{i,0}^{\mathrm{t}}-\underline{t}_{i,0})\alpha_{i,0}+\lambda\underline{t}_{i,0}]x_{i,N}$, which is non-decreasing in $\alpha_{i,0}$ and if $\alpha_{i,0}\leq\lambda$, the terms read $[(\mu_{i,0}^{\mathrm{t}}-\overline{t}_{i,0})\alpha_{i,0}+\lambda\underline{t}_{i,0}]x_{i,N}$, which is non-increasing in $\alpha_{i,0}$. Plugging in the optimal $\alpha_{i,0}=\lambda$ to the objective function, we arrive at the desired reformulation. □

Appendix D. Proof of Lemma 1 Adapted from Jiang et al. (2019)

Lemma 1. (of Jiang et al. (2019)). *Suppose that $\mathcal{S}$ is non-empty, convex and compact. Then there exist nonegative constants c_1 and c_2 such that, for all $R \geq 1$ and $\beta \in (0, \min\{1, c_1\})$,*

$$\mathbb{P}^R_{\boldsymbol{\xi}}\Big\{W_p(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}}) \leq \epsilon_R(\beta)\Big\} \geq 1 - \beta$$

where $\mathbb{P}^R_{\boldsymbol{\xi}}$ represents the product measure of R copies of $\mathbb{P}_{\boldsymbol{\xi}}$ and $\epsilon_R(\beta) = \left[\frac{\log(c_1\beta^{-1})}{c_2 R}\right]^{\frac{1}{\max\{3p,n\}}}$.

For completeness, we provide the proof of Lemma 1 of Jiang et al. (2019) as detailed in their paper, which is adapted from Theorem 2 in Fournier and Guillin (2015).

Proof. Notice that by the compactness assumption of the support, there exist $\alpha > p$ and $\gamma > 0$ such that $\mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}[\exp\{\gamma||\xi||_p^\alpha\}] < \infty$. By Theorem 2 in Fournier and Guillin (2015), for any $R \geq 1$ and $\epsilon \in (0, \infty)$, there exist positive constants c and C depending only on p, n, α and γ such that

$$\mathbb{P}^R_{\boldsymbol{\xi}}\left(W_p(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}}) \geq \epsilon^{1/p}\right) \leq a(R,\epsilon)\mathbb{1}(\epsilon \leq 1) + b(R,\epsilon),$$

where

$$a(R,\epsilon) = C\begin{cases} \exp\left\{-cR\epsilon^2\right\} & \text{if } p > n/2 \\ \exp\left\{-cR(\epsilon/\log(2+1/\epsilon))^2\right\} & \text{if } p = n/2 \\ \exp\left\{-cR\epsilon^{n/p}\right\} & \text{if } p \in [1, n/2) \end{cases} \tag{D.1}$$

and $b(R,\epsilon) = C\exp\{-CR\epsilon^{\alpha/p}\}\mathbb{1}(\epsilon > 1)$ with n being the dimension of the random vector $\boldsymbol{\xi}$. We first bound the term $a(R,\epsilon)$ for $\epsilon \in (0,1]$. Notice that

$$\left[\frac{\epsilon}{\log(2+1/\epsilon)}\right]^2 \geq \frac{\epsilon^3}{[\log(3)]^2},$$

which immediately gives $\epsilon[\log(2+1/\epsilon)]^2 \leq [\log(3)]^2$. Then, we have

$$a(R,\epsilon) \leq C\exp\left\{-\frac{c}{[\log(3)]^2}R\epsilon^{\max\{3,n/p\}}\right\}.$$

Next, to bound the term $b(R,\epsilon)$, let $\alpha = \max\{3p, n\} > p$. Then, we have

$$b(R,\epsilon) \leq C\exp\left\{-cR\epsilon^{\max\{3,n/p\}}\right\}.$$

To sum up, we arrive at

$$\mathbb{P}^R_{\boldsymbol{\xi}}\left(W_p(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}}) \geq \epsilon^{1/p}\right) \leq c_1\exp\left\{-c_2R\epsilon^{\max\{3,n/p\}}\right\},$$

where $c_1 = C$ and $c_2 = c/[\log(3)]^2$. By equating the right hand side of the last inequality to β, we obtain $\epsilon = [(c_2R)^{-1}\log(c_1\beta^{-1})]^{-\max\{3,n/p\}^{-1}}$. Plugging in this expression to the last inequality gives the desired result. □

Appendix E. Proof of Theorem 1 from Jiang et al. (2019)

Theorem 1. (Asymptotic consistency, adapted from Jiang et al., 2019 and Theorem 3.6 of Mohajerin Esfahani and Kuhn, 2018). *Suppose that the support $\mathcal{S}$ is non-empty, convex and compact. Consider a sequence of confidence levels $\{\beta_R\}_{R\in\mathbb{R}}$ such that $\sum_{R=1}^{\infty}\beta_R<\infty$ and $\lim_{R\to\infty}\epsilon_R(\beta_R)=0$, and let $(\hat{\boldsymbol{x}}(R,\epsilon_R(\beta_R)),\hat{\boldsymbol{a}}(R,\epsilon_R(\beta_R)))$ represent an optimal solution to W-DHRAS with the ambiguity set $\mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}},\epsilon_R(\beta_R))$. Then, $\mathbb{P}^{\infty}_{\boldsymbol{\xi}}$- almost surely we have $\hat{Z}(R,\epsilon_R(\beta_R))\to Z^*$ as $R\to\infty$. In addition, any accumulation points of $\{(\hat{\boldsymbol{x}}(R,\epsilon_R(\beta_R)),\hat{\boldsymbol{a}}(R,\epsilon_R(\beta_R)))\}_{R\in\mathbb{N}}$ is an optimal solution of (4) $\mathbb{P}^{\infty}_{\boldsymbol{\xi}}$-almost surely.*

Proof. Recall the dual of $f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})$.

$$
\begin{aligned}
f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi}) = \max_{\boldsymbol{y}} & \left(\sum_{i=1}^{N} t_{0,i}x_{i,1}-a_1\right)y_1+\sum_{j=2}^{N+1}\left(a_{j-1}-a_j+\sum_{i=1}^{N}d_i x_{i,j-1}+\sum_{i=1}^{N}\sum_{i'\neq i}t_{i,i'}x_{i,j-1}x_{i,j}\right)y_j \\
\text{s.t. } & \boldsymbol{y}\in\mathcal{Y}=\{\boldsymbol{y}:\ y_{N+1}\le c^{\text{o}},\ -c^{\text{u}}_j\le y_j\le c^{\text{w}}_j+y_{j+1},\ \forall j\in[N]\}.
\end{aligned}
$$

For a fixed pair of $(\boldsymbol{x},\boldsymbol{a})\in\mathcal{X}\times\mathcal{A}$, this is an LP in $\boldsymbol{y}$. Note that $\mathcal{Y}$ is bounded, which implies that $f(\boldsymbol{a},\boldsymbol{x},\boldsymbol{\xi})$ is finite. Also, since $\mathcal{X}\times\mathcal{A}$ and $\mathcal{S}$ are bounded sets, the function $f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})$ is bounded on $\mathcal{X}\times\mathcal{A}\times\mathcal{S}$. Immediately, we have $|f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})|\le M(1+\|\boldsymbol{\xi}\|)$, where M is just the upper bound on f. Next, we claim that $f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})$ is continuous on $\mathcal{X}\times\mathcal{A}\times\mathcal{S}$. For simplicity, write the objective function of the dual problem as $b(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})^\top y$, where the i-th entry of $b(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})$ is the coefficient associated to y_i. By fundamental theorem of LP, instead of maximizing over the entire $\mathcal{Y}$, we can maximize over the set of finite extreme points of $\mathcal{Y}$. Hence, $f(\boldsymbol{x},\boldsymbol{a},\boldsymbol{\xi})$ is the maximum of finitely many linear functions and the continuity follows. Finally, the result follows directly from Theorem 3.6 of Mohajerin Esfahani and Kuhn (2018), where the required conditions are verified. ☐

Appendix F. Proof of Theorem 2 in Jiang et al. (2019)

Theorem 2. (Finite-data guarantee, adapted from Jiang et al., 2019 and Theorem 3.5 in Mohajerin Esfahani and Kuhn, 2018). *For any $\beta\in(0,1)$, let $(\hat{\boldsymbol{x}}(R,\epsilon_R(\beta_R)),\hat{\boldsymbol{a}}(R,\epsilon_R(\beta_R)))$ represent an optimal solution of W-DHRAS with ambiguity set $\mathcal{F}_p(\hat{\mathbb{P}}^R_{\boldsymbol{\xi}},\epsilon_R(\beta_R))$. Then,*

$$
\mathbb{P}^R_{\boldsymbol{\xi}}\Big\{\mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}}[f(\hat{\boldsymbol{x}}(R,\epsilon_R(\beta_R)),\hat{\boldsymbol{a}}(R,\epsilon_R(\beta_R)),\boldsymbol{\xi})])\le\hat{Z}(R,\epsilon_R(\beta_R))\Big\}\ge 1-\beta.
$$

Proof. (adapted from Jiang et al., 2019). By the assumption on support $\mathcal{S}$ and Lemma 1, all conditions of Theorem 3.5 in Mohajerin Esfahani and Kuhn (2018) are satisfied. Therefore, the conclusions of Theorem 2 hold valid. ☐

Appendix G. Proof of Proposition 4

Proof. Recall that $\hat{\mathbb{P}}^R_{\boldsymbol{\xi}} = \frac{1}{R}\sum_{r=1}^R \delta_{\hat{\boldsymbol{\xi}}^r}$. For any $\mathbb{P}_{\boldsymbol{\xi}} \in \mathcal{P}(\mathcal{S})$, we can rewrite the joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}})$ by the conditional distribution of $\boldsymbol{\xi}$ given $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}^r$ for $r = 1, \ldots, R$, denoted as $\mathbb{Q}^r_{\boldsymbol{\xi}}$. That is, $\Pi = \frac{1}{R}\sum_{r=1}^R \mathbb{Q}^r_{\boldsymbol{\xi}}$. Notice that if we find one joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}})$ such that $\int ||\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}||_1 d\Pi \le \epsilon$, then $W_1(\mathbb{P}_{\boldsymbol{\xi}}, \hat{\mathbb{P}}^R_{\boldsymbol{\xi}}) \le \epsilon$. Hence, we can drop the infimum operator in Wasserstein distance and arrive at the following equivalent problem

$$\sup_{\mathbb{Q}^r_{\boldsymbol{\xi}} \in \mathcal{P}(\mathcal{S}), r\in[R]} \frac{1}{R}\sum_{r=1}^R \int_{\mathcal{S}} f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) d\mathbb{Q}^r_{\boldsymbol{\xi}} \tag{G.1a}$$

$$\text{s.t.} \quad \frac{1}{R}\sum_{r=1}^R \int_{\mathcal{S}} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 d\mathbb{Q}^r_{\boldsymbol{\xi}} \le \epsilon. \tag{G.1b}$$

Using a standard strong duality argument, we can reformulate the problem by its dual, i.e.,

$$\begin{aligned}
&\inf_{\rho \ge 0} \sup_{\mathbb{Q}^r_{\boldsymbol{\xi}} \in \mathcal{P}(\mathcal{S}), r\in[R]} \left\{ \frac{1}{R}\sum_{r=1}^R \int_{\mathcal{S}} f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) d\mathbb{Q}^r_{\boldsymbol{\xi}} + \rho \left[\epsilon - \frac{1}{R}\sum_{r=1}^R \int_{\mathcal{S}} \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 d\mathbb{Q}^r_{\boldsymbol{\xi}} \right] \right\} \\
&= \inf_{\rho\ge 0} \left\{ \epsilon\rho + \frac{1}{R}\sum_{r=1}^R \sup_{\mathbb{Q}^r_{\boldsymbol{\xi}} \in \mathcal{P}(\mathcal{S})} \int_{\mathcal{S}} \left[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 \right] d\mathbb{Q}^r_{\boldsymbol{\xi}} \right\} \\
&= \inf_{\rho\ge 0} \left\{ \epsilon\rho + \frac{1}{R}\sum_{r=1}^R \sup_{\boldsymbol{\xi}\in\mathcal{S}} \left\{ f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho \|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1 \right\} \right\}.
\end{aligned} \tag{G.2}$$

This completes the proof. □

Appendix H. Proof of Proposition 5

Proof. With reference to (B.1), we can reformulate $\max_{\boldsymbol{\xi}\in\mathcal{S}}\{f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi}) - \rho\|\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}^r\|_1\}$ as

$$\begin{aligned}
\max \; & \sum_{v=1}^{N+2} -a_1 \pi_{1,v} b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} (a_{j-1} - a_j)\pi_{j,v} b_{k,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N} d_i x_{i,j-1}\pi_{j,v} b_{k,v} \\
& + \sum_{v=1}^{N+2}\sum_{i=1}^{N} t_{0,i}\pi_{1,v} x_{i,1} b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'}\pi_{j,v} x_{i,j-1} x_{i',j} b_{k,v} \\
& + \lambda\sum_{i=1}^{N} t_{0,i} x_{i,1} + \lambda \sum_{i=1}^{N} t_{i,0} x_{i,N} + \lambda \sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i'\neq i} t_{i,i'} x_{i,j-1} x_{i',j} - \rho \sum_{j=2}^{N+1}\sum_{i=1}^{N} |d_i - \hat{d}^r_i| x_{i,j-1} \\
& - \rho \sum_{j=2}^{N+1}\sum_{i=1}^{N}\sum_{i'\neq i} |t_{i,i'} - \hat{t}^r_{i,i'}| x_{i,j-1} x_{i',j} - \rho \sum_{i=1}^{N} |t_{0,i} - \hat{t}^r_{0,i}| x_{i,1} - \rho \sum_{i=1}^{N} |t_{i,0} - \hat{t}^r_{i,0}| x_{i,N}
\end{aligned} \tag{H.1a}$$

$$\text{s.t.} \; \sum_{k=1}^{j}\sum_{v=j}^{N+2} b_{k,v} = 1, \qquad \forall j \in [N+2]. \tag{H.1b}$$

$$b_{k,v} \in \{0,1\}, \quad \forall k \in [N], \forall v \in [k, N+2]_{\mathbb{Z}}. \tag{H.1c}$$

As in the proof of Proposition 2, we group terms with the same $t_{i,i'}$ and d_i and find an explicit formula for their supremum.

First, we define the following quantities for $r \in [R]$.

$$u^r_{i,0} = \max\{\lambda \hat{t}^r_{i,0}, \lambda \bar{t}_{i,0} - \rho(\bar{t}_{i,0} - \hat{t}^r_{i,0})\} \tag{H.2a}$$

$$u^r_{0,i,1,v} = \max\{(\pi_{1,v} + \lambda)\underline{t}_{0,i} - \rho(\hat{t}^r_{0,i} - \underline{t}_{0,i}), (\pi_{1,v} + \lambda)\hat{t}^r_{0,i}, (\pi_{1,v} + \lambda)\bar{t}_{0,i} - \rho(\bar{t}_{0,i} - \hat{t}^r_{0,i})\} \tag{H.2b}$$

$$u^r_{i,i',j,v} = \max\{(\pi_{j,v} + \lambda)\underline{t}_{i,i'} - \rho(\hat{t}^r_{i,i'} - \underline{t}_{i,i'}), (\pi_{j,v} + \lambda)\hat{t}^r_{i,i'}, (\pi_{j,v} + \lambda)\bar{t}_{i,i'} - \rho(\bar{t}_{i,i'} - \hat{t}^r_{i,i'})\} \tag{H.2c}$$

$$\nu^r_{i,j,v} = \max\{\pi_{j,v}\underline{d}_i - \rho(\hat{d}^r_i - \underline{d}_i), \pi_{j,v}\hat{d}^r_i, \pi_{j,v}\bar{d}_i - \rho(\bar{d}_i - \hat{d}^r_i)\} \tag{H.2d}$$

Note that for the terms with $t_{i,0}$,

$$\sup_{t_{i,0} \in [\underline{t}_{i,0}, \bar{t}_{i,0}],\, i \in [N]} \sum_{i=1}^{N} \left[\lambda t_{i,0} - \rho|t_{i,0} - \hat{t}^r_{i,0}|\right] x_{i,N} = \sum_{i=1}^{N} \sup_{t_{i,0} \in [\underline{t}_{i,0}, \bar{t}_{i,0}]} \left[\lambda t_{i,0} - \rho|t_{i,0} - \hat{t}^r_{i,0}|\right] x_{i,N}. \tag{H.3}$$

Since $t_{i,0} - \rho|t_{i,0} - \hat{t}^r_{i,0}|$ is a piecewise linear function on the intervals $[\underline{t}_{i,0}, \hat{t}^r_{i,0}]$ and $[\hat{t}^r_{i,0}, \bar{t}_{i,0}]$, and it is strictly increasing on $[\underline{t}_{i,0}, \hat{t}^r_{i,0}]$, the maximum is attained either at $\hat{t}^r_{i,0}$ or $\bar{t}_{i,0}$. That is,

$$\sup_{t_{i,0} \in [\underline{t}_{i,0}, \bar{t}_{i,0}]} \left[\lambda t_{i,0} - \rho|t_{i,0} - \hat{t}^r_{i,0}|\right] = u^r_{i,0}. \tag{H.4}$$

For the terms with $t_{0,i}$, we have

$$\begin{aligned}
&\sup_{t_{0,i} \in [\underline{t}_{0,i}, \bar{t}_{0,i}],\, i \in [N]} \sum_{i=1}^{N} \sum_{v=1}^{N+2} \left[(\pi_{1,v} + \lambda)t_{0,i} - \rho|t_{0,i} - \hat{t}^r_{0,i}|\right] b_{i,v} x_{i,1} \\
&= \sum_{i=1}^{N} \sum_{v=1}^{N+2} \sup_{t_{0,i} \in [\underline{t}_{0,i}, \bar{t}_{0,i}]} \left[(\pi_{1,v} + \lambda)t_{0,i} - \rho|t_{0,i} - \hat{t}^r_{0,i}|\right] b_{i,v} x_{i,1} \\
&= \sum_{i=1}^{N} \sum_{v=1}^{N+2} u^r_{0,i,1,v} b_{i,v} x_{i,1},
\end{aligned} \tag{H.5}$$

where the last equality follows from the fact that $(\pi_{1,v} + \lambda)t_{0,i} - \rho|t_{0,i} - \hat{t}^r_{0,i}|$ is a piecewise linear function on the interval $[\underline{t}_{0,i}, \hat{t}^r_{0,i}]$ and $[\hat{t}^r_{0,i}, \bar{t}_{0,i}]$ and the maximum is attained when $t_{0,i}$ takes any one of the values from $\underline{t}_{0,i}$, $\hat{t}^r_{0,i}$ and $\bar{t}_{0,i}$. Using the same argument, we obtain the expressions for the terms with $t_{i,i'}$ and d_i as follows.

$$\begin{aligned}
&\sup_{t_{i,i'} \in [\underline{t}_{i,i'}, \bar{t}_{i,i'}], i \in [N], i' \neq i} \sum_{i=1}^{N} \sum_{i' \neq i} \sum_{j=2}^{N+1} \sum_{k=1}^{j} \sum_{v=j}^{N+2} \left[(\pi_{j,v} + \lambda)t_{i,i'} - \rho|t_{i,i'} - \hat{t}^r_{i,i'}|\right] x_{i,j-1} x_{i',j} b_{k,v} \\
&= \sum_{i=1}^{N} \sum_{i' \neq i} \sum_{j=2}^{N+1} \sum_{k=1}^{j} \sum_{v=j}^{N+2} u^r_{i,i',j,v} x_{i,j-1} x_{i',j} b_{k,v}
\end{aligned} \tag{H.6}$$

$$\sup_{d_i \in [\underline{d}_i, \bar{d}_i], i \in [N]} \sum_{i=1}^{N} \sum_{j=2}^{N+1} \sum_{k=1}^{j} \sum_{v=j}^{N+2} \left[\pi_{j,v} d_i - \rho|d_i - \hat{d}^r_i|\right] x_{i,j-1} b_{k,v}$$

$$= \sum_{i=1}^{N} \sum_{j=2}^{N+1} \sum_{k=1}^{j} \sum_{v=j}^{N+2} \nu^r_{i,j,v} x_{i,j-1} b_{k,v} \tag{H.7}$$

With the use of (H.3) to (H.7), we arrive at the following integer program.

$$\begin{aligned}
\max_{\boldsymbol{b}} \sum_{v=1}^{N+2} -a_1\pi_{1,v}b_{1,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}(a_{j-1}-a_j)\pi_{j,v}b_{k,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N}\nu^r_{i,j,v}x_{i,j-1}b_{k,v} \\
+ \sum_{v=1}^{N+2}\sum_{i=1}^{N} u^r_{0,i,1,v}x_{i,1}b_{i,v} + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2}\sum_{i=1}^{N}\sum_{i'\neq i} u^r_{i,i',j,v}x_{i,j-1}x_{i',j}b_{k,v} + \sum_{i=1}^{N} u^r_{i,0}x_{i,N} \qquad \text{(H.8a)}
\end{aligned}$$

$$\text{s.t.} \sum_{k=1}^{j}\sum_{v=j}^{N+2} b_{k,v} = 1, \qquad \forall j \in [N+2], \tag{H.8b}$$

$$b_{k,v} \in \{0,1\}, \quad \forall k \in [N], \forall v \in [k, N+2]_{\mathbb{Z}}. \tag{H.8c}$$

Since the coefficient matrix associated to the constraints (H.8b) is totally unimodular, we can relax the integral constraints on $b_{k,b}$ by $b_{k,v} \geq 0$. Taking the dual of the resulting LP in (H.8) with β^r_j being the dual variable associated to the constraint (H.8b), we arrive at the following LP.

$$\min_{\boldsymbol{\beta}} \sum_{j=1}^{N+2} \beta^r_j + \sum_{i=1}^{N} u^r_{i,0}x_{i,N} \tag{H.9a}$$

$$\text{s.t. } \beta^r_1 \geq -a_1\pi_{1,1} + \sum_{i=1}^{N} u^r_{0,i,1,1}x_{i,1}, \tag{H.9b}$$

$$\begin{aligned}
\sum_{j=1}^{v}\beta^r_j \geq -a_1\pi_{1,v} + \sum_{j=2}^{\min(v,N+1)}(a_{j-1}-a_j)\pi_{j,v} + \sum_{i=1}^{N} u^r_{0,i,1,v}x_{i,1} \\
+ \sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i} u^r_{i,i',j,v}x_{i,j-1}x_{i',j} + \sum_{j=2}^{\min(v,N+1)}\sum_{i=1}^{N}\nu^r_{i,j,v}x_{i,j-1}, \quad \forall v \in [2, N+2]_{\mathbb{Z}},
\end{aligned} \tag{H.9c}$$

$$\begin{aligned}
\sum_{j=k}^{v}\beta^r_j \geq \sum_{j=k}^{\min(v,N+1)}(a_{j-1}-a_j)\pi_{j,v} + \sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\nu^r_{i,j,v}x_{i,j-1} \\
+ \sum_{j=k}^{\min(v,N+1)}\sum_{i=1}^{N}\sum_{i'\neq i} u_{i,i',j,v}x_{i,j-1}x_{i',j}, \quad \forall k \in [2, N+1]_{\mathbb{Z}}, \forall v \in [k, N+2]_{\mathbb{Z}},
\end{aligned} \tag{H.9d}$$

$$\beta^r_{N+2} \geq 0. \tag{H.9e}$$

Finally, we can treat $\boldsymbol{u}^r$ and $\boldsymbol{\nu}^r$ as variables and introduce the following constraints.

$$u^r_{i,0} \geq \lambda\hat{t}^r_{i,0}, \quad u^r_{i,0} \geq \lambda\bar{t}_{i,0} - \rho(\bar{t}_{i,0} - \hat{t}^r_{i,0}), \quad \forall i \in [N] \tag{H.10a}$$

$$\begin{aligned}
u^r_{0,i,1,v} \geq (\pi_{1,v}+\lambda)\underline{t}_{0,i} - \rho(\hat{t}^r_{0,i} - \underline{t}_{0,i}), \quad u^r_{0,i,1,v} \geq (\pi_{1,v}+\lambda)\hat{t}^r_{0,i}, \\
u^r_{0,i,1,v} \geq (\pi_{1,v}+\lambda)\bar{t}_{0,i} - \rho(\bar{t}_{0,i} - \hat{t}^r_{0,i}), \quad \forall i \in [N],\, v \in [N+2]
\end{aligned} \tag{H.10b}$$

$$u^r_{i,i',j,v} \geq (\pi_{j,v}+\lambda)\underline{t}_{i,i'} - \rho(\hat{t}^r_{i,i'} - \underline{t}_{i,i'}), \quad u^r_{i,i',j,v} \geq (\pi_{j,v}+\lambda)\overline{t}_{i,i'} - \rho(\overline{t}_{i,i'} - \hat{t}^r_{i,i'}),$$

$$u^r_{i,i',j,v} \geq (\pi_{j,v}+\lambda)\hat{t}^r_{i,i'}, \quad \forall i \in [N],\ i' \in [N]\setminus\{i\},\ j \in [2, N+1]_{\mathbb{Z}},\ v \in [j, N+2]_{\mathbb{Z}} \tag{H.10c}$$

$$\nu^r_{i,j,v} \geq \pi_{j,v}\underline{d}_i - \rho(\hat{d}^r_i - \underline{d}_i), \quad \nu^r_{i,j,v} \geq \pi_{j,v}\overline{d}_i - \rho(\overline{d}_i - \hat{d}^r_i),$$

$$\nu^r_{i,j,v} \geq \pi_{j,v}\hat{d}^r_i, \quad \forall i \in [N],\ j \in [2, N+1]_{\mathbb{Z}},\ v \in [j, N+2]_{\mathbb{Z}} \tag{H.10d}$$

This completes the proof. □

Appendix I. Tight McCormick

For the model (16), we can derive upper and lower bounds for the values of $\boldsymbol{\rho}$ and $\boldsymbol{\alpha}$.

Proposition 6. *Let $P^u_1 = \max_{j\in[2,N+1]_{\mathbb{Z}},\, v\in[j,N+2]_{\mathbb{Z}}} \pi_{j,v}$ and $P^l_1 = \min_{j\in[2,N+1]_{\mathbb{Z}},\, v\in[j,N+2]_{\mathbb{Z}}} \pi_{j,v}$. Also, let $P^u_2 = \max_{v\in[N+2]} \pi_{1,v}$ and $P^l_2 = \min_{v\in[N+2]} \pi_{1,v}$. Then, the lower bounds are $\underline{\rho}_i = P^l_1$, $\underline{\alpha}_{i,i'} = P^l_1 - \lambda$ and $\underline{\alpha}_{0,i} = P^l_2 - \lambda$ while the upper bounds are $\overline{\rho}_i = P^u_1$, $\overline{\alpha}_{i,i'} = P^u_1 + \lambda$ and $\overline{\alpha}_{0,i} = P^u_2 + \lambda$.*

Proof. We only present the case for ρ_i and the remaining two cases are similar. From the model in Proposition 1 and its reformulation (B.2) with (B.4) to (B.7), the terms involving ρ_i can be summarized as

$$\mu^{\mathrm{d}}_i \rho_i + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} \left[(\pi_{j,v} - \rho_i)\underline{d}_i + \Delta d_i(\pi_{j,v} - \rho_i)^+\right] x_{i,j-1} b_{k,v}.$$

If $\rho_i \geq P^u_1$, we can simplify the terms as

$$\mu^{\mathrm{d}}_i \rho_i + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} (\pi_{j,v} - \rho_i)\underline{d}_i x_{i,j-1} b_{k,v} = (\mu^{\mathrm{d}}_i - \underline{d}_i)\rho_i + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} x_{i,j-1}\pi_{j,v} b_{k,v},$$

which is non-decreasing in ρ_i since $\mu^{\mathrm{d}}_i - \underline{d}_i \geq 0$. Similarly, if $\rho_i \leq P^l_1$, then the terms read

$$\mu^{\mathrm{d}}_i \rho_i + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} (\pi_{j,v} - \rho_i)\overline{d}_i x_{i,j-1} b_{k,v} = (\mu^{\mathrm{d}}_i - \overline{d}_i)\rho_i + \sum_{j=2}^{N+1}\sum_{k=1}^{j}\sum_{v=j}^{N+2} x_{i,j-1}\pi_{j,v} b_{k,v},$$

which is non-increasing in ρ_i since $\mu^{\mathrm{d}}_i - \overline{d}_i \leq 0$. This shows that there exists a minimizer over ρ_i lying in $[P^l_1, P^u_1]$ and we can set $\underline{\rho}_i = P^l_1$ and $\overline{\rho}_i = P^u_1$. □

Corollary 7. *We have lower bounds $\underline{\delta}_{i,j,v} = 0$, $\underline{\gamma}_{i,i',j,v} = 0$ and $\underline{\gamma}_{0,i,1,v} = 0$ with upper bounds $\overline{\delta}_{i,j,v} = \pi_{j,v} - P^l_1$, $\overline{\gamma}_{i,i',j,v} = \pi_{j,v} + 2\lambda - P^l_1$ and $\overline{\gamma}_{0,i,1,v} = \pi_{1,v} + 2\lambda - P^l_2]$.*

Proof. We only present the case for $\delta_{i,j,v}$ and the remaining two cases are similar. Recall $\delta_{i,j,v} = (\pi_{j,v} - \rho_i)^+$ and from Proposition 6, we have

$$\pi_{j,v} - P^u_1 \leq \pi_{j,v} - \rho_i \leq \pi_{j,v} - P^l_1\,,$$

where the lower bound is no greater than 0 and the upper bound is no less than 0. □

Next, we consider the model (26). From the proof of Proposition 5 in Appendix H, we have the explicit formula for the quantities $\boldsymbol{u}^r$ and $\boldsymbol{\nu}^r$ in (H.2). Immediately, the lower bounds are $\underline{u}^r_{i,0} = \lambda \hat{t}^r_{i,0}$, $\underline{u}^r_{0,i,1,v} = (\pi_{1,v} + \lambda)\hat{t}^r_{0,i}$, $\underline{u}^r_{i,i',j,v} = (\pi_{j,v} + \lambda)\hat{t}^r_{i,i'}$ and $\underline{\nu}^r_{i,j,v} = \pi_{j,v}\hat{d}^r_i$. Since the expressions in (H.2) are non-increasing in ρ and we have $\rho \geq 0$, the upper bounds are simply $\overline{u}^r_{i,0} = \lambda \overline{t}_{i,0}$, $\overline{u}^r_{0,i,1,v} = \max\{(\pi_{1,v} + \lambda)\underline{t}_{0,i}, (\pi_{1,v} + \lambda)\overline{t}_{0,i}\}$, $\overline{u}^r_{i,i',j,v} = \max\{(\pi_{j,v} + \lambda)\underline{t}_{i,i'}, (\pi_{j,v} + \lambda)\overline{t}_{i,i'}\}$ and $\overline{\nu}^r_{i,j,v} = \max\{\pi_{j,v}\underline{d}_i, \pi_{j,v}\overline{d}_i\}$.

Appendix J. Sample averaging approximation approach

In the numerical experiments, we compare DRO models with the sample averaging approximation (SAA) approach. Suppose we have a set of R scenarios $\{\hat{\boldsymbol{\xi}}^1, \ldots, \hat{\boldsymbol{\xi}}^R\}$. The SAA approach is to solve the original problem by replacing the true distribution with the empirical distribution $\widehat{\mathbb{P}}^R_{\boldsymbol{\xi}}$. That is,

$$
\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{a},\boldsymbol{u},\boldsymbol{w}} \ & \frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{N}\left[(c^{\mathrm{w}}_j w^r_j + c^{\mathrm{u}}_j u^r_j) + c^{\mathrm{o}} w^r_{N+1} + \lambda A^r\right] \\
\text{s.t. } & \boldsymbol{x} \in \mathcal{X}, \quad \boldsymbol{a} \in \mathcal{A}, \\
& w^r_1 - u^r_1 = \sum_{i=1}^{N} t^r_{0,i} x_{i,1} - a_1, \\
& w^r_j - w^r_{j-1} - u^r_j = a_{j-1} - a_j + \sum_{i=1}^{N} d^r_i x_{i,j-1} + \sum_{i=1}^{N}\sum_{i'\neq i} t^r_{i,i'} x_{i,j-1} x_{i,j}, \ \forall j \in [2, N], \\
& w^r_{N+1} - w^r_N - u^r_{N+1} = a_N - a_{N+1} + \sum_{i=1}^{N} d^r_i x_{i,N}, \\
& A^r = \sum_{j=2}^{N}\sum_{i=1}^{N}\sum_{i'\neq i} t^r_{i,i'} x_{i,j-1} x_{i',j} + \sum_{i=1}^{N}(t^r_{0,i} x_{i,1} + t^r_{i,0} x_{i,N}), \\
& (w^r_j, u^r_j) \geq 0, \quad \forall j \in [N+1].
\end{aligned}
$$

Appendix K. Symmetry-breaking constraints

To enhance the tractability of the models, we introduce the symmetry-breaking constraints. The idea behind these constraints follows from prior appointment scheduling observations that customers of the same type (i.e., requesting the same service) have the same service time distribution and the observation from Nikzad et al. (2021) that customers within a service region form a basic unit or cluster that shares the same travel time distribution. In this case, the route within the same group does not matter since the service and travel time distributions are the same. This means, with the presence of homogeneous groups, we could be able to eliminate some of the equialent routes and hopefully, improve the model's tractability. In particular, we focus on the presence of one homogeneous group, which is common when services are provided within a service region.

Suppose that except the depot (node 0), the N customers form a homogeneous group. That is, d_i are distributionally the same over $i \in [N]$ and $t_{i,i'}$ are distributionally the same over $i \in [N]$ and $i' \in [N] \setminus \{i\}$. In this case, we argue that the service provider only needs to make decision on the nodes departing from and entering to the depot. We summarize this result formally in the following lemma.

Lemma 8. *Assume that there is only one homogeneous group. For a fixed route $(i_1, \dots, i_N)$ denoted as $\boldsymbol{x}$, i.e. $x_{i_j,j} = 1$ for $j \in [N]$, define $\boldsymbol{x}^\Pi$ by another route with $x_{\Pi(i_j),j}$, where $(\Pi(i_j))_{j\in[N]}$ is a permutation of $[N]$ with $\Pi(i_1) = i_1$ and $\Pi(i_N) = i_N$. Then, $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ is distributionally the same as $f(\boldsymbol{x}^\Pi, \boldsymbol{a}, \boldsymbol{\xi})$ for any $\boldsymbol{a} \in \mathcal{A}$.*

Proof. Recall that $f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})$ is the second-stage cost, which is a function of the idling time u_i for $i \in [N]$, the waiting time and overtime w_i for $i \in [N+1]$. Define the same quantities u_i^Π and w_i^Π for the decision $\boldsymbol{x}^\Pi$. We will show, by induction, that w_j and u_j are distributionally the same as w_j^Π and u_j^Π respectively. Note that w_1 and u_1 are distributionally the same as w_1^Π and u_1^Π respectively since by assumption, the first visiting customer is the same. Next, assume that w_{j-1} and u_{j-1} are distributionally the same as w_{j-1}^Π and u_{j-1}^Π respectively. Note that

$$w_j = \max\left\{ w_{j-1} + a_{j-1} - a_j + \sum_{i=1}^{N} d_i x_{i,j-1} + \sum_{i=1}^{N} \sum_{i' \neq i} t_{i,i'} x_{i,j-1} x_{i',j}, 0 \right\}.$$

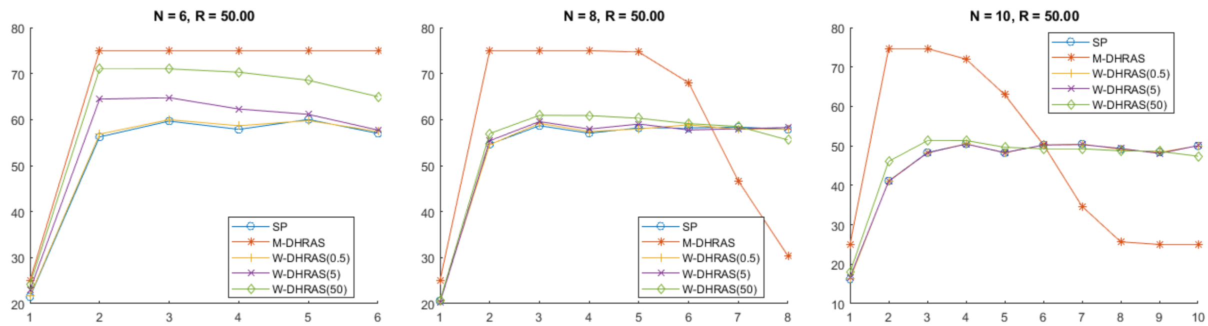
By induction assumption and the assumption that the distributions of travel and service times are the same, w_j equals w_j^Π in distribution. It is easy to see that, with the same argument, u_j equals u_j^Π in distribution. This completes the proof. □

From Lemma 8, we have $\mathbb{E}[f(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{\xi})] = \mathbb{E}[f(\boldsymbol{x}^\Pi, \boldsymbol{a}, \boldsymbol{\xi})]$ for any permutation with fixed entering and departing node. As an illustration of the symmetry, consider that we have a homogeneous group of 6 customers. Then, the routes $(2, 1, 3, 4, 6, 5)$ and $(2, 4, 3, 6, 1, 5)$ give the same expected cost. We can impose the lexicographic order of the route for the middle customers (not the first and the last customer), i.e., the service provider will first visit customers with a smaller index i. In this example, we only consider the route $(2, 1, 3, 4, 6, 5)$ and eliminate any other possible permutations of the middle customers. Hence, we can introduce the following symmetry-breaking constraints.

$$x_{1,j} \geq x_{1,j+1}, \quad \forall j \in [2, N-2]_{\mathbb{Z}} \tag{K.1a}$$

$$x_{i,j} \leq \sum_{l=1}^{i-1} x_{l,j-1}, \quad \forall i \in [2, N]_{\mathbb{Z}},\ j \in [3, N-1]_{\mathbb{Z}} \tag{K.1b}$$

If customer 1 is neither the first nor the last customer, constraints (K.1a) enforces customer 1 to be the second customer (see example matrix X^1, where each entry represents $x_{i,j}$). Otherwise, the



(a) $N = 6$ (b) $N = 8$ (c) $N = 10$

Figure L.13: Mean inter-arrival times with $R = 50$ under $(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$ and $\lambda = 0.5$

constraint is satisfied since $x_{1,j} = 0$ for all $j \in [2, N-1]_{\mathbb{Z}}$ (see X^2 and X^3). Constraints (K.1b) enforces the ordering of the third to the $(N-1)$-th customer. If customer i is the j-th customer, i.e. $x_{i,j} = 1$ for some $j \in [3, N-1]_{\mathbb{Z}}$, we require that a customer with index less than i must be served at position $j-1$ (see X^1, X^2 and X^3). By imposing these constraints, the solver solely determines the entering and departing node without optimizing any permutations for the middle customers.

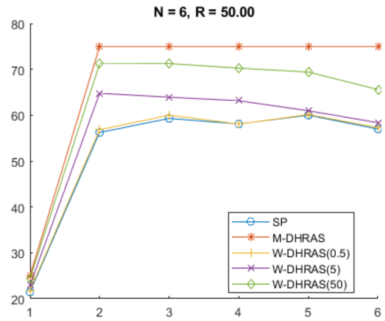
$$X^1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad X^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad X^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Appendix L. Additional results on appointment time structure

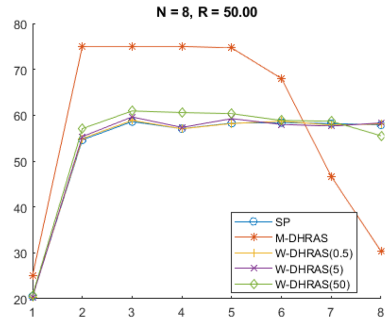
In this section, we provide additional results for the appointment time structure with the two other choices of λ, namely 0.5 and 1. Figures L.13 and L.14 show the inter-arrival times for the cost structure $(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (2, 1, 20)$ while Figures L.15 and L.16 show the results for the cost structure $(c^{\text{w}}_j, c^{\text{u}}_j, c^{\text{o}}) = (5, 1, 7.5)$. We observe similar patterns for the three choices of λ.

Appendix M. Additional results on out-of-sample performance

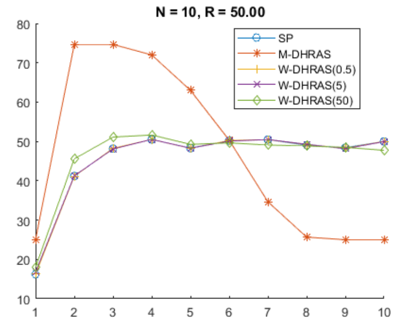
We provide additional out-of-sample performance results in this appendix. All the settings are the same as described in Section 4 and we only present the results for $\lambda = 2$ since similar patterns could be observed for other two choices of λ. We show the results of the 9 out-of-sample testing sets in each figure in the following order: (Row 1) Set 1, Set 2, Set 3, (Row 2) Set 4 with



(a) $N = 6$

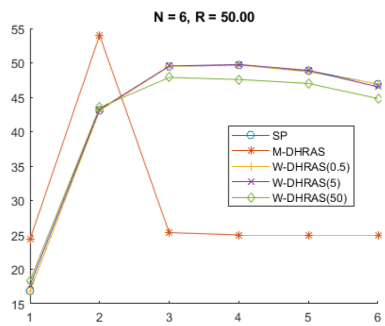


(b) $N = 8$

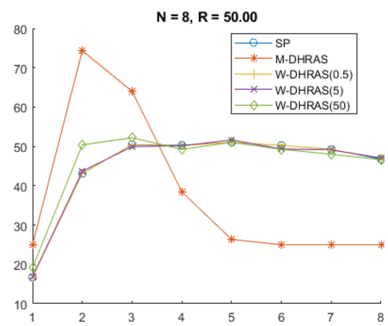


(c) $N = 10$

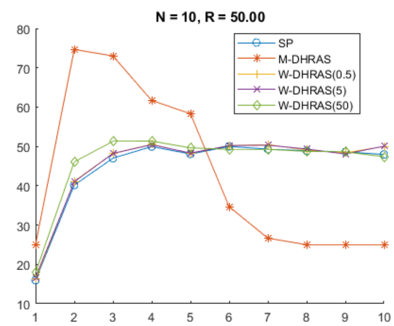
Figure L.14: Mean inter-arrival times with $R = 50$ under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (2, 1, 20)$ and $\lambda = 1$



(a) $N = 6$

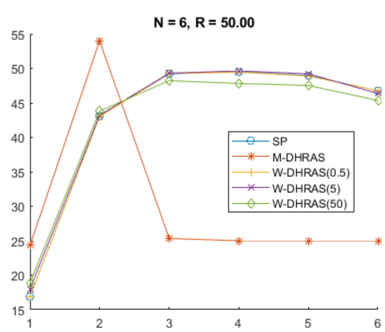


(b) $N = 8$

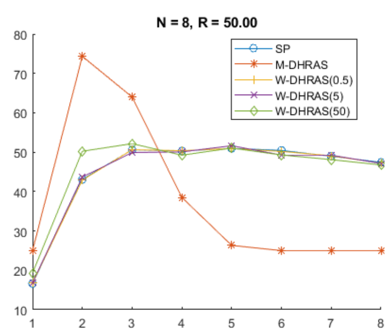


(c) $N = 10$

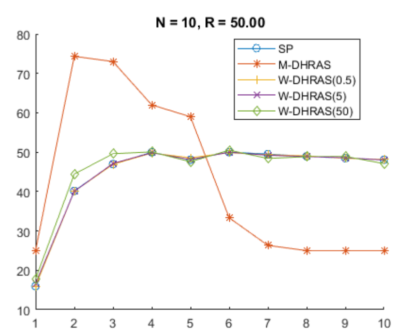
Figure L.15: Mean inter-arrival times with $R = 50$ under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ and $\lambda = 0.5$



(a) $N = 6$



(b) $N = 8$



(c) $N = 10$

Figure L.16: Mean inter-arrival times with $R = 50$ under $(c_j^{\mathrm{w}}, c_j^{\mathrm{u}}, c^{\mathrm{o}}) = (1, 5, 7.5)$ and $\lambda = 1$

$\delta \in \{0.1, 0.25, 0.5\}$ and (Row 3) Set 5 with $\delta \in \{0.1, 0.25, 0.5\}$. Figures M.17 and M.18 show the corresponding performance for $N = 8$ and $N = 10$ under cost structure (a). We only present the choice of $\epsilon \in \{0.05, 0.5, 5, 50\}$ such that W-DHRAS gives the best overall performance. In this case, we choose $\epsilon = 0.05$ and $\epsilon = 50$ for $N = 8$ and $N = 10$ respectively.

Similar patterns could be observed as in the case with $N = 6$. First, M-DHRAS model is the most conservative model, which yields the highest out-of-sample cost for almost all the cases. Second, we observe that W-DHRAS shares a similar out-of-sample performance as SP (mainly due to a small choice of ϵ) for $N = 8$ while for $N = 10$, the W-DHRAS model consistently produces a lower out-of-sample cost than the SP model. Indeed, for $N = 10$, the operator has a very tight schedule. A small deviation from the in-sample distribution (from the empirical data) results in a significant amount of overtime. Therefore, W-DHRAS could perform better with its ability to hedge against unfavorable scenarios.

Figures M.19 to M.21 show the results under cost structure (b) with $\lambda = 2$ for the three choices of N. We only present the W-DHRAS model with the best performing ϵ, which are 0.05, 0.5 and 50 respectively. Similar patterns are observed as in cost structure (a). First, M-DHRAS gives the largest out-of-sample costs for almost all the cases. Second, we observe that W-DHRAS and SP have similar performance with $N = 6$ and $N = 8$, though the performance of W-DHRAS is slightly better than SP when the sample size is small. Finally, for $N = 10$, W-DHRAS consistently performs the best among the three models.

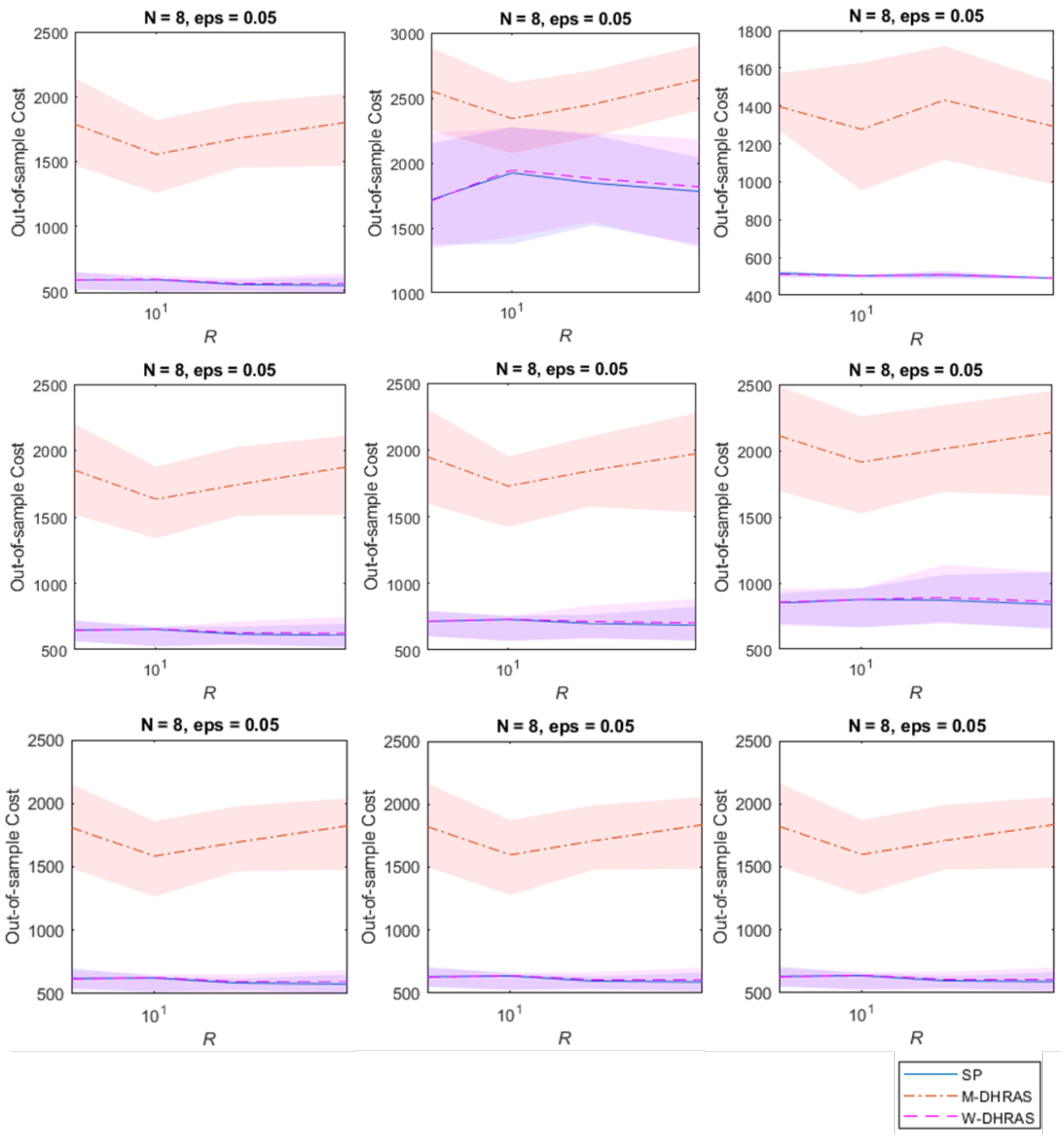


Figure M.17: Out-of-sample performance for $N = 8$ with cost structure (a) and $\lambda = 2$

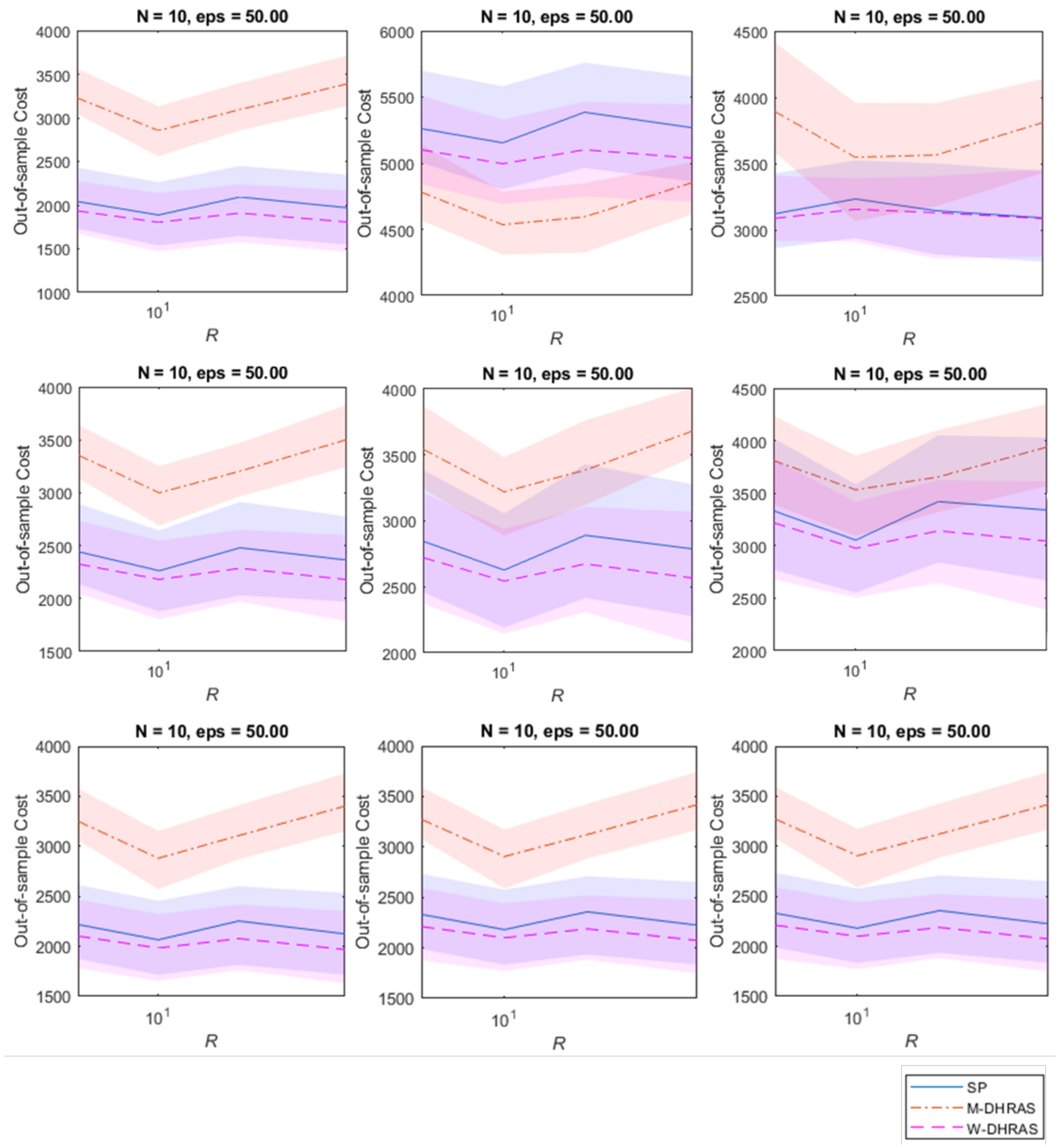


Figure M.18: Out-of-sample performance for $N = 10$ with cost structure (a) and $\lambda = 2$

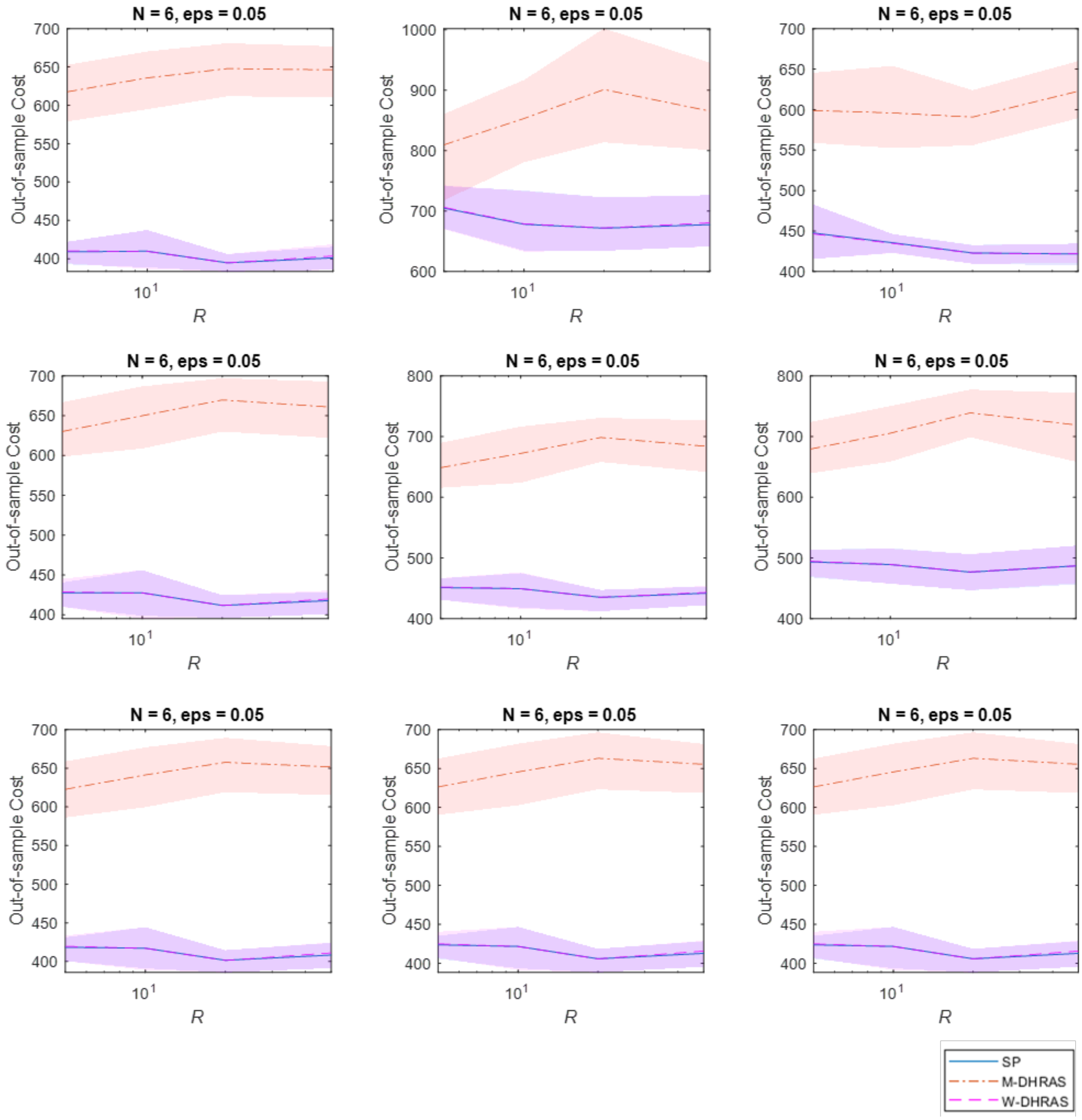


Figure M.19: Out-of-sample performance for $N = 6$ with cost structure (b) and $\lambda = 2$

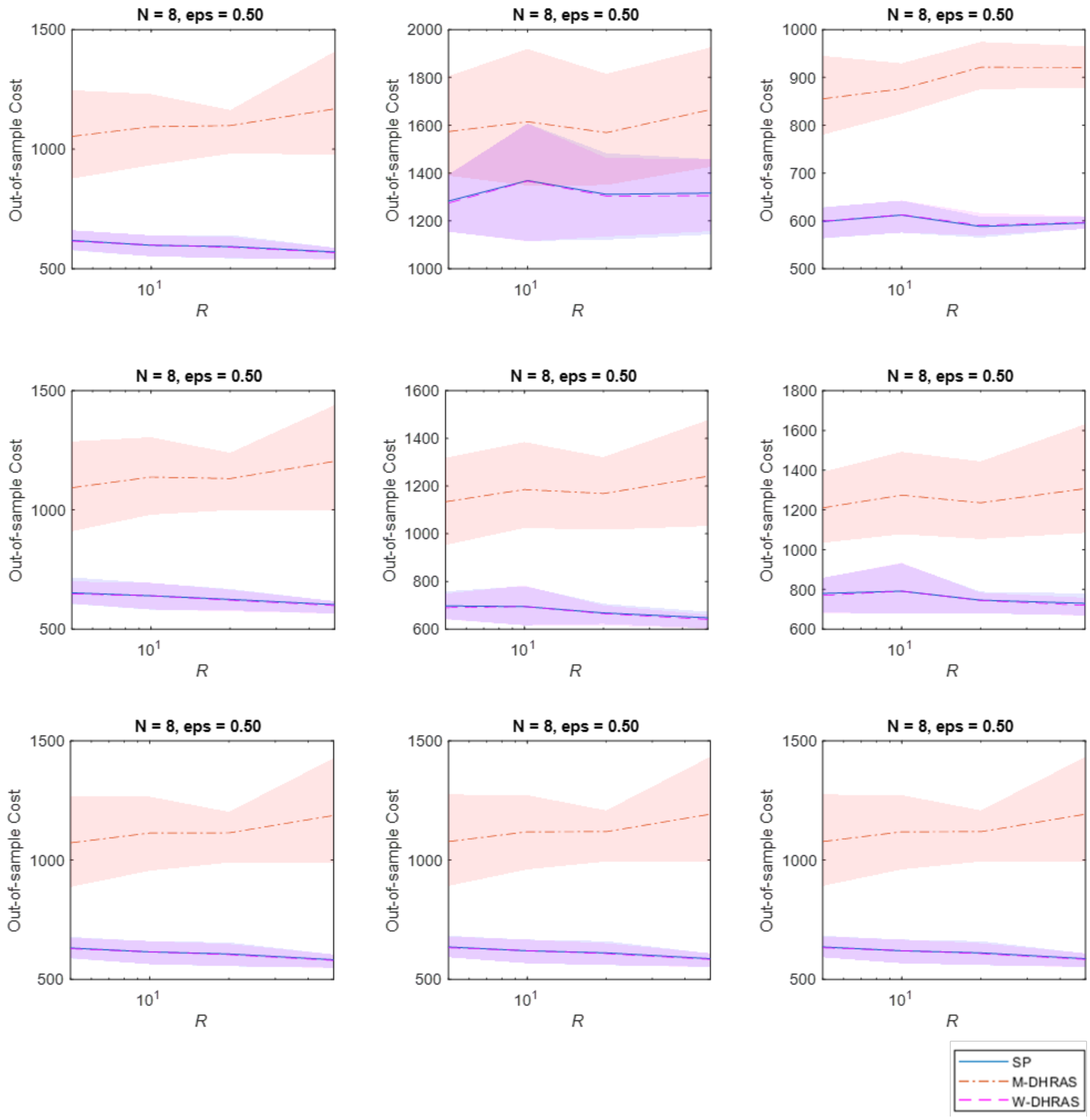


Figure M.20: Out-of-sample performance for $N = 8$ with cost structure (b) and $\lambda = 2$

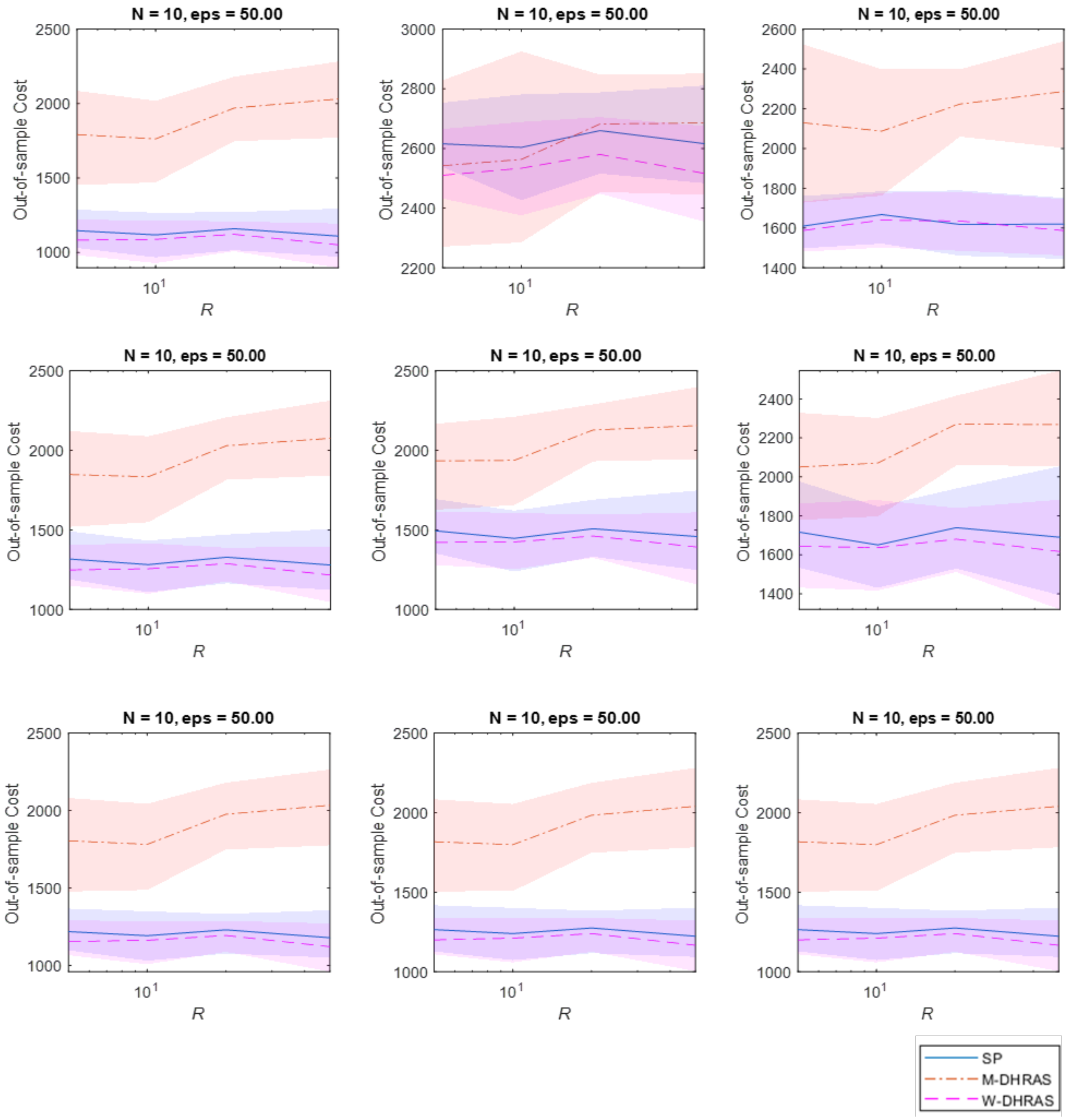


Figure M.21: Out-of-sample performance for $N = 10$ with cost structure (b) and $\lambda = 2$

References

Ahmadi-Javid, A., Jalali, Z., Klassen, K. J., 2017. Outpatient appointment systems in healthcare: A review of optimization studies. European Journal of Operational Research 258 (1), 3–34.

Artigues, C., Koné, O., Lopez, P., Mongeau, M., 2015. Mixed-integer linear programming formulations. In: Handbook on Project Management and Scheduling Vol. 1. Springer, pp. 17–41.

Ben-Tal, A., Den Hertog, D., Vial, J.-P., 2015. Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical Programming 149 (1), 265–299.

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.

Braekers, K., Hartl, R. F., Parragh, S. N., Tricoire, F., 2016. A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience. European Journal of Operational Research 248 (2), 428–443.

Chen, Z., Sim, M., Xiong, P., 2020. Robust stochastic optimization made easy with rsome. Management Science 66 (8), 3329–3339.

Cook, W. J., 2011. In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press.

Costa, L., Contardo, C., Desaulniers, G., 2019. Exact branch-price-and-cut algorithms for vehicle routing. Transportation Science 53 (4), 946–985.

Creemers, S., Lambrecht, M. R., Beliën, J., Van den Broeke, M., 2021. Evaluation of appointment scheduling rules: A multi-performance measurement approach. Omega 100, 102231.

Delage, E., Saif, A., 2018. The value of randomized solutions in mixed-integer distributionally robust optimization problems. GERAD HEC Montréal.

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

Demirbilek, M., Branke, J., Strauss, A., 2019. Dynamically accepting and scheduling patients for home healthcare. Health Care Management Science 22 (1), 140–155.

Duque, Daniel, M. S., Morton, P. D., 2020. Distributionally robust two-stage stochastic programming. Optimization Online E-print.

Eveborn, P., Flisberg, P., Rönnqvist, M., 2006. Laps care—an operational system for staff planning of home care. European Journal of Operational Research 171 (3), 962–976.

Fikar, C., Hirsch, P., 2017. Home health care routing and scheduling: A review. Computers & Operations Research 77, 86–95.

Fournier, N., Guillin, A., 2015. On the rate of convergence in Wasserstein distance of the empirical measure. Probability Theory and Related Fields 162 (3-4), 707–738.

Grieco, L., Utley, M., Crowe, S., 2020. Operational research applied to decisions in home health care: A systematic literature review. Journal of the Operational Research Society, 1–32.

Gutiérrez, E. V., Vidal, C. J., 2013. Home health care logistics management problems: A critical review of models and methods. Revista Facultad de Ingeniería Universidad de Antioquia (68), 160–175.

Hanasusanto, G. A., Kuhn, D., 2018. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. Operations Research 66 (3), 849–869.

Jiang, R., Ryu, M., Xu, G., 2019. Data-driven distributionally robust appointment scheduling over Wasserstein balls. arXiv preprint arXiv:1907.03219.

Jiang, R., Shen, S., Zhang, Y., 2017. Integer programming approaches for appointment scheduling with random no-shows and service durations. Operations research 65 (6), 1638–1656.

Kandakoglu, A., Sauré, A., Michalowski, W., Aquino, M., Graham, J., McCormick, B., 2020. A decision support system for home dialysis visit scheduling and nurse routing. Decision Support Systems 130, 113224.

Klabjan, D., Simchi-Levi, D., Song, M., 2013. Robust stochastic lot-sizing by means of histograms. Production and Operations Management 22 (3), 691–710.

Klotz, E., Newman, A. M., 2013. Practical guidelines for solving difficult mixed integer linear programs. Surveys in Operations Research and Management Science 18 (1-2), 18–32.

Kong, Q., Lee, C.-Y., Teo, C.-P., Zheng, Z., 2013. Scheduling arrivals to a stochastic service delivery system using copositive cones. Operations Research 61 (3), 711–726.

Kuiper, A., de Mast, J., Mandjes, M., 2021. The problem of appointment scheduling in outpatient clinics: A multiple case study of clinical practice. Omega 98, 102122.

Lanzarone, E., Matta, A., 2014. Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care. Operations Research for Health Care 3 (2), 48–58.

Lee, S., Kang, Y., Prabhu, V. V., 2013. Continuous variable control approach for home care crew scheduling. In: 2013 Winter Simulations Conference (WSC). IEEE, pp. 2262–2273.

Liu, R., Yuan, B., Jiang, Z., 2019. A branch-and-price algorithm for the home-caregiver scheduling and routing problem with stochastic travel and service times. Flexible Services and Manufacturing Journal 31 (4), 989–1011.

Lunden, I., apr 2019. Super raises $20m to fix the home services and repairs market with its subscription service.
URL https://techcrunch.com/2019/04/17/super-raises-20m-to-fix-the-home-services-and-repairs

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

Mankowska, D. S., Meisel, F., Bierwirth, C., 2014. The home health care routing and scheduling problem with interdependent services. Health Care Management Science 17 (1), 15–30.

Milburn, A. B., 2012. Operations research applications in home healthcare. In: Handbook of Healthcare System Scheduling. Springer, pp. 281–302.

Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming 171 (1-2), 115–166.

NAHC, 2010. Basic statistics about home care.
URL http://www.nahc.org/assets/1/7/10hc_stats.pdf.

Nikzad, E., Bashiri, M., Abbasi, B., 2021. A matheuristic algorithm for stochastic home health care planning. European Journal of Operational Research 288 (3), 753–774.

Oyola, J., Arntzen, H., Woodruff, D. L., 2018. The stochastic vehicle routing problem, a literature review, part i: models. EURO Journal on Transportation and Logistics 7 (3), 193–221.

Pflug, G. C., Pohl, M., 2018. A review on ambiguity in stochastic portfolio optimization. Set-Valued and Variational Analysis 26 (4), 733–757.

Pinedo, M. L., 2016. Scheduling: Theory, Algorithms, and Systems. Springer.

Rahimi, I., Gandomi, A. H., 2020. A comprehensive review and analysis of operating room and surgery scheduling. Archives of Computational Methods in Engineering, 1–22.

Rahimian, H., Mehrotra, S., 2019. Distributionally robust optimization: A review. arXiv preprint arXiv:1908.05659.

Restrepo, M. I., Rousseau, L.-M., Vallée, J., 2020. Home healthcare integrated staffing and scheduling. Omega 95, 102057.

Robinson, L. W., Chen, R. R., 2003. Scheduling doctors' appointments: optimal and empirically-based heuristic policies. Iie Transactions 35 (3), 295–307.

Saif, A., Delage, E., 2021. Data-driven distributionally robust capacitated facility location problem. European Journal of Operational Research 291 (3), 995–1007.

Shang, C., You, F., 2018. Distributionally robust optimization for planning and scheduling under uncertainty. Computers & Chemical Engineering 110, 53–68.

Shehadeh, K. S., Cohn, A. E., Jiang, R., 2020. A distributionally robust optimization approach for outpatient colonoscopy scheduling. European Journal of Operational Research 283 (2), 549–561.

Shi, Y., Boudouh, T., Grunder, O., 2019. A robust optimization for a home health care routing and scheduling problem with consideration of uncertain travel and service times. Transportation Research Part E: Logistics and Transportation Review 128, 52–95.

Shi, Y., Boudouh, T., Grunder, O., Wang, D., 2018. Modeling and solving simultaneous delivery and pick-up problem with stochastic travel and service times in home health care. Expert Systems with Applications 102, 218–233.

Smith, J. E., Winkler, R. L., 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Management Science 52 (3), 311–322.

Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research 21 (5), 1154–1157.

Thiele, A., 2010. A note on issues of over-conservatism in robust optimization with cost uncertainty. Optimization 59 (7), 1033–1040.

Wang, S., Chen, Z., Liu, T., 2020. Distributionally robust hub location. Transportation Science 54 (5), 1189–1210.

Wang, Y., Zhang, Y., Tang, J., 2019. A distributionally robust optimization approach for surgery block allocation. European Journal of Operational Research 273 (2), 740–753.

Yuan, B., Liu, R., Jiang, Z., 2015. A branch-and-price algorithm for the home health care scheduling and routing problem with stochastic service times and skill requirements. International Journal of Production Research 53 (24), 7450–7464.

Zhan, Y., Wan, G., 2018. Vehicle routing and appointment scheduling with team assignment for home services. Computers & Operations Research 100, 1–11.

Zhan, Y., Wang, Z., Wan, G., 2021. Home service routing and appointment scheduling with stochastic service times. European Journal of Operational Research 288 (1), 98–110.

Zhang, Y., Shen, S., Erdogan, S. A., 2017. Distributionally robust appointment scheduling with moment-based ambiguity set. Operations Research Letters 45 (2), 139–144.

Zhou, S., Li, D., Yin, Y., 2021. Coordinated appointment scheduling with multiple providers and patient-and-physician matching cost in specialty care. Omega 101, 102285.